Modern Integration

Brent Baccala

Contents

Preface

In 1970, Robert Risch published [Ri70], which sketched in four pages how to bound the torsion of a divisor on an algebraic curve, and thus provided the “missing link” in a comprehensive algorithm that would either find an elementary form for a given integral, or prove that no such elementary form can exist. Risch’s method, suitably enhanced, is currently used in the symbolic integration routines of the most sophisticated computer algebra systems.

The goal of this book is to present the Risch integration algorithm in a manner suitable to be understood by undergraduate mathematics students, the prerequisites being calculus and abstract algebra, and the expected context being a senior-level university class.

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. One of the best ways to learn a subject 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 benefit of studying Risch integration 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 a specific goal.

Secondly, there is a sense of both historical and educational completion to be obtained. 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 without studying the Risch algorithm, the student is left with a vague sense that integration is nothing but a bag of tricks, and a real deficiency without understanding that the integration problem has been solved.

Third, an introduction to differential algebra may be quite appropriate at a point where 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, and that mystery has animated my own mathematical research for a number of years.

Furthermore, the available material on this subject is spread around among some terse research papers, some sparse lecture notes, and a single graduate level textbook ([Br05]), that while excellent, is unfortunately incomplete due to the untimely death of its author prior to completing an anticipated second volume. Having slowly assimilated this material over the course of years of study, and having given roughly a dozen lectures on Risch integration without the benefit of a textbook, the lack of a suitable text has become obvious. Although I began work on this textbook in 2006, I set it aside after a while and moved on to other interests. In the winter of 2016-17, I was once again preparing to lecture on Risch integration, and once again scrambling to pull everything together without a textbook.

Therefore, it seems appropriate to compile this knowledge together and offer it back to the mathematical community. Partly for the reasons I have listed above, and partly just to write something different from [Br05], I have decided to target this book at an undergraduate audience with some exposure to higher algebra.

I have liberally used the computer algebra system *Sage* in conjunction with the LaTeX package `pythontex`, which, suitable extended, allows *Sage* code in the LaTeX source to be automatically processed and typeset into the output. In keeping with my Christian religious principles, the book is freely available on the Internet, both in PDF form, and as LaTeX source in a `github` repository.

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 my dear friend Bruce Caslow, whose support and encouragement has been invaluable in this, as well as many other pursuits.

Chapter 1

Introduction

Who Wants to be a Mathematician?

$50,000 Question[1]

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

[1] The instructor may not necessarily possess a $50,000 prize fund.

The answer to this "\$50,000" question 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 6.6, 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 9.10:

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

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

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

$$y = \sqrt{x^4 + 4x^3 + 2x^2 + 1}$$

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

Integral (B), on the other hand, can not be "solved" in this manner, and example 7.5 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.

1.1 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}:

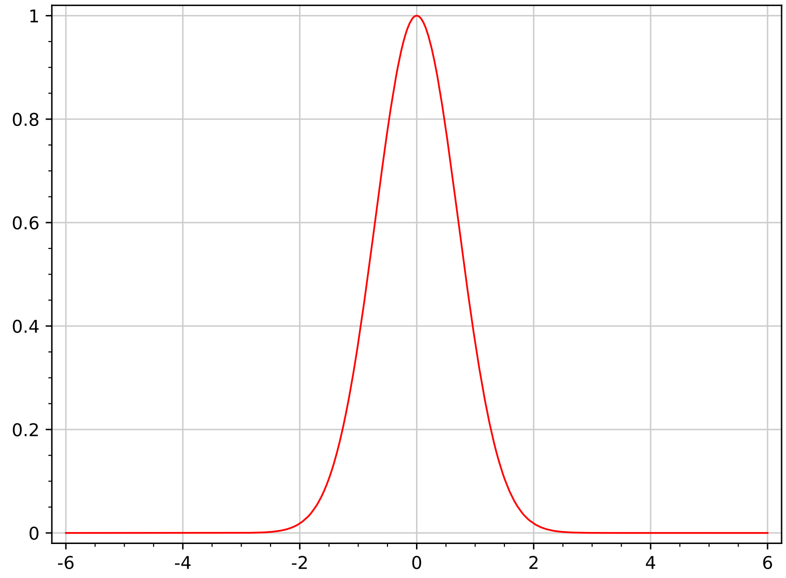


Figure 1.1: e^{-t^2}

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:

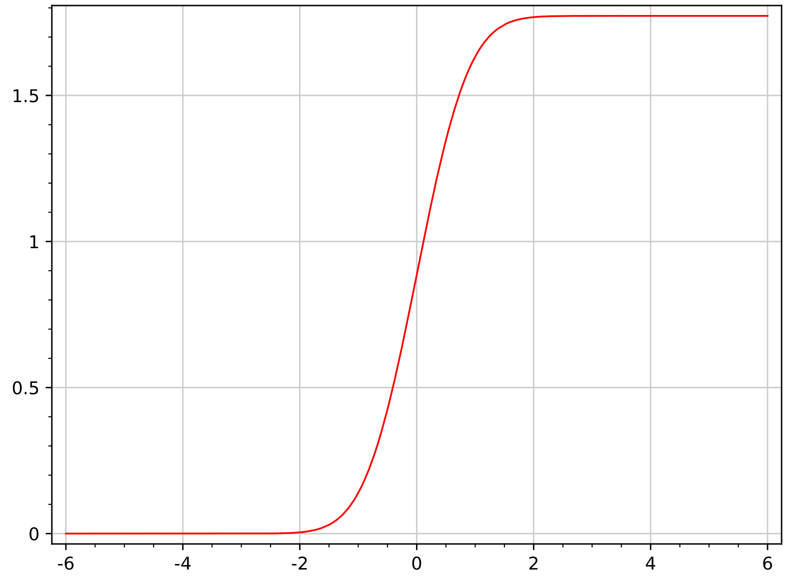


Figure 1.2: $\int_0^x e^{-t^2} dt$

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...

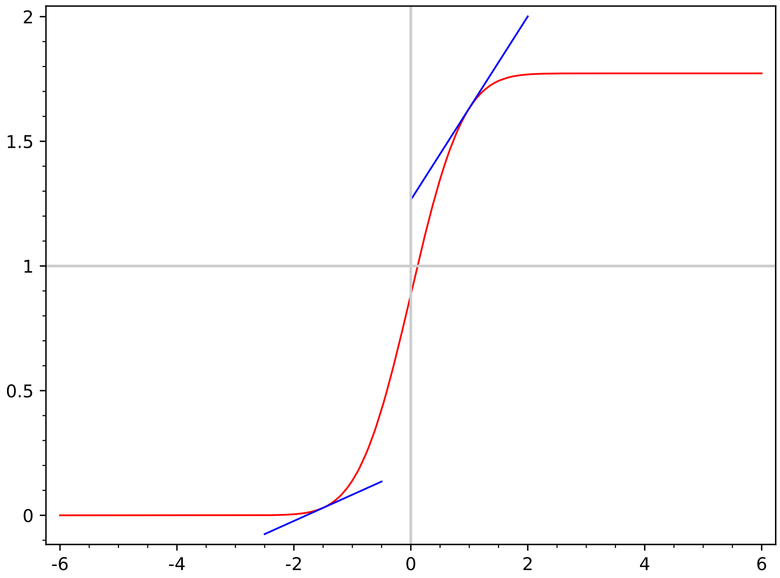


Figure 1.3: $\int_0^x e^{-t^2}\,dt$ (with tangent lines at $x_0 = -1.5$ and $x_0 = 1$)

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 $e^{-x_0^2}$.

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 probably 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 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 its anti-derivative is 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:

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

This is the cardinal sine, or *sinc* function, and it's easy to plot:

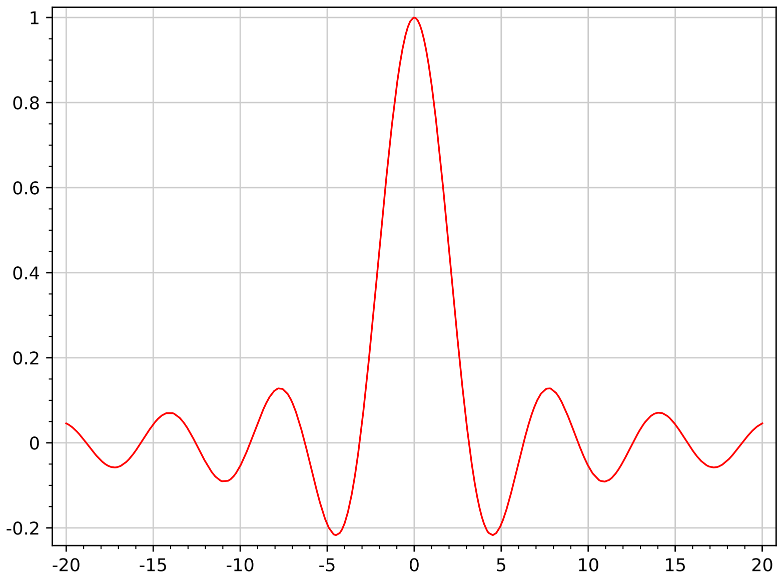


Figure 1.4: $\operatorname{sinc} t = \frac{\sin t}{t}$

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

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

...and we can plot it...

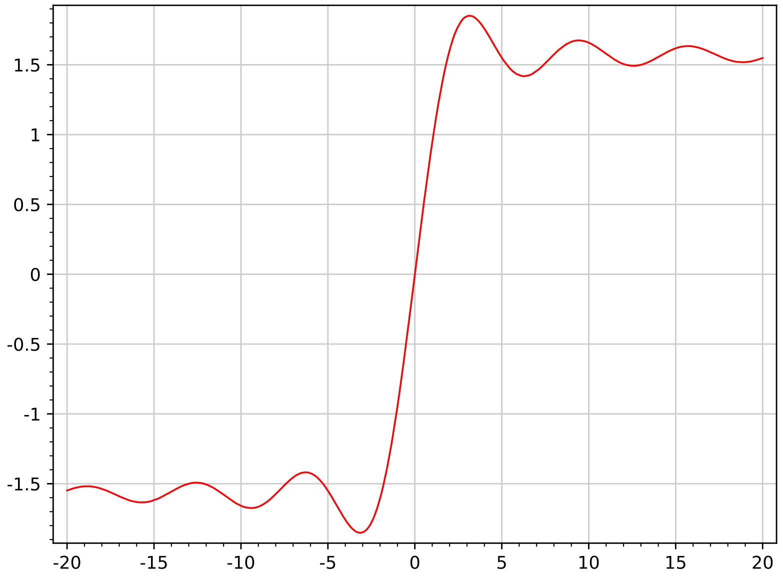


Figure 1.5: $\int_0^x \frac{\sin t}{t} dt$

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

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

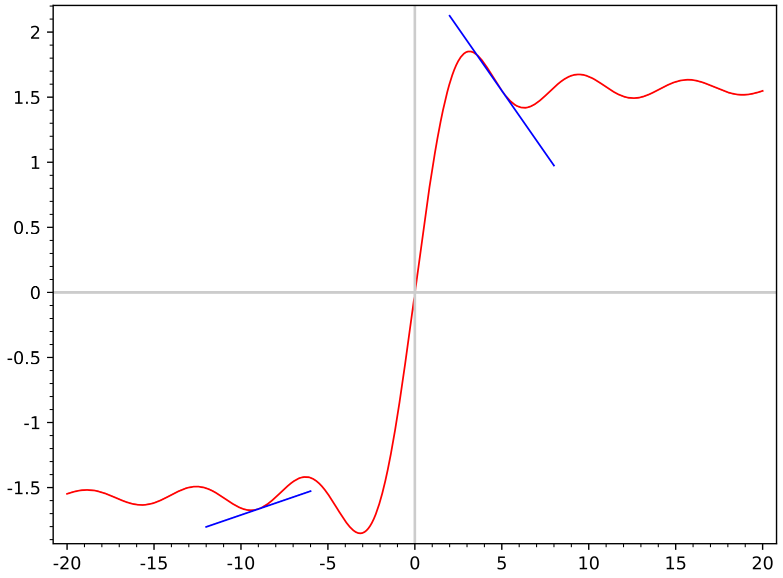


Figure 1.6: $\int_0^x \frac{\sin t}{t} dt$ (with tangent lines at $x_0 = -9$ and $x_0 = 5$)

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 develop a theory that lets us prove that these two integrals have no simple forms. One of the most surprising aspects of this theory is that it's based not on analysis, but rather algebra.

1.2 Algebra

In high school, we study what the Arabs called “al-jabr”, or what the Encyclopaedia Britannica 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:

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

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

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

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 into 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, anyway? Can we just make up new ones if the old ones weren't good enough?"

To which we wave our hands and reply, “A number is... 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 that 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 “crazy” integrals. Along the way, we will have

[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?

cause to at least survey some of the deepest waters of modern mathematics. Differential algebra is very much a 20$^{\mathrm{th}}$ 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 Richardson's Theorem

The absolute value function (in real analysis) or modulus function (in complex analysis) presents a serious obstacle to any attempt to develop an comprehensive algorithm for symbolic integration, as has been known since Daniel Richardson's proof of the following theorem in 1968:

Richardson's Theorem

Let E be a set of expressions representing real, single valued, partially defined functions of one real variable. E^* will be the set of functions represented by expressions in E.

If A is an expression in E, $A(x)$ is the function denoted by A.

It is assumed that E^* contains the identity function and the rational numbers as constant functions and that E^* is closed under addition, subtraction, multiplication and composition. In every case it is also supposed that given A and B in E there is an effective procedure for finding expressions in E to represent:

$$A(x) + B(x)$$
$$A(x) - B(x)$$
$$A(x) \cdot B(x)$$
$$A(B(x))$$

$A(x) \equiv B(x)$ will mean that $A(x)$ and $B(x)$ are defined at the same points and are equal wherever they are defined.

The integration problem for (E, E^*) is the problem of deciding, given A in E, whether there is a function $f(x)$ in E^* so that $f'(x) \equiv A(x)$.

If E^* satisfies conditions 1, 2, and 3, the integration problem for (E, E^*) will be unsolvable.

1. E^* contains $\ln 2$, π, e^x, $\sin x$.
2. There is a function, $\mu(x)$, in E^* so that $\mu(x) = |x|$ for $x \neq 0$.
3. There is a totally defined function, $\mathcal{B}(x)$, in E^* so that for no function, $f(x)$, in E^* and no interval I, is $f'(x) = \mathcal{B}(x)$ on I.

Daniel Richardson,
Some Undecidable Problems involving Elementary Functions of a Real Variable,
The Journal of Symbolic Logic, Volume 33, Number 4, Dec. 1968

Obviously, any algorithm purporting to work over real or complex numbers as its coefficient field will include π and $\ln 2$. Furthermore, we will see later in this text that $\sin x$ and e^x don't present any serious problem. The presence of the absolute value function is the critical component that leads to the undecidability of the problem. Why?

Not only is the absolute value function non-differentiable at the origin, but its obvious generalization to complex numbers, the modulus function, fails to be analytic *anywhere*. It also introduces an ordering on our field of constants, and allows the two square roots of a positive integer to be distinguished from each other.

Richardson's proof strategy is not based on differential algebra, but rather is developed from the undecidability proof of Hilbert's Tenth Problem, the so-called MRDP theorem, named after the surnames of its four principal contributors.

No attempt will be made to prove Richardson's Theorem here, as the proof of the MRDP theorem is notoriously difficult. I don't know it, and it would take us too far astray.

My approach in this textbook is simply to note the significance of Richardson's Theorem, discard absolute value and complex modulus from our menu of elementary operations, and move on.

1.4 Algorithms

Several algorithms are required, some of which are too complicated to discuss here in full detail. The difficult algorithms are:

- Polynomial Factorization
- Primary Decomposition
- Constructing the basis for a Riemann-Roch space
- Basic field operations on an algebraically closed field constants, including extension for polynomial factorization

A complete implementation of the theory described in this book would require implementations of all of these algorithms to be available, in both characteristic zero and prime characteristic.

1.5 Sage

To demonstrate the power of this theory, it's important to do some fairly difficult examples, such as integrals (C) and (D) in the "Who Wants to be a Mathematician?" question on the first page of this chapter. However, attempts to work such difficult integrals by hand quickly bogs down in a mess of algebra. Therefore, the judicious use of a computer algebra system is an essential learning tool in this subject. I've chosen the open source program Sage, and will use it liberally in the book's exercises.

In the course of writing the book, I've encountered a number of computation that Sage couldn't do, and have added code to Sage (taking advantage of its open source philosophy) to improve its functionality. While some of these features have been incorporated into the main Sage distribution, not all have, so the examples in the book are constructed using a customized version of Sage, the source code of while I maintain in a Github repository[7]. To do the exercises yourself, it's likely that you'll need to download and compile this customized version of Sage.

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

1.5.1 Tables

Sage has matrices where all entries share a common mathematical type, but lacks any native tool for formatting tables that contain entries with different mathematical types. The following code creates a `table` function that accepts a list of lists and formats the output as a table:

```
class LatexObject(SageObject):
    def __init__(self, str):
        self.str = str
    def _latex_(self):
        return self.str

def table(listOfLists):
    return LatexObject('\\begin{aligned}' + '\\\\'.join(['&\\quad&'.

# table1() is an attempt to create a table that contains series
# expansions aligned so that terms of the same power line up under
# each other.  We also have to deal with header cells that won't hav
# any alignment chars at all.  It's got a lot of problems.
#
# I've modified LaurentSeries to accept a 'table' parameter that
# causes TeX alignment characters to be inserted in the output.  We
# assume that each series has the same absolute precision, so any
# difference in the number of alignment characters is caused by the
```

[7] https://github.com/BrentBaccala/sage

```
# valuation of the series (i.e, the power of the leading term)
# being different, and we patch this up by inserting extra
# alignment characters at the beginning of the cell.

def table_latex(e):
    if isinstance(e, LaurentSeries):
       return e._latex_(table=True)
    else:
       return latex(e)

def fill_amps(target_amp_count, e):
    # If there are any TeX alignment characters in the cell,
    # this is (probably) a series with a higher valuation, so
    # pad it with enough &s to left-fill the cell.  Otherwise,
    # pad it with enough \omit\span's to span the cell, since
    # it's (probably) a header cell.  I don't use the suggested
    # practice of dropping the trailing \span, because that
    # seems to take us out of math mode and cause errors.
    if e.count('&') > 0:
        return (target_amp_count-e.count('&'))*'&' + e
    else:
        return (target_amp_count)*'\\omit\\span' + e

def table1(listOfLists):
    latex_listOfLists = [[table_latex(e) for e in l] for l in listOfList
    ampcount = map(max, zip(*[[e.count('&') for e in l] for l in latex_l
    latex_listOfLists_2 = [[fill_amps(ampcount[i], e) for i,e in enumera
    return LatexObject('\\begin{alignedat}{' + str(sum(ampcount)) + "}"
```

1.5.2 Arrays

This next function is convenient for displaying Sage arrays.

Here's a Python trick[8] to find out what variable name a caller uses:

```
import inspect

def varName(var):
    lcls = inspect.stack()[2][0].f_locals
    for name in lcls:
        if id(var) == id(lcls[name]):
            return name
    return None

def displayarray(b):
```

[8] https://stackoverflow.com/questions/2749796

```
  varname = varName(b)
  for k in sorted(b.keys()):
    print('$$', varname, '_{')
    if isinstance(k, tuple):
      print(','.join([latex(kk) for kk in k]))
    else:
      print(latex(k))
    print('}=', latex(b[k]), '$$')
```

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. Where I have omitted proofs, I have tried to provide references to [Go14], which is a good introductory algebra textbook that is freely available on-line.

2.1 Rings and Fields

[van der Waerden], §3.1
[Go14], §1.11

We begin with two key definitions that we will use throughout: the *ring* and the *field*.

A *ring* is a mathematical system where addition, subtraction, and multiplication are defined.

A *field* is a mathematical system where addition, subtraction, multiplication, and division are defined.

Both concepts are associated 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 called a field.

A *ring* $\mathcal{R}$ obeys the axioms in figure 2.1.

Notice the commutative law of multiplication (CR1), along with the existence of a multiplicative identity (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 require the existence of a multiplicative identity. Much of ring theory

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)

Figure 2.1: Ring axioms

All ring axioms, plus:

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

Figure 2.2: Field axioms

can be developed without this axiom, 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) through (R6) to obtain a *commutative ring with unity*, and call it 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 one additional axiom (Figure 2.1).

Informally, 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 **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 **Z**) is not necessarily invertible — there is no integer that when multiplied by 2 forms 1. (F1) is not satisfied. Thus **Z** forms a ring but not a field.

2.2 Fraction fields

The field Q

[van der Waerden], §3.3
[Go14], §6.4

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 *fraction 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 fraction 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:

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

Figure 2.3: Fraction field operations

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. 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 $\frac{2}{4}$ 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

All ring axioms, plus:

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

Figure 2.4: Integral domain axioms

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 identity 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 fraction 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, plus one more (figure 2.2) that guarantees the non-existence of zero divisors. All of the rings in this book are integral domains.

The fraction 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 fraction 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 fraction 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 polynomials in some variable with coefficients in the underlying field, all but a finite number of which must be zero.[1] 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 fraction 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, since it's equal to 1! 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 fraction 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 equivalence class that includes $\frac{2}{4}$.

It's convenient to select one unique representative of each equivalence class. In the case of fraction fields, we'll use the fraction with no common factors between the numerator and the denominator, i.e, $\frac{1}{2}$ instead of $\frac{2}{4}$ and $\frac{1}{x}$ instead of $\frac{x}{x^2}$. To reduce any given fraction to its canonical form, we need to compute the greatest common divisor of the numerator and denominator, and divide it out. The simplest way to do this involves long division.

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

2.4 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:

$$\begin{array}{r|r}
 & \frac{1}{2}x + \frac{3}{4} \\ \cline{2-2}
2x+1 & x^2 + 2x + 1 \\
 & -(x^2 + \frac{1}{2}x \qquad) \\ \cline{2-2}
 & \frac{3}{2}x + 1 \\
 & -(\frac{3}{2}x + \frac{3}{4}\,) \\ \cline{2-2}
 & \frac{1}{4}
\end{array}$$

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.

2.5 Greatest Common Divisors

[van der Waerden], §3.7, §3.8, §5.4 (multivariate rings)
[Ge92], 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.[2] 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 ([Go14] Theorem 1.8.16). This has been known since the time of Euclid, at least in the case of integers.

Nothing in our ring axioms guarantees the existence of GCDs. The problem is that there might be a lattice of divisors for a given element, instead of a strict ordering of them. However, GCDs exist in any integral domain in which the procedure described in the last paragraph can be carried out ([Go14] Theorem 6.5.8 and Lemma 6.6.2).

A *GCD domain* is an integral domain in which any two elements have a greatest common divisor.

A *unit* is an invertable element.

There is usually more than one GCD. 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.

[2] See [Ge92], for example

Example 2.1. Compute a GCD of $4x^4 + 13x^3 + 15x^2 + 7x + 1$ and $2x^3 + x^2 - 4x - 3$ in $\mathbf{Q}[x]$.

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

$$
\begin{array}{r|rrrr}
\multicolumn{1}{r}{} & & & \tfrac{4}{35}x & -\ \tfrac{6}{35} \\ \cline{2-5}
\tfrac{35}{2}x^2 + 35x + \tfrac{35}{2} & 2x^3 & +\ \ x^2 & -\ \ 4x & -\ \ 3 \\
\multicolumn{1}{r}{} & -(2x^3 & +\ 4x^2 & +\ \ 2x &) \\ \cline{2-5}
\multicolumn{1}{r}{} & & -\ 3x^2 & -\ \ 6x & -\ \ 3 \\
\multicolumn{1}{r}{} & & -(-3x^2 & -\ \ 6x & -\ \ 3\ ) \\ \cline{3-5}
\multicolumn{1}{r}{} & & & & 0
\end{array}
$$

The divisor of the last step, in this case $\frac{35}{2}x^2 + 35x + \frac{35}{2}$, is a GCD, and multiplying it 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.

The Sage function `gcd` compute GCDs.

```
sage: gcd(4*x^4 + 13*x^3 + 15*x^2 + 7*x + 1,
          2*x^3 + x^2 - 4*x - 3)
```

$$x^2 + 2\,x + 1$$

□

Multivariate GCDs

A few words are in order here about GCDs of multivariate polynomials. 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 fraction field with respect to one of the two variables and thus obtain a polynomial ring (in the other variable) over the fraction field. In the case of $\mathbf{Q}[x, y]$ we do not form a fraction 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 GCD domain is itself a GCD domain ([Go14] Lemma 6.6.2 and Theorem 6.6.7), so by induction any finite series of polynomial rings (like $\mathbf{Q}[x, y]$) is also a GCD domain. This implies that GCDs exist in multivariate polynomial rings. The problem is finding them, since the procedure described above requires long division, and this only works cleanly in an $\mathcal{F}[x]$-type system.

One solution is to pick one variable (say, x) to form coefficients (in the ring $\mathbf{Q}[x]$), then use the second variable (say, y) to form polynomials in $\mathbf{Q}[x][y]$.

Then we factor out of each $\mathbf{Q}[x][y]$ polynomial the GCD of the coefficients (calculated in $\mathbf{Q}[x]$), which we call the *content* of the polynomial, leaving a *primitive polynomial.* It can be shown ([Go14] Lemma 6.6.9) that if a primitive polynomial factors at all, then it factors into primitive polynomials. We thus can compute a GCD of the primitive parts and multiply this by the GCD of the contents to obtain a GCD in $\mathbf{Q}[x, y]$.

We will have little use for multivariate polynomial factorizations in this book, since invariably we will calculate GCDs with respect to one variable, and form fraction fields from any others, and thus always be working in $\mathcal{F}[x]$ systems.

Example 2.2. Compute the GCD of $5xy - 5y^2 - 7x + 7y$ and $2x^2 - yx - y^2$ in $\mathbf{Q}[x, y]$.

$\mathbf{Q}[x, y]$ is a multivariate polynomial ring, but 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 and our coefficients are in $\mathbf{Q}[x]$:

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

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 $\mathbf{Q}(x)[y]$:

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

$$\begin{array}{r l}
 & \frac{5}{2x+1}y - \frac{7}{2x+1} \\ \cline{2-2}
-(2x+1)y+(2x^2+x) & \big)\, -5y^2 + (5x+7)y - 7x \\
 & -(-5y^2 + 5xy\,) \\ \cline{2-2}
 & 7y - 7x \\
 & -(7y - 7x\,) \\ \cline{2-2}
 & 0
\end{array}$$

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 $\gcd(5y-7, 5y^2-7y) = 5y-7$. Dividing this out, we obtain:

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

and compute the GCD of these polynomials:

$$\begin{array}{r l}
 & 2x + y \\ \cline{2-2}
x-y & \big)\, 2x^2 - yx - y^2 \\
 & -(2x^2 - 2yx\,) \\ \cline{2-2}
 & yx - y^2 \\
 & -(yx - y^2\,) \\ \cline{2-2}
 & 0
\end{array}$$

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

```
sage: R.<x,y> = QQ[]
```

$$\mathbf{Q}[x, y]$$

```
sage: gcd(5*x*y - 5*y^2 - 7*x + 7*y,
          2*x^2 - y*x - y^2)
```

$$-x + y$$

□

2.6 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 **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:

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

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 that $\deg(t) < \deg(a)$.

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

Example 2.3. 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.

Using the notation:

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

we see that we are trying to solve $sa+tb=c$. a and b are the same polynomials used for the first GCD example. Recalling that GCD calculation:

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

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

The first division states that:

$$a \quad=\quad (2x+\frac{11}{2})b+(\frac{35}{2}x^2+35x+\frac{35}{2})$$

Rearranging this equation we get:

$$x^2+2x+1 \quad=\quad \frac{2}{35}a-\frac{1}{35}(4x+11)b$$

The second division isn't used in solving the polynomial Diophantine equation; it simply states that $\frac{35}{2}x^2 + 35x + \frac{35}{2}$ and $x^2 + 2x + 1$ are GCDs.

Having concluded that $x^2 + 2x + 1$ is a GCD of a and b, we divide it into c:

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

The remainder is zero, so the original problem has a solution (remember that c had to be a multiple of $\gcd(a, b)$).

We substitute our expansion for $x^2 + 2x + 1$ above into $c = (x + 3)(x^2 + 2x + 1)$ and obtain:

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

$\deg(\frac{2}{35}(x + 3)) = 1 < \deg(b) = 3$ and $\deg(\frac{1}{35}(4x^2 + 23x + 33)) = 2 < \deg(a) = 4$, so the degree bounds are met and we have our solution:

$$
\begin{aligned}
s &= \frac{2}{35}(x + 3) \\
t &= -\frac{1}{35}(4x^2 + 23x + 33)
\end{aligned}
$$

Now let's verify this solution using Sage:

```
sage: R.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: a = 4*x^4+13*x^3+15*x^2+7*x+1;

sage: b = 2*x^3+x^2-4*x-3;

sage: c = x^3+5*x^2+7*x+3;
```

```
sage: xgcd(a,b)
```

$$\left(x^{2} + 2x + 1, \frac{2}{35}, -\frac{4}{35}x - \frac{11}{35}\right)$$

```
def diophantine(a,b,c):
   (g,s,t) = xgcd(b,c)
   if not g.divides(a):
      raise ValueError("diophantine: a doesn't divide gcd(b,c)")
   s = s * (a//g)
   t = t * (a//g)
   # g = sb + ct
   # a = mg = (ms)b + (mc)t
   (q,r) = s.quo_rem(c)
   # s = r + qb, so r = s - qb
   return (r, t + q*b)
```

```
sage: (s,t) = diophantine(c,a,b)
```

$$\left(\frac{2}{35}x + \frac{6}{35}, -\frac{4}{35}x^{2} - \frac{23}{35}x - \frac{33}{35}\right)$$

```
sage: c == a*s + b*t
```

True

□

2.7 Partial Fractions Expansion

[van der Waerden], §5.10

As a first application of polynomial Diophantine equations, we use them to construct partial fractions expansions, which will be a major technical tool in the first half of the book.

Consider an element a from a polynomial fraction 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 ($\deg_x n < \deg_x d$), then we can construct a *partial fractions expansion* of a:

$$a = \frac{n}{d} = \sum_{i=1}^{n} \sum_{j=1}^{e_i} \frac{n_{i,j}}{d_i^{\,j}} \qquad \deg_x(n_{i,j}) < \deg_x(d_i) \tag{2.1}$$

We begin by computing an expansion in the form:

$$a = \frac{n}{d} = \sum_{i=1}^{n} \frac{n_i}{d_i^{\,e_i}} \qquad \deg_x(n_i) < e_i \deg_x(d_i) \tag{2.2}$$

i.e, each denominator factor is separated apart from the others, but there is only a single term for each denominator factor and the degree bounds are weaker.

n_i is found by solving the following polynomial Diophantine equation for n_i and r_i:

$$n \;=\; n_i \Big(\prod_{j \neq i} d_j^{\,e_j} \Big) + r_i (d_i^{e_i})$$

The degree bounds from the previous section guarantee that $\deg_x(n_i) < e_i \deg_x(d_i)$, and dividing through by d shows:

$$\frac{n}{d} \;=\; \frac{n_i}{d_i^{e_i}} + \frac{r_i}{\prod_{j \neq i} d_j^{\,e_j}}$$

We can now either ignore r_i and use this procedure to compute all of the n_i's, or we can note that 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.

Having computed an expansion of the form (2.2), a series of long divisions now suffices to seperate each of these fractions into a full partial fraction expansion in the form (2.1):

$$q_{i,e_i} = n_i \qquad q_{i,j} = q_{i,j-1} d_i + n_{i,j} \qquad q_{i,1} = n_{i,1}$$

The degree bounds on long division ensure that

$$\deg_x q_{i,j} < j \deg_x d_i \qquad \text{and} \qquad \deg_x n_{i,j} < \deg_x d_i$$

Assembling the $n_{i,j}$'s together, we obtain the form required by equation (2.1):

$$n_i = \sum_{j=1}^{e_i} n_{i,j}\, d_i^{e_i-j} \qquad \frac{n_i}{d_i^{e_i}} = \sum_{j=1}^{e_i} \frac{n_{i,j}}{d_i^j}$$

Theorem 2.4. *Partial fractions expansions that meet the minimal degree bounds are unique.*

Proof

If two different partial fractions expansions can be constructed for the same proper fraction a, then subtracting them from each other would yield a non-trivial partial fractions expansion for 0:

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{e_i} \frac{n_{i,j}}{d_i^j} \qquad \deg_x(n_{i,j}) < \deg_x(d_i) \tag{2.3}$$

We clear the denominators:

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{e_i} n_{i,j} \left(d_i^{e_i-j} \prod_{k \neq i} d_k^{e_k} \right) \tag{2.4}$$

Isolating the n_{1,e_i} term we obtain:

$$-n_{1,e_1} \prod_{k=2}^{n} d_k^{e_k} = \sum_{j=1}^{e_1-1} n_{1,j} \left(d_1^{e_1-j} \prod_{k=2}^{n} d_k^{e_k} \right) + \sum_{i=2}^{n} \sum_{j=1}^{e_i} n_{i,j} \left(d_i^{e_i-j} \prod_{k \neq i} d_k^{e_k} \right) \tag{2.5}$$

All of the terms on the right hand side include a d_1 factor, which implies that d_1 must also factor the left hand side. Since all of the d_i's are relatively prime, this d_1 factor can only come from n_{1,e_1} itself. Yet $\deg_x(n_{1,e_1}) < \deg_x(d_1)$, so this is impossible, and no such expansion can exist.

□

Example 2.5. Compute the partial fractions expansion of

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

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:

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

Next, we solve the polynomial Diophantine equation:

$$3x^2 - 4x + 2 = s(x-2)^2 + t(x+1) = sa + tb$$

First, we compute the GCD of $a = (x-2)^2 = x^2 - 4x + 4$ and $b = x + 1$. The actual result is obvious, but now we're interested in both the remainder and the quotient.

$$\begin{array}{r r}
 & x - 5 \\
x+1 & \overline{\big) x^2 - 4x + 4} \\
 & -(x^2 + \;\; x \qquad) \\
\cline{2-2}
 & -5x + 4 \\
 & -(-5x - 5\,) \\
\cline{2-2}
 & 9
\end{array}$$

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.

$$a = (x-5)b + 9$$

$$9 = a - (x-5)b$$

$$1 = \frac{1}{9}[a - (x-5)b]$$

$$3x^2 - 4x + 2 = \frac{1}{9}(3x^2 - 4x + 2)[a - (x-5)b]$$

$$3x^2 - 4x + 2 = \frac{1}{9}(3x^2 - 4x + 2)a - \frac{1}{9}(3x^3 - 19x^2 + 22x - 10)b$$

Our degree bounds aren't met yet, so we divide $b = x + 1$ into a's coefficient $3x^2 - 4x + 2$:

$$
\begin{array}{r|r}
 & 3x-7 \\ \cline{2-2}
x+1 & 3x^2-4x+2 \\
 & -(3x^2+3x) \\ \cline{2-2}
 & -7x+2 \\
 & -(-7x-7\,) \\ \cline{2-2}
 & 9
\end{array}
$$

The remainder becomes our new a coefficient, and after subtracting $(3x-7)a = 3x^3 - 19x^2 + 40x - 28$ from the b coefficient, we conclude that:

$$x^2 + 3x + 2 = a - (-2x+2)b$$

In other words (remember that $a = (x-2)^2$ and $b = x+1$),

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

Now we need only divide $(2x-2)$ by $(x-2)$:

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

so,

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

Here's Sage code to do a partial fractions expansion and save the results in an array:

```
def partfrac(num, den):
   b = {}
   factorization = factor(den)
   for f in factorization:
     (t,s) = diophantine(num, f[0]^f[1], den//(f[0]^f[1]))
     for i in range(f[1],1,-1):
       (s, b[f[0],i]) = s.quo_rem(f[0])
     b[f[0],1] = s
   return(b)
```

```
sage: b = partfrac(3*x^2 - 4*x + 2, x^3 - 3*x^2 + 4);
sage: displayarray(b);
```

$$b_{x-2,1} = 2$$

$$b_{x-2,2} = 2$$

$$b_{x+1,1} = 1$$

□

2.8 Resultants

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

At times, we will want a simple way of testing two polynomials in $\mathcal{F}[x]$ 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. The GCD exists in $\mathcal{F}[x]$, while the resultant is in $\mathcal{F}$.

For example, the polynomials $tx + x + t + 1$ and $ty + y$ share $t + 1$ as a GCD, so their t-resultant is zero. On the other hand, their x-resultant is not zero, because in the ring $\mathbf{C}(y, t)[x]$, $t + 1$ is a unit, so the GCD is 1. The resultant is constructed in an $\mathcal{F}[x]$-type system, so for multivariate polynomials, we always need to specify which variable is the ring variable; any remaining variables are implicitly field variables.

The resultant is defined[3] 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:

[3] There are other equivalent ways of defining the resultant.

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

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 (the f_i's are the bottom row of $S_x(P,Q)^{-1}$):

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

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 2.6. *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.

Example 2.7. Compute the $t-$resultant of $t^2 - 1 - x$ and $t^3 - t - y$.

```
sage: R.<x,y,t> = QQ[]
```

$$\mathbf{Q}[x, y, t]$$

```
sage: (t^2 - 1 - x).sylvester_matrix(t^3 - t - y, t)
```

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

```
sage: (t^2 - 1 - x).resultant(t^3 - t - y, t)
```

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

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 $t - 2$:

```
sage: (t^2-4).factor()
```

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

```
sage: (t^3-t-6).factor()
```

$$(t - 2) \cdot (t^2 + 2t + 3)$$

□

2.9 Algebraic Extensions

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

Both our fraction 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 fraction field construction with an algebraic extension; the ring is already a field.

Algebraic extensions are used to model fields where some algebraic expression (the minimal polynomial) is zero. Since adding any multiple of zero to an expression doesn't affect its value, all elements in an algebraic equivalence class are essentially the same value.

This isn't exactly the same as the real number system. Real numbers are typically constructed using Cauchy sequences of rational numbers. Thus, the square root of two is constructed as a real number using the Cauchy sequence $\{1, \frac{14}{10}, \frac{141}{100}, \frac{1414}{1000}, \frac{14142}{10000}, \cdots\}$, which obeys the algebraic relationship $\gamma^2 - 2 = 0$. The distinction is that the negative square root of two obeys the same algebraic relationship, but is constructed from a different Cauchy sequence: $\{-1, -\frac{14}{10}, -\frac{141}{100}, -\frac{1414}{1000}, -\frac{14142}{10000}, \cdots\}$. Thus γ, as a solution of $\gamma - 2 = 0$, could be either the positive or the negative square root of two; we can't distinguish them using only the four field operations (addition, subtraction, multiplication, and division). To distinguish between $+\sqrt{2}$ and $-\sqrt{2}$, we'd have to introduce an additional relationship – greater than or less than – which would produce an *ordered field*, and that's not part of our vocabulary. An ordered field can be treated as an unordered field by ignoring its ordering, which is fortunate for us, because it allows our integration theory to be applied to standard real numbers. For our purposes, γ, as a root of $\gamma^2 - 2 = 0$, is simply a field element that when squared equals two. It could be either $+\sqrt{2}$ or $-\sqrt{2}$; we can't tell the difference, but it isn't important because our results will be the same in either case.

Polynomial long division (section 2.4) can be used to divide any polynomial by the minimal polynomial, then we discard the quotient and use the remainder as the unique representative of our equivalence class. Thus, we can easily perform addition, subtraction, and multiplication in an algebraic extension field, by simply performing ordinary polynomial operations and then modulo reducing to a remainder, but how do we perform division? The polynomial Diophantine equation algorithm described in section 2.6 can be used to solve the following equation:

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

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

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.

We need to specify the variable in order to distinguish between operations in $\mathcal{C}(x)[y]$ mod $n(y)$ and $\mathcal{C}(y)[x]$ mod $n(x)$. For example, reducing mod $(y - x)$ effectively sets y equal to x, but it could also set x equal to y:

```
sage: R.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: S.<y> = Frac(R)[]
```

$$\mathrm{Frac}(\mathbf{Q}[x])[y]$$

```
sage: (x^3 + y) % (y-x)
```

$$x^3 + x$$

```
sage: R.<y> = QQ[]
```

$$\mathbf{Q}[y]$$

```
sage: S.<x> = Frac(R)[]
```

$$\mathrm{Frac}(\mathbf{Q}[y])[x]$$

```
sage: (x^3 + y) % (y-x)
```

$$y^3 + y$$

2.10 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 4 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.

$$\boxed{x^n + \underbrace{c_{n-1}}_{-\operatorname{Tr}(x)} x^{n-1} + \cdots + c_1 x + \underbrace{c_0}_{(-1)^n \mathrm{N}\,(x)} = 0}$$

The special significance of these elements becomes more obvious if we consider a *splitting field*, which is a further 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.

Let's write x's minimal polynomial in the form:

$$\prod_\sigma (t - \sigma x) = t^n - \sum_\sigma \sigma x t^{n-1} + \cdots + \prod_\sigma (-\sigma x)$$

where the sums and products are over all automorphisms that fix the base field.

Note that the $n-1^{\text{th}}$ coefficient in the polynomial is the sum of all the negatives of the conjugates, while the zeroth-order coefficient coefficient in the polynomial, is the product of all the negatives of the conjugates.

Definition 2.8. *The trace of an element x in E, written $\operatorname{Tr}(x)$, is the sum of all the conjugates of x.*

Definition 2.9. *The norm of an element x, written $\mathrm{N}(x)$, is the product of all the conjugates of x.*

The significance of these two functions is first, that they map elements from the extension field to the base field, and second, that trace commutes with addition and norm commutes with multiplication.

Example 2.10. Compute the norm and trace of $x + 1$ in $\mathbf{Q}[x] \mod x^2 - 2$.

The field $\mathbf{Q}[x] \mod x^2 - 2$ consists of the rational numbers, adjoined with $x = \sqrt{2}$. Actually, x is any square root of 2.

$$(x+1)^2 = x^2 + 2x + 1 \mod x^2 - 2 = 2x + 3 = 2(x+1) + 1$$

$$(x+1)^2 - 2(x+1) - 1 = 0$$

$$\operatorname{Tr}(x+1) = 2 \qquad \operatorname{N}(x+1) = -1$$

$$\operatorname{Tr}(x + (x+1)) = \operatorname{Tr} x + \operatorname{Tr}(x+1) = 0 + 2 = 2 = \operatorname{Tr} 1 = 2$$

□

Example 2.11. Prove that $(1+2i)$ is irreducible in $\mathbf{Q}[i] \mod i^2 + 1$.

The field $\mathbf{Q}[i] \mod i^2 + 1$ consists of the rational numbers, adjoined with i, the square root of -1. These form the *Gaussian integers*.

Since norm commutes with multiplication, any factorization of $(1+2i)$ must lead to a factorization of its norm, so let's compute the norm of $(1+2i)$:

$$(1+2i)^2 = 1 + 4i + 4i^2 \mod i^2 + 1 = -3 + 4i = 2(1+2i) - 5$$

$$(1+2i)^2 - 2(1+2i) + 5 = 0 \mod i^2 + 1$$

So, $\operatorname{N}(1+2i) = 5$, which is prime. Therefore, if $(1+2i)$ factors, at least one of its factors must have norm 1 or -1. There are no elements in this field with norm -1, and the only elements with norm 1 are 1, -1, i, and $-i$, all of which are units.

□

2.11 Hermite Normal Form

If we're given a ring R and a maximal ideal I, we know that $R \mod I$ will be a field, but how can we actually reduce an element e of R modulo I and find its **residue**, which is e's value in $R \mod I$, written $e \mod I$?

The answer to that question depends heavily on the ring R and how it and the ideal I are represented.

In Chapter 8, we'll need to perform this computation when R is a free $K[x]$-module (K is a field), which is to say that we'll have a set of basis elements b_i in a larger field, typically R's fraction field. Multiplying these basis elements by a vector in $K[x]$ will give us an element of R, and since the module is *free*, every vector in $K[x]$ will correspond to a unique element of R.

Since an element of R is specified as a vector in $K[x]$, we can form a square matrix corresponding to the element that, when multiplied by a vector v in $K[x]$ will return a vector in $K[x]$ that is exactly the multiplication of the element of R specified by the vector v by the element of R specified by the matrix. If we expand out the element horizontally (as a $K[x]$ vector), and list out its products with b_i vertically, we'll get a matrix that when multiplied on the left by v (as a row vector) will return the product as a row vector.

I will be given by a set of generators, also in R. Multiplying any one of these vectors by b_i will give us the corresponding generator. To form an element of I, we would then multiply each generator by an element of R and add the products.

In matrix notation, we can take the matrices that correspond to multiplying by each generator and stack them vertically. Now we'll left multiply by a row vector v in $K[x]$ with length ng where n is the degree of the $K[x]$-module and g is the number of generators. This corresponds to multiplying each generator by the corresponding element of R, of which there are g.

A matrix H (not necessarily square) with entries in $K[x]$ is in **Hermite Normal Form** if: (wikipedia, modified slightly)

1. H is upper triangular
2. The leading coefficient of a row is strictly to the right of the leading coefficent of the row above it, and it is monic
3. The elements below pivots are zero and elements above pivots are nonnegative and have degree strictly smaller than the pivot

A simple algorithm to convert a matrix M into HNF using elementary row operations works like this:

1. Let i run from 1 to the number of columns in M.

2. Use elementary row operations to move any rows with a zero in the i'th column, considering only rows i and lower, to the bottom of the matrix
3. Compute the monic GCD (use Section 2.5) of the entries in the i'th column, considering only rows i and lower (i.e, the diagonal element and everything below it), but ignore the rows with zeros
4. Use elementary row operations to construct the monic GCD in element e_{ii}.
5. Element e_{ii} will now be a factor of all elements beneath it, so use elementary row operations to replace these elements with zeros
6. Divide e_{ii} into each of the elements above it, and use elementary row operations to replace these elements by their remainders mod e_{ii}.

We can form HNF by left multiplying by an invertable matrix. So, we were left multiplying by an arbitrary row vector v of length ng to get vM, which is an element of the ideal, specified as a vector of length n. Using HNF, we now get $H = UM$, or $M = U^{-1}H$, or $vU^{-1}H$ as our result. Now the bottom $(n-1)g$ rows of the HNF will be all zeros, so the rightmost $(n-1)g$ elements of vU^{-1} will be ignored.

Given a HNF matrix H and an element e of R represented by a vector in $K[x]$, we can reduce modulo the ideal and obtain a vector u in $K[x]$ with each element of strictly lower degree than the corresponding diagonal element in H.

The algorithm to do this runs as follows:

1. Let i run from 1 to the number of columns in M.

HNF reduction will produce a row u with as many entries as rows in H, such that uH, dotted into b_i (call it uHB), will be an element of R that differs from e by less than the degree of the diagonal elements (clarify this). Since $vU^{-1} = u$, padding this vector with zeros and multiplying by U gives us $v = uU$, which is a list of elements that when multiplied by the ideal generators, would produce the same result.

If the ideal is maximal, the residue field is, in fact, a field, and will be K. This implies that the matrix H will have at most a single linear polynomial along the diagonal, all other entries in the matrix will be constants, and vHB will differ from e by a constant.

The bottom row is somewhat special, in that it will correspond to a generator using only a single basis element (the last one), and that generator will be the GCD of all elements in the ideal formed using only that basis element.

We'll usually want that final basis element to be 1, and the matrix element will be the GCD of all elements in the ideal formed with x and not y. In this case, we'll end up with $e - Hv$ (as vectors), being a vector with a single constant entry in the last element, and the rest of the vector zero.

2.12 Factorization

Some integers, as we know, are *prime* numbers, meaning that they can not be factored into a product of smaller integers, and the primes play a central role in number theory. In a more abstract setting, they generalize into the concepts of *irreducible elements* and *prime elements*, which are similar, but not identical. First, let me first introduce the simpler concept of a *unit*:

A *unit* is an invertable element.

Put another way, u is a unit if there exists v such that $uv = 1$. In a field, all elements except 0 are units, so units are only meaningful in the context of a ring.

An *irreducible element* is a non-zero, non-unit element that can not be written as the product of two non-units.

Any divisor can be transformed into another divisor by multiplying it by a unit, since if $uv = 1$, then $ab = (ua)(vb)$ for any a and b whatsoever. Therefore, factors are only unique up to units. For example, x can be "factored" as $1 \cdot x$ or $(-1) \cdot (-1) \cdot x$, but we don't regard these as factorizations for determining irreducibility, since 1 and -1 are units.

There exist rings in which indeterminates can be factored.

Example 2.12. In the ring $\mathcal{F}[x, y] \mod (y^2 - x)$, the element x is not irreducible.

$x \equiv y^2 \mod (y^2 - x)$, so x can be factored as $y \cdot y$.

Roughly speaking, y is the square root of x. □

A *prime element* p is a non-zero, non-unit element with the property that when p divides ab, p divides either a or b.

$$\forall a, b \quad p|ab \Rightarrow p|a \cup p|b$$

There exist rings in which these two concepts do not coincide.

Example 2.13. In the ring $\mathcal{F}[x, y, z] \mod (z^2 - xy)$, the element z is irreducible but not prime. [Wikipedia]

First, note that z is irreducible in this ring.

Then, consider the element xy. z divides xy since $xy \equiv z^2 \mod (z^2 - xy)$, so $\frac{xy}{z} \equiv z \mod (z^2 - xy)$. However, z divides neither x nor y. Therefore, z is not prime. □

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.

Unique Factorization

A *unique factorization domain* $\mathcal{U}$ is an integral domain in which every non-zero element can be written as a product of a unit and prime elements. [Wikipedia]

The *Fundamental Theorem of Arithmetic* states that $\mathbb{Z}$ is a unique factorization domain.

Theorem 2.14. *For any field K, the ring $K[x]$ is a unique factorization domain.*

Proof

Assume the contrary, that we're working in a unique factorization domain that is not an integral domain. Pick two elements c and d which are divisors 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).

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.

Algebraic Closure

[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 $\mathcal{F}[x]$.

A field $\mathcal{F}$ is *algebraically closed* if all polynomials in $\mathcal{F}[x]$ can be completely factored into factors linear (i.e, first degree) in x.

Put another way, a field $\mathcal{F}$ is algebraically closed if the only irreducible polynomials in $\mathcal{F}[x]$ are linear in x. These irreducible polynomials do not have to be linear in any other indeterminate that might exist in $\mathcal{F}$.

A polynomial is *monic* if its leading coefficient is 1. If $\mathcal{F}$ is algebrically closed, then the monic irreducible polynomials in $\mathcal{F}[x]$ are all of the form $(x - \lambda)$, where λ is some element of $\mathcal{F}$.

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.

In this book, the complex field $\mathbf{C}$ is the most common field that we'll use for our constants, precisely because it's algebraically closed and the irreducible polynomials of $\mathbf{C}[x]$ can be so easily characterized. It's major disadvantage is that it isn't *computable*, meaning that an arbitrary complex number can't be expressed in a finite form. This makes $\mathbf{C}$ fine for theoretical proofs and for specific examples, but unsuitable for algorithms designed to work with arbitrary inputs, for which smaller, more complicated fields are required.

I use the algebraic closure of $\mathbf{Q}$ extensively. This option causes Sage to print elements of $\overline{\mathbf{Q}}$ using radical notation, if possible:

```
QQbar.options.display_format = 'radical';
```

Square-free factorization

[Ge92], §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.

Full polynomial factorization (optional)

[van der Waerden], §5.6 [Ge92], 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 existence 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 $\phi_{x-a} : \mathbf{Z}[x] \to \mathbf{Z}$ from polynomials to integers. Not only is this a mapping, but it is a *homomorphism*, a mapping that preserves the operations, so $\phi_{x-a}(m+n) = \phi_{x-a}m \mathbin{\hat{+}} \phi_{x-a}n$ and $\phi_{x-a}(m \cdot n) = \phi_{x-a}m \mathbin{\hat{\cdot}} \phi_{x-a}n$, 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 $\mathbf{Z}[x]$.[5] 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 $-1 \cdot -3$.

2.13 Primary Decomposition

Primary decomposition generalizes polynomial factorization to systems of polynomial equations.

Ideals in rings generalize primes of integers. In the ring of integers, all ideals are principal, meaning that they are generated by a single element, so the ideals are in one-to-one correspondence with the integers themselves. The prime ideals are exactly those ideals generated by prime integers.

In more general rings, ideals are not necessarily principal, and factorization, while well defined, is not as simple as factoring an integer, although the RSA algorithm would be useless if factoring an integer was trivial.

Ideal in general factor in the following manner. An ideal can be written as the intersection of *primary* ideals, each of which is contained within a single *prime* ideal. While the

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

primary ideals are not uniquely defined, the associated prime ideals are. The radical of a primary ideal is its associated prime ideal, and if the original ideal is radical, then the primary ideals are themselves both radical and prime. Thus, while the factorization of an arbitrary ideal is only defined up to its associated prime ideals, a radical ideal is exactly equal to the intersection of its prime ideals.

How, then, do we compute a *primary decomposition* of an ideal?

Let's consider the easiest case first: a radical ideal in an polynomial ring over an algebraically closed field of constants, like $\mathbf{C}[x, y, z]$.

First, consider all possible subsets of the variables. Which subsets u, when intersected with the ideal $I \cup \mathbf{C}[u]$, form (0)? Is all we need to check are its generators? A largest such subset is called a *maximal independent set*; it may not be unique, but all maximal independent sets have equal size, and their size equals the dimension of the ideal.

Pick a maximal independent set u, form the fraction field $C(u)$, and consider the polynomial ring over that fraction field generated by the remaining elements $C(u)[x\backslash u]$. In this ring, the ideal is zero-dimensional. We can find the components of a zero-dimensional ideal by computing a Grobner basis using an elimination order and then factoring polynomials.

Ideals, however, can have components of varying dimension. We've only found the components of largest dimension. Once we've found the d-dimensional components, we can use *saturation* to divide them out. If the resulting ideal is (1), we're done, otherwise we continue going to find the remaining components, of smaller dimension. Obviously this process will eventually terminate.

Consider the ideal (xy, xz) in $\mathbf{C}(x, y, z)$. $\{y, z\}$ is a maximal independent set, because setting y and z to zero annihilates both generators. Consider the ideal in the ring $C(y, z)[x]$; it's generated by (x). Now we compute $(xy, xz) : (x)^{\text{inf}} = (y, z)$.

2.14 Exercises

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

1. $x^2 - 1$ (factors in all four rings)
2. $4x^2 - 1$ (factors in all rings except $\mathbf{Z}[x]$)
3. $x^2 - 2$ (factors in $\mathbf{R}[x]$ and $\mathbf{C}[x]$, but not in $\mathbf{Z}[x]$ or $\mathbf{Q}[x]$)
4. $x^2 + 1$ (factors only in $\mathbf{C}[x]$)

Write a computer program to factor the following polynomials:

5. $x^5 + 2x^4 + 2x^3 - x - 1$
   ```
   factor( x^5+2*x^4+2*x^3-x-1 )
   ```
   Ans: $(x^2 + x + 1) \cdot (x^3 + x^2 - 1)$
6. $34x^5 + 51x^4 + 60x^3 + 25x^2 + 8x - 1$
   ```
   factor( 34*x^5+51*x^4+60*x^3+25*x^2+8*x-1 )
   ```
   Ans: $(x^2 + x + 1) \cdot (34x^3 + 17x^2 + 9x - 1)$
7. $x^5 + 2x^4 + 2x^3 - x - 1$
   ```
   factor( x^5+2*x^4+2*x^3-x-1 )
   ```
   Ans: $(x^2 + x + 1) \cdot (x^3 + x^2 - 1)$

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

8. $2x^3 + 3x^2y - 7x^2 + 14xy + 21y^2 + 22x - 16y - 77$
   ```
   factor( 2*x^3+3*x^2*y-7*x^2+14*x*y+21*y^2+22*x-16*y-77 )
   ```
   Ans: $(2x + 3y - 7) \cdot (x^2 + 7y + 11)$
9. $x^2y^3 + 2x^4 + y^4 + 2x^2y + y^3 - x^2 - 3y - 3$
   ```
   factor( x^2*y^3+2*x^4+y^4+2*x^2*y+y^3-x^2-3*y-3 )
   ```
   Ans: $(x^2 + y + 1) \cdot (y^3 + 2x^2 - 3)$

Chapter 3

Algebraic Geometry

In pure algebra, such as we've studied in the last chapter, x is just x. It does not "take a value" and has no other interpretation. Everything we did in the last chapter is based solely on the axioms of a commutative ring or field.

We next want to consider what happens when we let x take a value, say a real number, and then conclude that the equation $x^2 - 1 = 0$ is true if x is either -1 or 1.

This is now *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, $\sum a_{ij}x^iy^j = 0$ is called an *algebraic curve*.

3.1 Solving systems of equations

Probably the most basic application of primary decomposition is solving systems of equations.

An ideal in a polynomial ring has an associated *variety*, which is the set of points that zero all of the polynomials in the ideal. A variety can be expressed as the intersection of its irreducible components, each of which corresponds precisely with a primary ideal in the primary decomposition.

If the underlying field of constants has no zero divisors (as is the case with ordinary real or complex numbers), then we are justified in taking the radical of the ideal, and then the primary decomposition is just the prime decomposition because our primary ideals are all prime ideals.

Returning to the (xy, xz) example, we found that this ideal's primary decomposition is $(x)\cap(y, z)$. In other words, the solution variety of the system of equations $xy = 0$, $xz = 0$ is formed by the z-axis $y = 0$ $z = 0$ and the $x - y$ plane $x = 0$.

Chapter 4

Differential Algebra

4.1 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:

$$\begin{array}{lll} \text{addition law of derivations} & \forall a, b \in \mathcal{D}, & D(a+b) = Da + Db \quad \text{(D1)} \\ \text{multiplication law of derivations} & \forall a, b \in \mathcal{D}, & D(ab) = aDb + bDa \quad \text{(D2)} \end{array}$$

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.

Let me immediately note the difference between differential algebra and algebraic geometry. In algebraic geometry, the indeterminates generally take the value of *constants*, typically real or complex numbers. In differential algebra, the indeterminates are now *functions* with non-trivial derivatives. In particular, we should not immediately expect that we can use differential algebra to solve a system of differential equations in the same manner as we can use algebraic geometry to solve a system of polynomial equations. Without introducing additional assumptions (which we absolutely can do), we're only going to find differential relationships between functions. The additional assumptions will typically take the form of assuming that the functions have some particular form, such as being elementary, and we're not going to be able to "solve" differential equations without these additional assumptions. If this seems at all puzzling, ask yourself what form those solutions would take, if an indeterminate is not a real or complex number, but rather some function with no additional assumptions about its structure.

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

Theorem 4.1.

- $D(0) = 0$
- $D(1) = 0$
- $D(\frac{1}{a}) = -\frac{1}{a^2} D(a)$
- $D(cx) = c\,D(x)$ *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 4.2. *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 4.3. *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 4.4. 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 7.7 for the actual integration.

□

That's it! The basic two differential axioms, algebraic extensions, fraction 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.

```
from sage.rings.fraction_field_element import is_FractionFieldElement
from sage.rings.polynomial.polynomial_element import is_Polynomial
from sage.rings.polynomial.multi_polynomial import is_MPolynomial

from sage.misc.latex import str_function

class Derivation(SageObject) :

    def __init__(self, parent, generator_map):
        self.generator_map = generator_map
        self.parent = parent

    def _repr_defn(self):
        gm = self.generator_map
        return '\n'.join(['%s |--> %s'%(i, gm[i]) for i in gm])

    def _repr_(self):
        return 'Derivation of %s\n  Defn: %s'%(self.parent, '\n         '.join(self._repr_

    def _latex_defn(self):
        gm = self.generator_map
        return '\\\\'.join(['%s & \\rightarrow & %s'%(latex(i), latex(gm[i])) for i in gm]

    def _latex_(self):
        return '\\begin{array}{rcl}\\multicolumn{3}{c}{' + str_function('Derivation of ') 

    def __call__(self, x):
        gm = self.generator_map
        if is_FractionFieldElement(x):
            n = x.numerator()
            d = x.denominator()
            return (self(n)*d - n*self(d))/(d^2)
        elif is_Polynomial(x):
            var = x.args()[0]
            result = 0
            for power in range(0, x.degree()+1):
                coeff = x[power]
                if power != 0:
                    result = result + power * coeff * var^(power-1) * gm[var]
                result = result + self(coeff) * var^power
            return self.parent(result)
        elif is_MPolynomial(x):
            vars = x.args()
            result = 0
            for (monomial, coeff) in x.dict().items():
                for i in range(len(monomial)):
                    power = monomial[i]
                    if power > 0:
                        var = vars[i]
                        Dvar = gm[var]
                        result = result + power * coeff * x.parent().monomial(*monomial) / var *
                        result = result + self(coeff) * x.parent().monomial(*monomial)
            return self.parent(result)
        elif x.parent() in [ZZ, QQ, AA, QQbar, RR, CC]:
            return 0
        else:
            raise NotImplementedError
```

```
sage: R.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: D = Derivation(R, {x: 1})
```

$$\begin{array}{l}\texttt{Derivation of }\mathbf{Q}[x]\\ x \;\rightarrow\; 1\end{array}$$

```
sage: D(x^2)
```

$$2x$$

```
sage: R.<x,theta,psi> = QQ[]
```

$$\mathbf{Q}[x,\theta,\psi]$$

```
sage: D = Derivation(R, {x: 1, theta: 1/x, psi: psi})
```

$$\begin{array}{lll}\multicolumn{3}{l}{\texttt{Derivation of }\mathbf{Q}[x,\theta,\psi]}\\ x & \rightarrow & 1\\ \theta & \rightarrow & \frac{1}{x}\\ \psi & \rightarrow & \psi\end{array}$$

```
sage: D(x*theta*psi)
```

$$x\theta\psi + \theta\psi + \psi$$

Definition 4.5. *An* **elementary extension** *of a differential field is a differential extension field constructed using a finite number of algebraic, logarithmic, and exponential exten-*

sions.

Definition 4.6. *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 4.7. 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 4.8. 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:

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

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

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

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.

4.2 Liouvillian Forms

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

4.3 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 4.9. *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* $\operatorname{Deg}_\theta p' = \operatorname{Deg}_\theta p - 1$
2. *If p's leading coefficient is not constant ($p'_n \neq 0$), then* $\operatorname{Deg}_\theta p' = \operatorname{Deg}_\theta p$
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':

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

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 4.1 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 2.7):

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

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 4.10. *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 4.11. 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 4.12. *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. $\operatorname{Tr}(Dx) = D(\operatorname{Tr} x)$,

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

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 4.3 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 an automorphism σ of A/K interacts with Tr. Remember that trace, in a Galois extension, can be written as a sum over all automorphisms that fix the base field:

$$\operatorname{Tr} x = \sum_\sigma \sigma x$$

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

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

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

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

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 $\operatorname{Tr}(Da) = nDa$, where n is the degree of the algebraic extension, by Theorem **??**. It follows that

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

Since $\operatorname{Tr} a \in K$, we have identified an element in K with the same derivation as a, which therefore can differ from a solely by an additive 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 4.13. 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:

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

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 4.14.

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

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 4.15. *(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 4.9(5), 4.10(4), and 4.12(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 $m_0 \in M$ and $M(\theta_i)$ are simple logarithmic extensions of M.

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

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

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

to reduce to the case where the m_i (except m_0) are all irreducible polynomials, so let's consider our three cases:

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

In this case, Theorem 4.9(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 4.9(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 4.9(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 4.10(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:

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

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

$$nl' = \mathrm{Tr}(m_0)' + \sum c_i \frac{\mathrm{N}\,(m_i)'}{\mathrm{N}\,(m_i)}$$

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

   Setting $k_0 = \frac{\mathrm{Tr}(m_0)}{n}$ and $k_i = \mathrm{N}\,(m_i)$, 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 5

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.

5.1 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:

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

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

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

5.2 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:

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

```
sage: R.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: a = x^4-3*x^2+6;

sage: b = x^6-5*x^4+5*x^2+4;
```

```
sage: gcd(b,b.differentiate())
```

$$1$$

```
sage: S.<z> = QQ[]
```

$$\mathbf{Q}[z]$$

```
sage: T.<x> = S[]
```

$$\mathbf{Q}[z][x]$$

```
sage: trager = (T(a) - z*T(b).differentiate()).resultant(b);
sage: trager.factor()
```

$$(2930944) \cdot (z^2 + \frac{1}{4})^3$$

```
sage: SS.<z> = QQbar[]
```

$$\overline{\mathbf{Q}}[z]$$

```
sage: SS(trager).roots()
```

$$\left[\left(-\frac{1}{2}i, 3\right), \left(\frac{1}{2}i, 3\right)\right]$$

These roots give us coefficients of the logarithmic terms.

```
sage: RR.<x> = QQbar[]
```

$$\overline{\mathbf{Q}}[x]$$

```
sage: loglist = [];

sage: for (r,m) in SS(trager).roots():
          s = (T(a)-z*T(b).differentiate()).map_coefficients(lambda v : v
          loglist.append(r * log(gcd(RR(b), RR(s))))
```

```
sage: sum(loglist)
```

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

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

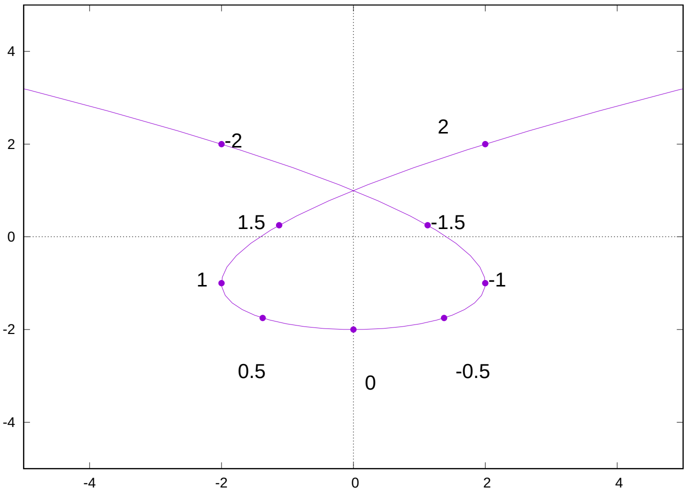


Figure 5.1: $(x^3 - 3x) + (x^2 - 2)i$

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:

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

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:

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

$$= -\frac{5\pi}{4} + \arctan\left(\frac{1}{2}\right) \approx -3.46$$

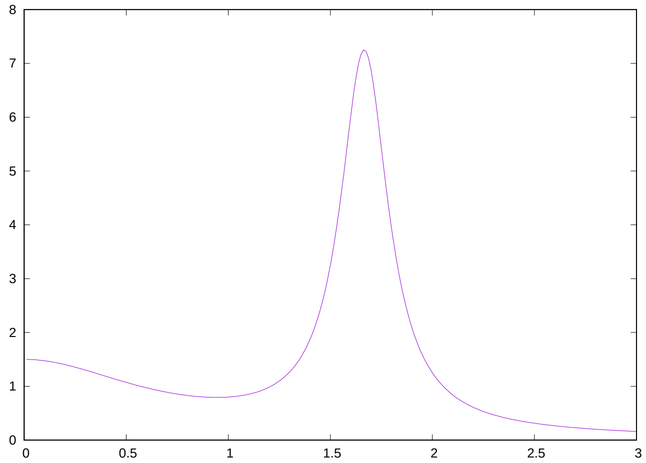
Since the integral was positive over the entire range of integration, this answer can not possibly be correct.

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

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

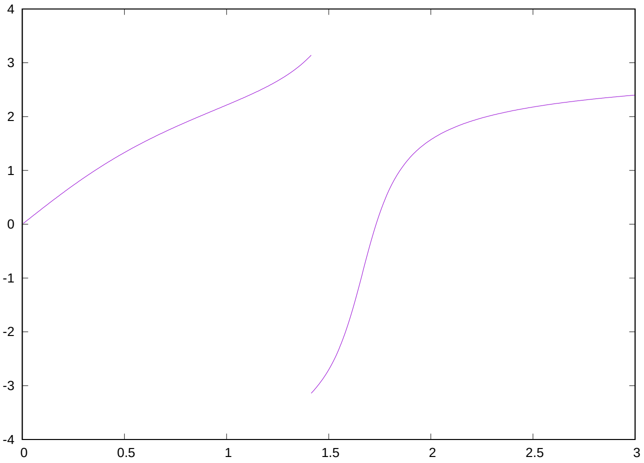
$$\int_1^2 \frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4} dx = \arctan 1 - \arctan 2 \approx -0.32$$

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



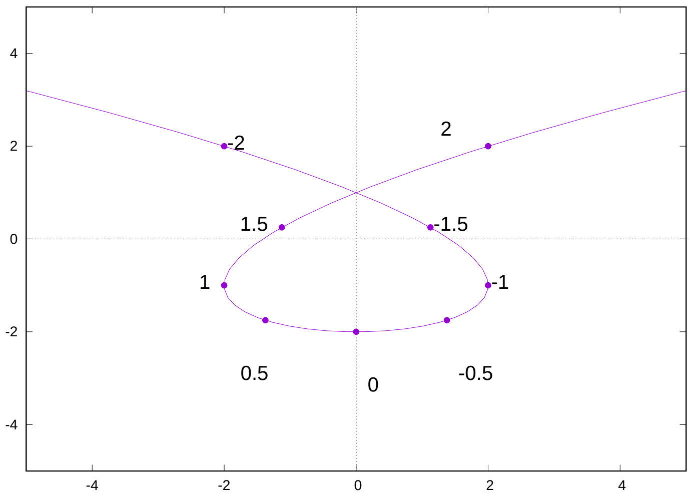
$$\frac{x^4 - 3x^2 + 6}{x^6 - 5x^4 + 5x^2 + 4}$$

And now the (indefinite) integral:



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

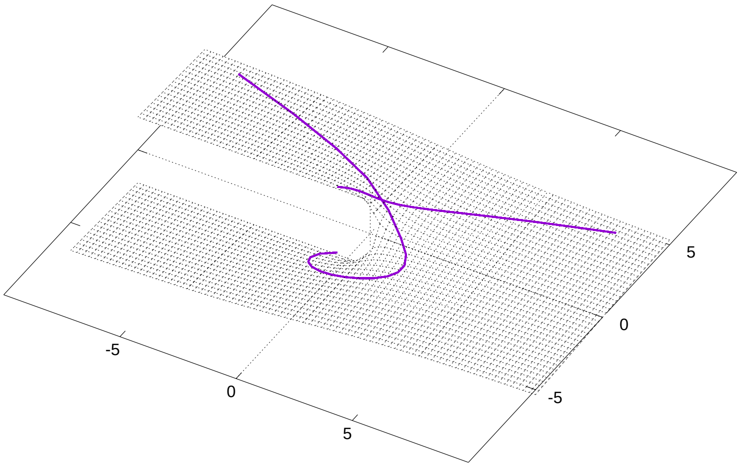
A discontinuity has appeared, seemingly out of nowhere. A closer inspection reveals that the break occurs suspiciously close to $\sqrt{2}$ — exactly where a plot of $(x^3 - 3x) + (x^2 - 2)i$ in the complex plane crosses the negative real axis:



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.

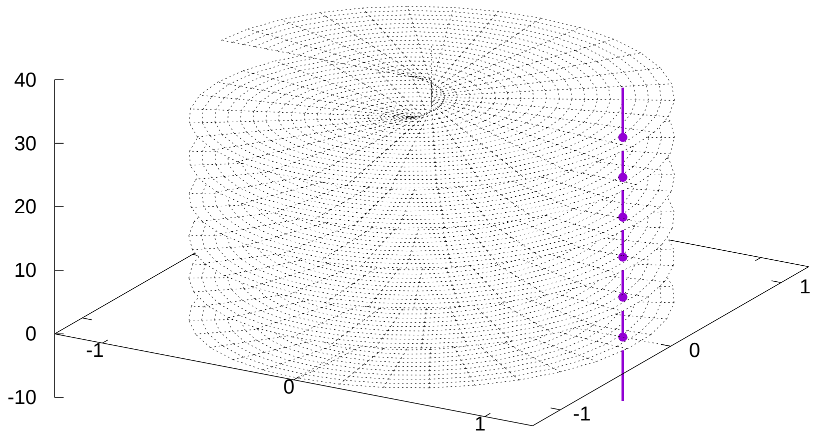
Where's the discontinuity in $\arctan$? Typically, the function is defined so that it ranges from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ and is continuous over finite numbers, so its discontinuity is where its argument becomes infinite; i.e, where its denominator goes to zero. In the $\ln(a+bi)$ expansion, 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 $(x^3-3x)+(x^2-2)i$.

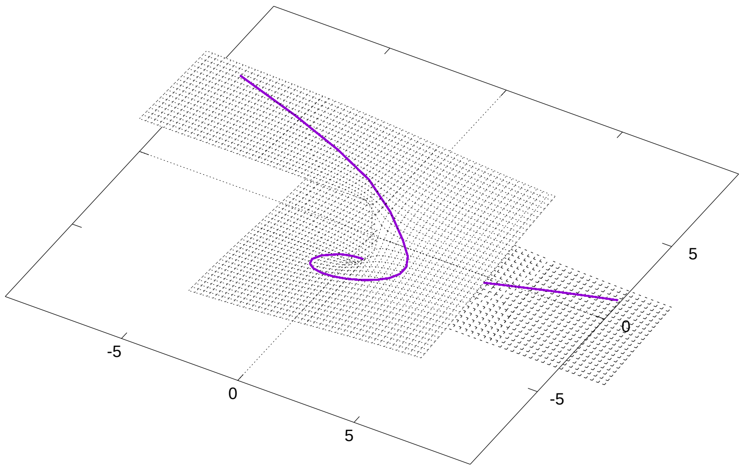
Of course, this isn't really $\arctan$'s Riemann surface, only a slice of it, and now we begin to find our solution. $\arctan$'s complete Riemann surface looks like an infinite screw, with an infinite series of blades, each spaced 2π apart in the z-direction. So, when we plot our function, what we really 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.



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

```
sage: (s,t) = diophantine(1, x^3-3*x, x^2-2)
```

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

```
sage: ((x^3-3*x)+(x^2-2)*QQbar(I)) * (t + s*QQbar(I)) * 2
```

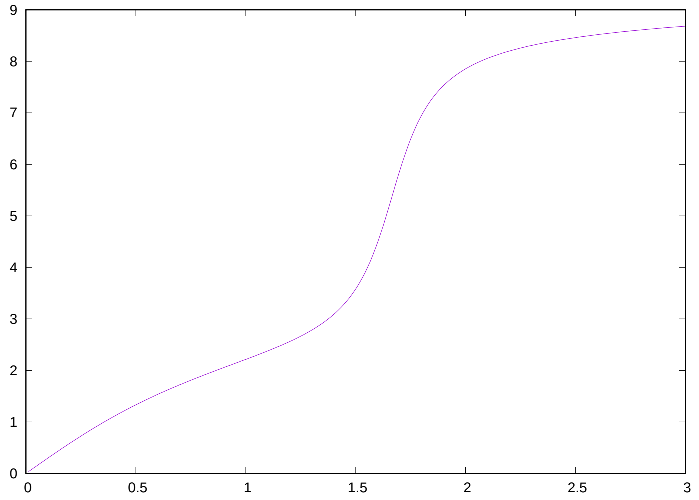
$$x^5-3x^3+x+2i$$

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

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

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

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


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

5.3 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 6

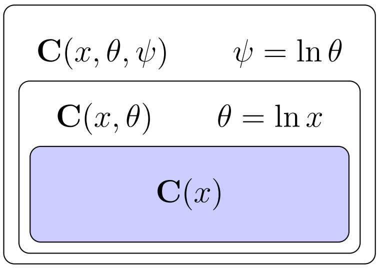
The Logarithmic Extension

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 6.7:

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

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.

6.1 The Logarithmic Integration Theorem

Theorem 6.1. *Let K be a differential field with $k \in K$, 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:*

$$f = \sum_{i=0}^{n} a_i \theta^i + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i} \frac{b_{i,j}(\theta)}{n_i(\theta)^j} \tag{6.1}$$

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

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

$$F = c_{n+1}\theta^{n+1} + \sum_{i=1}^{n} \left[A_i + c_i\right] \theta^i + A_0 + \sum_{i=1}^{\nu} \sum_{j=1}^{m_i - 1} \frac{B_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\nu} C_i \ln n_i(\theta) \tag{6.2}$$

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

and the following relationships hold:

$$\begin{aligned} (A_n + (n+1)c_{n+1}\theta)' &= a_n && && (6.3a) \\ (A_i + (i+1)c_{i+1}\theta)' &= a_i - (i+1)\frac{k'}{k} A_{i+1} && {\scriptstyle 0 \leq i < n} && (6.3b) \end{aligned}$$

$$\begin{aligned} R_{i,m_i-1}(\theta) &= b_{i,m_i}(\theta) && && (6.4a) \\ R_{i,j}(\theta) &= b_{i,j+1}(\theta) - B_{i,j+1}'(\theta) - Q_{i,j+1}(\theta) && {\scriptstyle 1 \leq j < m_i - 1} && (6.4b) \end{aligned}$$

$$B_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{j n_i'(\theta)} \mod n_i(\theta) \tag{6.5}$$

$$Q_{i,j}(\theta) = -\frac{R_{i,j}(\theta) + j B_{i,j}(\theta) n_i'(\theta)}{n_i(\theta)} \tag{6.6}$$

$$C_i = \frac{b_{i,1}(\theta) - B_{i,1}'(\theta) - Q_{i,1}(\theta)}{n_i'(\theta)} \qquad {\scriptstyle m_i > 1} \tag{6.7a}$$

$$C_i = \frac{b_{i,1}(\theta)}{n_i'(\theta)} \qquad {\scriptstyle m_i = 1} \tag{6.7b}$$

Proof

By Theorem 4.15, an elementary antiderivative of f can only exist in a finite logarithmic 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, then F becomes:

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

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 $\theta' = \frac{k'}{k}$:

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

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

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 $\deg R_{i,j}(\theta) < \deg n_i(\theta)$.

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

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 (6.3a) and (6.3b):

$$[A_n + (n+1)c_{n+1}\theta]' = a_n \tag{6.3a}$$

$$[A_i + (i+1)c_{i+1}\theta]' = a_i - (i+1)\frac{k'}{k}A_{i+1} \tag{6.3b}$$

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 4.15 tells us that integrals can generally include an arbitrary number of additional logarithms. The integrations required by (6.3a) and (6.3b) 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 (6.3b) reads:

$$[A_0 + c_1\theta]' = a_0 - \frac{k'}{k}A_1 \tag{6.8}$$

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 6.2. 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 (6.1) 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 (6.2) we get:

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

□

Example 6.3. 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 (6.1) 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 (6.2) we get:

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

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

□

Example 6.4. 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 4.2, 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 (6.1) with $n = 1$ and $a_1 = \frac{1}{2}i$, so Theorem 6.1 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 $A_1 = \frac{1}{2}ix$.

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

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

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

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

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

and therefore our solution is:

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

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

□

Example 6.5. 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 (6.3a) 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 6.6. Compute[1]

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

Due to the complexity of this integral, I'll use Sage for the entire calculation.

```
sage:  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)
```

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

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

```
sage: theta = var('theta');

sage: psi = var('psi');

sage: lintegrand = integrand.subs(
          {log(x+1) : theta,
           exp(x^2) : psi,
           exp(2*x^2) : psi^2,
           exp(3*x^2) : psi^3})
```

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

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 $\mathbb{C}(x, \psi)$!

Let's declare this field to Sage and tell it how our variables differentiate:

```
sage: F.<x,psi> = FractionField(ZZ['x', 'psi']);
```

[1] This integral [Ge92]'s Example 12.8.

```
sage: R.<theta> = F['theta']
```

$$\mathrm{Frac}(\mathbf{Z}[x,\psi])[\theta]$$

```
sage: D = Derivation(R, {x: 1, theta: 1/(x+1), psi: 2*x*psi})
```

$$\begin{array}{lcl}\multicolumn{3}{l}{\texttt{Derivation of }\mathrm{Frac}(\mathbf{Z}[x,\psi])[\theta]}\\ x & \rightarrow & 1\\ \theta & \rightarrow & \frac{1}{x+1}\\ \psi & \rightarrow & 2x\psi\end{array}$$

```
sage: num = R(lintegrand.numerator(False))
```

$$x\theta^4 - 2x^3\psi^2\theta^2 + \left(-4x^5\psi^3 - 4x^4\psi^3 - 2x^3\psi^3 - 2x^2\psi^3\right)\theta + x^5\psi^4 + 2x^3\psi^3$$

```
sage: den = R(lintegrand.denominator(False))
```

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

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

```
sage: (a,N) = num.quo_rem(den)
```

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

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

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

```
sage: A = integrate(SR(a), x)
```

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

Let's consider the fractional term. Polynomial factorization can be difficult, but this denominator is easy – it's just a difference of squares: $(\theta^2 - x^2\psi^2) = (\theta - x\psi)(\theta + x\psi)$. We'll factor the denominator into its irreducible factors and then perform a partial fractions expansion:

```
sage: n = [f[0] for f in factor(den)]
```

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

```
sage: b = partfrac(N, den);
sage: displayarray(b);
```

$$\begin{aligned} b_{\theta-x\psi,1} &= \frac{1}{-2x-2} \\ b_{\theta-x\psi,2} &= \frac{2x^4\psi^2 + 2x^3\psi^2 + x^2\psi^2 + x\psi^2 - x\psi}{-2x-2} \\ b_{\theta+x\psi,1} &= \frac{1}{2x+2} \\ b_{\theta+x\psi,2} &= \frac{2x^4\psi^2 + 2x^3\psi^2 + x^2\psi^2 + x\psi^2 + x\psi}{2x+2} \end{aligned}$$

Now we're ready to apply Theorem 6.1. We'll number our factors 0 and 1, since that's how Python like to do it, and start with $n_0(\theta)$. $m_0 = 2$, so

$$R_{0,1}(\theta) = b_{0,2}(\theta) = \frac{2x^4 + 2x^3 + x^2 + x}{2(x+1)}\psi^2 + \frac{x}{2(x+1)}\psi$$

and we wish to compute

$$B_{0,1}(\theta) \equiv -\frac{R_{0,1}(\theta)}{n_0'(\theta)} \mod n_0(\theta)$$

As modulo calculations go, this one is easy.

```
sage: R = {};
sage: B = {};
sage: R[0,1] = b[n[0],2]
```

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

```
sage: B[0,1] = - R[0,1] / D(n[0])
```

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

Now we wish to compute $Q_{0,1}(\theta)$:

$$Q_{0,1}(\theta) = -\frac{R_{0,1}(\theta) + B_{0,1}(\theta)n_0'(\theta)}{n_0(\theta)}$$

```
sage: Q = {};
sage: Q[0,1] = -(R[0,1] + B[0,1] * D(n[0])) / n[0]
```

$$0$$

This division to obtain $Q_{0,1}$ will always be exact. What might not be exact is the following division to obtain C_0. If the division isn't exact, or if C_0 isn't a constant, then the integral is not elementary.

$$C_0 = \frac{b_{0,1}(\theta) - B_{0,1}'(\theta) - Q_{0,1}(\theta)}{n_0'(\theta)}$$

```
sage: C = {};
sage: C[0] = (b[n[0],1] - D(B[0,1]) - Q[0,1]) / D(n[0])
```

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

A similar calculation handles the other irreducible factor:

```
sage: R[1,1] = b[n[1],2]
```

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

```
sage: B[1,1] = - R[1,1] / D(n[1])
```

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

```
sage: Q[1,1] = - (R[1,1] + B[1,1] * D(n[1])) / n[1]
```

$$0$$

```
sage: C[1] = (b[n[1],1] - D(B[1,1]) - Q[1,1]) / D(n[1])
```

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

Plugging everything together, we conclude that our solution is:

```
sage: lans = A + sum([B[i,1]/n[i] for i in range(2)]) \
        + sum([2 * C[i] * log(n[i]) for i in range(2)]).simplify_log()/2
```

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

Converting back to our original form:

```
sage: ans = lans.subs({theta : log(x+1), psi : exp(x^2)})
```

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

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

```
sage: bool(diff(ans,x) == integrand)
```

True

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

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

□

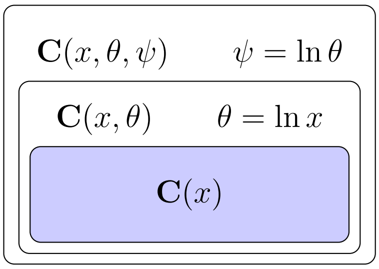
Example 6.7. Compute

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

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 (6.3a) reads:

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

We already performed this integration in example 6.3 with the result:

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

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

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

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

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:

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

□

6.2 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:

$$\begin{aligned} N &= VA' - (n-1)AV' + BV \\ N &= (A'+B)V - (n-1)AV' \end{aligned}$$

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 6.8. Redo Example 6.6 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:

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

```
sage: v = theta^2-x^2*psi^2
```

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

```
sage: vtick = D(v)
```

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

```
sage: (s,t) = diophantine(N/(x+1), v, vtick)
```

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

So, our solution is:

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

```
sage: A = -t
```

$$-x\psi\theta$$

```
sage: B = s - D(A)
```

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

$$A = -x\psi\theta$$
$$A' = -\psi\theta - x(2x)\psi\theta - x\psi\frac{1}{x+1}$$
$$A' + B = \frac{-2x}{x+1}\psi$$
$$B = (2x^2+1)\psi\theta - \frac{x}{x+1}\psi$$

Which reduces our integral to:

```
sage: integrate(SR(a),x) + A/v + integrate(SR(B/v), x, hold=True)
```

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

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

Now we can compute a Rothstein-Trager resultant:

```
sage: R1 = FractionField(ZZ['x','psi'])['z']
```

$$\mathrm{Frac}(\mathbf{Z}[x, \psi])[z]$$

```
sage: R = FractionField(R1)['theta']
```

$$\mathrm{Frac}(\mathrm{Frac}(\mathbf{Z}[x, \psi])[z])[\theta]$$

```
sage: r = R(v).resultant(R(B - R('z') * D(v)))
```

$$\left(\frac{16x^8\psi^4 + 32x^7\psi^4 + 32x^6\psi^4 + 32x^5\psi^4 + 20x^4\psi^4 + 8x^3\psi^4 + 4x^2\psi^4 - 4x^2\psi^2}{x^2 + 2x + 1}\right) z^2 + \frac{-4x^8\psi^4 - 8x^7\psi^4 - 8x^6\psi^4 - 8x^5\psi^4 - 5x^4\psi^4 - 2x^3\psi^4 - x^2\psi^4 + x^2\psi^2}{x^2 + 2x + 1}$$

This result is in $\mathbb{C}(x, \psi)[z]$, so the first two factors are just *content* (EXPLAIN THIS TERM).

```
sage: R1(r) / R1(r).factor().unit()
```

$$4z^2 - 1$$

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:

```
sage: gcd(B - (1/2)*D(v), v)
```

$$\theta + x\psi$$

```
sage: gcd(B + (1/2)*D(v), v)
```

$$\theta - x\psi$$

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

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

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

□

Chapter 7

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 4.10. *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. $\operatorname{Deg}_\theta p' = \operatorname{Deg}_\theta p$
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 4.10 (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:

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

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 $p_i' + Ap_i$, and equating the p' polynomial to the integrand's polynomial produces a differential equation of the form:

$$p_i' + Ap_i = B \qquad A, B, p_i \in K$$

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 2.6, 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.

7.1 The Exponential Integration Theorem

Theorem 7.1. *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:*

$$f = \sum_{i=0}^{n} a_i\theta^i + \sum_{j=1}^{l} \frac{b_j}{\theta^j} + \sum_{i=1}^{\nu}\sum_{j=1}^{m_i} \frac{c_{i,j}(\theta)}{n_i(\theta)^j} \tag{7.1}$$

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

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:

$$F = \sum_{i=0}^{n} A_i\theta^i + \sum_{j=1}^{l} \frac{B_j}{\theta^j} + \sum_{i=1}^{\nu}\sum_{j=1}^{m_i-1} \frac{C_{i,j}(\theta)}{n_i(\theta)^j} + \sum_{i=1}^{\nu} D_i \ln n_i(\theta) \tag{7.2}$$

$$A_i, B_j \in K \qquad C_{i,j}(\theta), n_i(\theta) \in K[\theta] \qquad D_i' = 0$$

and the following relationships hold:

$$A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} \tag{7.3}$$

$$A_i' + iA_ik' = a_i \qquad B_j' - jk'B_j = b_j \tag{7.4}$$

$$\begin{aligned} R_{i,m_i-1}(\theta) &= c_{i,m_i} && \text{(7.5a)}\\ R_{i,j-1}(\theta) &= c_{i,j} - C_{i,j}'(\theta) - Q_{i,j}(\theta) \qquad {\scriptstyle 1<j\le m_i} && \text{(7.5b)} \end{aligned}$$

$$C_{i,j}(\theta) \equiv -\frac{R_{i,j}(\theta)}{jn_i'(\theta)} \mod n_i(\theta) \tag{7.6}$$

$$D_i = \frac{c_{i,1} - C_{i,1}'(\theta) - Q_{i,1}(\theta)}{n_i'(\theta) - \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} n_i(\theta)} \tag{7.7}$$

Proof

By Theorem 4.15, 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 4.10):

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

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

$$F' = \sum_{i=0}^{N}(A_i'+iA_ik')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} + \sum_{i=1}^{\nu}\sum_{j=1}^{M_i} \frac{C_{i,j}'(\theta)n_i(\theta) - jC_{i,j}(\theta)n_i'(\theta)}{n_i(\theta)^{j+1}} + \sum_{i=1}^{\eta} D_i \frac{E_i'(\theta)}{E_i(\theta)}$$

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:

$$F' = A_0' + \sum D_k \frac{E_k'}{E_k} + \sum_{i=1}^{N}(A_i' + iA_ik')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} + \sum_{i=1}^{\nu}\sum_{j=1}^{M_i} \frac{C_{i,j}'(\theta)n_i(\theta) - jC_{i,j}(\theta)n_i'(\theta)}{n_i(\theta)^{j+1}} + \sum_{i=1}^{\eta} D_i \frac{n_i(\theta)'}{n_i(\theta)}$$

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.

$$\begin{aligned}F' = A_0' + \sum D_k \frac{E_k'}{E_k} + \sum D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} + \sum_{i=1}^{N}(A_i' + iA_ik')\theta^i + \sum_{j=1}^{L} \frac{B_j' - jk'B_j}{\theta^j} \\ + \sum_{i=1}^{\nu}\left[\frac{R_{i,M_i}(\theta)}{n_i(\theta)^{M_i+1}} + \sum_{j=2}^{M_i} \frac{C_{i,j}'(\theta) + Q_{i,j}(\theta) + R_{i,j-1}(\theta)}{n_i(\theta)^j}\right. \\ \left. + \frac{C_{i,1}'(\theta) + Q_{i,1}(\theta) + D_i\left[n_i(\theta)' - \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} n_i(\theta)\right]}{n_i(\theta)}\right]\end{aligned}$$

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:

$$A_i' + iA_ik' = a_i \qquad B_j' - jk'B_j = b_j$$

Normal denominators give rise to these terms:

$$\begin{aligned} R_{i,m_i}(\theta) &= c_{i,m_i+1} \\ C'_{i,j}(\theta) + Q_{i,j}(\theta) + R_{i,j-1}(\theta) &= c_{i,j} \qquad {\scriptstyle 1<j\leq m_i} \\ C'_{i,1}(\theta) + Q_{i,1}(\theta) + D_i n_i(\theta)' &= c_{i,1} \end{aligned}$$

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 2.6, 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 7.2. 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 7.1, 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 7.3. 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 (7.1) with $\nu = 2$, $n_1(\psi) = \psi - 1$, $n_2(\psi) = \psi + 1$, $m_1 = m_2 = 1$ and $c_{1,1} = c_{2,1} = i$.

$$D_1 = \frac{c_{1,1}}{n_1'(\theta) - \frac{\operatorname{lc} n_1'(\theta)}{\operatorname{lc} n_1(\theta)} n_1(\theta)} = \frac{i}{i\psi - i(\psi - 1)} = 1$$

$$D_2 = \frac{c_{2,1}}{n_2'(\theta) - \frac{\operatorname{lc} n_2'(\theta)}{\operatorname{lc} n_2(\theta)} n_2(\theta)} = \frac{i}{i\psi - i(\psi + 1)} = -1$$

so by Theorem 7.1 the integral can be written:

$$\int \left[\frac{i}{\psi - 1} + \frac{i}{\psi + 1}\right] dx = \ln(\psi - 1) - \ln(\psi + 1) = \ln\left(\frac{\psi - 1}{\psi + 1}\right)$$

$$= \ln\left(\frac{e^{ix} - 1}{e^{ix} + 1}\right) = \ln\left(\frac{e^{ix/2} - e^{-ix/2}}{e^{ix/2} + e^{-ix/2}}\right) = \ln i \frac{\sin\frac{x}{2}}{\cos\frac{x}{2}} = \ln \tan\frac{x}{2}$$

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 7.4. 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 (7.1) 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 (7.7), we obtain:

$$\begin{aligned}
D_1 &= \frac{c_{1,1}}{n_1(\theta)' - \frac{\operatorname{lc} n_1'(\theta)}{\operatorname{lc} n_1(\theta)} n_1(\theta)} = \frac{1}{i\theta - i(\theta - i)} = -1 \\
D_2 &= \frac{c_{2,1}}{n_2(\theta)' - \frac{\operatorname{lc} n_2'(\theta)}{\operatorname{lc} n_2(\theta)} n_2(\theta)} = \frac{-1}{i\theta - i(\theta + i)} = -1
\end{aligned}$$

Pluging this into equation (7.3), we obtain:

$$\begin{gathered}
A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_i'(\theta)}{\operatorname{lc} n_i(\theta)} \\
A_0' = -i + 2i = i \\
A_0 = ix
\end{gathered}$$

so our integral is:

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

We can convert this into a more familiar form by realizing that $ix = \ln \exp ix = \ln \theta$, so

$$= \ln \theta - \ln \left[\theta^2 + 1\right] = \ln \frac{\theta}{\theta^2 + 1} = \ln \frac{1}{\theta^{-1} + \theta}$$

$$\int \tan x \, \mathrm{d}x = \ln \csc x$$

□

7.2 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 7.1 then produces Risch equations in the underlying field; in this case, $\mathbb{C}(x)$.

Consider such a Risch equation:

$$r' + Sr = T \qquad S, T, r \in \mathbb{C}(x) \tag{7.8}$$

Recall that in $\mathbb{C}[x]$, all irreducible factors have the form $(x-\gamma)$, since $\mathbb{C}$ is algebraically closed. Futhermore, all irreducible factors in r's denominator increase in degree on differentiation, since all of $\mathbb{C}[x]$'s factors are normal. Therefore, the factors in r's denominator must appear in either S or T's denominator, because otherwise r' would have a factor in its denominator that could not cancel anything else in the equation. Thus, we can easily identify which factors can appear in r's denominator, and we next wish to calculate the multiplicities with which they appear.

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 (7.8), we find:

$$\frac{-ja}{(x-\gamma)^{j+1}} + \cdots + \frac{ba}{(x-\gamma)^{j+k}} + \cdots = \frac{c}{(x-\gamma)^l} + \cdots$$

We know k (the pole's multiplicity in S's denominator) and l (the pole's multiplicity in T's denominator), and we wish to determine j, the pole's multiplicity in r's denominator.

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, where $p, d \in \mathbb{C}[x]$:

$$\frac{p'd - pd'}{d^2} + S\frac{p}{d} = T$$

$$(p'd - pd') + Spd = Td^2$$

We still might have lingering denominators, which can be cleared by multiplying through by D, the least common multiple of the denominators of Spd and Td^2:

$$D(p'd - pd') + SDpd = TDd^2$$

$$(Dd)p' + (SDd - Dd')p = TDd^2$$

Setting $A = Dd$, $B = D(Sd - d')$ and $C = TDd^2$, we now have a polynomial Risch equation that must be satisfied by the numerator p:

$$Ap' + Bp = C \qquad A, B, C, p \in \mathbb{C}[x] \tag{7.9}$$

Our next aim is to upper bound the degree of p, and there are again three cases.

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:

$$A = a_j x^j + \cdots \qquad B = b_{j-1}x^{j-1} + \cdots \qquad p = p_k x^k + \cdots$$

$$Ap' + Bp = (ka_j p_k + b_{j-1}p_k)x^{j+k-1} \cdots$$

In order for this coefficient to be zero, $k = -b_{j-1}/a_j$. So, if these conditions are met:

$$\deg A = \deg B + 1 \qquad k = -\frac{b_{j-1}}{a_j} = -\frac{\operatorname{lc} B}{\operatorname{lc} A}$$

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

$$\deg p = \deg C - \max(\deg A - 1, \deg B)$$

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.

Summarizing, equation (7.9) can be solved only if one of these three conditions is met:

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 $pB = C$

They are not mutually exclusive. In particular, if both of the first two cases can be met, then the larger of the two values for $\deg p$ should be used.

Having determined the degree of p, we can now determine its coefficients by hypothesizing p as a polynomial of the calculated degree and plugging it into equation (7.9).

Example 7.5. 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 7.1 that our solution, if it exists, must have the form $A_1\psi$, where $A_1 \in \mathbb{C}(x)$, and A_1 must satisfy equation (7.4):

$$A_1' - 2xA_1 = 1$$

This is already a polynomial Risch equation, and A_1' has only a constant coefficient, so A_1 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 $\deg A \neq \deg B + 1$ (case 1), the leading terms on the left hand side can not cancel, so they must match the leading term on the right. We compute:

$$\deg C - \max(\deg A - 1, \deg B) = 0 - 1 = -1$$

so that doesn't work (case 2). Futhermore, C is not a constant multiple of B, so a constant can't solve our equation (case 3).

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 7.6. 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 $\mathbb{C}(x)$:

$$\frac{1}{2i}\int \left[\frac{1}{x}\psi - \frac{1}{x}\frac{1}{\psi}\right]\,dx$$

Applying theorem 7.1, we see that 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 equations (7.4):

$$A_1' + iA_1 = \frac{1}{x} \qquad A_{-1}' - iA_{-1} = -\frac{1}{x}$$

In both cases, we've got an equation of the form (7.8) with 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 7.7. Compute $\int \frac{4^x+1}{2^x+1}\mathrm{d}x$

We'll use the field $\mathbb{C}(x, \theta = \exp(x \ln 2))$; $\theta' = (\ln 2)\theta$ and the representation (see Example 4.4):

$$\frac{\theta^2+1}{\theta+1} = \theta - 1 + \frac{2}{\theta+1}$$

This integrand has the form of equation (7.1) with $a_1 = 1$, $a_0 = -1$, $n_1(\theta) = \theta + 1$, $n_1'(\theta) = (\ln 2)\theta$ and $c_{1,1} = 2$.

We'll start with equation (7.7):

$$D_1 = \frac{c_{1,1}}{n_1(\theta)' - \frac{\mathrm{lc}\, n_1'(\theta)}{\mathrm{lc}\, n_1(\theta)} n_1(\theta)} = \frac{2}{(\ln 2)\theta - \frac{\ln 2}{1}(\theta+1)} = -\frac{2}{\ln 2}$$

Now, equation (7.4) yields:

$$A_1' + (\ln 2)A_1 = 1 \qquad \Longrightarrow \qquad A_1 = \frac{1}{\ln 2}$$

and equation (7.3) yields:

$$A_0' = a_0 - D_1 \frac{\mathrm{lc}\, n_i'(\theta)}{\mathrm{lc}\, n_i(\theta)} = -1 + \frac{2}{\ln 2}\ln 2 = 1$$

$$A_0 = x$$

Plugging everything into equation (7.2), then substituting 2^x for θ, we obtain our solution:

$$\int \frac{4^x+1}{2^x+1}\mathrm{d}x = \frac{1}{\ln 2}\theta + x - \frac{2}{\ln 2}\left[\ln(\theta+1)\right] = \frac{2^x}{\ln 2} + x - \frac{2}{\ln 2}\ln(2^x+1)$$

□

Example 7.8. Redo Example 6.6 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:

```
sage: F.<x,theta> = FractionField(ZZ['x', 'theta']);
sage: Ring.<psi> = F['psi']
```

$$\operatorname{Frac}(\mathbf{Z}[x,\theta])[\psi]$$

```
sage: D = Derivation(Ring, {x: 1, theta: 1/(x+1), psi: 2*x*psi})
```

$$\begin{array}{lll} \multicolumn{3}{l}{\texttt{Derivation of } \operatorname{Frac}(\mathbf{Z}[x,\theta])[\psi]} \\ x & \to & 1 \\ \theta & \to & \frac{1}{x+1} \\ \psi & \to & 2x\psi \end{array}$$

```
sage: num = Ring(lintegrand.numerator(False))
```

$$x^5\psi^4 + \left(-4x^5\theta - 4x^4\theta - 2x^3\theta + 2x^3 - 2x^2\theta\right)\psi^3 - 2x^3\theta^2\psi^2 + x\theta^4$$

```
sage: den = Ring(lintegrand.denominator(False))
```

$$\left(x^5 + x^4\right)\psi^4 + \left(-2x^3\theta^2 - 2x^2\theta^2\right)\psi^2 + x\theta^4 + \theta^4$$

```
sage: (a,N) = num.quo_rem(den)
```

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

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:

```
sage: n = [f[0] for f in factor(den)]
```

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

```
sage: c = partfrac(N, den);
sage: displayarray(c);
```

$$\begin{aligned}
c_{x\psi-\theta,1} &= \frac{2x^3\theta + 2x^2\theta + x\theta - x + \theta}{-x^2 - x}\\
c_{x\psi-\theta,2} &= \frac{-2x^3\theta^2 - 2x^2\theta^2 - x\theta^2 + x\theta - \theta^2}{2x^2 + 2x}\\
c_{x\psi+\theta,1} &= \frac{-2x^3\theta - 2x^2\theta - x\theta + x - \theta}{x^2 + x}\\
c_{x\psi+\theta,2} &= \frac{2x^3\theta^2 + 2x^2\theta^2 + x\theta^2 - x\theta + \theta^2}{2x^2 + 2x}
\end{aligned}$$

Now Theorem 7.1 tells us that

```
sage: R = {};
sage: R[0,1] = c[n[0],2]
```

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

$$C_{0,1}(\psi) = -\frac{R_{0,1}(\psi)}{n_0'(\psi)} \mod n_0(\psi)$$

This time, we do need to perform a modulo reduction, but as modulo reductions go, it's trivial. We construct a quotient ring, map our operands into it, and perform the division there.

```
sage: F2 = Ring.quo(n[0]);
sage: C[0,1] = - F2(R[0,1]) / F2(D(n[0]))
```

$$\frac{\theta}{2}$$

Now we wish to compute $Q_{0,1}(\psi)$. There is no modulo reduction in this division, and it should always be exact. We `lift` the result of the modulo reduction back into its parent ring, and use `//` to perform exact division (as opposed to working in a fraction field).

$$Q_{0,1}(\psi) = -\frac{R_{0,1}(\theta) + C_{0,1}(\theta)n_0'(\theta)}{n_0(\theta)}$$

```
sage: Q = {};
sage: Q[0,1] = - (R[0,1] + C[0,1].lift() * D(n[0])) // n[0]
```

$$\frac{-2x^2\theta - \theta}{2x}$$

Having computed $C_{0,1}$ and $Q_{0,1}$, we are now able to compute D_0:

$$D_0 = \frac{c_{0,1} - C'_{0,1}(\psi) - Q_{0,1}(\psi)}{n'_0(\psi) - \frac{\operatorname{lc} n'_0(\psi)}{\operatorname{lc} n_0(\psi)} n_0(\psi)}$$

```
sage: myD = {};
sage: myD[0] = (c[n[0],1] - D(C[0,1].lift()) - Q[0,1]) \
        / (D(n[0])-D(n[0]).lc()/n[0].lc()*n[0])
```

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

A similar calculation for $n_1(\psi) = x\psi + \theta$ yields yields

```
sage: R[1,1] = c[n[1],2]
```

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

```
sage: F2.<zbar> = Ring.quo(n[1]);
sage: C[1,1] = - F2(R[1,1]) / F2(D(n[1]))
```

$$\frac{-\theta}{-2}$$

```
sage: Q[1,1] = - (R[1,1] + C[1,1].lift() * D(n[1])) // n[1]
```

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

```
sage: myD[1] = (c[n[1],1] - D(C[1,1].lift()) - Q[1,1]) \
        / (D(n[1])-D(n[1]).lc()/n[1].lc()*n[1])
```

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

In an exponential extension, our D coefficients can affect our A_0 term...

$$A_0' = a_0 - \sum_{i=1}^{\nu} D_i \frac{\operatorname{lc} n_1'(\psi)}{\operatorname{lc} n_1(\psi)}$$

```
sage: a = a - sum([myD[i]*D(n[i]).lc()/n[i].lc() for i in range(2)])
```

$$\frac{x}{x+1}$$

```
sage: A = integrate(a,x)
```

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

We end up with the same result that we obtained from the logarithmic theory:

```
sage: lans = A + sum([C[i,1].lift()/n[i] for i in range(2)]) \
        + sum([2 * myD[i] * log(n[i]) for i in range(2)]).simplify_log(
```

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

```
sage: ans = lans.subs({theta : log(x+1), psi : exp(x^2)})
```

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

```
sage: bool(diff(ans,x) == integrand)
```

True

□

7.3 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:

$$r' + Sr = T \qquad S, T, r \in K(\theta) \tag{7.10}$$

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 $n(\theta)$:

$$R_1(\theta) = -ja(\theta)n'(\theta) \mod n(\theta)$$
$$R_2(\theta) = a(\theta)b(\theta) \mod n(\theta)$$

$$\frac{R_1(\theta)}{n(\theta)^{j+1}} + \cdots + \frac{R_2(\theta)}{n(\theta)^{j+k}} + \cdots = \frac{c(\theta)}{n(\theta)^l} + \cdots$$

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 $j = l - 1$. (7.11a)
2. $k = 1$ and either $j = l - 1$ or $j = \frac{b(\theta)}{n'(\theta)} \mod n(\theta)$. (7.11b)
3. $k > 1$ and $j = l - k$. (7.11c)

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 of the form:

$$Ap' + Bp = C \qquad A, B, C, p \in K[\theta] \tag{7.12}$$

In addition to the three cases that we considered for $\mathbb{C}[x]$, there is a fourth case that must be considered. p can be zeroth order in θ, yet not be a constant. So our cases become:

1. $\deg_\theta A = \deg_\theta B + 1$ and $\deg_\theta p = -\frac{\operatorname{lc} B}{\operatorname{lc} A}$
2. $\deg_\theta p = \deg_\theta C - \max(\deg_\theta A - 1, \deg_\theta B)$
3. $\deg_\theta p = 0$
4. p is a constant and $pB = C$

Example 7.9. 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 $\mathbb{C}(x, \theta, \psi)$, where $\theta = \ln x$ and $\psi = \exp x\theta$. So, we're trying to compute

$$\int \psi \, dx$$

Applying theorem 7.1, we obtain the following Risch equation in $\mathbb{C}(x, \theta)$:

$$A_1' + k' A_1 = 1$$

$$A_1' + (\theta + 1) A_1 = 1 \tag{7.13}$$

$$A = 1 \qquad B = \theta + 1 \qquad C = 1$$

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.

If $\deg_\theta A_1$ is zero, then $A_1 \in \mathbb{C}(x)$ and equation (7.13) splits into two Risch equations in $\mathbb{C}(x)$, one for first degree terms in θ and one for zeroth degree terms in θ:

$$\begin{aligned} A_1 &= 0 \\ A_1' + A_1 = 1 \qquad &\Longrightarrow \qquad A_1 = -1 \end{aligned}$$

These equations have no simultaneous solution.

We conclude that equation (7.13) has no solution in $\mathbb{C}(x, \theta)$, and that the original integral is not elementary.

□

Example 7.10. Integrate $\int \log(x) e^{-x^2} dx$

It's noted in the Sage documentation that Sage can't simplify this integral, though both Maple and Mathematica can (`calculus/tests.html`). □

7.4 Normal Risch Equations in Sage

To facilitate use of Sage, I've written a Risch equation solver in Sage.

First, we want a partial fractions expansion that's a little bit different from the standard function `partfrac`. If *expr* isn't a fraction, or if its denominator doesn't involve the variable *var*, we want an empty list returned, otherwise return a list of the fractions in the expansion. We just want the highest-powered term in the expansion.

```
def partfrac1(num, den):
    b = {}
    if den != 1:
      factorization = factor(den)
      for f in factorization:
        (t,s) = diophantine(num, f[0]^f[1], den//(f[0]^f[1]))
        (q,r) = s.quo_rem(f[0])
        b[f[0]] = [r, f[1]]
    return(b)
```

Next, given a Risch equation in the form

$$r' + Sr = T$$

`normal_risch_factors` returns a list of sublists, with one sublist for each irreducible factor in S or T's denominator. Each sublist contains the factor along with a pair of lists, one for S and one for T, containing the highest power in that variable's partial fractions expansion, along with the corresponding numerator, in the following format:

`[factor, [S-numerator, S-power], [T-numerator, T-power]]`

```
def normal_risch_factors(S, T):
    result = {}
    Sb = partfrac1(S.numerator(), S.denominator())
    Tb = partfrac1(T.numerator(), T.denominator())
    for f in set(Sb.keys()).union(Tb.keys()):
       result[f] = [Sb.get(f, [0,1]), Tb.get(f, [0,1])]
    return result
```

`normal_risch_denominator` uses `normal_risch_factors` to compute the solutions's denominator.

```
def is_integer(P):
    try:
       int(P)
       return True
    except (TypeError, ValueError):
      return False

def normal_risch_denominator1(L, D):
   (f, [[Snum, k], [Tnum, l]]) = L
   if k == 0:
      return f^max(0,l-1)
   elif k == 1:
      R = f.parent().quo(f)
      j = R(Snum)/R(D(f))
      if is_integer(j):
         return f^max(l-1, j, 0)
      else:
         return f^max(l-1, 0)
   else:
      return f^max(0,l-k)

def normal_risch_denominator(S, T, D):
  return prod(map(lambda L: normal_risch_denominator1(L, D), normal_
```

The next function, `normal_risch_polynomial`, given S, T, and the denominator computed by `normal_risch_denominator`, returns the components of the polynomial equation satisfied by the solution's numerator:

$$An' + Bn = C$$

in the form `[A, B, C]`.

```
def normal_risch_polynomial(S, T, D1, D):
  D2 = (S*D1).denominator() * (T*D1).denominator()
  # XXX check to make sure that denominator is 1
  return [D1*D2, S*D1*D2 - D(D1)*D2, (T*D1^2*D2).numerator()]
```

Next, given A, B, and C, we wish to compute the maximum degree of the numerator, or `None` if we can determine at this point that there is no solution.

```
def is_constant(P, D):
   return (D(P) == 0)

def normal_risch_degree(A, B, C, D):

   # XXX What if B is zero?

   if C != 0:
      k = C.degree() - max(A.degree() - 1, B.degree())
   else:
      k = -1
   if is_constant(C/B, D):
      k = max(k,0)

   if A.degree() == B.degree() + 1:
      p = - B.lc() / A.lc()
      if is_integer(p):
         k = max(k,p)

   return k
```

Finally, we combine all these pieces together into a function that either returns a solution to a normal Risch equation, or `None` if no such solution exists.

```
def normal_risch_equation(S, T, D):
  D1 = normal_risch_denominator(S, T, D)
  [A, B, C] = normal_risch_polynomial(S, T, D1, D)

  if is_constant(C/B, D):
     return C/B/D1
  elif normal_risch_degree(A,B,C,D) == -1:
     return None
  else:
     raise NotImplementedError
```

Let's demonstrate the use of this code by using it to solve the Risch equations that we've already seen in this chapter's examples so far.

```
sage: R.<x> = QQbar[]
```

$$\overline{\mathbf{Q}}[x]$$

```
sage: I = R(sqrt(R(-1)))
```

$$i$$

```
sage: D = Derivation(R, {x: 1})
```

Derivation of $\overline{\mathbf{Q}}[x]$
$x \rightarrow 1$

```
sage: normal_risch_equation(I, 1/(2*I), D)
```

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

```
sage: normal_risch_equation(-2*x, R(1), D)
```

None

```
sage: normal_risch_equation(I, 1/x, D)
```

None

```
sage: C2.<log2> = QQbar[]
```

$$\overline{\mathbf{Q}}[log_2]$$

```
sage: C2._latex_names[0] = '{\\ln 2}'
```

```
{\ln 2}
```

```
sage: R2.<x> = C2[]
```

$$\overline{\mathbf{Q}}[\ln 2][x]$$

```
sage: D = Derivation(R2, {x: 1, log2: 0})
```

$$\begin{array}{lcl} \multicolumn{3}{l}{\texttt{Derivation of } \overline{\mathbf{Q}}[\ln 2][x]} \\ x & \to & 1 \\ \ln 2 & \to & 0 \end{array}$$

```
sage: normal_risch_equation(R2(log2), R2(1), D)
```

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

```
sage: R2.<theta> = R[]
```

$$\overline{\mathbf{Q}}[x][\theta]$$

```
sage: D = Derivation(R2, {x: 1, theta: 1/x})
```

$$\begin{array}{lcl} \multicolumn{3}{l}{\texttt{Derivation of } \overline{\mathbf{Q}}[x][\theta]} \\ x & \to & 1 \\ \theta & \to & \frac{1}{x} \end{array}$$

```
sage: normal_risch_equation(theta+1, R2(1), theta)
```

None

7.5 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 $\theta' = k'\theta$.

$$r' + Sr = T \qquad S, T, r \in K(\theta) \tag{7.14}$$

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 $j = \frac{a'+ab}{ak'}$
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:

$$Ar' + Br = C \qquad A, B, C \in K[\theta] \quad r \in K(\theta) \tag{7.15}$$

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:

$$[-jak' + a']\, A(0) + aB(0) = 0$$

$$jk' - \frac{a'}{a} = \frac{B(0)}{A(0)}$$

Integrating, we obtain:

$$jk - \ln a = \int \frac{B(0)}{A(0)} dx \tag{7.16}$$

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:

$$Ap' + Bp = C \qquad A, B, C, p \in K[\theta] \tag{7.17}$$

$$p = p_n\theta^n + \cdots \qquad p' = (p_n' + np_nk')\theta^n + \cdots$$

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 7.16, so again, we integrate:

$$nk + \ln p_n = -\int \frac{\operatorname{lc} B}{\operatorname{lc} A} dx \tag{7.18}$$

If this integral has this desired form, then cancellation is possible between the terms on the left hand side of 7.17, and n is the θ-degree of the solution polynomial.

Otherwise, there is no cancellation and $\deg_\theta p = \deg_\theta C - \max(\deg_\theta A, \deg_\theta B)$

This degree can be negative, so long as it is no lower than the lower degree bound determined earlier.

Example 7.11. ([Br05] examples 6.2.1, 6.3.3, 6.4.2) Integrate

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

We'll use the differential field $\mathbb{C}(x, \theta, \psi)$ where $\theta = \exp x$ and $\psi = \exp\left(\frac{x^2-1}{x} + \frac{1}{\theta+x}\right)$.

```
sage: var('x', 'theta', 'psi');

sage: integrand =                                 \
         (exp(x) - x^2 + 2*x)                     \
            / ((exp(x) + x)^2 * x^2)              \
         * exp((x^2-1)/x + 1/(exp(x)+x))
```

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

```
sage: exponent = (x^2-1)/x + 1/(theta+x)
```

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

```
sage: lintegrand = integrand.subs(
         {exp(x)          : theta,
          exp(exponent)  : psi})
```

$$-\frac{(x^2 - \theta - 2\,x)\psi}{(\theta + x)^2 x^2}$$

```
sage: F.<x,theta> = FractionField(ZZ['x', 'theta']);

sage: R.<psi> = F[]
```

$$\mathrm{Frac}(\mathbf{Z}[x, \theta])[\psi]$$

```
sage: D1 = Derivation(R, {x: 1, theta: theta})
```

$$\begin{array}{l} \texttt{Derivation of } \mathrm{Frac}(\mathbf{Z}[x, \theta])[\psi] \\ x \ \rightarrow \ 1 \\ \theta \ \rightarrow \ \theta \end{array}$$

```
sage: D = Derivation(R, {x: 1, theta: theta, psi: D1(R(exponent))*psi})
```

$$\begin{array}{lcl} \texttt{Derivation of } \mathrm{Frac}(\mathbf{Z}[x,\theta])[\psi] \\ x & \to & 1 \\ \theta & \to & \theta \\ \psi & \to & \left(\frac{x^4+2x^3\theta+x^2\theta^2-x^2\theta+2x\theta+\theta^2}{x^4+2x^3\theta+x^2\theta^2}\right)\psi \end{array}$$

The integrand has the form $a_1\psi$, where $a_1 \in \mathbb{C}(x,\theta)$, so we're ready to apply Theorem 7.1. Equation (7.4) gives:

```
sage: a1 = lintegrand/psi
```

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

```
sage: A1f = function('A1', nargs=1);
```

```
sage: A1 = A1f(x)
```

$$A_1\left(x\right)$$

```
sage: eq = diff(A1,x) + D(psi)/psi * A1 - a1
```

$$\frac{(\theta^2x^2+2\,\theta x^3+x^4-\theta x^2+\theta^2+2\,\theta x)A_1\left(x\right)}{\theta^2x^2+2\,\theta x^3+x^4}+\frac{x^2-\theta-2\,x}{\left(\theta+x\right)^2x^2}+\frac{\partial}{\partial x}A_1\left(x\right)$$

This is already in the form required by equation 7.14:

$$r' + Sr = T \qquad S, T, r \in K(\theta)$$

with S and T expanded in partial fractions. None of the denominator factors are special (only factors of the form θ^n are special), and both $(\theta + x)$ and x are normal polynomials with $k = 2$ (the multiplicity in S's denominator) and $l = 2$ (the multiplicity in T's denominator). Case 3 in 7.11 tells us that $j = l - k = 0$, so neither of these (normal) factors can appear in the denominator of r, and r's denominator can only involve special factors.

Next we can move on to equation 7.15:

$$Ar' + Br = C \qquad A, B, C \in K[\theta] \quad r \in K(\theta)$$

Let's use `numerator()` to collapse the entire equation into a single fraction and discard the denominator, and then extract the coefficients of A_1 and A_1'.

```
sage: [BC, _], [A, _] = \
         eq.numerator().coefficients(diff(A1,x));
sage: [C, _], [B, _] = \
         BC.numerator().coefficients(A1);
sage: A
```

$$\theta^2 x^2 + 2\,\theta x^3 + x^4$$

```
sage: B
```

$$\theta^2 x^2 + 2\,\theta x^3 + x^4 - \theta x^2 + \theta^2 + 2\,\theta x$$

```
sage: C
```

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

Now we want to integrate $A(0)/B(0)$, to determine the highest power of θ that can appear in the denominator:

```
sage: (A/B).subs(theta=0)
```

$$1$$

Since $\theta = e^x$, $k = x$, this solves equation 7.16 with $j = 1$ and a constant, which tells us that θ can appear in r's denominator with multiplicity one.

What about the polynomial degree?

```
sage: R2 = ZZ['x']['theta']
```

$$\mathbf{Z}[x][\theta]$$

```
sage: R2(B).lc() / R2(A).lc()
```

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

```
sage: integrate(R2(B).lc() / R2(A).lc(), x)
```

$$x - \frac{1}{x}$$

This does not have the desired form of 7.18, specifically it is not an integer multiple of x plus a logarithm, so we lack cancellation. This means that A_1's θ-degree is upper bounded by $\deg_\theta C - \max(\deg_\theta A, \deg_\theta B) = -1$. Since the lower bound is also -1, this gives us the following form for A_1:

$$A_1 = A_{1,-1}\theta^{-1} \qquad A_{1,-1} \in \mathbb{C}(x)$$

Sage doesn't know how `theta` differentiates[1], so I expanded it out by hand:

```
sage: A1m1f = function('A1m1', latex_name='A_{1,-1}', nargs=1);
sage: A1m1 = A1m1f(x);
sage: eq2 = eq.subs({A1 : A1m1/theta,
        diff(A1,x) : (diff(A1m1,x)*theta - A1m1*D(theta))/theta^2})
```

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

```
sage: eq2.numerator()
```

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

This can also be done with more sophisticated tools. We create a differential ring involving A_1, $A_{1,-1}$ and θ, with x as its sole derivation. An elimination ordering is used (the default), with A_1 appearing earlier in the list than $A_{1,-1}$, thus eliminating A_1 in favor of $A_{1,-1}$. Then we construct a differential ideal showing how θ differentiates and how A_1 is related to $A_{1,-1}$. Finally, we reduce our equation modulo the single differential chain in the differential ideal, producing a normal form with the desired result.[2]

$A_{1,-1}$ is a rational functions in $\mathbb{C}(x)$. In particular, it does not involve θ, so each power of θ in this equation must be zero. Let's extract the leading θ coefficient and set it equal to zero.

This Risch equation has an obvious solution. Now we're solving in $\mathbb{C}(x)$, so if it were difficult, we could use the theory of the previous section:

```
sage: R2.<x> = QQbar[];
sage: D = Derivation(R2, {x: 1});
```

[1] Maxima is somewhat better in this regard. It's `gradef` command lets the user declare how variables differentiate.

[2] This is also somewhat painful with Sage, requiring us to convert θ into a "function" so that we can take its derivative.

```
sage: normal_risch_equation(1/x^2, 1/x^2, D)
```

$$1$$

Having solved for $A_{1,-1}$, we've now solved for A_1:

```
sage: A1 = R(1/theta)
```

$$\frac{1}{\theta}$$

Since our integrand was of the form $a_1\psi$, our integral is of the form $A_1\psi$, so we've completed our calculation. Is it correct?

```
sage: lans = psi/theta
```

$$\frac{1}{\theta}\psi$$

```
sage: ans = lans.subs({psi : exp(exponent), theta : exp(x)})
```

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

```
sage: bool(diff(ans,x) == integrand)
```

True

□

Chapter 8

Algebraic Curves

THIS CHAPTER IS INCOMPLETE.

Having addressed logarithmic and exponential extensions, we now turn to the algebraic extension, which at first appears to be completely different in character from the two transcendental cases. The differences stem largely from the lack of unique factorization in the algebraic case; algebraic extensions are not, in general, unique factorization domains.

There are four basic operations we perform on a rational function in order to integrate it:

1. We **factor** its numerator and denominator
$$\frac{x^3+x^2-5x+3}{x^2-5x+6} = \frac{(x+3)\cdot(x-1)^2}{(x-3)\cdot(x-2)}$$
From the factorization, we can read off the locations of the function's zeros and poles. In this example, our zeros are at -3 and 1 (multiplicity 2), and our poles are at 2 and 3.

2. We construct a **partial fractions expansion**
$$\begin{aligned}\frac{x^3+x^2-5x+3}{x^2-5x+6} &= \quad x+6 \quad + \frac{24}{x-3} - \frac{5}{x-2} \\ &= \frac{1}{1/x} + 6 + \frac{24}{x-3} - \frac{5}{x-2}\end{aligned}$$

3. We reconstruct a function given a factorization of its numerator and denominator (or equivalently, a list of poles and zeros along with their multiplicities)

4. We reconstruct a function given a partial fractions expansion (or equivalently, a set of principal parts expansions at the function's poles)

In this chapter, we'll develop a basic set of technical tools for working in the simplest kind of algebraic extension, an extension of $\mathrm{C}(x)$. This will prepare us for the next chapter, where we'll study *Abelian integrals*, which are integrals whose integrands are formed from polynomials and roots of polynomials. In other words, integrands in an algebraic extension of $\mathbf{C}(x)$.

How might we handle an algebraic extension of $\mathbf{C}(x)$? 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 an algebraic 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 such *modern algebraic geometry*, however, let's see how far we can get with the classical algebraic geometry of the nineteenth century.

[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.

8.1 The Topology of an Algebraic Curve

Recall that the graph of a bivariate polynomial, $\sum a_{ij}x^iy^j = 0$, is called an *algebraic curve*, and will be our main focus of attention in this chapter.

One of the first problems we face when dealing with algebraic curves is the multi-valued nature of their solutions. Consider the algebraic algebraic curve $y^2 = x^2 - 1$. For almost any given x, there are two seperate y values that solve this equation. 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 appears multiple times in the curve's defining polynomial, something like $y^3+x^2y^2-x+4y = 0$?

Our solution to this problem is to regard the entire algebraic curve as a two-dimensional surface in a four-dimensional x-y space. Why four dimensions? Well, just as in the univariate case, we find it convenient to work with complex numbers because of their property of algebraic closure. Most algebraic geometers talk about dimension with respect to the coefficient field, so we can consistently say that an algebraic curve is a one dimensional curve in two dimensional space. 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.

Once we have defined an algebraic curve, we will want to consider *algebraic functions* on the curve, which are simply rational functions (ratios of polynomials) in the two variables x and y. We will only be interested the values of these functions on the curve; the curve acts as a restriction on the possible values of x and y. If this seems at all puzzling, consider an integrand involving both x and $\sqrt{x^2+1}$. We'll define $y^2 = x^2 + 1$ and replace all instances of the square root with y. Then we'll have an integrand involving both x and y (the algebraic function), but y is always the square root of $x^2 + 1$, so it makes no sense to consider values of the algebraic function where $y^2 = x^2 + 1$ is not satisfied.

Given an algebraic curve and an algebraic function, we'd like to construct a series expansion of the algebraic function at each point of the curve. How can we construct series expansions of rational functions that involve both x and y? At each point of the curve, we seek to find a single *uniformizing variable* that is suitable for constructing power series expansions that converge in an open neighborhood of the point. This is also called a *local uniformizer*, since no one function is a suitable uniformizing variable at all points of an algebraic curve.

Do local uniformizing variables always exist at every point of an algebraic curve?

The answer is a qualified "yes". At most points on an algebraic curve, the answer is an unqualified "yes" as a result of the Implicit Function Theorem:

Theorem 8.1. Implicit Function Theorem

The two-dimensional complex analytic version:

Let $f(x,y)$ be an analytic mapping of an open set $E \in \mathbb{C}^2$ into $\mathbb{C}$, such that $f(a,b) = 0$ and $\frac{df}{dy}(a,b) \neq 0$, then open sets U and V exist in $\mathbb{C}$ such that $a \in U$, $b \in V$, and an analytic function $g(x)$ exists that maps U into V such that $f(x, g(x)) = 0$.

What does it mean for a $\mathbb{C}^2 \to \mathbb{C}$ function to be analytic?

[Ru76] Theorem 9.28 is a real version of the theorem that is based on the inverse function theorem, but only establishes that the implicit function is continuously differentiable, whereas we want to show that it's analytic. Of course, for a complex version of this proof, differentiable would immediately imply analytic.

The Wikipedia article on the implicit function theorem includes a proof for the two-dimensional case (they only case we care about here) that derives a differential equation that the implicit function must satisfy and uses the Picard-Lindelöf theorem to show its existence, but this also only establishes continuous differentiability.

[Gu05] Lecture 7 starts with a complex version of the theorem.

See `https://math.stackexchange.com/questions/489789`

`https://math.stackexchange.com/a/289640/71520` *discusses the difficulty of moving along the transition $R \to R^n \to C \to C^n$.*

□

This theorem applies everywhere on the curve that $\frac{df}{dy} \neq 0$, which is everywhere except a finite number of points. Where does it fail? Those points at which both $f(x,y) = 0$ and $\frac{df}{dy}$ equals zero. In other words, points at which the defining function and its derivative with respect to one of its variables share a zero. At these points, the curve exhibits a behavior called *ramification*.

If the derivative with respect to one variable is zero, could we try to apply the theorem using the derivative with respect to the other variable? The answer is often yes, but not always. Ramification thus occurs with respect to a specific variable, more generally with respect to a specific mapping.

On some curves, however, there are points, called *singularities*, at which the derivative with respect to both variables is zero. One might hope to pick a uniformizing variable different from either coordinate variable that would allow the Implicit Function Theorem to be applied, but this turns out to be impossible. The derivative with respect to any algebraic function could be expanded using partial derivatives with respect to the coordinate variables, and the resulting derivative would necessarily be zero:

$$\frac{df}{dz} = \frac{df}{dx}\frac{dx}{dz} + \frac{df}{dy}\frac{dy}{dz} = 0\frac{dx}{dz} + 0\frac{dy}{dz} = 0$$

So there is no variable at a singularity that meets the requirements of the Implicit Function Theorem. Something more is required.

Since both f and $\frac{df}{dy}$ share a zero at a given value of x, f's zero must be at least second order, so singularities are necessarily multiple points of the curve.

Consider a specific multiple point where n sheets of the curve meet at a single point P. A small circle at distance ϵ from P will map to n values of the curve. Without loss of generality, let's assume that P is the origin.

[Silverman] Theorem 10.13: If a function is analytic on every point of a open disk $K : |z-a|< R$, then it has a power series expansion centered on a that converges everywhere in K.

By this theorem, we see that we can surround P with disks of radius ϵ; a finite number of them will circle around P. These disks don't have to cover every point in a neighborhood of P; they need only circle P so that we cleanly identify a permutation of the sheets.

Label the r values of y at $(x+\epsilon)$ as $y_1(x), \ldots, y_n(x)$. As we trace along the circle defined by $x = \epsilon e^{i\theta}$ these values deform continuously as θ goes from 0 to 2π. Once we reach 2π, we have come full circle and the y values match up, but permuted. Call the permutation σ, so $y_1(x), \ldots, y_n(x)$ map to $y_{\sigma 1}(x), \ldots, y_{\sigma n}(x)$. The permutation σ of the n sheets can decomposed into k disjoint cycles. We seek to show that a Riemann surface can be obtained by replacing the singular point with k distinct points.

Consider a single cycle of length r, the *ramification index*, and an open disk in the t-plane, $\Delta = \{t : |t|< \epsilon\}$. We wish to exhibit an bihomomorphism from Δ to the cycle. Set $x(t) = t^r$, which is clearly holomorphic in Δ (in fact, everywhere on the t-plane). The function

$$y(t) = y_{\sigma^{\lfloor (r \ \arg t)/(2\pi) \rfloor}}(t^r)$$

where $\arg t$ is the complex argument, with range $[0, 2\pi)$, $\lfloor \cdot \rfloor$ is the integer floor function, and powering σ by an integer applies the permutation that many times. The permutation ensures continuity of the function at each transition between the various y_n functions. $y(t)$ is obviously holomorphic away from the origin, but is it also holomorphic at the origin? Yes, according to Riemann's theorem on removable singularities.

Riemann removable singularity theorem

Theorem 8.2. *Let $D \subset \mathbb{C}$ be an open subset of the complex plane, $a \in D$ a point of D and f a holomorphic function defined on the set $D \setminus \{a\}$. The following are equivalent:*

1. *f is holomorphically extendable over a.*
2. *f is continuously extendable over a.*
3. *There exists a neighborhood of a on which f is bounded.*
4. $\lim_{z \to a}(z-a)f(z) = 0$.

Proof

The implications $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$ are trivial. To prove $4 \Rightarrow 1$, we first recall that the holomorphy of a function at a is equivalent to it being analytic at a, i.e. having a power series representation. Define

$$h(z) = \begin{cases} (z-a)^2 f(z) & z \neq a, \\ 0 & z = a. \end{cases}$$

Clearly, h is holomorphic on $D\backslash\{a\}$, and there exists $h'(a) = \lim_{z\to a} \frac{(z-a)^2 f(z) - 0}{z-a} = \lim_{z\to a}(z-a)f(z) = 0$ by 4, hence h is holomorphic on D and has a Taylor series about a:

$$h(z) = c_0 + c_1(z-a) + c_2(z-a)^2 + c_3(z-a)^3 + \cdots .$$

We have $c_0 = h(a) = 0$ and $c_1 = h'(a) = 0$; therefore

$$h(z) = c_2(z-a)^2 + c_3(z-a)^3 + \cdots .$$

Hence, where $z \neq a$, we have:

$$f(z) = \frac{h(z)}{(z-a)^2} = c_2 + c_3(z-a) + \cdots .$$

However,

$$g(z) = c_2 + c_3(z-a) + \cdots .$$

is holomorphic on D, thus an extension of f.

□

Source: copied verbatim from Wikipedia

To see that function $y(t)$ is bounded on Δ, pick a δ so that $f(0,\delta)$ is G, then ensure that ϵ is small enough to ensure that $f(x,\delta) \neq 0 \forall x \in \Delta_\epsilon$. Since $y(t)$ is continuous on Δ_ϵ, if it were not bounded, then $f(t^r, y(t)) = \delta$ for some value t, contridicting our assumption. Thus, $y(t)$ is bounded on Δ, so it can be holomorphically extended to the origin by the previous theorem.

Lemma. Given a polynomial $f(x,y)$, then at any given point (x_0, y_0) and any given real number $\delta > 0$, there exists a real number ϵ such that

$$f(x,y) \neq f(x_0,y_0) \quad \forall x,y; |x-x_0|< \epsilon; |y-y_0|= \delta$$

Proof. Consider $g(y) = f(x_0,y) - f(x_0,y_0)$, a polynomial in y with a zero at y_0. Pick a number a such that no other zero is within $\pm a$ of y_0. Consider the complex circle of radius a centered at y_0. $g(y)$ on this circle must have a minimum value that is not zero; call this minimum value m. Now, at any point $y_0 + a$, $f(x_0, y_0 + a)$ by **complex** continuity will have a value ϵ such that x_0 can be varied up to ϵ without changing the function by more than m. But do all of these values have a minimum greater than zero?

In short, to obtain a complex manifold, we need to modify our curve slightly by adding additional points at singularities. Theorem 8.2 tells us that no additional information is need to specify the behavior of holomorphic functions at those additional points – their behavior at an isolated point is completely determined by their behavior in an open neighborhood surrounding that point.

What about meromorphic functions? They can just be promoted to holomorphic functions by multiplying them by a suitable power of the uniformizing variable (t, in the above treatment), and Theorem 8.2 again tells us that their behavior is completely specified.

Interestingly enough, a rational function can have different values at the same points over a singularity.

Example 8.3. Example: $y^2 = x^3 + x^2$ has a singularity at the origin, since $f(x,y) = y^2 - x^3 - x^2$, $\frac{df}{dx} = -3x^2 - 2x$, $\frac{df}{dy} = 2y$, and $\frac{df}{dx}(0,0) = 0$ and $\frac{df}{dy}(0,0) = 0$.

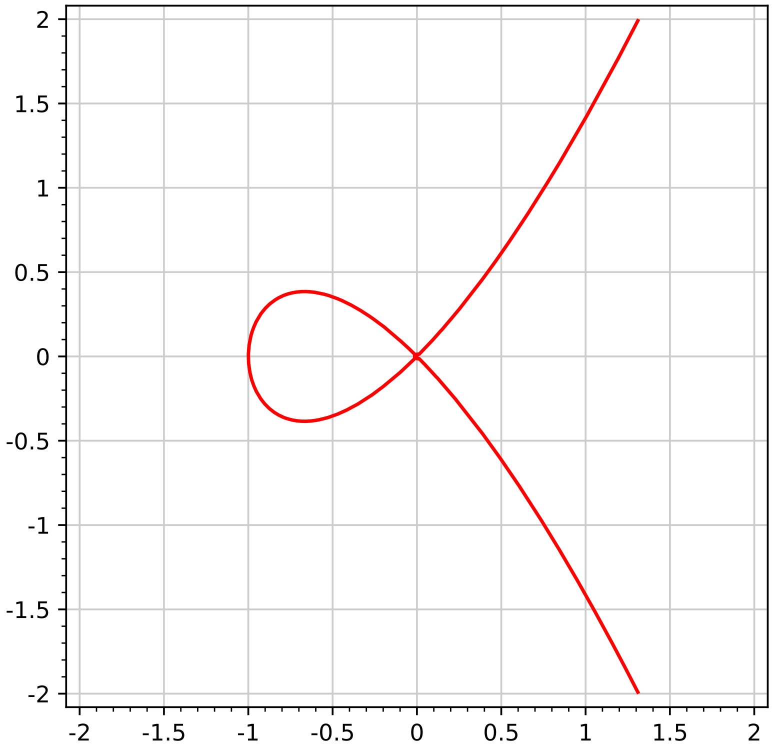


Figure 8.1: $y^2 = x^3 + x^2$

The function y/x has the value 1 on one branch and -1 on the other. It is also possible, straightforward even, to construct functions with a zero on only one branch ($y/x - 1$) or the other ($y/x + 1$), or a pole on only one branch ($x/(y-x)$), or a pole on one branch and zero on the other ($x/(y-x) + 1/2$). □

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. (PROOF)

Resolving our singularities in this manner creates a complex manifold, but it lacks a crucial property: topological compactness. In order to apply Theorem 8.5, we need a *compact* manifold. We fix this problem by embedding our algebraic curve in projective space, using a standard construction.

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.

Definition 8.4. *A topological space X is compact if any open covering of X admits a finite subset that covers X.*

There are several crucial theorems that depend on the topological property of *compactness*. The complex plane is not compact; we remedy this by adding a point at infinity to obtain the *Riemann sphere*. Likewise, two-dimensional complex space is not compact, either; we remedy this by adding a line at infinity and obtaining *projective space*.

Theorem: Projective space is compact

Munkres Theorem 26.2: Every closed subspace of a compact space is compact.

Theorem: The Zariski topology is the coursest topology in which singletons are closed.

Varieties are closed in the Zariski topology.

Standard topology is finer than Zariksi, so all open sets in Zariski are open in standard, and all Zariki-closed sets are closed in standard.

So, varieties are closed in the standard topology, and are therefore compact in projective space.

Projective space. Compactness.

Another highly desirable property is to be locally isomorphic to Euclidean space. A differentiable surface that is everywhere locally Euclidean is called a *manifold*.

By adding a line at infinity and resolving our singularities, we can coax our algebraic curve into a compact, connected, complex manifold. The primary utility of this construction is embodied in the following theorems.

Regular functions. Holomorphic functions. Identical on complex projective varieties.

Theorem 8.5. *Every holomorphic function $M \to C$ on a compact, connected, complex manifold M is constant.* [2]

Proof

`https://math.stackexchange.com/questions/881742`

[Gu05] Lecture 2 contains a proof of the Maximum Modulus Priciple.

□

Definition 8.6. *The principal part of an algebraic function at a pole is the portion of its corresponding Laurent series with negative exponents.*

Theorem 8.7. *An algebraic function on an algebraic curve is completely characterized, up to an additive constant, by its principal parts.*

Proof

[2] In fact, we don't need the complex structure and can make the stronger statement that the only regular functions on any projective variety are the constants. See Proposition 4.2 in `https://www.math.utah.edu/ bertram/6030/Projective.pdf`, or Hartshorne Theorem (I, 3.4a), or `https://math.stackexchange.com/questions/56236`. If the base field is not algebraically closed, however, new constants may appear as algebraic functions. For example, the polynomial $(x + y)^2 + 1$ is irreducible in the ring $\mathbf{Q}[x, y]$, so we can use it to construct an algebraic curve, but $x + y$ is a constant on this curve, a square root of -1, in fact.

Consider two algebraic functions f and g with identical principal parts. Taking the difference between them, we obtain a function $f - g$ with no principal parts, i.e, a holomorphic function. By Theorem 8.5, $f - g$ must be constant.

□

Theorem 8.8. *An algebraic function on an algebraic curve is completely characterized, up to a multiplicative constant, by its divisor.*

Proof

Consider two algebraic functions f and g with identical divisors. Dividing f/g we obtain a function with no poles (or zeros), a holomorphic function. By Theorem 8.5, f/g must be constant.

□

Related: Hartshorne Corollary 6.10. A principle divisor on a complete nonsingular curve has degree zero.

Given the importance of an algebraic function's principal parts, we will now develop tools to calculate them.

8.2 Puiseux Expansions

The previous section showed that, in complex projective space, a covering surface can be constructed for an algebraic curve, which is isomorphic to the curve except at singularities (where additional points must be introduced), such that:

- the covering surface is a complex manifold (a *Riemann surface*),
- the curve's rational functions are meromorphic functions on that manifold, and
- the rational functions admit series expansions at every point of the manifold.

Our next task is to compute those series expansions.

Since the rational functions in the curve's function field are formed as rational functions in x and y (or whatever our variables are named), our primary goal is to compute series expansions for x and y at arbitrary points on the curve. With such expansions in hand, it is straightforward to construct expansions for any algebraic function, simply by substituting in the x and y expansions.

At any point where the discriminant is non-zero and y obtains a finite value, a series expansion for y exists as a power series in $(x - \alpha)$, which, as we have seen, is a straightforward application of the Implicit Function Theorem. At these *ordinary* points, we need only postulate a Taylor series for y in powers of $(x - \alpha)$, substitute this into the curve's defining polynomial, and equate like powers to obtain a set of equations to be solved simultaneously. Multiple solutions will typically be found, corresponding to multiple branches of the curve.

At ramification points, the series expansion exists in terms of fractional powers of $(x-\alpha)$, where the denominator of the fractions is the ramification index. Issac Newton, in 1676, first proposed a method of computing the ramification index using what are now called *Newton polygons*.

Let's assume that we're expanding around the point $(0, 0)$, as this simplifies the analysis with no loss of generality. Consider factoring the defining polynomial of the algebraic curve:

$$p_n y^n + p_{n-1} y^{n-1} + \cdots + p_0 = (y - r_1)(y - r_2) \cdots (y - r_n)$$

How might we do this, if the polynomial is irreducible? We need to extend to a larger field where the polynomial's roots exist. The analysis above shows that Puiseux series form a suitable extension.

For each root r_i, define its *order* as the lowest power of t that appears in its Puiseux expansion, divided by its ramification index. Multiplying factors together adds their orders, so p_0's order will be the sum of all of the r_i's orders.

Now let's consider increasing i by one. How does p_0's order change? p_1 is formed by adding together all products of $n-1$ roots, so p_1's order will be lower than p_0's order by the largest of r_i's orders, unless there are multiple r_i's with the same order. In this case, cancellation between these multiple terms could result in p_1 having a larger order than otherwise expected.

If there are j r_i's with the same largest order, increasing i by j will lower p_i's order by j times that largest order.

The Newton polygon is formed by plotting the orders of the p_i coefficients, with i varying along the horizonal axis and the order plotted vertically. The easiest way to do this is to plot the powers of the monomials that appear in the equation, and construct the polygon's lower convex hull.

Thus, a segment on the lower convex hull of the Newton polygon will correspond to as many solutions as the width of the line segment, each with order equal to the change in height divided by the width, i.e, the negative slope of the line segment. The denominator of the slope will be the ramification index, and the numerator of the slope will be the lowest exponent expected in the expansion of y.

Consider a Puiseux series corresponding to a single line segment of the Newton polygon. Letting α be the x exponent and β be the y exponent, so the monomials in f have the form $x^\alpha y^\beta$, then the equation of the line segment is $r\alpha + s\beta = p$, where r, s, and p are integers and r and s are relatively prime. Making the substitution $x = t^r$ and $y = t^s u(t)$, we obtain:

$$\begin{aligned} f(x,y) &= \sum A_{\alpha\beta} x^\alpha y^\beta \\ &= \sum A_{\alpha\beta} t^{r\alpha} t^{s\beta} u(t)^\beta \\ &= t^p \underbrace{\sum A_{\alpha\beta} t^{r\alpha+s\beta-p} u(t)^\beta}_{g(t,u)} \end{aligned}$$

At least two of the $(r\alpha+s\beta-p)$ exponents will be zero (those monomials corresponding to the endpoints of the line segment on the Newton polygon); all of the remaining exponents will be positive. This means that if we expand $u(t)$ in a power series in t:

$$u(t) = u_0 + u_1 t + u_2 t^2 + u_3 t^3 + \cdots$$

then any power of $u(t)$ will have the form:

$$u(t)^\beta = U_0(u_0) + U_1(u_0,u_1)t + U_2(u_0,u_1,u_2)t^2 + U_3(u_0,u_1,u_2,u_3)t^3 + \cdots$$

and $g(t,u)$ will also have the form:

$$g(t,u) = G_0(u_0) + G_1(u_0,u_1)t + G_2(u_0,u_1,u_2)t^2 + G_3(u_0,u_1,u_2,u_3)t^3 + \cdots$$

In order for $g(t,u) = 0$ at $t = 0$, $G_0(u_0)$ must be zero, and since $G_0(u_0)$ is a polynomial in u_0, this gives us a finite number of values for u_0 that can solve our equation.

Now, by setting $g(t,u) = 0$, can we obtain $u(t)$ as a function of t?

The Implicit Function Theorem states that we can, if $\frac{\delta g}{\delta u}$ is not zero at the point we wish to expand around.

$$\frac{\delta g}{\delta u}(0,u_0) = \frac{\delta}{\delta u}G_0(u_0)$$

In short, the roots of $G_0(u_0)$ give us the starting values for our series expansion, and if the roots are simple, then the Implicit Function Theorem guarantees that we'll have a unique series expansion for $u(t)$ as a function of t. If any of the roots are not simple, then we can repeat this procedure for $g(t,u)$. It can be shown ([Bl47] §15) that this procedure always terminates.

Example 8.9. Construct Puiseux expansions of y at the multiple points of the curve $y^2 = 1 - x^2$

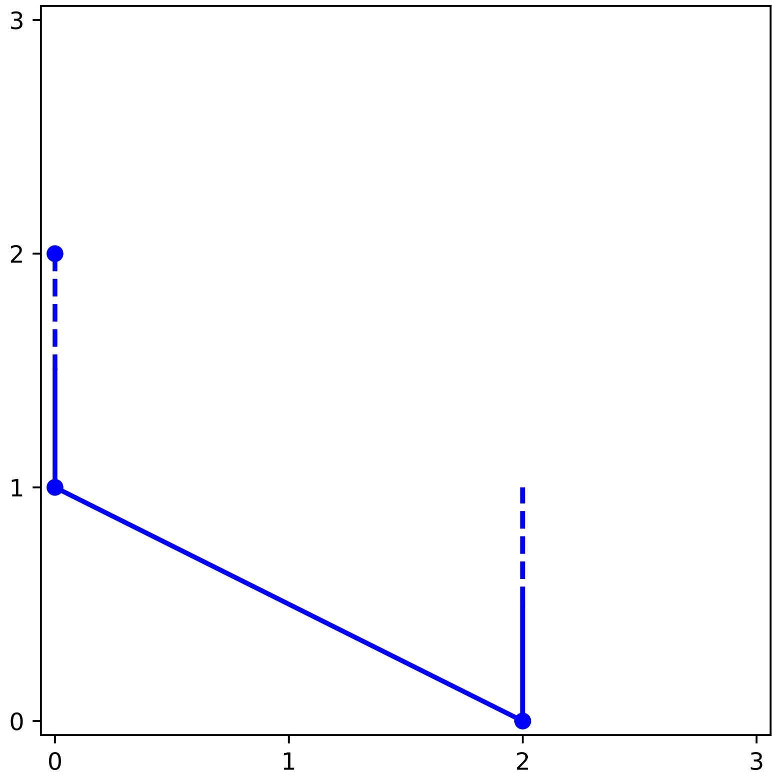
We normalize the defining polynomial by writing it as $y^2 + x^2 - 1 = 0$. Where does it have multiple points? We compute the discriminant:

```
sage: R.<x,y> = QQbar[];
sage: (y^2 + x^2 - 1).discriminant(y).factor()
```

$$(-4) \cdot (x - 1) \cdot (x + 1)$$

The multiple points of $y^2 = 1 - x^2$ lie at the roots of the discriminant, which are $x = \pm 1$. In both cases, $y = 0$ is the only solution, so the curve has multiple points at $(x, y) = (\pm 1, 0)$. The partial derivative of the defining polynomial with respect to x is $2x$, which is non-zero, so neither of these multiple points are singular; we'll get ramification instead. 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, $y^2 + (x-1)^2 + 2(x-1) = 0$. Next, we construct the Newton polygon by plotting the monomial powers, putting the y exponents on the horizontal axis and the $(x - 1)$ exponents on the vertical:



The only segment on the Newton polygon's lower convex hull has slope $-1/2$ and width

2, telling us that two of our roots (the width of the segment) will require a single Puiseux series with ramification index 2 (the denominator of the slope):

$$x = t^2 + 1$$

We know that y can be expressed as a power series in t with initial exponent 1 (the numerator of the slope):

$$y = a_1 t + a_2 t^2 + a_3 t^3 + \cdots$$

Now, substituting these expressions for x^2 and y^2 into the curve's defining equation $y^2 + x^2 - 1 = 0$ and setting all coefficients of t to zero, we find:

```
sage: var('x, y, t, a1, a2, a3');
sage: f = y^2 + x^2 - 1;
sage: exp = f.subs({x: t^2+1,
                    y: a1*t + a2*t^2 + a3*t^3});
sage: exp.collect(t)
```

$$a_3^2 t^6 + 2\,a_2 a_3 t^5 + 2\,a_1 a_2 t^3 + \left(a_2^2 + 2\,a_1 a_3 + 1\right) t^4 + \left(a_1^2 + 2\right) t^2$$

$$2 + a_1^2 = 0 \qquad 2a_1 a_2 = 0 \qquad 1 + 2a_1 a_3 + a_2^2 = 0$$

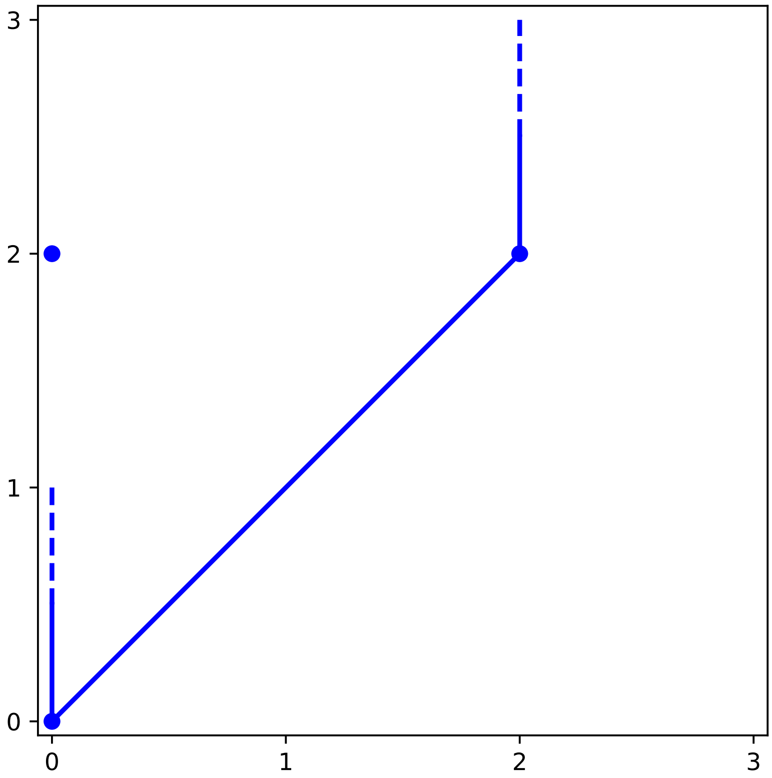
The first equation tells us that $a_1 = \pm\sqrt{2}i$, the second equation tells us that $a_2 = 0$ and the third equation tells us that $a_3 = \pm\frac{\sqrt{2}}{4}i$, so

$$x = t^2 + 1; \qquad y = \pm\left[\sqrt{2}it + \frac{\sqrt{2}}{4}it^3 + \cdots\right] \tag{8.1}$$

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 point at infinity. We move infinity to a finite point (0) with the substitution $x = u^{-1}$, then combine all of our terms over a common denominator and discard the denominator. Our curve becomes:

$$y^2 u^2 + 1 - u^2 = 0$$



The Newton polygon's lower convex hull has a single line segment, slope 1, length 2, telling us that we'll have two separate poles, each with ramification index 1. Thus, u can be used directly as a uniformizing variable, and we postulate an expansion for y in the form:

$$y = a_{-1}\frac{1}{u} + a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \cdots$$

Plugging this into $y^2u^2 + 1 - u^2$ and setting all the resulting coefficients to zero, we conclude:

```
sage: var('u, y, t, a0, a1, a2, a3');
sage: var('an1', latex_name='a_{-1}');
sage: f = y^2*u^2 + 1 - u^2
```

$$u^2y^2 - u^2 + 1$$

```
sage: exp = f.subs({y: an1*(1/u) + a0 + a1*u + a2*u^2 + a3*u^3});
```

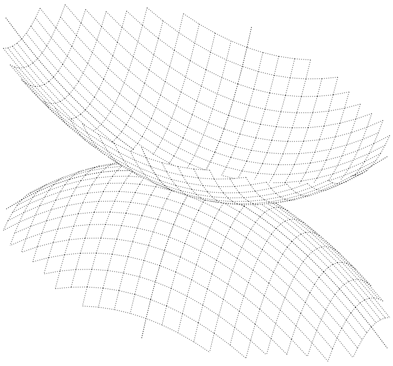
```
sage: exp.collect(u)
```

$$
\begin{aligned}
&a_3^2u^8 + 2\,a_2a_3u^7 + \left(a_2^2 + 2\,a_1a_3\right)u^6 + 2\,(a_1a_2 + a_0a_3)u^5\\
&\quad + \left(a_1^2 + 2\,a_0a_2 + 2\,a_3a_{-1}\right)u^4 + 2\,(a_0a_1 + a_2a_{-1})u^3\\
&\quad + 2\,a_0a_{-1}u + \left(a_0^2 + 2\,a_1a_{-1} - 1\right)u^2 + {a_{-1}}^2 + 1
\end{aligned}
$$

$$a_{-1} = \pm i; \qquad a_0 = 0; \qquad a_1 = \mp\frac{1}{2}i; \qquad a_2 = 0; \qquad a_3 = \mp\frac{1}{8}i$$

$$y = \pm i\frac{1}{u} \mp \frac{1}{2}iu \mp \frac{1}{8}iu^3 + \cdots$$

This time, there is no ramification, since u, and not a power of u, is $\frac{1}{x}$. We actually have two distinct series that will yield two different values of y for each value of u. Geometrically, we have two sheets that approach each other and touch at a singular point where the curve is not locally Euclidean, in a manner somewhat like this:



□

We've seen how to construct Puiseux expansions at arbitrary points of an algebraic curve, but some points have multiple expansions, corresponding to multiple cycles on their corresponding surfaces. These points are the curve's *singularities*, which are the only points where the curve is not locally Euclidean, and behaves instead like in the graphic above, where several Euclidean sheets touch at a single point.

We now seek some mechanisms for distinguishing between multiple cycles at a single point. To this end, we introduce the concept of a *place*. Intuitively speaking, a place is a cycle, and places are in one-to-one correspondence with cycles. Therefore, we can handle singularities by thinking in terms of places. Non-singular points have a unique place associated with them, while there are multiple places (and multiple cycles) associated with a singular point.

8.3 Valuation Rings and Orders

Driven in no small part by the difficulty in visualizing higher dimensional geometric shapes, mathematicians have developed increasingly algebraic techniques to manipulate geometric objects. In particular, the techniques of the previous section were presented largely for educational purposes, as they are now considered obsolete. The current state of the art is to use the tools of abstract algebra developed in the early twentieth century, such as rings, ideals, and fields.

We can't use ideals directly in a function field, or any field for that matter, because there are only two ideals in any field – the zero ideal, which is the ideal containing only the zero element, and the unit ideal, which is the entire field. This follows directly from the invertability of field elements. As all field element possess inverses, any non-zero ideal generator can be multiplied by its multiplicative inverse to generate 1, which then generates the entire field.

The obvious choice would be to use ideals of the *coordinate ring*, which is simply the polynomial ring modulo the defining polynomial of the curve, but these ideals are in one-to-one correspondence with the *points* of the curve, and we need some way to represent *places*.

To solve this problem, we introduce the concept of *valuation rings* and decompose the function field into *orders*, of which there are two of primary interest: the maximal finite order and the maximal infinite order. The ideals of the maximal finite order correspond to finite places, and the ideals of the maximal infinite order correspond to infinite places. These maximal orders can be constructed using the algebraic process of *normalization*. Using one or the other, we obtain ideals that represent all places in the function field.

Definition 8.10. *The* **coordinate ring** *of the algebraic curve is the ring* $K[x,y] \mod p(x,y)$. *If* $p(x,y)$ *is irreducible, then the coordinate ring is an integral domain.*

K is the curve's field of constants. The easy case is to let K be the complex numbers $\mathbb{C}$. Since $\mathbb{C}$ is algebraically closed, many of the theorems are simplified. Even if we're looking for real solutions, it is often simplest to work with complex numbers because of the simplification of the theory, then restrict to real solutions only at the end. For example, the entire argument of the last section is based on the curve having n solutions at all but a finite number of values for x; this is only the case if we're working over $\mathbb{C}$, or some other algebraically closed field, such as $\overline{\mathbf{Q}}$, the algebraic number field, which is the algebraic closure of $\mathbb{Q}$. For actually computations, however, $\mathbb{C}$ is often not the best choice because both the difficulty of expressing an arbitrary complex number, and due to the added complexity of having to fully factor a polynomial. See Example 8.21 for a problem where the results are simplified by using a different field of constants.

What do we do if $p(x,y)$ is not irreducible? In this case, we're trying something like computing an integral with a root of a polynomial that can be factored. We factor the polynomial, producing two separate roots. We then use the theorem of the primitive element (NEED TO PUT IT IN CHAPTER 2) to construct a single algebraic extension defined by a single irreducible polynomial in which we can construct both roots. So, for the remainder of this discussion, we need only consider the case where $p(x, y)$ is irreducible.

Definition 8.11. *The fraction field of the coordinate ring is called the curve's* **function field***, and its elements are called* **algebraic functions***.*

A curve's function field contains all rational functions in x and y, grouped together into equivalence classes by the relation $p(x, y) = 0$.

Definition 8.12. *A* **valuation ring** *of a function field F/K is a subring $O \in F$ such that $K \subset O \subset F$ (both proper inclusions), and for every $z \in F$, either $z \in O$ or $z^{-1} \in O$*

Both ideals and valuation rings are subrings, but there are two different conditions that they must satisfy. Fields admit only two ideals (the unit ideal and the trivial ideal), but there are typically an infinite number of valuation rings within a given function field.

We now wish to show that valuation rings are in one-to-one correspondence with places. The key observation is that an element z in is a valuation ring O only if z has non-negative valuation (i.e, is finite) at the corresponding place.

The condition that either z or z^{-1} must be in the valuation ring does not preclude the possibility that both will be in the valuation ring. In fact, these are precisely the elements of valuation zero, that have neither a zero nor a pole at the place in question. The elements of O whose inverses are not in O are the elements of positive valuation, that have a zero at the place in question. They form an ideal P of the valuation ring. O is, in fact, a *local ring*, and P is its unique maximal ideal.

Given a valuation ring O and its maximal prime ideal P, we define the **residue class field** of the valuation as $F_P := O/P$. Given any element of O, its value at the corresponding place is its image in F_P. The easy case in when the field of constants is algebraically closed. Otherwise, we can have non-rational places whose residue fields have degree (over the constants) greater than one.

A valuation ring might not be **discrete valuation ring** (DVR), which is a valuation ring with value group isomorphic to the integers under addition. All of our valuation rings are DVRs, mainly because we're working on an algebraic curve, which has Krull dimension one.[3] [St09] Theorem 1.1.6 shows that the valuation ring of a function field is a DVR by establishing one of a DVR's key properties: it's a principal ideal domain with a unique maximal ideal.

Theorem 8.13. *The valuation rings of a curve's function field are in one-to-one correspondence with its places.*

Proof

1. Every cycle induces a valuation ring

[3] `https://en.wikipedia.org/wiki/Discrete_valuation_ring` contains a list of equivalent conditions that a DVR satisfies, including "R is an integrally closed Noetherian local ring with Krull dimension one".

To do this, we just use the Puiseux expansion of a function field element to compute its valuation. If it is finite, it is in the valuation ring; if it is a pole, it is not in the valuation ring.

Show that valuation of the Puiseux expansion is independent of the choice of uniformizing variable (PUT THIS IN THE PREVIOUS SECTION)

Show compatibility of the Puiseux expansions – inverting one inverts the valuation (PUT THIS IN THE PREVIOUS SECTION)

2. Every valuation ring induces a cycle

   How to compute a Puiseux expansion from a valuation ring? Since we have a discrete valuation ring, which is a principal ideal domain, there exist elements which generate the DVR's unique maximal ideal P. Every such element $t \in P$ such that $P = tO$ is a uniformizing parameter.

   Show that $O \mod P$ is isomorphic to the constants. $x \mod P$ and $y \mod P$ give us the coordinates of the point corresponding to the place.

   Given an element in O, mod out by P to find its value. Subtract this value to get an element in P. Divide by a uniformizing parameter. Repeat to construct a Puiseux expansion.

   Show that all elements have a well-defined valuation (P^v contains them but P^{v+1} does not).

   If the element is not in O, its inverse must be, show that its inverse respects valuation and we can multiply by t^{-v}, where t is a uniformizing parameter, to move the inverse into O.

□

Since there are an infinite number of valuation rings, it is most convenient to separate them into two sets (the finite and the infinite), each of which can be represented as ideals of a particular subring (an order).

Definition 8.14. *A valuation ring O is* **finite** *if both $x \in O$ and $y \in O$. Otherwise, it is* **infinite**. *(or is it finite if only $x \in O$?)*

Definition 8.15. *An* **order** *$\mathcal{O}$ of a ring R (also called an R-order) is a subring of R such that*

1. *$\mathcal{O}$ is a finite-dimensional algebra over the field $\mathbb{Q}$ of rational numbers*
2. *$\mathcal{O}$ spans R over $\mathbb{Q}$, and*
3. *$\mathcal{O}$ is a $\mathbb{Z}$-lattice in R.*

[wikipedia]

Definition 8.16. *The* **maximal finite order** *of a function field F is the intersection of all its finite valuation rings.*

Show that it's an order.

Show that valuation rings are maximal ideals of the order. (are there other maximal ideals?)

Definition 8.17. *The* **maximal infinite order** *of a function field F is the intersection of all its infinite valuation rings.*

Show that it's an order.

All places correspond to a maximal ideal in one of these two orders. That ideal consists of all algebraic functions which are zero at that place and, in the finite case, have no finite poles. How can they be characterized in the infinite case? Equivalently, it consists of all integral elements (that satisify a monic polynomial) which are zero at that place. Two different places, even if they're at the same point, correspond to two distinct maximal ideals of the order.

Definition 8.18. *The* **integral closure** *of an integral domain R in its field of fractions F are the elements of F with no finite poles, or that admit monic defining polynomials.*

Theorem: Every element of an R-order is integral over R.

Theorem: The integral closure of an integral domain R in its field of fractions is an R-order, and in fact is maximal (due to the last theorem).

Corollary: The maximal finite order is the integral closure of the coordinate ring.

We might need all this stuff to show that integral closure is the same as the intersection of all the finite valuation rings.

[SwHu06] Proposition 6.8.14 Let R be an integral domain. Then the integral closure of the ring R equals $\cap_V V$, where V varies over all the valuation domains between R and its field of fractions. If R is Noetherian, all the V may be taken to be Noetherian.

[Ko07] Definition 1.23 Let S be an integral domain with quotient field $Q(S)$. The *normalization* of S in $Q(S)$, denoted by $\overline{S}$, is the unique largest subring $\overline{S} \subset Q(S)$ such that every homomorphism $\phi : S \to R$ to a DVR extends to a homomorphism $\overline{\phi} : \overline{S} \to R$.

[Ko07] Lemma 1.24 A unique factorization domain is normal. In particular, any polynomial ring over a field is normal.

[Ko07] Lemma 1.25. Assume that $t \in Q(S)$ satisfies a monic equation with coefficients in S (i.e, t is integral). Then $t \in \overline{S}$.

[Ko07] Defintion 1.27. Let S be an integral domain. The *normalization* of S is its integral closure in its quotient field

[Ko07]: "The easy argument that every normal integral domain S is the intersection of all the valuation rings sitting between it and its quotient field is given in [AM69, 5.22].

Working only with discrete valuation rings is a bit harder. The strongest theorem in this direction is Serre's condition for normality; see [Mat70, 17.1] or [Mat89, 23.8]."

[Mat89]:

R_i condition: A_p is regular for all $P \in \mathrm{Spec}A$ with $\mathrm{ht}P \leq i$

S_i condition: depth $A_p \geq \min(\mathrm{ht}P, i)$ for all $P \in \mathrm{Spec}A$

(S_0) always holds. (S_1) says that all the associated primes of A are minimal. $(R_0) + (S_1)$ is n.a.s.c. for A to be reduced.

Theorem 23.8: $(R_1) + (S_2)$ are n.a.s.c. for a Notherian ring A to be normal.

What do we need this for?

A **Dedekind domain** is a integral domain in which all ideals factor into a product of prime ideals.

Theorem: such a factorization is necessarily unique.

[Wiki Dedekind domain] Theorem: Let R be a Dedekind domain with fraction field K. Let L be a finite degree field extension of K and denote by S the integral closure of R in L. Then S is itself a Dedekind domain.

Theorem (Kummer): If $\{1, y, ..., y^{n-1}\}$ is a local integral basis at some prime polynomial $\mathfrak{p}$ in x, then we can factor the ideal $\mathfrak{p}^e$ by factoring the field's defining polynomial mod $\mathfrak{p}$. [Stichtenoch Theorem 3.3.7]

An **Artinian ring** satisfied the descending chain condition on ideals.

Theorem: the modulo ring constructed from a Dedekind domain and a proper ideal is an Artinian ring.

Show that the maximal finite order is a finitely generated $R[x]$-module. See Section 11.4.

To work with the maximal finite order of an algebraic curve, or the integral closure of the coordinate ring (it's same thing), we'd like to compute a basis for the order as an $R[x]$-module, called an **integral basis**. Computing an integral basis is not trivial, but algorithms to do this have been known for over a hundred years. Chapter 5 in [Al14] describes Trager's algorithm, which was originally published in [Tr84]. If our algebraic curve is formed as the root of a polynomial, i.e, $y^n = p(x)$, then an integral basis has a particularly simple form, given in [Tr84] pp. 30-31.

Once we have an integral basis, we can represent places as ideals within the order using their ideal generators, of which there are also a finite number. This gives us the most convenient way of working with places on an algebraic curve: an integral basis to describe the maximal finite order, and a set of ideal generators to describe a place as an ideal in said order.

As a basic (and important) example of the technique, I'll show how to evaluate an algebraic function at a place, given a representation of the place as an ideal of an order.

Algorithm 8.19. *Evaluation of an algebraic function at a place on a curve*

Given: a rational function f and a prime ideal I in a maximal order O.

Goal: compute the value of $f \bmod I$.

Step 0: Precompute a $k[x]$-module basis for I in Hermite normal form (doable since I is finitely generated over O and O is finitely generated over $k[x]$) and α, a rational function with a simple pole at the place corresponding to I and no other finite poles.

Step 1: Compute the valuation of f's denominator (how?); call it ν. Multiply both f's numerator and denominator by α^ν. Now both the numerator and denominator are in O, and the denominator is not in I (i.e, it has a finite value). So we've reduced to the case of computing the residue of an element of O, as we now can divide by the denominator's residue, since we know it's not zero.

Step 2: Use the techniques from Section 2.11 to compute residues of both the numerator and the denominator, and divide them to obtain the result.

Note: If the field k is not algebraically closed, then find a primitive element of the residue field and call it g. Compute g's minimal polynomial and use it to construct the residue field as an algebraic extension of the constant base field. We can represent g and its powers w.r.t. the HNF basis, construct a matrix that converts from g-basis to HNF basis, then invert it to obtain a matrix that converts from HNF basis to g-basis. Use reduction mod I's HNF basis, then multiply by this inverse matrix to obtain an element in the residue field, expressed in g-basis.

Source: Sage's `FunctionFieldPlace_polymod._residue_field`

Example 8.20. Compute the valuation of y/x at the single place lying over the point $(1, 1)$ in the function field $y^3 = x$.

The answer is obviously 1, but let's see how to compute this using the above algorithm.

First, we know that an integral basis for this function field is $(1, y, y^2)$ [Tr84] pp. 30-31.

We'll use the ordering $(y^2, y, 1)$, because we want the 1 to be last.

Our first generator $x - 1$ can be represented by the matrix

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

Remember that the columns correspond to the integral basis $(y^2, y, 1)$, and the rows are the generators $x - 1$, $(x - 1)y$, and $(x - 1)y^2$.

Our second generator $y-1$ can be represented by the matrix

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

Putting this second matrix in HNF we obtain:

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

Combining the two matrices, we obtain:

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

Reducing further using HNF elementary row operations yields:

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

Since the two three rows are identical to those from the $y-1$ matrix, this implies that the ideal is principal and $y-1$ is the only generator required and our ideal's HNF matrix is:

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

x is represented as $(0,0,x)$. Applying the HNF reduction algorithm to this vector, we reduce by left-multiplying the matrix by $(0,0,-1)$ to obtain $(0,0,1-x)$. Adding this vector to $(0,0,x)$ gives us $(0,0,1)$, which, when dot-multiplied by $(y^2,y,1)$, yields 1, our remainder.

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^3 - x)

         Function field in y defined by y^3 - x

sage: F.maximal_order().ideal(x-1, y-1)
```

$$(y-1)$$

```
sage: F.maximal_order().ideal(y-1).hnf()
```

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

□

Conceptually, the various rings, fields, and inclusion maps look like this:

$$\begin{tikzcd}
\textbf{Rational Functions} & \textbf{Algebraic Curve} & \\
C(x) \arrow[r] & C(x,y) \mod p(x,y) & \text{Fraction Fields} \\
 & \mathcal{I} \arrow[u] & \text{Maximal Finite Order / Integral Basis} \\
C[x] \arrow[uu] \arrow[r] & C[x,y] \mod p(x,y) \arrow[u] & \text{Coordinate Rings}
\end{tikzcd}$$

The coordinate rings are subsets of their fraction fields, and can be mapped into them. The univariate coordinate rings and fields (the rational functions in x) are subsets of their counterparts on the algebraic curve, and can be mapped into them. Finally, the algebraic curve's integral basis is a subset of its fraction field and a superset of its coordinate ring.

[St09] Corollary 1.2.3 shows that all places of $C(x)$ (the rational function field) are in 1-to-1 correspondence with $C \cup \inf$ (as places).

We have the function field $C(x)$ and its places, and also the function field defined by the curve, and its places. Since all the rational functions in $C(x)$ are also rational functions in $C(x, y) \mod p(x, y)$ (the coordinate ring), it follows that the function field defined by the curve is a field extension of the rational function field.

The maximal finite order of an algebraic curve is an extension of the curve's coordinate ring, which is an extension of the univariate polynomial ring. The maximal ideals in the univariate ring correspond to coordinates in a single variable x, the maximal ideals in the coordinate ring correspond to the points on the curve, and the maximal ideals in the maximal finite order correspond to the places of the curve. By computing how these ideals decompose under extension, we can compute the points that lie over a given x coordinate, as well as the places that lie over a given point, or a given x coordinate.

All points correspond to maximal ideals in the coordinate ring. For points, our ideal generators are just $x - x_0$ and $y - y_0$, and this ideal is maximal in the *coordinate ring*. If (x_0, y_0) was *not* on the curve, then the ideal would be $(x - x_0, y - y_0, p(x, y))$, which would simplify to $p(x, y)$ being a non-zero constant, and this would generate the unit ideal.

At a non-singular point, the ideal $(x - x_0, y - y_0)$ is prime not only in the coordinate ring, but also in the maximal finite order. How can we show this?

At a singularity, the ideal $(x - x_0, y - y_0)$ will be prime in the coordinate ring, but will decompose in the maximal finite order into multiple prime (and maximal) ideals, corresponding to the places that lie at that point.

1. All places of an extension are "over" a place of the underlying field
2. $v_{P'}(x) = e \cdot v_P(x)$, so e is the ramification index (and is a finite integer)

Do we need this discussion here? Only if we need to compute ramification indices, I think.

[Wiki Ideal] Theorem: if $f : A \to B$ is surjective and $\mathfrak{a} \supseteq \ker f$ then:

- $\mathfrak{a}^{ec} = \mathfrak{a}$ and $\mathfrak{b}^{ce} = \mathfrak{b}$
- $\mathfrak{a}$ is a prime ideal in $A \Leftrightarrow \mathfrak{a}^e$ is a prime ideal in B.
- $\mathfrak{a}$ is a maximal ideal in $A \Leftrightarrow \mathfrak{a}^e$ is a maximal ideal in B.

So, to factor $\mathfrak{p}$, we need to find all of the prime/maximal ideals in $R \mod \mathfrak{p}$.

Constructing $R \mod \mathfrak{p}$ as a finite dimensional algebra, we have an algorithm to enumerate all of the algebra's maximal ideals. This is the same as the ideal decomposition algorithm.

We also want to find each ideal's ramification index and relative degree. I wrote a paper on this.

The ramification indices are the powers of the prime ideals. Cohen Theorem 4.8.3:

$$p\mathbb{Z}_K = \prod_{i=1}^{g} \mathfrak{p}_i^{e_i}$$

Stichnoch defines the ramification index as the integer such that $\nu_{P'}(x) = e\nu_P(x)$ for all x in the base field. (P' lies over P)

We can find the ramification index in the modulo ring by raising each maximal ideal to successive powers and determining when it stabilizes. The Artinian condition guarantees that it will eventually stabalize.

The relative (or residual) degree of $\mathfrak{p}$ is defined (Cohen Definition 4.8.4):

$$f_i = [\mathbb{Z}_K/\mathfrak{p}_i : \mathbb{Z}/p\mathbb{Z}]$$

This seems to be the nullity of the matrix that defines the ideal in the algebra.

Why? The dimension of the algebra (over $\mathbb{Z}/p\mathbb{Z}$) is the dimension of $\mathbb{Z}_K$. The number of basis elements in the ideal is the dimension of $\mathfrak{p}_i$. The dimension of $\mathbb{Z}_K/\mathfrak{p}_i$ is the difference of these two numbers (the dimension of the algebra minus the dimension of the ideal), which is the dimension of the vector space minus the dimension of the ideal's basis matrix's kernel, or the nullity of this matrix.

8.4 Sage's FunctionField code

Sage has built-in routines to compute Puiseux expansions without having to construct Newton polygons and substitute trial expansions. In fact, there are three different constructions in Sage for algebraic curves – polynomial rings modulo an ideal, function fields, and curves in projective space. We'll mostly use the `FunctionField` code.

Instead of constructing a polynomial ring the usual way (which is?), we'll construct a univariate `FunctionField` over a single variable, specifying its constant field, then form its `extension` modulo the curve's defining polynomial, which will yield another, bivariate, `FunctionField`.

A Sage `FunctionField` has methods to construct both the maximal finite order and the maximal infinite order. We can construct ideals of these orders, and from them construct the corresponding places. Given a place in the underlying univariate `FunctionField`, we have a method that will extend the corresponding ideal to the bivariate `FunctionField`, factor (decompose) it, and return a list of all `places_above` the original place.

The standard way to construct Puiseux expansions in Sage is the `completion` method, defined on `FunctionField`'s, but this method has some limitations. It doesn't work on differentials (only on functions), doesn't allow a uniformizing parameter to be specified, and doesn't allow absolute precision to be specified (only relative precision). To overcome this problems, I've written the following function that's basically just a wrapper around `completion`. Its default behavior is to compute either the principal part of a Puiseux expansion for poles, or a single term of the Puiseux expansion for finite values. The main thing I don't like about it is that when you pass it a differential, it returns a `LaurentSeries`, when it should actually return a `LaurentSeries` times ds, where s is the uniformizing variable.

If a uniformizing variable is specified, either as a function field element or as a `LaurentSeries`, we reverse the series, which only works if the series has valuation 1, giving us a new series that expresses s in powers of the uniformizing variable, then then substitute the reversed series into the original s-expansion of the function field element, which gives us an expansion of the element in powers of the uniformizing variable.

```
from sage.rings.function_field.differential import FunctionFieldDifferential
def puiseux(F, pl, uvar=None, absprec=0):
    def puiseux2(f):
       if isinstance(f, FunctionFieldDifferential):
          base_differential = f.parent()._gen_base_differential
          f = f._f
          is_differential = True
          valuation = f.valuation(pl) + base_differential.valuation(pl) - 1
       else:
          is_differential = False
          valuation = f.valuation(pl)
       series = F.completion(pl, prec=max(1, absprec-valuation))
       if uvar:
          if isinstance(uvar, LaurentSeries):
             uvar_series = uvar.reverse()
          else:
            uvar_series = series(uvar).reverse()
         f_series = series(f)(uvar_series)
      else:
         f_series = series(f)
      if is_differential:
         if uvar:
            f_series *= series(base_differential).derivative()(uvar_series)
         else:
            f_series *= series(base_differential).derivative()
      return f_series
   return puiseux2
```

Example 8.9 cont. Construct Puiseux expansions of y at the multiple points of the curve $y^2 = 1 - x^2$

This is Example 8.9, redone now using Sage `FunctionField`.

First we create a function field in one variable with coefficients in $\overline{\mathbf{Q}}$:

```
sage: R.<x> = FunctionField(QQbar)

Rational function field in x over Algebraic Field
```

Next we create a ring of polynomials in y with coefficients in the function field, which allows us to write the minimal polynomial of the algebraic curve. We then create a new function field that is an extension of the rational function field:

```
sage: L.<y> = R[]
Rational function field in x over Algebraic Field[y]
sage: F.<y> = R.extension(y^2 + x^2 - 1)

  Function field in y defined by y^2 + x^2 - 1
```

Recall from Example 8.9 that our multiple points lie at $x = \pm 1$. Since we wish to construct our series expansion at a point with finite coordinates (remember that this is projective space, so we also have points at infinity), we use the finite maximal order, construct the ideal corresponding the desired point, then construct the unique place at that point:

```
sage: O = F.maximal_order()
Maximal order of Function field in y defined by y^2 + x^2 - 1
sage: pl = O.ideal(x-1, y).place()
```

$$(x - 1, y)$$

Finally, we construct the Puiseux expansion of y at the place, specifying the desired precision of the expansion. Notice that `puiseux` actually returns a function, to which we pass the function field element (y) that we wish to expand:

```
sage: puiseux(F, pl, absprec=4)(y)
```

$$s + O(s^4)$$

This answer differs from the one we computed in Example 8.9 because the choice of uniformizing variable is not unique, and the computer made a different choice than we did. Our "t" variable in Example 8.9 was roughly $\sqrt{x-1}$. Even though square roots don't exist in this field, we can construct a series expansion of $x-1$, and construct the square root of the series expansion as another series expansion.

```
sage: xminus1 = puiseux(F, pl, absprec=5)(x-1)
```

$$-\frac{1}{2}s^2 - \frac{1}{8}s^4 + O(s^5)$$

```
sage: t = xminus1^(1/2)
```

$$\frac{1}{2}i\sqrt{2}s + \frac{1}{8}i\sqrt{\frac{1}{2}}s^3 + O(s^4)$$

Using this series as a uniformizing variables gives us y in powers of t, even though the computer still prints the results using s.

```
sage: puiseux(F, pl, absprec=4, uvar=t)(y)
```

$$-i\sqrt{2}s - \frac{1}{2}i\sqrt{\frac{1}{2}}s^3 + O(s^4)$$

Comparing this to equation 8.1, we see that they're the same.

Now let us turn to the point at infinity. We begin by constructing the maximal infinite order of the underlying rational function field $C(x)$, because its structure is quite simple and we know that it will only have a single point, and a single place, at infinity.

```
sage: ROinf = R.maximal_order_infinite();

sage: Rinf = ROinf.ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

Now we'll use the `places_above` method to obtain a list of all places in an extension field that lie over a given place in the underlying function field. Recall from Example 8.9 that this curve has a singular point at infinity with two separate cycles and therefore two separate places.

```
sage: Pinf = F.places_above(Rinf)
```

$$\left(\left(\frac{1}{x}, \frac{1}{x}y - i\right), \left(\frac{1}{x}, \frac{1}{x}y + i\right)\right)$$

Having obtained these places, represented as ideals in the maximal infinite order, it is straightforward to use the `puiseux` function to construct Puiseux series for x and y.

```
sage: [puiseux(F, pl, absprec=5)(x) for pl in Pinf]
```

$$\left[\frac{1}{s} + O(s^5), \frac{1}{s} + O(s^5)\right]$$

```
sage: [puiseux(F, pl, absprec=5)(y) for pl in Pinf]
```

$$\left[\frac{i}{s} - \frac{1}{2}is - \frac{1}{8}is^3 + O(s^5), \frac{-i}{s} + \frac{1}{2}is + \frac{1}{8}is^3 + O(s^5)\right]$$

Comparing this to Example 8.9, we see that the results are the same.

□

Example 8.21. [Bl47] §68 Compute expansions at all multiple points of

$$y^3 + x^3y + x = 0$$

We begin by computing the discriminant of the equation, which gives us the locations of the multiple points.

```
sage: R.<x,y> = QQbar[];
sage: f = y^3 + x^3*y + x
```

$$x^3y + y^3 + x$$

```
sage: f.discriminant(y).factor()
```

$$\begin{aligned}&(-4) \cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(-\frac{6}{7} i \pi\right)}) \cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(\frac{6}{7} i \pi\right)}) \\ &\cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(-\frac{4}{7} i \pi\right)}) \cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(\frac{4}{7} i \pi\right)}) \\ &\cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(-\frac{2}{7} i \pi\right)}) \cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(\frac{2}{7} i \pi\right)}) \cdot (x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}}) \cdot x^2\end{aligned}$$

That result is rather confusing. Let's try factoring over $\mathbf{Q}$ instead of $\overline{\mathbf{Q}}$:

```
sage: f.discriminant(y).change_ring(QQ).factor()
```

$$(-1) \cdot x^2 \cdot (4x^7 + 27)$$

The multiple points lie over the roots of this equation: $x = 0$ and the seven roots of $4x^7 + 27 = 0$. Infinity also needs to be examined. We begin with $x = 0$:

```
sage: R.<x> = FunctionField(QQbar)
```

```
Rational function field in x over Algebraic Field
```

```
sage: L.<y> = R[]
```

Rational function field in x over Algebraic Field$[y]$

```
sage: F.<y> = R.extension(y^3 + x^3*y + x)
```

```
Function field in y defined by y^3 + x^3*y + x
```

```
sage: O = F.maximal_order()
```

```
Maximal order of Function field in y defined by y^3 + x^3*y + x
```

```
sage: F.places_above(R.maximal_order().ideal(x).place())
```

$$((y))$$

```
sage: O.ideal(x).factor()
```

$$((1)) \cdot (y)^3$$

```
sage: pl = O.ideal(x, y).place()
```

$$(y)$$

```
sage: xseries = puiseux(F, pl)(x)
```

$$-s^3 + O(s^4)$$

```
sage: yseries = puiseux(F, pl)(y)
```

$$s + O(s^2)$$

```
sage: yseries((xseries^(1/3)).reverse())
```

$$\left(-\frac{1}{2} i \sqrt{3} + \frac{1}{2}\right) s + O\left(s^{2}\right)$$

This result shows that we have a single cycle at $(x, y) = (0, 0)$ with three sheets. Now, let's look at a specimen root of $4x^7 + 27 = 0$:

```
sage: g = QQbar(-27/4)^(1/7)
```

$$\frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} \left(-1\right)^{\frac{1}{7}}$$

```
sage: pl = O.ideal(x-g, y+3/(2*g^2)).place()
```

$$\left(x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(-\frac{6}{7} i \pi\right)}, y + \frac{1}{8} \cdot 8^{\frac{6}{7}} 3^{\frac{1}{7}} e^{\left(-\frac{2}{7} i \pi\right)}\right)$$

```
sage: xseries = F.completion(pl, prec=3)(x)
```

$$\frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} \left(-1\right)^{\frac{1}{7}} + \left(\frac{3}{7} \cdot 48^{\frac{1}{7}} \left(-1\right)^{\frac{1}{7}} e^{\left(\frac{4}{7} i \pi\right)}\right) s^{2} + O\left(s^{3}\right)$$

```
sage: yseries = F.completion(pl, prec=2)(y)
```

$$\frac{1}{8} \cdot 8^{\frac{6}{7}} 3^{\frac{1}{7}} \left(-1\right)^{\frac{1}{7}} e^{\left(\frac{4}{7} i \pi\right)} + s + O\left(s^{2}\right)$$

```
sage: pl = O.ideal(x-g, y-3/g^2).place()
```

$$\left(x + \frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} e^{\left(-\frac{6}{7} i \pi\right)}, y + 48^{\frac{1}{7}} \left(-1\right)^{\frac{1}{7}} e^{\left(\frac{4}{7} i \pi\right)}\right)$$

```
sage: xseries = F.completion(pl, prec=3)(x)
```

$$\frac{1}{4} \cdot 27^{\frac{1}{7}} 4^{\frac{6}{7}} \left(-1\right)^{\frac{1}{7}} + s + O\left(s^{3}\right)$$

```
sage: yseries = F.completion(pl, prec=2)(y)
```

$$48^{\frac{1}{7}} e^{\left(-\frac{2}{7} i \pi\right)} + \left(\frac{10}{81} \cdot 64^{\frac{1}{7}} 9^{\frac{6}{7}} \left(-1\right)^{\frac{1}{7}} e^{\left(-\frac{4}{7} i \pi\right)}\right) s + O\left(s^{2}\right)$$

```
sage: yseries((xseries-g).reverse())
```

$$48^{\frac{1}{7}}e^{\left(-\frac{2}{7}i\,\pi\right)} + \left(\frac{10}{81}\cdot 64^{\frac{1}{7}}9^{\frac{6}{7}}\left(-1\right)^{\frac{1}{7}}e^{\left(-\frac{4}{7}i\,\pi\right)}\right)s + O(s^2)$$

```
sage: R1.<g> = QQ[]
```

$$\mathbf{Q}[g]$$

```
sage: S.<g> = NumberField(4*g^7+27)
```

$$\mathbf{Q}[g]/(4g^7+27)$$

```
sage: R.<x> = FunctionField(S)
```

```
Rational function field in x over Number Field in g with defining polyno
```

```
sage: L.<y> = R[]
```

```
Rational function field in x over Number Field in g with defining polyno
```

```
sage: F.<y> = R.extension(y^3 + x^3*y + x)
```

```
     Function field in y defined by y^3 + x^3*y + x
```

```
sage: O = F.maximal_order()
```

```
Maximal order of Function field in y defined by y^3 + x^3*y + x
```

```
sage: O.ideal(x-g).factor()
```

$$((1))\cdot(\left(x-g, y+\frac{4}{9}g^5\right))\cdot(\left(x-g, y-\frac{2}{9}g^5\right))^2$$

```
sage: pl = O.ideal(x-g, y+3/(2*g^2)).place()
```

$$\left(x-g, y-\frac{2}{9}g^5\right)$$

```
sage: xseries = F.completion(pl, prec=3)(x)
```

$$g+\frac{4}{21}g^{5}s^{2}+O(s^{3})$$

```
sage: yseries = F.completion(pl, prec=2)(y)
```

$$\frac{2}{9}g^{5}+s+O(s^{2})$$

```
sage: xseries((yseries+3/(2*g^2)).reverse())
```

$$g+\frac{4}{21}g^{5}s^{2}+O(s^{3})$$

```
sage: pl = O.ideal(x-g, y-3/g^2).place()
```

$$\left(x-g,y+\frac{4}{9}g^{5}\right)$$

```
sage: xseries = F.completion(pl, prec=2)(x)
```

$$g+s+O(s^{2})$$

```
sage: yseries = F.completion(pl, prec=2)(y)
```

$$-\frac{4}{9}g^{5}-\frac{40}{81}g^{4}s+O(s^{2})$$

```
sage: yseries((xseries-g).reverse())
```

$$-\frac{4}{9}g^{5}-\frac{40}{81}g^{4}s+O(s^{2})$$

We have one sheet of two cycles at $(g,-3/(2g^{2}))$ and an ordinary point at $(g,3/g^{2})$.

Finally, let’s look at what happens when x goes to infinity:

```
sage: Rinf = R.maximal_order_infinite().ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

```
sage: Pinf = F.places_above(Rinf)
```

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

```
sage: xseries = puiseux(F, Pinf[0])(x)
```

$$\frac{1}{s} + O(1)$$

```
sage: yseries = puiseux(F, Pinf[0])(y)
```

$$-s^2 + O(s^3)$$

```
sage: xseries = puiseux(F, Pinf[1])(x)
```

$$\frac{-1}{s^2} + O(1)$$

```
sage: yseries = puiseux(F, Pinf[1])(y)
```

$$\frac{1}{s^3} + O(1)$$

Here we have an ordinary point at $(\infty, 0)$ and a single cycle of two sheets at (∞, ∞).

We have examined all of this curve's ramification points, including those at infinity (since we analyzed all of its points at infinity), and found that all of them admitted a single Puiseux expansion.

Therefore, this curve is *non-singular*, and according to the genus-degree formula (MORE INFO), its geometric and arithmetic genus are the same. Its arithmetic genus is $\frac{1}{2}(d-1)(d-2) = 3$, where $d = 4$ is the degree of the defining polynomial. Computing the geometric genus is more difficult[4], but we can verify our information with Sage, being careful to work in *projective* space:

```
sage: PP.<x,y,z> = ProjectiveSpace(QQ, 2)
```

$$\mathbf{P}^2_{\mathbf{Q}}$$

[4] https://www.singular.uni-kl.de/Overview/Examples/Genus/genus1.html
https://en.wikipedia.org/wiki/Algebraic_curve#Classification_of_singularities
https://math.stackexchange.com/questions/150840
http://mathforum.org/library/drmath/view/71229.html

```
sage: C = Curve(y^3*z + x^3*y + x*z^3)
```

$$x^3y + y^3z + xz^3$$

```
sage: C.is_singular()
```

False

```
sage: C.arithmetic_genus()
```

3

```
sage: C.geometric_genus()
```

3

□

Example 8.22. Find the principal parts of $\frac{1}{y}$ on the curve $y^2 = 1 - x^2$

The *principal part* of an algebraic function is the part of its series expansion with negative exponents. Theorem 8.7 states that an algebraic function is completely determined, up to adding a constant, by its principal parts.

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 $x = \pm 1$. Now, if we use `puiseux`, we can just request a series truncated at the -1 term:

```
sage: R.<x> = FunctionField(QQbar)

    Rational function field in x over Algebraic Field

sage: L.<y> = R[]
```

Rational function field in x over Algebraic Field$[y]$

```
sage: F.<y> = R.extension(y^2 + x^2 - 1)

    Function field in y defined by y^2 + x^2 - 1

sage: D = (1/y).divisor()
```

$$\left(\frac{1}{x}, \frac{1}{x}y - i\right) + \left(\frac{1}{x}, \frac{1}{x}y + i\right) - (x - 1, y) - (x + 1, y)$$

```
sage: table([[p, F.completion(p, prec=1)(1/y)] \
          for p,m in D.list() if m < 0])
```

$$\begin{array}{ll} (x-1,y) & \frac{1}{s} + O(1) \\ (x+1,y) & \frac{1}{s} + O(1) \end{array}$$

□

Example 8.23. 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, due to the presence of the differential, which must be adjusted at ramification points.

Let's expand $\frac{x}{y}$ at $x = 1$:

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 + x^2 - 1)

     Function field in y defined by y^2 + x^2 - 1

sage: O = F.maximal_order()

Maximal order of Function field in y defined by y^2 + x^2 - 1

sage: pl = O.ideal(x-1, y).place()
```

$$(x-1, y)$$

```
sage: F.completion(pl, prec=6)(x)
```

$$1 - \frac{1}{2}s^2 - \frac{1}{8}s^4 + O(s^6)$$

```
sage: F.completion(pl, prec=6)(x/y)
```

$$\frac{1}{s} - \frac{1}{2}s - \frac{1}{8}s^3 + O(s^5)$$

Now $x = t^2 + 1$, so $\mathrm{d}x = 2t\,\mathrm{d}t$. Thus, multiplying $\frac{x}{y}$ by $\mathrm{d}x$ and changing our variable to t will multiply all of the terms in our expansion by $2t$:

```
sage: dx = F.completion(pl, prec=6)(x).derivative()
```

$$-s - \frac{1}{2}s^3 + O(s^5)$$

```
sage: F.completion(pl, prec=6)(x/y) * dx
```

$$-1 + O(s^4)$$

Even though $\frac{x}{y}$ has a pole at $x = 1$, $\frac{x}{y}\,\mathrm{d}x$ does not!

Its behavior at infinity also requires analysis.

```
sage: F.maximal_order().ideal(x - sqrt(QQbar(-1)));

sage: Rinf = R.maximal_order_infinite().ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

```
sage: Pinf = F.places_above(Rinf)
```

$$\left(\left(\frac{1}{x}, \frac{1}{x}y - i\right), \left(\frac{1}{x}, \frac{1}{x}y + i\right)\right)$$

```
sage: F.completion(Pinf[0], prec=2)(x)
```

$$\frac{1}{s} + O(s)$$

```
sage: F.completion(Pinf[0], prec=2)(y)
```

$$\frac{i}{s} + O(s)$$

```
sage: F.completion(Pinf[0], prec=2)(x/y)
```

$$-i + O(s^2)$$

```
sage: dxinf = F.completion(Pinf[0], prec=2)(x).derivative()
```

$$\frac{-1}{s^2} + O(1)$$

```
sage: F.completion(Pinf[0], prec=2)(x/y) * dxinf
```

$$\frac{i}{s^2} + O(1)$$

$\frac{x}{y}$ has no poles at infinity, and approaches the limiting values $\pm i$ as x and y approach infinity. The differential $\frac{x}{y}\,\mathrm{d}x$, on the other hand, requires us to multiply by $\mathrm{d}x$, and since $x = \frac{1}{t}$, $\mathrm{d}x = -\frac{1}{t^2}\,\mathrm{d}t$.

In short, while $\frac{x}{y}$ has poles only at $(\pm 1, 0)$, $\frac{x}{y}\,\mathrm{d}x$ has poles only at infinity.

```
sage: D = (x/y).divisor() + F(x).differential().divisor()
```

$$-2\left(\frac{1}{x}, \frac{1}{x}y - i\right) - 2\left(\frac{1}{x}, \frac{1}{x}y + i\right) + (x, y - 1) + (x, y + 1)$$

```
sage: table([[p, F.completion(p, prec=2)(x/y) * F.completion(p)(x).deri
```

$$\begin{array}{ll} \left(\frac{1}{x}, \frac{1}{x}y - i\right) & \frac{i}{s^2} + O(1) \\ \left(\frac{1}{x}, \frac{1}{x}y + i\right) & \frac{-i}{s^2} + O(1) \end{array}$$

□

Consider, for example the lemniscate of Bernoulli, defined by the equation

$$(x^2 + y^2)^2 - (x^2 - y^2) = 0$$

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]
```

```
sage: F.<y> = R.extension((x^2+y^2)^2 - (x^2-y^2))

Function field in y defined by y^4 + (2*x^2 + 1)*y^2 + x^4 - x^2

sage: O = F.maximal_order()

Maximal order of Function field in y defined by y^4 + (2*x^2 + 1)*y^

sage: I = O.ideal(x,y)
```

$$(x, y)$$

```
sage: I.factor()
```

$$((1)) \cdot (\left(x, \frac{1}{x}y^3 + \frac{1}{x}y - 1\right)) \cdot (\left(x, \frac{1}{x}y^3 + \frac{1}{x}y + 1\right))$$

```
sage: O.basis()
```

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

One characterization of the maximal finite order O is that it contains all functions with no poles at finite places. The first three elements in the basis are obvious, but why is the fourth so complicated? Isn't $\frac{y}{x}$ in O?

```
sage: y/x in O
```

False

Doesn't $\frac{y}{x}$ approach either 1 or -1 as it approaches the origin? Remember that we're working in complex space. We've got four roots, not two. Let's look at some examples of limiting values using some numerical examples:

```
sage: set_verbose(-1)
```

None

```
sage: R.<u,v> = CC[]
```

$$\mathbf{C}[u, v]$$

```
sage: ideal(((u^2+v^2)^2 - (u^2-v^2)), (u-.0001))
```

$$\left(u^4 + 2.00000000000000u^2v^2 + v^4 - u^2 + v^2, u - 0.000100000000000000\right)\mathbf{C}[u, v]$$

```
sage: ideal(((u^2+v^2)^2 - (u^2-v^2)), (u-.0001)).variety()
```

```
[{u:  0.000100000000000000, v:  0.0000999999980000001},{v:  1.00000001500
```

```
sage: [d[v] for d in ideal(((u^2+v^2)^2 - (u^2-v^2)), (u-.0001)).variety
```

$$[0.0000999999980000001, 1.00000001500000i,\\ -1.00000001500000i, -0.0000999999980000001]$$

```
sage: [d[v]/.0001 for d in ideal(((u^2+v^2)^2 - (u^2-v^2)), (u-.0001)).v
```

$$[0.999999980000001, 10000.0001500000i, -10000.0001500000i, -0.999999980000001]$$

```
sage: [d[v]/.00001 for d in ideal(((u^2+v^2)^2 - (u^2-v^2)), (u-.00001))
```

$$[0.999999999800000, 100000.000015000i, -100000.000015000i, -0.999999999800000]$$

```
sage: [(d[v]^3+d[v])/.00001 for d in ideal(((u^2+v^2)^2 - (u^2-v^2)), (u
```

$$[0.999999999900000, -0.0000300000024822111i, 0.0000300000024822111i,\\ -0.999999999900000]$$

```
sage: Fs = O.ideal(x).factor()
```

$$((1)) \cdot ((x, y - i)) \cdot (\left(x, \frac{1}{x}y^3 + \frac{1}{x}y - 1\right)) \cdot (\left(x, \frac{1}{x}y^3 + \frac{1}{x}y + 1\right)) \cdot ((x, y + i))$$

```
sage: F.completion(Fs[0][0].place(), prec=7)(y)
```

$$i + \frac{3}{2}is^2 - \frac{25}{8}is^4 + \frac{203}{16}is^6 + O(s^7)$$

```
sage: F.completion(Fs[1][0].place(), prec=7)(y)
```

$$s - 2s^3 + 6s^5 - 28s^7 + O(s^8)$$

```
sage: F.completion(Fs[2][0].place(), prec=7)(y)
```

$$-s + 2s^3 - 6s^5 + 28s^7 + O(s^8)$$

```
sage: F.completion(Fs[3][0].place(), prec=7)(y)
```

$$-i - \frac{3}{2}is^{2} + \frac{25}{8}is^{4} - \frac{203}{16}is^{6} + O(s^{7})$$

How does one of these ideal generators factor?

```
sage: A.<a> = QQbar[]
```

$$\overline{\mathbf{Q}}[a]$$

```
sage: (1/2*a^2+1/2*sqrt(QQbar(-1))*a+1).factor()
```

$$\left(\frac{1}{2}\right) \cdot (a - i) \cdot (a + 2i)$$

8.5 Divisors

Given a function on an algebraic curve, we can ask at which places it has poles and zeros. The location and strengths of a function's poles and zeros are called its *divisor*.

For non-singular curves, the points and places are in one-to-one correspondence, and a function's divisor can be described in terms of the points where its poles and zeros lie.

Thus, one way of defining a divisor is to associate integers (positive for zeros, negative for poles) with each point of the curve, subject to the stipulation that all but a finite number of those integers is zero. Such a description is called a *Weil divisor*, and is most suitable for working in *intersection theory*.

?? Definition 1.5.6 defines a Weil *differential*.

For singular curves, the situation is more complicated. A divisor needs to be associated with places, not points. Such a divisor is called a *Cartier divisor*, and is more suitable for our purposes.

8.6 Riemann-Roch spaces

A *Riemann-Roch space* is a subspace of an algebraic curve's function field characterized by specifying a minimum order that the function must obtain at all of the curve's points. Aside from having great theoretical significance, Riemann-Roch spaces are practically useful because they are finite dimensional, and algorithms exist for constructing Riemann-Roch bases. Finding a basis for a Riemman-Roch space in a crucial first step in solving a Mittag-Leffler problem.

Numerous algorithms have been developed for computing bases of Riemann-Roch spaces. Sage uses an implementation of Hess's algorithm from [He02].

Here's a simple example[5] of a Riemann-Roch space calculation:

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 - x^3 + x)


      Function field in y defined by y^2 - x^3 + x

sage: O = F.maximal_order()

Maximal order of Function field in y defined by y^2 - x^3 + x

sage: P = O.ideal(x,y)

                          (x, y)

sage: D = P.divisor()


                          (x, y)

sage: D.basis_function_space()

                           [1]
```

[5] From https://math.stackexchange.com/questions/294644

```
sage: (2*D).basis_function_space()
```

$$\left[1, \frac{1}{x}\right]$$

```
sage: (3*D).basis_function_space()
```

$$\left[1, \frac{1}{x}, \frac{1}{x^2}y\right]$$

```
sage: (4*D).basis_function_space()
```

$$\left[1, \frac{1}{x}, \frac{1}{x^2}, \frac{1}{x^2}y\right]$$

Here are the examples from [Al14] §6.3:

```
sage: R.<x> = FunctionField(QQbar)
```

```
    Rational function field in x over Algebraic Field
```

```
sage: L.<y> = R[]
```

Rational function field in x over Algebraic Field$[y]$

```
sage: F.<y> = R.extension(y^2 - x^3 - 1)
```

```
      Function field in y defined by y^2 - x^3 - 1
```

```
sage: O = F.maximal_order()
```

```
Maximal order of Function field in y defined by y^2 - x^3 - 1
```

```
sage: P1 = O.ideal(x-2,y-3)
```

$$(x-2, y-3)$$

```
sage: P2 = O.ideal(x-2,y+3)
```

$$(x-2, y+3)$$

```
sage: Rinf = R.maximal_order_infinite().ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

```
sage: Pinf = F.places_above(Rinf)[0]
```

$$\left(\frac{1}{x^2}y\right)$$

```
sage: D1 = P1.divisor()
```

$$(x-2, y-3)$$

```
sage: D2 = P2.divisor()
```

$$(x-2, y+3)$$

```
sage: Dinf = Pinf.divisor()
```

$$\left(\frac{1}{x^2}y\right)$$

```
sage: (Dinf-D1).basis_function_space()
```

$$[]$$

```
sage: (2*Dinf-D1).basis_function_space()
```

$$[x-2]$$

```
sage: (3*Dinf-D1).basis_function_space()
```

$$[x-2, y-3]$$

```
sage: (4*Dinf-D1).basis_function_space()
```

$$\left[x^2-2x, x-2, y-3\right]$$

```
sage: (Dinf).basis_function_space()
```

$$[1]$$

```
sage: (2*Dinf).basis_function_space()
```

$$[x, 1]$$

```
sage: (3*Dinf).basis_function_space()
```

$$[x, 1, y]$$

8.7 Mittag-Leffler Problems

Theorem 8.7 tells us that a rational function on an algebraic curve is completely characterized, up to an additive constant, by the principal parts of the Puiseux expansions at its poles. Note that Theorem 8.7 does not guarantee the existence of a function with specified principal parts. It only shows that any two such functions, *if they exist*, differ by at most a constant.

A *Mittag-Leffler problem* is the practical application of this theorem – given a set of principal parts, find a function that matches them all, or prove that no such function exists.

The first step in solving a Mittag-Leffler problem is to identify the maximum strengths of the poles, and construct a basis for a Riemann-Roch space that includes all functions with poles of such strength. We now have a finite basis for a vector space that must include the function we are looking for. We construct Puiseux expansions for the basis functions, and use them to construct a matrix equation that, when solved, gives the coefficients needed to form the function we seek from the basis functions.

The input data is a set of principal parts or, alternately, a divisor combined with a vector of coefficients.

Let's assume that we've got our data in the latter form, so we can run `riemannroch` on the divisor and obtain a set of basis functions. Now let's construct a Sage function to extract the principal parts of the basis functions and form them into a matrix:

```
def principal_parts_matrix(div, basis):
    F = div.parent().function_field()
    coeffs = [(puiseux(F,p), i) for p,m in div.list() for i in range
    return matrix([[c[0](b)[c[1]] for c in coeffs] for b in basis]).
```

Given a vector `b` of coefficients, we now want to solve a matrix equation:

$$m \cdot v = b$$

This will typically be an overspecified system – a non-square matrix that may or may not have a solution. That's fine; since some integrals have no elementary form, this doesn't represent a limitation in our theory. Failure to solve this matrix equation would only show that no function exists on this curve with the coefficients `b`.

Example 8.24. Let's say that we've identified a divisor on an algebraic curve (example **??**):

We now compute its principal parts matrix:

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field
```

```
sage: L.<y> = R[]
```

Rational function field in x over Algebraic Field$[y]$

```
sage: F.<y> = R.extension(y^2 - x^8 - 1)
```

```
Function field in y defined by y^2 - x^8 - 1
```

```
sage: y.divisor();
```

```
sage: O = F.maximal_order()
```

```
Maximal order of Function field in y defined by y^2 - x^8 - 1
```

```
sage: Oinf = F.maximal_order_infinite()
```

```
Maximal infinite order of Function field in y defined by y^2 - x^8 - 1
```

```
sage: Dfinite = add([O.ideal(x-a*QQbar(-1).sqrt(), y-b*QQbar(2).sqrt())
```

$$\left(x-i,y-\sqrt{2}\right)+\left(x-i,y+\sqrt{2}\right)+\left(x+i,y-\sqrt{2}\right)+\left(x+i,y+\sqrt{2}\right)$$

```
sage: Rinf = R.maximal_order_infinite().ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

```
sage: Dinf = add([pl.divisor() for pl in F.places_above(Rinf)])
```

$$\left(\frac{1}{x},\frac{1}{x^4}y-1\right)+\left(\frac{1}{x},\frac{1}{x^4}y+1\right)$$

```
sage: D1 = Dfinite + 2*Dinf
```

$$2\left(\frac{1}{x},\frac{1}{x^4}y-1\right)+2\left(\frac{1}{x},\frac{1}{x^4}y+1\right)+\left(x-i,y-\sqrt{2}\right)$$
$$+\left(x-i,y+\sqrt{2}\right)+\left(x+i,y-\sqrt{2}\right)+\left(x+i,y+\sqrt{2}\right)$$

```
sage: basis = Dfinite.basis_function_space()
```

$$\left[\frac{x^2}{x^2+1}, \frac{x}{x^2+1}, \frac{1}{x^2+1}\right]$$

```
sage: D1.basis_function_space()
```

$$\left[\frac{x^4}{x^2+1}, \frac{x^3}{x^2+1}, \frac{x^2}{x^2+1}, \frac{x}{x^2+1}, \frac{1}{x^2+1}, \left(\frac{1}{x^2+1}\right) y\right]$$

```
sage: principal_parts_matrix(D1, basis)
```

$$\left(\begin{array}{rrr}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\frac{1}{2} i & \frac{1}{2} & -\frac{1}{2} i \\
\frac{1}{2} i & \frac{1}{2} & -\frac{1}{2} i \\
-\frac{1}{2} i & \frac{1}{2} & \frac{1}{2} i \\
-\frac{1}{2} i & \frac{1}{2} & \frac{1}{2} i
\end{array}\right)$$

TODO

Introduce a sample vector `b` and show how to proceed.

□

8.8 Parallels with the Transcendental Cases

At this point, it may seem that we've spent this entire chapter developing a suite of technical tools that appear completely different from everything that came before them. Why should the algebraic case be so much different from the transcendental cases? What would happen if we used here the same kind of techniques from earlier in the book?

First, the key difference in the algebraic case is the lack of unique factorization. Algebraic extensions are not, in general, unique factorization domains, a classic example being the factorization of 6 into either $3 \cdot 2$ or $(1+\sqrt{-5}) \cdot (1-\sqrt{-5})$ in the ring $\mathbf{Z}[\sqrt{-5}]$. You can show that all four numbers 3, 2, $(1+\sqrt{-5})$ and $(1-\sqrt{-5})$ are all prime in $\mathbf{Z}[\sqrt{-5}]$, so we have two distinct factorizations in this ring.

Show an example in a function field.

The main problem with our earlier tools is the difficulty in defining factorization. How, for example, do you construct a partial fractions expansion? A review of Theorems 6.1 and 7.1 reveals that both depend not merely on the construction of a partial fractions expansions, but also on its *uniqueness*. Without unique factorization, how can you possibly have a unique partial fractions expansion?

The primary goal of this chapter is to develop techniques to carry out the same kinds of operations we did earlier, but without relying on unique factorization.

For example, a principle parts expansion of a function on an algebraic curve is exactly analogous to a partial fractions expansion of a rational function.

Demonstrate

Although we began our development using infinite series expansions, we ultimately concluded that we can completely specify a function (up to an additive constant), using only a finite number of constants – the principle parts coefficients, which turn out to align precisely with the coefficients in a partial fractions expansion.

Reassembling a partial fractions expansion into a rational function is easy – you just promote all the fractions to a common denominator, add up the terms, and cancel any common factors that remain between the numerator and the denominator. Solving a Mittag-Lefler problem is considerably more difficult, but is in principle the same operation – given the principle parts coefficients (resp. the partial fractions expansion), construct a single rational function that matches. The major caveat here is that, unlike reassembling a partial fractions expansion, there might be not solution. Not every principle parts expansion has a matching algebraic function.

Likewise, finding an algebraic function's divisor is exactly analogous to factoring the numerator and denominator of a rational function. You get a finite set of poles and zeros with their locations and multiplicities. Again, in the algebraic case, it's more complicated – you might have singularities with multiple places lying over a single point; the "coordinate" is more complicated that a simple (x, y) coordinate pair, but the principle is the same.

And finding a function with a specified set of poles and zeros is the same as taking a rational function in factored form and multiplying the factors together again. Again, there's a caveat – in the algebraic curve case there might be no solution.

So, if the tools we've developed in this chapter parallel neatly with the tools we used in Chapter 5 to solve integrals of rational functions, can we generalize these tools to handle more complicated transcendental fields, like we did in Chapters 6 and 7? And do we have anything like the Hermite reduction procedure we developed at the end of Chapter 5?

The answers to both of these questions is 'yes'. For the purpose of a clear exposition, I've developed this theory so far in its simplest form, and if you're seeing it for the first time, I suspect that you already appreciate not having met it in its full generality! We can drop the assumption of an algebraically closed coefficient field and lose very little except simplicity; this will be the subject of Chapter 11. Barry Trager showed in [Tr84] how the Hermite reduction can be performed in an algebraic extension; it's now called

Hermite-Trager reduction and I'll present it at the end of Chapter 9.

However, continuing with the intent of presenting the theory in its simplest form first, we'll begin the next chapter by looking at how to use these tools to integrate Abelian integrals, much like we first met partial fractions expansion when learning to integrate in first year calculus, and only later generalized it into a form suitable for integrating in arbitrary transcendental extensions. We'll find that completely solving integrals in even this simplest of algebraic extensions will require a significant excursion into modern algebraic geometry, so much so that the entirety of Chapter 10 will be devoted to proving the book's most exotic theorem.

If there's a lesson to be learned from Chapter 8, though, it's this:

Divisors, principle parts expansions, Riemann-Roch spaces, and Mittag Leffler problems are how we do factorization, partial fractions expansions, and their inverse operations in algebraic extensions where we've lost unique factorization.

Chapter 9

Abelian Integrals

THIS CHAPTER IS INCOMPLETE.

We can now use the machinery developed in the previous chapter to solve Abelian integrals in general.

These techniques, taken together with Liouville's Theorem, provide our method of attack for integration of Abelian integrals. We find all of the poles of the differential, construct Puiseux expansions of the differential there, and split the principal parts easily into two sets. The first order poles arise from logarithm components in the solution; all the higher order poles must come from the rational function. To find the rational function, we integrate term-wise to obtain its principal parts, and then solve a Mittag-Leffler problem to see if such a function exists. For the logarithmic components, it's a little more complicated.

Once we've calculated the principal parts of an integrand, we can integrate the resulting series to obtain the principal parts of the integral. This is possible due to the simple but crucial observation that poles in the integral can only appear where there are poles in the integrand. Once we've determined the principal parts of the integral, we need to solve a Mittag-Leffler problem to find an algebraic function that matches the given principal parts. This can be done by finding a basis for a suitable Riemann-Roch space.

Thus, having identified the locations and orders of the integral's principal parts, we can compute a Riemann-Roch basis for all functions with suitable poles at those locations. Having done so, it then becomes a straightforward exercise in linear algebra to find a combination of those basis functions that match a specified set of principal parts.

9.1 The Abelian Integration Theorem

Theorem 9.1. *Let* $\mathbf{C}$ *be the complex field, let* $\mathbf{C}(x)$ *be the rational function field in* x *over* $\mathbf{C}$*, let* $\mathbf{C}(x,y)$ *be an algebraic extension of* $\mathbf{C}(x)$*, and let* f *be an element of* $\mathbf{C}(x,y)$ *with pole divisor* $(f)_\infty$ *and principle parts expansion:*

$$\forall P \in (f)_\infty \qquad f = \sum_{\nu_P < i < 0} a_{P,i}\, t^i + \cdots$$

If f *has an elementary anti-derivative* F*, then* $F \in \mathbf{C}(x,y,\Psi)$*, where* $\mathbf{C}(x,y,\Psi)$ *is a finite logarithm extension of* $\mathbf{C}(x,y)$*,* F *can be written as the sum of an algebraic function in* $A \in \mathbf{C}(x,y)$ *and logarithms of algebraic functions* $B_j \in \mathbf{C}(x,y)$*:*

$$F = A + \sum_j c_j \ln B_j$$

A's pole divisor is a subset of f*'s pole divisor, and its principle parts expansion has the form:*

$$\forall P \in (f)_\infty \qquad A = \sum_{(\nu_P+1) < i < 0} A_{P,i}\, t^i + \cdots$$

$$A_{P,i} = \frac{1}{i} a_{P,i-1}$$

(what about ramification and infinity?)

The residues $a_{P,-1}$ *can be written in terms of a* $\mathbf{Z}$*-module in* $\mathbf{C}$*, with basis* $\{c_j\}$*:*

$$a_{P,-1} = \sum_j b_{P,j} c_j \qquad b_{P,j} \in \mathbf{Z}$$

and the divisors of the B_j *can be written:*

$$(B_j) = \sum_P b_{P,j} P$$

(and any such basis is as good as any other)

Proof

By Theorem 4.15, an elementary antiderivative of f can only exist in a finite logarithm extension $\mathbf{C}(x, y, \Psi)$ of $\mathbf{C}(x, y)$ and therefore must have the form:

$$F = A + \sum_j c_j \ln B_j$$

where $A, B_j \in \mathbf{C}(x, y)$, and the c_j are constants. Differentiating, we obtain:

$$f = \frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\mathrm{d}A}{\mathrm{d}x} + c_j \frac{\frac{\mathrm{d}B_j}{\mathrm{d}x}}{B_j}$$

Consider a place P with uniformizing variable s.

$$f = \frac{\mathrm{d}A}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}x} + c_j \frac{\frac{\mathrm{d}B_j}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}x}}{B_j}$$

Let A's Pusieux expansion at P be $A_a s^a + A_{a+1} s^{a+1} + \cdots$, let B_j's Pusieux expansions at P be $B_{j,k} s^{k_j} + B_{j,k+1} s^{k_j+1} + \cdots$, and let x's Pusieux expansion at P be $C_c s^c + C_{c+1} s^{c+1} + \cdots$. Substituting:

$$f = \frac{aA_a s^{a-1} + (a+1)A_{a+1}s^a + \cdots}{cC_c s^{c-1} + (c+1)C_{c+1}s^c + \cdots}$$

$$+ \frac{k_j B_{j,k} s^{k_j-1} + (k_j+1)B_{j,k+1}s^{k_j} + \cdots}{cC_c s^{c-1} + (c+1)C_{c+1}s^c + \cdots} \frac{1}{B_{j,k}s^{k_j} + B_{j,k+1}s^{k_j+1} + \cdots}$$

□

9.2 Some Simple Examples

Example 9.2. Evaluate $\int \frac{x}{\sqrt{x^2+1}}\,dx$

This is a simple integral that can be easily solved using first year calculus techniques, but let's see how to attack it using the more sophisticated techniques of this chapter.

First, we convert the the integrand into a rational function on an algebraic curve:

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 - (x^2+1))


      Function field in y defined by y^2 - x^2 - 1
```

Next, we identify the poles of the integrand. The finite poles can only be located where the denominator is zero, which is where $x^2 + 1 = 0$, over the points $x = \pm i$.

Let's compute the principal parts of $\frac{x}{y}$ at $x = i$. Calling `puiseux` with `deg=-1` computes just the principal part of the differential:

```
sage: D = (x/y).divisor()
```

$$-(x-i,y)+(x,y-1)+(x,y+1)-(x+i,y)$$

```
sage: p = D.support()[0]
```

$$(x-i,y)$$

```
sage: m = F.completion(p, prec=4)

Completion map:
  From:  Function field in y defined by y^2 - x^2 - 1
  To:      Laurent Series Ring in s over Algebraic Field
```

```
sage: m(x/y)
```

$$\frac{i}{s} - \frac{1}{2}is + O(s^3)$$

There appears to be a pole here, but appearances are deceptive. We need to expand the *differential*, not the *integrand*:

```
sage: m(x/y) * m(x).derivative()
```

$$1 + O(s^2)$$

So, even though the *integrand* has a pole at $(i, 0)$, the *differential* does not... and the differential is what matters!

The only other place we might have a pole is infinity.

```
sage: D2 = (x/y).divisor() + F(x).differential().divisor()
```

$$-2\left(\frac{1}{x}, \frac{1}{x}y - 1\right) - 2\left(\frac{1}{x}, \frac{1}{x}y + 1\right) + (x, y - 1) + (x, y + 1)$$

```
sage: F.maximal_order_infinite().decomposition()
```

$$\left[\left(\left(\frac{1}{x}, \frac{1}{x}y - 1\right), 1, 1\right), \left(\left(\frac{1}{x}, \frac{1}{x}y + 1\right), 1, 1\right)\right]$$

Now it appears that we have two sheets with no poles, the expansions indicating simply that $\lim_{x\to\infty} \frac{x}{\sqrt{x^2+1}} = \pm 1$, depending on whether we use the positive or negative square root, but again we have to take the differential into account. Since $x = \frac{1}{t}$, $dx = -\frac{1}{t^2}dt$, and we actually have second order poles at infinity.

```
sage: m2 = F.completion(D2.support()[0], prec=2)

Completion map:
  From:  Function field in y defined by y^2 - x^2 - 1
  To:      Laurent Series Ring in s over Algebraic Field

sage: m3 = F.completion(D2.support()[1], prec=2)

Completion map:
  From:  Function field in y defined by y^2 - x^2 - 1
  To:      Laurent Series Ring in s over Algebraic Field
```

```
sage: m2(x/y) * m2(x).derivative()
```

$$\frac{-1}{s^2} + O(1)$$

```
sage: m3(x/y) * m2(x).derivative()
```

$$\frac{1}{s^2} + O(1)$$

Integrating termwise, we see that since our differential has second order poles at infinity, our integral must have first order poles at infinity, and theorem ? states that this completely characterizes the integral.

What functions have first order poles at infinity and nowhere else?

```
sage: D3 = add([-(m+1)*p for p,m in D2.list() if m < 0])
```

$$\left(\frac{1}{x}, \frac{1}{x}y - 1\right) + \left(\frac{1}{x}, \frac{1}{x}y + 1\right)$$

```
sage: D3.basis_function_space()
```

$$[x, 1, y]$$

Our solution, if it exists, is in the vector space spanned by these three basis elements. 1 is in the list, and we expect it to be there, because of the presence of the constant of integration, so we can always add a multiple of 1 to our solution and get another solution.

What about x and y? We want to combine them in such a way as to match the principal parts of our differential.

Let's expand them:

```
sage: [m2(x), m3(x)]
```

$$\left[\frac{1}{s} + O(s), \frac{1}{s} + O(s)\right]$$

```
sage: [m2(y), m3(y)]
```

$$\left[\frac{1}{s} + O(s), \frac{-1}{s} + O(s)\right]$$

No linear algebra games are required to see that y will match the principal parts of the differential if it is itself differentiated. Therefore, $y = \sqrt{x^2+1}$ is our solution.

□

Example 9.3. Compute $\int \frac{1}{\sqrt{1-x^2}}\,dx$

This is a familiar example from first year calculus, but let's approach it using the techniques of this book. We'll use the algebraic extension $y^2 = 1 - x^2$ and integrate $\frac{1}{y}\,dx$.

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 - (1-x^2))

     Function field in y defined by y^2 + x^2 - 1

sage: integrand = 1/y * x.differential()
```

$$\left(\left(\frac{-1}{x^2-1}\right)y\right)\,dx$$

```
sage: D = (1/y).divisor() + F(x).differential().divisor()
```

$$-\left(\frac{1}{x},\frac{1}{x}y-i\right)-\left(\frac{1}{x},\frac{1}{x}y+i\right)$$

```
sage: table([[pl, F.completion(pl)(1/y, prec=2) * F.completion(pl)(x
         for pl,m in D.list() if m < 0])
```

$$\begin{array}{ll}
\left(\frac{1}{x},\frac{1}{x}y-i\right) & \frac{i}{s}+O(s) \\
\left(\frac{1}{x},\frac{1}{x}y+i\right) & \frac{-i}{s}+O(s)
\end{array}$$

Our only poles are at infinity, and they're first order poles, so this solution will be a logarithm.

The poles' residues (coefficients) are $-i$ and i. These exist in the field $\mathbf{Q}[i]$, 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), with the signs flipped due to the effect of differentiation of a negative power.

So, we have a singular point at infinity, and we want a function with a simple zero on one cycle and a simple pole on the other. The basis will have either one element or no elements, depending on whether an algebraic function exists with the desired properties.

```
sage: def logarithmic_divisor(differential, component):
          return add([- ZZ(differential.residue(pl) / component) * pl.d
```

```
sage: D = logarithmic_divisor(integrand, QQbar(-1).sqrt())
```

$$-\left(\frac{1}{x}, \frac{1}{x}y - i\right) + \left(\frac{1}{x}, \frac{1}{x}y + i\right)$$

```
sage: D.basis_function_space()
```

$$[y - ix]$$

Yes, it does exist. Remembering that our residues came multiplied by a factor of i, we conclude that our solution is $i\ \ln(y - ix)$, or:

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

where I used the negative of a logarithm being the logarithm of the inverse, and the last transformation came from section 4.2.

□

Example 9.4. 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 with its other sides 2 and x, but we'll attack this integral using the methods of this chapter.

First, transform the problem into an algebraic curve:

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 - (4-x^2))

      Function field in y defined by y^2 + x^2 - 4
```

With no denominator, there can be no poles at finite points, so we just need to check infinity:

```
sage: integrand = y * x.differential()
```

$$(y)\ dx$$

```
sage: D = sum([-m * pl.divisor() for pl,m in integrand.divisor().lis
```

$$3\left(\frac{1}{x}, \frac{1}{x}y - i\right) + 3\left(\frac{1}{x}, \frac{1}{x}y + i\right)$$

```
sage: table([[pl, F.completion(pl)(y, prec=2) * F.completion(pl)(x)
         for pl in D.support()])
```

$$\begin{array}{ll} \left(\frac{1}{x}, \frac{1}{x}y - i\right) & \frac{-i}{s^3} + O(s^{-1}) \\ \left(\frac{1}{x}, \frac{1}{x}y + i\right) & \frac{i}{s^3} + O(s^{-1}) \end{array}$$

Since these polar expansions have components at both t^{-3} and t^{-1}, we'll get a result with both algebraic and logarithmic components.

Let's start with the algebraic part. The integration step lowers the order of the poles by one, so we want a divisor at the same places as the divisor of the poles, but with multiplicity one less:

```
def reduced_divisor(D):
    return D - add([pl.divisor() for pl in D.support()])
```

```
sage: D2 = reduced_divisor(D)
```

$$2\left(\frac{1}{x}, \frac{1}{x}y - i\right) + 2\left(\frac{1}{x}, \frac{1}{x}y + i\right)$$

We need a Riemann-Roch space with at most second order poles at infinity:

```
sage: basis = D2.basis_function_space()
```

$$\left[x^2, x, 1, xy, y\right]$$

Now let’s construct a matrix of principal parts:

```
sage: m = principal_parts_matrix(D2, basis)
```

$$\begin{pmatrix} 1 & 0 & 0 & i & 0 \\ 0 & 1 & 0 & 0 & i \\ 1 & 0 & 0 & -i & 0 \\ 0 & 1 & 0 & 0 & -i \end{pmatrix}$$

Another Sage function will extract the principal parts of the differential, and divide them by the powers needed for term-wise integration:

```
def solution_vector(div, differential):
    F = differential.parent().function_field()
    coeffs = [(puiseux(F, p), i) for p,m in div.list() for i in range(-m
    return matrix([[c[0](differential)[c[1]-1]/c[1] for c in coeffs]]).t
```

```
sage: b = solution_vector(D2, integrand)
```

$$\begin{pmatrix} \frac{1}{2}i \\ 0 \\ -\frac{1}{2}i \\ 0 \end{pmatrix}$$

```
sage: v = m.solve_right(b)
```

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

```
sage: m * v == b
```

$$\mathrm{True}$$

```
sage: algebraic_solution = (matrix(basis) * v)[0,0]
```

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

This is the rational part of our answer. To find the logarithmic part, let's return to the series expansion of the differential, note that the residues are $-2i$ and $2i$, which can be placed in the Z-module $\{2i\}$, and search for:

```
sage: pls = D.support()
```

$$\left[\left(\frac{1}{x}, \frac{1}{x}y - i\right), \left(\frac{1}{x}, \frac{1}{x}y + i\right)\right]$$

```
sage: basis2 = (pls[0].divisor() - pls[1].divisor()).basis_function_
```

$$[y + ix]$$

```
sage: log_solution = basis2[0]
```

$$y + ix$$

The logarithmic component of our solution is thus:

```
sage: var('x')
```

$$x$$

```
sage: F = FractionField(PolynomialRing(QQbar, x))
```

$$\mathrm{Frac}(\overline{\mathbf{Q}}[x])$$

```
sage: log_solution.element().change_ring(F).subs({y:sqrt(4-x^2)})
```

$$i\,x + \sqrt{-x^2 + 4}$$

Remembering from the previous example that $i\ln\bigl(ix - \sqrt{1-x^2}\bigr) = \arcsin x$, and that logarithmic components are only specified up to a multiplicative constant (it disappears into the constant of integration), we can rewrite this:

$$2i\ln\left(\sqrt{4-x^2} - ix\right) = 2i\ln\left(\sqrt{4-4\left(\frac{x}{2}\right)^2} - 2i\left(\frac{x}{2}\right)\right)$$

$$= 2i\ln\left(i\frac{x}{2} - \sqrt{1-\left(\frac{x}{2}\right)^2}\right) = 2\arcsin\frac{x}{2}$$

Adding in the rational component we computed earlier, the final answer is:

$$\int \sqrt{4-x^2}\,dx = 2\arcsin\frac{x}{2} + \frac{x\sqrt{4-x^2}}{2}$$

□

9.3 An integral Sage can't solve

Example 9.5. Integrate $\int \frac{x^9+2x^7-x}{(x^4+2x^2+1)\sqrt{x^8+1}}\,dx$

When I say that Sage can't solve this integral, I mean that its built-in integration routine can't solve the integral:

```
sage: y = sqrt(x^8+1);
sage: integrand = (x^9+2*x^7-x)/((x^4+2*x^2+1)*y)
```

$$\frac{x^9 + 2\,x^7 - x}{\sqrt{x^8+1}(x^4 + 2\,x^2 + 1)}$$

```
sage: integrate(integrand,x)
```

$$\int \frac{(x^8 + 2\,x^6 - 1)x}{\sqrt{x^8+1}(x^2+1)^2}\,dx$$

Now let's attack the problem using the techniques of this book.

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]
```

Rational function field in x over Algebraic Field$[y]$

```
sage: F.<y> = R.extension(y^2 - (x^8+1))

     Function field in y defined by y^2 - x^8 - 1

sage: integrand = (x^9+2*x^7-x)/((x^4+2*x^2+1)*y)
```

$$\left(\frac{x^9 + 2x^7 - x}{x^{12} + 2x^{10} + x^8 + x^4 + 2x^2 + 1}\right) y$$

Where are our poles?

```
def differential_divisor_of_poles(differential):
    return sum([-m * pl.divisor() for pl,m in differential.divisor()
sage: D = differential_divisor_of_poles(integrand * x.differential())
```

$$3\left(\frac{1}{x}, \frac{1}{x^4}y - 1\right) + 3\left(\frac{1}{x}, \frac{1}{x^4}y + 1\right) + 2\left(x - i, y - \sqrt{2}\right) \\ + 2\left(x - i, y + \sqrt{2}\right) + 2\left(x + i, y - \sqrt{2}\right) + 2\left(x + i, y + \sqrt{2}\right)$$

```
sage: table([[pl, puiseux(F, pl)(integrand * x.differential())] for pl i
```

$$\begin{array}{cc}
\left(\frac{1}{x}, \frac{1}{x^4}y - 1\right) & \frac{-1}{s^3} + O(1) \\
\left(\frac{1}{x}, \frac{1}{x^4}y + 1\right) & \frac{1}{s^3} + O(1) \\
\left(x - i, y - \sqrt{2}\right) & \frac{\frac{1}{2}i\sqrt{\frac{1}{2}}}{s^2} + O(1) \\
\left(x - i, y + \sqrt{2}\right) & \frac{-\frac{1}{2}i\sqrt{\frac{1}{2}}}{s^2} + O(1) \\
\left(x + i, y - \sqrt{2}\right) & \frac{-\frac{1}{2}i\sqrt{\frac{1}{2}}}{s^2} + O(1) \\
\left(x + i, y + \sqrt{2}\right) & \frac{\frac{1}{2}i\sqrt{\frac{1}{2}}}{s^2} + O(1)
\end{array}$$

We've found four second order poles at the ordinary points $(\pm i, \pm\sqrt{2})$, as well as two third order poles at a singular point with two sheets at infinity.

Our next goal is to construct a basis for a suitable Riemann-Roch space. We invert the signs of the finite poles, since the convention for Riemann Roch spaces is their functions must have order greater than the *negative* of a divisor, remember that poles decrease in order by 1 when they are integrated, and conclude that the Riemann-Roch space that we're interested in is:

$$\mathcal{L}(Z(i, \sqrt{2})Z(-i, \sqrt{2}), Z(i, -\sqrt{2})Z(-i, -\sqrt{2})Z^2(\infty, \infty))$$

i.e, all the functions on our algebraic curve with at most first order poles at $(\pm i, \pm\sqrt{2})$, no other finite poles, and at most second order poles at infinity.

```
sage: D = reduced_divisor(D)
```

$$2\left(\frac{1}{x}, \frac{1}{x^4}y - 1\right) + 2\left(\frac{1}{x}, \frac{1}{x^4}y + 1\right) + \left(x - i, y - \sqrt{2}\right) + \left(x - i, y + \sqrt{2}\right) + \left(x + i, y - \sqrt{2}\right) + \left(x + i, y + \sqrt{2}\right)$$

We now wish to see if a linear combination of these basis functions will match the poles in the differential. This problem is a bit more complicated than the last one, so let's use the tools we developed in the last chapter for solving Mittag-Leffler problems:

```
sage: basis = D.basis_function_space()
```

$$\left[\frac{x^4}{x^2 + 1}, \frac{x^3}{x^2 + 1}, \frac{x^2}{x^2 + 1}, \frac{x}{x^2 + 1}, \frac{1}{x^2 + 1}, \left(\frac{1}{x^2 + 1}\right) y\right]$$

```
sage: m = principal_parts_matrix(D, basis)
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{2}i & -\frac{1}{2} & \frac{1}{2}i & \frac{1}{2} & -\frac{1}{2}i & -\frac{1}{2}i\sqrt{2} \\ -\frac{1}{2}i & -\frac{1}{2} & \frac{1}{2}i & \frac{1}{2} & -\frac{1}{2}i & \frac{1}{2}i\sqrt{2} \\ \frac{1}{2}i & -\frac{1}{2} & -\frac{1}{2}i & \frac{1}{2} & \frac{1}{2}i & \frac{1}{2}i\sqrt{2} \\ \frac{1}{2}i & -\frac{1}{2} & -\frac{1}{2}i & \frac{1}{2} & \frac{1}{2}i & -\frac{1}{2}i\sqrt{2} \end{pmatrix}$$

```
sage: b = solution_vector(D, integrand * x.differential())
```

$$\begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ 0 \\ -\frac{1}{2}i\sqrt{\frac{1}{2}} \\ \frac{1}{2}i\sqrt{\frac{1}{2}} \\ \frac{1}{2}i\sqrt{\frac{1}{2}} \\ -\frac{1}{2}i\sqrt{\frac{1}{2}} \end{pmatrix}$$

Now we have a matrix equation that we want to solve:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{1}{2}i & -\frac{1}{2} & \frac{1}{2}i & \frac{1}{2} & -\frac{1}{2}i & -\frac{1}{2}i\sqrt{2} \\ -\frac{1}{2}i & -\frac{1}{2} & \frac{1}{2}i & \frac{1}{2} & -\frac{1}{2}i & \frac{1}{2}i\sqrt{2} \\ \frac{1}{2}i & -\frac{1}{2} & -\frac{1}{2}i & \frac{1}{2} & \frac{1}{2}i & \frac{1}{2}i\sqrt{2} \\ \frac{1}{2}i & -\frac{1}{2} & -\frac{1}{2}i & \frac{1}{2} & \frac{1}{2}i & -\frac{1}{2}i\sqrt{2} \end{pmatrix} \cdot v = \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ 0 \\ -\frac{1}{2}i\sqrt{\frac{1}{2}} \\ \frac{1}{2}i\sqrt{\frac{1}{2}} \\ \frac{1}{2}i\sqrt{\frac{1}{2}} \\ -\frac{1}{2}i\sqrt{\frac{1}{2}} \end{pmatrix}$$

To find out if there actually is a solution, we simply have to check:

```
sage: v = m.solve_right(b)
```

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix}$$

```
sage: m * v == b
```

$$\mathrm{True}$$

So, yes, this system does have a solution. Now let's multiply our solution vector by the original basis:

```
sage: solution = (matrix(basis) * v)[0,0]
```

$$\left(\frac{\frac{1}{2}}{x^2 + 1}\right) y$$

...and convert back to our original form:

```
sage: var('x')
```

$$x$$

```
sage: F = FractionField(PolynomialRing(QQbar, x))
```

$$\mathrm{Frac}(\overline{\mathbf{Q}}[x])$$

```
sage: solution = solution.element().change_ring(F).subs({y:sqrt(x^8
```

$$\frac{\frac{1}{2}\,\sqrt{x^{8}+1}}{x^{2}+1}$$

Now we can verify the solution:

```
sage: integrand = integrand.element().change_ring(F).subs({y:sqrt(x
```

$$\frac{\left(x^{9}+2\,x^{7}-x\right)\sqrt{x^{8}+1}}{x^{12}+2\,x^{10}+x^{8}+x^{4}+2\,x^{2}+1}$$

```
sage: bool(solution.diff(x) == integrand)
```

True

□

9.4 Geddes's example

Example 9.6. Compute $\int \frac{1}{x\sqrt{x^4+1}}\,dx$

Two traditional solution techniques are either the substitution $x^2 = \tan u$, followed by the half angle formula for $\tan$, or $x^4+1 = u^2$, which leads to a rational function and a partial fractions expansion.

We'll use $\mathbf{C}(x, y); y^2 = x^4 + 1$ and integrate $\frac{1}{xy}$.

```
sage: R.<x> = FunctionField(QQbar)

    Rational function field in x over Algebraic Field

sage: L.<y> = R[]

   Rational function field in x over Algebraic Field[y]

sage: F.<y> = R.extension(y^2 - (x^4+1))

       Function field in y defined by y^2 - x^4 - 1

sage: integrand = 1/(x*y) * x.differential()
```

$$\left(\left(\frac{1}{x^5+x}\right)y\right)\,dx$$

```
sage: D = differential_divisor_of_poles(integrand)
```

$$(x, y-1) + (x, y+1)$$

```
sage: table([[pl, puiseux(F, pl)(integrand)] for pl in D.support()])
```

$$\begin{array}{ll} (x, y-1) & \frac{1}{s} + O(1) \\ (x, y+1) & \frac{-1}{s} + O(1) \end{array}$$

Our only poles are a pair of first order poles at $x = 0$. This will give rise to a logarithmic term; can we find a function that matches?

```
sage: pls = D.support()
```

$$[(x, y-1), (x, y+1)]$$

```
sage: D2 = (pls[0].divisor() - pls[1].divisor())
```

$$(x, y - 1) - (x, y + 1)$$

```
sage: D2.basis_function_space()
```

$$[]$$

Since we can't find a function that matches that divisor; let's try doubling the strength of the poles.

```
sage: log_solution = (2*D2).basis_function_space()[0]
```

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

This is the function we are looking for inside our logarithm.

```
sage: var('x')
```

$$x$$

```
sage: F = FractionField(PolynomialRing(QQbar, x))
```

$$\mathrm{Frac}(\overline{\mathbf{Q}}[x])$$

```
sage: 1/2*log(log_solution.element().change_ring(F).subs({y:sqrt(x^4
```

$$\frac{1}{2}\log\left(\frac{\sqrt{x^4+1}}{x^2} + \frac{1}{x^2}\right)$$

Sage's built-in integrator produces the result in a different form:

```
sage: integrate(1/(x*sqrt(x^4+1)),x)
```

$$-\frac{1}{4}\log\left(\sqrt{x^4+1}+1\right) + \frac{1}{4}\log\left(\sqrt{x^4+1}-1\right)$$

We can obtain this result from our algorithm by factoring $\frac{1}{4}$ out of the residues and using fourth-order poles:

```
sage: (4*D2).basis_function_space()
```

$$\left[\frac{1}{x^4}y + \frac{\frac{1}{2}x^4 + 1}{x^4}\right]$$

which is another way of writing $\frac{y-1}{y+1}$ (remember that $y^2 = x^4 + 1$).

We conclude that:

$$\int \frac{1}{x\sqrt{x^4+1}}\,dx = \frac{1}{2}\ln\frac{\sqrt{x^4+1}-1}{x^2} = \frac{1}{4}\ln\frac{\sqrt{x^4+1}-1}{\sqrt{x^4+1}+1}$$

□

9.5 Holliman’s Integral

Compute:

$$\int \frac{1}{(x^2+1)^{3/4}}\,dx$$

```
sage: R.<x> = FunctionField(QQbar)

   Rational function field in x over Algebraic Field

sage: L.<y> = R[]

  Rational function field in x over Algebraic Field[y]

sage: root = x^2+1
```

$$x^2+1$$

```
sage: F.<y> = R.extension(y^4 - root)

     Function field in y defined by y^4 - x^2 - 1

sage: integrand = y/(x^2+1) * x.differential()
```

$$\left(\left(\frac{1}{x^2+1}\right)y\right)\,dx$$

```
sage: D = integrand.divisor()
```

$$0$$

Any non-constant function will have poles that will increase in order under derivation, and any constant function will have a zero differential, which has no divisor.

Therefore, this integrand is not an exact differential (i.e, it is not the differential of a function), and the integral is non-elementary.

If we're not just going to take the computer's word that the integrand's divisor is zero, how can we compute it by hand?

First, looking at the integrand, we see that it's only possible finite poles are over the ramification points of the root, and we know that y is a uniformizing variable at those points. So, changing variables to y, we see...

$$y^4 = (x^2+1) \qquad \Longrightarrow \qquad 4y^3\,dy = 2x\,dx \qquad \Longrightarrow \qquad dx = \frac{2y^3}{x}dy$$

$$\int \frac{y}{x^2+1}dx = \int \frac{y}{x^2+1}\frac{2y^3}{x}dy = \int \frac{2}{x}dy$$

The ramification points are at $x = \pm i$, so the denominator (x) is finite and not zero, so there are no poles at the ramification points.

What about infinity? Here, $\frac{1}{y}$ is a uniformizing variable (WHY?), so let $v = \frac{1}{y}$, and

$$\int \frac{2}{x}\,dy = \int \frac{2}{x}\left(-\frac{1}{v^2}dv\right) = -\int \frac{2}{xv^2}\,dv$$

$$y^4 = x^2+1 \qquad \Longrightarrow \qquad \frac{1}{v^4} = x^2+1 \qquad \Longrightarrow \qquad \frac{1}{x} = \frac{v^2}{\sqrt{v^4-1}}$$

$$\frac{1}{x} = \pm\, v^2\left[i + \frac{i}{2}v^4 + \cdots\right] \qquad \Longrightarrow \qquad \frac{2}{xv^2}\,dv = \pm\left[2i + iv^4 + \cdots\right]\,dv$$

so we see that the differential has no pole at infinity, just as the computer reported.

We've got a singularity at infinity with two places, $\frac{1}{x}$ being doubly ramified at both:

```
sage: Rinf = R.maximal_order_infinite().ideal(1/x).place()
```

$$\left(\frac{1}{x}\right)$$

```
sage: pls = [pl for pl in F.places_above(Rinf)]
```

$$\left[\left(\frac{1}{x}y, \frac{1}{x}y^2 - 1\right), \left(\frac{1}{x}y, \frac{1}{x}y^2 + 1\right)\right]$$

```
sage: Dinf = add([pl.divisor() for pl in F.places_above(Rinf)])
```

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

```
sage: F.completion(pls[0])(1/x)
```

$$s^2 - \frac{1}{2}s^6 + \frac{7}{8}s^{10} - \frac{33}{16}s^{14} + \frac{715}{128}s^{18} + O(s^{22})$$

```
sage: F.completion(pls[1])(1/x)
```

$$-s^2 + \frac{1}{2}s^6 - \frac{7}{8}s^{10} + \frac{33}{16}s^{14} - \frac{715}{128}s^{18} + O(s^{22})$$

Just checking that we've got the same result as from the kash code:

```
sage: F.completion(pls[0])(1/x)(F.completion(pls[0])(1/y).reverse())
```

$$s^2 + \frac{1}{2}s^6 + \frac{3}{8}s^{10} + \frac{5}{16}s^{14} + \frac{35}{128}s^{18} + O(s^{22})$$

```
sage: F.completion(pls[1])(1/x)(F.completion(pls[1])(1/y).reverse())
```

$$-s^2 - \frac{1}{2}s^6 - \frac{3}{8}s^{10} - \frac{5}{16}s^{14} - \frac{35}{128}s^{18} + O(s^{22})$$

9.6 Chebyshev’s Integral

Example 9.7. Compute:

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

```
sage: R.<x> = FunctionField(QQbar)

    Rational function field in x over Algebraic Field

sage: L.<y> = R[]

   Rational function field in x over Algebraic Field[y]

sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 + 4x^3 + 2x^2 + 1$$

```
sage: F.<y> = R.extension(y^2 - root)

Function field in y defined by y^2 - x^4 - 4*x^3 - 2*x^2 - 1

sage: num = 6*x^2 + 5*x +7
```

$$6x^2 + 5x + 7$$

```
sage: den = 2*x^6 + 8*x^5 + 3*x^4 + - 4*x^3 - 1
```

$$2x^6 + 8x^5 + 3x^4 - 4x^3 - 1$$

```
sage: integrand = y*num/den * x.differential();
sage: D = differential_divisor_of_poles(integrand)
```

$$\left(x - \frac{1}{2}\sqrt{2}, y - \sqrt{2} - \frac{1}{2}\right) + \left(x - \frac{1}{2}\sqrt{2}, y + \sqrt{2} + \frac{1}{2}\right)$$
$$+ \left(x + \frac{1}{2}\sqrt{2}, y - \sqrt{2} + \frac{1}{2}\right) + \left(x + \frac{1}{2}\sqrt{2}, y + \sqrt{2} - \frac{1}{2}\right)$$

```
sage: table([[pl, puiseux(F,pl)(integrand)] for pl in D.support()])
```

$$\begin{array}{cc}
\left(x - \frac{1}{2}\sqrt{2}, y - \sqrt{2} - \frac{1}{2}\right) & \frac{\frac{5}{2}}{s} + O(1) \\
\left(x - \frac{1}{2}\sqrt{2}, y + \sqrt{2} + \frac{1}{2}\right) & \frac{-\frac{5}{2}}{s} + O(1) \\
\left(x + \frac{1}{2}\sqrt{2}, y - \sqrt{2} + \frac{1}{2}\right) & \frac{-\frac{5}{2}}{s} + O(1) \\
\left(x + \frac{1}{2}\sqrt{2}, y + \sqrt{2} - \frac{1}{2}\right) & \frac{\frac{5}{2}}{s} + O(1)
\end{array}$$

We have poles at $x = \pm\frac{1}{\sqrt{2}}$ and nowhere else.

Now we can see that all of our residues are $\pm\frac{5}{2}$, so we need only a single logarithmic term. It must have two poles and two zeros, at the coordinates we just calculated, and the obvious choice is for each of these poles and zero to have degree one. Let's use our Riemann-Roch basis space algorithm to see if such a function exists:

```
sage: D2 = add([QQ(integrand.residue(pl)).sign() * pl for pl in D.su
```

$$\left(x - \frac{1}{2}\sqrt{2}, y - \sqrt{2} - \frac{1}{2}\right) - \left(x - \frac{1}{2}\sqrt{2}, y + \sqrt{2} + \frac{1}{2}\right) - \left(x + \frac{1}{2}\sqrt{2}, y - \sqrt{2} + \frac{1}{2}\right) + \left(x + \frac{1}{2}\sqrt{2}, y + \sqrt{2} - \frac{1}{2}\right)$$

```
sage: D2.basis_function_space()
```

$$[]$$

No such function exists. However, a function may exist with higher degree poles and zero. Let's see... how about a function with second degree poles and zeros? Third degree?

```
sage: (2*D2).basis_function_space()
```

$$[]$$

```
sage: (3*D2).basis_function_space()
```

$$[]$$

```
sage: (4*D2).basis_function_space()
```

$$[]$$

```
sage: res = (5*D2).basis_function_space()
```

$$\Bigg[\Bigg(\frac{8832144405782038592248267177275666366163674962715995632074977306653940238509894494379110300209483153718734316011631386}{54452646977767999887956444300007264946891217495505640729169989333675512121334073234070902649759444530540734772537783}$$

$$+\frac{30176493386421965190181579522358526751059222789279651742922839131067629148242139522461960192382400775205675579706407237}{1856835261941888796179314750630247734688990516596742348864696636278334963337491897281817780356797058491439055743538400}$$

Turns out that fifth degree poles and zeros give us our solution. Leaving aside for a moment the question of finding this degree in some other way than multiplying the degree of the divisor by every larger integers (and how would we ever know when to stop?), let's continue with the problem by coaxing our result into a more reasonable form:

```
sage: num2 = R._to_bivariate_polynomial(res[0])[0]
```

$$\frac{30176493386421965190181579522358526751059222789279651742922839131067629148242139522}{1856835261941888796179314750630247734688990516596742348864696636278334963337491897} + \frac{8832144405782038592248267177275666366163674962715995632074977306653940238509894494}{54452646977767999887956444300007264946891217495505640729169989333675512121334073} + \frac{2649643321734611577674480153182699909849102488814798689622493191996182071552968348}{27226323488883999943978222150003632473445608747752820364584994666837756060667036} + \frac{1510296693388728599274453687314138948613988418624435253084821119437823780785191958}{23210440774273609952241434382878096683612381457459279360808707953479187041718648} + \frac{1168787109698489773707520689792813182455659653399416755311255330247204758229476038}{5989791167554479987675208873000799144158033924505620480208698826704306333346748} + \frac{1207059735456878607607263180894341070042368911571186069716913565242705165929685580}{14974477918886199969188022182501997860395084811264051200521747066760765833366870} + \frac{7498490600508950764818778833507040744872960043345880291631655733349195262494900425}{46420881548547219904482868765756193367224762914918558721617415906958374083437297} + \frac{3211956515569401368047619830135917335161523128107717078197933413853149600071431631}{928417630970944398089657375315123867344449525829837117443234831813916748166874594} + \frac{1325704875307883992696464903309077521561167611903670944374454093728756429800335163}{18568352619418887961793147506302477346889905165967423488646966362783349633374918} + \frac{3385655355549781460361835751289005440362742069041131658962074634217343758095459556}{265262180277412685168473535804321104955570073799534621266385233754047851905355598} + \frac{4613323427953484824684344888930323065259492888858655051820496479842241451248334890}{928417630970944398089657375315123867344449525829837117443234831813916748166874594} + \frac{1148178772751665016992274733045836627601277745153079432169747049865012231006286284}{11605220387136804976120717191439048341806190728729639680404353976739593520859324} + \frac{6682989267041742534801188830805254217063847388455103361603399495368148113805820167}{265262180277412685168473535804321104955570073799534621266385233754047851905355598} + \frac{1472024067630339765374711196212611061027279160452665938679162884442323373084982415}{928417630970944398089657375315123867344495258298371174432348318139167481668745948} + \frac{1472024067630339765374711196212611061027279160452665938679162884442323373084982415}{66315545069353171292118383951080276238892518449883655316596308438511962976338996} + \frac{2944048135260679530749422392425222122054558320905331877358325768884646746169964831}{663155450693531712921183839510802762388925184498836553165963084385119629763389963} + \frac{2944048135260679530749422392425222122054558320905331877358325768884646746169964831}{105502003519425499782915610831264075834601733897542178912766854333996304735084766} + \frac{2944048135260679530749422392425222122054558320905331877358325768884646746169964831}{928417630970944398089657375315123867344495258298371174432348318139167481668745948}$$

```
sage: den2 = R._to_bivariate_polynomial(res[0])[1]
```

$$x^{10} - \frac{5}{2}x^{8} + \frac{5}{2}x^{6} - \frac{5}{4}x^{4} + \frac{5}{16}x^{2} - \frac{1}{32}$$

```
sage: ld = lcm([QQ(c).denominator() for c in den2.coefficients()]);
```

```
sage: ln = lcm([QQ(c).denominator() for c in num2.coefficients()]);
```

```
sage: R2.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: F2 = Frac(R2)
```

$$\mathrm{Frac}(\mathbf{Q}[x])$$

```
sage: R3.<y> = F2[]
```

$$\mathrm{Frac}(\mathbf{Q}[x])[y]$$

```
sage: f = y^2 - (x^4 + 4*x^3 + 2*x^2 + 1)
```

$$y^2 - x^4 - 4x^3 - 2x^2 - 1$$

```
sage: Q = R3.quo(ideal(f))
```

```
Univariate Quotient Polynomial Ring in ybar over Fraction Field of Univa
```

```
sage: R2.<x> = ZZ[]
```

$$\mathbf{Z}[x]$$

```
sage: F2 = Frac(R2)
```

$$\mathrm{Frac}(\mathbf{Z}[x])$$

```
sage: R3.<y> = F2[]
```

$$\mathrm{Frac}(\mathbf{Z}[x])[y]$$

```
sage: f = y^2 - (x^4 + 4*x^3 + 2*x^2 + 1)
```

$$y^2 - x^4 - 4x^3 - 2x^2 - 1$$

```
sage: Q2 = R3.quo(ideal(f))
```

```
Univariate Quotient Polynomial Ring in ybar over Fraction Field of Univa
```

```
sage: Q2(Q(res[0].element()*(ln/ld)))
```

$$\Bigg(\frac{3011761242371675159956659107451002230861813162286154510537567261568993621331874022}{} \\ + \frac{3017649338642196519018157952235852675105922278927965174292283913106762914824213952}{}$$

This is a lot more complicated than what Geddes found:

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

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

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

$$y = \sqrt{x^4 + 4x^3 + 2x^2 + 1}$$

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

□

9.7 The Risch Theorem: A First Look

At this point, there is only one major missing piece in our integration theory for Abelian integrals — 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 principal?

We'll attack this problem the way Robert Risch discovered in 1970, by mapping into a finite field, solving the corresponding problem there, then lifting the result back to the original field. The finite field will be the integers modulo a prime. Reducing modulo some primes changes the genus of the algebraic curve, while other primes leave the genus unchanged. Those primes at which the genus of the curve remains unchanged yield *good reduction*

The key theorems as stated by [Tr73] on page 67:

Theorem 9.8. *The homomorphism between divisor class groups under good reduction is an isomorphism when restricted to divisors whose orders are relatively prime to the characteristic of the reduced function field.* □

Let p be the characteristic of the reduced field,

Theorem 9.9. *if the divisor D has order $p^k n$ where $\gcd(n, p) = 1$, then the reduction of D must have order $p^j n$ for some $j \leq k$.*

Proof Let the order of the reduction be $p^j m$. Since reduction is a group homomorphism, we must have $m|n$ and $j \leq k$. Since D^{p^k} has order exactly n, its reduction must have order exactly n. But the order of its reduction is a divisor of m and thus $n|m$ and so finally we have $n = m$.

□

Since good reduction preserves the part of the divisor's order relatively prime to the characteristic, by picking two different primes we can completely reconstruct the divisor's order in characteristic zero.

A property of algebraic curves over finite fields is that *all* divisors of total degree zero have some multiple that is principal (proof in Chapter 10 when we construct the Jacobian variety). Thus, to carry out this program, we need to compute bases for Riemann-Roch spaces in prime characteristic. Then, we can keep raising a divisor to higher and higher powers until we find its order. Do this for two different primes, each exhibiting good reduction, and then we can find the order of the original divisor in characteristic zero.

I'll prove these claims in the next chapter (not yet written), but for now let's just see how to use these theorems to fix the degree of the divisor in the last section. First, let's determine the genus of the curve over $\mathbf{Q}$ (what about $\overline{\mathbf{Q}}$?):

```
sage: PP.<x,y,z> = ProjectiveSpace(QQ, 2)
```

$$\mathbf{P}_{\mathbf{Q}}^{2}$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^{4} + 4x^{3} + 2x^{2} + 1$$

```
sage: C = Curve((y^2-root).homogenize(z))
```

$$-x^{4} - 4x^{3}z - 2x^{2}z^{2} + y^{2}z^{2} - z^{4}$$

```
sage: C.geometric_genus()
```

$$1$$

Now, let’s check reduction mod 2:

```
sage: PP.<x,y,z> = ProjectiveSpace(GF(2), 2)
```

$$\mathbf{P}_{\mathbf{F}_{2}}^{2}$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^{4} + 1$$

```
sage: C = Curve((y^2-root).homogenize(z))
```

$$x^{4} + y^{2}z^{2} + z^{4}$$

```
sage: C.geometric_genus()
```

$$0$$

The genus changed, so we don’t have good reduction mod 2.

What about mod 3 and mod 5?

```
sage: PP.<x,y,z> = ProjectiveSpace(GF(3), 2)
```

$$\mathbf{P}_{\mathbf{F}_{3}}^{2}$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 + x^3 - x^2 + 1$$

```
sage: C = Curve((y^2-root).homogenize(z))
```

$$-x^4 - x^3z + x^2z^2 + y^2z^2 - z^4$$

```
sage: C.geometric_genus()
```

$$1$$

```
sage: PP.<x,y,z> = ProjectiveSpace(GF(5), 2)
```

$$\mathbf{P}^2_{\mathbf{F}_5}$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 - x^3 + 2x^2 + 1$$

```
sage: C = Curve((y^2-root).homogenize(z))
```

$$-x^4 + x^3z + 3x^2z^2 + y^2z^2 - z^4$$

```
sage: C.geometric_genus()
```

$$1$$

Since the genus is unchanged, both of these primes exhibit good reduction. Let's do the divisor calculation, first modulo 3:

```
sage: A.<a> = GF(3)[]
```

$$\mathbf{F}_3[a]$$

```
sage: B.<b> = GF(3^2, modulus=a^2-2)
```

$$\mathbf{F}_{3^2}$$

```
sage: R.<x> = FunctionField(B)
```

```
Rational function field in x over Finite Field in b of size 3^2
```

```
sage: L.<y> = R[]
```

Rational function field in x over Finite Field in b of size 3^2$[y]$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 + x^3 + 2x^2 + 1$$

```
sage: F.<y> = R.extension(y^2 - root)
```

```
Function field in y defined by y^2 + 2*x^4 + 2*x^3 + x^2 + 2
```

```
sage: num = 6*x^2 + 5*x +7
```

$$2x + 1$$

```
sage: den = 2*x^6 + 8*x^5 + 3*x^4 + - 4*x^3 - 1
```

$$2x^6 + 2x^5 + 2x^3 + 2$$

```
sage: integrand = y*num/den * x.differential()
```

$$\left(\left(\frac{x+2}{x^6 + x^5 + x^3 + 1}\right) y\right)\, dx$$

```
sage: D = integrand.divisor()
```

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

```
sage: D2 = add([ZZ(1 if B(integrand.residue(pl))==1 else -1) * pl fo
```

$$-(x+b, y+b+2) + (x+b, y+2b+1) + (x+2b, y+b+1) - (x+2b, y+2b+2)$$

```
sage: D2.basis_function_space()
```

$$[]$$

```
sage: (2*D2).basis_function_space()
```

$$[]$$

```
sage: (3*D2).basis_function_space()
```

$$[]$$

```
sage: (4*D2).basis_function_space()
```

$$[]$$

```
sage: (5*D2).basis_function_space()
```

$$\left[\left(\frac{x^{6} + 2x^{5} + 2x^{4} + 2x^{2} + 2x}{x^{10} + 2x^{8} + x^{6} + x^{4} + 2x^{2} + 1}\right) y + \frac{x^{10} + 2x^{8} + x^{5} + x^{4} + x^{3} + x^{2} + 1}{x^{10} + 2x^{8} + x^{6} + x^{4} + 2x^{2} + 1}\right]$$

So, we see that the divisor has order 5, mod 3. This means that the original divisor's order is 5 times some multiple of 3.

Again, let's find the divisor's order mod 5:

```
sage: A.<a> = GF(5)[]
```

$$\mathbf{F}_{5}[a]$$

```
sage: B.<b> = GF(5^2, modulus=a^2-2)
```

$$\mathbf{F}_{5^{2}}$$

```
sage: R.<x> = FunctionField(B)
Rational function field in x over Finite Field in b of size 5^2
sage: L.<y> = R[]
```

Rational function field in x over Finite Field in b of size 5^2$[y]$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^{4} + 4x^{3} + 2x^{2} + 1$$

```
sage: F.<y> = R.extension(y^2 - root)
Function field in y defined by y^2 + 4*x^4 + x^3 + 3*x^2 + 4
sage: num = 6*x^2 + 5*x +7
```

$$x^{2} + 2$$

```
sage: den = 2*x^6 + 8*x^5 + 3*x^4 + - 4*x^3 - 1
```

$$2x^6 + 3x^5 + 3x^4 + x^3 + 4$$

```
sage: integrand = y*num/den * x.differential()
```

$$\left(\left(\frac{3}{x^4 + 4x^3 + 2x^2 + 1}\right) y\right)\, dx$$

```
sage: D = integrand.divisor()
```

$$0$$

What does this mean? Let's take a look at how this function behaves over **Q**:

```
sage: R.<x> = FunctionField(QQ)
```

```
    Rational function field in x over Rational Field
```

```
sage: L.<y> = R[]
```

```
   Rational function field in x over Rational Field[y]
```

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 + 4x^3 + 2x^2 + 1$$

```
sage: F.<y> = R.extension(y^2 - root)
```

```
Function field in y defined by y^2 - x^4 - 4*x^3 - 2*x^2 - 1
```

```
sage: num = 6*x^2 + 5*x +7
```

$$6x^2 + 5x + 7$$

```
sage: den = 2*x^6 + 8*x^5 + 3*x^4 + - 4*x^3 - 1
```

$$2x^6 + 8x^5 + 3x^4 - 4x^3 - 1$$

```
sage: integrand = y*num/den * x.differential()
```

$$\left(\left(\frac{3x^2 + \frac{5}{2}x + \frac{7}{2}}{x^6 + 4x^5 + \frac{3}{2}x^4 - 2x^3 - \frac{1}{2}}\right) y\right)\, dx$$

```
sage: D = integrand.divisor()
```

$$-\left(x^2-\frac{1}{2},y-2x-\frac{1}{2}\right)-\left(x^2-\frac{1}{2},y+2x+\frac{1}{2}\right)+\left(x^2+\frac{5}{6}x+\frac{7}{6}\right)$$

```
sage: ppoly = [pl.prime_ideal().gens()[0] for pl,m in D.list() if m < 0
sage: zpoly = [pl.prime_ideal().gens()[0] for pl,m in D.list() if m > 0
sage: R2.<x> = QQ[]
```

$$\mathbf{Q}[x]$$

```
sage: zpoly5 = R2(str(zpoly)).numerator().change_ring(GF(5))
```

$$x^2+2$$

```
sage: ppoly5 = R2(str(ppoly)).numerator().change_ring(GF(5))
```

$$2x^2+4$$

Pay attention to how the ideals' polynomials behaves when reduced mod 5. $x^2+\frac{5}{6}x+\frac{7}{6} \mod 5 = x^2+2$, and $x^2-\frac{1}{2} \mod 5 = 2x^2+4$, so the pole set and the zero set collapse together mod 5, and we get a zero divisor.

Since a zero divisor is principal (it matches any constant, non-zero function), we conclude that D is principal mod 5, and the original divisor's order is 1 times some multiple of 5. Combine this with the previous fact that the divisor's order is 5 times some multiple of 3, and we conclude that the divisor's order is exactly 5 – the only divisor we actually need to check over $\overline{\mathbf{Q}}$.

Let's take a look at mod 7:

```
sage: PP.<x,y,z> = ProjectiveSpace(GF(7), 2)
```

$$\mathbf{P}^2_{\mathbf{F}_7}$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4+4x^3+2x^2+1$$

```
sage: C = Curve((y^2-root).homogenize(z))
```

$$-x^4+3x^3z+5x^2z^2+y^2z^2-z^4$$

```
sage: C.geometric_genus()
```

$$1$$

So, we have good reduction mod 7, as well as the interesting property that $\sqrt{2}$ already exists in this field, since $3^2 = 2 \mod 7$. This causes two of the places in our $\overline{\mathbf{Q}}$ divisor to collapse together and cancel each other out, so we get a divisor supported over only two places.

```
sage: R.<x> = FunctionField(GF(7))
```

```
Rational function field in x over Finite Field of size 7
```

```
sage: L.<y> = R[]
```

Rational function field in x over Finite Field of size $7[y]$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4 + 4x^3 + 2x^2 + 1$$

```
sage: F.<y> = R.extension(y^2 - root)
```

```
Function field in y defined by y^2 + 6*x^4 + 3*x^3 + 5*x^2 + 6
```

```
sage: num = 6*x^2 + 5*x +7
```

$$6x^2 + 5x$$

```
sage: den = 2*x^6 + 8*x^5 + 3*x^4 + - 4*x^3 - 1
```

$$2x^6 + x^5 + 3x^4 + 3x^3 + 6$$

```
sage: integrand = y*num/den * x.differential()
```

$$\left(\left(\frac{3x}{x^5 + 2x^4 + x^3 + 3x^2 + x + 5}\right) y\right) dx$$

```
sage: D = integrand.divisor()
```

$$(x, y+1) + (x, y+6) - (x+5, y+1) - (x+5, y+6)$$

```
sage: [(integrand.residue(pl), pl) for pl,m in D.list() if m < 0]
```

$$[(1, (x+5, y+1)), (6, (x+5, y+6))]$$

```
sage: D2 = add([ZZ(1 if (integrand.residue(pl))==1 else -1) * pl for pl,
```

$$(x+5, y+1) - (x+5, y+6)$$

```
sage: D2.basis_function_space()
```

$$[]$$

```
sage: (2*D2).basis_function_space()
```

$$[]$$

```
sage: (3*D2).basis_function_space()
```

$$[]$$

```
sage: (4*D2).basis_function_space()
```

$$[]$$

```
sage: (5*D2).basis_function_space()
```

$$\left[\left(\frac{2x^3+5x^2}{x^5+4x^4+5x^3+4x^2+3x+3}\right)y+\frac{x^5+6x^4+3x^3+x^2+6x+6}{x^5+4x^4+5x^3+4x^2+3x+3}\right]$$

The divisor also has order 5, mod 7. This means that the original divisor's order is 5 times some multiple of 7. Combine this with the previous fact that the divisor's order is 5 times some multiple of 3, and we conclude that the divisor's order is exactly 5 – the only divisor we actually need to check over $\overline{\mathbf{Q}}$.

9.8 Hermite reduction

The previous sections in this chapter have laid out the theoretical framework for the solution of Abelian integrals and presented the simplest computational algorithm that I could formulate to acheive that goal. It is by no means the most efficient algorithm, and should not be used as a basis for a professional implementation. The remaining sections of this chapter showcase optimization techniques, and can be skipped without loss of continuity in the text. In particular, Puiseux expansions can be completely avoided, which offers significant savings in computational complexity.

[Tr84] shows how to extend Hermite reduction into the algebraic case. This offers a means of calculating the rational parts of the integrals without going through the calculations described in this chapter.

Example 9.10. 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:

```
sage: R.<x,y> = QQ[]
```

$$\mathbf{Q}[x,y]$$

```
sage: F=Frac(R)
```

$$\mathrm{Frac}(\mathbf{Q}[x,y])$$

```
sage: root = x^4+4*x^3+2*x^2+1
```

$$x^4+4x^3+2x^2+1$$

```
sage: num = 2*x^6 + 4*x^5 + 7*x^4-3*x^3-x^2-8*x-8
```

$$2x^6+4x^5+7x^4-3x^3-x^2-8x-8$$

```
sage: den = 2*x^2-1
```

$$2x^2-1$$

```
sage: root.factor()
```

$$(x+1)\cdot(x^3+3x^2-x+1)$$

…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 from section 2.1 of his "Symbolic Integration Tutorial" with $v = 2x^2-1$ to eliminate this square in the denominator:

```
sage: D = Derivation(F, {x: 1, y: root.derivative(x)/2*y/root})
```

$$\begin{array}{l} \texttt{Derivation of } \mathrm{Frac}(\mathbf{Q}[x,y]) \\ \begin{array}{lcl} x & \to & 1 \\ y & \to & \frac{2x^3y+6x^2y+2xy}{x^4+4x^3+2x^2+1} \end{array} \end{array}$$

```
sage: D(y)
```

$$\frac{2x^3y+6x^2y+2xy}{x^4+4x^3+2x^2+1}$$

```
sage: U = root
```

$$x^4+4x^3+2x^2+1$$

```
sage: V = den
```

$$2x^2-1$$

```
sage: S2 = U*V^2*D(y/V)
```

$$-4x^4y-6x^3y-6x^2y-6xy$$

Now we want to solve $f_2S_2 = A_2y$ where A_2y is our numerator.

```
sage: f2 = num*y/S2
```

$$\frac{2x^6+4x^5+7x^4-3x^3-x^2-8x-8}{-4x^4-6x^3-6x^2-6x}$$

```
sage: T2 = f2.numerator()
```

$$2x^6+4x^5+7x^4-3x^3-x^2-8x-8$$

```
sage: Q = f2.denominator()
```

$$-4x^4-6x^3-6x^2-6x$$

```
sage: R2 = QQ['x']
```

$$\mathbf{Q}[x]$$

```
sage: (A,R) = diophantine(R2(1),R2(V),R2(Q))
```

$$\left(-\frac{36}{49}x^3 - \frac{38}{49}x^2 - \frac{48}{49}x - 1, -\frac{18}{49}x + \frac{8}{49}\right)$$

```
sage: (Q2,B2) = (T2*R).quo_rem(V)
```

$$\left(-\frac{18}{49}x^5 - \frac{4}{7}x^4 - \frac{8}{7}x^3 + \frac{41}{49}x^2 - \frac{31}{49}x + \frac{177}{98}, x + \frac{1}{2}\right)$$

```
sage: h = A*num*y/(V*U) - (D(V)*Q2+D(B2))*y/V + Q2*D(y)
```

$$\frac{6x^2y + 5xy + 7y}{2x^6 + 8x^5 + 3x^4 - 4x^3 - 1}$$

```
sage: F3 = Frac(ZZ['x']['y'])
```

$$\mathrm{Frac}(\mathbf{Z}[x][y])$$

$$\begin{aligned}&\int \frac{(2x^6 + 4x^5 + 7x^4 - 3x^3 - x^2 - 8x - 8)\,y}{2x^2 - 1}\,\mathrm{d}x\\ &= \frac{(2x+1)\,y}{4x^2 - 2} + \int \frac{(6x^2 + 5x + 7)\,y}{2x^6 + 8x^5 + 3x^4 - 4x^3 - 1}\,\mathrm{d}x\end{aligned}$$

We solved this new integral in example 9.7. The final answer is:

$$A(x) = 1023x^8 + 4104x^7 + 5048x^6 + 2182x^5 + 805x^4 + 624x^3 + 10x^2 + 28x$$
$$B(x) = 1025x^{10}+6138x^9+12307x^8+10188x^7+4503x^6+3134x^5+1598x^4+140x^3+176x^2+2$$
$$C(x) = 32x^{10} - 80x^8 + 80x^6 - 40x^4 + 10x^2 - 1$$

$$y = \sqrt{x^4 + 4x^3 + 2x^2 + 1}$$

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

□

9.9 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

The Risch Theorem

THIS CHAPTER IS VERY VAGUE AND INCOMPLETE.

We now turn to the task of proving the Risch theorem.

There are three key steps in this proof.

1. Establish the existence of the Jacobian variety.

   It's not that hard to see that divisor classes on an algebraic curve can be composed and inverted, forming the *divisor class group*. It's not nearly so obvious that these divisor classes can be identified with the points on an algebraic variety. The resulting *Jacobian variety* is thus both a group and an algebraic variety, and the group operations can be represented by rational functions on the algebraic variety, thus forming an *Abelian variety*, which is an algebraic variety equipped with a group structure.

2. Prove that the following diagram commutes:

$$\begin{tikzcd}
C \arrow[r] \arrow[d] & \mathrm{J}(C) \arrow[d] \\
C_p \arrow[r] & \mathrm{J}(C_p)
\end{tikzcd}$$

   The objects on the left are algebraic curves; the objects on the right are Abelian varieties. The horizontal arrows correspond to constructing a curve's associated Jacobian variety; the vertical arrows correspond to reduction modulo a prime.

   At the end of the previous chapter, we were working on the lower-left half of this diagram. We reduced a curve modulo a prime, then computed the order of various divisors on the reduced curve, which essentially computes orders of points on the reduced curve's Jacobian variety. We now wish to show that those group orders are the same as if we had constructed the Jacobian of the original curve and then reduced the Jacobian modulo the prime.

3. Study the effects of reduction modulo a prime on an Abelian variety.

   At this point, we no longer need any special properties of the Jacobian variety; it suffices to study how reduction modulo a prime affects the group structure on an Abelian variety. In particular, we will find that the factor of an element's order coprime to the reduction prime is preserved.

10.1 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 10.1. (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 11.22, 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 10.2.

$$l(\mathfrak{p}) \leq \deg \mathfrak{p} + 1$$

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 10.3. *If $\mathfrak{A}$ is divisible by $\mathfrak{B}$, i.e, if $\mathfrak{A}\mathfrak{B}^{-1} \subseteq \mathcal{I}$, then*

$$n(\mathfrak{A}) - l(\mathfrak{A}) \leq n(\mathfrak{B}) - l(\mathfrak{B})$$

$$n(\mathfrak{p}) \equiv \text{deg}\mathfrak{p}$$

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

□

It immediately follows (from $\mathfrak{b} = \mathbf{0}$) that $l(\mathfrak{c}) = g$, which can be taken as the definition of the genus.

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.

The oldest construction of the Jacobian variety uses integrals and only works over the complex field **C**, and requires us to work on a non-singular model of the curve.

We can pick an arbitrary origin and $2g$ closed paths from that origin that form a basis for the surface's holonomy group. We can also pick g independent holomorphic differentials and evaluate them over the $2g$ closed paths to get a lattice Λ^g in $\mathbf{C}^g$. Given any point on the surface, we can now evaluate the differentials along a path from the origin to that point and thus map into the torus $\mathbf{C}^g/\Lambda^g$.

Not all complex torii are algebraic varieties, but this one is (PROOF NEEDED).

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 $\mathbf{C}^g/\Lambda^g$.

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, principal); this is Abel's theorem.

Pulling this all together, we need to show that the group operation is defined by rational functions.

For our purposes, it isn't enough to construct the Jacobian over the complex field **C**; we also need Jacobians for curves in prime characteristic. So the traditional construction using integrals isn't available, we need something different.

Andre Weil, in Courbes Algébriques et Variétés Abéliennes (1948), constructs the Jacobian variety over arbitrary fields using the g-fold symmetric product of the curve. See

Michel Raynaud, “André Weil and the Foundations of Algebraic Geometry” for an broad overview of this French text.

J.S. Milne’s “Jacobian Varieties” (jmilne.org/math) also uses the symmetric power of the variety.

Greg Anderson, “Abeliants and their application to an elementary construction of Jacobians” uses equivalence classes of matrices constructed using a Riemann-Roch space. Equivalence of matrices defined over a vector space, as Joshua Grochow taught me[1], is the same as simultaneous equivalence of matrices, a notoriously hard problem. Sergeichuk, “Canonical matrices for linear matrix problems” shows how to convert such a matrix into an canonical form, but there’s no obvious way to introduce the rank 1 condition required by Anderson’s construction, nor is the canonical form easily parametized.

10.3 Simple Algebraic Extensions over Finite Fields

Let’s start with a simple but crucial observation:

Theorem 10.4. *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 principal.* 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 10.5. *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)*

[1] https://mathoverflow.net/questions/303627

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 10.6. *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 10.7. Show that some multiple of $\mathrm{Z}(1,1)$ is principal 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 principal 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:

$$\begin{array}{rcl}
(0,0) & : & (-1)^{24} + (-1)^{24} = 4^{24} + 4^{24} = 1 + 1 = 2 \\
(1,4) & : & 3^{24} + 0^{24} = 1 + 0 = 1 \\
(2,\gamma) & : & (\gamma - 1)^{24} + (2 - 1)^{24} = 1 + 1 = 2, \quad \text{since :} \\
 & & \\
 & & (\gamma - 1)^2 = (\gamma^2 - 2\gamma + 1) = 3 - 2\gamma \\
 & & (\gamma - 1)^4 = (3 - 2\gamma)^2 = (9 - 12\gamma + 4\gamma^2) = 2 - 2\gamma \\
 & & (\gamma - 1)^8 = (2 - 2\gamma)^2 = (4 - 8\gamma + 4\gamma^2) = 2 - 3\gamma \\
 & & (\gamma - 1)^{12} = (2 - 2\gamma)(2 - 3\gamma) = (4 - 10\gamma + 6\gamma^2) = 1
\end{array}$$

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.

□

10.4 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 **Z** can be mapped homomorphicaly[2] 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 **Z** is commutative, the subring $[m]$ it maps to is also commutative, even though $\mathrm{End}(G)$ may not be.

[2] An easy consequence of **Z**'s repelling universal property in the category of rings, see [Lang], p. ?

Chapter 11

Algebraic Extensions

THIS CHAPTER IS VERY VAGUE AND INCOMPLETE.

We now turn to the algebraic extension in general. 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've studied in the last two chapters, 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: $\mathbf{C}(x, y)$. However, many of these 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.

The most important change that occurs when shifting from $\mathbf{C}(x, y)$ to $\mathrm{K}(\theta, y)$ is the loss of algebraic closure. The complex field is algebraically closed, so the irreducible polynomials is $\mathbf{C}[x]$ are simply the linear factors $(x - \gamma)$, and there's a one-to-one correspondence between these linear factors and the complex numbers.

An arbitrary differential field K, however, is not likely to be algebraically closed, so we need to shift our viewpoint from thinking in terms of places over *points* to thinking in terms of places over *irreducible polynomials*. We've already looked (WHERE?) at factoring polynomials into their irreducible factors, and now we wish to reformulate everything we've done in the past three chapters to adapt.

11.1 Exponential Extensions over Simple Algebraic Extensions

To analyze integrands formed from exponential extensions over algebraic extensions, we need to solve Risch equations in algebraic extensions. We haven't yet developed the tools to handle this problem in general algebraic extensions, but we can use the tools of Chapter 8 to solve Risch equations in a simple algebraic extension of $\mathbf{C}(x)$. This will allow us to solve integrals involving exponentials of polynomial roots. While such integrals aren't terribly common, this exercise will reinforce the principles of working over algebraic extensions.

Recall that a Risch equation has the form

$$r' + Sr = T \qquad S, T, r \in K \tag{11.1}$$

K is a differential field, that in our case will be an algebraic extension of $\mathbf{C}(x)$:

$$\frac{dR}{dx} + SR = T \qquad S, T, r \in \mathbf{C}(x)[y]/p(x, y) \tag{11.2}$$

Our approach in Chapter 7 was to identify the poles in S and T and then use this information to restrict r enough to solve it.

This same approach works with algebraic extensions; we just need to work with places and principal parts expansions instead of denominator factors and partial fractions expansions. Actually, working with respect to a local uniformizing variable simplies the theory, since we don't have to treat infinity separately.

Multiplying through by dx, we obtain:

$$dR + R(Sdx) = Tdx \tag{11.3}$$

Taking the differential of an algebraic function causes its poles to increase in order by exactly one, and introduces no new poles (Theorem ?). Therefore, we can't have any poles (finite or infinite) at places where there aren't poles in either $S\,dx$ or $T\,dx$ (the differentials, not the functions).

Once we find the poles of $S\,dx$ and $T\,dx$, there are three cases, as before:

1. $\nu(S\,dx) > -1 \quad \Longrightarrow \quad \nu(R) \geq \nu(T\,dx) + 1.$ (11.4a)
2. $\nu(S\,dx) = -1 \quad \Longrightarrow \quad \nu(R) = \nu(T\,dx) + 1$ or $\nu(R) = -\pi(S\,dx).$ (11.4b)
3. $\nu(S\,dx) < -1 \quad \Longrightarrow \quad \nu(R) \geq \nu(T\,dx) - \nu(S\,dx).$ (11.4c)

Once we've figured out the maximum orders of all of R's poles, it's then a straightforward matter to construct a suitable divisor and compute a basis for its Riemann-Roch space. Then we can use fairly straightfoward techniques of linear algebra to find the coefficients of the solutions, *if* they exist.

Example 11.1. Integrate $\int \left(\frac{5x^4+2x-2}{x^2}\left(1+\frac{1}{\sqrt{x^3+1}}\right)+\frac{x}{\sqrt{x^3+1}}\right)e^{x\sqrt{x^3+1}}dx$

From [Br91]

The integrand can be expressed using the tower $\mathbf{C}(x) \subset \mathbf{C}(x, y) \subset \mathbf{C}(x, y, \psi)$, where $y^2 = x^3 + 1$, $y' = \frac{3x^2}{2y}$, and $\psi = \exp xy$:

Integrate $\int \left(\frac{5x^4+2x-2}{x^2} \left(1+\frac{1}{y}\right) + \frac{x}{y} \right) \psi \, dx$

Since the top-most extension is exponential, we apply Theorem 6.1, and see that $k = xy$, $k' = xy' + y = (\frac{3x^3}{2y} + y)$, and $a_1 = \left(\frac{5x^4+2x-2}{x^2} \left(1+\frac{1}{y}\right) + \frac{x}{y} \right)$, so equation 6.4 reads:

$$A_1' + (\frac{3x^3}{2y} + y)A_1 = \left(\frac{5x^4+2x-2}{x^2} \left(1+\frac{1}{y}\right) + \frac{x}{y} \right)$$

This equation is in an algebraic form of equation 6.8, so we want to analyse the poles of the S and R functions.

```
sage: R.<x> = FunctionField(QQbar)
```

```
   Rational function field in x over Algebraic Field
```

```
sage: L.<y> = R[]
```

```
  Rational function field in x over Algebraic Field[y]
```

```
sage: F.<y> = R.extension(y^2 - (x^3+1))
```

```
     Function field in y defined by y^2 - x^3 - 1
```

```
sage: S = 3*x^3/(2*y) + y
```

$$\left(\frac{\frac{5}{2}x^3 + 1}{x^3 + 1} \right) y$$

```
sage: T = (5*x^4+2*x-2)/(x^2)*(1+1/y) + x/y
```

$$\left(\frac{5x^3 - 4x^2 + 4x - 2}{x^4 - x^3 + x^2} \right) y + \frac{5x^4 + 2x - 2}{x^2}$$

```
sage: differential_divisor_of_poles(S*x.differential())
```

$$6 \left(\frac{1}{x^2} y \right)$$

```
sage: Tdp = differential_divisor_of_poles(T*x.differential())
```

$$7 \left(\frac{1}{x^2} y \right) + 2\,(x, y - 1)$$

We have two places at which either $S\,dx$ or $T\,dx$ have poles, so we need to consider them both.

At $(0,1)$, $\nu(S\,dx) = 0$, so $\nu(R) >= \nu(T\,dx) + 1 = -2 + 1 = -1$.

At infinity, $\nu(S\,dx) = -6$, so $\nu(R) >= \nu(T\,dx) - \nu(S\,dx) = -7 - (-6) = -1$.

Our solution, if it exists, has at most first-order poles at $(0,1)$ and ∞.

```
sage: sum(Tdp.support())
```

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

```
sage: BFS = sum(Tdp.support()).basis_function_space()
```

$$\left[1, \frac{1}{x} y + \frac{1}{x}\right]$$

Is there some linear combination of these basis functions that solves our Risch equation?

```
sage: R = 2 * BFS[1]
```

$$\frac{2}{x} y + \frac{2}{x}$$

```
sage: R.differential() + R * S * x.differential() == T * x.differential
```

$$\mathrm{True}$$

This solves the Risch equation (?), giving us the A_1 term in our solution. Since there's no term in the integrand of ψ-degree zero, the A_1 term is the only term in our solution, and we conclude:

```
sage: var('x,y')
```

$$(x, y)$$

```
sage: integrand = eval(preparse(str(T))).subs({y:sqrt(x^3+1)}) * exp(x*
```

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

```
sage: ans = eval(preparse(str(R))).subs({y:sqrt(x^3+1)}) * exp(x*sq
```

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

```
sage: bool(diff(ans,x) == integrand)
```

True

□

[Br91] gives a more sophisticated algorithm for attacking these kinds of problems.

11.2 Integral Elements

An important component of Hess's algorithm for computing bases of Riemann-Roch spaces is the construction of maximal orders. A moment's thought suggests that the key characteristic of the finite maximal order is that it contains all of the functions with no finite poles. In fact, that is an important characterization of the finite maximal order, one that we now wish to define more concretely.

Definition 11.2. *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.

11.3 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 11.3. *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 11.4. *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 distributive property to factor out the coefficients from of each basis element and then performing the addition in the ring R:

$$(a_1 b_1 + a_2 b_2 + ... + a_n b_n) + (c_1 b_1 + c_2 b_2 + ... + c_n b_n)$$

[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.

$$= (a_1 + c_1)b_1 + (a_2 + c_2)b_2 + ... + (a_n + c_n)b_n$$

So the elements generated from a basis clearly form a module. R operates on them by multiplication by every coefficient.

Example 11.5.

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 $\mathrm{K}(\theta, y)$.

11.4 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 11.5), 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 11.6. *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 $\mathrm{Tr}(w_i u_j) = \delta_{ij}$. ([Lang] Corollary VI.5.3)*

Proof

Consider the following matrix:

$$M = \begin{pmatrix} \mathrm{Tr}(w_1 w_1) & & \\ \vdots & \ddots & \\ \mathrm{Tr}(w_1 w_n) & \cdots & \mathrm{Tr}(w_n w_n) \end{pmatrix}$$

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 11.7. *$\mathcal{I}$ 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 11.6, 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 11.8. *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 11.9. *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 11.8, 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 11.7 and 11.9 demonstrate that $\mathcal{I}$ is a free $\mathrm{K}[\theta]$-module.

Definition 11.10. *A basis for $\mathcal{I}$ will be called an* **integral basis**.

Theorem 11.11. *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.

11.5 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 11.12. *The* **conjugates** *of a rational function $\eta(x, y)$ in $\mathrm{C}(x, y)$ 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 11.13. *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 11.14. *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 11.15. *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 11.14, 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 11.16. 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 11.14 and 11.15, 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.

11.6 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 11.17. *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 11.18. *$\mathcal{I}$ 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 11.19. *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 11.20. *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 11.19 to check if the function has a simple zero, and if not, choose a different value for γ.

□

Theorem 11.21. *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 11.19 to check for any extra zeros. Since only a finite number of γ values produce extra zeros, this process is guaranteed to terminate.

□

Theorem 11.22. *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 11.20 or 11.21 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 11.23. *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 11.24. *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 11.22 with the zeros and poles required by $\mathcal{D}$, adding $\mathfrak{p}$ to Σ if necessary.

□

Theorem 11.25. *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 11.18), $\mathcal{I}\{r^{-1}\}$ 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 11.24 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 11.22) 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 11.25 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 11.26. *Given a divisor $\mathcal{D}$, a pair of functions can always be constructed that generate the divisor's associated integral module.*

Proof

Use Theorem 11.22 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 11.25, (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 principal 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 11.26 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 11.27. *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 $\mathfrak{p}$

$$f\,dx = rt^{r-1}fg\,dt + t^r f(\frac{dg}{dt})dt$$

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 $rt^r fg = r(x - x_0)f$.

□

Chapter 12

Notes

For a while I was thinking that the product of *prime* ideals is equal to their *intersection*. This is true in principal ideal domains, but not in general.

For example, consider $I = (x, z)$ and $J = (x + z)$ in $K[x, z]$. Now, $x + z \in I \cap J$, but $x + z \notin I \cdot J$.

Sage was useful in puzzling this out:

```
sage: R.<x,z> = QQ[];

sage: I = Ideal(x,z);
sage: J = Ideal(x+z);

sage: I.is_prime()

                         True

sage: J.is_prime()

                         True

sage: I.intersection(J) == I*J

                         False
```

Looking at the primary decomposition, we see that the product is smaller than the intersection, because there's an extra ideal that needs to be intersected (the original ideal is the intersection of the ideals in its primary decomposition). Sage's comparison operator for ideals also shows us that the product is contained in the intersection.

```
sage: I.intersection(J).primary_decomposition()
```

$$\left[\left(x + z\right) \mathbf{Q}[x, z]\right]$$

```
sage: (I*J).primary_decomposition()
```

$$\left[\left(x + z\right) \mathbf{Q}[x, z], \left(z, x^{2}\right) \mathbf{Q}[x, z]\right]$$

```
sage: I.intersection(J) > I*J
```

True

So now it's a question of finding something in the intersection that isn't in the product. The ideal quotient isn't useful for this, probably because of its Zariski closure property.

```
sage: I.intersection(J).quotient(I*J)
```

$$\left(1\right) \mathbf{Q}[x, z]$$

```
sage: diff(sqrt(x^4+1),x)
```

$$\frac{2\,x^{3}}{\sqrt{x^{4}+1}}$$

```
sage: diff(sqrt(x^4+1),x,2)
```

$$-\frac{4\,x^{6}}{\left(x^{4}+1\right)^{\frac{3}{2}}}+\frac{6\,x^{2}}{\sqrt{x^{4}+1}}$$

```
sage: diff(sqrt(x^4+1),x,3)
```

$$\frac{24\,x^{9}}{\left(x^{4}+1\right)^{\frac{5}{2}}}-\frac{36\,x^{5}}{\left(x^{4}+1\right)^{\frac{3}{2}}}+\frac{12\,x}{\sqrt{x^{4}+1}}$$

```
sage: diff(sqrt(x^4+1),x,4)
```

$$-\frac{240\,x^{12}}{\left(x^{4}+1\right)^{\frac{7}{2}}}+\frac{432\,x^{8}}{\left(x^{4}+1\right)^{\frac{5}{2}}}-\frac{204\,x^{4}}{\left(x^{4}+1\right)^{\frac{3}{2}}}+\frac{12}{\sqrt{x^{4}+1}}$$

```
sage: diff(sqrt(x^3+1),x)
```

$$\frac{3\,x^{2}}{2\,\sqrt{x^{3}+1}}$$

```
sage: diff(sqrt(x^3+1),x,2)
```

$$-\frac{9\,x^{4}}{4\,\left(x^{3}+1\right)^{\frac{3}{2}}}+\frac{3\,x}{\sqrt{x^{3}+1}}$$

```
sage: diff(sqrt(x^3+1),x,3)
```

$$\frac{81\,x^{6}}{8\,\left(x^{3}+1\right)^{\frac{5}{2}}}-\frac{27\,x^{3}}{2\,\left(x^{3}+1\right)^{\frac{3}{2}}}+\frac{3}{\sqrt{x^{3}+1}}$$

12.1 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:

$$\begin{array}{lll}
\text{positivity} & \forall a \in \mathbf{K}, & \phi(a) \geq 0 \\
\text{definiteness} & \forall a \in \mathbf{K}, & \phi(a) > 0 \iff a \neq 0 \\
\text{homomorphism (on the multiplicative group)} & \forall a, b \in \mathbf{K}, & \phi(ab) = \phi(a)\phi(b) \\
\text{subadditivity (or triangle inequality)} & \forall a, b \in \mathbf{K}, & \phi(a+b) \leq \phi(a) + \phi(b)
\end{array}
\begin{array}{l}
\text{(V1)} \\ \text{(V2)} \\ \text{(V3)} \\ \text{(V4)}
\end{array}$$

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 $w(a) = -\ln \phi(a)$:

$$\begin{array}{lll}
\forall a \in \mathbf{K}, & w(a) \in (-\infty, \infty] & \text{(E1)} \\
\forall a \in \mathbf{K}, & w(a) = \infty \iff a = 0 & \text{(E2)} \\
\forall a, b \in \mathbf{K}, & w(ab) = w(a) + w(b) & \text{(E3)} \\
\forall a, b \in \mathbf{K}, & w(a+b) \geq \min(w(a), w(b)) & \text{(E4)}
\end{array}$$

12.2 Notes on Harris - Geometry of Alg Curves - Harvard 287

Abel's theorem – p. 29

Classical Jacobian discussed on pp. 28-31

"As we've defined it, the Jacobian is only a complex torus so far. Note that a general complex torus is not embeddable in projective space. However, it turns out that the Jacobian has enough meromorphic functions to embed in projective space, so it is a projective variety."

12.3 Notes on [Fu08]

[Fu08] is a good, freely available introduction to algebraic geometry.

[Fu08] assumes an algebraically closed coefficient field throughout (except Chapter 1).

```
Riemann-Roch Theorem

Let C be an algebraic curve, let X be its non-singular model, and let
K be its function field.

Proposition 8.4.  Let x \in K, x \notin k. Let (x)_0 be the divisor
of zeros of x and let n=[K:k(x)].  Then

  1) (x)_0 is an effective divisor of degree n,
  2) There is a constant \tau such that l(r(x)_0) \ge rn-\tau \forall r.

Proof

Prop 6.9. K is an algebraic function field in one variable over k.  By
definition, this means that exists some t such that K is algebraic
over k(t).  So x \in K is algebraic over k(t), and \exists F \ in
k[X,T] such that F[x,t] = 0.  x is not algebraic over k (1-48), so t
must appear in F, so t is algebraic over k(x), and therefore k(x,t) is
algebraic over k(x) (1-50), so K is algebraic over k(x) (1-46).



Problem 1-54: If R is a domain with quotient field K, and L is a
finite algebraic extension of K, then there exists a basis for L over
K such that each basis element is integral over R.

Proof 1-54: Let {w_1, ..., w_n} be any basis for L over K.  Since each
basis element is algebraic over K, by clearing denominators we can
write:

a_{i0} w_i^{n_i} + a_{i1} w_{n_i-1} + \cdots = 0       a_{ij} \in R

We can pull a_{i0} into w_i, and thus adjust the w_i's to be integral
over R by multiplying each one by something in R.  Since anything
in L can be written

   l = \sum c_i w_i        c_i \in K

it can also be written

   l = \sum (c_i / r_i) w'_i

where the r_i adjust the w_i to be integral and c_i/r_i is still in K.



INTEGRAL ELEMENTS: w integral over k[x] means that w is finite
everywhere x is, and has poles only where x does.  w integral over
k[x^{-1}] means that w is finite everywhere x^{-1} is, and has poles
```

only where x has zeros.

So, k[x^{-1}] is a domain with quotient field k(x), and K is a finite algebraic extension of k(x), so there exists a basis for K over k(x) such that each basis element is integral over k[x^{-1}].

Let {w_1, ..., w_n} be such a basis for K over k(x). We will show that the poles of these functions must lie over the roots of x.

w_i^{n_i} + a_{i1} w_{n_i-1} + \cdots = 0   a_{ij} \in k[x^{-1}]

So, ord_P(a_ij) \ge 0 if P \ne S (zero set of x), since x^{-1} and thus anything in k[x^{-1}] is finite away from S.

Problem 2-29: if for some i, ord(a_i) < ord(a_j) \forall j \ne i, then a_1 + \cdots + a_n \ne 0

Proof of 2-29: Assume the contrary. Then we can write a_i = \sum_{i\ne j} a_j. Taking ord of both sides, and using ord(a+b) \ge min(ord(a), ord(b)), we see this is impossible.

Therefore, ord_P(w_i) \ge 0 if P \ne S, since otherwise ord_P(w_i^{n_i}) < ord_P(a_ij w^{n_i-j}) \forall j.

Therefore, the poles of w_i are isolated at the zeros of x, and since there are only a finite number of w_i and each has a finite number of poles, then for some t, div(w_i) + tZ > 0 \forall i.

So w_i \in L(tZ), and if j \le r, then w_i x^{-j} \in L((r+t)Z)

Now w_i are independent over k(x) and 1, x^{-1}, ..., x^{-r} are independent over k, so l((r+t)Z) \ge n(r+1).

Now, l((r+t)Z) = l(rZ) + dim(L((r+t)Z) / L(rZ))

dim(L((r+t)Z) / L(rZ)) \le tm (Prop 3-1), where m is the degree of Z.

So, l(rZ) \ge n(r+1) - tm, so pick \tau = tm-n, and

l(rZ) \ge nr - \tau, \forall r

Riemman-Roch

l(D) = deg(D) + 1 - g + l(W-D)

deg(W) = 2g-2   l(W) = g

l(0) = 1

l(D) = 0 if deg(D) < 0

If deg(D) > 2g-2, then l(D) = deg(D) + 1 - g

If deg(D) = 2g-2, then deg(W-D) = 0, and l(W-D) = 1 iff D-W is principal, otherwise l(W-D) =

Let D=W+X, where deg(X)=0, then l(D) = 2g-2 + 1 - g + (1/0) depending on whether X is principal
   l(D) = g (X is principal) or l(D) = g - 1 (X is not principal)

Goal: an g-dimensional algebraic variety that represents Pic0

Consider (2g-2)-dimensional symmetric space. Each point corresponds to an effective divisor of degree (2g-2).

Fix a (g-2)-tuple. We're left with g free points.

Milne's construction

Use r-dimensional symmetric space, with r > 2g-2. Pick an (r-g) tuple.

12.4 Notes on [Sh61]

A function (Y -> k) is regular at a point on a variety if there exists an open neighborhood of the point where the function is given by a rational function of polynomials, the denominator never zero. (relation to coordinate ring?)

A map (Y -> k) is regular on Y if it is regular at every point of Y.

A morphism (X -> Y) between varieties is a continuous map (in the Zariski topology) such that pullbacks of regular functions are regular.

A rational map is a morphism defined only on an open subset.

Given a rational map from affine varieties X to Y, if [x,y,z] is Y's coordinate system, then we pullback to a regular function, which is a rational function in X's function field. So Y's coordinates are given by rational functions in X's coordinates.

A group variety is an algebraic variety equipped with a group structure, where the group operation and group inversion are rational maps. If the group operation is commutative, it's an abelian variety.

A homomorphism between abelian varieties is a rational map that commutes with the group operation.

An endomorphism is a homomorphism from the variety to itself.

Endomorphisms of a group form a ring. Addition is performed by mapping through both endomorphisms, then applying the group operation, which is commutative, so endomorphism addition is commutative. Multiplication is performed by composition, and need not be commutative.

Using just the addition structure, we get an abelian group that can be structued as a Z-module. Multiplication by an integer is just repeated application of the endomorphism.

We can promote the endomorphism ring into an algebra by tensoring with Q, call this EndQ(A). Elements of EndQ(A) are basically endomorphisms with an associated 1/n denominator (numerators can be sucked into the endomorphism).

A lattice is a free Z-module of the same rank as the algebra over Q.

An order is a lattice that is also a subring and contains the identity.

The order of A, written t, is the image of the endomorphism ring in the endomorphism algebra.

a is a lattice in the endomorphism algebra contained in t, so it's a collection of actual endomorphisms.

g(a,A) is the set of points on A mapped to 0 by every element of a.

Prop 16. Let a be an integral ideal (p. 49) of F. Reduction mod p defines a homomorphism of g(a,A) onto g(a,A^). If a is prime to the characteristic of k^, this homomorphism is an isomorphism.

Div(C) is group of divisors
Div0(C) is group of degree 0 divisors
P(C) is group of principal divisors

P(C) in Div0(C) in Div(C)

define Pic(C) = Div(C)/P(C)
define Pic0(C) = Div0(C)/P(C)

Pic0(C) is isomorphic to the Jacobian variety with a point O.

Given a degree zero divisor in Div0(C), we can use the group law on the Jacobian to construct a single point corresponding to the divisor. Asking if the divisor is principal is asking if this point is O. Asking if any multiple of the divisor is principal is asking if any multiple of this point is O.

We can define an endomorphism to be addition by a point, using the group law. NO - doesn't take identity to identity.

Let's consider [n], the endomorphism defined by applying the group operation n times (n is an integer). This should generate a lattice contained in t, so it's an integral ideal. g([n], A) is the set of points whose n-multiples are principal.

Given a divisor whose (k p^q)-multiple is principal, let's multiply by p^q and get a divisor whose k-multiple is principal. Endomorphism ideal [k] is prime to p. Then this divisor point will be in g(a,A) and g(a,A^), where a is [k].

We can determine that the divisor point is in g(a,A^) for a=[k] by determining that the divisor's k-multiple is principal on the module curve. Since this is an isomorphism, the divisor point is also in g(a,A) for a=[k].

Given a non-singular algebraic curve C, we reduce mod p to get Cp. Good reduction implies that Cp is non-singular with the same genus. C has an associated Jacobian J. Cp also has an associated Jacobian Jp.

PROBLEM: (hopefully) Show that J mod p is Jp.

Construct J as follows. Pick r > 2g-2. Symmetric group J^(r). Pick r-g extra points and find an open covering of J^(r). Within each open set, construct J locally.

12.5 Notes on Bronstein

Definition 3.4.1. We say that $t \in K$ is a monomial over k (w.r.t. D), if

1. t is transcendental over k,
2. $Dt \in k[t]$.

Definition 3.4.3. We say that $u \in k$ is a logarithmic derivative of a k-radical if there exist $v \in k^*$ and an integer $n \neq 0$ such that $nu = Dv/v$.

Definition 5.1.1. $t \in K$ is a primitive over k if $Dt \in k$. $t \in K^*$ is an hyperexponential over k if $Dt/t \in k$. $t \in K$ is Liouvillian over k if t is either algebraic, or a primitive or an hyperexponential over k. K is a Liouvillian extension of k if there are $t_1, ..., t_n$ in K such that $K = k(t_1, ..., t_n)$ and t_i is Liouvillian over $k(t_1, ..., t_{i-1}$ for i in $1, ..., n$.

Theorem 5.1.1. If t is a primitive over k and Dt is not the derivative of an element of k, then t is a monomial over k, Const(k(t)) = Const(k), and S = k (i.e. $S^{irr} = S_1^{irr} = 0$). Conversely, if t is transcendental and primitive over k, and Const(k(t)) = Const(k), then Dt is not the derivative of an element of k.

Theorem 5.1.2. If t is an hyperexponential over k and Dt/t is not a logarithmic derivative of a k-radical, then t is a monomial over k, Const(k(t)) = Const(k), and $S^{irr} = S_1^{irr} = t$. Conversely, if t is transcendental and hyperexponential over k, and Const(k(t)) = Const(k), then Dt/t is not a logarithmic derivative of a k-radical.

This next corollary ($E_{K/C(x)}$ and $L_{K/C(x)}$ are the exponential and logarithm elementary monomials in a tower of elementary extensions) leads directly to an algorithm for testing new logarithmic and exponential primitives to see if they are monomials.

Corollary 9.3.1. Let $C, x, K, E_{K/C(x)}$ and $L_{K/C(x)}$ be as in Theorem 9.3.1, $a \in K^*$ and $b \in K$. Then,

1. Da/a is the derivative of an element of K if and only if there are $r_i \in \mathbf{Q}$ such that

$$\sum_{i \in L_{K/C(x)}} r_i Dt_i + \sum_{i \in E_{K/C(x)}} r_i \frac{Dt_i}{t_i} = \frac{Da}{a}$$

2. Db is the logarithmic derivative of a K-radical if and only if there are $r_i \in \mathbf{Q}$ such that

$$\sum_{i \in L_{K/C(x)}} r_i Dt_i + \sum_{i \in E_{K/C(x)}} r_i \frac{Dt_i}{t_i} = Db$$

12.6 Function Fields (Stichtenoth)

Stichtenoth, “Algebraic Function Fields and Codes”.

Let K be a field.

F/K is called *rational* if $F = K(x)$ for some $x \in F$

A valuation ring of a function field F/K is a ring $O \in F$ such that $K \in O \in F$ (both proper inclusions) for every $z \in F$, either $z \in O$ or $z^{-1} \in O$

For the rational function field $K(x)$, there's a valuation ring for each irreducible polynomial

A place of a function field F/K is the unique maximal ideal of a valuation ring O of F/K.

Every element of $t \in P$ such that $P = tO$ is called a prime element, or local paramter, or uniformizing variable of P).

P determines O uniquely. O_P is called the valuation ring of the place P

Definition 1.4.1 - a divisor is a formal sum over the places of F/K

12.7 The Riemann-Roch Theorem

Various proofs of the Riemann-Roch Theorem:

Fulton (Chapter 8) - uses nonsingular model of curve

Milne (Theorem 14.6) - assumes curve is nonsingular

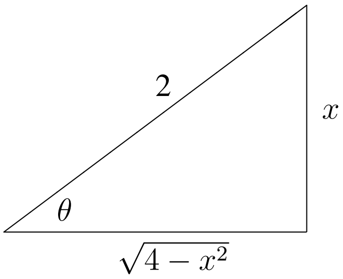
Stichtenoth - constructs divisor using places (Cartier divisor)

Vakil (Eq 18.4.2.1) - unclear to me if he assumes regularity

12.8 Examples

Example 12.1. 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:

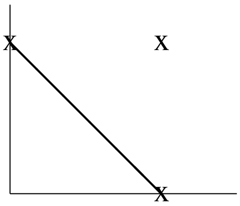
$$x = \frac{1}{u} \qquad y = \frac{1}{v}$$

$$\left(\frac{1}{v}\right)^2 = 4 - \left(\frac{1}{u}\right)^2 \Longrightarrow 4u^2v^2 - v^2 - u^2 = 0$$

$$\int \frac{1}{v}\, d\left(\frac{1}{u}\right) \Longrightarrow -\int \frac{1}{vu^2}\, du$$

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

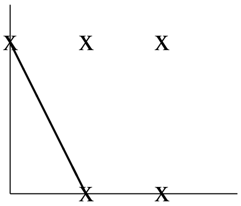
$$\frac{1}{v} = \pm(-i\frac{1}{u} + 2iu + \cdots) \qquad @(0,0)$$

$$\frac{1}{vu^2}\,du = \pm\left[-i\frac{1}{u^3} + 2i\frac{1}{u} + \cdots\right] du \qquad @(0,0)$$

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

$$(u-\frac{1}{2}) = c_1 y + c_2 y^2 + \cdots$$
$$(u-\frac{1}{2})^2 = c_1^2 y^2 + \cdots$$

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:

$$\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \begin{array}{l} \leftarrow (0,0)_a \\ \leftarrow (0,0)_b \end{array}$$

$$\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = 0$$

The solution shows us how to modify the basis:

$$\{u, \frac{iu + (4u^2 - 1)v}{u}\} = \{u, i + \frac{(4u^2 - 1)v}{u}\}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \leftarrow (0,0)_a \\ \leftarrow (0,0)_b \end{array}$$

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

$$\{u, i\frac{1}{u} + \frac{(4u^2 - 1)v}{u^2}\}$$

$$\begin{vmatrix} 1 & -2i \\ 0 & 2i \end{vmatrix} = 2i$$

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

$$2i\ln(ix-y) = 2i\ln\Big(\frac{y}{2} - i\frac{x}{2}\Big) + 2i\ln(-2)$$
$$= 2i\ln\left(\sqrt{1-\Big(\frac{x}{2}\Big)^2} - i\frac{x}{2}\right) = 2i(-i\arcsin\frac{x}{2}) = 2\arcsin\frac{x}{2}$$

(the constant disappears into the constant of integration) and the final answer is:

$$\int\sqrt{4-x^2}\,dx = 2\arcsin\frac{x}{2} + \frac{x\sqrt{4-x^2}}{2}$$

□

12.9 arcsin

Example 12.2. 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 11.27:

$$\begin{array}{lll}
(0,i) & -\dfrac{1}{(u^2-1)}v \text{ @ } (0,i) & = i \\
(0,-i) & -\dfrac{1}{(u^2-1)}v \text{ @ } (0,-i) & = -i \\
(1,0) & -2\dfrac{1}{u(u+1)}v \text{ @ } (1,0) & = 0 \\
(-1,0) & -2\dfrac{1}{u(u-1)}v \text{ @ } (-1,0) & = 0
\end{array}$$

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 11.21 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 11.21 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:

$$
\begin{aligned}
\int \frac{1}{\sqrt{1-x^2}}\,dx &= i\ \ln\Big(\sqrt{1-x^2} - ix\Big) \\
&= -i\ \ln\left(\frac{1}{\sqrt{1-x^2} - ix}\right) \\
&= -i\ \ln\left(\frac{\sqrt{1-x^2} + ix}{1 - x^2 + x^2}\right) \\
&= -i\ \ln\Big(\sqrt{1-x^2} + ix\Big) \\
&= \arcsin x
\end{aligned}
$$

where I used the negative of a logarithm being the logarithm of the inverse, and the last transformation came from section 4.2.

□

Bibliography

[Al14] Paraskevas Alvanos; *Riemann-Roch Spaces and Computation*. De Gruyter Open, 2014. ISBN 978-3-11-042613-7, e-ISBN 978-3-11-042612-0

Free PDF download available from publisher's web site

[Bl47] Gilbert Ames Bliss; *Algebraic Functions*. American Mathematical Society, 1947 ISBN 0821874535, 9780821874530

[Br91] Bronstein, Manuel; *The Risch Differential Equation on an Algebraic Curve*. 1991.

[Br00] Bronstein, Manuel; *Symbolic Integration Tutorial*. 2000.

[Br05] Bronstein, Manuel; *Symbolic Integration I*. Springer, 2005.

[BrCo10] Briggs, Cochran, *Calculus*, 2^{nd} Edition. Pearson, 2010. ISBN-10: 0321336119

[Co93] Henri Cohen, *A Course in Computational Algebraic Number Theory*. Graduate Texts in Mathematics 138. Springer, 1993.

[Co99] Henri Cohen, *Advanced Topics in Computational Number Theory*. Graduate Texts in Mathematics 193. Springer, 1999.

[DeWe82] Dedekind, R., Weber, H. (1882). Theorie der algebraischen Functionen einer Veanderlichen. *J. reine angew. Math.*, **92**:181-290.

[Fu08] Fulton, William, *Algebraic Curves, An Introduction to Algebraic Geometry*, 3^{rd} Edition, 2008.

`http://www.math.lsa.umich.edu/∼wfulton/CurveBook.pdf`

[Ge92] Geddes, Czapor, Labahn, *Algorithms for Computer Algebra*. Springer. ISBN: 978-0-585-33247-5

[Go14] Goodman, *Algebra: Abstract and Concrete*, 6^{th} Edition. SemiSimple Press. ISBN: 978-0-9799142-1-8

[Gu05] Victor Guillemin. 18.117 Topics in Several Complex Variables. Spring 2005. Massachusetts Institute of Technology: MIT OpenCourseWare, https://ocw.mit.edu. License: Creative Commons BY-NC-SA.

[He02] Hess, F. (2002). Computing Riemann-Roch Spaces in Algebraic Function Fields and Related Topics, Journal of Symbolic Computation, 33(4), 425-445.

[Ko07] Kollár, János (2007), Lectures on Resolution of Singularities, Princeton: Princeton University Press, ISBN 978-0-691-12923-5

Kollár's Resolution of Singularities – Seattle Lecture, `https://arxiv.org/abs/math/0508332`, is a more advanced treatment of resolution of singularities on varieties in general, while his Princeton University Press book contains two introductory chapters on resolution for curves (the only one relevant here) and resolution for surfaces before it gets into the more general theory.

[Sh61] Shimura, Goro; Taniyama, Yutaka (1961), *Complex multiplication of abelian varieties and its applications to number theory*, Publications of the Mathematical Society of Japan, 6, Tokyo: The Mathematical Society of Japan, MR 0125113

Later expanded and published as Shimura (1997)

[St09] Stichtenoth, Henning. *Algebraic Function Fields and Codes*, 2nd edition. Springer-Verlag. Graduate Texts in Mathematics 254.

[SwHu06] Swanson, Irena; Huneke, Craig (2006), *Integral Closure of Ideals, Rings, and Modules*, Cambridge University Press, Cambridge.

Available on-line at `https://www.math.purdue.edu/ iswanso/book/`

[Ri70] Robert Risch; *The Solution of the Problem of Integration in Finite Terms*. Bulletin A.M.S, Vol 76, pp. 605-608.

[Ru76] Rudin, Walter (1976). *Principles of mathematical analysis (Third ed.)*. New York. ISBN 0-07-054235-X.

[Tr84] Trager, Barry; *Integration of Algebraic Functions*, Ph.D. thesis, Massachusetts Institute of Technology, 1984.