Modern Integration

Brent Baccala

Contents

Preface

This book grew out of an abortive class in Risch Integration that I taught at University of Maryland at College Park in the spring of 2006,[1] which I canceled after three weeks when I had no students left. Aside from the lack of student interest (it was a non-credit class), another deficiency in the class became apparent to me — the lack of a good textbook. So I am writing this book to fill this perceived gap, the need for a senior level undergraduate text on differential algebra, developing the subject so far as the solution of the problem of integration in finite terms (the integration problem), the theory's most famous application to date.

Why, first of all, should math students study this subject, and why near the end of an undergraduate mathematics program?

First and foremost, for pedagogical reasons. Almost all modern college math curricula include higher algebra, yet this subject seems to be taught in a very abstract context. The integration problem puts this abstraction into concrete form. We have a specific goal in mind — the development of an algorithm that, given an integral constructed from elementary functions, either solves that integral by expressing it using elementary functions, or else proves that no such expression is possible. One of the best ways to learn a subject, or at least to convince yourself that you understand it, is to apply it in a specific and concrete way. The greatest difficulties I have encountered in math is when faced with abstract concepts lacking concrete examples. Such, in my mind, is the primary goal of studying differential algebra near the end of an undergraduate program. The student has no doubt been exposed to higher algebra, now we want to make sure we understand it by taking all those rings, fields, ideals, extensions and what not and applying them to this specific goal.

Secondly, there is a sense of both historical and educational completion to be obtained here. Not only has the integration problem challenged mathematicians since the development of the calculus, but there is a real danger of getting through an entire calculus sequence and be left thinking that if you really want to solve an integral, the best way is to use a computer! Due to the intricacy of the calculations involved, the best way probably is to use a computer, but the student is left at a vague but quite definite disadvantage without the understanding that the integration problem has been solved and without some familiarity with the techniques used to solve it.

Third, an introduction to differential algebra may be quite appropriate at a point where

[1] I am not a professor at UMCP, and am not affiliated with the University of Maryland in any way other than having studied physics there as an undergraduate and being a member of the University Alumni Association.

students are starting to think about research interests. Though this field has profitably engaged the attentions of a number of late twentieth century mathematicians, it is still a young field that may turn out to be a major breakthrough in the solution of differential equations. It may also turn out to be a dead end ("interesting but not compelling" in the words of one commentator), which I why I hesitate to list this reason first on my list. The big question, in my mind, is whether this theory can be suitably extended to handle partial differential equations, as both integrals and ordinary differential equations can now be adequately handled using numerical techniques. This question remains unanswered at this time.

Finally, I have a strong personal motivation in writing this book. I am not an expert in this field, really a student myself at this point. Another very good way to learn a subject, or at least to convince two people that you understand it, is to explain it to somebody else.

Since the available material on this subject is too sparsely spread around among a variety of texts and research papers, I decided for all of these reasons to compile, more so than write, a book targeted at an undergraduate audience with some exposure to higher algebra. However, in keeping with my primarily pedagogical aims, I re-introduce all the key concepts of algebra as they are needed. This serves both to refresh and reinforce concepts already learned and also to act a convenient reference without having to flip constantly back and forth between books. This book should not be taken as a substitute for a broader theory text, as I introduce only the concepts needed for my particular application, and only at a level of detail that seems appropriate.

Since the book is still a work in progress, I can't hope to properly conclude this preface at this time. I would, however, like to specifically thank Dr. Denny Gulick, Undergradate Chair of the UMCP Mathematics Department, for giving me the opportunity to teach the class which lead directly to this book.

Chapter 1

Introduction

Who Wants to be a Mathematician?

$50,000 Question

Which of the following integrals can *not* be expressed as an elementary function?

A. $\int \sin x \ \mathrm{d}x$

B. $\int e^{-x^2} \, \mathrm{d}x$

C. $\int \frac{x\{(x^2e^{2x^2}-\ln^2(x+1))^2+2xe^{3x^2}(x-(2x^3+2x^2+x+1)\ln(x+1))\}}{(x+1)(\ln^2(x+1)-x^2e^{2x^2})^2} \, \mathrm{d}x$

D. $\int \frac{2x^6+4x^5+7x^4-3x^3-x^2-8x-8}{(2x^2-1)^2\sqrt{x^4+4x^3+2x^2+1}} \, \mathrm{d}x$

The answer to this "$50,000" question[1] is, somewhat surprisingly, (B). Simplifying (A) as $\int \sin \, \mathrm{d}x = -\cos x + C$ is an easy exercise from a first year calculus course. (C) and (D), while appearing more formidable, are solvable using the techniques of this book.

(C) is example 5.6, and can be written as:

$$\int \frac{x\{(x^2e^{2x^2} - \ln^2(x+1))^2 + 2xe^{3x^2}(x - (2x^3+2x^2+x+1)\ln(x+1))\}}{(x+1)(\ln^2(x+1) - x^2e^{2x^2})^2} dx$$

$$= x - \ln(x+1) - \frac{xe^{x^2}\ln(x+1)}{\ln^2(x+1) - x^2e^{2x^2}} + \frac{1}{2}\ln\frac{\ln(x+1) + xe^{x^2}}{\ln(x+1) - xe^{x^2}}$$

(D) is example 9.30:

$$A(x) = 1023x^8 + 4104x^7 + 5048x^6 + 2182x^5 + 805x^4 + 624x^3 + 10x^2 + 28x$$

$$B(x) = 1025x^{10}+6138x^9+12307x^8+10188x^7+4503x^6+3134x^5+1598x^4+140x^3+176x^2+2$$

$$C(x) = 32x^{10} - 80x^8 + 80x^6 - 40x^4 + 10x^2 - 1$$

$$\int \frac{(2x^6+4x^5+7x^4-3x^3-x^2-8x-8)y}{(2x^2-1)^2(x^4+4x^3+2x^2+1)}\,\mathrm{d}x = \frac{(x+\frac{1}{2})y}{2x^2-1} + \frac{1}{2}\ln\frac{A(x)y - B(x)}{C(x)}$$

Integral (B), on the other hand, can not be "solved" in this manner, and example 6.5 proves this claim of impossibility.

What does it mean to "solve" an integral?

Is there a formal procedure, an algorithm, that lets us solve any integral, or prove that such a solution is impossible?

These questions have puzzled mathematicians for over 300 years, since the invention of calculus, so much so that an introductory calculus sequence can start to seem like a series of puzzle problems, each chapter harder than the last.

This book aims to present our most sophisticated integration theory that provides definitive answers to these questions, but the existance of integrals like $\int e^{-x^2}\,\mathrm{d}x$ without any elementary form shows that any such theory has severe limitations. Futhermore, the development of the electronic computer, coupled with sophisticated numerical integration techniques, has provided us with powerful approximation methods that significantly reduce the importance of solving integrals. Nevertheless, more difficult differential equations continue to elude easy analysis, so perhaps the greatest benefit of studying integration is the insight it provides to solving differential equations in general.

[1] The author does not actually possess a $50,000 prize fund.

1.1 Calculus

Let's consider again the integral $\int e^{-x^2} dx$. We can *trivially* construct an anti-derivative as follows:

$$E(x) = \int_0^x e^{-t^2} dt$$

I claim that $E(x)$ is an anti-derivative of e^{-x^2}. Let's see...

First, is $E(x)$ well defined? Let's recall some material from a standard introductory calculus textbook, say, [BrCo10]:

> [BrCo10] Definition - Definite Integral (p. 324)
>
> A function f defined on $[a, b]$ is **integrable** on $[a, b]$ if $\lim_{\Delta \to 0} \sum_{k=1}^{n} f(\bar{x}_k)\Delta x_k$ exists and is unique over all partitions of $[a, b]$ and all choices of $\bar{x}_k$ on a partition. This limit is the **definite integral of f from a to b**, which we write
>
> $$\int_a^b f(x)dx = \lim_{\Delta \to 0} \sum_{k=1}^{n} f(\bar{x}_k)\Delta x_k$$
>
> .

> [BrCo10] Theorem 5.2 - Integrable Functions (p. 325)
>
> If f is continuous on $[a, b]$ or bounded on $[a, b]$ with a finite number of discontinuities, then f is integrable on $[a, b]$.

So, $E(x) = \int_0^x e^{-t^2} dt$ is *integrable* on $[0, x]$ if e^{-t^2} is continuous on $[0, x]$, and e^{-t^2} is continuous everywhere on the real line. We can easily plot e^{-t^2}:

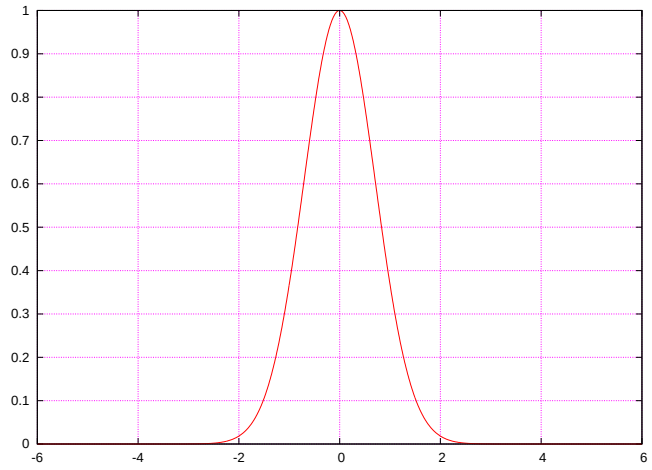


Figure 1.1: e^{-t^2}

It's obviously continuous, so Theorem 5.2 tells us that $E(x)$ is well defined for any real number x – the limit used to construct the Riemann sum exists and is unique. We can also plot $E(x)$, using a numerical integration routine to approximate the integral at each point of the graph:

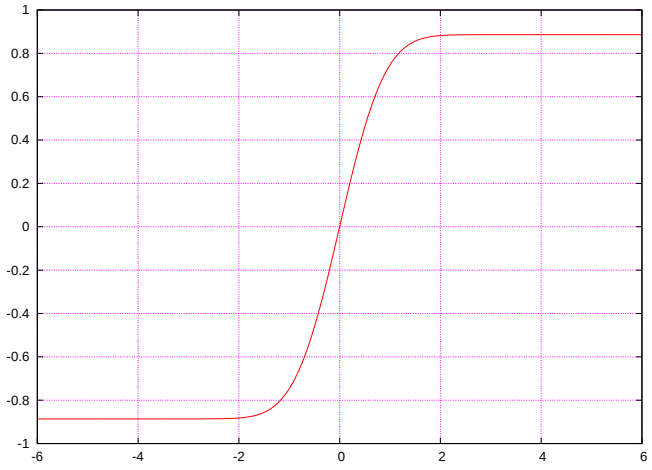


Figure 1.2: $\int_0^x e^{-t^2} dt$

We're plotting the *integral* now... the height of each point on the graph was calculated by numerically approximating a Riemann sum.

Is this an anti-derivative of e^{-t^2}? Plotting various tangent lines suggests that it *might* be...

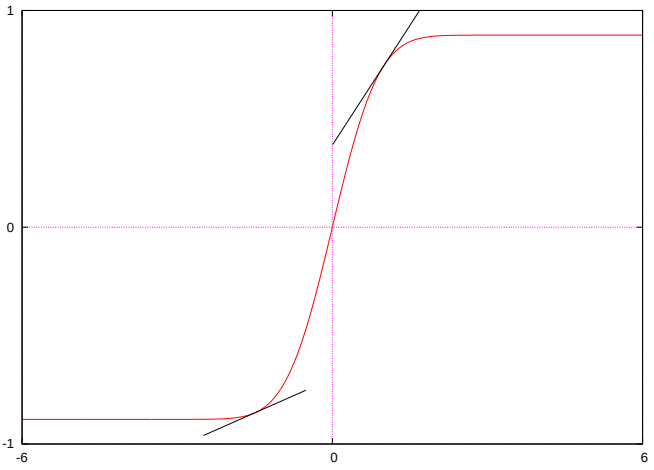


Figure 1.3: $\int_0^x e^{-t^2}\,dt$ (with tangent lines at $x_0 = -1.5$ and $x_0 = 1$)

The tangent lines in the graph were plotted using this formula:

$$y(x) = E(x_0) + e^{-x_0^2}(x - x_0)$$

i.e, the point-slope equation of a straight line, with point $(x_0, E(x_0))$ and slope $e^{-x_0^2}$.

See, I used the $E(x) = \int_0^x e^{-t^2}\,dt$ formula for the y-coordinate of the point, and the e^{-x^2} formula for the slope. If $E(x)$ is an anti-derivative of e^{-x^2}, then the derivative of $E(x)$ is e^{-x^2}, and the formula will produce tangent lines for any value of x_0. If $E(x)$ were *not* an anti-derivative of e^{-x^2}, we'd get lines, but they wouldn't be tangent lines.

Varying the value of x_0 produces different lines (the two lines in the graph were generated using $x_0 = -1.5$ and $x_0 = 1$), and they *appear* to be tangent lines, so perhaps $E(x)$ is an anti-derivative of e^{-x^2}.

In fact, we can do far better than guess. Remember the Fundamental Theorem of Calculus?

[BrCo10] Theorem 5.3 (part 1) - Fundemental Theorem of Calculus (p. 338)

If f is continuous on $[a, b]$, then the area function

$$A(x) = \int_a^x f(t)\,dt \qquad \text{for} \quad a \le x \le b$$

is continuous on $[a, b]$ and differentiable on (a, b). The area function satisfies $A'(x) =$

$f(x)$; or, equivalently,

$$A'(x) = \frac{d}{dx}\int_a^x f(t)\,dt = f(x),$$

which means that the area function of f is an antiderivative of f.

Pay particular attention to that last formula – it says that the derivative of an integral with respect to its upper bound of integration is just the integrand, with the name of the variable changed.

So, $E(x)$, defined like this:

$$E(x) = \int_0^x e^{-t^2} dt$$

is *trivially* an anti-derivative of e^{-x^2}, because the Fundamental Theorem of Calculus tells us that:

$$E'(x) = \frac{d}{dx}\int_0^x e^{-t^2} dt = e^{-x^2}$$

[BrCo10] Theorem 5.2 tells us that $E(x)$ *exists* (because e^{-x^2} is continuous), and [BrCo10] Theorem 5.3 tells us that $E(x)$ is an anti-derivative of e^{-x^2}.

Of course, we had something else in mind when we asked for an anti-derivative of e^{-x^2}. We wanted a simplified form, something like this:

$$\int x^2\,dx = \frac{1}{3}x^3 + C$$

not some mathematical smart aleck telling us that the answer is $\int x^2\,dx$!

The problem is that $\int e^{-x^2} dx$ doesn't have a simplified form. It has an anti-derivative (we plotted it, remember?), and it's completely well-defined as a mathematical function, but we can't simplify it in the way that we can simplify $\int x^2\,dx$.

Another example is $\int \frac{\sin x}{x} dx$. It's also continuous everywhere. The only point where that's at all in question is $x = 0$, but L'Hospital's Rule[2] tells us that:

[2]Using L'Hospital's Rule here is actually a circular argument, because we had to evaluate this limit to prove that sine's derivative is cosine.

$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = \frac{\cos 0}{1} = 1$$

which means that the division by zero in $\frac{\sin x}{x}$ is a *removable discontinuity*. We can patch up our function like this:

$$f(x) = \begin{cases} \frac{\sin x}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

It's usually called the *sinc* function, and it's easy to plot:

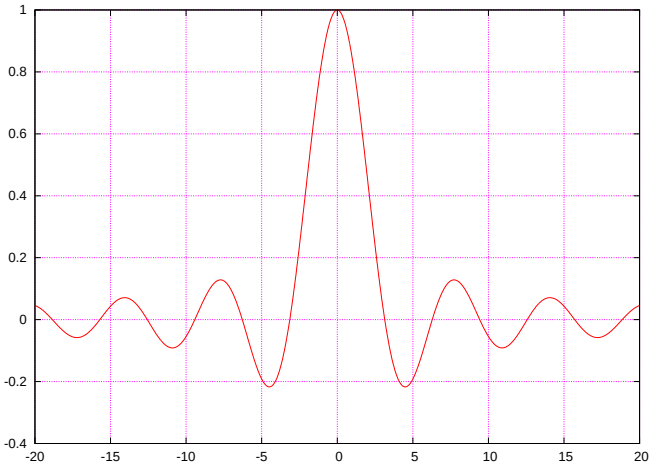


Figure 1.4: $\operatorname{sinc} t = \frac{\sin t}{t}$

Since sinc is continuous everywhere, this integral is well defined everywhere:

$$\operatorname{Si}(x) = \int_0^x \frac{\sin(t)}{t} dt$$

...and we can plot it...

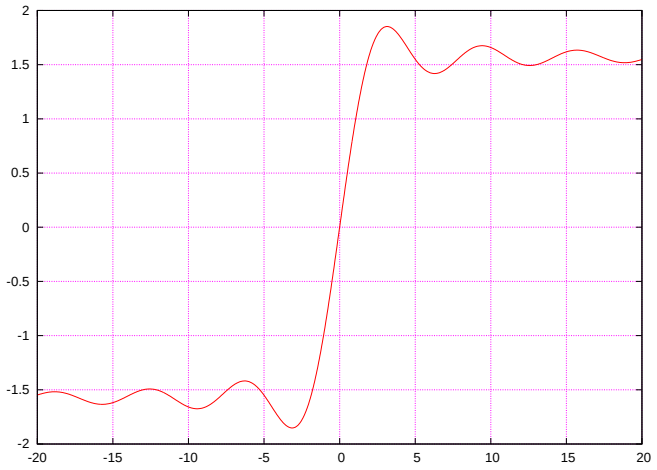


Figure 1.5: $\int_0^x \frac{\sin t}{t} dt$

...and we can check some of its tangent lines, using the formula:

$$y(x) = \mathrm{Si}(x_0) + \frac{\sin x_0}{x_0}(x - x_0)$$

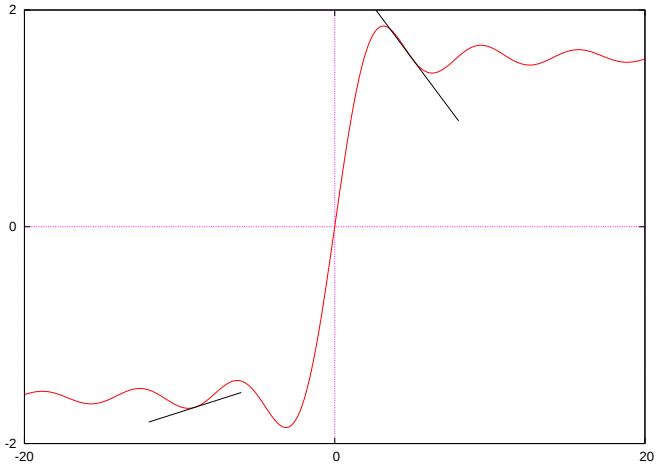


Figure 1.6: $\int_0^x \frac{\sin t}{t} dt$ (with tangent lines at $x_0 = -9$ and $x_0 = 5$)

Again, it's *trivial* that $\mathrm{Si}(x)$:

1. **exists**, by [BrCo10] Theorem 5.2 and the continuity of $\frac{\sin x}{x}$, and

2. is an **anti-derivative** of $\frac{\sin x}{x}$, by [BrCo10] Theorem 5.3 and the definition of $\mathrm{Si}(x)$:

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t} dt \qquad \Longrightarrow \qquad \mathrm{Si}'(x) = \frac{d}{dx} \int_0^x \frac{\sin(t)}{t} dt = \frac{\sin(x)}{x}$$

Yet, again, we have no simple closed form for $\mathrm{Si}(x)$.

Let's see... how could we find simple expressions for $\int e^{-x^2} dx$ and $\int \frac{\sin x}{x} dx$?

Could we try...

1. Integration by Parts
2. Trigonometric Substitution
3. Partial Fractions
4. ...some clever change of variables...
5. Google

How about this instead – let's prove that these two integrals have no simple forms.

1.2 Algebra

In high school, we study what the Arabs called “al-jabr”, or what the Encyclopaedia Britanncia calls “a generalization and extension of arithmetic”. “Elementary algebra," the encyclopedia goes on, “is concerned with properties of arbitrary numbers,” and cites the commutative law of addition $(a+b = b+a)$ as an example of such a property. We use only a few others: the commutative law of multiplication; associative laws of both addition and multiplication; the distributive law. The key point is that all of these laws are valid for any numbers whatsoever, so we are justified in applying them to unknown numbers.

In addition to these basic laws, there is a language to be learned, as well as the more general Principle of Equality: given two identical quantities, the same operation applied to both must given identical results. This hold true no matter what the operation is, so long as it is deterministic (i.e, has no randomness). Thus, combining the Principle of Equality with the commutative law of addition, I can conclude that $\sin(a+b) = \sin(b+a)$, without any additional knowledge of what “sin” might be.

For example, consider the following sequence:

$$
\begin{aligned}
(ax+\tfrac{b}{2})^2 &= (ax+\tfrac{b}{2})(ax+\tfrac{b}{2}) && \text{definition of square}\\
&= ax(ax+\tfrac{b}{2}) + \tfrac{b}{2}(ax+\tfrac{b}{2}) && \text{distributive law}\\
&= axax + ax\tfrac{b}{2} + \tfrac{b}{2}(ax+\tfrac{b}{2}) && \text{distributive law}\\
&= axax + ax\tfrac{b}{2} + \tfrac{b}{2}ax + \tfrac{b}{2}\tfrac{b}{2} && \text{distributive law}\\
&= aaxx + \tfrac{1}{2}abx + \tfrac{1}{2}abx + \tfrac{b}{2}\tfrac{b}{2} && \text{commutative law of multiplication (3 times)}\\
&= a^2x^2 + \tfrac{1}{2}abx + \tfrac{1}{2}abx + \tfrac{b^2}{4} && \text{definition of square}\\
&= a^2x^2 + (\tfrac{1}{2}+\tfrac{1}{2})abx + \tfrac{b^2}{4} && \text{distributive law}\\
&= a^2x^2 + abx + \tfrac{b^2}{4} && \text{basic arithmetic}\\
(ax+\tfrac{b}{2})^2 - \tfrac{b^2}{4} + ac &= a^2x^2 + abx + \tfrac{b^2}{4} - \tfrac{b^2}{4} + ac && \text{principle of equality}\\
(ax+\tfrac{b}{2})^2 - \tfrac{b^2}{4} + ac &= a^2x^2 + abx + ac && \text{definition of subtraction}
\end{aligned}
$$

So, if $ax^2 + bx + c = 0$, then

$$
\begin{aligned}
ax^2 + bx + c &= 0 && \\
a(ax^2 + bx + c) &= 0a && \text{principle of equality} \\
a(ax^2 + bx + c) &= 0 && \text{zero theorem}[3] \\
a^2x^2 + abx + ac &= 0 && \text{distributive law} \\
(ax + \tfrac{b}{2})^2 - \tfrac{b^2}{4} + ac &= 0 && \text{principle of equality}[4] \\
(ax + \tfrac{b}{2})^2 - \tfrac{b^2}{4} + ac + \tfrac{b^2}{4} - ac &= \tfrac{b^2}{4} - ac && \text{principle of equality} \\
(ax + \tfrac{b}{2})^2 &= \tfrac{b^2}{4} - ac && \text{definition of subtraction} \\
4(ax + \tfrac{b}{2})^2 &= 4\tfrac{b^2}{4} - 4ac && \text{principle of equality} \\
4(ax + \tfrac{b}{2})^2 &= b^2 - 4ac && \text{definition of division} \\
2^2(ax + \tfrac{b}{2})^2 &= b^2 - 4ac && \text{definition of square} \\
(2(ax + \tfrac{b}{2}))^2 &= b^2 - 4ac && \text{commutative law of multiplication}[5] \\
(2ax + 2\tfrac{b}{2})^2 &= b^2 - 4ac && \text{distributive law} \\
(2ax + b)^2 &= b^2 - 4ac && \text{definition of division} \\
\sqrt{(2ax + b)^2} &= \sqrt{b^2 - 4ac} && \text{principle of equality} \\
(2ax + b) &= \sqrt{b^2 - 4ac} && \text{!?!?!??!} \\
(2ax + b) - b &= \sqrt{b^2 - 4ac} - b && \text{principle of equality} \\
2ax &= \sqrt{b^2 - 4ac} - b && \text{definition of subtraction} \\
\tfrac{1}{2a}2ax &= \tfrac{1}{2a}(\sqrt{b^2 - 4ac} - b) && \text{principle of equality} \\
x &= \tfrac{1}{2a}(\sqrt{b^2 - 4ac} - b) && \text{definition of division}
\end{aligned}
$$

At each step in the sequence (except one), we're just applying one of the basic rules above. The problem with the "mystery step" isn't so much that we're taking the square root, since the principle of equality tells us that we can perform the same operation on both sides of the equal sign, but rather that it cancels out the square in some undefined way. So, assuming that we can perform the mystery step, and noting that the division in the next to last step is only defined if $a \neq 0$, we can legitimately conclude that the final result is true for any a, b, and c whatsoever.

The mystery step leads us to introduce complex numbers, typically when we want to use this equation to solve polynomials such as $x^2 + 1 = 0$. At this point, the alert student, having been lured in to a false sense of security by the encyclopedia's "numbers", and now finding himself facing a whole new type of number entirely, can rightly ask, "What is a number?"

To which we wave our hands and reply, "It's, you know, a number!" I am reminded of the time that I was asked to sub in a seventh grade pre-algebra class, and was promptly asked by one of the students to explain the difference between "3" and "2.9999999…" I think I mumbled something lame like "I don't know, what do you think?" I certainly hadn't come to class prepared to discuss Cauchy sequences!

In college we are no longer satisfied with this answer, and here is really the launching point for "higher" algebra. Our "numbers" become objects in a set, and our simple concepts of addition and multiplication morph into operations which map pairs of objects into other objects. When asked, "What is a number?", we now confidently reply, "Anything whose operations obey the axioms!", which really isn't all that surprising an answer (anymore) because our entire theory had been built around those axioms to begin with.

The program of higher algebra (in fact much of modern mathematics) goes thus. We postulate the existance of one or more sets of objects and one or more operations, which are simply mappings defined on the objects of those sets. We write out a list of axioms that we assume those sets and operations obey. Which axioms are those? Whichever we find useful (or at least interesting). Then we develop as little or much of a theory as we can, reasoning always from the base axioms. Finally, we take some specific set of objects (like the integers), demonstrate that they obey our set of axioms, and conclude that the entire theory developed for those axioms must apply, therefore, to the integers. Sometimes we reverse the process by finding axioms obeyed by some specific set of objects we wish to study, then developing a theory around them.[6]

The most important (i.e, repeatedly used) sets of axioms are given names, or more precisely the sets and operators which obey them are given names. Thus, a "group" is any set and operator which obey three or four certain axioms. A "ring" is any set and pair of operators which obey about six axioms. Add another axiom or two and it becomes a "field". If a different axiom is obeyed, it is a "Noetherian ring".

It's easy to get bogged down with terminology, especially in a classroom environment where you can't raise your hand during a test and ask, "Excuse me, what's a semigroup again?" Far more important, I think, is to grasp the central idea that any of these terms refers simultaneously to three things: a set of axioms, a theory logically developed from those axioms, and any particular object(s) that obeys those axioms, and therefore the theory. The ultimate goal is to develop far more sophisticated theories than are possible using the "numbers" of elementary algebra.

Our goal in this book is the development of an algebraic system that allows us to represent as a single object any expression written using elementary functions, putting $\sqrt{1+\sin x}$ on par with $\frac{3}{2}$, introducing the concept of a derivative so that we can write differential equations using these objects (it now becomes *differential* algebra), and equipping this system with a theory powerful enough to either integrate anything so expressed, or prove that it can't be done, at least not using elementary functions. This is how computer programs like *Mathematica* or *AXIOM* solve "impossible" integrals. Along the way, we will

[6] How do we demonstrate that a certain set obeys certain axioms? By using more axioms, of course! Mathematics is probably the most self-contained of all major academic fields of study. Many other fields use its results, but math itself references nothing. It's impossible to get started without assuming *something*, so the entire process becomes a bit of a chicken-and-egg operation, which leads you to wonder… which *did* come first?

have cause to at least survey some of the deepest waters of modern mathematics. Differential algebra is very much a 20^{th} century theory — the integration problem was not solved until roughly 1970; a really workable algorithm for the toughest cases wasn't available until 1990; a key sub-problem (testing the equivalence of constants) remains unsolved still. Yet one thing is for sure. Three hundred years after the development of calculus, one of its most basic and elusive problems has finally yielded not to limits, sums, and series, but to rings, fields and polynomials. Quite a triumph for "al-jabr".

1.3 Maxima

Many of the more complex examples in the book will be solved using the open source computer algebra system Maxima.

In the section, I'll collect several useful functions that I'll use throughout the book.

First, TeX has two different ways of displaying the Greek letter θ, and Maxima lets us select which we will use. I prefer θ.

```
(%i1)   theta;

(%o1)
                                 ϑ


(%i2)   :lisp (defprop $theta "\\theta" texword)

(%i2)   theta;

(%o2)
                                 θ
```

I use lambda expressions with a single variable so much that I find it useful to create a version of `map` that treats its first argument as a lambda expression in u.

```
(%i3)   mapu(func, expr) ::=
           map(buildq([func], lambda([u], func)),
               ev(expr))$

(%i4)   mapu(u+4, [1,2,3]);

(%o4)
                              [5, 6, 7]


(%i5)   mapu(denom(u), [x,1,1/x,1/x^2]);

(%o5)
                             [1, 1, x, x^2]
```

1.3.1 List Functions

Here's a function that finds the index of an element in a list:

```
(%i6)   which(n,u) :=
           sublist_indices(n, lambda([x], x=u))[1]$


(%i7)   which([a,b,c,d,e], b);

(%o7)
                                2
```

This function converts the parts of a mathematical expression into a list:

```
(%i8)   partlist(expr) := block(
           [partswitch: true, result:[]],
           for i:1 step 1 unless part(expr,i) = end
              do result:cons(part(expr,i), result),
           result)$

(%i9)   partlist(a+b+c);

(%o9)
                              [a, b, c]
```

1.3.2 Array Functions

This next function is convenient for displaying Maxima arrays:

```
(%i10)  displayarray(b) :=
           map(display,
              map(lambda([u], arraymake(b,u)),
                  rest(arrayinfo(b),2)))$
```

1.3.3 Simplification

The Maxima function `ratsimp` simplifies a rational function to its simplest form, and I use it so much, let's define a shorthand notation to simplify an expression and assign it to

a variable:

```
(%i11)  infix(":::", 180, 20)$

(%i12)  (x ::: y) := x :: ratsimp(y)$

(%i13)  (x+1)/(x^2-1) - 1/(x-1);

(%o13)
```

$$\frac{1+x}{x^2-1}-\frac{1}{x-1}$$

```
(%i14)  a ::: (x+1)/(x^2-1) - 1/(x-1);

(%o14)
```

$$0$$

```
(%i15)  infix("===", 20, 20)$

(%i16)  (x === y) := is(ratsimp(x - y) = 0)$
```

Chapter 2

Commutative Algebra

In this chapter I will outline the basic algebraic structures necessary to carry out the program sketched out in the previous chapter. This material is included mainly to provide a starting point for the rest of the book. The pace of this chapter is deliberately quick; I omit a lot of the more basic proofs and doubt that this will substitute for a good introductory text on higher algebra.[1]

2.1 Rings and Fields

[van der Waerden], §3.1

We begin with two key definitions that we will use throughout: the *ring* and the *field*. As I explained in the previous chapter, both are associated primarily with sets of axioms. Any algebraic system that obeys the ring axioms is called a ring; any algebraic system that obeys the field axioms is a field.

Both rings and fields are defined over sets with two binary operators, conventionally called addition and multiplication. It will appease the nervous reader to know that for our purposes, addition is addition and multiplication is multiplication — the same addition and multiplication we learned in grade school. Of course, in the general case, any pair of operations that obey the axioms will suffice to form a ring or a field, but we won't need to concern ourselves with this.

I'm going to use basic set-theoretic notation to define the axioms. Read $\forall$ as "for all" and $\exists$ as "there exists". Each symbol is immediately followed by the new variable or a list of new variables that it qualifies. Their ordering is significant. The rule is that each variable is assigned in left-to-right order. So, "$\exists a, \forall b$" means "there exists an a (independent of b, because b hasn't appeared yet) so that for all $b\ldots$", while "$\forall b, \exists a$" means "for all b, there exists an a (possibly different for each b) so that…". In the later case, a can be a function of b, but not in the first case. Sometimes I'll add the set inclusion symbol $\in$, read "in",

[1] Need a good reference to such a text here. I want to footnote all the missing proofs with references back to two or three texts; maybe Lang, van der Waerden, and an introductory undergraduate text of some kind.

such as "$\forall a, b, c \in \mathcal{R}$...", reading "for all a, b, and c that are members of $\mathcal{R}$...", but I'll omit this from the next table for simplicity, since everything is a member of $\mathcal{R}$.

A *commutative ring with unity* $\mathcal{R}$ obeys the following axioms:

associative law of addition	$\forall a, b, c, (a + b) + c = a + (b + c)$	(R1)
associative law of multiplication	$\forall a, b, c, (ab)c = a(bc)$	(R2)
commutative law of addition	$\forall a, b, \quad a + b = b + a$	(R3)
distributive law	$\forall a, b, c, a(b + c) = ab + bc$	(R4)
existence of an additive identity (zero)	$\exists 0, \forall a, \, 0 + a = a$	(R5)
invertibility of addition	$\forall a, \exists b, \, a + b = 0$	(R6)
commutative law of multiplication	$\forall a, b, \quad ab = ba$	(CR1)
existence of a multiplicative identity (unity)	$\exists 1, \forall a, \, 1a = a$	(RwU1)

You notice the commutative law of multiplication at the end as (CR1), along with the existance of a unity element 1, labeled (RwU1). A substantial theory has been developed around *non-commutative* rings, probably because matrix multiplication (a critically important example) is non-commutative. Most of our rings are commutative ring, or *abelian* (the terms are synonymous). Also, I required the existence of a multiplicative identity. Much of ring theory can be developed without this axion, but some theorems require it, and I don't want to belabor the point, since all of our rings will have a unity element. Therefore, I will omit any additional terminology, adopt the CR1 and RwU1 axioms along with ring axioms R1-R6 to obtain a *commutative ring with unity*, and call it informally a "ring" for the rest of the book.

A *field* $\mathcal{F}$ obeys all the ring axioms (thus all fields are also rings), as well as the following axiom:

$$\text{invertibility of multiplication} \quad \forall a \neq 0, \exists b, ab = 1 \quad \text{(F1)}$$

In plain English, rings are mathematical systems in which addition and multiplication are cleanly defined. Subtraction is also defined due to (R6), the invertibility of addition, which allows a subtraction problem to be turned into an addition problem. Division, however, is not, since it requires (F1). Since the ring axioms do not require multiplication to be invertible, there is no guarantee that we can carry out division in a ring. A field, on the other hand, is a mathematical system in which all four elementary operations — addition, subtraction, multiplication, and division — are defined.

The simplest example of a ring is the set of integers, which I shall denote as $\mathbf{Z}$ (after the German word for number, *zahl*). A pair of integers can be added, subtracted or multiplied to form another integer. Note, however, that a pair of integers can not necessarily be divided to form another integer. $\frac{3}{2}$ is not an integer, because multiplication (in $\mathbf{Z}$) is not necessarily invertible — there is no integer that when multiplied by 2 forms 1. (F1) is not satisfied. Thus $\mathbf{Z}$ forms a ring but not a field.

2.2 Quotient fields

The field Q

[van der Waerden], §3.3

We can remedy our inability to divide using only integers by moving on the *rational numbers*, traditionally denoted **Q** (probably for *quotient*). This is the simplest example of a *quotient field*, in this case formed over the integers, **Z**. It is also our first example of a theme we'll use repeatedly in this book, that of using a simple algebraic system to construct a more complex one.

To form a quotient field from a ring, we take pairs of elements from the ring (conventionally arranged into fractions) and establish an *equivalence relationship* between them. We also require that the second element in the pair (the denominator) can not be zero. Two pairs (a, b) and (c, d) are equivalent if $ad = bc$, and we write them $\frac{a}{b}$ and $\frac{c}{d}$. We group equivalent pairs together into *equivalence classes* and define our basic field operations as follows:

$$\begin{aligned}
\frac{a}{b} + \frac{c}{d} &= \frac{ad + bc}{bd} \\
\frac{a}{b} - \frac{c}{d} &= \frac{ad - bc}{bd} \\
\frac{a}{b}\frac{c}{d} &= \frac{ac}{bd} \\
\frac{\frac{a}{b}}{\frac{c}{d}} &= \frac{ad}{bc}
\end{aligned}$$

The additive identity element is $\frac{0}{1}$ and the multiplicative identity is $\frac{1}{1}$, using the original identities 0 and 1 from the base ring.[2] Note that the division by zero is not defined, nor do our field axioms require it to be.

Notice that although we define all four field operations, we only use the three ring operations to do it! I.e, when we divide $\frac{a}{b}$ by $\frac{c}{d}$, we need only to form ad and bc in order to form the (ad, bc) pair, which we write as $\frac{ad}{bc}$. We thus divide $\frac{1}{2}$ by $\frac{2}{3}$ to obtain $\frac{3}{4}$ without ever having to divide the *integers* — only multiplying them ($1 \cdot 3 = 3$ and $2 \cdot 2 = 4$).

In general, there is no guarantee that this kind of construction will work. We can't just pair numbers up however we want and call it a field. Several other conditions have to be met. First of all, we have to ensure that the equivalence relationship is well-defined. If $x = y$ (in the sense of equivalence) and $y = z$, then we must have $x = z$, otherwise we can't even cleanly establish the critical notion of an *equivalence class* (which says that $\frac{1}{2}$ and $\frac{2}{4}$

[2] Actually, a ring doesn't have to have a multiplicative identity (check the axioms). But all of our rings will have 1, and it's not that difficult to extend an arbitrary ring into a larger one that does.

are basically the same thing). I emphasize here that the new field, and its new operations, are defined using the equivalence classes, although we muddle this distinction by using the smallest fraction in a class to represent it. Strictly speaking, the multiplicative identity is not $\frac{1}{1}$ but $\{\frac{1}{1}, \frac{2}{2}, \frac{3}{3}, \ldots\}$, the additive identify is not $\frac{0}{1}$ but $\{\frac{0}{1}, \frac{0}{2}, \frac{0}{3}, \ldots\}$, and my example in the last paragraph should have read "we thus divide $\{\frac{1}{2}, \frac{2}{4}, \frac{3}{6}, \ldots\}$ by $\{\frac{2}{3}, \frac{4}{6}, \frac{6}{9}, \ldots\}$ to obtain $\{\frac{3}{4}, \frac{6}{8}, \frac{9}{12}, \ldots\}$..."

Having cleanly established equivalence classes, we have to make sure that our operations actually work consistently on them, since they are defined in terms of fractions within the classes. We need to verify that taking any fraction from one equivalence class and any fraction from other, then applying one of four operations to them, we always get an answer in a third equivalence class. The actual answer can (and will) vary depending on the choice of representative fractions, but it has to always be in the same class. In this way, we confirm that the operations are cleanly defined not just on the fractions, but on the classes. I'm not going to actually make this verification, but leave it as an exercise.

Which is why I excluded zero as a possible denominator. We do this because otherwise our operations aren't cleanly defined on these equivalence classes. $\frac{1}{0}$ is not equivalent to $\frac{0}{1}$ (since $1 \cdot 1 \neq 0 \cdot 0$), so $\frac{1}{0}$ must have a multiplicative inverse (by axiom F1); i.e, some fraction $\frac{a}{b}$ must exist which when multiplied by $\frac{1}{0}$ produces $\frac{1}{1}$, yet by the zero theorem, no such element b can exist in the base ring so that $1b = 0$. Excluding zero as a possible denominator ensures that our field axioms are satisfied.

Yet in the quotient field operations, where we multiply two denominators together to get the result's denominator, what would happen if two non-zero elements can be multiplied to form zero, producing a zero denominator? Nothing in the ring axioms prevents this from happening, so we add an additional axiom.

An *integral domain* $\mathcal{I}$ obeys all the ring axioms, as well as:

$$\text{non-existence of zero divisors} \quad \forall a \neq 0, b \neq 0 \in \mathcal{I}, \quad ab \neq 0 \quad \text{(I1)}$$

The quotient field construction is only defined on integral domains, and I'll leave it as an exercise to show that $\mathbf{Z}$ is an integral domain. The main point of this section is to recognize that the quotient field construction can be performed not only on the integers $\mathbf{Z}$ to obtain the rationals $\mathbf{Q}$, but on any integral domain to obtain its quotient field.

2.3 Polynomial rings and rational function fields

The ring $\mathcal{F}[x]$ and the field $\mathcal{F}(x)$

[van der Waerden], §3.4

Having built a field from a ring, can we build a ring from a field? The answer is yes, and the most important such construction is a *polynomial ring*, whose elements are poly-

nomials in some variable with coefficients in the underlying field, all but a finite number of which must be zero.[3] We write this ring using the underlying field, brackets and the variable, so $\mathcal{F}[x]$ is the ring of polynomials in x with coefficients from the field $\mathcal{F}$.

$\mathcal{F}[x]$ is a ring but not a field. It is, however, an integral domain (left as an exercise), so we can form a quotient field from it, which we write using parenthesis instead of brackets: $\mathcal{F}(x)$. Elements in $\mathcal{F}(x)$ are fractions, both the numerator and denominator of which are polynomials in x. So, for example, $\frac{x}{x-1}$ is a element of $\mathbf{Q}(x)$. Fractions of polynomials are called *rational functions*, so $\mathbf{Q}(x)$ is the *field of rational functions in x over the rational numbers*.

Now, you might ask, "Can't $(x-1)$ be zero? Say, if x is 1?". The answer is *no*. x is not 1 or any other number. x is x, and in $\mathbf{Q}[x]$, $(x-1)$ is as different from zero as $\frac{3}{2}$ is. $(x-1)$, and things like it, are *completely distinct elements* in the algebraic systems in which they are defined.

Now, obviously, we can set x to be 1. But now we are no longer working in $\mathbf{Q}[x]$ — for starters, there is no longer a distinct element x! Now we are working in $\mathbf{Q}$. Setting x equal to 1 mapped everything from $\mathbf{Q}[x]$ into $\mathbf{Q}$. This is a simple example of an *evaluation homomorphism* — a homomorphism (a mapping which preserves operations) from one system to another created by setting an independent variable equal to some constant value.

So, you ask, "what about $\frac{1}{2}$ and $\frac{2}{4}$? Are they distinct elements as well?" *No*. This time we are dealing with elements that are basically the same. This is where the technical details of the quotient field construction become significant. Strictly speaking, we are not working with elements like $\frac{1}{2}$ at all. We are working with the *equivalence classes* defined above. $\frac{1}{2}$ is a *representative* of an equivalance class that includes $\frac{2}{4}$.

If all this seems a bit arbitrary, well, it is. I could easily pick two numbers from $\mathbf{Z}$ and pair them into equivalence classes in some other way than for $\mathbf{Q}$; if my basic axioms were satisified I would even have a field! Whether it would be useful for something other than puzzling the few readers who would bother is a different story. Let me briefly cite a few other examples of *useful* equivalence classes. Take pairs of elements from a field f and g in the expression $f\,dg$ and form equivalence classes based on whether the expression can be transformed to $f'\,dg'$; this is one way to introduce differentials into an algebraic context. Take points in a Cartesian geometry $(x_1, x_2, \ldots x_n)$ and group them together if they are related by a simple constant multiple $(\lambda x_1, \lambda x_2, \ldots \lambda x_n)$; you now have lines through the origin and the basis for projective geometry. Take infinite convergent sequences of rational numbers (from $\mathbf{Q}$) and group them together if the differences between them converge to zero; the equivalence classes are the real numbers. I could keep going. These constructions are easy to form; their utility lies in our ability to relate them to real world problems.

[3] If we relax the finiteness requirement and allow infinite polynomials, we obtain a ring of *power series* over the field, typically written $\mathcal{F}[[x]]$. We will have little use for power series in this book.

2.4 Algebraic Extensions

The field $\mathcal{F}[x]$ mod $n(x)$

Both our quotient field and polynomial ring constructions are examples of *extensions*. Simply put, when we use an algebraic system to construct a new algebraic system that includes the original system as a subset, then the new system is an *extension* of the original.[4] So, $\mathbf{Z}[x]$ is an extension of $\mathbf{Z}$ because we can identify $\mathbf{Z}$ as a subset of $\mathbf{Z}[x]$ and, in fact, even homomorphicly map $\mathbf{Z}$ into $\mathbf{Z}[x]$. Such an *inclusion homomorphism* should not be confused with an evaluation homomorphism, which would map the other way (from $\mathbf{Z}[x]$ into $\mathbf{Z}$).

The only remaining type of extension that will be important to us is the *algebraic extension*. It is another equivalence class construction that we build starting with a polynomial ring over a field, say $\mathcal{F}[x]$. Our equivalence classes are all elements in $\mathcal{F}[x]$ whose differences are multiples of some distinguished irreducible polynomial in $\mathcal{F}[x]$, called the *minimal polynomial* of the field. And I do say *field*, because we don't need to use the quotient field construction with an algebraic extension; the ring is already a field.

Polynomial long division (section 2.6) can be used to reduce modulo polynomials, which allows us to perform addition, subtraction, and multiplication in an modulo field, but how do we perform division? The polynomial Diophantine equation algorithm described in section 2.8 can be used to solve the following equation:

$$sx + tn = \gcd(x, n)$$

Now, if n is irreducible, its GCD with any polynomial of lower degree will be 1, so our equation becomes:

$$sx + tn = 1$$

Reducing mod n, we obtain

$$sx \equiv 1 \mod n$$

So s, when multiplied by x, yields 1, which means that s and x are multipicative inverses. If n was not irreducible, this construction would not work for all values of x, and we would have a ring but not a field.

[4] The key property is that the one system is a subset of the other, not the exact method of construction.

2.5 Trace and Norm

[van der Waerden], §5.7

Given an algebraic extension E of a field F, two of the coefficients in the minimal polynomial have a special significance that often makes them particularly useful. Our only use of them will come in Chapter 3 in the proof of Liouville's theorem.

First, let's note that while we used a minimal polynomial to construct the extension field, in fact, every element in the extension field has a minimal polynomial associated with it, which can be constructed by raising the element to successive powers until one of the powers is a linear combination of the lower powers.

Definition 2.1. *The trace of an element x in E, written $\mathrm{Tr}(x)$, is the coefficient of the second-highest power in its minimal polynomial.*

Definition 2.2. *The norm of an element, written $\mathrm{N}(x)$, is its zeroth-order coefficient.*

The special significance of these elements becomes more obvious if we consider a *splitting field*, which is a futher field extension (an extension of the extension) in which the minimal polynomial factors completely into linear factors. The element x is itself one of the roots in these factors; the remaining roots are called the *conjugates* of x.

Write x's minimal polynomial in the form:

$$\prod_{i=1}^{n}(t - x_i) = \sum_{k=0}^{n}(-1)^k a_k t^{n-k}$$

and note that $\mathrm{Tr}(x)$, the $n - 1^{\mathrm{th}}$ coefficient in the polynomial, is the sum of all the conjugates, while $\mathrm{N}(x)$, the zeroth order coefficient in the polynomial, is the product of all the conjugates.

2.6 Long Division

[van der Waerden], §3.4

As we all learned in grade school, polynomials can be divided using long division. To generalize this in our more abstract context, let's consider a very simple calculation of this type:

$$\begin{array}{r} \frac{1}{2}x+\frac{3}{4} \\ 2x+1 \overline{\big|\; x^2+2x+1} \\ -(x^2+\frac{1}{2}x\quad) \\ \hline \frac{3}{2}x+1 \\ -(\frac{3}{2}x+\frac{3}{4}) \\ \hline \frac{1}{4} \end{array}$$

Each step starts by dividing the leading terms, i.e, x^2 is divided by $2x$ to form $\frac{x}{2}$. Actually, we can be a bit more precise. Each step starts by dividing the leading *coefficients*, since the variables are divided just by subtracting their powers. x^2 divided by x is just x. We divide 1 by 2 to form $\frac{1}{2}$ and in this manner obtain $x \cdot \frac{1}{2} = \frac{x}{2}$.

Next, we multiply this value by the divisor to obtain a polynomial that we will subtract from the dividend (or what remains of it after prior steps). Again, let's be more precise. We multiply the polynomial variable just by adding its powers. What we really have to *multiply* are the *coefficients*. To multiple $\frac{x}{2}$ by $2x+1$ we multiply $\frac{1}{2}$ by 2 to obtain 1, add the powers of x and x to obtain x^2, and arrive at the first term $1 \cdot x^2 = x^2$. Next, we multiply $\frac{1}{2}$ by 1, get $\frac{1}{2}$, add the powers of 1 and x to obtain x, and have the second term $\frac{1}{2} \cdot x = \frac{1}{2}x$. Adding these terms we get $x^2+\frac{1}{2}x$ — the first of the intermediate polynomials.

To perform the third step, we don't have to do anything with the variables. We just subtract the coefficients. These three steps are repeated until we are left with a remainder of lower degree than the divisor.

So, to summarize, working with the polynomial variable is easy — we just add or subtract its integer powers. We perform polynomial long division by dividing, multiplying, and subtracting the *coefficients*. Now, these are three of the basic four operations provided by a field. It follows, therefore, that we can perform polynomial long division on polynomials whose coefficients lie in any field whatsoever. Given $\mathcal{F}[x]$, a polynomial ring over a field, we can use the field operations provided by $\mathcal{F}$ to divide any two elements from $\mathcal{F}[x]$ using polynomial long division and obtain a remainder and a quotient.

We can even say a bit more. Just like with grade school long division, we know that the degree of the quotient will be the difference in degrees of the dividend and the divisor, and that the degree of the remainder will be less than the degree of the divisor. We just need to keep in mind that these degrees are measured relative to the polynomial ring variable, not any other variable that might appear as part of the underlying field.

2.7 Greatest Common Divisors

[van der Waerden], §3.7, §3.8, §5.4 (multivariate rings)
[Geddes], Ch. 7

One of the most important uses of polynomial long division is to compute greatest common divisors (GCDs), at least in theory. In practice, there are other, more efficient algorithms.[5] However, because long division is a simple and straightforward way to compute GCDs, because it provides a theoretical underpinning for other methods, and because it leads us directly to solving polynomial diophantine equations, I'll present it here in this section.

The first thing to observe is that the long division equation, $D = qd + r$ (dividend equals quotient times divisor plus remainder), can be rearranged to read $r = D - qd$, which shows that any common divisor of the dividend and the divisor can be divided out from the right hand side of the equation, so must divide the left hand side also. Thus, common divisors of the dividend and divisor are preserved in the remainder.

Furthermore, since the remainder is always of lower degree than the divisor, we can repeat the long division with the divisor as the new dividend and the remainder as the new divisor. The new remainder will also preserve common divisors of the original dividend and divisor, and will be of lower degree than the original remainder. This process can repeated, lowering the degree of the remainder at each step, until we are left with a zero remainder, i.e. $D' = q'd'$, where I've used primes to emphasize that we are no longer dealing with the original dividend and divisor. Since common divisors have been preserved throughout by D' and d', it follows that d', the divisor of the last step, must be a common divisor. It is, in fact, a greatest common divisor, if GCDs exist in this ring. This has been known since the time of Euclid, at least in the case of integers.

I did say "if" GCDs exist, because nothing in our axioms guarantee their existance. The problem is that there might be a lattice of divisors for a given element, instead of a strict ordering of them. We'll remedy this, again, by introducing a new axiom.

A *unique factorization domain* $\mathcal{U}$ obeys all the ring axioms, as well as:

unique factorization
$$\forall a, b, c, d \in \mathcal{U},$$
$$ab = cd \implies \exists x \in \mathcal{U}, ax = c \cup ax = d \cup bx = c \cup bx = d \quad \text{(U1)}$$

U1 implies I1. Take a unique factorization domain, and pick two elements c and d which are multiples of zero, $cd = 0$. Obviously, we can pick $a = 0$ and $b = 0$ and have $ab = 0 = cd$. So, by U1, x exists so that $0x = c$ or $0x = d$, which by the zero theorem implies that either $c = 0$ or $d = 0$. This proves I1. Thus, a unique factorization domain is also an integral domain. Also, $\mathcal{F}[x]$ is a unique factorization domain (proof omitted).

U1 also implies the existence of GCDs. Consider an element x with two different factors,

[5] See [Geddes], for example

say a and c, so $x = ab$ and $x = cd$. U1 immediately implies that either a is a factor of c, if $aa' = c$, or that c is a factor of a, if $aa' = d$ and $x = caa' = ab$ and $ca' = b$. FIX THIS.

Not all rings satisfy U1. Consider, for example, $\mathbf{Z}[i]; i^2 = -1$, the Gaussian integers. This ring differs from the polynomial ring $\mathbf{Z}[x]$ because polynomials of degree two and higher don't exist since the square of i is -1; i is thus *algebraic* (see below) and this makes all the difference. The number 9 can be factored two different ways in this ring: $9 = 3 \cdot 3 = (4 - i)(4 + i)$. It's not too hard to see that 3 can't be multiplied by any Gaussian integer to form either $(4 - i)$ or $(4 + i)$, so U1 is not satisfied. The Gaussian integers form an integral domain, but not a unique factorization domain.

I did say a greatest common divisor, not the greatest common divisor, because there can be more than one.[6] A *unit* is an invertable element. Put another way, u is a unit if there exists u' such that $uu' = 1$. Now, any divisor can be transformed into another divisor by multiplying it by a unit, since if $uu' = 1$, then $ab = (ua)(u'b)$ for any a and b whatsoever. In particular, a greatest common divisor can be transformed into another greatest common divisor by multiplying it by a unit. I leave without proof the claims that in $\mathcal{F}[x]$, the units are all elements in $\mathcal{F}$, and that all GCDs differ from each other by a unit multiple.

A few words are in order here about GCDs in systems of the form $\mathcal{U}[x]$, i.e, where the coefficients come from a unique factorization domain that is not a field. A factorization in $\mathbf{Q}(x)[y]$ or $\mathbf{Q}(y)[x]$ (both of the form $\mathcal{F}[x]$) is superficially so similar to a factorization in $\mathbf{Q}[x, y]$ (*not* of the form $\mathcal{F}[x]$), that the distinction should be noted. In both of the first two cases, we form a quotient field with respect to one of the two variables and thus obtain a polynomial ring (in the other variable) over the quotient field. In the case of $\mathbf{Q}[x, y]$ we do not form a quotient field with respect to either variable; thus we have a polynomial ring over not a field, but over another polynomial ring.

Now a polynomial ring over a unique factorization domain $\mathcal{U}[x]$ itself satisfies U1 (proof omitted), so by induction any finite series of such polynomial rings over a unique factorization domain (like $\mathbf{Q}[x, y]$) is also a unique factorization domain. This implies that GCDs exist in $\mathcal{U}[x]$-type systems. The problem is finding them, since long division only works cleanly in an $\mathcal{F}[x]$-type system.

The solution, invented by Gauss[7], is to first factor out of each polynomial the GCD of the coefficients (calculated in $\mathcal{U}$) which we call the *content* of the polynomial, leaving a *primitive polynomial*. It can be shown[8] that if a primitive polynomial factors at all, then it factors into primitive polynomials. We thus can compute a primitive GCD of the primitive parts and multiply this by the GCD of the contents to obtain a GCD in $\mathcal{U}[x]$. A LITTLE UNCLEAR.

We will have little use for $\mathcal{U}[x]$ factorizations in this book, since invariably we will calculate GCDs with respect to one variable, and form quotient fields from any others, and thus always be working in $\mathcal{F}[x]$ systems. I mention this mainly to avoid confusion between factoring in $\mathcal{F}[x, y]$ and $\mathcal{F}(x)[y]$, and have thus omitted the proofs of Gauss' method; see the references for details.

[6] Actually, I haven't even proved that GCDs exist at all, and in some algebraic systems, they don't!

[7] check this

[8] van der Waerden

Example 2.3. Compute a GCD of $4x^4+13x^3+15x^2+7x+1$ and $2x^3+x^2-4x-3$ in $\mathbf{Q}[x]$.

$$
\begin{array}{rl}
 & \qquad\qquad\qquad\qquad\qquad\qquad\quad 2x+\frac{11}{2} \\
\cline{2-2}
2x^3+x^2-4x-3 & \big| \; 4x^4+13x^3+15x^2+\;7x+\;1 \\
 & -(4x^4+\;2x^3-\;8x^2-\;6x\qquad) \\
\cline{2-2}
 & \qquad\qquad 11x^3+23x^2+13x+\;1 \\
 & \qquad -(11x^3+\tfrac{11}{2}x^2-22x-\tfrac{33}{2}) \\
\cline{2-2}
 & \qquad\qquad\qquad \tfrac{35}{2}x^2+35x+\tfrac{35}{2}
\end{array}
$$

$$
\begin{array}{rl}
 & \qquad\qquad\qquad\qquad\quad \frac{4}{35}x-\frac{6}{35} \\
\cline{2-2}
\frac{35}{2}x^2+35x+\frac{35}{2} & \big| \; 2x^3+\;x^2-\;4x-\;3 \\
 & -(2x^3+4x^2+\;2x\qquad) \\
\cline{2-2}
 & \qquad\qquad -3x^2-\;6x-\;3 \\
 & \qquad\quad -(-3x^2-\;6x-\;3) \\
\cline{2-2}
 & \qquad\qquad\qquad\quad 0
\end{array}
$$

The divisor of the last step, in this case $\frac{35}{2}x^2+35x+\frac{35}{2}$, is the GCD, or I should say a GCD, since multiplying by any unit will produce a different GCD. In the case of a polynomial ring over a field, the units are the elements of the underlying field, so we can multiply by anything in $\mathbf{Q}$ (i.e, any rational number) and get another GCD. For this example, the obvious thing to multiply by is $\frac{2}{35}$, which both clears the denominators and divides out the common factor in the numerators to produce x^2+2x+1. Both answers are acceptable.

```
(%i17)  gcd(4*x^4 + 13*x^3 + 15*x^2 + 7*x + 1,
            2*x^3 + x^2 - 4*x - 3);

(%o17)
                                  x^2 + 2 x + 1
```

□

Example 2.4. Compute the GCD of $5xy - 5y^2 - 7x + 7y$ and $2x^2 - yx - y^2$ in $\mathbf{Q}[x, y]$.

This is a $\mathcal{U}[x]$-type system, so we'll need to work in a $\mathcal{F}[x]$-type system to perform the computation. Our choices are $\mathbf{Q}(x)[y]$ and $\mathbf{Q}(y)[x]$.

Let's start with $\mathbf{Q}(x)[y]$, and rearrange the polynomials so that y is the polynomial variable:

$$-5y^2 + (5x + 7)y - 7x \qquad \text{and} \qquad -y^2 - xy + 2x^2$$

The first step is to compute the content (GCD of the coefficients) of each polynomial. Clearly, the GCD of -5, $(5x + 7)$, and $-7x$ is 1 and the GCD of -1, $-x$, and $2x^2$ is also 1, so both polynomials are already primitive and we can just proceed with the GCD calculation in $\mathbf{Q}(x)[y]$:

$$\begin{array}{r|l}
 & \frac{1}{5} \\
-5y^2 + (5x+7)y - 7x & -y^2 - xy + 2x^2 \\
 & -(-y^2 + (x+\frac{7}{5})y - \frac{7}{5}x) \\
\hline
 & -(2x+\frac{7}{5})y + (2x^2 + \frac{7}{5}x)
\end{array}$$

$$\begin{array}{r|l}
 & \frac{25}{10x+7}y + \frac{35}{-10x-7} \\
-(2x+\frac{7}{5})y + (2x^2+\frac{7}{5}x) & -5y^2 + (5x+7)y - 7x \\
 & -(-5y^2 + 5xy) \\
\hline
 & 7y - 7x \\
 & -(7y - 7x) \\
\hline
 & 0
\end{array}$$

This leads us to conclude that the last divisor, $-(2x+\frac{7}{5})y+(2x^2+\frac{7}{5}x)$ is a GCD in $\mathbf{Q}(x)[y]$. Now we need to remove its content, which is the GCD of $-(2x+\frac{7}{5})$ and $(2x^2+\frac{7}{5}x)$, or $(2x+\frac{7}{5})$. Dividing through by this polynomial (a polynomial in $\mathbf{Q}[x]$, and thus a unit in $\mathbf{Q}(x)[y]$) we obtain $-y+x$. We now multiply by the GCD of our original contents, but they were just 1, so we conclude that $x - y$ is our GCD in $\mathbf{Q}[x, y]$.

Now let's do all that again in $\mathbf{Q}(y)[x]$. Our polynomials become:

$$(5y - 7)x - (5y^2 - 7y) \qquad \text{and} \qquad 2x^2 - yx - y^2$$

The second one has unit content (the GCD of 2, $-y$, and $-y^2$), but the first one's content is $\gcd_y(5y - 7, 5y^2 - 7y) = 5y - 7$. Dividing this out, we obtain:

$$x - y \qquad \text{and} \qquad 2x^2 - yx - y^2$$

and compute:

$$
\begin{array}{r}
2x + \ y \\
\hline
x - y \,\big|\, 2x^2 - \ yx - y^2 \\
-(2x^2 - 2yx \quad\ ) \\
\hline
yx - y^2 \\
-(yx - y^2\ ) \\
\hline
0
\end{array}
$$

Thus, $x - y$ is the GCD of the primitive polynomials, and it has unit content $\gcd_y(1, -y)$. The GCD of the original contents (1 and $5y - 7$) is 1, so the final result is again $x - y$.

```
(%i18)  gcd(5*x*y - 5*y^2 - 7*x + 7*y,
            2*x^2 - y*x - y^2);

(%o18)
                                 y - x
```

□

2.8 Polynomial Diophantine Equations

The same long division procedure used for GCD computations can also be used solve a certain class of *polynomial Diophantine equations*. A Diophantine equation is one whose variables are restricted to be integers. The most famous example is Fermat's equation, $x^n+y^n = z^n$; Fermat's theorem states that this equation has no solutions $x, y, z, n \in \mathbf{Z}$ for $n > 2$. A generalized Diophantine equation is one whose variables are restricted to some algebraic system, not necessarily $\mathbf{Z}$. A polynomial Diophantine equation is one whose variables are restricted to be polynomials of some form, and the one we will consider here is this:

$$sa + tb = c; \qquad a, b, c \in \mathcal{F}[x] \text{ given}; \qquad s, t \in \mathcal{F}[x] \text{ unknown}$$

Thus, xxx is an equation of this form.

Let's begin by noting that any common divisor of a and b, and in particular $\gcd(a, b)$, can be divided out from the left hand side of the equation, and thus must also divide the right hand side, so c must be be a multiple of $\gcd(a, b)$, or the equation has no solution.

This necessary condition is also sufficient, and the simplest way to demonstrate this is to use the GCD computation in a constructive proof. Note that first step in computing $\gcd(a, b)$ is to solve $a = qb+r$. Rearranging this as $r = a - qb$ we see how the remainder can be expressed in the Diophantine form $sa + tb$. More generally, at each step of the calculation, we solve $D = qd+r$, where D and d are each either a, b, or a remainder from a previous step, so using $r = D - qd$ we can write each remainder in the form $sa + tb$. At the end of the calculation, we will have expressed $\gcd(a, b)$ in the form $sa + tb$.

We now use long division to divide c by $\gcd(a, b)$. Because of the necessity demonstrated above, the division must be exact (i.e, zero remainder) or the equation has no solution. Having computed both $\gcd(a, b) = sa + tb$ and $c = q \gcd(a, b)$ we can now combine these expression to form $c = (qs)a + (qt)b$, which solves the original equation.

This solution is not unique. Given a solution to $c = sa + tb$, we can form any multiple of ab, say mab, and write another solution $c = (s - mb)a + (t + ma)b$. Note however, that $(s - mb)$ has the form of a remainder after dividing s by b (m is the quotient). Since the degree of a remainder is always less than the degree of the divisor, it follows that if $sa + tb = c$ can be solved, then we can always compute an s of lower degree than b, or a t of lower degree than a.

If $\deg(c) < \deg(a) + \deg(b)$, then these conditions are not exclusive; finding an s of lower degree than b implies a t of lower degree than a. To see this, simply note that if $\deg(s) < \deg(b)$, then $\deg(sa) = \deg(s) + \deg(a) < \deg(b) + \deg(a)$. Since $tb = c - sa$, if $\deg(c) < \deg(a) + \deg(b)$ and $\deg(sa) < \deg(a) + \deg(b)$, then $\deg(tb) < \deg(a) + \deg(b)$, which implies $\deg(t) < \deg(a)$.

We will make repeated use of this polynomial Diophantine equation throughout the book.

Example 2.5. Solve:

$$s(4x^4+13x^3+15x^2+7x+1)+t(2x^3+x^2-4x-3)=x^3+5x^2+7x+3$$

for $s,t\in\mathbf{Q}[x]$ satisfying minimal degree bounds.

$$\begin{array}{r|l}
 & 2x+\frac{11}{2} \\
\hline
2x^3+x^2-4x-3 & 4x^4+13x^3+15x^2+7x+1 \\
 & -(4x^4+2x^3-8x^2-6x\ \ \ \ ) \\
\hline
 & 11x^3+23x^2+13x+1 \\
 & -(11x^3+\tfrac{11}{2}x^2-22x-\tfrac{33}{2}) \\
\hline
 & \tfrac{35}{2}x^2+35x+\tfrac{35}{2}
\end{array}$$

$$\begin{array}{r|l}
 & \frac{4}{35}x-\frac{6}{35} \\
\hline
\frac{35}{2}x^2+35x+\frac{35}{2} & 2x^3+x^2-4x-3 \\
 & -(2x^3+4x^2+2x\ \ \ \ ) \\
\hline
 & -3x^2-6x-3 \\
 & -(-3x^2-6x-3) \\
\hline
 & 0
\end{array}$$

$a=4x^4+13x^3+15x^2+7x+1;\qquad b=2x^3+x^2-4x-3;\qquad c=x^3+5x^2+7x+3$

$$\begin{aligned}
a &= (2x+\frac{11}{2})b+(\frac{35}{2}x^2+35x+\frac{35}{2}) \\
x^2+2x+1 &= \frac{2}{35}a-\frac{1}{35}(4x+11)b
\end{aligned}$$

$$\begin{array}{r|l}
 & x+3 \\
\hline
x^2+2x+1 & x^3+5x^2+7x+3 \\
 & -(x^3+2x^2+x\ \ \ \ ) \\
\hline
 & 3x^2+6x+3 \\
 & -(3x^2+6x+3) \\
\hline
 & 0
\end{array}$$

$$c=(x+3)(x^2+2x+1)=\frac{2}{35}(x+3)a-\frac{1}{35}(4x^2+23x+33)b$$

```
(%i19)  a: 4*x^4+13*x^3+15*x^2+7*x+1 $

(%i20)  b: 2*x^3+x^2-4*x-3 $

(%i21)  c: x^3+5*x^2+7*x+3 $
```

Maxima's `gcdex(a,b)` function returns a list `[s,t,gcd]` satisfying

$$as + bt = \gcd(a, b)$$

```
(%i22)  gcdex(a,b);

(%o22)
```

$$\left[1, -\frac{11+4\,x}{2}, \frac{35+70\,x+35\,x^2}{2}\right]$$

If the c polynomial divides $\gcd(a, b)$, then we can just multiply through by the quotient:

```
(%i23)  divide((x^3+5*x^2+7*x+3), %[3]);

(%o23)
```

$$\left[\frac{6+2\,x}{35}, 0\right]$$

```
(%i24)  [s,t,_] : %th(2)*%[1];

(%o24)
```

$$\left[\frac{6+2\,x}{35}, -\frac{33+23\,x+4\,x^2}{35}, x^3+5\,x^2+7\,x+3\right]$$

```
(%i25)  a*s + b*t;

(%o25)
```

$$x^3+5\,x^2+7\,x+3$$

```
(%i26)  kill(a,b,c,s,t)$
```

□

2.9 Square-free factorization

[Geddes], §8.2

A *square-free polynomial* is one with no repeated factors. $x^2 - 1$ is square-free because it factors as $(x - 1)(x + 1)$. $x^2 + 2x + 1$ is not square-free because it factors as $(x + 1)^2$.

Whether or not a polynomial is square-free is independent of the field in which the factorization occurs.

A *square-free factorization* of a polynomial is a factorization into square-free factors, each of which appears at a different power. It is much easy to compute than a full factorization into irreducible factors and for this reason will be quite useful to us.

Surprisingly, a polynomial's square-free factorization is independent of its algebraic system! For example, $x^2 + 1$ is irreducible in $\mathbf{R}[x]$, so its square-free factorization is simply $x^2 + 1$. On the other hand, in $\mathbf{C}[x]$, $x^2 + 1$ factors as $(x + i)(x - i)$. Yet both of these factors combine together in the square-free factorization (since they both appear to the first power), so $x^2 + 1$'s square-free factorization in $\mathbf{C}[x]$ is... $x^2 + 1$.

To compute square-free factorizations, we'll use an operation that, for lack of a better word, I'll call "differentiation." We "differentiate" a polynomial by multiplying each term by its power and then lowering the power by one.

This "differentiation" not to be confused with the field operation that I will define in the next chapter. "Differentiation" is simply a mechanical procedure of lowering powers and multiplying by constants. In particular, no attempt is made to "differentiate" the coefficients *even if they are not constants*.

To compute a square-free factorization, first we "differentiate" the polynomial. The result is a second polynomial with the degree of all factors reduced by one. Note in particular that any factors of unit degree (and only those factors) disappear completely. Dividing this into the original polynomial, we obtain a polynomial with no square factors — all factors now appear with unit degree. Computing the GCD of this polynomial with the original one also produces a polynomial with only factors of unit degree, except that the original unit degree factors are missing. We can divide this last two polynomials into each other to determine the original unit degree factors. Going back to the "differentiation" step, we can keep repeating the process until we have obtained all the square-free factors.

2.10 Partial Fractions Expansion

[van der Waerden], §5.10

As a first application of polynomial Diophantine equations, we use them to construct partial fractions expansions. Consider an element a from a polynomial quotient field $\mathcal{F}(x)$. We can write $a = \frac{n}{d}$ where $n, d \in \mathcal{F}[x]$. If we are now given a factorization of $d = d_1^{e_1} d_2^{e_2} \cdots d_k^{e_k}$, where $d_i \in \mathcal{F}[x]$ and $\gcd_{\mathcal{F}[x]}(d_i, d_j) = 1$ if $i \neq j$, and assuming that a is a proper fraction ($\deg_{\mathcal{F}[x]} n < \deg_{\mathcal{F}[x]} d$), then we can construct a *partial fractions expansion* of a:

$$a = \frac{n}{d} = \sum_{i=1}^{n} \sum_{j=1}^{e_i} \frac{n_{ij}}{{d_i}^j} \qquad \deg_{\mathcal{F}[x]}(n_{ij}) < \deg_{\mathcal{F}[x]}(d_i)$$

We begin by computing an expansion in the form:

$$a = \frac{n}{d} = \sum_{i=1}^{n} \frac{n_i}{{d_i}^{e_i}} \qquad \deg_{\mathcal{F}[x]}(n_i) < e_i \deg_{\mathcal{F}[x]}(d_i)$$

n_1 is found by solving the following polynomial Diophantine equation for n_1 and r_1:

$$n = n_1 \Big(\prod_{j \neq 1} {d_j}^{e_j} \Big) + r_1 (d_1^{e_1})$$

Our degree bounds guarantee that $\deg(n_1) < e_1 \deg(d_1)$, and dividing through by d shows:

$$\frac{n}{d} = \frac{n_1}{d_1^{e_1}} + \frac{r_1}{\prod_{j \neq 1} {d_j}^{e_j}}$$

The second term on the right is a fraction in the original form, but with one less factor in the denominator, so we can recurse and separate out all the ${d_i}^{e_i}$ into seperate fractions. Simple long division now suffices to seperate these fractions:

$$\begin{aligned} n_i &= q_{ij} d_i + r_{ij} \\ \frac{n_i}{d_i^j} &= \frac{q_{ij}}{d_i^{j-1}} + \frac{r_{ij}}{d_i^j} \end{aligned}$$

The r_{ij} are our original n_{ij}, and the degree bounds on long division ensure that $\deg_{\mathcal{F}[x]}(n_{ij}) < \deg_{\mathcal{F}[x]}(d_i)$. We recurse on the first term until we have completed the desired construction.

Example 2.6. Compute the partial fractions expansion of

$$\frac{x^2+3x+2}{x^3-3x^2+4}$$

We begin by factoring the denominator. While factoring can be quite complicated in practice, in this case we need only try some small integers to discover that either 2 or -1 solve the denominator, leading directly to a factorization:

$$\frac{x^2+3x+2}{x^3-3x^2+4} = \frac{x^2+3x+2}{(x-2)^2(x+1)}$$

Next, we solve the polynomial Diophantine equation:

$$x^2+3x+2 = s(x-2)^2 + t(x+1) = sa + tb$$

Compute the GCD of $a = (x-2)^2 = x^2-4x+4$ and $b = x+1$:

$$\begin{array}{r|l} & x-5 \\ \hline x+1 & x^2-4x+4 \\ & -(x^2+\ x\ \ \ \ ) \\ \hline & -5x+4 \\ & -(-5x-5\ ) \\ \hline & 9 \end{array}$$

Since the remainder, 9, is a unit, a and b have no common factors and their GCD is 1. Of course, this result is hardly surprising since $(x-2)$ and $(x+1)$ have no common factor between them.

$$a = (x-5)b + 9$$
$$9 = a - (x-5)b$$
$$1 = \frac{1}{9}[a-(x-5)b]$$
$$x^2+3x+2 = \frac{1}{9}(x^2+3x+2)a - \frac{1}{9}(x^2+3x+2)(x-5)b$$
$$x^2+3x+2 = \frac{1}{9}(x^2+3x+2)a - \frac{1}{9}(x^3-2x^2-13x-10)b$$

Our degree bounds aren't met yet, so we divide $b = x+1$ into x^2+3x+2:

$$\begin{array}{r|l} & x+2 \\ \hline x+1 & x^2+3x+2 \\ & -(x^2+\ x\ \ \ \ ) \\ \hline & 2x+2 \\ & -(2x+2\ ) \\ \hline & 0 \end{array}$$

The zero remainder means that the a term drops away completely, and after subtracting $(x+2)a = x^3 - 2x^2 - 4x + 8$ from the b coefficient, we conclude that:

$$x^2 + 3x + 2 = -\frac{1}{9}(-9x - 18)b = (x+2)b$$

In other words (remember that $b = x + 1$),

$$\frac{x^2 + 3x + 2}{(x-2)^2(x+1)} = \frac{(x+2)(x+1)}{(x-2)^2(x+1)} = \frac{(x+2)}{(x-2)^2}$$

We need only divide $(x + 2)$ by $(x - 2)$:

$$\begin{array}{r|l} & 1 \\ \cline{2-2} x-2 & x+2 \\ & -(x-2) \\ \cline{2-2} & 4 \end{array}$$

so,

$$\frac{x^2 + 3x + 2}{x^3 - 3x^2 + 4} = \frac{4}{(x-2)^2} + \frac{1}{(x-2)}$$

Again, let's verify this result using Maxima:

```
(%i27) partfrac((x^2 + 3*x + 2)/(x^3 - 3*x^2 + 4), x);

(%o27)
```

$$\frac{1}{x-2} + \frac{4}{(x-2)^2}$$

□

2.11 Resultants

[van der Waerden], §5.8; [Lang], §IV.8

At times, we will want a simple way of testing two polynomials over a field to see if they have a GCD, without actually computing it. This is more than just a computational convenience. The presence of the polynomial's variable in the GCD often encumbers us. On the other hand, the *resultant* yields a simple element from the underlying field that is zero if the polynomials have a non-trivial GCD and non-zero otherwise.

For example, the polynomials $tx + x + t + 1$ and $ty + y$ share $t + 1$ as a GCD, so their t-resultant is 0. MORE HERE - why a t-resultant, not an x-resultant?

The resultant is defined as the determinant of the Sylvester matrix $S_x(P, Q)$, which is the $m + n \times m + n$ matrix constructed from two polynomials (in $\mathcal{F}[x]$) P and Q of degrees m and n (all the blanks are zeros):

$$P = \sum_{i=0}^{m} p_i \, x^i \qquad Q = \sum_{i=0}^{n} q_i \, x^i$$

$$S_x(P,Q) = \begin{pmatrix} p_m & p_{m-1} & \cdots & p_0 & & & \\ & p_m & p_{m-1} & \cdots & p_0 & & \\ & & \cdots & & \cdots & & \\ & & & p_m & p_{m-1} & \cdots & p_0 \\ \vdots & & & \vdots & & & \vdots \\ q_n & q_{n-1} & \cdots & q_0 & & & \\ & q_n & q_{n-1} & \cdots & q_0 & & \\ & & \cdots & & \cdots & & \\ & & & q_n & q_{n-1} & \cdots & q_0 \end{pmatrix}$$

In plain English, the matrix is constructed by forming the first row from the first polynomial coefficients, adding $n - 1$ trailing zeros at the end of the row. The second row is formed by shifting the first row one position to the right. This shifting is repeated a total of $m - 1$ times to obtain the first m rows. The last n rows are constructed in the same way from the second polynomial.

Now consider the following straightforward matrix identity:

$$S_x(P,Q) \begin{pmatrix} x^{n+m-1} \\ x^{n+m-2} \\ \vdots \\ x \\ 1 \end{pmatrix} = \begin{pmatrix} Px^{m-1} \\ Px^{m-2} \\ \vdots \\ Qx \\ Q \end{pmatrix}$$

If $\det S_x(P, Q)$ is non-zero, then the Sylvester matrix is invertible, and we can form the

following equation:

$$\begin{pmatrix} x^{n+m-1} \\ x^{n+m-2} \\ \vdots \\ x \\ 1 \end{pmatrix} = S_x(P,Q)^{-1} \begin{pmatrix} Px^{m-1} \\ Px^{m-2} \\ \vdots \\ Qx \\ Q \end{pmatrix}$$

Since the matrix is formed exclusively from the polynomials' underlying field $\mathcal{F}$, its inverse must also be formed from $\mathcal{F}$. Now consider the bottom element in the last equation. It must have the following form:

$$1 = f_0 P x^{m-1} + f_1 P x^{m-2} + \ldots + f_{n+m-1} Q x + f_{n+m} Q \qquad f_i \in \mathcal{F}$$
$$1 = AP + BQ \qquad A, B \in \mathcal{F}[x]$$

The only way this statement can be true is if P and Q (viewed as polynomials in x) have a trivial GCD, so a non-zero determinant of $S_x(P,Q)$ imply that P and Q have only a trivial GCD.

Conversely, assume that $\gcd_x(P,Q) = 1$. Then we can solve a series of polynomial Diophantine equations to express $1, x, \ldots, x^{n+m-1}$ as $AP+BQ$, where $\deg A < \deg Q = n$ and $\deg B < \deg P = m$, which suffices to construct an inverse of the Sylvester matrix.

We have thus proved:

Theorem 2.7. *The resultant is zero iff the two polynomials have a non-trivial GCD.* □

Let me note two points. First, though we used a field construction for the proof, determinants are constructed using only ring operations, so resultants can be computed in any ring and the result will be a ring element. The proof depends on the ring having a well-defined fraction field, but since UFDs are integral domains, the GCD concept doesn't make sense without a fraction field anyway.

Second, if the underlying ring involves multiple variables, the net effect of the resultant is to eliminate one of them. To see this, imagine arbitrary values being assigned to the other variables. The resultant yields a condition on the remaining variables for the original system to be solvable.

Here's a Maxima function that constructs a Sylvester matrix:

```
(%i28)  syl_elem(a,b,t,i,j) :=
           if (i > hipow(b,t))
              then coeff(b,t,i-j)
              else coeff(a,t,i-j+hipow(a,t))$

(%i29)  sylvester(a,b,t) := block(
           [size: hipow(a,t) + hipow(b,t)],
           genmatrix(lambda([i,j], syl_elem(a,b,t,i,j)),
                     size, size)
        )$
```

Example 2.8. Compute the $t-$resultant of $t^2 - 1 - x$ and $t^3 - t - y$.

```
(%i30)  sylvester(t**2 - 1 - x, t**3 - t - y, t);
```

(%o30)

$$\begin{pmatrix} 1 & 0 & -x-1 & 0 & 0 \\ 0 & 1 & 0 & -x-1 & 0 \\ 0 & 0 & 1 & 0 & -x-1 \\ 1 & 0 & -1 & -y & 0 \\ 0 & 1 & 0 & -1 & -y \end{pmatrix}$$

```
(%i31)  ratsimp(determinant(%));
```

(%o31)

$$y^2 - x^3 - x^2$$

More succinctly, we can use the Maxima function `resultant`:

```
(%i32)  resultant(t**2 - 1 - x, t**3 - t - y, t);
```

(%o32)

$$y^2 - x^3 - x^2$$

This result implies that the polynomials $t^2 - 1 - x$ and $t^3 - t - y$ have a common factor when $y^2 - x^3 - x^2 = 0$. For example, this condition is satisfied when $x = 3$ and $y = 6$, and the two polynomials become $t^2 - 4$ and $t^3 - t - 6$, which have the common factor $t - 2$:

```
(%i33)  factor(t**2-4);

(%o33)
```

$$(t-2)\,(t+2)$$

```
(%i34)  factor(t**3-t-6);

(%o34)
```

$$(t-2)\,\left(t^2+2\,t+3\right)$$

□

2.12 Algebraic Closure

The fields R and C

[van der Waerden], §11

The most important way to characterize a field $\mathcal{F}$ is to extend it to a polynomial ring $\mathcal{F}[x]$ and then study how polynomials factor in $\mathcal{F}[x]$.

There exist fields $\mathcal{F}$ where all polynomials in $\mathcal{F}[x]$ can be completely factored into linear factors. Such a field is said to be *algebraically closed*. The **Fundamental Theorem of Algebra** states that the complex field $\mathbf{C}$ is algebraically closed. There are several proof routes for this theorem. The most common involves complex analysis, and can be found in any standard complex analysis text, usually near Liouville's Theorem on the behavior of bounded entire functions. I won't go into it here.

I do want to note the difference between $\mathbf{C}$ and $\mathbf{C}(x)$. Both are fields. $\mathbf{C}$ is algebraically closed because any polynomial in $\mathbf{C}[x]$ can be completely factored into linear factors, i.e, $x^3 - 3x^2 + 4 = (x-2)^2(x+1)$. $\mathbf{C}(x)$ is *not* algebraically closed. If it were, then all polynomials in $\mathbf{C}(x)[y]$ could be completely factored, but in fact, there exist polynomials such as $y^2 - x$ that can not be factored, so $\mathbf{C}(x)$ is not algebraically closed.

2.13 Polynomial factorization (optional)

[van der Waerden], §5.6 [Geddes], Chs. 5, 6, 8

Let's conclude this chapter by taking a least a brief look at fully factoring a polynomial into its irreducible factors. There are several reasons to do this.

First of all, it's easy to declare that "the Fundamental Theorem of Algebra tells us that any polynomial in **C**[x] can be factored into linear factors", and maybe even prove it. That's true, but when it comes time to actually do a computation, how do we proceed? How do we actually factor a polynomial? Call it the price of success. Differential algebra is solid enough to actually compute integrals, so existance theorems don't cut it. We need constructive algorithms.

Second, it's a surprisingly difficult problem. An appreciation of its difficulty now will motivate the discussion later when I show various techniques that have been developed to avoid full factorization whenever possible. Yet the fact remains that it is at times unavoidable.

Finally, both techniques that I will discuss here work according to a basic principle that we'll use again later in a more advanced context, so it makes sense to present it now in a simpler form. Specifically, we'll solve a difficult problem in an algebraic system by using a homomorphism to map into a different algebraic system where we can solve the

problem more easily, then find some way of "lifting" this answer back into the original system. This is one of the most powerful solution methods in algebra, and has been used to solve problems once thought impossible.

Let's start simple. We want to factor a polynomial in $\mathbf{Z}[x]$, the ring of polynomials with integer coefficients. If such a polynomial has a factorization in $\mathbf{Z}[x]$, for example $x^2-1 = (x+1)(x-1)$, we want to find it. If it has no such factorization, for example x^2+1 (which would require at least $\mathbf{Z}[i,x]; i^2 = -1$ to factor), we want to prove this.

Now consider what happens when we set x equal to some specific integer value, say a. Any polynomial in $\mathbf{Z}[x]$ will be transformed into an integer. Thus, we have a mapping $\phi_{x-a} : \mathbf{Z}[x] \rightarrow \mathbf{Z}$ from polynomials to integers. Not only is this a mapping, but it is a *homomorphism*, a mapping that preserves the operations, so $\phi_{x-a}(m+n) = \phi_{x-a}m \hat{+} \phi_{x-a}n$ and $\phi_{x-a}(m \cdot n) = \phi_{x-a}m \hat{\cdot} \phi_{x-a}n$, where I have used $\hat{+}$ and $\hat{\cdot}$ to emphasize that these operations are operations in $\mathbf{Z}$, and distinct from $+$ and $\cdot$, which are operations in $\mathbf{Z}[x]$.[9] I leave it an exercise to actually prove this is a homomorphism.

Since ϕ_{x-a} is a homomorphism, any factorization of a polynomial in $\mathbf{Z}[x]$ must map into a factorization of its image integer in $\mathbf{Z}$. In other words, if a polynomial factors into smaller polynomials (all with integer coefficients), then setting the variable equal to some specific integer causes all the polynomials to evaluate into integers, which must themselves factor. Consider $x^2 - 1 = (x+1)(x-1)$. If we set $x = 2$ (the evaluation homomorphism ϕ_{x-2}), then the equation becomes $3 = 3 \cdot 1$. This happens irregardless of our choice of integer. Choosing $x = 3$ (ϕ_{x-3}) transforms $x^2 - 1 = (x+1)(x-1)$ into $8 = 4 \cdot 2$.

Thus, we have our homomorphism, which maps our problem from $\mathbf{Z}[x]$ into $\mathbf{Z}$ and transforms the factorization of polynomials into a factorization of integers. Although factoring integers is certainly not trivial (the security of the RSA cryptosystem depends on its near impossibility for large numbers), it is much easier than factoring polynomials. Not only easier, but *finite*. There are only a certain number of ways any given integer can factor, and for relatively small integers, we can enumerate them by computing a prime factorization and then listing the finite number of ways that the primes can be combined into factors. The number 3, for example, can be split into two integer factors in only one of four ways: $3 \cdot 1$, $1 \cdot 3$, $-3 \cdot -1$, and $-1 \cdot -3$.

[9] The symbols $\cdot$ and $\hat{\cdot}$ represent multiplication, which we normally omit entirely, but I have written explicitly here to make this point.

2.14 Exercises

Factor the following polynomials in $\mathbf{Z}[x]$, $\mathbf{Q}[x]$, $\mathbf{R}[x]$, and $\mathbf{C}[x]$:

1. $x^2 - 2$ (factors in $\mathbf{R}[x]$ and $\mathbf{C}[x]$, but not in $\mathbf{Z}[x]$ or $\mathbf{Q}[x]$)
2. $x^2 + 1$ (factors in $\mathbf{C}[x]$, but not in any other ring)

Write a computer program to factor the following polynomials:

3. $x^5 + 2\,x^4 + 2\,x^3 - x - 1$

   ```
   expand((x^2+x+1)*(x^3+x^2-1));
   ```

4. $34\,x^5 + 51\,x^4 + 60\,x^3 + 25\,x^2 + 8\,x - 1$

   ```
   factor(34*x^5+51*x^4+60*x^3+25*x^2+8*x-1);
   ```

   Ans: $(x^2 + x + 1)\,(34\,x^3 + 17\,x^2 + 9\,x - 1)$

Write a computer program to factor the following polynomials:
(Hint: You'll need a 2 dimensional grid)

5. $2x^3 + 3x^2y - 7x^2 + 14xy + 22x + 21y^2 - 16y - 77$

   2*x**3 + 3*x**2*y - 7*x**2 + 14*x*y + 22*x + 21*y**2 - 16*y - 77

   Ans: $(2x + 3y - 7)\,(x^2 + 7y + 11)$

   ```
   (%i35)  factor(2*x**3 + 3*x**2*y - 7*x**2 + 14*x*y
                  + 22*x + 21*y**2 - 16*y - 77);

   (%o35)
                        (3 y + 2 x - 7) (7 y + x^2 + 11)
   ```

6. $2x^4 + x^2y^3 + 2x^2y - x^2 + y^4 + y^3 - 3y - 3$

   2*x**4 + x**2*y**3 + 2*x**2*y - x**2 + y**4 + y**3 - 3*y - 3

   Ans: $(x^2 + y + 1)\,(2x^2 + y^3 - 3)$

Chapter 3

Differential Algebra

3.1 Differential Fields

The advent of the modern, axiomatized approach to mathematics at the turn of the twentieth century led directly to the development of abstract algebra, with its rings and fields, in the 1920s. By 1940, a Columbia University professor named J.F. Ritt had proposed the concepts of *differential rings* and *differential fields*. They are exactly analogous to ordinary rings and fields, except that they are equipped with a third basic operator (addition and multiplication are the first two), called *derivation*. A derivation is a unary operator (the other two are binary), which we shall typically denote by D. Since algebra is fundamentally concerned with how operators commute with each other, the first question we are lead to ask are, "How does derivation commute with addition and multiplication?" The answer is to be found in two basic axioms:

addition law of derivations	$\forall a, b \in \mathcal{D},$	$D(a+b) = Da + Db$	(D1)
multiplication law of derivations	$\forall a, b \in \mathcal{D},$	$D(ab) = aDb + bDa$	(D2)

Neither axiom should come as any great surprise. After all, these are just the basic addition and multiplication rules we learned in first year calculus. Yet note how they are being presented; not as results derived from some theorem involving fractions and limits, but as axioms that are assumed true from the start. One of the great themes of differential algebra is that we purge from the subject almost any mention of limits; for us, derivation is just a mapping in a field that carries an object a to another object b. Integration, then, is little more than the inversion of derivation: given an object b, can we find an object a which maps into b?

Yet the connection to calculus should be made clear. Since derivation (in the calculus sense) obeys these two axioms for derivation (in the algebra sense), the calculus derivation will always behave as an algebra derivation, so any theory we develop for the algebra derivation will apply immediately to the calculus version.

What can we determine from these two axioms? A surprising lot, in my opinion.

Theorem 3.1.

- $D(0) = 0$
- $D(1) = 0$
- $D(\frac{1}{a}) = -\frac{1}{a^2}D(a)$
- $D(cx) = c\,D(x)$ *if* $D(c) = 0$

Proof

$$D(0) = D(0) + D(0) - D(0) = D(0+0) - D(0) = D(0) - D(0) = 0$$

$$D(1) = D(1 \cdot 1) = D(1) + D(1)$$

$$D(1) = D(1) + D(1) - D(1) = D(1) - D(1) = 0$$

$$0 = D(1) = D(a \cdot \frac{1}{a}) = \frac{1}{a}D(a) + aD(\frac{1}{a})$$

$$aD(\frac{1}{a}) = -\frac{1}{a}D(a)$$

$$D(\frac{1}{a}) = -\frac{1}{a^2}D(a)$$

□

It follows immediately from this theorem that our entire prime subfield, as well as any purely algebraic extension thereof, must map to zero under derivation.

Theorem 3.2. *The set of all elements in a differential field which map to zero under derivation forms a subfield.*

□

The subfield which maps to zero is called the *constant subfield*. It necessarily includes the prime subfield and any elements algebraic over the prime subfield, but may include other transcendental elements as well. For example, consider $\mathbf{R}$, the real numbers. 2 is in the prime subfield, so $D(2) = 0$; $\sqrt{2}$ is algebraic over the prime subfield, so $D(\sqrt{2}) = 0$; π is transcendental over the prime subfield, so doesn't *have* to map to zero, but we will (obviously) set $D(\pi) = 0$. All three elements — 2, $\sqrt{2}$, π — are in the constant subfield.

Theorem 3.3. *The derivation of an algebraic extension is determined uniquely by the derivation of its subfield.*

Proof

The derivation of an algebraic element is completely defined by the subfield's derivation and the element's minimal polynomial.

Given an element ξ, let its minimal polynomial be:

$$\sum_i a_i \xi^i = 0$$

Differentiating this polynomial (using the D1 and D2 axioms), we obtain:

$$\sum_i (a_i' \xi^i + i a_i \xi^{i-1} \xi') = 0$$

$$\sum_i i a_i \xi^{i-1} \xi' = -\sum_i a_i' \xi^i$$

$$\xi' = -\frac{\sum_i a_i' \xi^i}{\sum_i i a_i \xi^{i-1}}$$

□

The upshot of all this is that our basic D1 and D2 axioms completely define a derivation both for our prime subfield as well as any purely algebraic extensions. It therefore follows that we need only specify the behavior of a derivation on transcendental elements and we will have completely defined the derivation.

We will use four types of transcendental elements in our theory:

1. Constants. $D(c) = 0$

2. The distinguished variable of integration. $D(x) = 1$

Since this is an O.D.E. theory, and particularly an integration theory, we are always integrating with respect to some variable of integration. There is no loss of generality in labeling it x. By setting $D(x) = 1$ we establish that our derivation is in fact a derivative and not a differential.

Incidently, Ritt had already conceived back in the 1940s of equipping a differential field with multiple derivations, one for each of a set of independent variables. This corresponds nicely to what is needed for a P.D.E. theory. Thus, given variables x, y and z, we could construct derivatives D_x, D_y and D_z so that $D_x(x) = 1$, $D_x(y) = 0$, $D_x(z) = 0$ and

so on. Since our focus is on integration, I'll have nothing more to say about fields with multiple derivations.

3. Logarithmic extensions. $D(\theta) = \frac{D(\phi)}{\phi}$

4. Exponential extensions. $D(\theta) = \theta D(\phi)$

ϕ, in both of these cases, is some element in the underlying field.

These two extentions clearly correspond to $\theta = \ln \phi$ (in the logrithmic case) and $\theta = \exp \phi$ (in the exponential case). The key point I want to made immediately is that these are *transcendental* extensions... and not all logarithms and exponentials are transcendental! Transcendental extensions are defined by exclusion — any extension that isn't algebraic is transcendental. If we're dealing with an algebraic extension, even if defined using logarithms and exponentials, we have to use our algebraic theory.

Example 3.4. Represent $\dfrac{4^x + 1}{2^x + 1}$ in Liouvillian form

There are three ways to do this — the easy way, the hard way, and the wrong way.

Let me first note that $4^x = (2^2)^x = (2^x)^2$. The existence of this algebraic relationship between 4^x and 2^x means that we *can not* use two seperate transcendental extensions. So this:

$$\frac{\theta + 1}{\phi + 1}; \theta = \exp(x \ln 4); \phi = \exp(x \ln 2)$$

is the *wrong* way.

The *easy* way is to set up 2^x first and then construct 4^x as its square:

$$\frac{\phi^2 + 1}{\phi + 1}; \phi = \exp(x \ln 2)$$

You can also do this the *hard* way, setting up 4^x first and then using an additional algebraic extension to get its square root, 2^x:

$$\frac{\theta + 1}{\phi + 1}; \theta = \exp(x \ln 4); \phi^2 = \theta$$

See Example 6.7 for the actual integration.

□

That's it! The basic two differential axioms, algebraic extensions, quotient fields, and these four types of transcendentals, round out the entire base algebraic structure we'll need to construct our theory. We do need to be careful, though, as the last example illustrated. In these simple examples, figuring out which elements are algebraic and which are transcendental is easy, but in more complex expressions this may not be obvious. We'll discuss in Chapter ?? how to test new elements for transcendence.

Definition 3.5. *An* **elementary extension** *of a differential field is a differential extension field constructed using a finite number of algebraic, logarithmic, and exponential extensions.*

Definition 3.6. *A* **elementary function** *of a single variable x over a specified field of constants K is a function in an elementary extension of the rational function field $K(x)$.*

What about sines and cosines, all those arc-functions, raising things to powers, and all that? Turns out we can express all those operations using just our basic extensions. The key here is Euler's famous identity $e^{i\theta} = i\sin\theta + \cos\theta$.

Example 3.7. Express $\sin x$ in Liouvillian form

Euler's identity immediately gives:

$$\sin x = -i\,\frac{e^{ix} - e^{-ix}}{2}$$

Therefore, starting from $\mathbf{C}(x)$, we add the exponential extension $\theta = \exp(ix)$, and conclude that $\sin x$ can be expressed as the rational function:

$$\frac{\theta^2 - 1}{2i\theta}$$

in the field $\mathbf{C}(x, \theta); \theta = \exp(ix)$.

□

If trigonometric functions can be represented using complex exponentials, then it should come as no real surprise that inverse trigonometric functions can be represented with complex logarithms.

Example 3.8. Represent $\arcsin x$ in Liouvillian form

Let's start with Euler's identity and take its logarithm:

$$e^{i\theta} = i\sin\theta + \cos\theta$$

$$i\theta = \ln(i\sin\theta + \cos\theta)$$

Now, if $\theta = \arcsin x$, then $x = \sin\theta$, and we can use the basic $\sin^2\theta + \cos^2\theta = 1$ identity to compute $\cos\theta = \sqrt{1 - \sin^2\theta} = \sqrt{1 - x^2}$. Substituting above:

$$i\theta = \ln(ix + \sqrt{1 - x^2})$$

$$\theta = -i\ln(ix + \sqrt{1 - x^2})$$

$$\arcsin x = -i\ln(ix + \sqrt{1 - x^2})$$

Thus, we need first an algebraic extension to construct $\phi = \sqrt{1 - x^2}$, followed by a logarithm extension to construct $\arcsin x = -i\ln(ix + \phi)$.

□

I think the details of further constructions along these lines are straightforward enough that I will simply summarize them in a table.

3.2 Liouvillian Forms

Expression	Liouvillian Form	Expression	Liouvillian Form
f^g	$e^{g \ln f}$		
$\sin x$	$-i\,\dfrac{e^{ix}-e^{-ix}}{2}$	$\sinh x$	$\dfrac{e^{x}-e^{-x}}{2}$
$\cos x$	$\dfrac{e^{ix}+e^{-ix}}{2}$	$\cosh x$	$\dfrac{e^{x}+e^{-x}}{2}$
$\tan x$	$-i\,\dfrac{e^{ix}-e^{-ix}}{e^{ix}+e^{-ix}}$	$\tanh x$	$\dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$
$\sec x$	$\dfrac{2}{e^{ix}+e^{-ix}}$	$\operatorname{sech} x$	$\dfrac{2}{e^{x}+e^{-x}}$
$\csc x$	$\dfrac{2i}{e^{ix}-e^{-ix}}$	$\operatorname{csch} x$	$\dfrac{2}{e^{x}-e^{-x}}$
$\cot x$	$i\,\dfrac{e^{ix}+e^{-ix}}{e^{ix}-e^{-ix}}$	$\coth x$	$\dfrac{e^{x}+e^{-x}}{e^{x}-e^{-x}}$
$\arcsin x$	$-i\,\ln(ix+\sqrt{1-x^2})$	$\sinh^{-1} x$	$\ln(x+\sqrt{x^2+1})$
$\arccos x$	$-i\,\ln(x+i\sqrt{1-x^2})$	$\cosh^{-1} x$	$\ln(x+\sqrt{x^2-1})$
$\arctan x$	$\dfrac{1}{2}\,i\,\ln\dfrac{ix-1}{ix+1}$	$\tanh^{-1} x$	$\dfrac{1}{2}\ln\dfrac{1+x}{1-x}$
$\sec^{-1} x$	$-i\,\ln\dfrac{1+i\sqrt{x^2-1}}{x}$	$\operatorname{sech}^{-1} x$	$\dfrac{1}{2}\ln\dfrac{1+\sqrt{1-x^2}}{1-\sqrt{1-x^2}}$
$\csc^{-1} x$	$-i\,\ln\dfrac{i+\sqrt{x^2-1}}{x}$	$\operatorname{csch}^{-1} x$	$\dfrac{1}{2}\ln\dfrac{\sqrt{1+x^2}+1}{\sqrt{1+x^2}-1}$
$\cot^{-1} x$	$\dfrac{1}{2}\,i\,\ln\dfrac{i+x}{i-x}$	$\coth^{-1} x$	$\dfrac{1}{2}\ln\dfrac{x+1}{x-1}$

3.3 Liouville's Theorem

The next problem we must confront is to limit the number of possible fields in which we can find solutions to our problem. So far, we have seen how to construct an algebraic system to express any elementary function, but there are an infinity of such systems. Searching them exhaustively for the solution to a given integral is out of the question. Fortunately, it's been known for almost 200 years that there are severe restrictions on what extensions can appear in an integral above and beyond those used in the original integrand.

For example, consider the expression e^x. Differentiating it yields, well, e^x. Now the key thing to note is that the exponential does not disappear after differentiation. This, in fact, is a general property of exponentials — differentiation never makes them disappear. They can change around, to be sure, $\frac{d}{dx}e^{2x} = 2e^{2x}$, but notice that the exponential is still present in the result. Therefore, since the solution to our integral must differentiate into the original integrand, we conclude that no new exponentials can appear in the integral beyond those in the integrand. If there were new exponentials in the result, then they would have to appear in the integrand as well, since they can never disappear under differentiation.

The same thing happens with roots. Differentiate $\sqrt{x}$ and you get $\frac{1}{2\sqrt{x}}$. This time the root moves from the numerator to the denominator, but again, it doesn't completely disappear. This is a general property of roots, algebraic extensions in general, in fact.

Logarithms are different, though. Differentiate $\ln x$ to get $\frac{1}{x}$. The logarithm is gone. So new *logarithms* can appear in integrals, because they can disappear under differentiation to recover the original integrand. Even here, though, there are important restrictions. The logarithms have to appear with constant coefficients (because something like $x \ln x$ would differentiate into $1 + \ln x$), can not appear in powers or in denominators ($\frac{d}{dx}\ln^2 x = 2\frac{\ln x}{x}$), and can not be nested ($\frac{d}{dx}\ln(\ln x) = \frac{1}{x \ln x}$).

These examples are all special cases of *Liouville's Theorem* — the only new extensions that can appear in an integral are simple logarithms with constant coefficients. Let's begin by stating and proving some basic properties of our three basic types of extensions.

Theorem 3.9. *Let $E = K(\theta)$ be a simple transcendental logarithmic extension of a differential field K with the same constant subfield as K, let $p = \sum p_i\theta^i$ be a polynomial in $K[\theta]$, and let $r = a/b$ be a rational function in $K(\theta)$ ($a, b \in K[\theta]$). Then:*

1. *If p's leading coefficient is constant ($p'_n = 0$), then $\mathrm{Deg}_\theta\, p' = \mathrm{Deg}_\theta\, p - 1$*
2. *If p's leading coefficient is not constant ($p'_n \neq 0$), then $\mathrm{Deg}_\theta\, p' = \mathrm{Deg}_\theta\, p$*
3. *If p is monic and irreducible, then $p' \nmid p$*
4. *If an irreducible factor appears in r's denominator with multiplicity m, then it appears in r''s denominator with multiplicity $m + 1$*
5. *$r' \in K$ if and only if r has the form $c\theta + k$, where c is a constant*

Proof

The first two statements follow easily from considering p':

$$p' = \sum_{i=0}^{n} (p_i' \theta^i + i p_i \theta' \theta^{i-1}) = \sum_{i=0}^{n} \left(p_i' + (i+1)\, p_{i+1} \theta'\right) \theta^i$$

Note that since $K(\theta)$ is a logarithmic extension, $\theta' \in K$, so for all i the entire expression $(p_i' + (i+1)p_{i+1}\theta')$ is in K. In particular, since p_{n+1} is zero, the n^{th} coefficient of p' is just p_n' and the θ-degree of p' will be n if p_n' is non-zero. On the other hand, if p_n' is zero, then the $n-1^{\text{th}}$ coefficient of p' is $(p_{n-1}' + np_n\theta')$ which would be zero only if $\theta' = -\frac{p_{n-1}'}{np_n} = (-\frac{p_{n-1}}{np_n})'$ (by Theorem 3.1 since p_n is constant), which implies an algebraic relationship between θ and $-\frac{p_{n-1}}{np_n}$ (specifically, they differ only by a constant, which must be in K), contradicting the transcendence of E over K.

Next, if p is monic and irreducible, then $\operatorname{Deg}_\theta p' = \operatorname{Deg}_\theta p - 1$, and no lower degree polynomial can divide an irreducible polynomial, establishing the third claim.

Now consider a rational function $r = a(\theta)/b(\theta)$ in its normalized form, so $\gcd(a, b) = 1$ and b is monic. Now we can factor b into irreducible factors ($b = \prod b_i(\theta)^{m_i}$) and expand r using partial fractions (Section 2.10):

$$r = a_0(\theta) + \sum_{i=1}^{\mu} \sum_{j=1}^{m_i} \frac{a_{ij}(\theta)}{b_i(\theta)^j}$$

where $a_0, a_{ij}, b_i \in K[\theta]$ and $\operatorname{Deg}_\theta a_{ij} < \operatorname{Deg}_\theta b_i$. Now let's differentiate:

$$r' = a_0'(\theta) + \sum_{i=1}^{\mu} \sum_{j=1}^{m_i} \left[\frac{a_{ij}'(\theta)}{b_i(\theta)^j} - \frac{j\, a_{ij}(\theta)\, b_i'(\theta)}{b_i(\theta)^{j+1}} \right]$$

a_{ij} does not divide b_i (since $\operatorname{Deg}_\theta a_{ij} < \operatorname{Deg}_\theta b_i$, and we proved above that b_i' does not divide b_i (since b_i is monic and irreducible), so there is exactly one term on the right hand side with $b_i(\theta)^{m_i+1}$ in its denominator and no other terms with higher powers. Therefore, r' must have a $b_i(\theta)^{m_i+1}$ in its denominator, establishing the fourth claim.

Finally, since the hypothesis of the fifth claim states that r' is in K, it can not have any θ terms in its denominator (or anywhere else), so there can not be any $b_i(\theta)$ factors, and r must be a polynomial. Futhermore, our first two claims imply that if $\operatorname{Deg}_\theta r' = 0$ (since $r' \in K$), then $\operatorname{Deg}_\theta r$ can be at most 1, and its leading coefficient must be constant.

□

Theorem 3.10. *Let $E = K(\theta)$ be a simple transcendental exponential extension of a differential field K with the same constant subfield as K, let $p = \sum p_i \theta^i$ be a polynomial in $K[\theta]$, and let $r = a/b$ be a rational function in $K(\theta)$ ($a, b \in K[\theta]$). Then:*

1. $\mathrm{Deg}_\theta\, p' = \mathrm{Deg}_\theta\, p$
2. *$p' \mid p$ if and only if p is monomial (i.e, has the form $p_i\theta^i$)*
3. *If an irreducible factor other than θ appears in r's denominator with multiplicity m, then it appears in r''s denominator with multiplicity $m+1$*
4. *$r' \in K$ if and only if $r \in K$*

Proof

Again,

$$p' = \sum_{i=0}^{n} (p_i'\theta^i + ip_i\theta'\theta^{i-1})$$

This time, however, $\theta' = k'\theta$, so

$$p' = \sum_{i=0}^{n} (p_i' + ip_ik')\theta^i$$

Assume that one of these coefficients, say $(p_i' + ip_ik')$, was zero but p_i was non-zero. Then $D(p_i\theta^i) = (p_i' + ip_ik')\theta^i = 0$, so $p_i\theta^i$ would be a constant, which must be in K, contradicting the transcendence of E. Therefore, none of these coefficients can be zero, establishing the first claim.

To establish the second claim, assume first that $p' \mid p$. Since p' has the same degree as p (by the first claim), it can only divide p if it has the form mp, where $m \in K$. Equating coefficients of θ in the above sums leads us to conclude that

$$m = (\frac{p_i'}{p_i} + ik')$$

If p was not monomial, then all of its coefficients must yield the same value for m, i.e,

$$m = (\frac{p_i'}{p_i} + ik') = (\frac{p_j'}{p_j} + jk')$$

$$p_i'p_j - p_ip_j' + (i-j)k'p_ip_j = 0$$

$$\frac{p_i'p_j - p_ip_j'}{p_j^2} + (i-j)k'\frac{p_i}{p_j} = 0$$

$$\left(\frac{p_i}{p_j}\right)' + (i-j)k'\frac{p_i}{p_j} = 0$$

Then $D(\frac{p_i}{p_j}\theta^{j-i}) = \left(\frac{p_i}{p_j}\right)' + (i-j)\frac{p_i}{p_j}k' = 0$, again contradicting the transcendence of E over K. So p must be monomial.

Conversely, if p is monomial, say $a\theta^n$, then $p' = (a' + nak')\theta^n = \frac{a'+nak'}{a}p$ and $p' \mid p$.

To prove the final two claims, we proceed as before, expanding r using partial fractions:

$$r = a_0(\theta) + \sum_{i=1}^{\mu}\sum_{j=1}^{m_i}\frac{a_{ij}(\theta)}{b_i(\theta)^j}$$

and taking the derivative:

$$r' = a_0'(\theta) + \sum_{i=1}^{\mu}\sum_{j=1}^{m_i}\left[\frac{a_{ij}'(\theta)}{b_i(\theta)^j} - \frac{j\,a_{ij}(\theta)\,b_i'(\theta)}{b_i(\theta)^{j+1}}\right]$$

θ is the only irreducible monomial, so if a b_i is not θ, then it will not be canceled by b_i', and again we'll have a single term on the R.H.S. with b_i^{j+1} in the denominator, so the L.H.S. must also have a b_i^{j+1} in its denominator.

This time, however, b_i' can divide b_i if b_i is monomial. When r' is in K', all other possibilities are excluded as before, so r must now have the form:

$$r = \sum_{i=-m}^{n} r_i\theta^i$$

where $r_i \in K$. We've already established that if r_i is non-zero, then the corresponding term in the derivative is also non-zero, so the only way for r' to be in K is if r is in K.

□

Example 3.11. Let $p = xe^x$. Then $\frac{d}{dx}xe^x = e^x + xe^x = (x+1)e^x = \frac{x+1}{x}e^x$.

We start with the rational function field $K = \mathbb{C}(x)$, and extend by the transcendental exponential $\theta = \exp(x)$ to form the ring $K[\theta]$. Both p and p' are in $K[\theta]$; $p = x\theta$ is monomial (in θ); note that $\frac{x+1}{x} \in K$, so $p' \mid p$ in this ring.

□

Theorem 3.12. *Let A be an algebraic extension of a differential field K with the same constant subfield as K, let σ be an automorphism of A/K, and let a be an element of A.*

1. $D(\sigma x) = \sigma(Dx)$,

2. $\mathrm{Tr}\,(Dx) = D(\mathrm{Tr}\,x)$,

3. $\mathrm{Tr}\,(\frac{Dx}{x}) = \frac{\mathrm{N}\,(x)'}{\mathrm{N}\,(x)}$

4. $a' \in K \leftrightarrow a \in K$.

Proof

1. Consider an automorphism σ of A/K, i.e, an automorphism of A that fixes the differential field K, so that $\sigma x = x$ for $x \in K$. Writing the minimal polynomial of x as $\sum_i a_i x^i = 0$, applying σ to this equation, and remembering that automorphism commutes with multiplication and addition and that $a_i \in K$, we obtain $\sum_i a_i(\sigma x)^i = 0$, i.e, σx has the same minimal polynomial as x; we say that σx is a *conjugate* of x.

Theorem 3.3 now gives us the derivation of σx:

$$D(\sigma x) = -\frac{\sum_i a_i'(\sigma x)^i}{\sum_i i a_i(\sigma x)^{i-1}}$$

Applying the operators in the other direction, however, and again using the fact that automorphism commutes with our field operators, we obtain:

$$\sigma(Dx) = \sigma\left(-\frac{\sum_i a_i' x^i}{\sum_i i a_i x^{i-1}}\right) = -\frac{\sum_i a_i'(\sigma x)^i}{\sum_i i a_i(\sigma x)^{i-1}}$$

i.e, $D(\sigma x) = \sigma(DX)$; automorphisms that fix the base field of an algebraic extension commute with derivation.

2. Now let's consider how σ interacts with Tr. Remember that trace, in a Galois extension, can be written as a sum over all automorphisms that fix the base field:

$$\mathrm{Tr}\,x = \sum_\sigma \sigma x$$

We extend A, if necessary, into a Galois extension, and use the commutation relationship we just proved to establish:

$$D(\mathrm{Tr}\,x) = D\left(\sum_\sigma \sigma x\right) = \sum_\sigma D(\sigma x) = \sum_\sigma \sigma(Dx) = \mathrm{Tr}\,(Dx)$$

3. Using the commutation relationship we just proved, along with the definitions of Tr and N :

$$\mathrm{Tr}\,(\frac{Dx}{x}) = \sum_\sigma \sigma\left(\frac{Dx}{x}\right) = \sum_\sigma \frac{\sigma Dx}{\sigma x} = \sum_\sigma \frac{D\sigma x}{\sigma x} = \frac{D\prod_\sigma \sigma x}{\prod_\sigma \sigma x} = \frac{D(\mathrm{N}\,(x))}{\mathrm{N}\,(x)}$$

4. The right-to-left implication is obvious (since differential fields are closed under derivation), so we need only to prove the left-to-right implication.

Consider a, with $Da \in K$, so $\mathrm{Tr}\,(Da) = nDa$, where n is the degree of the algebraic extension, by Theorem **??**. It follows that

$$Da = \frac{1}{n}\mathrm{Tr}\,(Da) = \frac{1}{n}D(\mathrm{Tr}\,a)$$

Since $\mathrm{Tr}\,a \in K$, we have identified an element in K with the same derivation as a, which therefore can differ from a solely by a constant. Since A and K have the same constant subfield, all of our constants are in K, so a is therefore also in K.

□

Example 3.13. Explain the "disappearance" of the square root in:

$$\int \frac{1}{\sqrt{1-x^2}} = \arcsin x$$

Finding $\arcsin x$ in the table, we see that:

$$\arcsin x = -i\,\ln(ix + \sqrt{1-x^2})$$

That's where it went! It "disappeared" into the complex logarithm that $\arcsin x$ is formed from. New logarithms, of course, are acceptable. Notice that the new logarithm has a constant coefficient $(-i)$, is not nested, and appears to the first power.

□

Notice the extra condition on the algebraic extension, that the extension has to preserve the constant subfield. The theorem would fail without this condition, as shown by numerous examples of roots appearing in integrals where only rational numbers were needed in the integrand. The simplest way to handle this situation is to use an algebraically closed constant subfield (like $\mathbb{C}$), but this is not always practical.

Example 3.14.

$$\begin{aligned}\int \frac{1}{x^2-2}dx &= \int \frac{1}{2\sqrt{2}}\left[\frac{1}{x-\sqrt{2}} - \frac{1}{x+\sqrt{2}}\right]dx \\ &= \frac{1}{2\sqrt{2}}\left[\ln(x-\sqrt{2}) - \ln(x+\sqrt{2})\right]\end{aligned}$$

This integrand can be expressed in $\mathbb{Q}(x)$, but the integral requires $\mathbb{Q}(x,\xi,\theta,\psi)$; ξ is algebraic with minimal polynomial $\xi^2 - 2 = 0$; θ and ψ are logarithmic transcendental with $\theta' = 1/(x-\xi)$ and $\psi' = 1/(x+\xi)$.

□

Finally, we want to prove the full Liouville theorem, establishing that the only new extensions that can appear in an integral are logarithmic ones.

Theorem 3.15. *(Liouville) Let L be an elementary extension of a differential field K with the same constant subfield as K. Then $\forall l \in L$, $l' \in K$ iff l has the form:*

$$k + \sum_{i=1}^{n} c_i \theta_i$$

where $k \in K$, $K(\theta_i)$ are simple logarithmic extensions of K and c_i are constants.

Proof

By the definition of an elementary extension, L has the form $K(t_1, \ldots, t_n)$ where each t_i is a simple elementary extension of $K(t_1, \ldots, t_{i-1})$.

We'll proceed by induction on the number of extensions n. Theorems 3.9(5), 3.10(4), and 3.12(4) establish the theorem for $n = 1$. So assume that the theorem is true for all $i < n$.

Let $M = K(t_1)$, so $L = M(t_2, \ldots, t_n)$, and the induction hypothesis implies that if $l' \in M$, then l has the form:

$$l = m_0 + \sum c_i \theta_i$$

where $\theta_i \in M$, so

$$l' = m_0' + \sum c_i \frac{m_i'}{m_i}$$

$l' \in K$, and m_0 and the various m_i are rational functions in $K(t_1)$. We can use our basic logarithm identities:

$$\frac{(ab)'}{ab} = \frac{a'}{a} + \frac{b'}{b} \qquad\qquad \frac{(\frac{1}{a})'}{\frac{1}{a}} = \frac{-\frac{a'}{a^2}}{\frac{1}{a}} = -\frac{a'}{a}$$

to reduce to the case where the m_i are all irreducible polynomials, so let's consider our three cases:

1. $K(t_1)$ is logarithmic over K.

   In this case, Theorem 3.9(4) states that if m_0 has a non-trivial denominator with an irreducible factor p, then p appears in m_0''s denominator with multiplicity at least two. Since $l' \in K$, this implies that the sum must contribute a denominator with the same multiplicity in order to achieve cancellation, so p must be one of the m_i's.

However, since p is irreducible, Theorem 3.9(4) implies that the multiplicity of p in the sum's denominator can be no more than one, so m_0 must be a polynomial.

Futuremore, there is no cancellation between a normal irreducible polynomial and its derivative, so none of the m_i's can be polynomials in $K[t_1]$; they must exist in K, since otherwise l' would have a non-trivial denominator in $K(t_1)$, and by hypothesis, $l' \in K$.

Finally, Theorem 3.9(5) states that m_0 must have the form $k_0 + c_0 t_1$. In other words, a simple logarithm extension can contribute a single term to the sum in the statement of the theorem.

2. $K(t_1)$ is exponential over K.

   This case is similar to the logarithmic one, except that we must now consider the possibility of special polynomials in the m_i's. Actually, there is only one special irreducible polynomial, t_1 itself. If one of the m_i's was t_1, then it would cancel with its derivative as follows:

$$\frac{t_1'}{t_1} = k'$$

   Since both l' and k' are in K, we can collect them together as follows:

$$l' - c_1 k' = m_0' + \sum c_i \frac{m_i'}{m_i}$$

   and proceed with the proof as before. This time, however, Theorem 3.10(4) tells us that m_0 can only have the form k_0, so exponential extensions contribute nothing to the sum in the statement of the theorem.

3. $M = K(t_1)$ is algebraic over K

   Applying the trace map to our induction equation:

$$\mathrm{Tr}\left(l'\right) = \mathrm{Tr}\left(m_0'\right) + \sum c_i \mathrm{Tr}\left(\frac{m_i'}{m_i}\right)$$

   Since $l' \in K$, $\mathrm{Tr}\left(l'\right) = dl'$, where d is the degree of the algebraic extension $K(t_1)$ over K, and:

$$dl' = \mathrm{Tr}\left(m_0\right)' + \sum c_i \frac{\mathrm{N}\left(m_i\right)'}{\mathrm{N}\left(m_i\right)}$$

$$l' = \left[\frac{\mathrm{Tr}\left(m_0\right)}{d}\right]' + \sum \frac{c_i}{d} \frac{\mathrm{N}\left(m_i\right)'}{\mathrm{N}\left(m_i\right)}$$

   Setting $k_0 = \frac{\mathrm{Tr}\left(m_0\right)}{d}$ and $k_i = \mathrm{N}\left(m_i\right)$, we see that l' can be written:

$$l' = k_0' + \sum \frac{c_i}{d} \frac{k_i'}{k_i}$$

   establishing that l has the form required by the theorem.

□

Chapter 4

Integration of Rational Functions

Since our strategy will be to reduce integrals in complex fields by stripping away their extensions and obtaining integrals in the simpler underlying fields, it follows that we should start this discussion by describing how to integrate in $\mathbf{C}(x)$, the field that underlies all the others.

Perhaps this seems pedantic. After all, didn't we go over all this in first year Calculus? Don't we already know everything we need to about integrating rational functions? We just factor the denominator, do a partial fractions expansion, plug in some simple known integrals, and we're done, right?

Not so fast. To begin with, there's that business of "just" factoring the denominator. As we've already seen, factoring a large polynomial can be quite a daunting undertaking. Techniques have been developed to avoid it as much as possible. Also, if you're seeing Liouville's theorem for the first time, then a whole new dimension to things like $\arctan$ open up when you regard them as complex logarithms. And finally, the multi-valued nature of complex logarithms and related functions make them very slippery little beasts. It's easy to get nonsense answers from the simplest calculations if you're not careful.

4.1 Logarithms and related functions

Let's start with a simple calculation, one we learned back in Calc I:

$$\int \frac{1}{x^2-1} dx = \arctan x$$

Now, we've already learned that the way to handle arctangents and the like is to convert them to E-L-R form, which for $\arctan$ is:

$$\arctan x = \frac{1}{2}\, i\, \ln \frac{ix-1}{ix+1}$$

Interesting, but not very illuminating. Nevertheless, as Sherlock Holmes was wont to say, "once you have eliminated the impossible…"

Let's examine the improbable remains in an attempt to find the truth:

$$\int \frac{1}{x^2-1}dx \;=\; \int \frac{1}{(x+i)(x-i)}dx$$

Applying one or another of our techniques for partial fractions expansion, we compute:

$$\begin{aligned}\frac{1}{(x+i)(x-i)} &= \frac{1}{2}\Big[\frac{i}{x+i}-\frac{i}{x-i}\Big]\\ &= -\frac{1}{2}i\Big[\frac{i}{-ix+1}-\frac{i}{-ix-1}\Big]\end{aligned}$$

On that last step, I multiplied through by 1 in the form $\frac{-i}{-i}$ for reasons that I'll explain later. But now, since each numerator is just the negative of its denominator's derivative, we proceed:

$$\begin{aligned}\int \frac{1}{x^2-1}dx &= -\frac{1}{2}i\int\Big[\frac{i}{-ix+1}-\frac{i}{-ix-1}\Big]dx\\ &= -\frac{1}{2}i\Big[-\ln(-ix+1)+\ln(-ix-1)\Big]\\ &= -\frac{1}{2}i\Big[\ln(-ix-1)-\ln(-ix+1)\Big]\end{aligned}$$

How do we evaluate something like $\ln(-ix-1)$? Well, Euler's identity is a good place to start:

$$\begin{aligned}e^{i\theta} &= i\sin\theta+\cos\theta\\ i\theta &= \ln[i\sin\theta+\cos\theta]\end{aligned}$$

We can always factor a complex number into its modulus and its angle:

$$a+bi \;=\; \sqrt{a^2+b^2}\Big(\frac{a}{\sqrt{a^2+b^2}}+\frac{b}{\sqrt{a^2+b^2}}i\Big)$$

Now, the expression in parenthesis on the right is just an x-y coordinate pair on the unit circle, which form the sine and cosine of an angle. What angle? Well, the tangent of the angle is going to be the ratio between the y (imaginary) and x (real) coordinates, so its $\frac{b}{a}$, and therefore the angle must be $\arctan\frac{b}{a}$. Combining this logic with the last two equations and the addition law of logarithms lets us obtain a general expression for imaginary logarithms:

$$\ln(a+bi) \;=\; \ln\sqrt{a^2+b^2} + i\arctan\frac{b}{a}$$

Plugging this back into our integral, and using the fact that $\arctan$ is an odd function, we conclude:

$$\begin{aligned}
\int \frac{1}{x^2-1}dx &= -\frac{1}{2}i\Big[\ln(-ix-1) - \ln(-ix+1)\Big] \\
&= -\frac{1}{2}i\Big[\big(\ln\sqrt{x^2+1} + i\arctan x\big) - \big(\ln\sqrt{x^2+1} + i\arctan(-x)\big)\Big] \\
&= -\frac{1}{2}i\Big[2i\arctan x\Big] \\
&= \arctan x
\end{aligned}$$

4.2 Multi-valued logarithms

The complex logarithm is a multi-valued function, since adding $2\pi i$ to any logarithm produces another power for the same value.

In *Symbolic Integration I*, Manuel Bronstein gave a detailed analysis, which was so enlightening to me that I will repeat and expand it here, of the following definite integral:

$$\int \frac{x^4-3x^2+6}{x^6-5x^4+5x^2+4}dx$$

```
(%i165) a: x^4-3*x^2+6;

                                        4      2
(%o165)                                 x  - 3 x  + 6
(%i166) b: x^6-5*x^4+5*x^2+4;

                                      6      4      2
(%o166)                               x  - 5 x  + 5 x  + 4
(%i167) gcd(b,diff(b,x));

(%o167)                                       1
```

```
(%i168) m: sylvester[a-z*diff(b,x),b,x];

         [ - 6 z    1    20 z    - 3   - 10 z    6       0       0       0       0     0 ]
         [                                                                               ]
         [   0    - 6 z    1    20 z    - 3    - 10 z    6       0       0       0     0 ]
         [                                                                               ]
         [   0      0    - 6 z    1     20 z    - 3    - 10 z    6       0       0     0 ]
         [                                                                               ]
         [   0      0      0    - 6 z    1      20 z    - 3    - 10 z    6       0     0 ]
         [                                                                               ]
         [   0      0      0      0    - 6 z     1      20 z    - 3    - 10 z    6     0 ]
         [                                                                               ]
(%o168)  [   0      0      0      0      0     - 6 z     1      20 z    - 3    - 10 z  6 ]
         [                                                                               ]
         [   1      0     - 5     0      5       0       4       0       0       0     0 ]
         [                                                                               ]
         [   0      1      0     - 5     0       5       0       4       0       0     0 ]
         [                                                                               ]
         [   0      0      1      0     - 5      0       5       0       4       0     0 ]
         [                                                                               ]
         [   0      0      0      1      0      - 5      0       5       0       4     0 ]
         [                                                                               ]
         [   0      0      0      0      1       0      - 5      0       5       0     4 ]
(%i169) d: ratsimp(determinant(m));

                                      6             4           2
(%o169)                       2930944 z  + 2198208 z  + 549552 z  + 45796
(%i170) e: gcd(d,diff(d,z));

                                        4           2
(%o170)                              732736 z  + 366368 z  + 45796
(%i171) gcd(e, diff(e,z));

                                           2
(%o171)                                183184 z  + 45796
(%i172) expand(gcd(e, diff(e,z)) / 45796);

                                           2
(%o172)                                4 z  + 1
(%i173) expand((4*z^2+1)^3*45796);

                                      6             4           2
(%o173)                       2930944 z  + 2198208 z  + 549552 z  + 45796
```

```
(%i225) ratexpand((-%i*256/963)*gcdex(a-(%i/2)*diff(b,x),b,x))[3];

                                       3       2
(%o225)                               x  + %I x  - 3 x - 2 %I
(%i226) ratexpand((%i*256/963)*gcdex(a+(%i/2)*diff(b,x),b,x))[3];

                                       3       2
(%o226)                               x  - %I x  - 3 x + 2 %I
```

$$\int \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx = \frac{i}{2} \ln(x^3 + ix^2 - 3x - 2i) - \frac{i}{2} \ln(x^3 - ix^2 - 3x + 2i)$$

$$= \tan^{-1}(\frac{x^3 - 3x}{x^2 - 2})$$

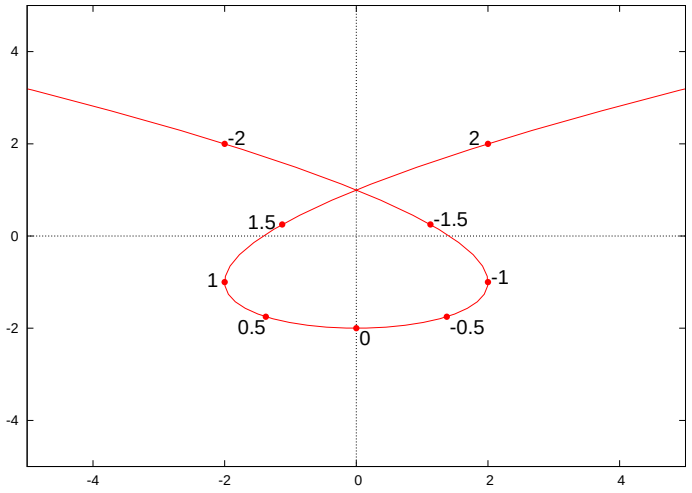


Figure 4.1: $(x^3 - 3x) + (x^2 - 2)i$

Let us note briefly that the integrand is clearly positive over the entire real line. Now, using a straightforward application of the method of partial fractions, we conclude:

$$\int \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx = \sum_{\alpha | 4\alpha^2 + 1 = 0} \alpha \log(x^3 + 2\alpha x^2 - 3x - 4\alpha)$$

Since the zeros of $4\alpha^2 + 1$ are $\alpha = \pm i/2$, we evaluate the definite integral using the indefinite integral, expand the complex logarithms, and obtain:

$$\int_1^2 \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx$$
$$= \left(\frac{i}{2}\log(2 + 2i) - \frac{i}{2}\log(2 - 2i)\right) - \left(\frac{i}{2}\log(-2 - i) - \frac{i}{2}\log(-2 + i)\right)$$
$$= -\frac{5\pi}{4} + \arctan(\frac{1}{2}) \approx -3.46$$

Since the integral was positive over the entire range of integration, this answer can not possibly be correct.

Alternately, we can apply the $\arctan$ identity from the last chapter and conclude:

$$\int \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx$$

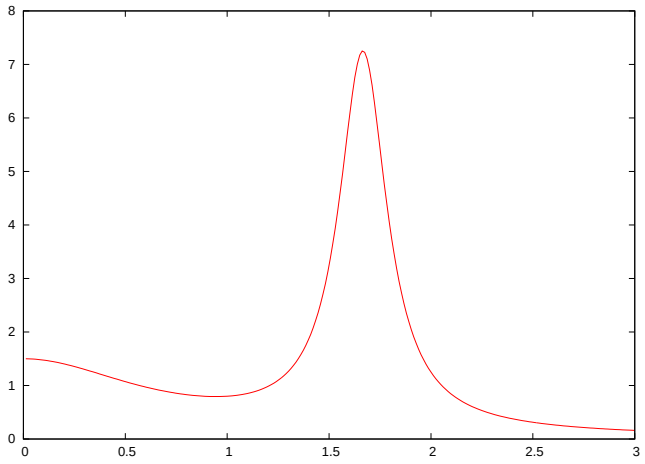
$$= \sum_{\alpha | 4\alpha^2 + 1 = 0} \alpha \log(x^3 + 2\alpha x^2 - 3x - 4\alpha)$$

$$= \frac{i}{2} \ln(x^3 + ix^2 - 3x - 2i) - \frac{i}{2} \ln(x^3 - ix^2 - 3x + 2i) = \frac{i}{2} \ln\left(\frac{x^3 - 3x + (x^2 - 2)i}{x^3 - 3x - (x^2 - 2)i}\right)$$

$$= \frac{i}{2} \ln\left(\frac{(x^3 - 3x)i - (x^2 - 2)}{(x^3 - 3x)i + (x^2 - 2)}\right) = \arctan\left(\frac{x^3 - 3x}{x^2 - 2}\right)$$

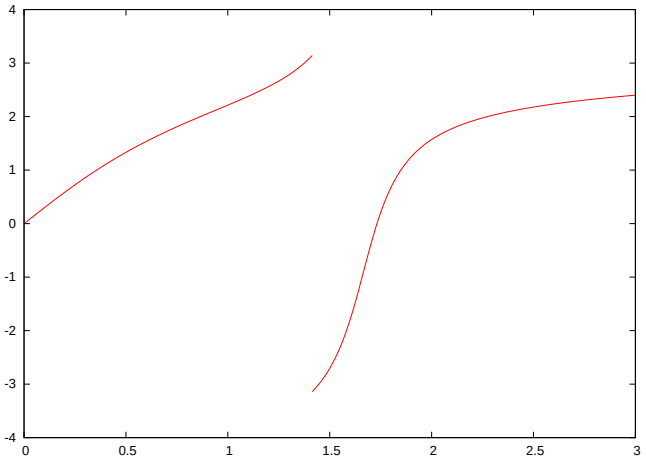
$$\int_1^2 \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx = \arctan 1 - \arctan 2 \approx -0.32$$

What went wrong? We gain a key insight, as is so often the case, by graphing first the integrand:



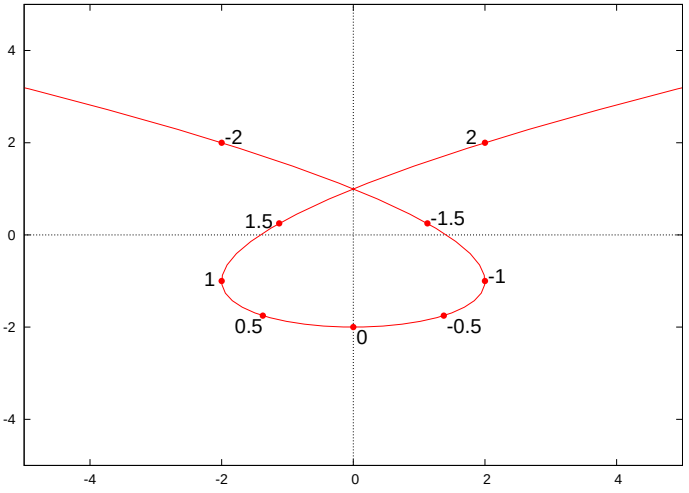
$$\frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4}$$

And now the (indefinite) integral:



$$\int \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx = \tan^{-1}\left(\frac{x^3 - 3x}{x^2 - 2}\right)$$

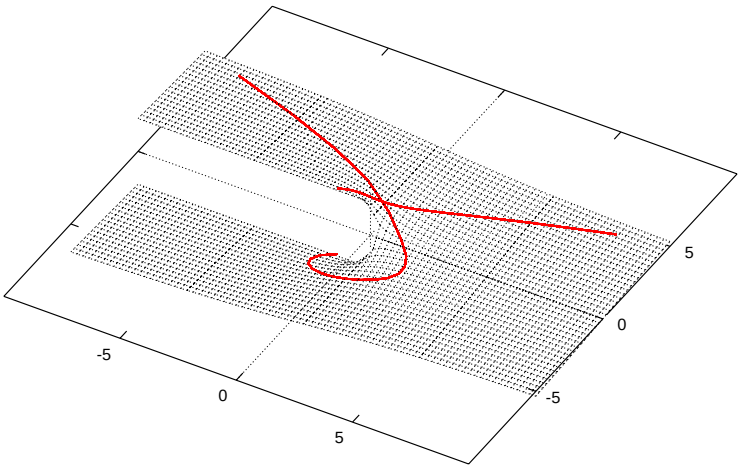
A discontinuity has appeared, seemingly out of nowhere. A closer inspection reveals that the break occurs suspiciously close to $\sqrt{2}$ — exactly where a plot of $(x^3 - 3x) + (x^2 - 2)i$ in the complex plane crosses the negative real axis:



OK, so here's what happened. When we evaluate a complex logarithm, we implicitly select one blade of $\ln$'s Riemann surface. There's no way to avoid this. $\ln$ is a multi-valued function, the different values corresponding to different blades of the Riemann surface, and if we want to actually obtain a numerical result from $\ln$ we need to pick one of those values. In the context of an indefinite integral, which is only defined within an unspecified additive constant, the choice is completely arbitrary.

So far, so good. Yet now we want a definite integral. Now we're going to evaluate not just a single logarithm, but we're going to trace out a curve along the Riemann surface. As I noted above, if we assign a single fixed value to $\ln x$, then there's no way to avoid a discontinuity *somewhere* in the Riemann surface. In this specific case, we used the identity $\ln(a+bi) = \ln\sqrt{a^2+b^2} + i\arctan\frac{b}{a}$, which just converts the discontinuity in $\ln$ to one expressed in terms of $\arctan$, also a multi-valued function. See, there's no way to completely avoid this.

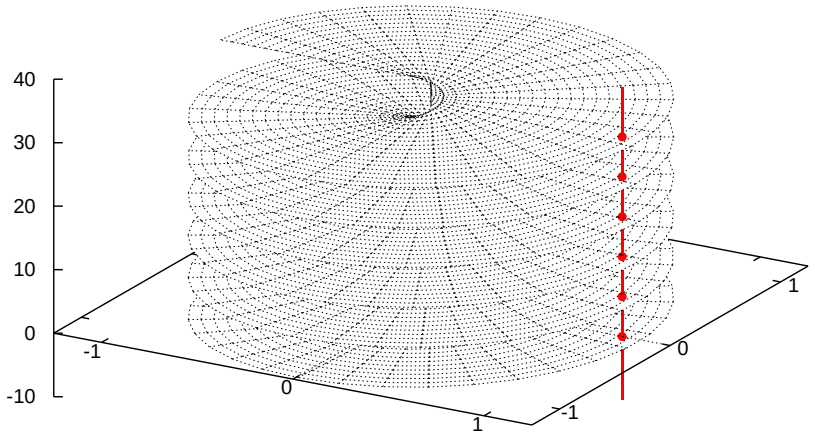
Where's the discontinuity in $\arctan$? Typically, the function is defined so that it ranges from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ and is continuous over finite numbers, so its discontinuity is where its argument becomes infinite; i.e, where its denominator goes to zero. In the $\ln(a+bi)$ expansion, this occurs where a becomes zero; i.e, where $a+bi$ crosses the negative real axis. This is easier to visualize if we reproduce that last graph in 3-D, superimposing $(x^3-3x)+(x^2-2)i$ on $\arctan$'s Riemann surface:



The z-axis is now the value of $\arctan$. Along the x-axis, it is zero, and it grows positive as we move counter-clockwise, and negative as we move clockwise. The resulting discontinuity along the negative real axis, where $\arctan$ jumps from $\frac{\pi}{2}$ to $-\frac{\pi}{2}$, clearly creates a matching discontinuity in the plot of $(x^3-3x)+(x^2-2)i$.

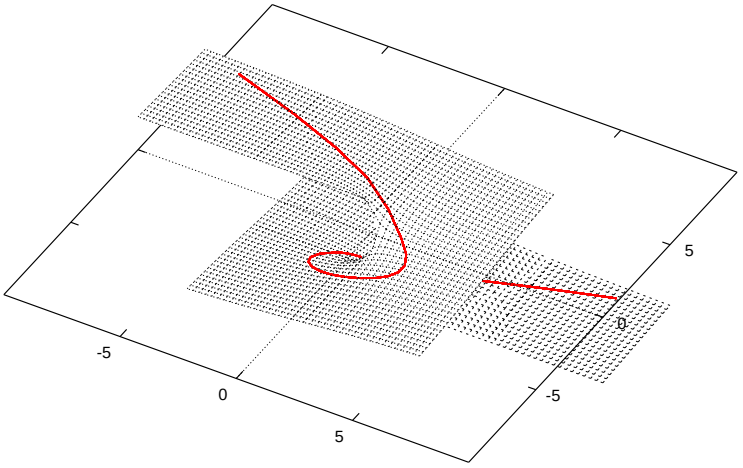
Of course, this isn't really $\arctan$'s Riemann surface, only a slice of it, and now we begin to find our solution. $\arctan$'s complete Riemann surface looks like an infinite screw, with an infinite series of blades, each spaced 2π apart in the z-direction. So, when we plot our function, what we really want is something more like this, moving smoothly along the Riemann surface without any discontinuity.

Alas, while easy enough to graph, and easy enough to understand once you're thinking about the continuity of a Riemann surface, this is just asking a bit much from poor old $\ln$



(or $\arctan$). There's no way for the function to know which result is needed based solely on a single value as the argument.

However, there is a way out, at least in the case of rational functions. The method, due to R. Rioboo, is to take advantage of two things. First, the addition law of logarithms ($\ln ab = \ln a + \ln b$), combined with Euler-?? interpretation of complex numbers, lets us split a complex logarithm into two logarithms, each of which require only half the range of angles of the original logarithm. Second, since a rational function only intersects the real axis in a finite number of points (the finite number of zeros of its real component), a finite (and easily computed) number of reductions converts the logarithm into a sum of logarithms, none of whose arguments cross the real axis.



$$\int \frac{x^4-3x^2+6}{x^6-5x^4+5x^2+4}dx = \frac{i}{2}\ln(x^3+ix^2-3x-2i) - \frac{i}{2}\ln(x^3-ix^2-3x+2i)$$

```
(%i205) gcdex(x^3-3*x,x^2-2);
```

```
                                                    2
                                               x   x  - 1
(%o205)/R/                                [- -, ------, 1]
                                               2     2
```

```
(%i214) expand(((x^3-3*x)+(x^2-2)*%i)*((x^2-1)-x*%i));
```

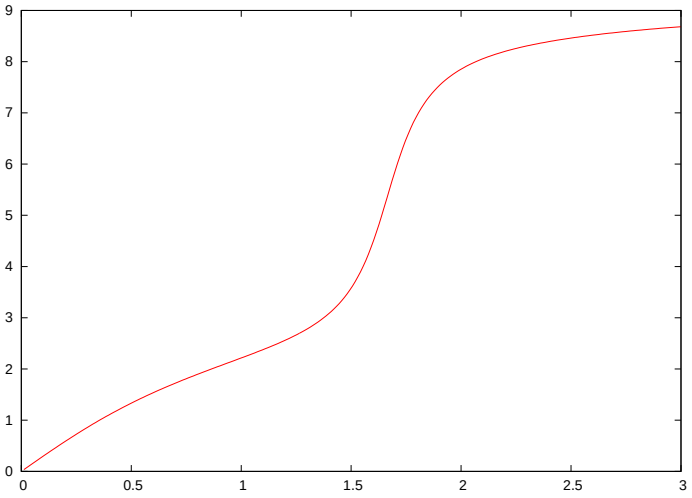
```
                                             5       3
(%o214)                                     x  - 3 x  + x + 2 %I
```

$$\ln\Big[(x^3-3x)+(x^2-2)i\Big] = \ln\Big[x^5-3x^3+x+2i\Big] - \ln\Big[(x^2-1)-xi\Big]$$

$$\ln(x^3+ix^2-3x-2i) = \ln(x^5-3x^3+x+2i) - \ln(x^3-i) + \ln(x+i)$$

$$\ln(x^3-ix^2-3x+2i) = \ln(x^5-3x^3+x-2i) - \ln(x^3+i) + \ln(x-i)$$

$$\int \frac{x^4-3x^2+6}{x^6-5x^4+5x^2+4}dx = \tan^{-1}(\frac{x^5-3x^2+x}{2}) + \tan^{-1}(x^3) + \tan^{-1}(x)$$


9
8
7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
2.5
3

4.3 A Bit Of Perspective

Now, something interesting happened during the course of this chapter. Something subtle enough that I didn't notice it for the first year or so that I worked with this theory. Maybe you're more cleaver than I am and have already seen it (or maybe I just organized this book well enough that it's more obvious to you that it was to me).

How did we start with a problem that was pure algebra, and end up with a solution that involved topology?

Let's review. We defined a differential field using mapping between elements as the derivation. No topology yet — the field was purely discrete, with no concept of "closeness" between the elements. We went looking for an element that got mapped onto some specified element. Again, no topology. And finally, we extended to a logarithmic extension, defined purely as a transcendental extension with a specific differential mapping. Still no topology.

So why have we spent the last ? pages discussing topology?

The change happened when we went from regarding "$\ln x$" as a transcendental element over the field $\mathbf{C}(x)$ to regarding it as $\ln(x)$, a function mapping one complex number to another...

In conclusion, we need to be careful when working with multi-valued functions like $\ln$ and $\arctan$, but the concept of a Riemann surface provides an important conceptual tool for dealing with them. Liouville's theorem tells us that additional logarithms can be introduced by integration, and the multi-valued nature of the complex logarithm leaves us with a choice as to which branch of the function (or blade of the Riemann surface) to use. Yet fundamental principles of calculus tell us that an indefinite integral is only defined within a constant of integration, so it doesn't matter exactly which value of the logarithm we choose to use. Having made that choice, however, we then need to remain consistent by preserving continuity on the Riemann surface during the evaluation of any definite integral. In the specific case of rational function integration, Rioboo's method gives us an algorithm that automatically preserves this continuity. For more complicated integrals that introduce new logarithms, we apply the more general concept of continuity on the Riemann surface.

Chapter 5

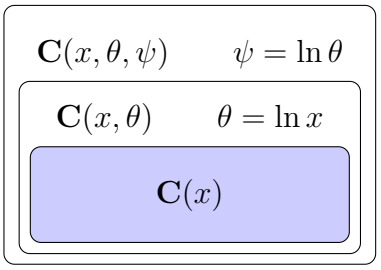
The Logarithmic Extension

When we express a function in Liouvillian form, we construct a tower of nested fields, starting with $\mathbb{C}(x)$ at the bottom, and building up to the extension required to express our integral.

For example, consider example 5.7:

$$\int \left[(\ln(\ln x))^2 \ln x - \frac{2}{\ln x}(\ln(\ln x) + 1) \right] dx$$

To express this integral, we need to construct $\ln(\ln x)$, which requires $\ln x$ to be constructed first. Thus, we obtain nested fields according to the following structure:



Our integrand can now be expressed as a rational function in the top-most field:

$$\int \frac{\theta^2\psi^2 - 2\psi - 2}{\theta} dx$$

Our basic strategy for integration is to always work in the top most field extension, reducing to some kind of problem that must be solved in the next lower extension. Since we only have three basic types of field extension, our aim is to develop a theory to handle each type.

In this chapter, we'll analyze the logarithmic extension.

For logarithmic extensions, the problem is particularly easy, since integration leads only to futher integration steps in the base field.

5.1 The Logarithmic Integration Theorem

Theorem 5.1. *Let K be a differential field, let $K(\theta = \ln k)$ be a simple logarithmic extension of K, let $n_i(\theta)$ be irreducible polynomials in $K[\theta]$, and let f be an element of $K(\theta)$ with partial fractions expansion:*

$$f = \sum_{i=0}^{n} a_i \theta^i + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{b_{i,j}(\theta)}{n_i(\theta)^j} \tag{5.1}$$

$$a_i \in K \qquad b_{i,j}(\theta), n_i(\theta) \in K[\theta]$$

If f has an elementary anti-derivative F, then $F \in K(\theta, \Psi)$, where $K(\theta, \Psi)$ is a finite logarithmic extension of $K(\theta)$, F has a partial fractions expansion of the form:

$$F = c_{n+1}\theta^{n+1} + \sum_{i=1}^{n} \left[A_i + c_i\right] \theta^i + A_0 + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i - 1} \frac{B_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\nu} C_i \ln n_i(\theta) \tag{5.2}$$

$$A_0 \in K(\Psi) \qquad \underset{i \neq 0}{A_i \in K} \qquad B_{i,j}(\theta), n_i(\theta) \in K[\theta] \qquad c_i' = C_i' = 0$$

and the following relationships hold:

$$\left[A_n + (n+1)c_{n+1}\theta\right]' = a_n \tag{5.3a}$$

$$\left[A_i + (i+1)c_{i+1}\theta\right]' = a_i - (i+1)\frac{k'}{k}A_{i+1} \qquad {\scriptstyle 0 \leq i < n} \tag{5.3b}$$

$$R_{i,m_i-1}(\theta) = b_{i,m_i}(\theta) \tag{5.4a}$$

$$R_{i,j}(\theta) = b_{i,j+1}(\theta) - B_{i,j+1}'(\theta) - Q_{i,j+1}(\theta) \qquad {\scriptstyle 1 \leq j < m_i - 1} \tag{5.4b}$$

$$B_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{j n_i'(\theta)} \mod n_i(\theta) \tag{5.5}$$

$$Q_{i,j}(\theta) = -\frac{R_{i,j}(\theta) + j B_{i,j}(\theta) n_i'(\theta)}{n_i(\theta)} \tag{5.6}$$

$$C_i = \frac{b_{i,1}(\theta) - B_{i,1}'(\theta) - Q_{i,1}(\theta)}{n_i'(\theta)} \qquad {\scriptstyle m_i > 1} \tag{5.7a}$$

$$C_i = \frac{b_{i,1}(\theta)}{n_i'(\theta)} \qquad {\scriptstyle m_i = 1} \tag{5.7b}$$

Proof

By Theorem 3.15, an elementary antiderivative of f can only exist in a finite logarithm extension $K(\theta, \Psi)$ of $K(\theta)$ and therefore must have the form:

$$F = R + \sum_{i=1}^{\eta} C_i \Psi_i$$

where $R \in K(\theta)$, the C_i are constants, and the Ψ_i are logarithms.

We can perform a partial fractions expansion on R:

$$F = \sum_{i=0}^{N} A_i \theta^i + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{B_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\eta} C_i \Psi_i$$

It will be convenient later to separate a constant from the A_i terms, so let's do that now:

$$F = \sum_{i=0}^{N} (A_i + c_i) \theta^i + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{B_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\eta} C_i \Psi_i$$

Our basic logarithmic relationships:

$$\ln ab = \ln a + \ln b \qquad\qquad \ln \frac{a}{b} = \ln a - \ln b$$

allow us to assume, without loss of generality, that the Ψ_i's are logarithms of irreducible polynomials. Some of those irreducible polynomials will exist solely in our underlying field K, and those we collapse into A_0, noting that this makes A_0 unique among the A_i's because it can include additional logarithms. The remaining irreducible polynomials (those involving θ) can be identified as $n_i(\theta)$'s, simply by increasing i and adding new $n_i(\theta)$'s if needed.

Now let's differentiate, remembering that $\theta' = \frac{k'}{k}$:

$$F' = \sum_{i=0}^{N} \left[A_i' \theta^i + i \frac{k'}{k} (A_i + c_i) \theta^{i-1} \right] + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{B_{i,j}'(\theta) n_i(\theta) - j B_{i,j}(\theta) n_i'(\theta)}{n_i(\theta)^{j+1}} + \sum_{i=1}^{\nu} C_i \frac{n_i'(\theta)}{n_i(\theta)}$$

$$F' = A'_N\theta^N + \sum_{i=0}^{N-1}\left[A'_i + (i+1)\frac{k'}{k}(A_{i+1} + c_{i+1})\right]\theta^i$$
$$+ \sum_{i=1}^{\nu}\left[\frac{-M_i B_{i,M_i}(\theta)n'_i(\theta)}{n_i(\theta)^{M_i+1}} + \sum_{j=1}^{M_i-1}\frac{B'_{i,j+1}(\theta) - jB_{i,j}(\theta)n'_i(\theta)}{n_i(\theta)^{j+1}} + C_i\frac{n'_i(\theta)}{n_i(\theta)}\right]$$

F' has the form of a partial fractions decomposition, but it is not a partial fractions decomposition because the numerators in the B terms violate the partial fractions degree bounds. The problem is that the product $B_{i,j}(\theta)n'_i(\theta)$ might have degree greater than $\deg_\theta n_i(\theta)$. To fix this, let's divide the $-jB_{i,j}(\theta)n'_i(\theta)$ terms by $n_i(\theta)$ (think polynomial long division) and rewrite them as a quotient and a remainder:

$$-jB_{i,j}(\theta)n'_i(\theta) = Q_{i,j}(\theta)n_i(\theta) + R_{i,j}(\theta)$$

This fixes the B terms, since $\deg Q_{i,j}(\theta) < \deg n_i(\theta)$ and $\deg R_{i,j}(\theta) < \deg n_i(\theta)$.

$$F' = A'_N\theta^N + \sum_{i=0}^{N-1}\left[A'_i + (i+1)\frac{k'}{k}(A_{i+1} + c_{i+1})\right]\theta^i$$
$$+\sum_{i=1}^{\nu}\left[\frac{R_{i,M_i}(\theta)}{n_i(\theta)^{M_i+1}} + \sum_{j=1}^{M_i-1}\frac{B'_{i,j+1}(\theta) + Q_{i,j+1}(\theta) + R_{i,j}(\theta)}{n_i(\theta)^{j+1}} + \frac{B'_{i,1}(\theta) + Q_{i,1}(\theta) + C_i n'_i(\theta)}{n_i(\theta)}\right]$$

What is the degree of F'? It's N, if A_N is not constant. If A_N is constant and not zero, then the degree of F' is $N-1$, since otherwise the $N-1$ coefficient would be zero:

$$A'_{N-1} + N\frac{k'}{k}A_N = 0 \qquad \Longrightarrow \qquad A'_{N-1} = (-NA_N)\frac{k'}{k}$$

Since A_N is constant, this could only be satisfied by $A_{N-1} = -NA_N\theta$, contradicting the assumption that $A_{N-1} \in K$.

Performing a partial fractions decomposition of f:

$$f = \sum_{i=0}^{n} a_i\theta^i + \sum_{i=1}^{\nu}\sum_{j=1}^{m_i}\frac{b_{i,j}(\theta)}{n_i(\theta)^j}$$

setting $F' = f$ and equating like terms, we establish the relationships listed in the statement of the theorem.

For the degree of F' to be n, either $\deg_\theta F = n+1$ and A_{n+1} is constant, or $\deg_\theta F = n$.

Since the highest order denominators in F' have order $M_i + 1$, and they must match with f's denominators of order m_i, we conclude that $M_i = m_i - 1$.

To establish the remaining relationships, let's remember the definition of $R_{i,j}$ and $Q_{i,j}$:

$$-jB_{i,j}(\theta)n_i'(\theta) = Q_{i,j}(\theta)n_i(\theta) + R_{i,j}(\theta)$$

Reducing this equation modulo $n_i(\theta)$, we obtain:

$$-jB_{i,j}(\theta)n_i'(\theta) \equiv R_{i,j}(\theta) \mod n_i(\theta)$$

Now we use the fact that $n_i(\theta)$ is *irreducible*, and invoke Theorem ??, which states the quotient ring modulo a prime ideal is a field, so we can perform division:

$$B_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{jn_i'(\theta)} \mod n_i(\theta)$$

This equation seems to identify $B_{i,j}(\theta)$ up to a multiple of $n_i(\theta)$, but if we remember our degree bound on partial fractions expansions, $\deg_\theta B_{i,j}(\theta) < \deg_\theta n_i(\theta)$, we see that in fact we've completely determined $B_{i,j}(\theta)$ from $R_{i,j}(\theta)$.

□

A few comments are in order. First, let's recall equations (5.3a) and (5.3b):

$$[A_n + (n+1)c_{n+1}\theta]' = a_n \tag{5.3a}$$

$$[A_i + (i+1)c_{i+1}\theta]' = a_i - (i+1)\frac{k'}{k}A_{i+1} \tag{5.3b}$$

These equations both require integration of the right hand side in the underlying field K. The resulting integral must take the form of a element of K (that's the A_i), plus possibly a constant times a single additional logarithm, θ itself. Liouville's theorem 3.15 tells us that integrals can generally include an arbitrary number of additional logarithms. The integrations required by (5.3a) and (5.3b) are more restrictive; the only additional logarithm they can include is θ, which makes sense because θ is not part of the underlying field K, but could easily appear in the final result because it already appears in the original integrand.

The exception is A_0, since it can include arbitrary additional logarithms, not just θ. It's version of equation (5.3b) reads:

$$[A_0 + c_1\theta]' = a_0 - \frac{k'}{k}A_1 \tag{5.8}$$

So, this integral can include a logarithm θ, just like the others, as well as additional logarithms collapsed into the A_0 term. The c_0 term, by the way, does not appear in any of the theorem's equations, but a moment's thought shows that c_0 is merely the constant of integration.

Example 5.2. Compute $\int \frac{1}{x \ln x} \mathrm{d}x$

Operating in $\mathbf{C}(x, \theta = \ln x)$, we evaluate:

$$\int \frac{1}{\theta x} \, \mathrm{d}x = \int \frac{\frac{1}{x}}{\theta} \, \mathrm{d}x$$

This has the form of equation (5.1) with $n_1(\theta) = \theta$, $m_1 = 1$ and $b_{1,1}(\theta) = \frac{1}{x}$.

$$C_1 = \frac{b_{1,1}(\theta)}{n_1'(\theta)} = \frac{\frac{1}{x}}{\frac{1}{x}} = 1$$

Plugging C_1 into equation (5.2) we get:

$$\int \frac{1}{x \ln x} \mathrm{d}x = \ln n_1(\theta) = \ln \ln x$$

□

Example 5.3. Compute $\int \ln x \, \mathrm{d}x$

Again we'll use $\mathbf{C}(x, \theta = \ln x)$

$$\int \theta \, \mathrm{d}x$$

This has the form of equation (5.1) with $n = 1$ and $a_1 = 1$, so

$$[A_1 + 2c_2\theta]' = a_1 = 1$$

$$A_1 + 2c_2\theta = x$$

Since $A_1 \in \mathbf{C}(x)$, it can not involve θ, so $A_1 = x$ and $c_2 = 0$.

$$[A_0 + c_1\theta]' = a_0 - \frac{k'}{k} A_1 = 0 - \frac{1}{x} x = -1$$

$$A_0 + c_1\theta = -x$$

So $A_0 = -x$ and $c_1 = 0$. Plugging A_0, A_1, c_1 and c_2 into equation (5.2) we get:

$$\int \theta \, \mathrm{d}x = x\theta - x$$

$$\int \ln x \, \mathrm{d}x = x \ln x - x$$

□

Example 5.4. Compute $\int \tan^{-1} x \, \mathrm{d}x$

Using the differential algebra identity $\tan^{-1} x = \frac{1}{2}\, i \, \ln \frac{ix-1}{ix+1}$ from Section 3.2, we use the differential field $\mathbb{C}(x, \theta)$; $\theta = \ln \frac{ix-1}{ix+1}$; $\theta' = -\frac{2i}{x^2+1}$ and compute

$$\int \frac{1}{2}\, i\, \theta \, \mathrm{d}x$$

This is in the form of equation (5.1) with $n = 1$ and $a_1 = \frac{1}{2}i$, so Theorem 5.1 tells us that the θ-degree of our integral is at most two.

$$[A_1 + 2c_2\theta]' = a_1 = \frac{1}{2}\, i$$

$$A_1 + 2c_2\theta = \frac{1}{2}\, ix$$

Since $A_1 \in \mathbb{C}(x)$, c_2 is zero and $A_1 = \frac{1}{2}ix$.

$$[A_0 + c_1\theta]' = a_0 - A_1\frac{k'}{k} = -\frac{1}{2}ix\frac{-2i}{x^2+1}$$

$$[A_0 + c_1\theta]' = -\frac{x}{x^2+1}$$

$$A_0 + c_1\theta = -\frac{1}{2}\ln(x^2+1)$$

Remembering that A_0 can include new logarithmic extensions, we conclude that

$$c_1 = 0 \qquad A_0 = -\frac{1}{2}\ln(x^2+1)$$

and therefore our solution is:

$$\int \frac{1}{2}\, i\, \theta \, \mathrm{d}x = \frac{1}{2}\, i\, x\theta - \frac{1}{2}\ln(x^2+1)$$

$$\int \tan^{-1} x \, \mathrm{d}x = x\, \tan^{-1} x - \frac{1}{2}\ln(x^2+1)$$

□

Example 5.5. Determine if $\sum_{n=0}^{\infty} \frac{1}{n^2} x^n$ is an elementary function.

We can differentiate the series term wise, and if the resulting series can be identified with an elementary function, then we need only to decide if that derivative can be integrated into an elementary function. Using a standard identity for the Taylor series of $\ln(1-x)$, we determine that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[\sum_{n=0}^{\infty} \frac{1}{n^2} x^n\right] = \sum_{n=1}^{\infty} \frac{1}{n} x^{n-1} = \frac{1}{x}\ln(1-x)$$

so we need to determine if $\int \frac{1}{x}\ln(1-x)\,\mathrm{d}x$ is elementary. Working in the differential field $\mathbb{C}(x, \theta = \ln(1-x))$, we're trying to integrate $\frac{1}{x}\theta\,\mathrm{d}x$. Equation (5.3a) reads:

$$[A_1 + 2c_2\theta]' = \frac{1}{x}$$

$$A_1 + 2c_2\theta = \ln x$$

$\ln x$ is required to express A_1, but new logarithms are not allowed at this point in the algorithm. Therefore, $\sum_{n=0}^{\infty} \frac{1}{n^2} x^n$ is not elementary.

□

Example 5.6. Compute

$$\int \frac{x\{(x^2e^{2x^2} - \ln^2(x+1))^2 + 2xe^{3x^2}(x - (2x^3+2x^2+x+1)\ln(x+1))\}}{(x+1)(\ln^2(x+1) - x^2e^{2x^2})^2}dx$$

```
(%i36)  integrand:
           x*((x^2*exp(2*x^2)-log(x+1)^2)^2
              +2*x*exp(3*x^2)*(x-(2*x^3+2*x^2+x+1)*log(x+1)))
           / ((x+1)*(log(x+1)^2 - x^2*exp(2*x^2))^2);

(%o36)
```

$$\frac{x\left(2\,x\,e^{3\,x^2}\,\left(x-\left(1+x+2\,x^2+2\,x^3\right)\,\log\left(1+x\right)\right)+\left(x^2\,e^{2\,x^2}-\log^2\left(1+x\right)\right)^2\right)}{\left(x+1\right)\,\left(\log^2\left(1+x\right)-x^2\,e^{2\,x^2}\right)^2}$$

We begin by putting our integral into Liouvillian form, assigning $\psi = \exp(x^2)$ and $\theta = \ln(x+1)$ to obtain:

$$\int \frac{x\{(x^2\psi^2 - \theta^2)^2 + 2x\psi^3(x - (2x^3+2x^2+x+1)\theta)\}}{(x+1)(\theta^2 - x^2\psi^2)^2}dx$$

```
(%i37)  lintegrand:
           subst([log(x+1) = theta,
                  exp(x^2) = psi,
                  exp(2*x^2) = psi^2,
                  exp(3*x^2) = psi^3],
              integrand);

(%o37)
```

$$\frac{x\left(2\,\psi^3\,x\,\left(x-\theta\,\left(1+x+2\,x^2+2\,x^3\right)\right)+\left(\psi^2\,x^2-\theta^2\right)^2\right)}{\left(x+1\right)\,\left(\theta^2-\psi^2\,x^2\right)^2}$$

Since ψ and θ are both simple extensions of $\mathbb{C}(x)$, we can operate in either $\mathbb{C}(x,\psi,\theta)$ or $\mathbb{C}(x,\theta,\psi)$. Since $\mathbb{C}(x,\theta,\psi)$ is an exponential extension of $\mathbb{C}(x,\theta)$, working in $\mathbb{C}(x,\theta,\psi)$ would require evaluating an integral in an exponential extension, which we won't study until the next chapter. Therefore, we'll work in $\mathbb{C}(x,\psi,\theta)$, a logarithmic extension of $\mathbb{C}(x,\psi)$, and hope that we won't have to do anything too complicated in $\mathbb{C}(x,\psi)$!

To use Maxima for this calculation, let's begin by defining a function that performs our differentiation.

```
(%i38)  depends([theta,psi], x)$

(%i39)  D(f) := subst(['diff(psi,x) = 2*x*psi,
                       'diff(theta,x) = 1/(x+1)],
                      diff(f,x))$
```

We need to put our integrand into partial fractions form, so let's begin by using Maxima's `divide` function, which performs polynomial long division with respect to a specific variable, and obtain:

```
(%i40)  [a, N]:
           divide(num(lintegrand),
                  denom(lintegrand), theta);

(%o40)
```

$$\left[\frac{x}{x+1}, \theta\left(-2\,\psi^3\,x^2-2\,\psi^3\,x^3-4\,\psi^3\,x^4-4\,\psi^3\,x^5\right)+2\,\psi^3\,x^3\right]$$

$$\int \frac{x}{x+1}+\frac{(-4x^5-4x^4-2x^3-2x^2)\psi^3\theta+2x^3\psi^3}{(x+1)(\theta^2-x^2\psi^2)^2}dx$$

The $a_0 = \frac{x}{x+1}$ term is easy.

```
(%i41)  A : integrate(a, x);

(%o41)
```

$$x-\log\left(1+x\right)$$

Let's consider the fractional term. Polynomial factorization can be difficult, but this denominator is easy – it's just a difference of squares: $(\theta^2 - x^2\psi^2) = (\theta - x\psi)(\theta + x\psi)$. Maxima's `partfrac` function performs partial fractions expansion with respect to a given variable, in this case θ.

```
(%i42)  pf: partfrac(N/denom(lintegrand), theta);

(%o42)
```

$$\frac{1}{\left(2\,x+2\right)\,\left(\psi\,x+\theta\right)}-\frac{1}{\left(2\,x+2\right)\,\left(\theta-\psi\,x\right)}+\frac{\left(\psi^2+\psi\right)\,x+\psi^2\,x^2+2\,\psi^2\,x^3+2\,\psi^2\,x^4}{\left(2\,x+2\right)\left(\theta+\psi\,x\right)^2}-\frac{(\psi$$

$$\int \frac{x}{x+1} + \frac{\frac{2x^4+2x^3+x^2+x}{2(x+1)}\psi^2 + \frac{x}{2(x+1)}\psi}{(\theta+x\psi)^2} + \frac{\frac{1}{2(x+1)}}{(\theta+x\psi)} - \frac{\frac{2x^4+2x^3+x^2+x}{2(x+1)}\psi^2 - \frac{x}{2(x+1)}\psi}{(\theta-x\psi)^2} - \frac{\frac{1}{2(x+1)}}{(\theta-x\psi)} dx$$

Next, we want to extract the irreducible factors from the denominators of our partial fractions expansion. To facilitate this, let's start by defining some helper functions:

```
(%i43)  is_power(expr) := is(part(expr, 0) = "^")$

(%i44)  base(e) :=
           if is_power(e) then part(e,1) else e$

(%i45)  power(e) :=
           if is_power(e) then part(e,2) else 1$
```

Now we can extract the portions of the denominators that involve the variable θ:

```
(%i46)  den(pl,v) := partition(denom(pl),v)[2]$

(%i47)  denoms: map(lambda([u], den(u,theta)),
                    partlist(pf));

(%o47)
```

$$\left[(\theta-\psi\,x)^2, (\theta+\psi\,x)^2, \theta-\psi\,x, \psi\,x+\theta\right]$$

We extract the irreducible factors:

```
(%i48)  n: unique(map(base, denoms));

(%o48)
```

$$[\theta-\psi\,x, \psi\,x+\theta]$$

Now, for each fraction in the partial fractions expansion, we determine which irreducible factor is in its denominator, to what power it appears, and use this information to assign the rest of the fraction to the variable $b_{i,j}$:

```
(%i49)  map(lambda([u], which(n, base(u))), denoms);
```

```
(%o49)
```

$$[1, 2, 1, 2]$$

```
(%i50)  map(power, denoms);

(%o50)
```

$$[2, 2, 1, 1]$$

```
(%i51)  map(lambda([u],
               b[which(n, base(den(u,theta))),
                 power(den(u,theta))]: u*den(u,theta)),
            partlist(pf))$

(%i52)  displayarray(b)$
```

$$b_{1,1} = -\frac{1}{2\,x+2}$$

$$b_{1,2} = -\frac{\left(\psi^2-\psi\right)\,x+\psi^2\,x^2+2\,\psi^2\,x^3+2\,\psi^2\,x^4}{2\,x+2}$$

$$b_{2,1} = \frac{1}{2\,x+2}$$

$$b_{2,2} = \frac{\left(\psi^2+\psi\right)\,x+\psi^2\,x^2+2\,\psi^2\,x^3+2\,\psi^2\,x^4}{2\,x+2}$$

We have two irreducible factors: $n_1(\theta) = \theta - x\psi$ and $n_2(\theta) = \theta + x\psi$. Each has θ-degree 1, so the numerators in our partial fractions expansions have θ-degree 0, as expected.

Now we're ready to apply Theorem 5.1. Let's start with $n_1(\theta)$. $m_1 = 2$, so

$$R_{1,1}(\theta) = b_{1,2}(\theta) = \frac{2x^4+2x^3+x^2+x}{2(x+1)}\psi^2 + \frac{x}{2(x+1)}\psi$$

```
(%i53)  R[1,1] : b[1,2];
```

```
(%o53)
```

$$-\frac{\left(\psi^2-\psi\right)\,x+\psi^2\,x^2+2\,\psi^2\,x^3+2\,\psi^2\,x^4}{2\,x+2}$$

and we wish to compute

$$B_{1,1}(\theta) \equiv -\frac{R_{1,1}(\theta)}{n_1'(\theta)} \mod n_1(\theta)$$

As modulo calculations go, this one is easy.

```
(%i54)  B[1,1] ::: - R[1,1] / D(n[1]);

(%o54)
```

$$-\frac{\psi\,x}{2}$$

Now we wish to compute $Q_{1,1}(\theta)$:

$$Q_{1,1}(\theta) = -\frac{R_{1,1}(\theta) + B_{1,1}(\theta)n_1'(\theta)}{n_1(\theta)}$$

```
(%i55)  Q[1,1] :::
           - (R[1,1] + B[1,1] * D(n[1])) / n[1];

(%o55)
```

$$0$$

This division to obtain $Q_{1,1}$ will always be exact. What might not be exact is the following division to obtain C_1. If the division isn't exact, or if C_1 isn't a constant, then the integral is not elementary.

$$C_1 = \frac{b_{1,1}(\theta) - B_{1,1}'(\theta) - Q_{1,1}(\theta)}{n_1'(\theta)}$$

```
(%i56)  C[1] :::
           (b[1,1] - D(B[1,1]) - Q[1,1]) / D(n[1]);

(%o56)
```

$$-\frac{1}{2}$$

A similar calculation handles the other irreducible factor:

```
(%i57)  R[2,1] : b[2,2];
```

(%o57)

$$\frac{\left(\psi^2+\psi\right)\,x+\psi^2\,x^2+2\,\psi^2\,x^3+2\,\psi^2\,x^4}{2\,x+2}$$

```
(%i58)  B[2,1] ::: - R[2,1] / D(n[2]);
```

(%o58)

$$-\frac{\psi\,x}{2}$$

```
(%i59)  Q[2,1] :::
           - (R[2,1] + B[2,1] * D(n[2])) / n[2];
```

(%o59)

$$0$$

```
(%i60)  C[2] :::
           (b[2,1] - D(B[2,1]) - Q[2,1]) / D(n[2]);
```

(%o60)

$$\frac{1}{2}$$

Plugging everything together, we conclude that our solution is:

```
(%i61)  A + ratsimp(sum(B[i,1]/n[i],i,1,2))
           + logcontract(2*sum(C[i] * log(n[i]),i,1,2))/2;
```

(%o61)

$$\frac{\log\left(-\frac{\theta+\psi\,x}{\psi\,x-\theta}\right)}{2}-\log\left(1+x\right)+\frac{\psi\,\theta\,x}{\psi^2\,x^2-\theta^2}+x$$

Converting back to our original form:

```
(%i62)  subst([theta=log(x+1), psi=exp(x^2)], %);
```

```
(%o62)
```

$$\frac{\log\left(-\frac{x\,e^{x^2}+\log(1+x)}{x\,e^{x^2}-\log(1+x)}\right)}{2}+\frac{x\,e^{x^2}\,\log\left(1+x\right)}{x^2\,e^{2\,x^2}-\log^2\left(1+x\right)}-\log\left(1+x\right)+x$$

Finally, we verify that this is, in fact, an anti-derivative of the original integrand:

```
(%i63)  diff(%,x) === integrand;

(%o63)
```

true

$$\int \frac{x\{(x^2e^{2x^2}-\ln^2(x+1))^2+2xe^{3x^2}(x-(2x^3+2x^2+x+1)\ln(x+1))\}}{(x+1)(\ln^2(x+1)-x^2e^{2x^2})^2}dx$$

$$=x-\ln(x+1)-\frac{xe^{x^2}\ln(x+1)}{\ln^2(x+1)-x^2e^{2x^2}}+\frac{1}{2}\ln\frac{\ln(x+1)+xe^{x^2}}{\ln(x+1)-xe^{x^2}}$$

□

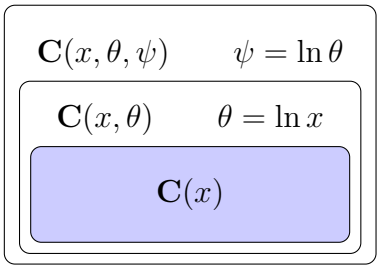
Example 5.7. Compute

$$\int \left[(\ln(\ln x))^2 \ln x - \frac{2}{\ln x}(\ln(\ln x) + 1) \right] dx$$

We'll use the extension $\mathbf{C}(x, \theta, \psi)$ where $\theta = \ln x$ and $\psi = \ln \theta = \ln \ln x$. Converting to Liouvillian form, our integral becomes

$$\int \left[\psi^2 \theta - \frac{2}{\theta}(\psi + 1) \right] dx = \int \left[\theta \psi^2 - \frac{2}{\theta}\psi - \frac{2}{\theta} \right] dx$$

Since we need θ to construct ψ, our extensions nest in only a single way:



So, we'll work in $\mathbf{C}(x, \theta, \psi)$, recursing into $\mathbf{C}(x, \theta)$ and $\mathbf{C}(x)$. In $\mathbf{C}(x, \theta, \psi)$, $n = 2$, and we identify the coefficients of our integrand:

$$a_2 = \theta \qquad a_1 = -\frac{2}{\theta} \qquad a_0 = -\frac{2}{\theta}$$

So equation (5.3a) reads:

$$[A_2 + 3c_3\psi]' = a_2 = \theta$$

We already performed this integration in example 5.3 with the result:

$$A_2 + 3c_3\psi = x\theta - x$$

In this example, $A_2 \in \mathbb{C}(x, \theta)$, so $A_2 = x\theta - x$, $c_3 = 0$.

$$[A_1 + 2c_2\psi]' = a_1 - 2\frac{\frac{1}{x}}{\theta}[x\theta - x]$$

$$= -\frac{2}{\theta} - \frac{2}{x\theta}[x\theta - x] = -2$$

$$A_1 + 2c_2\psi = -2x$$

So $A_1 = -2x$ and $c_2 = 0$. Finally,

$$A_0' = a_0 - \frac{1}{x\theta}(-2x) = -\frac{2}{\theta} + \frac{2}{\theta} = 0$$

So $A_0 = C$ and our result becomes:

$$\int \left[\theta\psi^2 - \frac{2}{\theta}(\psi + 1)\right] dx = (x\theta - x)\psi^2 - 2x\psi$$

□

5.2 Hermite Reduction

Another, more efficient, approach to handling the polynomials in denominators is to reduce their order until our denominator has only factors of multiplicity one. We're attempting to do this:

$$\int \frac{N}{V^n} = \frac{A}{V^{n-1}} + \int \frac{B}{V^{n-1}}$$

Differentiating:

$$\frac{N}{V^n} = \frac{A'}{V^{n-1}} - (n-1)\frac{AV'}{V^n} + \frac{B}{V^{n-1}}$$

and multiplying through by V^n:

$$N = VA' - (n-1)AV' + BV$$
$$N = (A' + B)V - (n-1)AV'$$

This equation has the form of a polynomial Diophantine equation, and since we know N, V and V', we can use the extended Euclidian algorithm to find $(n-1)A$ and $(A' + B)$, which easily gives us A and B. So long as V is square-free, we know that $\gcd(V, V') = 1$ (EXPLAIN), so we're guaranteed a solution (STATE THEOREM).

Example 5.8. Redo Example 5.6 using Hermite reduction.

$$\int \frac{x\{(x^2e^{2x^2} - \ln^2(x+1))^2 + 2xe^{3x^2}(x - (2x^3 + 2x^2 + x + 1)\ln(x+1))\}}{(x+1)(\ln^2(x+1) - x^2e^{2x^2})^2}dx$$

We proceed as before, putting the integral into Liouvillian form and dividing out $\frac{x}{x+1}$ to obtain a proper fraction:

$$\int \frac{x}{x+1} + \frac{(-4x^5 - 4x^4 - 2x^3 - 2x^2)\psi^3\theta + 2x^3\psi^3}{(x+1)(\theta^2 - x^2\psi^2)^2}dx$$

Now we apply the Hermite reduction, using:

$$V = \theta^2 - x^2\psi^2 \qquad V' = \frac{2}{x+1}\theta - (2x + 4x^3)\psi^2$$

$$N = \frac{-4x^5 - 4x^4 - 2x^3 - 2x^2}{x+1}\psi^3\theta + \frac{2x^3}{x+1}\psi^3$$

Our polynomial Diophantine equation is:

$$\frac{-4x^5-4x^4-2x^3-2x^2}{x+1}\psi^3\theta+\frac{2x^3}{x+1}\psi^3=sV+tV'$$

```
(%i64)  v: theta^2-x^2*psi^2;
```

(%o64)

$$\theta^2-\psi^2\,x^2$$

```
(%i65)  vtick: D(v);
```

(%o65)

$$\frac{2\,\theta}{x+1}-4\,\psi^2\,x^3-2\,\psi^2\,x$$

```
(%i66)  N/denom(lintegrand);
```

(%o66)

$$\frac{2\,\psi^3\,x^3+\theta\,\left(-4\,\psi^3\,x^5-4\,\psi^3\,x^4-2\,\psi^3\,x^3-2\,\psi^3\,x^2\right)}{\left(x+1\right)\,\left(\theta^2-\psi^2\,x^2\right)^2}$$

```
(%i67)  r: gcdex(v,vtick,theta) * N/(x+1)$
(%i68)  S::: r[1] - divide(r[1],vtick,theta)[1]*vtick;
```

(%o68)

$$-\frac{2\,\psi\,x}{x+1}$$

```
(%i69)  T::: r[2] + divide(r[1],vtick,theta)[1]*v;
```

(%o69)

$$\psi\,\theta\,x$$

So, our solution is:

$$\frac{-4x^5-4x^4-2x^3-2x^2}{x+1}\psi^3\theta+\frac{2x^3}{x+1}\psi^3=\frac{-2x}{x+1}\psi V+x\psi\theta V'$$

```
(%i70)  A: -T;
```

```
(%o70)
```

$$-\psi\,\theta\,x$$

```
(%i71)  B: S - D(A);

(%o71)
```

$$-\frac{\psi\,x}{x+1}+2\,\psi\,\theta\,x^2+\psi\,\theta$$

$$A=-x\psi\theta$$

$$A'=-\psi\theta-x(2x)\psi\theta-x\psi\frac{1}{x+1}$$

$$A'+B=\frac{-2x}{x+1}\psi$$

$$B=(2x^2+1)\psi\theta-\frac{x}{x+1}\psi$$

Which reduces our integral to:

```
(%i72)  A[0] + A/v + 'integrate(B/v, x);

(%o72)
```

$$\int\frac{\psi\,\theta+2\,\psi\,\theta\,x^2-\frac{\psi\,x}{x+1}}{\theta^2-\psi^2\,x^2}\,dx-\frac{\psi\,\theta\,x}{\theta^2-\psi^2\,x^2}+\left(-\psi\,\theta\,x\right)_0$$

$$x-\theta-\frac{x\psi\theta}{\theta^2-x^2\psi^2}+\int\frac{(2x^2+1)\psi\theta-\frac{x}{x+1}\psi}{\theta^2-x^2\psi^2}$$

Now we can compute a Rothstein-Trager resultant:

```
(%i73)  resultant(B - z * D(v), v, theta);

(%o73)
```

$$\psi^2\,x^2\,\left(4\,\psi^2\,x^6+8\,\psi^2\,x^5+8\,\psi^2\,x^4+8\,\psi^2\,x^3+5\,\psi^2\,x^2+2\,\psi^2\,x+\psi^2-1\right)\,\left(4\,z^2-1\right)$$

This result is in $\mathbb{C}(x,\psi)[z]$, so the first two factors are just *content* (EXPLAIN THIS TERM).

```
(%i74)  partition(%,z)[2];

(%o74)
```

$$4\,z^2-1$$

The result is really just $4z^2-1$, which has two solutions: $\pm\frac{1}{2}$. Substituting in these two values for z, we obtain the corresponding logarithms:

```
(%i75)  gcd(B - (1/2)*D(v), v);

(%o75)
```

$$\frac{\theta+\psi\,x}{x+1}$$

```
(%i76)  gcd(B + (1/2)*D(v), v);

(%o76)
```

$$\frac{\psi\,x-\theta}{x+1}$$

$$x-\theta-\frac{x\psi\theta}{\theta^2-x^2\psi^2}+\frac{1}{2}\ln(\frac{\theta+x\psi}{x+1})-\frac{1}{2}\ln(\frac{\theta-x\psi}{x+1})$$

$$\int\frac{x(x+1)\{(x^2e^{2x^2}-\ln^2(x+1))^2+2xe^{3x^2}(x-(2x^3+2x^2+x+1)\ln(x+1))\}}{((x+1)\ln^2(x+1)-(x^3+x^2)e^{2x^2})^2}dx$$

$$=x-\theta-\frac{x\psi\theta}{\theta^2-x^2\psi^2}+\frac{1}{2}\ln(\theta+x\psi)-\frac{1}{2}\ln(\theta-x\psi)$$

$$=x-\ln(x+1)-\frac{xe^{x^2}\ln(x+1)}{\ln^2(x+1)-x^2e^{2x^2}}$$
$$+\frac{1}{2}\ln\left[\ln(x+1)+xe^{x^2}\right]-\frac{1}{2}\ln\left[\ln(x+1)-xe^{x^2}\right]$$

□

Chapter 6

The Exponential Extension

The two distinctive features of integration in exponential extensions are the presence of *special* polynomials, which divide their own derivatives, and the appearance of the Risch differential equation.

Let's recall our basic theorem on the behavior of exponential extensions:

Theorem 3.10. *Let $E = K(\theta)$ be a simple transcendental exponential extension of a differential field K with the same constant subfield as K, let $p = \sum p_i \theta^i$ be a polynomial in $K[\theta]$ ($p_i \in K$), and let r be a rational function in $K(\theta)$. Then:*

1. $\mathrm{Deg}_\theta\, p' = \mathrm{Deg}_\theta\, p$
2. *If p is monic and irreducible, then $p \mid p'$ if and only if $p = \theta$.*
3. *If an irreducible monic factor other than θ appears in r's denominator with multiplicity m, then it appears in r''s denominator with multiplicity $m + 1$*
4. *$r' \in K$ if and only if $r \in K$*

Contrast this theorem with the behavior of the ordinary polynomials that we're accustomed to. Ordinary irreducible polynomials never divide their own derivatives in the manner described in (2); polynomials that do are called *special*. Instead, ordinary polynomials always behave in the way described in (3); such polynomials are called *normal*.

Irreducible polynomials are characterized as either normal or special, depending on whether they divide their own their derivatives. Theorem 3.10 (2) states that in a exponential extension, the only special irreducible polynomial is θ itself.

We attack integrands in exponential extensions in much the same way as we attack ordinary polynomials: we factor the denominator into irreducible factors and perform a partial fractions expansion. In this case, however, we have to classify the denominator factors as either normal or special. Normal factors can be handled in much the same way as we're used to, but special factors are treated in a manner similar to polynomials.

For example, let $p = \sum p_i \theta^i$ be a polynomial in $K[\theta]$, with $\theta = \exp k$, and take its derivation:

$$p' = \sum_{i=0}^{n} (p_i' + i p_i k') \theta^i$$

Notice that, unlike the logarithm or rational cases, there is no interdependence between the various terms of the sum; each term is completely independent of the others. Instead, each coefficient of θ^i has the form $p_i' + A p_i$, and equating the p' polynomial to the integrand's polynomial produces a differential equation of the form:

$$p_i' + A p_i = B \qquad A, B, p_i \in K$$

This is called a *Risch equation* and is a primary object of our study. Solving Risch equations in a differential extension is the principle problem that we need to solve in order to carry out our program of symbolic integration.

Special factors in the denominator behave in almost exactly the same way as polynomials. They both give rise to Risch equations that need to be solved in the underlying field. On the other hand, partial fractions terms involving normal polynomials give rise to rational functions and logarithms in the result that can be solved using the extended Euclidean algorithm, again operating in the underlying field.

We've already studied, in Section 2.8, how to use the extended Euclidean algorithm over an arbitrary field, so the primary additional tool we need to develop is the ability to solve Risch equations in arbitrary differential fields, or at least in the differential fields that arise in the course of our study. Once we can do that, we can evaluate integrals in complicated extension fields by "peeling away" the extensions, and solving equations in successively simplier extensions until we've reached the rational function field $\mathbb{C}(x)$.

I'll begin by presenting the basic integration theorem for exponential extensions, then we'll consider how to solve Risch equations in $\mathbb{C}(x)$, which is a simplified case that lets us solve integrals in *simple* exponential extensions. Finally, we'll study solving the Risch equation more generally.

6.1 The Exponential Integration Theorem

Theorem 6.1. *Let K be a differential field, let $K(\theta = \exp k)$ be a simple exponential extension of K, let $n_i(\theta)$ be normal irreducible polynomials in $K[\theta]$, and let f be an element of $K(\theta)$ with partial fractions expansion:*

$$f = \sum_{i=0}^{n} a_i \theta^i + \sum_{j=1}^{l} \frac{b_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{c_{i,j}(\theta)}{n_i(\theta)^j} \tag{6.1}$$

$$a_i, b_j \in K \qquad c_{i,j}(\theta), n_i(\theta) \in K[\theta]$$

If f has an elementary anti-derivative F, then $F \in K(\theta, \Psi)$, where $K(\theta, \Psi)$ is a finite logarithm extension of $K(\theta)$, F has a partial fractions expansion of the form:

$$F = \sum_{i=0}^{n} A_i \theta^i + \sum_{j=1}^{l} \frac{B_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i - 1} \frac{C_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\nu} D_i \ln n_i(\theta) \tag{6.2}$$

$$A_i, B_j \in K \qquad C_{i,j}(\theta), n_i(\theta) \in K[\theta] \qquad D_i' = 0$$

and the following relationships hold:

$$A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} \tag{6.3}$$

$$A_i' + i A_i k' = a_i \qquad B_j' - j k' B_j = b_j \tag{6.4}$$

$$\begin{aligned} R_{i,m_i-1}(\theta) &= c_{i,m_i} && \text{(6.5a)} \\ R_{i,j-1}(\theta) &= c_{i,j} - C_{i,j}'(\theta) - Q_{i,j}(\theta) \qquad {\scriptstyle 1<j\le m_i} && \text{(6.5b)} \end{aligned}$$

$$C_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{j n_i'(\theta)} \mod n_i(\theta) \tag{6.6}$$

$$D_i = \frac{c_{i,1} - C_{i,1}'(\theta) - Q_{i,1}(\theta)}{n_i'(\theta) - \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} n_i(\theta)} \tag{6.7}$$

Proof

By Theorem 3.15, an elementary antiderivative of f can only exist in a finite logarithm extension $K(\theta, \Psi)$ of $K(\theta)$ and therefore must have the form:

$$F = R + \sum_{i=1}^{\eta} D_i \Psi_i$$

where $R \in K(\theta)$, and the D_i are constants.

Constructing a partial fractions expansion of R, separating the normal and special components of its denominator, and using the fact that $s_1 = \theta$ is the only special irreducible polynomial (Theorem 3.10):

$$F = \sum_{i=0}^{N} A_i \theta^i + \sum_{j=1}^{L} \frac{B_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{C_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\eta} D_i \Psi_i$$

Now let's differentiate, remembering that $\theta' = k'\theta$:

$$F' = \sum_{i=0}^{N} (A_i' + iA_i k')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{C_{i,j}'(\theta)n_i(\theta) - jC_{i,j}(\theta)n_i'(\theta)}{n_i(\theta)^{j+1}} + \sum_{i=1}^{\eta} D_i \frac{E_i'(\theta)}{E_i(\theta)}$$

Let's examine the logarithmic elements $E_i(\theta)$. If an $E_i(\theta)$ doesn't involve θ, i.e, $E_i \in K$, then we can collapse $D_i \Psi_i$ into A_0, with the understanding that when we recurse into K to solve for A_0, additional logarithm terms are allowed.

So now let's consider what happens if $E_i(\theta)$ is a polynomial in $K[\theta]$. If it's reducible, then the basic properties of logarithms let us split it into multiple irreducible elements. Otherwise, it's irreducible and therefore either normal or special. If it's special, then it must be θ itself and $\ln \theta = \ln \exp k = k$, which contracts the transcendence of the logarithm extension Ψ. So all of the $E_i(\theta)$'s must be normal, and therefore F' must have the form:

$$F' = A_0' + \sum D_k \frac{E_k'}{E_k} + \sum_{i=1}^{N} (A_i' + iA_i k')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{M_i} \frac{C_{i,j}'(\theta)n_i(\theta) - jC_{i,j}(\theta)n_i'(\theta)}{n_i(\theta)^{j+1}} + \sum_{i=1}^{\eta} D_i \frac{n_i(\theta)'}{n_i(\theta)}$$

F' has the form of a partial fractions decomposition, but it is not a partial fractions decomposition because the numerators in the C and D terms violate the partial fractions degree

bounds. To fix this, let's divide the $-jC_{i,j}(\theta)n_i'(\theta)$ terms by $n_i(\theta)$ (think polynomial long division) and rewrite them as a quotient and a remainder:

$$-jC_{i,j}(\theta)n_i'(\theta) = Q_{i,j}(\theta)n_i(\theta) + R_{i,j}(\theta)$$

This fixes the C terms, since $\deg Q_{i,j}(\theta) < \deg n_i(\theta)$ and $\deg R_{i,j}(\theta) < \deg n_i(\theta)$.

We can fix the D terms by noting that $\deg n_i'(\theta) = \deg n_i(\theta)$, so by subtracting an appropriate multiple of $n_i(\theta)$ we can ensure the cancellation of the leading terms, achieving our degree bounds.

$$\begin{aligned} F' = A_0' + \sum D_k \frac{E_k'}{E_k} + \sum D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} + \sum_{i=1}^{N} (A_i' + iA_i k')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} \\ + \sum_{i=1}^{\nu} \left[\frac{R_{i,M_i}(\theta)}{n_i(\theta)^{M_i+1}} + \sum_{j=2}^{M_i} \frac{C_{i,j}'(\theta) + Q_{i,j}(\theta) + R_{i,j-1}(\theta)}{n_i(\theta)^j} \right. \\ \left. + \frac{C_{i,1}'(\theta) + Q_{i,1}(\theta) + D_i \left[n_i(\theta)' - \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} n_i(\theta) \right]}{n_i(\theta)} \right] \end{aligned}$$

Now F' is an actual partial fractions decomposition. It not only has the right form, but all of the other conditions, specifically the degree bounds, are met. Therefore, we can perform a partial fractions decomposition of f:

$$f = \sum_{i=0}^{n} a_i \theta^i + \sum_{j=1}^{l} \frac{b_j}{\theta^j} + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{c_{ij}(\theta)}{n_i(\theta)^j}$$

Setting $F' = f$ and equating like terms, we establish that $n = N$, $l = L$, $M_i + 1 = m_i$, and the relationships listed in the statement of the theorem. The zero-order term:

$$A_0' + \sum D_k \frac{E_k'}{E_k} + \sum D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} = a_0$$

Polynomial terms and special denominators give rise to Risch equations:

$$A_i' + iA_i k' = a_i \qquad B_j' - jk'B_j = b_j$$

Normal denominators give rise to these terms:

$$\begin{aligned} R_{i,m_i}(\theta) &= c_{i,m_i+1} \\ C'_{i,j}(\theta) + Q_{i,j}(\theta) + R_{i,j-1}(\theta) &= c_{i,j} \qquad 1<j\leq m_i \\ C'_{i,1}(\theta) + Q_{i,1}(\theta) + D_i n_i(\theta)' &= c_{i,1} \end{aligned}$$

Remember the definition of $R_{i,j}$ and $Q_{i,j}$:

$$-jC_{i,j}(\theta)n'_i(\theta) = Q_{i,j}(\theta)n_i(\theta) + R_{i,j}(\theta)$$

Reducing this equation modulo $n_i(\theta)$, we obtain:

$$-jC_{i,j}(\theta)n'_i(\theta) \equiv R_{i,j}(\theta) \mod n_i(\theta)$$

Now we use the fact that $n_i(\theta)$ is ***irreducible***, and invoke Theorem ??, which states the quotient ring modulo a prime ideal is a field, so we can perform division:

$$C_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{jn'_i(\theta)} \mod n_i(\theta)$$

This equation seems to identify $C_{i,j}(\theta)$ up to a multiple of $n_i(\theta)$, but if we remember our degree bound on partial fractions expansions, $\deg_\theta C_{i,j}(\theta) < \deg_\theta n_i(\theta)$, we see that in fact we've completely determined $C_{i,j}(\theta)$ from $R_{i,j}(\theta)$.

□

We'll begin discussing the Risch equation in the next section, which is how we obtain the A_i's and B_i's.

How can we calculate the $C_{i,j}(\theta)$'s?

The highest order term in the partial fractions expansion gives us an $R_{i,j}$ directly. Then we use the extended Euclidean algorithm from Section 2.8, which is our major computational tool for calculating inverses in quotient rings, to calculate $C_{i,j}(\theta)$. A simple long division step then gives us the quotient $Q_{i,j}(\theta)$, and we just move down the list, solving this system of equations from highest order terms to lowest. Once we get to the end, we need to see if the bottom equation can be solved using a constant D_i. If not, then the equation has no solution.

Once all of the D_i's have been calculated, then we have all of the information needed to determine A'_0, and an integration step in the underlying field yields A_0 itself.

Example 6.2. Compute $\int \sin x \, \mathrm{d}x$

We'll operate in $\mathbb{C}(x, \theta = \exp\ ix)$ and evaluate

$$\int \frac{\theta - \theta^{-1}}{2i}\, dx$$

The first step is write the integrand in partial fractions form:

$$\int \left[\frac{1}{2i}\theta - \frac{1}{2i}\frac{1}{\theta}\right]\, dx$$

By Theorem 6.1, the integral must have the form $a_1\theta + a_{-1}\frac{1}{\theta}$ with $a_1, a_{-1} \in \mathbb{C}(x)$ and must satisfy the equations:

$$\frac{1}{2i} = a_1' + ia_1 \qquad -\frac{1}{2i} = a_{-1}' - ia_{-1}$$

These are very simple Risch equations that can be solved by inspection to obtain $a_1 = a_{-1} = -\frac{1}{2}$, so

$$\int \frac{\theta - \theta^{-1}}{2i}\, dx = -\frac{1}{2}(\theta + \theta^{-1}) = -\frac{1}{2}(e^{ix} + e^{-ix}) = -\cos x$$

□

Example 6.3. Compute $\int \csc x \, \mathrm{d}x$

We'll operate in $\mathbb{C}(x, \psi = \exp\, ix)$ and evaluate

$$\int \frac{2i}{\psi - \psi^{-1}}\, dx = \int 2i \frac{\psi}{\psi^2 - 1}\, dx$$

Now we want a partial fractions expansion. We could use a resultant, or the Euclidean G.C.D. algorithm, but it's simpler to just note that the denominator's a difference of squares and compute:

$$\frac{c_1}{\psi - 1} + \frac{c_2}{\psi + 1} = \frac{\psi}{\psi^2 - 1}$$

$$c_1(\psi + 1) + c_2(\psi - 1) = \psi \qquad c_1 = c_2 = \frac{1}{2}$$

giving us

$$\int \left[\frac{i}{\psi - 1} + \frac{i}{\psi + 1} \right] dx$$

Now we have an integral in the form of equation (6.1) with $\nu = 2$, $n_1(\psi) = \psi - 1$, $n_2(\psi) = \psi + 1$, $m_1 = m_2 = 1$ and $c_{1,1} = c_{2,1} = i$.

$$D_1 = \frac{c_{1,1}}{n_1'(\theta) - \frac{\mathrm{lc}\, n_1'(\theta)}{\mathrm{lc}\, n_1(\theta)} n_1(\theta)} = \frac{i}{i\psi - i(\psi - 1)} = 1$$

$$D_2 = \frac{c_{2,1}}{n_2'(\theta) - \frac{\mathrm{lc}\, n_2'(\theta)}{\mathrm{lc}\, n_2(\theta)} n_2(\theta)} = \frac{i}{i\psi - i(\psi + 1)} = -1$$

so by Theorem 6.1 the integral can be written:

$$\int \left[\frac{i}{\psi - 1} + \frac{i}{\psi + 1} \right] dx = \ln(\psi - 1) - \ln(\psi + 1) = \ln(\frac{\psi - 1}{\psi + 1})$$

$$= \ln(\frac{e^{ix} - 1}{e^{ix} + 1}) = \ln(\frac{e^{ix/2} - e^{-ix/2}}{e^{ix/2} + e^{-ix/2}}) = \ln i \frac{\sin \frac{x}{2}}{\cos \frac{x}{2}} = \ln \tan \frac{x}{2}$$

where I dropped the i at the end because, as a constant multiple inside a logarithm, it disappears into the constant of integration, and we conclude that

$$\int \csc x \, \mathrm{d}x = \ln \tan \frac{x}{2}$$

□

Example 6.4. Compute $\int \tan x \, \mathrm{d}x$

We'll operate in $\mathbb{C}(x, \theta = \exp ix)$ and evaluate

$$-i \int \frac{\theta - \theta^{-1}}{\theta + \theta^{-1}} \, dx$$

The first step is write the integrand in partial fractions form:

$$\frac{\theta - \theta^{-1}}{\theta + \theta^{-1}} = \frac{\theta^2 - 1}{\theta^2 + 1} = 1 - \frac{2}{\theta^2 + 1} = 1 - \frac{2}{(\theta + i)(\theta - i)}$$

$$= 1 + \frac{i}{\theta - i} - \frac{i}{\theta + i}$$

$$-i \frac{\theta - \theta^{-1}}{\theta + \theta^{-1}} = -i + \frac{1}{\theta - i} - \frac{1}{\theta + i}$$

This integrand is now in the form of equation (6.1) with $n_1(\theta) = \theta - i$ and $n_2(\theta) = \theta + i$, $c_{1,1} = 1$, $c_{2,1} = -1$, $a_0 = 1$ and $k = ix$, so plugging these values into equation (6.7), we obtain:

$$D_1 = \frac{c_{1,1}}{n_1(\theta)' - \frac{\operatorname{lc} n_1'(\theta)}{\operatorname{lc} n_1(\theta)} n_1(\theta)} = \frac{1}{i\theta - i(\theta - i)} = -1$$

$$D_2 = \frac{c_{2,1}}{n_2(\theta)' - \frac{\operatorname{lc} n_2'(\theta)}{\operatorname{lc} n_2(\theta)} n_2(\theta)} = \frac{-1}{i\theta - i(\theta + i)} = -1$$

Pluging this into equation (6.3), we obtain:

$$A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)}$$

$$A_0' = -i + 2i = i$$

$$A_0 = ix$$

so our integral is:

$$-i \int \frac{\theta - \theta^{-1}}{\theta + \theta^{-1}} \, dx = ix - \ln(\theta + i) - \ln(\theta - i)$$

$$= ix - \ln\left[(\theta + i)(\theta - i)\right]$$

$$= ix - \ln\left[\theta^2 + 1\right]$$

We can convert this into a more familiar form by realizing that $ix = \ln \exp ix = \ln \theta$, so

$$= \ln \theta - \ln\left[\theta^2 + 1\right] = \ln \frac{\theta}{\theta^2 + 1} = \ln \frac{1}{\theta^{-1} + \theta}$$

$$\int \tan x \, \mathrm{d}x = \ln \csc x$$

□

6.2 Risch Equations in $\mathbb{C}(x)$

Risch equations in $\mathbb{C}(x)$ arise when our integrand exists in a *simple* exponential extension of $\mathbb{C}(x)$, i.e, an integrand formed as a rational function of x and a single exponential of a rational function of x. Theorem 6.1 then produces Risch equations in the underlying field; in this case, $\mathbb{C}(x)$.

Consider such a Risch equation:

$$r' + Sr = T \qquad S, T, r \in \mathbb{C}(x)$$

The only poles that can appear in r's denominator must appear in either S or T's denominator, so let's consider a single pole at γ, expand S, T, and r using partial fractions, and look at the highest powers of $(x - \gamma)$ in the denominator:

$$r = \frac{a}{(x-\gamma)^j} + \cdots \qquad r' = \frac{-ja}{(x-\gamma)^{j+1}} + \cdots$$

$$S = \frac{b}{(x-\gamma)^k} + \cdots$$

$$T = \frac{c}{(x-\gamma)^l} + \cdots$$

Combining everything into the Risch equation, we find:

$$\frac{-ja}{(x-\gamma)^{j+1}} + \cdots + \frac{ba}{(x-\gamma)^{j+k}} + \cdots = \frac{c}{(x-\gamma)^l} + \cdots$$

We can classify the equation into three basic cases, based on the value of k:

1. $k = 0$. In this case, the $\frac{-ja}{(x-\gamma)^{j+1}}$ term dominates the left hand side, and $j = l - 1$ in order to match the right hand side.
2. $k = 1$. Here, the high order terms on the left are equal, so either $j = l - 1$ in order to match the right hand side, or $j = b$ and $j > l - 1$ in order for the left hand terms to exactly cancel.
3. $k > 1$. Now the $\frac{ba}{(x-\gamma)^{j+k}}$ term dominates the left hand side, so $j = l - k$ in order to match the right hand side.

By checking all of S's and T's poles using this technique, we can identify all the poles in r's denominator and determine the multiplicity with which they appear. This determines r's denominator d completely.

Having done so, we can replace r with p/d, which clears the denominators and produces a polynomial Risch equation:

$$Ap' + Bp = C \qquad A, B, C, p \in \mathbb{C}[x]$$

There are three cases now.

First, the highest terms on the left can have higher degree than any term on the right, and so must cancel against each other. For this to occur, $\deg A = \deg B + 1$ (since $\deg p$ drops by one on differentiation), and we can determine $\deg p$ by looking at the leading coefficients in A and B:

$$A = a_j x^j + \cdots \qquad B = b_{j-1} x^{j-1} + \cdots \qquad p = p_k x^k \cdots$$

$$Ap' + Bp = (k a_j p_k + b_{j-1} p_k) x^{j+k-1} \cdots$$

In order for this coefficient to be zero, $k = -b_{j-1}/a_j$. So, if these conditions are met:

$$\deg A = \deg B + 1 \qquad k = -\frac{b_{j-1}}{a_j}$$

then p may exist as a k^{th} degree polynomial.

Otherwise, the leading terms of $Ap' + Bp$ do not cancel out, so they must match the leading term of C. This can only occur if

$$\deg p = \deg C - \max(\deg A - 1, \deg B)$$

Having determined the degree of p, we can now determine its coefficients.

The final case we need to consider is when p is a constant, which would solve the Risch equation if and only if C was a constant multiple of B, irregardless of A.

1. $\deg A = \deg B + 1$ and $\deg p = -\frac{\operatorname{lc} B}{\operatorname{lc} A}$
2. $\deg p = \deg C - \max(\deg A - 1, \deg B)$
3. p is a constant and $pB = C$

Example 6.5. Prove that $\int e^{-x^2}\,\mathrm{d}x$ has no elementary form

We'll use $\mathbb{C}(x, \psi = \exp\, -x^2)$, so $\psi' = -2x$ and study

$$\int \psi\,\mathrm{d}x$$

We know from Theorem 3.10 that our solution, if it exists, must have the form $a\psi$, where $a \in \mathbb{C}(x)$, and a must satisfy the Risch equation:

$$a' - 2xa = 1$$

This is already a polynomial Risch equation, and a' has only a constant coefficient, so a can not have a non-trivial denominator. Futhermore, identifying A as 1, B as $-2x$, and C as 1, we see that $\deg A = 0$ and $\deg B = 1$. Since $\deg A \neq \deg B + 1$, the leading terms on the left hand side can not cancel, so they must match the leading term on the right. We compute:

$$\deg p = \deg C - \max(\deg A - 1, \deg B) = 0 - 1 = -1$$

so that doesn't work. Futhermore, C is not a constant multiple of B, so a constant p can't solve our equation.

We conclude that no solution to this Risch equation exists in $\mathbb{C}(x)$, so the integral can not be expressed in elementary form.

□

Example 6.6. Prove that $\int \frac{\sin x}{x}\,\mathrm{d}x$ has no elementary form

As we often do with trigonometric integrals, we'll operate in $\mathbb{C}(x, \psi = \exp\, ix)$, use Euler's relationship $e^{ix} = i\sin x + \cos x$, and evaluate

$$\int \frac{\psi - \psi^{-1}}{2ix}\,dx$$

Let's begin by writing the integrand in the form of a rational function in $\mathbb{C}(x)(\psi)$, i.e, a ratio of ψ-polynomials, with coefficients in $\mathbb{C}(x)$:

$$\frac{1}{2i}\int \left[\frac{1}{x}\psi - \frac{1}{x}\frac{1}{\psi}\right]\,dx$$

The integral must have the form $a_1\psi + a_{-1}\frac{1}{\psi}$ with $a_1, a_{-1} \in \mathbb{C}(x)$ and must satisfy the equations:

$$\left[a_1\psi + a_{-1}\frac{1}{\psi}\right]' = (a_1' + ia_1)\psi + (a_{-1}' - ia_{-1})\frac{1}{\psi} = \left[\frac{1}{x}\psi - \frac{1}{x}\frac{1}{\psi}\right]$$

$$\frac{1}{x} = a_1' + ia_1 \qquad -\frac{1}{x} = a_{-1}' - ia_{-1}$$

$$a_1' + a_1 = \frac{1}{x}$$

We've got a single pole in the denominator of T, so $k = 0$, $l = 1$, and $j = l - 1 = 0$, so there are no poles in the denominator of our solution. However, there is then no way to produce the denominator on the right, so the Risch equation has no solution in $\mathbb{C}(x)$.

Thus, the integral can not be expressed in elementary form.

□

Partial fractions terms involving normal polynomials are handled the same way as, well, normal polynomials. Terms with simple denominators give rise to logarithms in the solution, while terms with higher powered denominators give rise to rational functions in the solution.

One unusual feature of exponential extensions is that the numerator of a derivative will have the same degree as the denominator, so a long division step is needed to make the fraction proper, and this will produce a constant that will modify the integrand. For this reason, it's best to handle the denominator's normal factors first,

Example 6.7. Compute $\int \frac{4^x+1}{2^x+1}\mathrm{d}x$

We'll use the field $\mathbb{C}(x, \Psi = \exp(x \ln 2))$; $\Psi' = (\ln 2)\Psi$ and the representation (see Example 3.4):

$$\frac{\Psi^2+1}{\Psi+1} = \Psi - 1 + \frac{2}{\Psi+1}$$

We'll start, as before, with equation (6.7):

$$D_1 = \frac{c_{1,1}}{n_1(\theta)' - \frac{\operatorname{lc} n_1'(\theta)}{\operatorname{lc} n_1(\theta)} n_1(\theta)} = \frac{2}{(\ln 2)\Psi - \ln 2(\Psi+1)} = -\frac{2}{\ln 2}$$

Now, equation (6.4) yields:

$$A_1' + (\ln 2)A_1 = 1 \qquad \Longrightarrow \qquad A_1 = \frac{1}{\ln 2}$$

and equation (6.3) yields:

$$A_0' = a_0 - D_1 \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} = -1 + \frac{2}{\ln 2}\ln 2 = 1$$
$$A_0 = x$$

So our solution is

$$\int \frac{4^x+1}{2^x+1}\mathrm{d}x = \frac{1}{\ln 2}\Psi + x - \frac{2}{\ln 2}\left[\ln(\Psi+1)\right] = \frac{2^x}{\ln 2} + x - \frac{2}{\ln 2}\ln(2^x+1)$$

□

Example 6.8. Redo Example 5.6 using the exponential theory.

$$\int \frac{x\{(x^2e^{2x^2} - \ln^2(x+1))^2 + 2xe^{3x^2}(x - (2x^3+2x^2+x+1)\ln(x+1))\}}{(x+1)(\ln^2(x+1) - x^2e^{2x^2})^2}dx$$

We proceed as before, putting the integral into Liouvillian form and dividing out $\frac{x}{x+1}$ to obtain a proper fraction:

```
(%i77)  divide(num(lintegrand), denom(lintegrand), psi);
```

(%o77)

$$\left[\frac{x}{x+1}, \psi^3\left(-4\,\theta\,x^5 - 4\,\theta\,x^4 + (2-2\,\theta)\,x^3 - 2\,\theta\,x^2\right)\right]$$

This time, we'll operate in $\mathbb{C}(x, \theta = \ln(x+1), \psi = \exp x^2)$, treating this as an exponential extension of $\mathbb{C}(x, \theta)$. We'll begin again by computing a partial fractions expansion, this time with respect to ψ:

```
(%i78)  pf: partfrac(N/denom(lintegrand), psi);
```

(%o78)

$$\frac{\theta^2 + (\theta^2-\theta)\,x + 2\,\theta^2\,x^2 + 2\,\theta^2\,x^3}{(\theta+\psi\,x)^2\,(2\,x^2+2\,x)} - \frac{\theta^2 + (\theta^2-\theta)\,x + 2\,\theta^2\,x^2 + 2\,\theta^2\,x^3}{(\psi\,x-\theta)^2\,(2\,x^2+2\,x)} - \frac{\theta + (\theta-1)\,x}{(\psi\,x+\theta}$$

```
(%i79)  denoms: map(lambda([u], den(u,psi)),
                    partlist(pf));
```

(%o79)

$$\left[\psi\,x-\theta, \psi\,x+\theta, (\psi\,x-\theta)^2, (\theta+\psi\,x)^2\right]$$

```
(%i80)  n: unique(map(base, denoms));
```

(%o80)

$$[\psi\,x-\theta, \psi\,x+\theta]$$

```
(%i81)  map(lambda([u],
              c[which(n, base(den(u,psi))),
                power(den(u,psi))]: u*den(u,psi)),
            partlist(pf))$
```

```
(%i82)  displayarray(c)$
```

$$c_{1,1} = -\frac{\theta + (\theta - 1)\ x + 2\,\theta\,x^2 + 2\,\theta\,x^3}{x^2 + x}$$

$$c_{1,2} = -\frac{\theta^2 + (\theta^2 - \theta)\ x + 2\,\theta^2\,x^2 + 2\,\theta^2\,x^3}{2\,x^2 + 2\,x}$$

$$c_{2,1} = -\frac{\theta + (\theta - 1)\ x + 2\,\theta\,x^2 + 2\,\theta\,x^3}{x^2 + x}$$

$$c_{2,2} = \frac{\theta^2 + (\theta^2 - \theta)\ x + 2\,\theta^2\,x^2 + 2\,\theta^2\,x^3}{2\,x^2 + 2\,x}$$

Now Theorem 6.1 tells us that

```
(%i83)  R[1,1] : c[1,2];

(%o83)
```

$$-\frac{\theta^2 + (\theta^2 - \theta)\ x + 2\,\theta^2\,x^2 + 2\,\theta^2\,x^3}{2\,x^2 + 2\,x}$$

$$C_{1,1}(\psi) = -\frac{R_{1,1}(\psi)}{n_1'(\psi)} \mod n_1(\psi)$$

```
(%i84)  modulo(r,p,v) :=
           divide(num(r),p,v)[2]
             * gcdex(denom(r),p,v)[1]$

(%i85)  kill(C)$

(%i86)  C[1,1] ::: modulo( - R[1,1] / D(n[1]),
                           n[1], psi);

(%o86)
```

$$\frac{\theta}{2}$$

This time, we do need to perform a modulo reduction, but as modulo reductions

go, it's trivial. Operating $\mod(\theta - x\psi)$ means that we're operating in a field where $\theta - x\psi = 0$, so $\psi = \frac{\theta}{x}$. (EXPLAIN WHY WE'RE REDUCING w.r.t ψ). Now we wish to compute $Q_{1,1}(\psi)$. There is no modulo reduction in this division, and it should always be exact.

$$Q_{1,1}(\psi) = -\frac{R_{1,1}(\theta) + C_{1,1}(\theta)n_1'(\theta)}{n_1(\theta)}$$

```
(%i87)  kill(Q)$

(%i88)  Q[1,1] ::: - (R[1,1] + C[1,1] * D(n[1]))
                                / n[1];

(%o88)
                                  -(θ + 2 θ x^2)/(2 x)
```

Having computed $C_{1,1}$ and $Q_{1,1}$, we are now able to compute D_1:

$$D_1 = \frac{c_{1,1} - C_{1,1}'(\psi) - Q_{1,1}(\psi)}{n_1'(\psi) - \frac{\mathrm{lc}\, n_1'(\psi)}{\mathrm{lc}\, n_1(\psi)} n_1(\psi)}$$

```
(%i89)  lc(a,v) := coeff(a, v, hipow(a,v))$

(%i90)  D[1] ::: (c[1,1] - D(C[1,1]) - Q[1,1])
        / (D(n[1])-lc(D(n[1]),psi)/lc(n[1],psi)*n[1]);

(%o90)
                                  -1/2
```

A similar calculation for $n_2(\psi) = \theta + x\psi$ yields

```
(%i91)  R[2,1] : c[2,2];

(%o91)
                  (θ^2 + (θ^2 - θ) x + 2 θ^2 x^2 + 2 θ^2 x^3)/(2 x^2 + 2 x)

(%i92)  C[2,1] ::: modulo( - R[2,1] / D(n[2]),
                                n[2], psi);
```

```
(%o92)
```

$$\frac{\theta}{2}$$

```
(%i93)  Q[2,1] ::: - (R[2,1] + C[2,1] * D(n[2]))
                              / n[2];

(%o93)
```

$$-\frac{\theta + 2\,\theta\,x^2}{2\,x}$$

```
(%i94)  D[2] ::: (c[2,1] - D(C[2,1]) - Q[2,1])
        / (D(n[2])-lc(D(n[2]),psi)/lc(n[2],psi)*n[2]);

(%o94)
```

$$\frac{1}{2}$$

In an exponential extension, our D coefficients can affect our A_0 term...

$$A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_1'(\psi)}{\operatorname{lc} n_1(\psi)}$$

```
(%i95)  a: a - sum(D[i]*lc(D(n[i]),psi)/lc(n[i],psi),i,1,2);

(%o95)
```

$$\frac{x}{x+1}$$

```
(%i96)  A: integrate(a,x);

(%o96)
```

$$x - \log\left(1 + x\right)$$

We end up with the same result that we obtained from the logarithmic theory:

```
(%i97)  A + ratsimp(sum(C[i,1]/n[i],i,1,2))
          + logcontract(2*sum(D[i]*log(n[i]),i,1,2))/2;
```

(%o97)

$$\frac{\log\left(\frac{\theta+\psi\,x}{\psi\,x-\theta}\right)}{2}-\log\left(1+x\right)+\frac{\psi\,\theta\,x}{\psi^2\,x^2-\theta^2}+x$$

```
(%i98)  subst([theta=log(x+1), psi=exp(x^2)], %);
```

(%o98)

$$\frac{\log\left(\frac{x\,e^{x^2}+\log(1+x)}{x\,e^{x^2}-\log(1+x)}\right)}{2}+\frac{x\,e^{x^2}\,\log\left(1+x\right)}{x^2\,e^{2\,x^2}-\log^2\left(1+x\right)}-\log\left(1+x\right)+x$$

```
(%i99)  diff(%,x) === integrand;
```

(%o99)

true

□

6.3 Risch Equations over Normal Polynomials

Now let's expand our study to include Risch equations in more complicated differential fields, starting with normal polynomials, which will allow us to handle logarithmic extensions.

Consider again such a Risch equation, this time in a simple transcendental extension $K(\theta)$ of an arbitrary differential field K:

$$r' + Sr = T \qquad S, T, r \in K(\theta)$$

Once again, we can perform a partial fractions expansion on S, T, and r, except that this time our irreducible polynomials are more complicated that $(x - \gamma)$. Consider one such normal irreducible polynomial $n(\theta)$:

$$S = \frac{b(\theta)}{n(\theta)^k} + \cdots \qquad T = \frac{c(\theta)}{n(\theta)^l} + \cdots$$

$$r = \frac{a(\theta)}{n(\theta)^j} + \cdots \qquad r' = \frac{a'(\theta)n(\theta) - ja(\theta)n'(\theta)}{n(\theta)^{j+1}} + \cdots = \frac{-ja(\theta)n'(\theta)}{n(\theta)^{j+1}} + \cdots$$

Plugging this all into the Risch equation, we obtain:

$$\frac{-ja(\theta)n'(\theta)}{n(\theta)^{j+1}} + \cdots + \frac{a(\theta)b(\theta)}{n(\theta)^{j+k}} + \cdots = \frac{c(\theta)}{n(\theta)^l} + \cdots$$

Both of the numerators on the left hand side could have θ-degree greater than $\deg_\theta n(\theta)$, so we divide them by $n(\theta)$:

$$R_1(\theta) = -ja(\theta)n'(\theta) \mod n(\theta)$$
$$R_2(\theta) = a(\theta)b(\theta) \mod n(\theta)$$

$$\frac{R_1(\theta)}{n(\theta)^{j+1}} + \frac{R_2(\theta)}{n(\theta)^{j+k}} = \frac{c(\theta)}{n(\theta)^l}$$

Since $n(\theta)$ is irreducible, $K/(n_i)$ is a field, so it has no zero divisors, and neither $R_1(\theta)$ and $R_2(\theta)$ are zero.

Our three cases are as before, except that when $k = 1$, our cancellation condition becomes:

$$R_1(\theta) = -R_2(\theta) \mod n(\theta)$$

$$ja(\theta)n'(\theta) = a(\theta)b(\theta) \mod n(\theta)$$

Again, division is possible in a field, so

$$j = \frac{b(\theta)}{n'(\theta)} \mod n(\theta)$$

Our three cases become:

1. $k = 0$ and $j = l - 1$.
2. $k = 1$ and either $j = l - 1$ or $j = \frac{b(\theta)}{n'(\theta)} \mod n(\theta)$.
3. $k > 1$ and $j = l - k$.

Once we have determined the factors and multiplicities in the denominator, then we can proceed as before, clearing out the denominator and obtaining a polynomial equation.

Example 6.9. Determine if $\int x^x dx$ has an elementary form.

To handle this integral, we'll rewrite it as $\int e^{x \ln x} dx$ and operate in the field $\mathbb{C}(\theta, \psi)$, where $\theta = \ln x$ and $\psi = \exp x\theta$. So, we're trying to compute

$$\int \psi \, dx$$

Applying theorem 6.1, we obtain the following Risch equation in $\mathbb{C}(\theta)$:

$$A_1' + k' A_1 = 1$$

$$A_1' + (\theta + 1) A_1 = 1$$

$$A = 1 \qquad B = \theta + 1 \qquad C = 1$$

Since $\deg A = 0$ and $\deg B = 1$, $\deg A \neq \deg B + 1$, so we can't have cancellation on the left hand side. This means that

$$\deg A_1 = \deg C - \max(\deg A - 1, \deg B) = 0 - \max(-1, 1) = -1$$

which is impossible. The only remaining case to consider is if A_1 is constant, which is also impossible.

We conclude that this Risch equation has no solution in $\mathbb{C}(\theta)$, and that the original integral is not elementary.

□

6.4 Risch Equations over Special Polynomials

Finally, let's consider Risch equations over fields with special polynomials, i.e, exponential extensions with $\theta = \exp k$ and $\theta' = k'\theta$.

$$r' + Sr = T \qquad S, T, r \in K(\theta)$$

First, what happens when our partial fractions decomposition yields special polynomials in the denominators of S or T?

$$S = \frac{b}{\theta^k} + \cdots \qquad T = \frac{c}{\theta^l} + \cdots \qquad b, c \in K$$

$$r = \frac{a}{\theta^j} + \cdots \qquad r' = \frac{-jk'a + a'}{\theta^j} + \cdots \qquad a \in K$$

First, we should consider if the leading r' term actually exists. Could the numerator actually be zero? If so, then $jk'a = a'$, but this could only happen if a were a constant multiple of θ^j (PROVE THIS), which contradicts the transcendance of θ over K.

Our Risch equation becomes:

$$\frac{-jk'a + a'}{\theta^j} + \cdots + \frac{ab}{\theta^{k+j}} + \cdots = \frac{c}{\theta^l} + \cdots$$

There are two cases:

1. $k = 0$ and either $j = l$ or $j = \frac{a'+bc}{k'a}$
2. $k > 0$ and $j = l - k$.

Now we have computed j, the multiplicity of the special factor θ in the denominator, and our normal theory from the previous section gives us the denominator multiplicity of our normal factors, so we've computed d, the denominator of r, and thus can replace r with p/d, which will yield a polynomial Risch equation.

There are additional issues that arises with special polynomials when solving a polynomial Risch equation:

$$Ar' + Br = C \qquad A, B, C \in K[\theta] \quad r \in K(\theta)$$

If $K[\theta]$ has only normal polynomials, then this equation can be solved as described before, since r must be a polynomial. In the special case, however, r could have a special denominator. Expanding as before...

$$r = \frac{a}{\theta^j} + \cdots \qquad r' = \frac{-jk'a + a'}{\theta^j} + \cdots$$

If A and B have no θ factors, then their zeroth order coefficients will produce j-th order fractions:

$$A(0)\frac{-jk'a + a'}{\theta^j} + \cdots + B(0)\frac{a}{\theta^j} + \cdots = C$$

Since C is a polynomial, the fractions on the left must cancel, and we obtain:

$$\left[-jak' + a'\right] A(0) + aB(0) = 0$$

$$jk' - \frac{a'}{a} = \frac{B(0)}{A(0)}$$

Integrating, we obtain:

$$jk - \ln a = \int \frac{B(0)}{A(0)} dx$$

We don't know a, but A, B, and k are all known, so solving this equation amounts to an integration step that must result in a constant multiple of k plus a possible logarithm.

This completes our determination of the denominator of r, and we are now reduced to a polynomial equation:

$$Ap' + Bp = C \qquad A, B, C, p \in K[\theta]$$

$$p = p_n\theta^n + \cdots \qquad p' = (p_n' + np_nk')\theta^n + \cdots$$

So the leading term on the left hand side is:

$$\operatorname{lc} A(p_n' + np_nk') + \operatorname{lc} Bp_n = 0$$

$$nk' + \frac{p_n'}{p_n} = -\frac{\operatorname{lc} B}{\operatorname{lc} A}$$

This equation has the same form as ?, so again, we integrate:

$$nk + \ln p_n = -\int \frac{\operatorname{lc} B}{\operatorname{lc} A} dx$$

Example 6.10. (Bronstein examples 6.2.1, 6.3.3, 6.4.2) Integrate

$$\int \frac{e^x - x^2 + 2x}{(e^x + x)^2 x^2} e^{(x^2-1)/(x+1)/(e^x+x)} dx$$

$$\int \frac{e^x - x^2 + 2x}{(e^x + x)^2 x^2} e^{(x-1)/(e^x+x)} dx$$

We'll use the differential field $\mathbb{C}(\theta, \psi)$ where $\theta = \exp x$ and $\psi = \exp \frac{x-1}{\theta+x}$.

$$\int \frac{\theta - x^2 + 2x}{(\theta + x)^2 x^2} \psi dx$$

$k = \frac{x-1}{\theta+x}$ so $k' = \frac{(\theta+x)-(x-1)(\theta+1)}{(\theta+x)^2} = \frac{-x\theta+2\theta+1}{(\theta+x)^2}$. Equation ? gives:

$$A_1' + \frac{-x\theta + 2\theta + 1}{(\theta + x)^2} A_1 = \frac{\theta - x^2 + 2x}{(\theta + x)^2 x^2}$$

$(\theta + x)$ is a normal polynomial, so our normal theory tells us that since $k = 2$ and $l = 2$, then $j = l - k = 0$, so $(\theta + x)$ does not appear in the denominator of A_1. Likewise, x is a normal polynomial for which $k = 0$ and $l = 2$, so $j = l - 1 = 1$, and x can appear in A_1's denominator. Writing:

$$A_1 = \frac{S_1}{x} \qquad A_1' = \frac{S_1' x - S_1}{x^2}$$

and substituting:

$$\frac{S_1' x - S_1}{x^2} + \frac{-x\theta + 2\theta + 1}{(\theta + x)^2} \frac{S_1}{x} = \frac{\theta - x^2 + 2x}{(\theta + x)^2 x^2}$$

$$(\theta + x)^2 (S_1' x - S_1) + x S_1 (-x\theta + 2\theta + 1) = \theta - x^2 + 2x$$

$$(\theta + x)^2 (S_1' x - S_1) + (-x^2\theta + 2x\theta + x) S_1 = \theta - x^2 + 2x$$

$$(\theta + x)^2 x S_1' + (-\theta^2 - x^2\theta - x^2 + x) S_1 = \theta - x^2 + 2x$$

$$A = (\theta + x)^2 x \qquad B = (-\theta^2 - x^2\theta - x^2 + x) \qquad C = \theta - x^2 + 2x$$

Can we have cancelation? To determine if so, we need to integrate:

$$\int \frac{B(0)}{A(0)} dx = \int \frac{-x^2 + x}{x^3} dx = -\ln x - \frac{1}{x}$$

This does not have the form $jk + \ln a(\theta)$, where $k = x$, so cancellation can not occur and we conclude that we have no special denominator.

Now we're reduced to solving a purely polynomial Risch equation over $\mathbb{C}(\theta)$:

$$(\theta + x)^2 x P_1' + (-\theta^2 - x^2\theta - x^2 + x)P_1 = \theta - x^2 + 2x$$

$$A = (\theta + x)^2 x \qquad B = (-\theta^2 - x^2\theta - x^2 + x) \qquad C = \theta - x^2 + 2x$$

Can cancellation occur now? We integrate:

$$-\int \frac{\operatorname{lc} B}{\operatorname{lc} A} dx = \int \frac{x^2 + 1}{x} dx = \frac{1}{2}x^2 + \ln x$$

□

Chapter 7

Algebraic Curves

Having addressed logarithmic and exponential extensions, we now turn to the algebraic extension, which turns out to be completely different in character from the two transcendental cases.

How might we handle a simple algebraic extension? A crucial property of *algebraic functions*, as elements of an algebraic extension are called, is that they admit series expansions everywhere, including infinity, so long as we allow a finite number of negative exponents. Such functions are called *meromorphic*. The logarithm function fails to be meromorphic at the origin, and the exponential function fails to be meromorphic at infinity, but algebraic functions are meromorphic everywhere, including infinity.

This means that around any specific point, we can construct a series expansion of the integrand and integrate termwise to obtain a series expansion for the integral. At first this doesn't seem terribly useful, because series expansions are infinite and we're trying to construct closed-form solutions, but it turns out that only a finite number of places will have negative exponents in their series expansions and that the function is completely specified, up to an additive constant, by the coefficients of the negative powers.

Thus, the basic strategy is first to identify the function's *poles*, the places where its value becomes infinite, and compute the *principal part* of the series expansions there, which are the negative exponents and their coefficients. This is fairly straightforward, though there are issues of computational complexity that make it non-trivial. Then we integrate termwise, which is trivial, and obtain local series expansions at the poles of the solution. Next, we need to reassemble this local information into a global function (if one exists), a *Mittag-Leffler problem*, for which I will present a basic algorithm in this chapter, although more efficient techniques have been developed.

What about the logarithmic terms? This turns out to be the most difficult part of the problem. We can begin to analyze them using the same techniques, by noting that the t^{-1} terms in the principal parts of the integrand lead directly to logarithms in the integral, and furthermore that the coefficients of these terms give us the locations and orders of the poles and zeros in the logarithms. This information specifies an algebraic function up to

a multiplicative constant[1], and our algorithm can be adapted without too much trouble to handle this case.

The problem is that no algebraic function might exist that match a given set of zeros and poles, but increasing the order of the zeros and poles might produce a solution. This corresponds to raising the logarithm term to powers, i.e, $\ln f$ is the same as $\frac{1}{2}\ln f^2$, which is the same as $\frac{1}{3}\ln f^3$, except that in our case the lower powers might not exist in our function field, even though higher powers do. What powers should we use? We could go on raising to higher and higher powers, hoping that something will work, but the only known algorithm to limit this search requires reducing modulo a prime, and that requires techniques that weren't developed until the 1960s. Before heading into *modern algebraic geometry*, let's see how far we can get with the classical algebraic geometry of the nineteenth century.

7.1 Basic Algebraic Geometry

The roots of algebraic geometry lie in studying the zeros of polynomial equations. We began with a single polynomial in a single variable, and have learnt a great deal about it. We know how to solve it (at least in terms of radicals) if its degree is less than 5. Galois proved that no such solution (in radicals) exists (in the general case) for larger degree, though abstract algebra provides us with a suitable general theory to handle this case. Simple long division tells us that it can have no more roots than its degree, and Gauss showed that all of the roots exist as complex numbers — the Fundamental Theorem of Algebra.

The next logical step is to consider zeros of a single polynomial in two variables, and this equation has also received a great deal of attention from mathematicians. Like the univariate case, we have theories devoted to low-order special cases — *linear equations* (all terms first degree or constant), the *conic sections* (all terms second degree or less), and the *elliptic curves* (one term third degree; all others second degree or less). In the general case, $\sum a_{ij}x^i y^j = 0$ is called an *algebraic curve*, and a rational function in x and y is called an *algebraic function*. These will be our main focus of attention in this chapter.

The first problem we face when dealing with algebraic curves is the multi-valued nature of their solutions. Consider, once again, the algebraic function y defined on the algebraic curve $y^2 = x^2 - 1$. There are, in fact, two seperate algebraic functions that solve this equation — both y and $-y$ are solutions. Conventionally, we express this by writing something like $y = \pm\sqrt{x-1}$, but for higher degree curves this kind of notation becomes unsuitable. How, for example, do you express the three possible solutions to a cube root, and how do you deal with the general case where y can appear multiple times in the curve's defining polynomial?

Our solution to this problem is to regard the entire algebraic curve as a two-dimensional surface in a four-dimensional space. Why four dimensions? Well, just as in the univariate

[1]Of course. Due to the presence of a constant of integration, we expect to specify the main part of the integral up to an *additive* constand, and the logarithmic parts of the integral up to a *multiplicative* constant.

case, we find it convenient to work with complex numbers, so as to deal easily with roots of negative numbers. Regarding both x and y as complex numbers (two dimensions each), and plotting them against each other, we obtain a four dimensional space. Just as in the real case, where an equation like $x^2 + y^2 = 1$ defines a circle, an algebraic curve defines a surface, the loci of x and y that satisfy the defining polynomial.

The defining polynomial can be regarded as a polynomial in y, whose coefficients are polynomials in x, simply by collecting terms with like powers of y. For any given value of x, we have a polynomial in y with complex coefficients that yields at most n solutions. We can be more specific. For any given value of x, we have *exactly* n solutions for y *unless* one of two things happen. Either the leading (y^n) coefficient is zero, in which case we have less than n solutions due to having a polynomial of degree less than n, or the polynomial has multiple identical roots, a *multiple point* of the algebraic curve.

Now, the coefficient of y^n in the defining polynomial will be a polynomial in x, which has a finite number of roots at which it is zero, so there are only a finite number of points where the defining polynomial is of degree less than n in y. As x approaches one of these points, the value of the y^n coefficient approaches zero, which causes at least one of the roots to approach infinity. We'll deal with these points by introducing a line at infinity, forming *projective space* and creating a *compact* surface.

Likewise, there are only a finite number of multiple points with multiple identical roots, as can be seen by considering the discriminant of defining polynomial (Theorem ?), regarded as a polynomial in y, with coefficients in $\mathbf{C}[x]$. The discriminant, as the determinant of a matrix with coefficients in $\mathbf{C}[x]$, exists itself in $\mathbf{C}[x]$, and therefore will have only a finite number of points where it is zero. Thus, an algebraic curve has only a finite number of multiple points, which can be found by computing the zeros of simple, univariate polynomials.

Multiple points are further classified according to whether or not the curve is locally Euclidean in their neighborhood. Geometrically, this corresponds to looping around the point until you return to your starting point. If a single such *cycle* covers all the sheets of the curve, the curve is locally Euclidean, and we have an *ordinary point* of the curve, albeit one with *ramification*, the *ramification index* being how many times we had to circle the point. Otherwise, multiple cycles are required to cover all of the sheets, and we have a *singular point*. Analytically, both partial derivatives of the curve's polynomial are zero at a singular point, while at least one is non-zero at ordinary points. This analysis is facilitated by Newton polygons.

Example 7.1. Find the singular points of $y^2 = x^2 - 1$

We normalize the defining polynomial by writing it as $y^2 - x^2 + 1 = 0$, and begin by noting that the coefficient of y^2 is 1, so the defining polynomial is second degree for all finite values of x. Where does it have multiple roots? We compute the discriminant:

$$\mathrm{disc}_y(y^2 - x^2 + 1) = \mathrm{res}_y(y^2 - x^2 + 1, 2y) = \det \begin{vmatrix} 1 & 0 & -x^2+1 \\ 2 & 0 & 0 \\ 0 & 2 & 0 \end{vmatrix} = 4(x^2 - 1)$$

Thus, we conclude that the multiple points of $y^2 = x^2 - 1$ lie at the roots of the discriminant, which are $x = \pm 1$. The partial derivative of the polynomial with respect to x is $-2x$, which is non-zero, so neither of these multiple points are singular.

What about the points at infinity? Introducing the subsitutions $u = x^{-1}$ and $v = y^{-1}$, our curve becomes $u^2 - v^2 + u^2v^2 = v^2(u^2 - 1) + u^2 = 0$, which has a multiple point at $(0, 0)$, since when $u = 0$, both of the roots of $-v^2 = 0$ are identical. Also, both partial derivatives are zero, so this is a singular point of the curve.

□

Any ordinary point can be expanded using a power series in $(x - \alpha)$, which for non-ramified points is a straightforward application of the Implicit Function Theorem.

IFT: [Baby Rudin 9.28; 2-dim complex version] Let f be an analytic mapping of an open set $E \in \mathbb{C}^2$ into $\mathbb{C}$, such that $f(x, y) = 0$ and $\frac{df}{dx} \neq 0$, then an analytic function $g(y)$ exists such that $f(x, g(y)) = 0$.

For infinity and/or poles, substitute z=1/x or v=1/y.

For ramification points, we use substitution of the form $x = t^r + \alpha$, where r is the *ramification index*, then use composition of analytic functions (x is analytic everywhere; y is analytic as a function of x everywhere except at t=0, so y is analytic as a function of t everywhere except at t=0) to establish that y is analytic everywhere on the t-plane except possibly at the origin. Then use existence of the Laurent series (Silverman 11.2) and continuity of the roots (HOW?) to establish analyticity at the multiple point, and consequently existance of a power series, but in t, not $(x - \alpha)$, a *Puiseux series*.

Singular points will admit multiple Puiseux series, each one corresponding to a single cycle. The simplest way to compute Puiseux series is to use Newton polygons to determine ramification, then setup a trial series with the correct ramification and substitute it into the curve's defining equation.

Example 7.2. Find the Puiseux expansions of y at the multiple points of the curve $y^2 = 1 - x^2$

We'll start with the finite zeros of y at $(x, y) = (\pm 1, 0)$. The analysis is almost the same in both cases, so I'll just do (1,0). First, construction of the Newton polygon requires recasting the curve's polynomial into a form centered about the point being analyzed, i.e, $y^2 + (x-1)^2 + 2(x-1) = 0$. The polygon's only non-trivial line segment has slope -2 and width 1, telling us that we'll require a single Puiseux series with ramification index 2:

$$x = t^2 + 1; \qquad x^2 = t^4 + 2t^2 + 1$$

We know that y can be expressed as a power series in t, so we'll write it in that form:

$$y = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + \cdots$$

$$y^2 = a_0^2 + 2a_0a_1t + (2a_0a_2 + a_1^2)t^2 + (2a_0a_3 + 2a_1a_2)t^3 + (2a_0a_4 + 2a_1a_3 + a_2^2)t^4 + \cdots$$

Now, substituting these expressions for x^2 and y^2 into the curve's defining equation $y^2 + x^2 - 1 = 0$ and equating coefficients of like powers of t, we find:

$$a_0^2 = 0 \qquad 2a_0a_1 = 0 \qquad 2 + 2a_0a_2 + a_1^2 = 0$$
$$2a_0a_3 + 2a_1a_2 = 0 \qquad 1 + 2a_0a_4 + 2a_1a_3 + a_2^2 = 0$$

The first equation tells us that $a_0 = 0$, the second equation tells us nothing (becasue $a_0 = 0$), the third equation tells us that $a_1 = \pm\sqrt{2}i$, the fourth equation tells us that $a_2 = 0$ and the fifth equation tells us that $a_3 = \pm\frac{\sqrt{2}}{4}i$, so

$$x = t^2 + 1; \qquad y = \pm\left[\sqrt{2}it + \frac{\sqrt{2}}{4}it^3 + \cdots\right]$$

It would seem that we have two different series to chose from. This is not really the case, as they differ by only a 180° rotation in the t-plane, as can been seen by substituting $t = -t$, which transforms one of the y-series into the other, while leaving the x-series unchanged.

Now, let's analyze the singular point at infinity. Again, we move infinity to a finite point (0) with the substitutions $x = u^{-1}$ and $y = v^{-1}$. Our curve becomes:

$$(u^2 - 1)v^2 - u^2 = 0$$

The Newton polygon has a single line segment, slope -1, length 2, telling us that we'll have two separate cycles, each with ramification index 1. Thus, u can be used directly as a uniformizing variable, and we postulate an expansion for v in the form:

$$v = a_0 + a_1u + a_2u^2 + a_3u^3 + \cdots$$

$$v^2 = a_0^2 + 2a_0a_1u + (2a_0a_2 + a_1^2)u^2 + (2a_0a_3 + 2a_1a_2)u^3 + (2a_0a_4 + 2a_1a_3 + a_2^2)u^4 + \cdots$$

Plugging this into $(u^2 - 1)v^2 - u^2$ and setting all the resulting coefficients to zero, we conclude:

$$a_0 = 0; \qquad a_1 = \pm i; \qquad a_2 = 0; \qquad a_3 = \pm\frac{1}{2}i$$

$$v = \pm iu \pm \frac{1}{2}iu^3 + \cdots$$

This time, without ramification, we actually have two distinct series that will yield two different values of v for each value of u. Inverting back to our original coordinates, we obtain:

$$y^{-1} = \pm\left[ix^{-1} + \frac{1}{2}ix^{-3} + \cdots\right]$$

Yet this is not an expansion for y, nor is it a Puiseux series, since it has an infinite number of negative exponents. We can invert the series (HOW?) to obtain our final result:

□

Example 7.3. Find the principal parts of $\frac{1}{y}$ on the curve $y^2 = 1 - x^2$

Remember that principal parts of an algebraic function are the parts of its series expansion with negative powers at its poles. So, the first step is to locate the function's poles, which in this case is simply the places where the denominator is zero, and that's just $x = \pm 1$. We're already computed a series expansion for y at these points, so we can invert that solution to obtain a series for $\frac{1}{y}$. We do this by noting that since y has a simple zero at $(\pm 1, 0)$, $\frac{1}{y}$ must have a simple pole at these places, postulating a series expansion starting with t^{-1}, multiplying and equating terms:

$$\frac{1}{y} = \pm\left[-\frac{\sqrt{2}}{2}it^{-1} + \cdots\right]$$

□

Example 7.4. Find the principal parts of $\frac{x}{y}\,\mathrm{d}x$ on the curve $y^2 = 1 - x^2$

Differential forms are not functions, and have different series expansions. This is primarily due to the presence of the differential, which must be adjusted at ramification points. Thus, we've already computed a series expansion for $\frac{1}{y}$ at $x = \pm 1$, expressed in terms of t. Now $x = \pm 1 + t^2$, so $\mathrm{d}x = 2t\,\mathrm{d}t$

$$\frac{x}{y}\,\mathrm{d}x = \pm\left[\pm 1 + t^2\right]\left[-\frac{\sqrt{2}}{2}it^{-1} + \cdots\right] 2t\,\mathrm{d}t$$

$$\frac{x}{y}\,\mathrm{d}x = \left[-\sqrt{2}i + \cdots\right]\,\mathrm{d}t$$

In short, even though both $\frac{1}{y}$ and $\frac{x}{y}$ have poles at $x = \pm 1$, $\frac{x}{y}\,\mathrm{d}x$ does not! Its behavior at infinity also requires analysis, since x and $\mathrm{d}x$ both have poles at infinity, but $\frac{1}{y}$ does not (it has a zero at infinity). As before, we'll use $u = \frac{1}{x}$ for a uniformizing variable, so $\mathrm{d}x = -\frac{1}{u^2}\,\mathrm{d}u$

$$\frac{x}{y}\,\mathrm{d}x = \pm\frac{1}{u}\left[iu + \frac{1}{2}iu^3 + \cdots\right]\left[-\frac{1}{u^2}\,\mathrm{d}u\right]$$

$$= \mp \left[iu^{-2} + \frac{1}{2} i + \cdots \right] \mathrm{d}u$$

$\square$

Chapter 9

Simple Algebraic Extensions

We now turn to the algebraic extension. Theorem ? allows us to collapse any two adjacent algebraic extensions together, so we need only consider an algebraic extension over a transcendental extension. The most basic case, the one that we will study in this chapter, is when the integrand involves only polynomials and a single root, so we are integrating on an algebraic curve and our algebraic extension occurs directly over the variable of integration: $\mathrm{C}(x, y)$. However, many of this chapter's basic results are applicable in the more general case where we have a series of field extensions that end in a transcendental (exponential or logarithmic) extension followed by an algebraic extension. I'll use the notation $\mathrm{K}(\theta, y)$ to emphasize when this is the case.

9.1 Integral Elements

Definition 9.1. *An element $f \in \mathrm{K}(\theta, y)$ is* **integral** *if it satisfies a monic polynomial with coefficients in $\mathrm{K}[\theta]$.*

Intuitively, an integral element is one with no finite poles. To see this, at least in the case where $K = \mathbf{C}$, define $z = \frac{1}{f}$ and substitute this into f's monic polynomial:

$$f^n + a_{n-1}f^{n-1} + \cdots + a_1 f + a_0 = 0$$

$$z^{-n} + a_{n-1}z^{-n+1} + \cdots + a_1 z^{-1} + a_0 = 0$$

$$1 + a_{n-1}z + \cdots + a_1 z^{n-1} + a_0 z^n = 0$$

Now, if z is zero at a place p over $x = x_0$, at least one of this polynomial's roots must be zero at $x = x_0$. Since all of the a_i are finite at $x = x_0$ (they are polynomials), multiplying

any of them by zero yields zero, so substituting in $z = 0$ yields $1 = 0$. We conclude that z can not be zero, and thus f can not have a pole over p.

What's special about infinity? Why not exclude some other place? Well, nothing's all that special about infinity. We've already seen how a birational transformation can be used to swap infinity with any finite point. Demanding that a field element have no poles anywhere is too restrictive, because Theorem ? tells us that such an element must be constant. So we want to relax this requirement slightly by allowing poles over a single point. We use infinity because it's convenient.

It's not always obvious from inspection which functions are integral. Something like $\frac{y}{x}$, which appears to have a pole at $x = 0$, is actually integral if, for example, $y^2 = x^3$. Then we can consider squaring $\frac{y}{x}$ to obtain $\frac{y^2}{x^2} = \frac{x^3}{x^2} = x$. If the square is finite, then the original function had to be finite (you can't square infinity and get a finite value), so we conclude that $\frac{y}{x}$ is, in fact, globally integral in $\mathbf{C}(x, y); y^2 = x^3$, as it satisfies the monic polynomial $f^2 - x = 0$.

Unfortunately, we have no straightforward means to construct such a polynomial, or prove that one doesn't exist, for any particular function f. To test a function to determine if it is integral, we'll need a more advanced approach.

9.2 Modules

We'll resort now to *modules*, a fairly important algebra concept backed by a substantial body of theory, upon which I shall only draw as needed. General references include [Atiyah+McDonald] and [Lang].

Definition 9.2. *An R-module over a ring R is an additive group M acted on by R (i.e, there is a mapping $R \times M \to M$) in a distributive manner:*

$$(r_1 + r_2)m = r_1 m + r_2 m \qquad r_1, r_2 \in \mathrm{R};\ m \in \mathrm{M}$$

where we have adopted the usual convention of writing R's action on M as a multiplication.

Definition 9.3. *A free R-module is an R-module spanned by a linearly independent basis $\{b_1, b_2, ...b_n\}$. It consists of all elements formed as follows:* [1]

$$a_1 b_1 + a_2 b_2 + ... + a_n b_n; \qquad a_i \in R$$

Not all modules have a finite set of generators, and not all those have a linearly independent set of generators. Elements formed from a basis can be added by using the module's

[1] I'll also note that a multiplication rule needs to be specified between the basis elements and the elements of the ring, and an addition rule between the elements of the module. Also, the expression has to be *unique* — you can't be able to write an element two different ways. In our case, these rules are obvious, but that's not always the case.

distributive property to factor out the coefficients from of each basis element and then performing the addition in the ring R:

$$\begin{aligned}&(a_1b_1 + a_2b_2 + ... + a_nb_n) + (c_1b_1 + c_2b_2 + ... + c_nb_n)\\ &= (a_1 + c_1)b_1 + (a_2 + c_2)b_2 + ... + (a_n + c_n)b_n\end{aligned}$$

So the elements generated from a basis clearly form a module. R operates on them by multiplication by every coefficient.

Example 9.4.

An ideal I in a ring R is a R-module, but a subring S of R, in general, is not, because multiplication by an element of R might not produce a result in the subring. R, however, can always be viewed as an S-module.

□

Note that it is vitally important to specify the ring used for the coefficients. For example, consider the basis $\{1, y\}$. Treating this as a $\mathbf{C}(x)$-module, I can form $\frac{y}{x} = \frac{1}{x}y$, since $\frac{1}{x} \in \mathbf{C}(x)$. However, $\frac{y}{x}$ does *not* belong to the $\mathbf{C}[x]$-module generated by $\{1, y\}$. I would need to use polynomial coefficients to form a $\mathbf{C}[x]$-module, not the rational functions coefficients allowed in a $\mathbf{C}(x)$-module. We'll be primarily interested in $\mathrm{K}[\theta]$-modules, $\mathrm{K}(\theta)$-modules, and $\mathcal{I}$-modules, where $\mathcal{I}$ is the ring of integral elements in $\mathrm{K}(\theta, y)$.

9.3 The $\mathrm{K}[\theta]$-module $\mathcal{I}$

Since polynomials have no finite poles, they are integral elements, and thus $\mathrm{K}[\theta] \subseteq \mathcal{I}$. Thus, $\mathcal{I}$ (the ring of integral elements) is trivially a $\mathrm{K}[\theta]$-module (see Example 9.4), but what is not nearly so obvious is that it is also a free module, a fact which underlies a great deal of our theory. I'll prove this first by showing that $\mathcal{I}$ is finitely generated as a $\mathrm{K}[\theta]$-module, then showing the existance of a linearly independent set of generators.

Let's start with a preliminary theorem.

Theorem 9.5. *If $\{w_1, \ldots, w_n\}$ is a basis for a finite separable field extension E/K, then a dual basis $\{u_1, \ldots, u_n\}$ can be constructed such that $\mathrm{Tr}(w_iu_j) = \delta_{ij}$. ([Lang] Corollary VI.5.3)*

Proof

Consider the following matrix:

$$M = \begin{pmatrix} \mathrm{Tr}(w_1w_1) & & \\ \vdots & \ddots & \\ \mathrm{Tr}(w_1w_n) & \cdots & \mathrm{Tr}(w_nw_n) \end{pmatrix}$$

Now take an element $x \in E$, and represent it relative to the basis $\{w_1, \ldots, w_n\}$ as a row vector $X = (x_i)$. Multiplying XM produces a row vector whose j^{th} element can be written:

$$\sum_i x_i \mathrm{Tr}(w_j w_i) = \mathrm{Tr}(w_j \sum_i x_i w_i) = \mathrm{Tr}(w_j x) = \mathrm{Tr}_x(w_j)$$

where I used first the K-linearity and additive distributive properties of Tr, then wrote $\mathrm{Tr}_x : f(a) = \mathrm{Tr}(ax)$ to emphasize that I'm regarding Tr_x as a linear form in $\mathrm{Hom}_E(E, K)$. So, if M is singular, then there exists some non-zero element x such that Tr_x is zero for all of w_i, which form a basis set, so Tr_x must therefore be the zero map. This can only happen if Tr is identically zero, which would be the case for an inseparable extension. For the separable case, therefore, M must be invertible, and we can write:

$$M^{-1}M = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

A moment's thought now shows that the rows of M^{-1} are the desired dual basis elements, written with respect to $\{w_1, \ldots, w_n\}$.

□

Theorem 9.6. *$\mathcal{I}$ is a finitely generated $\mathrm{K}[\theta]$-module. ([A+MacD] Proposition 5.17; [Lang] Exercise VII.3)*

Proof

Regarding $\mathrm{K}(\theta, y)$ as a vector space over $\mathrm{K}(\theta)$, we can easily construct a basis of integral elements by starting with $\{1, y, \ldots, y^{n-1}\}$ and multiplying each element (if needed) by a polynomial in θ which cancels all of its poles:

$$\mathrm{K}(\theta, y) = \mathrm{K}(\theta)\{w_1, \ldots, w_n\} \qquad \forall i(w_i \in \mathcal{I})$$

Using Theorem 9.5, construct a dual basis $\{u_1, \ldots, u_n\}$ so that $\mathrm{Tr}(w_i u_j) = \delta_{ij}$. Take any $x \in \mathcal{I}$ and write it using the dual basis:

$$x = \sum_i a_i u_i \qquad a_i \in \mathrm{K}(x)$$

Now consider $\mathrm{Tr}(xw_j)$. Now, x and w_j are both in $\mathcal{I}$, so xw_j is in $\mathcal{I}$, and therefore has a monic minimum polynomial with coefficients in $\mathrm{K}[\theta]$. Since Tr equals some integer multiple of the negative of the second coefficient in a monic minimum polynomial, $\mathrm{Tr}(xw_j) \in \mathrm{K}[\theta]$. But also,

$$\mathrm{Tr}(xw_j) = \mathrm{Tr}(\sum_i a_i u_i w_j) = \sum_i \mathrm{Tr}(a_i u_i w_j) = \sum_i a_i \mathrm{Tr}(u_i w_j) = a_i$$

which establishes that $\forall i(a_i \in \mathrm{K}[\theta])$, so

$$\mathcal{I} \subseteq \mathrm{K}[x]\{u_1, \ldots, u_n\}$$

$\mathrm{K}[\theta]$ is a Noetherian ring (Theorem ??), so $\mathrm{K}[\theta]\{u_1, \ldots, u_n\}$ is a Noetherian module (Theorem ??), which means that $\mathcal{I}$, as a submodule, is Noetherian and thus finitely generated (Theorem ??).

□

Theorem 9.7. *Any submodule of a finite free module over a principal ideal ring is free. ([Lang] Theorem III.7.1)*

Proof

Let $F = R\{w_1, \ldots, w_n\}$ be a free R-module (R a principal ideal ring) with a submodule M. Consider $F_i = R\{w_1, \ldots, w_i\}$, the free R-module generated by the first i basis elements, and $M_i = F_i \cap M$. We will show inductively that all of the M_i are free R-modules, and since $F_i = F$ and $M_i = M$, this will prove the theorem.

First, consider $M_1 = R\{w_1\} \cap M$. If M_1 is not empty (and thus free), then any $m \in M_1$ can be written rw_1. Since a module forms an additive group, and we can operate on the module using all the elements of R, it follows that all the r's must form an ideal, and since R is principal, that ideal can be written with a single generator, say (r_1), and $M_1 = R\{r_1 w_1\}$ (or is empty).

Now, assume that M_j is free for all $j < i$. Consider all $x \in M_i$, which can be written $r_1 w_1 + \cdots + r_i w_i$. Either $M_i = M_{i-1}$ (and is therefore free), or at least some of the r_i are non-zero. By the same rationale as the last paragraph, these r_i form an ideal, which can be written (r_i). Take any element $x \in M_i$ with its i^{th} coefficient r_i and add it to M_{i-1}'s basis set to form a basis set for M_i, since some multiple of this element can be used to cancel any i^{th} coefficient from an element in M_i and leave an element in M_{i-1} which can be formed using the remaining basis elements.

□

A *torsion-free* module has no "zero divisors", in the sense that no non-zero element of its associated ring can operate on a non-zero element of the module and produce zero. Since fields are torsion-free, and all of our modules are subsets of the field $\mathrm{K}(\theta, y)$, they are all torsion-free.

Theorem 9.8. *Any finitely generated, torsion-free module M over a principal ideal ring R is free. ([Lang] Theorem III.7.3)*

Proof

Take a maximal set of M's linearly independent generators $\{w_1, \ldots, w_n\}$ and any remaining generators $\{y_{n+1}, \ldots, y_m\}$. Every y_i is therefore linearly dependant on $\{w_1, \ldots, w_n\}$:

$$a_i y_i + r_1 w_1 + \cdots + r_n w_n = 0 \qquad a_i \neq 0$$

Take the product of all the a_i's: $a = a_{n+1} a_{n+2} \cdots a_m$ and consider the mapping $x \mapsto ax$ which is injective, since $a \neq 0$ and the module is torsion-free, so therefore maps M to aM, an isomorphic image which is a submodule of the free module $R\{w_1, \ldots, w_n\}$. By Theorem 9.7, aM is therefore free, and since it is isomorphic to M, we can take a basis of aM, divide all of its basis elements by a (they are all multiples of a), and obtain a basis for M.

□

Now, since $\mathrm{K}[\theta]$ is a principal ideal ring (Theorem ??), Theorems 9.6 and 9.8 demonstrate that $\mathcal{I}$ is a free $\mathrm{K}[\theta]$-module.

Definition 9.9. *A basis for $\mathcal{I}$ will be called an* **integral basis**.

Theorem 9.10. *Any integral basis is also a basis for the $\mathrm{K}(\theta, y)$ field as a $\mathrm{K}(\theta)$-module.*

□

While the preceding theorems offer an existance proof for an integral basis, it is not immediately clear how to obtain one for any particular field, and in fact the calculation of an integral basis ultimately becomes one of the biggest computational barriers in this theory. Therefore, I will defer a more detailed discussion until a later chapter, and instead present a simple construction for the special case of a simple radical extension.

9.4 Basis for all Rational Functions

The first kind of basis we're interested in, a *basis for all rational functions*, is one than spans the entire $\mathcal{C}(x, y)$ field as a $\mathcal{C}(x)$-module. In other words, we're looking for a basis $\{b_1, b_2, ...b_n\}$ so that everything in $\mathcal{C}(x, y)$ can be expressed in the form:

$$a_1 b_1 + a_2 b_2 + ... + a_n b_n; a_i \in \mathcal{C}(x)$$

Such a basis will always have n elements, where n is the degree of the $\mathcal{C}(x, y)$ extension over $\mathcal{C}(x)$, and can be most conveniently characterized using its *conjugate matrix*:

Definition 9.11. *The* **conjugates** *of a rational function $\eta(x, y)$ in $\mathbf{C}(x, y)$ are the functions formed by replacing y with its conjugate values.*

The **trace** *of a rational function $\eta(x, y)$ is the sum of its conjugates:*

$$\mathrm{T}(\eta(x,y)) = \sum_i \eta(x,y_i)$$

The **norm** *of a rational function $\eta(x,y)$ is the product of its conjugates:*

$$\mathrm{N}(\eta(x,y)) = \prod_i \eta(x,y_i)$$

Both the trace and norm, as symmetric functions in $y_1, ..., y_n$, are functions in $\mathbf{C}(x)$.

The **conjugate matrix** $\mathbf{M}_\omega$ *of n elements ω_i in $\mathcal{C}(x,y)$, where n is the degree of $\mathcal{C}(x,y)$ over $\mathcal{C}(x)$, is the matrix whose each row consists of the n conjugate values of a single element, and whose n rows are formed in this way from the n elements.*

A set of n elements $\omega_i \in \mathbf{C}(x,y)$ form a **rational function basis** *for $\mathbf{C}(x,y)$ if the determinant of their conjugate matrix is non-zero: $|\mathbf{M}_\omega| \neq 0$*

Definition 9.12. *For any function $\eta \in \mathbf{C}(x,y)$ and any rational function basis ω_i, the* **trace vector** $\mathbf{T}_{\eta/\omega} = \Big(\mathrm{T}(\eta\omega_i)\Big)$ *of η relative to ω is formed from the traces of the n products of η with the n functions ω_i.*

The **conjugate vector** $\mathbf{C}_\eta = (\eta(x,y_i))$ *is formed from the n conjugates of η.*

Theorem 9.13. *For any function $\eta \in \mathbf{C}(x,y)$ and any rational function basis ω_i, if $\mathbf{T}_{\eta/\omega}$ is the zero vector, then η is zero.*

Proof

$\mathbf{T}_{\eta/\omega}$, $\mathbf{M}_\omega$ and $\mathbf{C}_\eta$ satisfy the matrix equation

$$\mathbf{T}_{\eta/\omega} = \mathbf{M}_\omega \mathbf{C}_\eta$$

since each row of this matrix equation has the form

$$\mathrm{T}(\eta\omega_i) = \sum_j \omega_i(x,y_j)\eta(x,y_j)$$

Since $\mathbf{M}_\omega$ is invertible (since its determinant is non-zero), if $\mathbf{T}_{\eta/\omega}$ is identically zero, then so must be $\mathbf{C}_\eta$, and η is the first element in $\mathbf{C}_\eta$.

□

Theorem 9.14. *A rational function basis ω_i spans $\mathbf{C}(x,y)$ as a $\mathbf{C}(x)$-module. ([Bliss], Theorem 19.1)*

Proof

Note that when we multiply $\mathbf{M}_\omega$ by its transpose $\mathbf{M}_\omega^T$, the ij^{th} element of $\mathbf{M}_\omega\mathbf{M}_\omega^T$ is:

$$\sum_k \omega_i(x, y_k)\omega_j(x, y_k) = \mathrm{T}(\omega_i\omega_j)$$

Since $|\mathbf{M}_\omega|$ is non-zero, $|\mathbf{M}_\omega^T|$ is non-zero, and $|\mathbf{M}_\omega\mathbf{M}_\omega^T|$ is non-zero, so given any function $\eta \in \mathbf{C}(x, y)$, we can solve the following equation for $\mathbf{R}$:

$$\mathbf{T}_\eta = \mathbf{M}_\omega\mathbf{M}_\omega^T\mathbf{R}$$

each of row of which reads:

$$\mathrm{T}(\eta\omega_i) = \sum_j \mathrm{T}(\omega_i\omega_j)r_j$$

Since both $\mathbf{T}_\eta$ and $\mathbf{M}_\omega\mathbf{M}_\omega^T$ are composed of nothing but traces, they exist in $\mathbf{C}(x)$, so $\mathbf{R}$ must also exist in $\mathbf{C}(x)$ and its elements therefore commute with the trace:

$$\mathrm{T}(\eta\omega_i) = \sum_j \mathrm{T}(r_j\omega_j\omega_i)$$

Since the trace of a sum is the sum of the traces:

$$\mathrm{T}(\eta\omega_i) = \mathrm{T}(\sum_j r_j\omega_j\omega_i)$$

$$\mathrm{T}((\eta - \sum_j r_j\omega_j)\omega_i) = 0$$

which implies that $\eta = \sum_j r_j\omega_j$, by Theorem 9.13, and since we've already shown that the r_j are rational functions in $\mathbf{C}(x)$, this proves the theorem.

□

Let me illustrate with a simple example.

Example 9.15. Consider the basis $\{1, y\}$ over the field $\mathcal{C}(x, y); y^2 = x$. The conjugate value of y is $-y$ (PROVE THIS), so the conjugate matrix is:

$$C = \begin{pmatrix} 1 & 1 \\ y & -y \end{pmatrix}$$

and its determinant:

$$\det C = \begin{vmatrix} 1 & 1 \\ y & -y \end{vmatrix} = -2y$$

Since $-2y$ is not zero, we conclude that $\{1, y\}$ is a basis for all rational functions over $\mathcal{C}(x, y); y^2 = x$.

□

Notice that I didn't ask whether $-2y$ was zero at some place in the field. The determinant of the conjugate matrix can be zero at certain places; in fact, often is. It just can't be *identically* zero; i.e, it can't be zero *everywhere*. If this isn't clear, reread Theorems 9.13 and 9.14, noting that all the matrices are defined over the *fields* $\mathbf{C}(x)$ and $\mathbf{C}(x, y)$, where the only zero element is 0.

9.5 Divisors and Integral Modules

In $\mathbf{C}(x)$, we were working with the quotient field of a principal ideal ring, so we could always find a single function to generate any finitely generated $\mathbf{C}[x]$-module, simply by putting all the generators over a common denominator, then taking the G.C.D. of the numerators.

In $\mathrm{K}(\theta, y)$, we are no longer working with a principal ideal ring, so we can't guarantee that any particular ideal can be generated by a single function, but it turns out that every ideal can be generated by a *pair* of functions. Our course of attack is first to construct that pair of functions, then use them to determine if in fact the ideal is principal.

Definition 9.16. *An* **integral module** *(or $\mathcal{I}$-module) is a module formed over $\mathcal{I}$, the ring of integral elements in* $\mathrm{K}(\theta, y)$.

Since $\mathcal{I}$ itself can be expressed as a $\mathrm{K}[\theta]$-module using an integral basis, any $\mathcal{I}$-module is also a $\mathrm{K}[\theta]$-module. Not all $\mathrm{K}[\theta]$-modules are $\mathcal{I}$-modules, however, since $\mathcal{I}$ is typically larger than $\mathrm{K}[x]$.

Some authors use the term *fractional ideal* to refer to an $\mathcal{I}$-module. I have avoided use of this term for two reasons. First, I wish to emphasize the concept of a module. Second, $\mathcal{I}$-modules are not ideals, either in the ring $\mathcal{I}$ (since they may contain elements not in $\mathcal{I}$), nor in the field $\mathrm{K}(\theta, y)$, since, as a field, $\mathrm{K}(\theta, y)$ has only the trivial ideals. The term *fractional ideal* is used because an $\mathcal{I}$-module can be regarded as a fraction of ideals in $\mathcal{I}$.

Theorem 9.17. *$\mathcal{I}$ is a Noetherian ring.*

Proof

Since K is a field, $\mathrm{K}[\theta]$ is a Noetherian ring by the Hilbert basis theorem, $\mathrm{K}[\theta] \subseteq \mathcal{I}$, and $\mathcal{I}$ is finitely generated as a $\mathrm{K}[\theta]$-module, so $\mathcal{I}$ is a Noetherian ring by [Atiyah+McDonald] Proposition 7.2.

□

Theorem 9.18. *The order of the norm of f at a point θ_0 is the sum of the orders of f at all places over θ_0.*

□

Theorem 9.19. *A function can always be constructed with a simple zero at a specified finite, ordinary place (α, β), zero order at an additional finite set of finite, ordinary places Σ, and non-negative order at all other finite places.*

Proof

Begin with the function $(x - \alpha)$, which is a uniformizing variable and thus has a simple zero at (α, β). If none of the other places in Σ have x-value α, then we are done, since $(x - \alpha)$ has no finite poles.

Otherwise, compute $\frac{(x-\alpha)}{(y-\beta)}$ at all places in Σ that do *not* have $y = \beta$. Select a number γ different from all of these values. The function $(x - \alpha) - \gamma(y - \beta)$ has no finite poles and is non-zero at all places in Σ, but it may now have a zero of higher order at (α, β). Consider a series expansion of y in terms of $(x - \alpha)$:

$$y = \beta + c_1(x - \alpha) + c_2(x - \alpha)^2 + \cdots$$

So long as γ is also selected different from c_1, $(x - \alpha) - \gamma(y - \beta)$ will have a first order zero at (α, β) and meet all requirements of the theorem. The simplest way to do this is to pick a value for γ, use Theorem 9.18 to check if the function has a simple zero, and if not, choose a different value for γ.

□

Theorem 9.20. *A function can always be constructed with a simple pole at a specified finite, ordinary place (α, β), zero order at an additional finite set of finite, ordinary places Σ, and non-negative order at all other places.*

Proof

Begin with the function:

$$\frac{f(\alpha, y)}{(x - \alpha)(y - \beta)}$$

where $f(x, y)$ is the minimum polynomial of the algebraic extension. Note that the division by $(y - \beta)$ will always be exact, since $f(\alpha, \beta) = 0$. So we have a rational function $\frac{P(y)}{(x-\alpha)}$, where $P(y)$ is a polynomial in y. It has a simple pole at (α, β), as can be seen from a series expansion in $x - \alpha$ (again, a uniformizing variable). Since the $y - \beta$ factor has been divided out of $f(\alpha, y)$, the numerator is non-zero at (α, β), so the leading term in the series expansion involves $(x - \alpha)^{-1}$, and the pole is thus simple.

This function is finite at all other places, which is obvious except when $x = \alpha$ and $y \neq \beta$, where it takes the form $\frac{0}{0}$, so we can expand it using L'Hôpital's rule:

$$\lim_{(x,y)\to p} \frac{P(y)}{(x-\alpha)} = \lim_{(x,y)\to p} \frac{\frac{dP(y)}{dx}}{\frac{d(x-\alpha)}{dx}} = P'(y)\,\frac{dy}{dx}$$

where $'$ denotes differentiation with respect to the polynomial's variable. $P'(y)$ is a polynomial, and is thus finite where y is finite, as is $\frac{dy}{dx}$ (consider a series expansion of y in terms of $(x-\alpha)$, since all places in Σ are finite and non-singular). It follows that the function is at least finite everywhere except at (α, β).

A more algebraic way to prove this is to note that $f(x, y)$ has a simple zero at every place over $x = \alpha$ (assuming there are no multiple points over $x = \alpha$), so $P(y)$ will have a simple zero at every place over $x = \alpha$ except (α, β), which will exactly cancel the simple pole from $(x - \alpha)$.

Now, compute the value of the function at all other places in Σ, using either L'Hôpital's rule or Puiseux expansion if some of these are over α. If the value of the function is non-zero at all of these places, then we are done. Otherwise, select a number γ different from all of these values. The function:

$$\frac{f(\alpha, y)}{(x-\alpha)(y-\beta)} - \gamma$$

has the desired properties, since it still has a simple pole at (α, β), has no other poles, and is now non-zero at all places in Σ.

We can avoid computing any expansions by picking random values of γ, and using Theorem 9.18 to check for any extra zeros. Since only a finite number of γ values produce extra zeros, this process is guaranteed to terminate.

□

Theorem 9.21. *A function can always be constructed with specified integer orders at a finite set of finite, non-singular places Σ and non-negative order at all other finite places.*

Proof

For each pole or zero, use Theorems 9.19 or 9.20 to construct a function with a simple pole or a simple zero at that place, zero order at all other places in Σ and non-negative order at all other finite places. Raise each of these function to the integer power that is the order of the corresponding pole or zero, then multiply them all together.

□

Definition 9.22. *A* **finite multiple** *of a divisor is a function with order equal to or greater than that required by the divisor at all finite places.*

For the remainder of this section, I'll assume that our divisors involve only finite, ordinary places, which can always be guaranteed in the case of the integration theory.

Theorem 9.23. *For any divisior $\mathcal{D}$ and any finite, non-singular place $\mathfrak{p}$, at least one finite multiple of $\mathcal{D}$ exists with order at $\mathfrak{p}$ exactly that required by $\mathcal{D}$.*

Proof

Use Theorem 9.21 with the zeros and poles required by $\mathcal{D}$, adding $\mathfrak{p}$ to Σ if necessary.

□

Theorem 9.24. *There is a one-to-one relationship between finitely generated integral modules and divisors. Such a module consists of all finite multiples of its associated divisor, and the order of a module's divisor at every finite place is the minimum of the orders of the module's generators at that place.*

Proof

For a given divisor $\mathcal{D}$, consider the set $\mathcal{M}(\mathcal{D})$ of all finite multiples of $\mathcal{D}$. Now, adding two elements can not reduce their order at any finite place, nor can multiplying an element by an integral element $i \in \mathcal{I}$, so $\mathcal{M}(\mathcal{D})$ is clearly an $\mathcal{I}$-module, but it might not be finitely generated.

Since $\mathcal{D}$ has only a finite number of poles, we can always construct a function with order equal or less than that of $\mathcal{D}$ at all finite places simply by taking the inverse of the polynomial $r = (x-p_1)^{m_1} \cdots (x-p_n)^{m_n}$ where p_i are the x-coordinates of the poles in $\mathcal{D}$ and m_i are their multiplicities. For any $m \in \mathcal{M}(\mathcal{D})$, mr is integral, so $\mathcal{M}(\mathcal{D}) \subseteq \mathcal{I}\{r^{-1}\}$, where $\mathcal{I}\{r^{-1}\}$ is the $\mathcal{I}$-module generated by r^{-1}. Now, since $\mathcal{I}\{r^{-1}\}$ is a finitely generated module over a Noetherian ring (remember Theorem 9.17), $\mathcal{I}\{r^{-1}\}$ is a Noetherian module by [Atiyah+McDonald] Proposition 6.5, and $\mathcal{M}(\mathcal{D})$ must also be a finitely generated $\mathcal{I}$-module by [Atiyah+McDonald] Proposition 6.2.

Let $(b_1, ..., b_n)$ be an $\mathcal{I}$-module basis for $\mathcal{M}(\mathcal{D})$. Since there is no way to lower the orders of an element using $\mathcal{I}$-module constructions, and by Theorem 9.23 for each place there is at least one function in $\mathcal{M}(\mathcal{D})$ with order exactly that required by $\mathcal{D}$, it follows that for each place there must be at least one basis element with exactly the order required by $\mathcal{D}$. Futhermore, no basis element can have an order less than required by $\mathcal{D}$ at any finite place, since that element would not be a finite multiple of $\mathcal{D}$. Therefore, at each place $\mathfrak{p}$, the minimum of the orders of the basis elements must be exactly the order required by $\mathcal{D}$.

Conversely, given a finitely generated $\mathcal{I}$-module M, construct its associated divisor $\mathcal{D}$ by taking at every place the minimum of the orders of the module's generators at that place. Clearly, $M \subseteq \mathcal{M}(\mathcal{D})$, but some finite multiple of $\mathcal{D}$ might not be in M.

To eliminate this possibility, take the module's generators, say $\{b_1, b_2, b_3\}$ and expand them into a set where each additional generator beyond the first lowers the module's order by one at a single place, say $\{b_1, b_2', b_2, b_3'', b_3', b_3\}$. These additional generators can be constructed by multiplying the original generators by integral elements (constructed using Theorem 9.21) to remove any additional poles, so $b_2' = i_2' b_2$, where $i_2' \in \mathcal{I}$.

This new module M' clearly has the same associated divisor as M, and I'll now show inductively that any $f \in \mathcal{M}(\mathcal{D})$ can be found in M'. Let $\mathcal{D}_n$ be the divisor associated with the first n basis elements of M'. Clearly, any finite multiple of $\mathcal{D}_1$ can be constructed as integral element times b_1, so let's now assume that any finite multiple of $\mathcal{D}_{n-1}$ can be constructed with the first $n-1$ generators, and consider the n^{th} generator g_n. It lowers the order by one at a single place, so any $f \in \mathcal{M}(\mathcal{D}_n) - \mathcal{M}(\mathcal{D}_{n-1})$ must have exactly the same order as g_n. Multiplying g_n by a suitable constant (the ratio of coefficients in f and g_n's series expansion at this place) will exactly cancel this pole, so $f - cg \in \mathcal{M}(\mathcal{D}_n)$.

So any $f \in \mathcal{M}(\mathcal{D})$ can be constructed using the integral module M'. Writing this construction in matrix form shows how f can be constructed as an M-module element:

$$\begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} \begin{pmatrix} b_1 \\ b'_2 \\ b_2 \\ b'_3 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & i'_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & i'_3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a'_1 & \cdots & a'_3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Consider such a finite multiple f. For every $m \in \mathcal{M}$ (in particular, its basis elements), f must have lower order than m at at least one finite place $\mathfrak{p}$, since otherwise $i = fm^{-1}$ would be integral and f would exist in $\mathcal{M}$ as mi. Yet $\mathcal{D}$, by definition, is the minimum of the orders of $\mathcal{M}$'s basis elements at every finite place. Therefore f can not have lower order than a basis element at any finite place and thus can not exist.

□

Theorem 9.24 shows that an $\mathcal{I}$-module is associated with every divisor, but now we need a constructive procedure for forming an $\mathcal{I}$-module basis for a given divisor.

Theorem 9.25. *Given a divisor $\mathcal{D}$, a pair of functions can always be constructed that generate the divisor's associated integral module.*

Proof

Use Theorem 9.21 to construct a function f with the divisor's required poles and zeros, zero order at all other places conjugate to those poles and zeros, and non-negative order elsewhere. Construct g, a polynomial in x with n-th order roots at all points under n-th order zeros. (f, g) is the required basis. The only finite poles are those of f and g has zero order everywhere except at f's zeros and their conjugates, so by Theorem 9.24, (f, g) forms a basis for $\mathcal{D}$'s associated $\mathcal{I}$-module.

□

Of course, the whole point here is to actually find a function with a specified set of zeros and poles, so once we have constructed a basis for a divisor's associated integral module, we need to determine if the module is principal. Since the total order of a field element is always zero, this only makes sense for divisors of zero order, since divisors of non-zero

order can never be principal. Futhermore, since an integral module corresponds to *finite* multiples of a divisor, we can't use this technique to find functions with poles or zeros at infinity, but that isn't a serious limitation since if we need such a function, we can just transform into a field with a different point at infinity.

Since a finite multiple of a divisor differs from an exact multiple in that the finite multiple can have additional finite zeros, and thus additional infinite poles (since they always balance), a zero order $\mathcal{I}$-module is principle iff it contains a function with no poles at infinity. We can determine this by expressing the $\mathcal{I}$-module as a $\mathrm{K}[x]$-module, simply by multiplying the $\mathcal{I}$-module basis through by an integral basis (remember that an integral basis is simply a basis for $\mathcal{I}$ as a $\mathrm{K}[x]$-module).

We now transform this $\mathrm{K}[x]$-module basis to make it *normal at infinity*, i.e, to ensure that poles don't cancel between terms. First, we use a series of row-equivalent transformations to reduce our $2n$ basis elements to n elements, then additional transformations to make it normal.

We then check these basis elements to see if one of them has no poles at infinity. The most straightforward way to do this is to invert the field using $z = \frac{1}{x}$, which swaps zero with infinity. We can then construct an integral basis for the inverse field, and express each of the module's basis elements (after inverting them) using this inverse basis. If there are any poles at infinity in the original field, they will appear as poles at zero in the inverse field, and can easily be detected by checking if $z = 0$ is a zero of the denominators.

Finally, let me note that since g in Theorem 9.25 is a polynomial, it always has poles at infinity (unless the divisor has no zeros, and is thus trivially constant), and can thus be excluded from consideration. We need only look at the function f, and perhaps not even all of its integral multiples (CHECK THIS).

Theorem 9.26. *Let $f\,dx$ be a differential with order greater than or equal to -1 at some place $\mathfrak{p}$ with branching index r centered at x_0. The residue of $f\,dx$ at $\mathfrak{p}$ is equal to the value of the function $r(x - x_0)f$ at $\mathfrak{p}$. ([Trager], p. 56, taken almost verbatim)*

Proof

Let t be a uniformizing parameter at $\mathfrak{p}$. Since $x - x_0$ has order r at $\mathfrak{p}$, it can be written as

$$x - x_0 = t^r g$$

where g has order zero at $\mathfrak{p}$

$$dx = (rt^{r-1}g + t^r \frac{dg}{dt})dt$$

Since $\frac{dg}{dt}$ has non-negative order at $\mathfrak{p}$, dx has order $r - 1$ at $\mathfrak{p}$ and f must have order greater than or equal to $-r$ at $\mathfrak{p}$

$$f\,dx = rt^{r-1} f g\,dt + t^r f(\frac{dg}{dt})dt$$

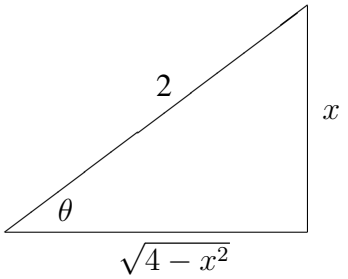
the second term on the right side is holomorphic at $\mathfrak{p}$ so the residue of $f\,dx$ at $\mathfrak{p}$ is the same as the residue of the first term on the right side. Since this term is expressed using the differential of a uniformizing paramter, its residue is the residue of $rt^{r-1}fg$, which is the value of $rt^r fg = r(x - x_0)f$.

□

9.6 Examples

Example 9.27. Compute $\int \sqrt{4-x^2}\,dx$

A solution method from first year calculus might be to note that this integrand forms one leg of a right triangle:



$$x = 2\sin\theta \qquad \sqrt{4-x^2} = 2\cos\theta \qquad dx = 2\cos\theta\,d\theta$$

$$\begin{aligned}
\int \sqrt{4-x^2}\,dx &= \int 4\cos^2\theta\,d\theta \\
&= \int (2 + 2\cos 2\theta)\;d\theta \\
&= 2\theta + \sin 2\theta \\
&= 2\theta + 2\sin\theta\cos\theta \\
&= 2\arcsin\frac{x}{2} + \frac{x\sqrt{4-x^2}}{2}
\end{aligned}$$

Now let's attack this integral using the methods of this chapter. First, transform the problem into an algebraic curve:

$$\int y\,dx \qquad y^2 = 4 - x^2$$

Since $\lim_{x\to\infty} y = \infty$, the integrand has poles at infinity. We want infinity to be an ordinary point of the curve (no ramification; no singularities) with no poles in the integrand. The simplest transformation is to exchange zero with infinity, and in this case zero is an ordinary point with places $(0, 2)$ and $(0, -2)$, neither of which is a pole of the integrand. So we'll invert x and y into u and v:

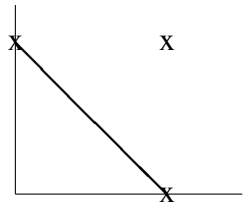
$$x = \frac{1}{u} \qquad y = \frac{1}{v}$$

$$\left(\frac{1}{v}\right)^2 = 4 - \left(\frac{1}{u}\right)^2 \Longrightarrow 4u^2v^2 - v^2 - u^2 = 0$$

$$\int \frac{1}{v}\, d\left(\frac{1}{u}\right) \Longrightarrow -\int \frac{1}{vu^2}\, du$$

The only poles in this integrand occur when either $u = 0$ or $v = 0$. Substituting these values into $4u^2v^2 - v^2 - u^2 = 0$, we see that these condiutions only occur at $(u, v) = (0, 0)$, so let's analyze our curve at that point, starting with the Newton polygon:

$$4u^2v^2 - v^2 - u^2 = 0$$



The Newton polygon has a single line segment of span 2 and slope -1, so we have two cycles, each with ramification index one: a singularity. Since there is no ramification, u is a uniformizing parameter and we expect to expand v as follows:

$$v = c_1u + c_2u^2 + c_3u^3 + \cdots$$
$$v^2 = c_1^2u^2 + 2c_1c_2u^3 + (2c_1c_3 + c_2^2)u^4 + \cdots$$

Substituting these expansions into $4u^2v^2 - v^2 - u^2 = 0$, we obtain:

$$4c_1^2u^4 + 8c_1c_2u^5 + (8c_1c_3 + 4c_2^2)u^6 + \cdots$$
$$-c_1^2u^2 - 2c_1c_2u^3 - (2c_1c_3 + c_2^2)u^4 + \cdots - u^2 = 0$$

Equating terms in u^2, we see that $c_1 = \pm i$. Each of these two values corresponds to one branch of the singularity. There is only a single term in u^3, which forces c_2 to be zero, and equating terms in u^4 produces $c_3 = 2c_1$, so

$$v = \pm(iu + 2iu^3 + \cdots) \qquad @(0, 0)$$

Inverting v and substituting into our 1-form, we obtain

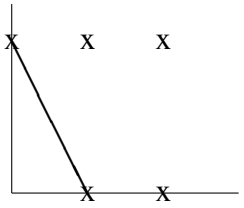
$$\frac{1}{v} = \pm(-i\frac{1}{u} + 2iu + \cdots) \qquad @(0, 0)$$

$$\frac{1}{vu^2}\, du = \pm\left[-i\frac{1}{u^3} + 2i\frac{1}{u} + \cdots\right] du \qquad @(0, 0)$$

The u^{-1} terms will integrate into logarithms, so let's ignore them for the moment and concentrate on the u^{-3} terms, which will integrate into u^{-2} terms, so we're looking for a function with second order poles at both places at the $(0, 0)$ singularity.

Starting with our standard basis for all rational functions, $\{1,\, v\}$, we seek to modify it into a basis for $\mathrm{P}^2(0, 0)_a\mathrm{P}^2(0, 0)_b$. Note first that v has poles at $u = \pm\frac{1}{2}$. Using $y = 1/u$, we analyze at $(\pm\frac{1}{2}, \infty)$ as follows:

$$y^2\left[(u-\tfrac{1}{2})^2+(u-\tfrac{1}{2})+\tfrac{1}{4}\right]-4(u-\tfrac{1}{2})^2-4(u-\tfrac{1}{2})$$



Our line segment has span 1 and slope -2, indicating a single place with ramification 2, and y as a uniformizing parameter. Setting

$$(u-\frac{1}{2})=c_1y+c_2y^2+\cdots$$
$$(u-\frac{1}{2})^2=c_1^2y^2+\cdots$$

Substituting, we find that $c_1=0$ and $c_2=\frac{1}{16}$, so

$$(u-\frac{1}{2})=\frac{1}{16}y^2+\cdots \qquad v=y^{-1} \qquad @(\frac{1}{2},\infty)$$

$$(u+\frac{1}{2})=\frac{1}{16}y^2+\cdots \qquad v=y^{-1} \qquad @(-\frac{1}{2},\infty)$$

In short, v has first order poles at $(\pm\frac{1}{2},\infty)$ and $(u\pm\frac{1}{2})$ has second order zeros, so we can adjust our basis accordingly and obtain $\{1,\ (4u^2-1)v\}$ for a basis with no finite poles. We can also use a theorem of Trager to shortcut this calculation.

Returning to our analysis at $(0,0)$, we see that 1 has zero order (obviously) and $(4u^2-1)v$ has a first order zero at both sheets there, since $4u^2-1=-1$ is finite and v has first order zeros. We also know that u is a uniformizing parameter, so it's easy to modify our basis and obtain

$$\left\{\frac{1}{u^2},\ \frac{4u^2-1}{u^3}v\right\} \text{ is a } \mathbf{C}[x]-\text{basis for } \mathrm{P}^2(0,0)_{\mathrm{a}}\mathrm{P}^2(0,0)_{\mathrm{b}}$$

Is this basis normal at infinity? Well, the representation order of $\frac{1}{u^2}$ is 2 and its u^-2 coefficients at $(\infty,\pm\frac{1}{2})$ are both 1, while the representation order of $\frac{4u^2-1}{u^3}v$ is 1, and its u^-1 coefficients are 2 and -2. Since

$$\det C=\begin{vmatrix}1 & 2\\ 1 & -2\end{vmatrix}=-4$$

is non-zero, the basis is normal at infinity.

The Riemann-Roch theorem says that the dimension of $\mathfrak{l}(D)$ is 5, $\frac{1}{u^2}$ can be multiplied by any polynomial up to second degree without introducing poles at infinity, and $\frac{4u^2-1}{u^3}v$ can be multiplied by any polynomial up to first degree, so

$$\left\{\frac{1}{u^2},\ \frac{1}{u},\ 1,\ \frac{4u^2-1}{u^3}v,\ \frac{4u^2-1}{u^2}v\right\}$$

is a $\mathcal{C}$-module basis for $\mathfrak{l}(D)$.

Any linear combination of these functions is a multiple of the divisor, but not all of them produce the correct residues. Looking at the residues, we see that only $\frac{4u^2-1}{u^3}v = \frac{1}{uv}$ has residues of $\pm i$ on the two sheets at the $(0,0)$ singularity. Dividing by 2 to correct for the 2 that will be introduced by the integration, we conclude that $\frac{1}{2uv} = \frac{xy}{2} = \frac{x\sqrt{4-x^2}}{2}$ is the desired function.

Next, we have to deal with the logarithms. Going back to the series expansions of our 1-form, we see that we have residues of $\pm 2i$ on our two sheets at $(0,0)$. The objective now is a bit different; we want a function with exactly the divisor $Z(0,0)_a P(0,0)_b$. Starting with an integral basis:

$$\{1, (4u^2-1)v\}$$

we want to modify these functions to make them multiples of $Z(0,0)_a P(0,0)_b$. The pole isn't a problem for an integral basis, and looking at the series expansion for v at $(0,0)$ we see that it (and therefore $(4u^2-1)v$) has a simple zero there, but 1 needs to be replaced with u:

$$\{u, (4u^2-1)v\}$$

Now we construct a matrix with the coefficients in the series expansions:

$$\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \begin{matrix} \leftarrow (0,0)_a \\ \leftarrow (0,0)_b \end{matrix}$$

$$\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = 0$$

The solution shows us how to modify the basis:

$$\{u, \frac{iu + (4u^2-1)v}{u}\} = \{u, i + \frac{(4u^2-1)v}{u}\}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \leftarrow (0,0)_a \\ \leftarrow (0,0)_b \end{matrix}$$

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

$$\{u, i\frac{1}{u} + \frac{(4u^2-1)v}{u^2}\}$$

$$\begin{vmatrix} 1 & -2i \\ 0 & 2i \end{vmatrix} = 2i$$

At the last step, the determinant is non-zero, which shows that we now have a basis for multiples of the divisor except at infinity. Is it normal at infinity? u's expansion at both places at infinity is $\left(\frac{1}{u}\right)^{-1}$, so its representation order

is -1, and the second element's expansion at infinity starts $\pm 2 + \cdots$, so its representation order is 0 and:

$$\det C = \begin{vmatrix} 1 & 2 \\ 1 & -2 \end{vmatrix} = -4$$

So the basis is normal at infinity. If an exact multiple of the divisor exists, it is one of the basis elements. It's not u, since u has a pole at infinity, but the second element is exact:

$$i\frac{1}{u} + \frac{(4u^2 - 1)v}{u^2} = i\frac{1}{u} - \frac{1}{v} = ix - y$$

The desired residues are $\pm 2i$, so the function we want is

$$2i \ln(ix - y) = 2i \ln(\frac{y}{2} - i\frac{x}{2}) + 2i \ln(-2)$$

$$= 2i \ln\left(\sqrt{1 - \left(\frac{x}{2}\right)^2} - i\frac{x}{2}\right) = 2i(-i \arcsin\frac{x}{2}) = 2 \arcsin\frac{x}{2}$$

(the constant disappears into the constant of integration) and the final answer is:

$$\int \sqrt{4 - x^2}\, dx = 2 \arcsin\frac{x}{2} + \frac{x\sqrt{4 - x^2}}{2}$$

□

9.7 arcsin

Example 9.28. Compute $\int \frac{1}{\sqrt{1-x^2}}\,dx$

The obvious attempt is to use the algebraic extension $y^2 = 1 - x^2$ and integrate $\frac{1}{y}\,dx$.

But we first need to determine if this differential has any poles at infinity, by inverting the field and looking for poles at zero. Setting $u = \frac{1}{x}$, we convert our minimal polynomial into $u^2y^2 = u^2 - 1$ (after multiplying through by u^2), and using $v = uy$ we obtain our inverse field $\mathbf{C}(u, v); v^2 = u^2 - 1$.

Since $x = \frac{1}{u}$ and $y = \frac{v}{u}$, we convert our differential as follows:

$$\frac{1}{y}\,dx = \frac{u}{v}(-\frac{1}{u^2}\,du) = -\frac{1}{uv}\,du$$

Now, $\{1, v\}$ is an integral basis for the inverse field, so we multiply through by $\frac{v}{v}$ to obtain:

$$= -\frac{v}{uv^2}du = -\frac{1}{u(u^2-1)}v\,du$$

which is now in normal form and clearly has a pole at $u = 0$, or $x = \infty$. Note that

$$\frac{1}{y} = \frac{u}{v} = \frac{uv}{v^2} = \frac{u}{u^2-1}v$$

has no pole at $u = 0$, a clear example of a differential having a pole at a place where its constituent function has none.

In any event, we clearly can not use the original field to conduct the integration, since it would require constructing a function with a pole at infinity, and our algorithm can't handle this. So we need to transform into a field where the differential has no pole at infinity.

Actually, we've already done this! Note that the integrand had no pole at zero in the original field:

$$\frac{1}{y}\,dx = \frac{y}{y^2}\,dx = \frac{1}{1-x^2}y\,dx$$

Since the inverse field swapped zero with infinity, it follows that there is no pole at infinity in the inverse field, so we can proceed to integrate $-\frac{1}{u(u^2-1)}v\,du$ in $\mathbf{C}(u, v); v^2 = u^2 - 1$.

Simple inspection of the integrand (already in normal form) shows that its poles are at $(0, i)$, $(0, -i)$, $(1, 0)$, and $(-1, 0)$. Remember that we're now working on the Riemann surface of an algebraic extension, so we need to specify *both* u and v to specify a place.

The next step is to compute the residues at each of these places, using Theorem 9.26:

$$
\begin{array}{lll}
(0,i) & -\dfrac{1}{(u^2-1)}v \ @ \ (0,i) & = i \\
(0,-i) & -\dfrac{1}{(u^2-1)}v \ @ \ (0,-i) & = -i \\
(1,0) & -2\dfrac{1}{u(u+1)}v \ @ \ (1,0) & = 0 \\
(-1,0) & -2\dfrac{1}{u(u-1)}v \ @ \ (-1,0) & = 0
\end{array}
$$

The poles with zero residues can be ignored. We're interested in the other two, which exist in $\mathbf{Q}[i]$, which can be regarded as a vector field over $\mathbf{Q}$ with basis $\{1, i\}$, and we want to construct a function whose poles and zeros match the i-component of the residues (the 1-component is uniformly zero).

We start by constructing an $\mathcal{I}$-module generator set for the divisor with a simple zero at $(0, i)$ and a simple pole at $(0, -i)$. Theorem 9.20 shows that:

$$f = \frac{v^2+1}{u(v+i)} = \frac{v-i}{u}$$

has a simple pole at $(0, -i)$. At $(0, i)$, L'Hôpital's rule gives:

$$\lim_{(u,v)\to(0,i)} \frac{v-i}{u} = \frac{(v-i)'}{u'}\frac{dv}{du} = \frac{dv}{du} = \frac{u}{v} = 0$$

where the last transformation was accomplished by differentiating the mimimal polynomial. So f has a zero at $(0, i)$, and I'll note that we've just stumbled into the solution. Theorem 9.20 already assures us that f has only a single finite simple pole, and we can see that its only zeros occur when $v - i = 0$, which, according to the minimum polynomial, can only occur at $u = 0$, thus $(0, i)$ is its only finite zero, and it is simple, as we can verify by showing that the corresponding pole in its inverse is simple:

$$\frac{1}{f} = \frac{u}{v-i} = \frac{u(v+i)}{v^2+1} = \frac{u(v+i)}{u^2} = \frac{1}{u}v + \frac{i}{u}$$

So we've found the function we're looking for by accident. Let's save the general case for the next example, and convert back to our original field:

$$\frac{v-i}{u} = x(\frac{y}{x} - i) = y - ix$$

Remembering that our residues came multiplied by a factor of i, we conclude that our solution is $i \ \ln(y - ix)$, or:

$$\begin{aligned}
\int \frac{1}{\sqrt{1-x^2}}\,dx &= i\,\ln(\sqrt{1-x^2}-ix) \\
&= -i\,\ln(\frac{1}{\sqrt{1-x^2}-ix}) \\
&= -i\,\ln(\frac{\sqrt{1-x^2}+ix}{1-x^2+x^2}) \\
&= -i\,\ln(\sqrt{1-x^2}+ix) \\
&= \arcsin x
\end{aligned}$$

where I used the negative of a logarithm being the logarithm of the inverse, and the last transformation came from section **??**.

□

9.8 Geddes's example

Example 9.29. Compute $\int \frac{1}{x\sqrt{x^4+1}}\,dx$

We'll use $\mathbf{C}(x,y); y^2 = x^4+1$ and integrate $\frac{1}{xy} = \frac{y}{x^5+x}$. Inverting this field ($z = \frac{1}{x}$) shows that this integrand has no poles at infinity, so we can proceed directly:

$$\frac{y}{x^5+x} = \frac{y}{x(x+\omega)(x-\omega)(x+i\omega)(x-i\omega)} \qquad \omega = \sqrt{i} = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$$

$$\begin{array}{lll}
(0,1) & \frac{y}{x^4+1} \text{ @ } (0,1) & = 1 \\
(0,-1) & \frac{y}{x^4+1} \text{ @ } (0,-1) & = -1 \\
(\omega,0) & 2\frac{y}{x(x^2+i)(x+\omega)} \text{ @ } (\omega,1) & = 0 \\
(-\omega,0), (i\omega,0), (-i\omega,0) & \cdots & = 0
\end{array}$$

We now use theorem 9.20 to construct a function with a simple pole at $(0,-1)$:

$$\frac{f(0,y)}{x(y+1)} = \frac{y^2-1}{x(y+1)} = \frac{y-1}{x}$$

This function has a zero at $(0,1)$, but, unfortunately, it is third order, as can be seen from either L'Hôpital's rule:

$$y^2 = x^4+1$$
$$2y\,dy = 4x^3\,dx$$
$$\frac{dy}{dx} = 2\frac{x^3}{y}$$

$$\frac{y-1}{x}@(0,1) = \lim \frac{dy}{dx} = 2\frac{x^3}{y} = 0$$
$$\frac{y-1}{x^2}@(0,1) = \lim \frac{1}{2x}\frac{dy}{dx} = \frac{x^2}{y} = 0$$
$$\frac{y-1}{x^3}@(0,1) = \lim \frac{1}{3x^2}\frac{dy}{dx} = \frac{2}{3}\frac{x}{y} = 0$$
$$\frac{y-1}{x^4}@(0,1) = \lim \frac{1}{4x^3}\frac{dy}{dx} = \frac{1}{2}\frac{1}{y} = \frac{1}{2}$$

... or from a series expansion of y at (0,1):

$$y^2 = x^4 + 1$$
$$(y-1)^2 = x^4 - 2(y-1)$$
$$(y-1) = \frac{1}{2}x^4 - \frac{1}{2}(y-1)^2$$
$$(y-1) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots$$
$$(y-1)^2 = c_0^2 + (2c_0c_1)x + (2c_0c_2 + c_1^2)x^2 + (2c_0c_3 + 2c_1c_2)x^3 + \cdots$$

$$c_0, c_1, c_2, c_3 = 0$$
$$c_4 = \frac{1}{2}$$
$$c_5, c_6, c_7 = 0$$
$$c_8 = -\frac{1}{8}$$

$$(y-1) = \frac{1}{2}x^4 - \frac{1}{8}x^8 + \cdots$$

$$\frac{(y-1)}{x} = \frac{1}{2}x^3 - \frac{1}{8}x^7 + \cdots$$

... or from the norm:

$$N(\frac{y-1}{x}) = \frac{y-1}{x} \cdot \frac{-y-1}{x} = -\frac{y^2-1}{x^2} = -\frac{x^4}{x^2} = -x^2$$

Since we know that the function has a simple pole at $(0,-1)$, so it must have a third order zero at $(0,1)$ to form a norm with a second order zero.

We can eliminate the inconvenient zero by adding a constant to the function, say 1: $\frac{x+y-1}{x}$. We can now use theorem 9.19 to create a simple zero at (0,1) by multiplying by $x+y-1$:

$$\frac{x+y-1}{x}(x+y-1) = \frac{2x-2}{x}y + \frac{x^4+x^2-2x+2}{x}$$

This function has a simple pole at (0,-1) and a simple zero at (0,1), but does it have other poles and zeros? If so, can it be modified to eliminate them? To find out, we form the generators of an $\mathcal{I}$-module:

$$\{\frac{2x-2}{x}y + \frac{x^4+x^2-2x+2}{x}, x\}$$

Noting that $\frac{x^4+x^2}{x} = x(x^2+1)$ and $x^2+1 \in \mathcal{I}$, we can simplify this:

$$\{\frac{2x-2}{x}y - \frac{2x-2}{x}, x\}$$

Using the integral basis $\{1, y\}$, we convert this to a $\mathbf{C}[x]$-module:

$$\{\frac{2x-2}{x}y - \frac{2x-2}{x}, \frac{2x-2}{x}(x^4+1) - \frac{2x-2}{x}y, x, xy\}$$

and since $(2x-2)\frac{x^4}{x} = x(2x^3-2x^2)$ and $2x^3-2x^2 \in \mathbf{C}[x]$, we simplify:

$$\{\frac{2x-2}{x}y - \frac{2x-2}{x}, \frac{2x-2}{x} - \frac{2x-2}{x}y, x, xy\}$$

and write it in matrix form:

$$\frac{1}{x}\begin{pmatrix} -(2x-2) & 2x-2 \\ 2x-2 & -(2x-2) \\ x^2 & 0 \\ 0 & x^2 \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix}$$

Elementary row operations[2] $R_{4,3,x^2}R_{1,3,(2x-2)}R_{3,4,-1}R_{3,1,\frac{1}{2}(x+1)}R_{2,1,1}$ yield:

$$\frac{1}{x}\begin{pmatrix} 1 & -1 \\ x^2 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ y \end{pmatrix} = \begin{pmatrix} \frac{1-y}{x} \\ x \end{pmatrix}$$

so $\{\frac{1-y}{x}, x\}$ forms a generator set for the $\mathbf{C}[x]$-module of the finite multiples of $Z(0,1)P(0,-1)$. We convert to a basis normal at infinity: $\{1, v\} = \{1, \frac{y}{x^2}\}$:

$$\begin{pmatrix} \frac{1-y}{x} \\ x \end{pmatrix} = \begin{pmatrix} \frac{1}{x} & -x \\ x & 0 \end{pmatrix}\begin{pmatrix} 1 \\ \frac{y}{x^2} \end{pmatrix} = \begin{pmatrix} x & \\ & x \end{pmatrix}\begin{pmatrix} \frac{1}{x^2} & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ \frac{y}{x^2} \end{pmatrix}$$

$$\det_{@\infty}\begin{pmatrix} \frac{1}{x^2} & -1 \\ 1 & 0 \end{pmatrix} = 1$$

so $\{\frac{1-y}{x}, x\}$ is normal at infinity. x clearly has a pole at infinity, so it can't be the function we're looking for, but what about $\frac{1-y}{x}$? Switching back to $\{u, v\}$ coordinates, we obtain $\frac{1-y}{x} = \frac{u^2-v}{u}$, which has $\{u, v\}$ poles at both $\{0, 1\}$ and $\{0, -1\}$, which translate into poles at $x = \infty$. Therefore, no rational function on this algebraic curve has a simple pole at (0, -1), a simple zero at (0,1), and no other poles or zeros.

[2]Read right to left; $R_{i,j,\lambda}$ adds λ times row j to row i; $R_{i,\alpha}$ multiplies row i by α (a unit)

So, let's try a double pole at (0,-1) and a double zero at (0,1). We can just square our previous generators: $\{\frac{1-y}{x}, x\}$ to obtain: $\{\frac{(1-y)^2}{x^2}, x^2\} = \{\frac{1-2y+x^4+1}{x^2}, x^2\}$ which simplifies to $\{\frac{1-y}{x^2}, x^2\}$. We again check for normalcy at infinity:

$$\begin{pmatrix} \frac{1}{x^2} & -1 \\ x^2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{y}{x^2} \end{pmatrix} = \begin{pmatrix} 1 & \\ & x^2 \end{pmatrix} \begin{pmatrix} \frac{1}{x^2} & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \frac{y}{x^2} \end{pmatrix}$$

$$\det_{@\infty} \begin{pmatrix} \frac{1}{x^2} & -1 \\ 1 & 0 \end{pmatrix} = 1$$

So, $\{\frac{1-y}{x^2}, x^2\}$ is a $\mathbf{C}[x]$-module, normal at infinity, containing the finite multiples of $Z^2(0,1)P^2(0,-1)$. x^2 has a pole at infinity, but does $\frac{1-y}{x^2}$? Switching to $x = \frac{1}{z}; y = \frac{u}{z^2}; u^2 = z^4 + 1$ and Writing it as $z^2(1 - \frac{u}{z^2}) = z^2 - u$ shows that it has no zero at $(z, u) = (0, \pm 1)$, and thus no pole at $x = \infty$. It is, therefore, the function we are looking for:

$$\int \frac{1}{x\sqrt{x^4+1}}\,dx = \frac{1}{2}\ln\frac{1-\sqrt{x^4+1}}{x^2}$$

I'll now point out to you what's been pointed out to me, and that is a traditional solution technique for this integral:

$$\int \frac{1}{x\sqrt{x^4+1}}\,dx$$

$$x^2 = u \qquad 2x\,dx = du$$

$$\int \frac{1}{2u\sqrt{u^2+1}}\,du$$

$$u = \tan z \qquad du = \sec^2 z dz$$

$$\frac{1}{2}\int \csc z\,dz = \frac{1}{2}\ln\tan\frac{z}{2}$$

$$= \frac{1}{2}\ln\frac{\sec z - 1}{\tan z}$$

$$= \frac{1}{2}\ln\frac{\sqrt{u^2+1}-1}{u}$$

$$= \frac{1}{2}\ln\frac{\sqrt{x^4+1}-1}{x^2}$$

$$\frac{1-\cos z}{\sin z} = \frac{1-\cos^2\frac{z}{2}+\sin^2\frac{z}{2}}{2\sin\frac{z}{2}\cos\frac{z}{2}} = \frac{2\sin^2\frac{z}{2}}{2\sin\frac{z}{2}\cos\frac{z}{2}} = \frac{\sin\frac{z}{2}}{\cos\frac{z}{2}}$$

We can also proceed like this:

$$\int \frac{1}{x\sqrt{x^4+1}}\,dx$$

$$x^4 = u \qquad 4x^3\,dx = du$$

$$\int \frac{1}{4u\sqrt{u+1}}\,du$$

$$v = u+1 \qquad dv = du$$

$$\int \frac{1}{4(v-1)\sqrt{v}}\,dv$$

$$z^2 = v \qquad 2z\,dz = dv$$

$$\int \frac{1}{4(z^2-1)z}\,2z\,dz$$

$$\int \frac{1}{2(z^2-1)}\,dz$$

$$\frac{1}{z^2-1} = \frac{1}{2}\frac{1}{z-1} - \frac{1}{2}\frac{1}{z+1}$$

$$\frac{1}{4}\int \frac{1}{z-1} - \frac{1}{z+1}\,dz$$

$$\frac{1}{4}\ln(z-1) - \ln(z+1)$$

$$\frac{1}{4}\ln\frac{z-1}{z+1}$$

$$\frac{1}{4}\ln\frac{\sqrt{v}-1}{\sqrt{v}+1}$$

$$\frac{1}{4}\ln\frac{\sqrt{u+1}-1}{\sqrt{u+1}+1}$$

$$\frac{1}{4}\ln\frac{\sqrt{x^4+1}-1}{\sqrt{x^4+1}+1}$$

so from the previous page and this one, we conclude

$$\int \frac{1}{x\sqrt{x^4+1}}\,dx = \frac{1}{2}\ln\frac{\sqrt{x^4+1}-1}{x^2} = \frac{1}{4}\ln\frac{\sqrt{x^4+1}-1}{\sqrt{x^4+1}+1}$$

Is this last equality true? Well, $\ln f^2 = 2\ln f$, so $\frac{1}{4}\ln f^2 = \frac{1}{2}\ln f$, and...

$$\left(\frac{\sqrt{x^4+1}-1}{x^2}\right)^2 = \frac{(\sqrt{x^4+1}-1)^2}{x^4}$$

$$\frac{\sqrt{x^4+1}-1}{\sqrt{x^4+1}+1}\cdot\frac{\sqrt{x^4+1}-1}{\sqrt{x^4+1}-1} = \frac{(\sqrt{x^4+1}-1)^2}{x^4+1-1}$$

□

9.9 Chebyshev's Integral

Example 9.30. Compute:

$$\int \frac{2x^6+4x^5+7x^4-3x^3-x^2-8x-8}{(2x^2-1)^2\sqrt{x^4+4x^3+2x^2+1}}\,\mathrm{d}x$$

The polynomial under the square root is square-free:

```
i8 : root = poly "x4+4x3+2x2+1"

      4     3     2
o8 = x  + 4x  + 2x  + 1

o8 : Rx

i9 : diff(x,root)

       3      2
o9 = 4x  + 12x  + 4x

o9 : Rx

i10 : gcd(o9, root)

o10 = 1

o10 : Rx
```

... so $y^2 = x^4+4x^3+2x^2+1$; $\{1, y\}$ is an integral basis; and our normal form for this integral is:

$$\int \frac{(2x^6+4x^5+7x^4-3x^3-x^2-8x-8)y}{(2x^2-1)^2(x^4+4x^3+2x^2+1)}\,\mathrm{d}x$$

Applying now Bronstein's Hermite reduction:

```
i28 : ytic=diff(x,root)//2/root*y

          3     2
        2x  + 6x  + 2x
o28 = ------------------*y
       4     3     2
      x  + 4x  + 2x  + 1

o28 : Fxy

i29 : U = root

       4     3     2
o29 = x  + 4x  + 2x  + 1

o29 : Rx

i30 : V=poly "2x2-1"

        2
```

```
o30 = 2x  - 1

o30 : Rx

i32 : S2=U*(V*ytic - diff(x,V)*y)

           4     3     2
o32 = (- 4x  - 6x  - 6x  - 6x)y

o32 : Fxy

i2 : num=poly "2x6+4x5+7x4-3x3-x2-8x-8"

       6     5     4     3    2
o2 = 2x  + 4x  + 7x  - 3x  - x  - 8x - 8

o2 : Rx

i6 : T2=num

       6     5     4     3    2
o6 = 2x  + 4x  + 7x  - 3x  - x  - 8x - 8

o6 : Rx

i79 : factor(U)

               3     2
o79 = (x + 1)(x  + 3x  - x + 1)

o79 : Expression of class Product

i92 : Q=S2

          4     3     2
o92 = - 4x  - 6x  - 6x  - 6x

o92 : Rx

i93 : A=(gcdCoefficients (V,Q))#1

        36  3   38  2   48
o93 = - --*x  - --*x  - --*x - 1
        49      49      49

o93 : Rx

i94 : R=(gcdCoefficients (V,Q))#2

        18      8
o94 = - --*x + --
        49     49

o94 : Rx

i96 : Q2=(T2*R)//V

        18  5   4  4   8  3   41  2   31     177
o96 = - --*x  - -*x  - -*x  + --*x  - --*x + ---
        49      7      7      49      49      98
```

```
o96 : Rx

i97 : B2=(T2*R)%V

           1
o97 = x + -
           2

o97 : Rx

i98 : h = A*num/(V*U) - (diff(x,V)*Q2+diff(x,B2))/V + Q2*ytic

                2
              6x  + 5x + 7
o98 = -------------------------
        6     5     4     3
      2x  + 8x  + 3x  - 4x  - 1

o98 : frac(Rx)

i100 : factor(denominator h)

                    2       3     2
o100 = (x + 1)(2x  - 1)(x  + 3x  - x + 1)

o100 : Expression of class Product
```

$$\int \frac{(2x^6+4x^5+7x^4-3x^3-x^2-8x-8)y}{(2x^2-1)^2(x^4+4x^3+2x^2+1)}\,\mathrm{d}x = \frac{(x+\frac{1}{2})y}{2x^2-1} + \int \frac{(6x^2+5x+7)y}{(2x^2-1)(x^4+4x^3+2x^2+1)}\,\mathrm{d}x$$

Non-zero residues

x	y	residue
$\frac{\sqrt{2}}{2}$	$\frac{1}{2}+\sqrt{2}$	$\frac{5}{2}$
$\frac{\sqrt{2}}{2}$	$-\frac{1}{2}-\sqrt{2}$	$-\frac{5}{2}$
$-\frac{\sqrt{2}}{2}$	$\frac{1}{2}-\sqrt{2}$	$\frac{5}{2}$
$-\frac{\sqrt{2}}{2}$	$-\frac{1}{2}+\sqrt{2}$	$-\frac{5}{2}$

$$A(x) = 1023x^8+4104x^7+5048x^6+2182x^5+805x^4+624x^3+10x^2+28x$$

$$B(x) = 1025x^{10}+6138x^9+12307x^8+10188x^7+4503x^6+3134x^5+1598x^4+140x^3+176x^2+2$$

$$C(x) = 32x^{10}-80x^8+80x^6-40x^4+10x^2-1$$

$$\int \frac{(2x^6+4x^5+7x^4-3x^3-x^2-8x-8)y}{(2x^2-1)^2(x^4+4x^3+2x^2+1)}\,\mathrm{d}x = \frac{(x+\frac{1}{2})y}{2x^2-1} + \frac{1}{2}\ln\frac{A(x)y-B(x)}{C(x)}$$

□

9.10 Señor Gonzalez, otra vez

The Rothstein-Trager resultant allows us to compute all the residues at once. Trager, in his Ph.D. thesis, then showed how to construct a function that is zero at all poles with a given residue, and non-zero at all other poles, as well as at all places conjugate to a pole.

Chapter 10

Good Reduction

At this point, there is only one major missing piece in our integration theory for simple radicals — how do we limit the multiples of a divisor to a testable set? We've seen how to repeatedly raise a divisor to higher and higher powers, but how do we know when to stop? At what point can we declare that a divisor has no multiple that is principle?

We'll attack this problem the same way we attacked polynomial factorization in Chapter ?, by mapping into a finite field, solving a corresponding problem there, then somehow lifting the result back to the original field. The details differ, but the basic idea is the same. Of course, divisors, like polynomials, behave somewhat different in finite fields, so our first task is to study some of their unique properties in this domain.

10.1 Simple Algebraic Extensions over Finite Fields

Let's start with a simple but crucial observation:

Theorem 10.1. *In an algebraic extension over a finite field, the evaluation field is also finite.*

Proof

Consider a finite field of constants $\mathcal{F}$, over which we'll extend first into a rational function field $\mathcal{F}(x)$ and then add an algebraic extension $\mathcal{F}(x, y)$, where y satisfies some minimial polynomial $f(x, y) = 0$. Start with the constant field, which gives us a finite number of values for x. Plugging each of these values into the minimal polynomial gives a finite set of polynomials $f(y_i) = 0$. By Theorem ?, we can extend $\mathcal{F}$ into a finite extension field $\mathcal{F}[\gamma]$ where all the roots of the polynomial exist. Since there a only a finite number of polynomials, we need at worst a finite set of extensions $\mathcal{F}[\gamma_1, ..., \gamma_k]$ to construct a field in which all the roots of all the polynomials exist. Using the Theorem of the Primitive Element, we can collapse all of these into a single finite extension field $\mathcal{F}[\phi]$. Since all values of x exist in $\mathcal{F}$, and all values of y exist in $\mathcal{F}[\phi]$, an evaluation homomorphism carries any rational function in x and y into $\mathcal{F}[\phi]$.

□

This theorem leads directly to the single more important difference (to us) between divisors in an infinite field versus those in a finite field. *In a finite field, some multiple of every divisor is principle.* The reason is that the multiplicative group of the evaluation field has finite order. The simplest way to demonstrate this is to construct theorems analogous to Theorems ? and ?:

Theorem 10.2. *In an algebraic extension of a finite field with characteristic greater than 2, a function can always be constructed with an m^{th}-order zero at a specified place (α, β) and zero order at all other finite places, where m is the multiplicative order of the evaluation field.*

Proof

The desired function is

$$(x - \alpha)^m + (y - \beta)^m$$

.

Clearly, this function is zero at (α, β) and of m^{th} order there (PROOF THIS). At all other places one of the two terms will be non-zero, and both exist in the evaluation field. By Theorem ?, any non-zero number raised to the multiplicative order of its field is one. Thus the value of this function will be either $1 + 0$, $0 + 1$, or $1 + 1 = 2$, all finite and non-zero, and thus of zero order.

□

Theorem 10.3. *In an algebraic extension of a finite field with characteristic greater than 2, a function can always be constructed with an m^{th}-order pole at a specified place (α, β) and zero order at all other finite places, where m is the multiplicative order of the evaluation field.*

Proof

The desired function is

$$\frac{f(\alpha, y)^m}{(x - \alpha)^m (y - \beta)^m} + 1$$

where the division by $(y - \beta)^m$ is exact. Clearly, this function has a pole at (α, β) and of m^{th} order there (PROOF THIS). CONSIDER OTHER PLACES OVER α. At all other places the denominator term will be non-zero, and thus one, and the numerator will be either zero or one (by Theorem ?) Thus the value of this function at these places will be either $0 + 1$ or $1 + 1 = 2$, both finite and non-zero, and thus of zero order.

□

Example 10.4. Show that some multiple of $\mathrm{Z}(1,1)$ is principle in $\mathbf{Z}_5(x,y); y^2 = x$.

Let's first construct a multiplication table for $\mathbf{Z}_5$:

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Now, let's list out the places on the Riemann surface for $\mathbf{Z}_5(x,y); y^2 = x$.

x	(x,y)		
0	(0,0)		
1	(1,1)	(1,4)	
2	$(2,\gamma)$	$(2,-\gamma)$;	$\gamma^2 - 2 = 0$
3	$(3,\theta)$	$(3,-\theta)$;	$\theta^2 - 3 = 0$
4	(4,2)	(4,3)	

It looks like we need $\mathbf{Z}_5[\gamma,\theta]$ to express these places. It's simplest to collapse γ and θ into a single algebraic extension. We could use the Theorem of the Primitive Element to do this, but in this case just looking at the multiplication table and noting that $3 = 2^3 = \gamma^6$ shows that $\theta = \pm\gamma^3$. So, in fact, we only need $\mathbf{Z}_5[\gamma]$:

x	(x,y)		
0	(0,0)		
1	(1,1)	(1,4)	
2	$(2,\gamma)$	$(2,-\gamma)$;	$\gamma^2 - 2 = 0$
3	$(3,\gamma^3)$	$(3,-\gamma^3)$	
4	(4,2)	(4,3)	

Since $\mathbf{Z}_5[\gamma]$ has $5^2 = 25$ elements, its multiplicative group has order one less than this. We conclude that 24 is our "magic" multiple, and that $\mathrm{Z}^{24}(1,1)$ must be principle in this field. Its generator should be simply $(x-1)^{24} + (y-1)^{24}$. Clearly this function is zero for $(x,y) = (1,1)$. Let's verify that it's non-zero for some other places on the Riemann surface:

$$
\begin{aligned}
(0,0) &: (-1)^{24} + (-1)^{24} = 4^{24} + 4^{24} = 1 + 1 = 2 \\
(1,4) &: 3^{24} + 0^{24} = 1 + 0 = 1 \\
(2,\gamma) &: (\gamma - 1)^{24} + (2-1)^{24} = 1 + 1 = 2, \quad \text{since :}
\end{aligned}
$$

$$
\begin{aligned}
(\gamma-1)^2 &= (\gamma^2 - 2\gamma + 1) = 3 - 2\gamma \\
(\gamma-1)^4 &= (3 - 2\gamma)^2 = (9 - 12\gamma + 4\gamma^2) = 2 - 2\gamma \\
(\gamma-1)^8 &= (2 - 2\gamma)^2 = (4 - 8\gamma + 4\gamma^2) = 2 - 3\gamma \\
(\gamma-1)^{12} &= (2 - 2\gamma)(2 - 3\gamma) = (4 - 10\gamma + 6\gamma^2) = 1
\end{aligned}
$$

In the final series of calculations, I used $\gamma^2 = 2$ and reduced mod 5 repeatedly. I think the pattern should be clear, and leave further verification as an exercise.

□

10.2 Jacobian Varieties

An algebraic extension is a simple example of what algebraic geometers term a *variety*, which is the zero locus of a set of polynomials defined over some field. Thus, for example, the unit circle is a variety (defined over the real numbers), because it is the zero locus of $x^2 + y^2 = 1$. But the points $(1, 0)$ and $(-1, 0)$ are also a variety, because they are the zero locus of the *set* of polynomials $\{x^2 = 1; y = 0\}$.

An *abelian variety* is a variety accompanied by a commutative group structure on its elements, which typically includes picking an arbitrary zero point as the identity element. The circle is an abelian variety, if we identify its points with their angles from the x-axis and make $(1, 0)$ our identity element. Now any two points can be "added" or "subtracted" (by adding or subtracting their respective angles) to obtain a third point, and each point has an inverse associated with it (its mirror image across the x-axis). It should be obvious that the choice of a zero point was totally arbitrary. Likewise, the points $(1, 0), (-1, 0)$ also form an abelian variety; their group structure is isomorphic to $\mathbf{Z}_2$ and the choice of one of them as the identity is, again, arbitrary.

Is every variety abelian? No, but any complete, non-singular variety can be homomorphicly mapped into an associated abelian variety (typically of higher dimension), called its *Jacobian variety*. This fact, combined with the extensive body of literature on abelian varieties ([Mumford], [Birkenhake and Lange], [Lang], to mention a few), makes the Jacobian variety an important object of study (though David Mumford, in the preface to [Mumford], described it as a "crutch").

We will be needing only a tiny bit of this theory here, so my goal in this section is only to demonstrate how the Riemann-Roch Theorem allows us to set up an abelian group structures on an algebraic extension.

10.3 The Riemann-Roch Theorem

The Riemann-Roch Theorem is one of the most celebrated theorems in mathematics. Not only does it provide a crucial tool in understanding the structure of algebraic extensions, but it does so by tying together algebra, analysis, and geometry in one equation.

First, let's review that equation:

Theorem 10.5. (Riemann-Roch)

For any divisor $\mathfrak{b}$*,*

$$l(\mathfrak{b}) = -\deg \mathfrak{b} + 1 - g + l(-\mathfrak{b} - \mathfrak{c})$$

where $l(\mathfrak{b})$ *is the dimension of the vector space* $L(\mathfrak{b})$ *of multiples of* $\mathfrak{b}$*,* g *is the genus of the extension, and* $\mathfrak{c}$ *is any divisor of the canonical class of differentials.*

□

Interrelated by this theorem is the purely algebraic concept of the dimension of the vector space of multiples of a divisor, the geometric concept of the genus, and the analytic concept of a differential.

However, this sophistication comes with a price. Specifically, we need a topology to define the genus, and we need a limit to define the differential.

André Weil showed how the Riemann-Roch theorem can be stripped of the analysis and the geometry, and proved as purely a result in algebra. The genus, instead of a topological invariant, now appears as merely a least upper bound on a divisor's degree of specialization, and a differential becomes an object in a dual space that maps a function into the field of constants. The advantage of this formulation is that does not require any topological structure, and is therefore well suited to use with finite fields. It is this formulation I will now adopt.

First, at any place in the function field, there is a local valuation ring with a maximal prime ideal. We can normalize the valuation (it is discrete) and pick a element of unit valuation to use as a uniformizing variable. By multiplying as necessary by some power of this element, we can adjust any field element to be a unit of the valuation ring and thus associate an order $\mathrm{ord}_{\mathfrak{p}}$ with that element. The valuation ring's units are a finite extension of the constant subfield; they are the constant subfield if it is algebraically closed. By subtracting out the remainder mod $\mathfrak{p}$, we get an element of higher order, which we can again subtract out, and so on, building a power series in the uniformizing variable. Each element of the function field thus has a power series associated with it at each place $\mathfrak{p}$.

A collection of such power series, one at each place, with arbitrary coefficients except that there are only a finite number of coefficients with negative powers, is called a *vector*. Each individual power series is called a *component* of the vector. Clearly, every function has a vector associated with it; but the converse is not necessarily true. The mapping from functions to vectors is injective, though. Any two different functions will have a non-zero difference that must therefore have a finite value, of finite order, at some place $\mathfrak{p}$, and their vectors will differ at that point.

We also have a dual space of *covectors*. The coefficients of a covector component at a place are dual to the constant field at that place; if the constant field is algebraically closed, then the covector coefficients are in the constant field. Like vectors, covectors can only have a finite number of negative power coefficients.

We define a dot product between a vector v and a covector λ:

$$v \cdot \lambda = \sum_{\mathfrak{p}} \sum_{i+j=-1} v_{\mathfrak{p},i} \lambda_{\mathfrak{p},j}$$

where $v_{\mathfrak{p},i}$ is the coefficient of the i^{th} power in v's component at $\mathfrak{p}$, and likewise for $\lambda_{\mathfrak{p},j}$. Notice that the second summation requires at least one of i or j to be negative, so there will only be a finite number of places for the first sum at which the second sum contributes anything at all.

Weil also requires the *Theorem of Independence*, which states that, although an arbitrary (full) vector may not have a function associated with it, a function can always be found which matches a set of finite prefixes at a finite number of places. This can be demonstrated using Theorem 9.21, repeatedly applied a finite number of times. We also need to know that a function without a pole is constant.

With this setup, we can now prove a series of theorems that lead up the the Riemann-Roch Theorem.

Theorem 10.6.

$$l(\mathfrak{p}) \leq \deg \mathfrak{p} + 1$$

i.e, $l(\mathfrak{p})$, the dimension (over the constants) of $L(\mathfrak{p})$, the multiples of $-\mathfrak{p}$, is no more than the degree of the divisor $\mathfrak{p}$, plus one.

Proof

Since $\deg -\mathfrak{p} = -\deg \mathfrak{p}$, there are at least $\deg \mathfrak{p}$ poles (counting multiplicities) in $-\mathfrak{p}$, and at least $\deg \mathfrak{p}$ coefficients with negative powers in the vectors corresponding to the elements in $L(\mathfrak{p})$

. We can impose

□

Theorem 10.7. *If $\mathfrak{A}$ is divisible by $\mathfrak{B}$, i.e, if $\mathfrak{A}\mathfrak{B}^{-1} \subseteq \mathcal{I}$, then*

$$n(\mathfrak{A}) - l(\mathfrak{A}) \leq n(\mathfrak{B}) - l(\mathfrak{B})$$

$$n(\mathfrak{p}) \equiv \deg \mathfrak{p}$$

Proof

Consider $\mathfrak{C} = \mathfrak{A}\mathfrak{B}^{-1}$. Now $n(\mathfrak{C}) = n(\mathfrak{A}) - n(\mathfrak{B})$ and since $\deg -\mathfrak{C} = -\deg \mathfrak{C}$, and $\mathfrak{C}$ is integral (by supposition), there are exactly $n(\mathfrak{C})$ poles (counting multiplicities) in $-\mathfrak{C}$. , and at least $\deg \mathfrak{p}$ coefficients with negative powers in the vectors corresponding to the elements in $L(\mathfrak{p})$

. We can impose

□

Back to the Riemann-Roch Theorem...

It immediately follows (from $\mathfrak{b} = \mathbf{0}$) that $l(\mathfrak{c}) = g$, which can be taken as the definition of the genus.

We can now pick g independent differentials from $\mathfrak{c}$ and use them (along with an arbitrary origin) to map into the torus $\mathbf{C}/\Lambda^g$.

Now, Abel's Theorem and the Jacobi inversion theorem ([Griffiths and Harris], p. 235) shows that Pic^0, the group of divisors of degree zero modulo linear equivalence is isomorphic to $\mathbf{C}/\Lambda^g$.

Alternately, ([Lang], II, §1, Theorem 3), we can factor a mapping of a product into an abelian variety into mappings on each factor.

Lang also characterizes Abel's theorem as follows:

> Let $\omega_1, ..., \omega_g$ be a basis for the differential forms on the first kind of V. If $\mathfrak{a} = \sum n_i P_i$ is a [divisor] of degree 0 on V, and P is a fixed point of V, then the map into $\mathbf{C}/\Lambda^g$ given by:
>
> $$\mathfrak{a} \to \sum n_i (\int_P^{P_i} \omega_1, ..., \int_P^{P_i} \omega_g)$$
>
> is well defined modulo the periods... the kernel consists of those divisors that are linearly equivalent to 0 (i.e, principle); this is Abel's theorem.

10.4 Endomorphism Rings

Any commutative group G induces a (non-commutative) ring structure on its endmorphisms, defined as follows (remember that an endomorphism is a homomorphism from an object to itself):

Two endmorphisms $\phi(g) : G \to G$ and $\gamma(g) : G \to G$ are added using G's group operation on the images: $(\phi + \gamma)(g) = \phi(g) \cdot \gamma(g)$, where $\cdot$ denotes the group operation. The additive identify is the endmorphism that maps the entire group onto its identity element.

Two endmorphisms $\phi(g) : G \to G$ and $\gamma(g) : G \to G$ are multiplied using composition of mappings: $(\phi\gamma)(g) = \phi(\gamma(g))$. The multiplicative identity is the endmorphism that maps every element in the group onto itself.

Let us now verify that these operations define a ring, the *endomorphism ring* of G, which we shall denote $\mathrm{End}(G)$. The properties of the identity elements are fairly obvious, I

think. Almost as obvious is that the associative and commutative properties of the underlying group translate directly into additive associative and commutative properties in the endmorphism ring. The multiplicative properties follow from composition of mappings being associative, but not necessarily commutative. The distributive law follows from the easily verified identity $\phi(\gamma(g) \cdot \mu(g)) = \phi(\gamma(g)) \cdot \phi(\mu(g))$, using the fact that ϕ is an endomorphism, and thus a homomorphism, and therefore maps the group operator through.

The ring of integers $\mathbf{Z}$ can be mapped homomorphicaly[1] into any ring, and an endomorphism ring is no exception. We'll denote by $[m]$ the endmorphism mapped to by the integer m. $[0]$ is clearly the additive identity mapping all elements to the group identity. $[1]$ is, of course, the multiplicative identity mapping all elements to themselves. $[2]$ is $[1] + [1]$, the endmorphism that composes each element with itself (using the group operator): $[2] : g \to g \cdot g$. $[3]$ composes each element with itself thrice: $[3] : g \to g \cdot g \cdot g$, etc.

Because $\mathbf{Z}$ is commutative, the subring $[m]$ it maps to is also commutative, even though $\mathrm{End}(G)$ may not be.

10.5 Good Reduction

[1] An easy consequence of $\mathbf{Z}$'s repelling universal property in the category of rings, see [Lang], p. ?

Chapter 11

Algebraic Geometry

```
(%i3) diff(sqrt(x^4+1),x);
                                      3
                                   2 x
(%o3)                            ------------
                                       4
                                 sqrt(x  + 1)
(%i4) diff(%o3,x);
                                 2             6
                              6 x           4 x
(%o4)                      ------------ - -----------
                                 4          4     3/2
                           sqrt(x  + 1)   (x  + 1)
(%i5) diff(%o4,x);
                                        5             9
                        12 x          36 x          24 x
(%o5)                ------------ - ----------- + -----------
                           4          4     3/2     4     5/2
                     sqrt(x  + 1)   (x  + 1)      (x  + 1)
(%i6) diff(%o5,x);
                                   4             8             12
                  12          204 x         432 x         240 x
(%o6)        ------------ - ----------- + ----------- - -----------
                   4          4     3/2     4     5/2     4     7/2
             sqrt(x  + 1)   (x  + 1)      (x  + 1)      (x  + 1)
```

```
(%i19) diff(sqrt(x^3+1),x);
                                      2
                                   3 x
(%o19)                           --------------
                                        3
                                 2 sqrt(x  + 1)
(%i20) diff(%o19,x);
                                             4
                                 3 x           9 x
(%o20)                     ------------ - -------------
                                 3             3     3/2
                           sqrt(x  + 1)   4 (x  + 1)
(%i21) diff(%o20,x);
                                        3              6
                        3             27 x           81 x
(%o21)             ------------ - ------------- + -------------
                         3             3     3/2        3     5/2
                   sqrt(x  + 1)   2 (x  + 1)      8 (x  + 1)
```

11.1 Valuations

[van der Waerden], §18.1

A *valuation* is a generalization of the absolute value. A *valuation* is a mapping ϕ from a field $\mathbf{K}$ to an ordered field $\mathcal{R}$ (typically the reals) obeying the following axioms:

positivity	$\forall a \in \mathbf{K},$	$\phi(a) \geq 0$	(V1)
definiteness	$\forall a \in \mathbf{K},$	$\phi(a) > 0 \Longleftrightarrow a \neq 0$	(V2)
homomorphism (on the multiplicative group)	$\forall a, b \in \mathbf{K},$	$\phi(ab) = \phi(a)\phi(b)$	(V3)
subadditivity (or triangle inequality)	$\forall a, b \in \mathbf{K},$	$\phi(a + b) \leq \phi(a) + \phi(b)$	(V4)

A moment's thought will show that the standard absolute value on the reals obeys these axioms, as does the modulus on the complex field. Valuations are similar to norms, except that norms are defined on vector spaces, while valuations are defined on fields.

A valuation is said to be *non-Archimedian* if it also satisfies the following axiom, stronger than V4:

$$\text{non-Archimedian axiom} \quad \forall a, b \in \mathbf{K}, \quad \phi(a + b) \leq \max(\phi(a), \phi(b)) \quad \text{(V4')}$$

In this case, we can switch from a multiplicative to an additive notation and obtain *exponential valuation* by replacing $\phi(a)$ with $w(a) = -\ln \phi(a)$:

$\forall a \in \mathbf{K},$	$w(a) \in (-\infty, \infty]$	(E1)
$\forall a \in \mathbf{K},$	$w(a) = \infty \Longleftrightarrow a = 0$	(E2)
$\forall a, b \in \mathbf{K},$	$w(ab) = w(a) + w(b)$	(E3)
$\forall a, b \in \mathbf{K},$	$w(a + b) \geq \min(w(a), w(b))$	(E4)

Bibliography

[BrCo10] Briggs, Cochran, *Calculus*, 2^{nd} Edition. Pearson, 2010. ISBN-10: 0321336119