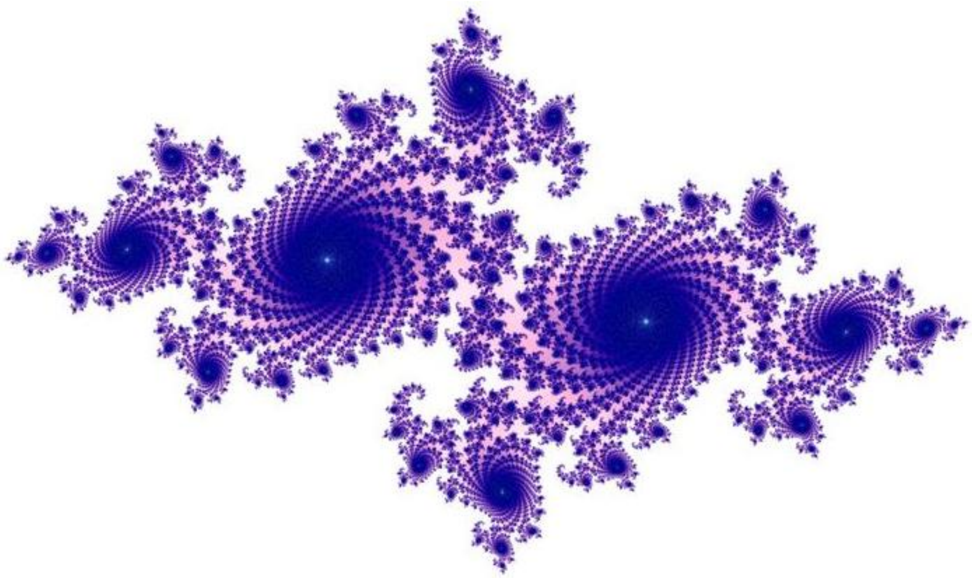
Advanced Topics in Information Dynamics

Christopher Goddard
University of Melbourne

Advanced Topics in Information Dynamics



Chris Goddard
February 17, 2010

Abstract

This work is sequel to the book "A Treatise in Information Geometry", submitted to vixra in late 2009. The aim of this dissertation is to continue the development of fractal geometry initiated in the former volume. This culminates in the construction of *first order self-referential geometry*, which is a special form of 8-tensor construction on a differential manifold with nice properties. The associated information theory has many powerful and interesting consequences.

Additionally within this treatise, various themes in modern mathematics are surveyed- Galois theory, Category theory, K-theory, and Sieve theory, and various connections between these structures and information theory investigated. In particular it is demonstrated that the exotic geometric analogues of these constructions - save for Category theory, which is foundational - form special cases of the self-referential calculus.

Contents

Preface

After I completed my PhD, and managed to bring my other work up to the standard I thought worthy of publication, I made both of these available on my webpage in early 2009. However, as is the way with these matters, there were a number of things that were left unfinished with my previous project [Go].

In particular I was still interested in understanding some of the dynamics of more complex phenomena, such as plasticity, and viscosity in fluid flow. And then there were deeper aspects, such as what I thought of at the time as rank manifolds. Furthermore I had a desire to push towards concrete applications, or at least to the point where I could make some significant progress towards some new developments. And then there were still a few unanswered questions, at least one fundamental - such as the question of the optimality of the Fisher information functional for a given geometric structure.

So this project grew out of a need to continue and resolve various aspects of the dialogue I began several years ago. As is the nature of such things, I was not entirely happy with some of the outcomes of the previous project, so as I was in the initial stages of writing this dissertation I revisited, and, in places, updated my previous writings. The main instance of this was, naturally, the turbulent geometry.

This project was also motivated by the observation that there are particular abstract areas of mathematics that have been recently developed, such as K theory, for which there does not seem to be a satisfactory treatment in the literature as to the connection with more concrete phenomena. Since furthermore I was not entirely familiar with these areas of mathematics, I saw this as a great opportunity to simultaneously study, and also synthesise this knowledge into a form compatible with appropriate information theoretic constructions.

Apart from the survey chapters, there are three major new strands to this dissertation. One, as mentioned, is that on turbulent geometry. Intuition was difficult to build here, other than that I was looking for a fully non-perturbative geometric structure associated to my previous ideas.

The next strand is the idea of multiplicative structure, which, roughly speaking, is a way of introducing a natural generalisation of the concept of self multiplication on a geometric structure. Understandably, the ultimate formulation is slightly detached from the original intuition. Nonetheless the construction of geometric precursors is quite straightforward - if f, g are signal functions then $\star(f; g)$ is just fg. This is related to fluid viscosity.

Another phenomena I examine is the notion of plasticity, in which one deals with a pair of embedded information functionals. This actually is another form of $\star$ structure in disguise, albeit with the self-multiplication in the stack, rather than the base.

To be more precise, recall that $\sigma_{ij} = \Lambda_{ijkl}\epsilon_{kl}$, where σ_{kl} is the *stress* tensor, ϵ_{kl} is the strain tensor, and Λ is the plasticity tensor. If the right hand side is viewed as a density it becomes logical to consider ϵ_{kl} as the Ricci-Cartan curvature tensor for some underlying metric τ, or the information density of an underlying structure. Then in turn it becomes logical to compute the information density associated to σ, and compute $I = \int_M R(\sigma) = \int_M R(\Lambda : R(\tau))$.

Transcendental geometry, the final strand, uses a quite similar idea - $\wedge(f; g) := f^g$, and is related to lattice dynamics, or "reverb". The general idea here is that one is considering processes that arise from fractional composition operators.

There are numerous applications of these new structural ideas, and perhaps some of the most interesting are associated with their synthesis into the topic of the last major chapter; that on self-referential geometry. It is to be emphasised however that I have really only indicated roughly how various problems - in economics, theoretical physics, and pure mathematics - might become more tractable upon application of the new techniques.

Consequently there is a lack of depth and exacting rigour to the results that might otherwise be expected from more conventional treatments. Regardless, I believe that the sketches that I have provided are instructive, and do at least make some partial progress in resolution of the associated issues. Hence, as with its companion volume [Go], I have decided to retain them in this dissertation.

In addition to my previous comments, this work, and its companion volume, were necessitated by three different points. One, my overriding curiosity regarding the connections between mathematics and physics. Two, my desire to communicate the value and utility of mathematics in being a universal language to solve all manner of practical and engineering problems. Third, but not least, I feel compelled by my sense of social responsibility to demonstrate that mathematics is very much a living discipline, and, if anything, there are more open problems today than there ever have been at any other point in the past. Throughout this document I have tried to indicate in places the incompleteness of any structural treatment of mathematics of finite complexity, and areas where further work could be done.

I will not pretend that this work deals with the concepts in question in a flawlessly complete and clean fashion. Nonetheless, I still feel that this dissertation communicates the key ideas I wish to get across to a level of clarity that I believe is sufficient.

Subject to peer review, I would also of course be quite interested to have this paper endorsed for submission to the arxiv, if it is judged to be of sufficient quality for publication there.

Acknowledgements

I would like to extend my thanks to the Department of Mathematics at the University of Melbourne for allowing me to extend my affiliation with them beyond the completion of my PhD - in particular the extension of my permissions regarding use of their considerably well resourced library facilities.

I would also like to thank my friends, and particularly my family, for their support while I was undertaking this project.

I should also acknowledge user *Saperaud*, who through Wikimedia Commons, has the image credit for the Julia set on the front page of this work.

Also, this undertaking was greatly assisted by the use of various internet resources. In particular, I have found wikipedia as a first stop a useful tool, followed by the arxiv (and related preprint or postprint servers, such as Project Euclid), in addition to numerous academic blogs and weblogs. I imagine that, as an academic resource, the web will only continue to get better, and I am quite excited by such developments as `bibalex.org`, Scholarpedia, and Citizendium.

Finally I would like to thank the vixra team for providing a publication platform whereby which freelance intelluctuals such as myself are able to make their work available.

Organisation and attribution of work

As in the previous book, my strategy is to divide the book into two main parts. In the first few chapters I survey Galois theory, Category theory, sieve theory, and K-theory, including Grothendieck's theory of schemes. These are primarily for the benefit of the curiosity of the author, though as a general comment Galois theory becomes particularly poignant when one examines the theory of exponential, or plastic geometry, and sieve theory takes on a new significance in the light of the meta-geometry.

My main sources for my initial overview of Galois theory are my third year algebra course notes, as well as Galois' original paper, and the book by Edwards. For category theory my sources are somewhat more extensive. I provide first an initial overview of the developments in axiomatic set theory at the turn of the 20th century, culminating in the work due to Paul Cohen and Kurt Gödel. Gödel's result in particular on incompleteness is extremely important, since it essentially establishes the incompleteness of any mathematical model of finite complexity - such as a physical theory.

I then turn to Goldblatt's book on Topoi for the information that I will need in respect to the theory of 1-categories. For advanced aspects of the theory I have referenced particular revision versions of articles on wikipedia. Next on the agenda is a brief mention of the work on 2-categories due to Daniel Mathews and its relationship with the study of twistors due to Roger Penrose. I then conclude my survey with a couple of applications, which conclude the discussion of various technical issues related to Professor Frieden's work incompletely covered in [Go]. These are the formal construction of the Cramer-Rao inequality for the particular structural categories that are of interest to me, and a proof of the Cencov representation theorem.

The chapter on K-theory is divided into several parts. First, I review the standard work on algebraic geometry, following Hartshorne, Lang, and Atiyah and Macdonald. This includes a review of the Etale cohomology theory due to Grothendieck, as well as Cech cohomology and the Leray spectral sequence. Following this I indicate how the ideas of Grothedieck extend those due to Galois, following the paper by Dubac and de la Vega. Next I finally examine the K-theory, following Max Karoubi's treatment, and culminating in an intuitive description of the Atiyah-Singer index theorem, again assisted to a certain extent by wikipedia.

I conclude the chapter by a brief mention of the L-theory, which is a deeper level of abstraction still.

Next is my review of Sieve theory. Here again my sources are quite varied. I found the Princeton companion to mathematics invaluable here, particularly in obtaining a vague understanding of solitons and sieves. The sieve theory is treated first, and, apart from the companion, my sources are the book by Glyn Harman and that by Alina Cojocaru.

My attention is then directed to the examination of integrability. Here my primary sources were the books by Ashok Das and A.V. Mikhailov. I examine the theory of solitons, followed by an examination of the current literature on instantons and its relationship to infinite hierarchies of structure. The chapter is concluded by a mention of the statistical information theory due to Amari, and a discussion of how (∞, n)-categories might be the required construction for the realisation of the n-categorical dual; which I refer to as n^*-categories, which would provide a foundation for the study of the statistical dual to geometric structures. This is of interest since we have the important observation that n-categories are a substructure of n^{**}-categories in general, which is related to quite deep structural theory, significantly beyond the scope of this dissertation.

Following this relatively substantial overview of the state of the field I then turn to the first original part of this work, the general theory of geometric turbulence. This is followed by further research chapters on viscoplastic geometry, as well as transcendental geometry. Ample applications of these techniques are developed concurrently with the structural theory.

The penultimate chapter focuses on the theory of self-referential geometry. This is a synthesis of the three separate structural streams subject to investigation in earlier chapters. Following this, the final chapter suggests further possible directions of research, together with some minor concluding remarks.

Chapter 1

A review of Galois Theory

Here I will provide an overview of the key methods and concepts underlying Galois theory, leading up to the key result due to the fellow, which was to settle the long standing open problem as to whether one can find solutions by radicals to algebraic equations of degree five or higher. My sources will be primarily my third year undergraduate algebra notes [Mill], together with the book by Edwards [Ed] which takes a rather less direct approach.

These results of course hold interest in and of themselves, and the theory itself is a glorious example of the power of creative mathematical invention. Nonetheless we will be more interested in the connection to transcendental number theory, which is important when I begin to discuss geometric viscosity at a later stage. Certain key problems, such as the Hodge conjecture, are known to essentially be statements about transcendental number theory. One would hope that a greater understanding of these matters might allow one to make some progress.

1.1 Introduction

1.1.1 Symmetric Polynomials

The beginnings of Galois theory go back to the following observation due to Newton, which I provide here without proof:

Theorem 1.1.1. *Any symmetric polynomial in the roots of an equation can be expressed in terms of the coefficients of that equation.*

This may be restated in a slightly more general form, by removing dependence on the nature of the roots, which may not be defined in terms of the ring one is working over (if it is not algebraically complete).

Theorem 1.1.2. *(Fundamental Theorem on symmetric polynomials). Any symmetric polynomial in $r_1, ..., r_n$ can be expressed as a polynomial in terms of the symmetric functions $\sigma_1, ..., \sigma_n$, where*

$$\sigma_i = \Sigma_{j=1}^{C(n,i)} S_j(i)$$

where $S_j(i)$ is the jth unique choice of i letters from the set $\{r_1, .., r_n\}$, and $C(n, i)$ is the number of ways i items can be removed from a bag containing n unique objects. Furthermore, if the symmetric polynomial has integer coefficients, then the corresponding polynomial in the σ_i will also have integer coefficients.

Proof. (pp 9-12, Edwards) □

1.1.2 Resolvents and solution of the cubic

Suppose that one has a cubic with roots x, y, z. Consider the quantity $t := x + \alpha y + \alpha^2 z$, where α is a (nontrivial) cube root of unity. Lagrange in his original work on this subject referred to this as the *resolvent*. Then t may take potentially six values, dependent on the ordering of the original roots x, y, z. These in turn will be the roots of a sixth degree equation, the *resolvent equation*

$$f(X) = \Pi_{p \in P(3)}(X - t(p))$$

where $P(3)$ is the 3rd order permutation group. But the coefficients of this equation will be symmetric and hence expressible in terms of the original coefficients of the given cubic by our theorem from before. This is a solvable equation since it is a quadratic equation in X^3 and can be solved by solving a quadratic and taking a cube root. (In the language of Galois theory still to come, we are finding the roots of our cubic by taking a field extension of degree 2 over the rational polynomial ring generated by our original cubic equation.)

1.1.3 Cyclotomic equations

A key question in the theory of the algebraic solution of equations is whether roots of unity, which have the transcendental expression $exp(2\pi ik/n)$ (for nth roots), can be expressed algebraically. This is a key question in taking roots more generally, since if $N^{1/k}$ is the kth root of N, then $\alpha N^{1/k}$ is also a kth root if α is a kth root of unity.

This leads one to the examination of the equation $x^n = 1$ and to try to find algebraic expressions for x. We can reduce this probably further, if we note that for $n = jk$ with $gcd(j, k) = 1$ then a primitive jth root of unity which is also a primitive kth root must be a primitive nth root of unity. It follows very easily that we need only consider the case where n is prime.

The first few primes can be shown to have algebraic roots via relatively simple calculation. For larger numbers, however, we have to be slightly cleverer.

Lemma 1.1.3. *Let p be prime. Then there is a integer g such that for all integers not congruent to* 0 *mod p, we have that they are congruent to a power of g mod p. (We will call g a primitive root mod p)*

Proof. By Fermat's little theorem, note that for any integer $k \neq 0$ mod p we have that k^{p-1} is congruent to 1 mod p. Then k is such an integer. □

In particular we have as a consequence that if α is a pth primitive root of unity then $\alpha, \alpha^g, ..., \alpha^{g^{p-2}}$ are all non-trivial roots of unity. We then may write the associated Lagrange resolvent as

$$t = \alpha + \beta\alpha^g + ... + \beta^{p-2}\alpha^{g^{p-2}}$$

where β is a $p-1$st root of unity. Then for $\alpha \mapsto \alpha^g$ we have $t \mapsto \beta{-}1t$, since we would have $\alpha^g + \beta\alpha^{g^2} + ... + \beta^{p-2}\alpha = \beta^{-1}(\beta^{p-1}\alpha + \beta\alpha^g + \beta^2\alpha^{g^2} + ... + \beta^{p-2}\alpha^{g^{p-2}}$.

Then since $t^{p-1} = (\beta^{-1}t)^{p-1} = t^{p-1}$, this is invariant under the transformation $\alpha \mapsto \alpha^g$. For similar reasons, t_it^{p-1-i} is also invariant, where t_i is t with $\beta \mapsto \beta^i$, since $\alpha + \beta^i\alpha^g + ... + \beta^{(p-2)i}\alpha^{g^{p-2}} \mapsto \alpha^g + \beta^i\alpha^{g^2} + ... + \beta^{(p-2)i}\alpha = \beta^{-i}t_i$, so that $(\beta^{-i}t_i)^{p-1-i} = (\beta^{-(p-1-i+i)}t_i^{p-1-i}) = t_i^{p-1-i}$.

Suppose that we know how to compute the t_i. Then I claim it follows that

$$\alpha = (p-1)^{-1}(t_1 + ... + t_{p-1})$$

since $\Sigma_i t_i = (p-1)\alpha + (\beta+\beta^2+...+\beta^{p-1})\alpha^g + ... + (\beta+\beta^2+...+\beta^{p-1})\alpha^{g^{p-2}}) = (p-1)\alpha$ as β is a primitive root of unity.

To complete our construction of α we need another, slightly harder lemma, which requires the full machinery of Galois theory to resolve.

Lemma 1.1.4. *(Edwards, p.27) Suppose p is prime, and α is a primitive pth root of unity. Let β be a primitive $p-1$st root of unity. If $P_1(\beta), ..., P_{p-1}(\beta)$ are polynomials in β with rational coefficients and if $P_1(\beta)\alpha + ... + P_{p-1}(\beta)\alpha^{p-1} = 0$ then $P_1(\beta) = 0, ..., P_{p-1}(\beta) = 0$.*

Now we may write t^{p-1} as a polynomial in the form of the lemma, plus some function of β times the trivial root, up to some reordering of $1, ..., p-1$ in powers of g. Since it is invariant under transformation $\alpha \mapsto \alpha^g$, we conclude by the lemma that all coefficients must be equal.

So then

$$t^{p-1} = P_0(\beta) + P_1(\beta)(\alpha + ... + \alpha^{p-1})$$

which is merely a function of β, since one can see via an easy geometric argument that the summation of all kth primitive roots of unity excluding the trivial root will give -1. So t can be computed as the $p-1$st root of a known quantity.

It follows similarly that the t_i can be computed and we are done.

1.2 Algebraic digression

I will assume that the reader is passingly familiar with groups, rings, ideals, and various other minor technicalities. However there are certain elementary concepts in algebraic theory that have a certain degree of importance to the treatment to follow, so I will deal with these quickly now. My source here will be a course I took in 2003 given by Chuck Miller [Mill].

1.2.1 Factorisation Domains

Definition 1. A ring R is an integral domain if $0 \neq 1$ and R contains no zero divisors.

Definition 2. Suppose R is an integral domain. Then an element u of R is a unit if there is some $v \in R$ with $uv = 1$.

Definition 3. If $s, t \in R$ we say s divides t or $s|t$ if $t = sa$ for some $a \in R$.

Definition 4. s and t are associates $s \sim t$ if $s|t$ and $t|s$.

Definition 5. (Principal Ideal). Let R be a commutative ring. Then if $s \in R$, the 2-sided ideal Rs is called the principal ideal associated to s, (s).

Definition 6. (Irreducible element). An element $r \in R$ is irreducible if it is not a unit, and if for every a, b such that $r = ab$ we must have that either a or b is a unit.

Definition 7. (Unique Factorisation Domain (UFD)). An integral domain R is a UFD if it satisfies the axioms

(UF1) Every nonzero $r \in R$ can be expressed as $up_1...p_n$ for $n \geq 0$, u a unit and the p_i irreducible,

(UF2) If $up_1...p_n = vq_1...q_m$ where u, v are units and the p_i, q_j are irreducible, then $n = m$ and the p_i, q_j are the same up to some reordering.

Examples:

(i) Z is a UFD.

(ii) Fields are UFDs.

(iii) $R = \{a + b\sqrt{-5} | a, b \in Z\} = Z[\sqrt{-5}]$ is not a UFD.

Definition 8. A nonzero element $p \in R$ is prime if it is not a unit and if $p|ab$ implies $p|a$ or $p|b$.

Theorem 1.2.1. *If R is an integral domain that satisfies UF1, then R satisfies UF2 iff all irreducibles in R are prime.*

The following notion is inspired via the Euclidean algorithm.

Definition 9. (Euclidean Domain (ED)). A Euclidean Domain is an integral domain R which has a Euclidean function $\phi : R^* := R\ \{0\} \to Z_{\geq 0}$ so that

(ED(1)) If $a|b$ then $\phi(a) \leq \phi(b)$.

(ED(2)) Given $a \in R$, $b \in R^*$ there are $q, r \in R$ so that $a = bq + r$ where either $r = 0$ or $\phi(r) < \phi(b)$.

Essentially a Euclidean Domain is a space "where the Euclidean algorithm works". In the case of the integers, $\phi(n) = |n|$. In the case of $F(x)$ with F a field, we have that this is a Euclidean Domain if we choose $\phi(p(x)) = deg(p(x))$.

Definition 10. (Principal Ideal Domain (PID)). An integral domain R is a PID if every ideal in R is principal.

Lemma 1.2.2. *If R is a ED, then R is a PID.*

Theorem 1.2.3. *If R is a PID, then R is a UFD.*

1.2.2 Fields

An example of a particular class of fields are the number fields - subfields of the complex numbers. These are fields such as the rationals, Q, and the real numbers, R. Then there are the finite fields, $Z_p = \{0, 1, ..., p-1\}$ where p is prime.

The definition of a field is relatively straightforward:

Definition 11. (Field). A field is a set that has operations of addition and multiplication, and is closed under both. It also is closed under the inverses of these operations.

Definition 12. (Subfield). A subfield is a subset of a field that remains closed under the restrictions of the original operations.

Definition 13. If the identity element of a field F is equal to the zero after a finite number k of additions, then this number is called the *characteristic* of the field, and, furthermore, it follows that Z_k is a *subfield* of F, $Z_k \subset F$. If such a k does not exist then the characteristic is said to be zero, and we have that $Q \subset F$.

For a more interesting example of the idea of a field, note that, starting with any field F one can take the associated polynomial ring $F[x]$. From this one can form the field of functions $F(x)$. Such fields are of considerable interest in Galois theory, as they are a natural structure within which one can build extensions to a field F.

1.2.3 Field Extensions

Definition 14. If F, K are fields with $F \subset K$ we will say that K is an extension of F.

Suppose now $\alpha \in K$. We note that there is a map $\phi : F[x] \to K$ such that it is the identity on F and it sends x to α. Furthermore, if f is any polynomial we note that $\phi \circ f(x) = f(\alpha)$.

If ϕ is injective, we call α transcendental over F, and if one were to denote the smallest subfield of K containing F and α to be $F(\alpha)$ we have that $F(\alpha)$ is isomorphic to the field of fractions of $F[x]$, with $F \subset F(\alpha) \subset K$.

Suppose on the other hand that ϕ is not injective. Then since $F[x]$ is a principal ideal domain (PID) we have that $ker(\phi) = (f(x))$ for some monic function f, and $f(\alpha) = \phi \circ f(\alpha) = 0$. In other words, α satisfies some finite degree polynomial over F, or is *algebraic* over F.

One of the key problems of Galois theory is to show when things are transcendental and when they are not. In particular to demonstrate that one can solve an equation by radicals one must have all field extensions algebraic.

Now $im(\phi) \subset K$ and hence is an integral domain. Therefore $f(x)$ is irreducible over F, via the lemma

Lemma 1.2.4. *If $I \neq R$ is an ideal, then I is a prime ideal iff R/I is an integral domain.*

Hence $F[x]/(f(x)) = F[x]/ker(\phi)$ is a field and so

$$F(\alpha) \cong F[x]/(f(x))$$

In particular we have a finite dimensional vector space over F with basis $1, \alpha, \alpha^2, ..., \alpha^{m-1}$ where $m = deg(f)$.

Definition 15. For any $F \subset K$, we call $dim_F K$ the degree of the extension, and define $[K : F] = dim_F K$.

Example. If α is algebraic over F with irreducible polynomial $f(x)$, then $deg(f(x)) = [F(\alpha) : F]$.

It is fairly easy to see that the following is true:

Proposition 1. *If $F \subset L \subset K$ are field extensions, then $[K : F] = [K : L][L : F]$*

Corollary 1.2.5. *If K is an extension of F of degree n and $\alpha \in K$, the $[F(\alpha) : F] | n$.*

Theorem 1.2.6. *(Field isomorphism theorem). Suppose that $\alpha \in K$ and $\beta \in L$ are two elements which are algebraic over F. Then there is an isomorphism $F(\alpha) \to F(\beta)$ iff the irreducible polynomials for α and β are the same.*

Proof. If such an isomorphism exists, then $f(\alpha) = 0$ iff $\theta \circ f(\alpha) = 0$ iff $f \circ \theta(\alpha) = 0$ iff $f(\beta) = 0$. Conversely, if they have the same irreducible polynomial $f(x) \in F[x]$, we have that there are isomorphisms $\psi : F[x]/(f(x)) \to F(\alpha) \subset K$, and $\phi : F[x]/(f(x)) \to F(\alpha) \subset L$. Then $\theta = \psi \circ \phi^{-1}$ is the desired isomorphism. □

Remark. More generally, if $\theta : K \to K'$ is an isomorphism of extensions K, K' of a field F, then α is a root of $f(x) \in F[x]$ in K iff $\theta(\alpha)$ is a root of $f(x)$ in K'.

Example. Consider the polynomial $x^3 - 2 \in Q[x]$. This is irreducible, as is the polynomial $x^4 - 5 \in Q[x]$. These have roots $\alpha = 2^{1/3}$ and $\beta = 5^{1/4}$. We are interested in the extension of the rationals with both these roots adjoined, $Q(\alpha, \beta)$. Then note that as $Q(\alpha), Q(\beta)$ are subfields of degree 3 and 4 respectively over Q that 3 and 4 must divide $[Q(\alpha, \beta) : Q]$. So 12 divides $[Q(\alpha, \beta) : Q]$. But it is clear that this is all there is, as $x^3 - 2$ and $x^4 - 5$ are irreducible and hence minimal polynomials. So this is the degree of our extension.

1.2.4 Ruler and Compass Constructions

As an application of these ideas, we will look at ruler and compass constructions. Points, lines, and circles are said to be constructible if, given two initial points at "unit" length, one can generate using a straight edge and compass such objects.

For instance, one can inscribe a circle of radius one using one point as the centre and the other as a point on the edge. One may then draw a line joining the two points intersecting the circle at a diametrically opposite point; these points then can be used to draw two circles of radius two, which intersect at points which in turn can be connected through the centre of the original circle. In this way, one can construct right angles. Furthermore, it is possible to bisect any angle using such methods.

Definition 16. A real number is constructible if a segment of length $|a|$ is constructible from the two original points (at distance 1).

Proposition 2. *The constructible numbers are a subfield of the real numbers.*

Proposition 3. *If $a \in R$ is constructible and $a \geq 0$, then $\sqrt{a}$ is constructible.*

Proof. For just construct a segment of length a in one direction, say "left", and a segment of length one towards the "right" from some reference point 0. Construct the midpoint of the segment of length $a + 1$, then build a circle with centre the midpoint. Then we will have two similar triangles, and the length of the vertical side can easily be seen to be $\sqrt{a}$. □

Proposition 4. *All field extensions due to construction are at most of degree* 2*.*

Proof. The equation of a line is linear. The equation of a circle is quadratic. The intersection of a line and a circle is again a quadratic equation. Hence all constructible numbers obtained as solutions to an extension induced by $\alpha x^2 + \beta x + c = 0$ will be of the form $F(\sqrt{\beta^2 - 4\alpha c})$. □

It is then easy to see as a consequence that

Theorem 1.2.7. *If $a_1, ..., a_n$ are constructible numbers in R then there is a chain of subfields $Q = F_0 \subset F_1 \subset ... \subset F_m \subset R$ so that (i) each $F_i \subset R$, (ii) $a_1, ..., a_n \in F_m$, (iii) F_{i+1} is obtained from F_i by a quadratic extension, (iv) $dim_Q K = dim_Q F_m = 2^m$, and (v) the a_i are all algebraic over Q.*

Consequently, the problem of trisecting the angle is impossible, since it requires a field extension of degree 3. Similarly, it is impossible to duplicate the cube, as again that requires a field extension of degree 3. (Note, however, that if one uses paper folds to construct numbers the range of possibilities is increased; in particular it *is* then possible to trisect the angle.)

1.3 Field Extensions and Galois Theory

It was my original intention to delve more deeply into Edwards' book on the subject, but I find that my algebra notes are somewhat more concise. So the following section will largely be a typset version of these, almost verbatim, with minor personal annotations and occasional commentary related to the books by Edwards.

1.3.1 Splitting Fields

Definition 17. (Splitting Field) Let F be a field, $f(x) \in F[x]$ a polynomial. An extension K of F is said to be a splitting field for $f(x)$ if

(i) in $K[x]$, $f(x) = c(x - u_1)...(x - u_n)$

(ii) $K = F(u_1, ..., u_n)$

Remark. In other words, not only does K contain all the roots of f, but, moreover, K is a minimal such extension over F.

Example. Let $f(x) = x^4 - x^2 - 2 = (x^2 - 2)(x^2 + 1)$ be a polynomial in $Q[x]$. Then a splitting field for f is $K := Q(\sqrt{2}, -\sqrt{2}, i, -i)$ since $f(x) = (x - \sqrt{(2)})(x + \sqrt{(2)})(x - i)(x + i)$ over K. In particular, we see that

$$F \subset F(\sqrt{2}) \subset F(\sqrt{2}, i)$$

Hence $[K : F] = 4$, and we have a basis $\{1, \sqrt{2}, i, i\sqrt{2}\}$ for K as a vector space over F. This is known as a biquadratic extension.

Theorem 1.3.1. *Suppose $f(x) \in F[x]$ is a nonconstant polynomial of degree n. Then there is a splitting field K of $f(x)$ over F such that $[K : F] \leq n!$.*

Proof. The proof is by induction on the degree of $f(x)$. For the base case, if $f(x) = c(x - u)$ is linear then F is already a splitting field and $[F : F] = 1$.

So suppose that we know the result for polynomials of degree $\leq n - 1$. In $F[x]$ let $p(x)$ be one of the prime divisors of $f(x)$ which we can assume is monic (leading coefficient is one), and we have certainly that $deg(p(x)) \leq n$.

Form $\frac{F[x]}{(p(x))} = F(u_n)$ with u_n a root of $p(x)$. Then in $F(u_n)[x]$ we have $f(x) = c(x - u_n)g(x)$ where $g(x)$ is monic of degree $n - 1$. Also note that $[F(u_n) : F] = deg(p(x)) \leq n$.

By the induction hypothesis applied to $g(x) \in F(u_n)[x]$ there is a splitting field $K_{n-1} = F(u_1, ..., u_{n-1})$ where $g(x) = (x - u_1)...(x - u_{n-1}) \in K_{n-1}[x]$, $u_i \in K_{n-1}$.

Then $K = K_{n-1}(u_n)$ is a splitting field for f over F. Also, it follows by the induction hypothesis that

$$[K : F(u_n)] \leq (n - 1)!$$

so

$$[K : F] = [K : F(u_n)][F(u_n) : F] \leq (n-1)!n = n!$$

□

Remark. As a consequence of the theorem above we have that all algebraic extensions have associated splitting fields.

Theorem 1.3.2. *Let $\sigma : F \to E$ be a field isomorphism, $f(x) \in F[x]$ a non-constant polynomial and $\sigma.f(x) \in E[x]$ its image. (If $f(x) = a_0 + ... + a_n x^n, \sigma.f(x) = \sigma(a_0) + ... + \sigma(a_n)x^n$.)*

Now if K is a splitting field for f over F and L is a splitting field for σf over E then there is an isomorphism from K to L extending σ, call it $\bar{\sigma} : K \to L$.

Proof. This is once more by induction on the degree of $f(x)$. If $f(x)$ has degree 1 then $f(x) = c(x-u)$ and $u \in F$, $K = F(u) = F$, $\sigma f(x) = \sigma(c)(x - \sigma(u))$ and $L = E(\sigma(u)) = E$.

Let $p(x)$ be a monic prime in $F[x]$ dividing $f(x)$. Then in K, $p(x)$ has some root, say u, so $\frac{F[x]}{(p(x))} \cong F(u) \subset K$ and $[F(u) : F] = deg(p(x))$.

Similarly, $\sigma p(x)$ is a monic prime in $E(x)$ and has a root v in L, where $\frac{E[x]}{(\sigma p(x))} \cong E(v) \subset L$.

As before, if we extend σ by putting $\bar{\sigma}(u) = v$ then we get an isomorphism $\bar{\sigma} : F(u) \to E(v)$.

In particular, in $F(u)[x]$ we can factor $f(x)$ as $f(x) = (x-u)g(x)$ where $g(x)$ has degree $n-1$, and it follows that in $E(v)[x]$, $\bar{\sigma} f(x)$ factors as $\bar{\sigma} f(x) = (x - \bar{\sigma}(u))g(x)$.

Now K is still a splitting field for $f(x)$ over $F(u)$ and L is a splitting field for $\sigma f(x)$ over $E(\bar{\sigma}(u)) = E(v)$. Hence K is a splitting field for $g(x)$ over $F(u)$ and L is a splitting field for $\bar{\sigma} g(x)$ over $E(v) = E(\bar{\sigma}(u))$.

Then by induction $\bar{\sigma}$ can be extended to an isomorphism $\hat{\sigma} : K \to L$ which restricts to $\bar{\sigma}$ on $F(u)$ and hence σ on F. □

1.3.2 Normal and Separable Extensions

Definition 18. An extension K of F is said to be normal if every irreducible polynomial $p(x) \in F[x]$ which has a root in K must split into linear factors in $K[x]$, i.e all roots of $p(x)$ lie in K.

Theorem 1.3.3. *The extension K of F is the splitting field of some polynomial in $F[x]$ iff K is a finite dimensional, normal extension of F.*

Proof. ($\leftarrow$) Since K is finite dimensional, it has a basis, say $\{u_1, ..., u_n\}$ over F. Then $K = F(u_1, ..., u_n)$ and each u_i is algebraic over F. Hence, u_i is the root of some irreducible $p_i(x) \in F[x]$.

Put $f(x) = p_1(x)...p_n(x)$.

Since K is normal, each $p_i(x)$ splits into linear factors in $K[x]$. Hence so does $f(x)$. Thus K is a splitting field for $f(x)$.

($\rightarrow$) Suppose K is a splitting field for $f(x) \in F[x]$. Then $K = F(u_1, ..., u_n)$ where u_i are roots of f and hence $[K : F] < \infty$ since each u_i is algebraic over F.

Suppose $p(x) \in F[x]$ is irreducible, and has a root $u \in K$. Let L be a splitting field for $p(x)$ over K. So

$$F \subset F(u) \subset K \subset L$$

Now suppose $w \in L$ is some other root of $p(x)$. Then we know that the identity on F extends to an isomorphism

$$K(w) = F(u_1, ..., u_n)(w) = F(u_1, ..., u_n, w) = F(w)(u_1, ..., u_n)$$

So $K(w)$ is a splitting field of $f(x)$ over $F(w)$. Also K is a splitting field for $f(x)$ over $F(u)$.

Hence the isomorphism $F(u) \rightarrow F(w)$ extends to an isomorphism $K \rightarrow K(w)$.

Now, this isomorphism restricts to the identity on F so

$$[K : F] = [K(w) : F]$$

Consequently $[K(w) : K] = 1$ and $w \in K$. $\square$

Definition 19. (Separable Polynomial). A polynomial $f(x) \in F[x]$ is said to be separable if all its roots in a splitting field are distinct (so it has no repeated roots).

Definition 20. (Separable Extension). An extension $K \supseteq F$ is said to be a separable extension if every $u \in K$ is the root of a separable polynomial $f(x) \in F[x]$.

Definition 21. (Derivatives). If $f(x) = a_0 + ... + a_n x^n$ then its derivative is $f'(x) = a_1 + 2a_2 x + ... + n a_n x^{n-1}$.

Lemma 1.3.4. *If $f(x)$ and $f'(x)$ are relatively prime in $F[x]$, then $f(x)$ is a separable polynomial.*

Lemma 1.3.5. *Suppose that $char(F) = 0$. If $f(x) \in F[x]$ is irreducible, then $f(x)$ is separable and hence every algebraic extension K of F is separable.*

Corollary 1.3.6. *If an extension K of F is not separable, it is not algebraic.*

Proof. (of lemma) Assume $f(x)$ is a non-constant irreducible (prime) polynomialof degree greater than one. Then $f'(x)$ is not constant and has smaller degree. Hence $f(x)$ and $f'(x)$ are relatively prime, and $f(x)$ is separable. □

1.3.3 Primitive elements

Definition 22. (Primitive element). Let $K \supseteq F$ be a finite dimensional algebraic extension. An element $u \in K$ is a primitive element for this extension if $K = F(u)$.

Remark. An example of this is given by the case when one is considering a nontrivial or *primitive* pth root of unity, where p is prime. These were mentioned earlier in my discussion of the cyclotomic equation.

Theorem 1.3.7. *(Existence of a primitive element). Let K be a finite dimensional, separable extension of F. Then K contains a primitive element, ie $\exists u \in K$ with $K = F(u)$.*

Proof. Since K is finite dimensional, $K = F(u_1, u_2, ..., u_n)$. If $n = 1$ there is nothing to prove.

Suffices to prove for $n = 2$:

Let $K = F(v, w)$. Want to find a primitive $u \in K$ so $K = F(u)$. We look for a primitive of the form $u = v + cw$.

For the purposes of this proof I shall assume that F is infinite.

Let $p(x) \in F[x]$ be a minimal polynomial for v with roots $v = v_1, ..., v_n$ in $L \supseteq F$. Similarly suppose $q(x) \in F[x]$ is a minimal polynomial for w, with roots $w = w_1, ..., w_m$ in $L \supseteq F$, where L is the splitting field for $p(x)q(x)$.

Since F is infinite, we can choose

$$c \neq \tfrac{v_i - v}{w - w_j}$$

for all $1 \leq i \leq n$, $1 < j \leq m$.

Then define $u = v + cw$. I claim that this is a primitive element for K, ie that $K = F(u)$.

Let $h(x) = p(u - cx) \in K[x]$. Then we note that $h(w) = p(u - cw) = p(v) = 0$ by definition of $p(x)$.

Now look at some $w_j \neq w$ ($j \neq 1$). If $p(u-cw_j) = 0$ then for some i, $u-cw_j = v_i$. Hence $v + cw - cw_j = v_i$ so

$$c = \tfrac{v_i - v}{w - w_j}$$

This contradicts our choice of c. Hence w must be the only common root of $q(x)$ and $h(x)$.

Let $r(x)$ be the minimal polynomial for w over $F(u)$. Then $r(x)|q(x)$ and $q(x)$ splits and has no repeated roots in $L[x]$. Therefore $r(x)$ has no repeated roots.

But also $r(x)|h(x)$ since $r(x)$ is a minimal polynomial and $h(x) \in F(u)[x]$. Hence $r(x)$ has w as a single root, $r(x) = b(x-w)$ and $w \in F(u)$. Hence $v = u-cw \in F(u)$ and $K = F(v, w) = F(u)$. □

Remark. In other words, what this is telling us is that separable extensions are special - in particular, that they are generated by a single function over the field F.

Example. *Consider $f(x) = x^3 - 2 \in Q[x]$. This happens to be irreducible. Let K be the splitting field of $f(x)$ in the complex numbers. Then for $w = exp(2\pi i/3)$, we have that $f(x) = (x - 2^{1/3})(x - w2^{1/3})(x + w2^{1/3})$.*

Now $\frac{Q[x]}{(x^3-2)} \cong Q(2^{1/3})$ has degree 3 over Q, with basis $1, 2^{1/3}, 2^{2/3}$. Over $Q(2^{1/3})$ we can compute that

$$f(x) = (x - 2^{1/3})(x^2 + 2^{1/3}x + 2^{2/3})$$

The second factor $g(x)$ is irreducible over $Q(2^{1/3})$. Then we can write $\frac{Q(2^{1/3})[x]}{(g(x))} \cong Q(2^{1/3}, w2^{1/3}) = K$ has degree 2 over $Q(2^{1/3})$ and has basis $1, w2^{1/3}$. So we have a basis for the splitting field K with $[K : Q] = 6$, $\{1, 2^{1/3}, 2^{2/3}, w2^{1/3}, w2^{2/3}, w2\}$.

Or alternatively, we see that $K = Q(2^{1/3}, w)$.

Now what this theorem above is telling us is that, since K is a finite dimensional splitting field over an infinite field, there is a number $u \in C$ with minimal polynomial over the rationals of degree 6 such that $K = Q(u)$.

Remark. There is furthermore a parallel with my remarks in the introduction. This minimal polynomial can be thought of as nothing other than the resolvent associated to the solution of the original equation.

1.3.4 Galois groups

Definition 23. (F - automorphism). Let $K \supseteq F$ be an extension. An isomorphism $\sigma : K \to K$ is an F - automorphism of K if $\sigma(c) = c$ for all $c \in F$. (F is fixed by σ).

Definition 24. (Galois group). The collection of all F-automorphisms of K is denoted $Gal(K/F)$, $G(K/F)$ or $Gal_F(K)$ and called the "Galois group of K over F".

Proposition 5. *$G(K/F)$ is a group with operation composition of functions.*

Proof. This is easy to check. □

Proposition 6. *(Permutation of roots). If $f(x) \in F[x]$ and K is a extension field of F containing a root $u \in K$ of $f(x)$, then for any $\sigma \in G(K/F)$, $\sigma(u)$ is also a root of $f(x)$.*

Example. *(i) $F = R, K = C.G(K/F) = \langle \sigma | \sigma^2 = 1 \rangle$ where $\sigma : C \to C$ is nothing other than complex conjugation.*

(ii) $G(C/Q)$ is very large (of cardinality exceeding aleph one).

(iii) $G(R/Q) \cong 1$.

Remark. For the purposes of the original motivation of Galois theory, we are interested in finite dimensional extensions, usually of Q. Note that if $f(x) \in F[x]$ has a root $u \in K$ and $\sigma \in G(K/F)$ then $\sigma(u)$ is also a root of $f(x)$ and σ permutes the roots of f. In fact, it turns out that this in fact completely characterises the behaviour of the Galois group over algebraic extensions.

Theorem 1.3.8. *(Uniqueness). Suppose $K = F(u_1, u_2, ..., u_n)$ is an algebraic extension. If $\sigma, \tau \in G(K/F)$ are such that $\sigma(u_i) = \tau(u_i)$ for $i = 1, ..., n$ then $\sigma = \tau$.*

Proof. Each element u_i is algebraic over F. u_i is a root of a polynomial with coefficients in F.

A spanning set for K can be obtained as 1 together with products of the various $u_{i_1}...u_{i_k}$, for $1 \leq k \leq n$. But σ, τ have the same value on such elements. Hence they must be equal. □

Remark. A consequence of the above is that any such $\sigma \in G(K/F)$ is a linear transformation of K as an F space.

The following theorem is of quite central importance to Galois' proof that polynomials of degree five or higher are insoluble by radicals.

Theorem 1.3.9. *(Permutation Theorem). If K is the splitting field of a polynomial $f(x) \in F[x]$ of degree n, and if $f(x)$ is separable, then $G(K/F)$ is isomorphic to a subgroup of S_n. (The Galois group is a permutation group).*

Remark. Recall that S_n is just the group of transformations on an alphabet of n letters.

Proof. The polynomial $f(x)$ has n distinct roots $u_1, ..., u_n$ in K and $K = F(u_1, ..., u_n)$. So any $\sigma \in G(K/F)$ just permutes the u_is. If $\sigma \in G(K/F)$ fixes all u_is, then σ is the identity (by the previous theorem). □

1.3.5 Intermediate fields and the Galois Correspondence

Definition 25. If $F \subset E \subset K$, then E is an intermediate field.

Suppose $H \subset Gal(K/F)$, $F \subset E \subset K$.

Define K^H to be the fixed field of H

$$= \{c = K | \sigma(c) = c \text{ for every } \sigma \in H\}$$

but this is a subfield of $K \supseteq F$.

Hence there is a correspondence

$$G(K/F) \supseteq H \rightsquigarrow K^H = \text{fixed field of } H$$

and also

$$E = \text{intermediate field} \rightsquigarrow G(K/E)$$

In particular note that $G(K/E)$ is a subgroup of $G(K/F)$. That is, $F \subset E \subset K \Rightarrow G(K/E) \subset G(K/F)$.

The main theorem of Galois theory says that (under certain conditions) these two correspondences are inverses of each other and so the subgroups of $G(K/F)$ correspond bijectively to intermediate fields.

The conditions are that (K is a finite, normal, separable extension) or ($char(K) = 0$, K is the splitting field of a polynomial) or ($[K : F] = |G(K/F)|$).

Theorem 1.3.10. *Let K = finite dimensional extension of F and H a subgroup of $G(K/F)$. Let $E = K^H$. Then K is a simple, normal separable extension of E. (simple means $K = E(u)$ for some $u \in K$)*

Proof. Each $u \in K$ is algebraic over F, and hence over E. Every $\sigma \in H$ sends u to some other root of its minimal polynomail $p(x) \in E[x]$.

So u has a finite number of distinct images $u = u_1, u_2, ..., u_t$ under H (orbit of u under H).

If $\sigma \in H$, then $\sigma(u_1), \sigma(u_2), ..., \sigma(u_t)$ is $\{u_1, ..., u_t\}$ in some order. Put $f(x) = (x - u_1)...(x - u_t)$.

The coefficients of this polynomial are then fixed by H, so $f(x) \in E[x]$.

Notice that the u_i's are distinct and so $f(x)$ is a separable polynomial. Hence we can find (by Theorem on primitive element) a single $u \in K$ with $K = E(u)$. (ie we have the conditions necessary to "compute a resolvent".)

Again the above setup holds. $f(x)$ splits into linear factors in $K[x]$ so the extension is also normal. □

Recall that if K is a finite dimensional extension of F and $H \leq G(K/F)$, $E = K^H$ is the fixed field of H.

Suppose $p(x)$ is the minimal polynomial of some element u over an intermediate field E, with degree n. We have that $K = E(u)$. Then $[K : E] = n$. Distinct elements of $G(K/E)$ map u to distinct elements of K since $K = E(u)$.

Hence $|G(K/E)| \leq n$. Since $H \subset G(K/E)$ we have $|H| \leq |G(K/E)| \leq n = [K : E]$.

For $f(x)$ as above, we have $p(x)|f(x)$. Now $f(x)$ has degree t so $|H| \geq t = deg(f) \geq deg(p) = n = [K : E]$.

Hence $H = G(K/E)$ and $|G(K/E)| = [K : E]$.

$E = K^H$ so $H = G(K/K^H)$.

Hence we have the Galois correspondence:

$$E \rightsquigarrow G(K/E) \subset G(K/F)$$

is surjective.

Definition 26. (Galois extension). $K \supseteq F$ is said to be a Galois extension if it is finite dimensional, normal and separable over F. (In characteristic zero, this means that K is the splitting field of some polynomial in $F[x]$.)

So assume that $K \supseteq F$ is a Galois extension. Let L be an intermediate field $F \subset L \subset K$. Put $L_0 = K^{G(K/l)} =$ fixed field of $G(K/L)$.

Clearly $L_0 \supseteq L$. We would like to show $L_0 = L$.

Suppose $u \notin L$. Now K is also a Galois extension of L. Also u is a root of a minimal polynomial $p(x)$ of degree ≥ 2 where $p(x) \in L[x]$.

The roots of $p(x)$ are distinct (separable) and lie in K (normality).

If v is any other distinct root of $p(x)$ then $\exists \sigma \in G(K/L)$ with $\sigma(u) = v$. Hence $v \notin L$, and also $u \notin L_0$.

Hence $L_0 = L$. That is, $L =$ fixed field of $G(K/L)$.

In particular, if $K \supseteq F$ is Galois then $F =$ fixed field of $G(K/F)$. Also $[K : L] = |G(K/L)|$ and $[L : F] = [G(K/F) : G(K/L)]$, where the notation on the right is the index of the subgroup $G(K/L)$ in $G(K/F)$.

This shows that the Galois correspondence is injective, and completes the proof that it is one-one. In particular if $K \supseteq F$ is a Galois extension then

$$H \subset G(K/F) \rightsquigarrow K^H = \text{fixed field of } H$$

and

$$\text{Intermediate field } L \rightsquigarrow G(K/L) \subset G(K/F)$$

Remark. (Fact). If $F \subset E \subset K$, then E is a normal extension of $F \Leftrightarrow G(K/E) \lhd G(K/F)$, and if so $G(E/F) \cong \frac{G(K/F)}{G(K/E)}$

So we have proved:

Theorem 1.3.11. *(The Fundamental Theorem). Let $K =$ Galois extension field of F (finite dimensional, normal, separable, and if $char(F) = 0$ the splitting field of a polynomial).*

(1) *Then* $E \rightsquigarrow G(K/E)$ *is a bijection between the set of intermediate fields* $F \subset E \subset K$ *and the subgroups of* $G(K/F)$.

Moreover, $[K : E] = |G(K/E)|$ *and* $[E : F] = [G(K/F) : G(K/E)]$ *and* $H = G(K/K^H)$ *and* $E = K^{G(K/E)}$.

(2) *An intermediate field* E *is a normal extension of* $F \Leftrightarrow G(K/E) \lhd G(K/F)$ *and if so then* $G(E/F) \cong \frac{G(K/F)}{G(K/E)}$.

Remark. Note that the Galois correspondence is order reversing; if $E \subset L$ then $G(K/E) \supseteq G(K/L)$.

1.3.6 Examples

As one example, consider $\mathcal{Q}(2^{1/3}) \cong \frac{\mathcal{Q}[x]}{(x^3-2)}$. Then $\mathcal{Q}(2^{1/3})$ is an extension of $\mathcal{Q}$ of order three, and $G(\frac{\mathcal{Q}(2^{1/3})}{\mathcal{Q}}) = \{1\}$.

As a second, consider this time the extension of the rationals $K = \mathcal{Q}(\sqrt{3}, \sqrt{5})$. Then this is the splitting field of $(x^2-3)(x^2-5)$, which has roots $\sqrt{3}, -\sqrt{3}, \sqrt{5}$, and $-\sqrt{5}$.

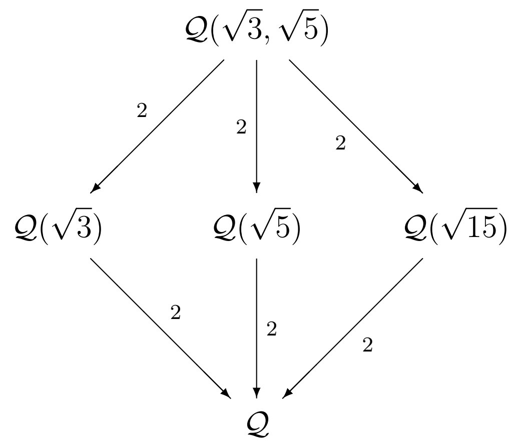
We furthermore clearly have a chain of inclusions $\mathcal{Q} \subset \mathcal{Q}(\sqrt{3}) \subset \mathcal{Q}(\sqrt{3}, \sqrt{5})$ such that the degree of each successive extension is order 2. Consequently $[\mathcal{Q}(\sqrt{3}, \sqrt{5}) : \mathcal{Q}] = 4$, and since this is a Galois extension, we have that $|G(\frac{K}{\mathcal{Q}})| = 4$.

To be more specific, we have that

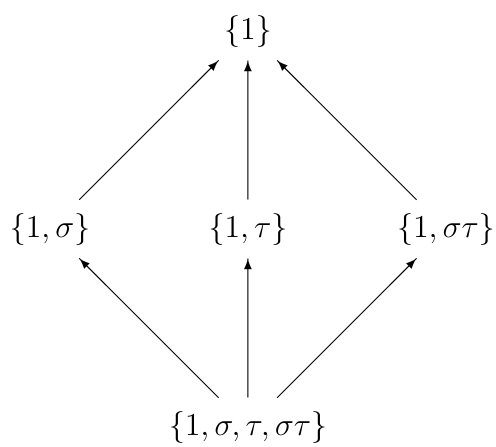
$$G(K/\mathcal{Q}) = \{\sigma^i \tau^j | \sigma^2 = 1 = \tau^2\} = C_2 \oplus C_2$$

where $\tau : \sqrt{3} \mapsto -\sqrt{3}$, $\sqrt{5} \mapsto \sqrt{5}$, and $\sigma : \sqrt{3} \mapsto \sqrt{3}$, $\sqrt{5} \mapsto -\sqrt{5}$ are maps on the roots of the splitting field K.

The structure of the space of subfields of the extension is as follows:



and the corresponding Galois groups have the following diagrammatic structure:



Recall that for our first example we were interested in $x^3 - 2$, but we did not have a splitting field. So now consider the corresponding splitting field over the complex polynomial ring, with roots $2^{1/3}\{1, \omega, \bar{\omega}\}$, where $\omega = exp(2\pi i/3)$.

In particular we have that $f(x) := x^3 - 2 = (x - 2^{1/3})(x - \omega 2^{1/3})(x - \bar{\omega} 2^{1/3})$. Consider the corresponding splitting field for f, K.

Then $[K : Q] = 6$, and since K is a Galois extension, we have that $|G(K/\mathcal{Q})| = 6$.

Then $G(K/\mathcal{Q}) \cong S_3 = \langle \sigma, \tau | \sigma^2 = 1 = \tau^3, \sigma\tau\sigma = \tau^{-1} \rangle$.

We have a basis for K over $\mathcal{Q}$ is $\{1, 2^{1/3}, 2^{2/3}, \omega 2^{1/3}, \omega^2 2^{1/3} = \bar{\omega} 2^{1/3}, \omega 2^{2/3}\}$.

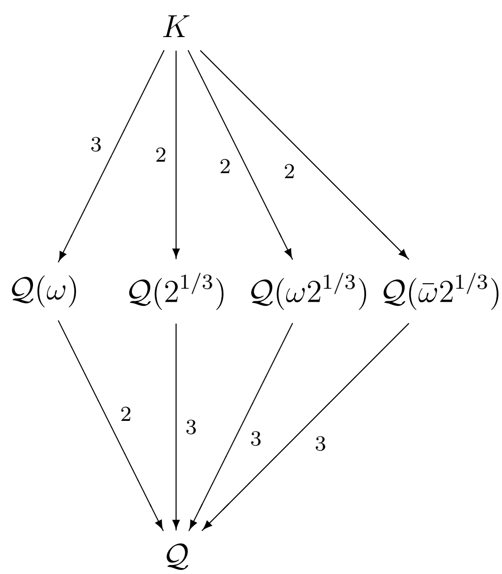
There is an action of $G(K/\mathcal{Q})$ on the generators of this extension, the roots $\{2^{1/3}, \omega 2^{1/3}, \bar{\omega} 2^{1/3}\}$, given by

$\sigma : 2^{1/3} \mapsto 2^{1/3}, \omega 2^{1/3} \mapsto \bar{\omega} 2^{1/3}, \bar{\omega} 2^{1/3} \mapsto \omega 2^{1/3}$ (complex conjugation), and

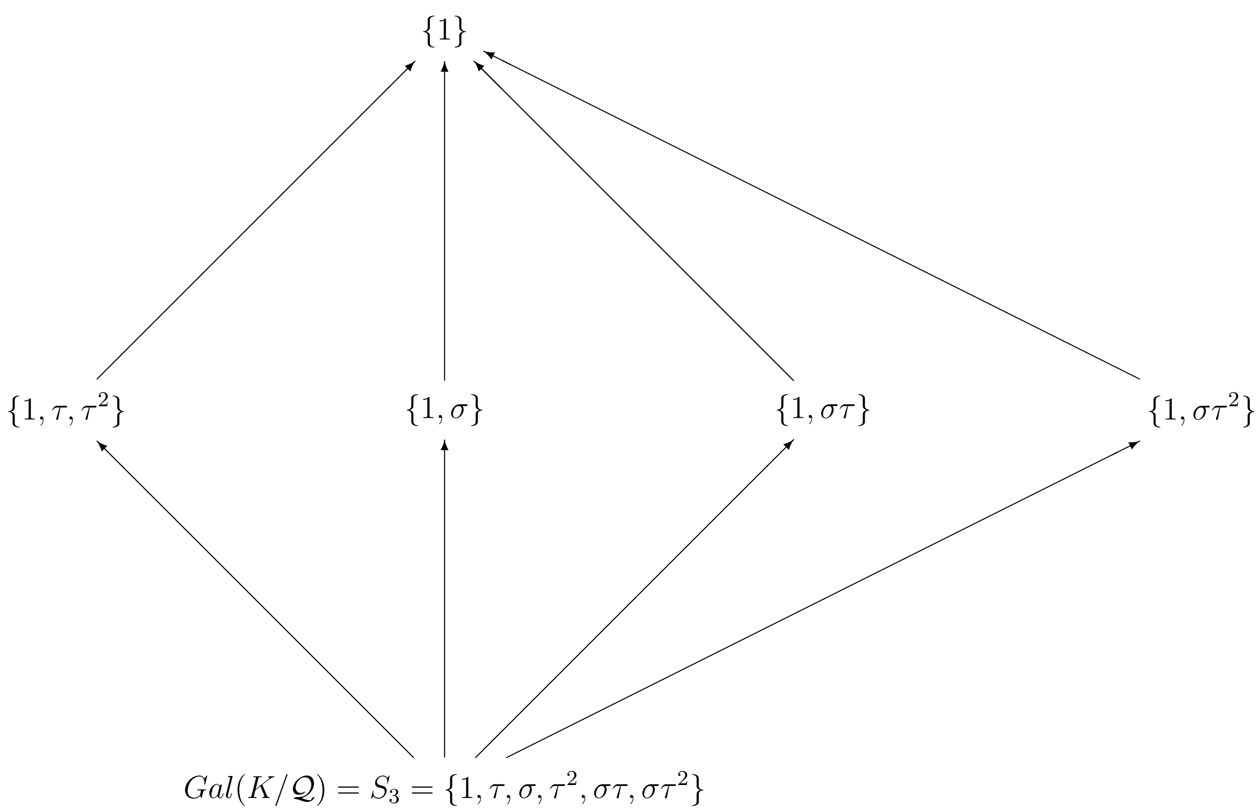
$\tau : 2^{1/3} \mapsto \omega 2^{1/3}, \omega 2^{1/3} \mapsto \bar{\omega} 2^{1/3}, \bar{\omega} 2^{1/3} \mapsto 2^{1/3}$ (rotation).

Note that τ keeps ω fixed, as can be verified without too much effort.

We then have a subfield structure along the following lines:



and corresponding Galois groups which have the following relationship:



In particular it is clear by examination that, for these examples, there is a correspondence between subfields of K and subgroups of $G(K/\mathcal{Q})$. This is in fact generally true, and is the celebrated Galois correspondence, which we established formally in the earlier section.

1.3.7 The problem of solution by radicals

It is well known that $ax^2 + bx + c = 0$ has solutions $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

We say that we can "solve an algebraic expression by radicals" if we can find a closed form for the solution which is a succession of nth roots and rational functions of its coefficients.

An algebraic equation is solvable by radicals iff its solutions are algebraic over the real numbers.

In fact it can be shown that one can find closed form solutions by radicals for the general cubic and quartic equations. However in the early 1800's Abel and Ruffin demonstrated that the quintic equation is not solvable by radicals.

More abstractly, an equation $f(x) \in F[x]$ is solvable by radicals if it has a splitting field contained in a radical extension

$$F = F_0 \subset F_1 \subset ... \subset F_t = K$$

where each $F_i = F_{i-1}(u_i)$ where $u_i^{n_i} - c = 0$ and $c \in F_{i-1}$.

The Galois group of a polynomial $f(x) \in F[x]$ is the Galois group of its splitting field over F.

A group G is said to be solvable if it has a series

$$G = G_0 \rhd G_1 \rhd ... \rhd G_s = 1$$

where G_i/G_{i+1} are abelian.

Lemma 1.3.12. *(Galois' Criterion). $f(x)$ is solvable by radicals $\Leftrightarrow$ its Galois group is solvable.*

Remark. (Fact). For $n \geq 5$, S_n is not solvable; for $n \geq 5$, we have that there is a simple group A_n of index 2 in S_n such that

$$S_n \rhd A_n \rhd 1$$

In particular A_n is not abelian.

Consequently, in general, no polynomial equation of degree exceeding four is solvable by radicals.

Chapter 2

Category Theory

Often in mathematics one comes across constructions and structures that are at first glance different, but are either in correspondence or are separate aspects of a deeper, underlying structure. Consequently one may often prove the same thing several times, when for book-keeping purposes one instance would have been sufficient. It is these considerations that motivate the development of the mathematical language known as category theory.

These notes are primarily based on a survey of [RG], as well as [Su].

2.1 Introduction

2.1.1 Set theoretic foundations and motivation

There are certainly a few famous paradoxes associated to the most naive formulations of set theory. It was originally hoped that perhaps set theory needed to be developed more formally. Hence, in the early years of the twentieth century there was a strong movement by David Hilbert and others to try to axiomatise the foundations of mathematics. Ultimately these efforts were not successful [Gö], but they were instructive.

It is fairly easy to describe what we mean by naive set theory. We have sets, consisting of zero, one or maybe more elements. We have logical operators, such as &, $\vee$, and $\neg$. We have implication symbols, such as $\rightarrow$, $\leftrightarrow$. Finally we have punctuation in the form of parentheses and quantifiers $\forall, \exists, E!$ that operate in the usual way.

The original intuitive development of set theory by Cantor implicitly assumed three axioms.

Axiom 2.1.1. *(Extensionality). Two sets are identical if they have the same elements.*

Axiom 2.1.2. *(Abstraction). For any property there exists a set consisting of the elements that have that property, for instance, the set of even numbers, or the set of numbers that have no factors save for themselves and one.*

Axiom 2.1.3. *(The Axiom of Choice). Informally put, this states that, given a potentially infinite (and maybe even uncountable) number of sets A_i, it is possible to create a new set A containing an element from each A_i.*

A consequence of the Axiom of Choice is that if a set has an ordering defined on it, then the axiom states that it is possible to choose a smallest element. This can be counterintuitive. For instance the set $(0,1)$ does not really seem to have a well defined smallest quantity.

In fact differing formulations of set theory may sometimes assume this last axiom, and sometimes not. For instance, in the Zermelo-Fraenkel formulation of set theory (ZF) the axiom of choice is actually independent of the ZF axioms, that is, neither it nor its negation is provable within ZF. This was demonstrated by Paul Cohen [Cn], using a mathematical technique called *forcing*, which I will not describe here.

The negation of the *continuum hypothesis*, which states that there is no intermediate infinity between the cardinality of the natural numbers and the real numbers, was shown to be unprovable by Gödel in 1940 using ZF, even additionally assuming the axiom of choice (ZFC). In 1963 Paul Cohen demonstrated that not only the negation, but also its converse was not provable in ZFC. Consequently, the continuum hypothesis is independent of the axioms of ZFC.

My personal opinion on the matter is that there should be intermediate infinities between $\aleph_0 := card(N)$ and $\aleph_1 := card(R)$, and that one should be able to construct these somehow by construction of some more abstract form of naive set theory, possibly by consideration of pairs of sets A, A' that are somehow identified to a fractal construction B that is not really a set in the standard sense, but has a well defined cardinality. This is motivated by my previous work on smooth fractal geometry, in [Go].

In particular it should be possible to abstract to ideas of infinity $\aleph_{(x_1,...,x_n)}$, where $x := (x_1, ..., x_n)$ is a point in an n-dimensional space. More formally, we might like to write $\aleph(x) := \aleph_x$, where x is an element of a set A, with respect to some abstraction of the concept of cardinality.

Even without the Axiom of Choice, however, there are difficulties with naive set theory. I will illustrate one now:

Remark. (Russell's Paradox). Consider the set of all sets that are not members of themselves. This is permitted by the Axiom of Abstraction. In particular, we have that we are allowed to write

$$(\exists y)(\forall x)(x \in y \rightarrow x \notin x)$$

Here $\phi(x) = x \notin x := \neg(x \in x)$ is a formula in which x is not free, so it is allowed. But then, for $x = y$, $y \in y \leftrightarrow y \notin y$, which is a contradiction.

It is then immediate that the Axiom of Abstraction is too strong, and needs to be restricted in generality. This ultimately leads to the reformulation of this Axiom as the Axiom Schema of Separation:

Axiom 2.1.4. *(Axiom of Separation).* $(\exists y)(\forall x)[x \in y \leftrightarrow x \in z \& \phi(x)]$

In other words, this axiom states that we must have both that $\phi(x)$ is true and x an element of the set z in order for the existence of y to be guaranteed.

There are many other paradoxes, of either this variety (logical and / or mathematical) or semantic ("I am lying"), but I will not go into these here. Nonetheless, there are many documented in the literature. These motivate the proper development of axiomatic set theory, to be described next.

2.1.2 Axiomatic Set Theory

I will give an indication of the Zermelo-Fraenkel (ZF) formalism here (which is intuitively easier to understand than the alternative approaches).

Definition 27. (Symbols). ZF set theory uses five forms of symbols: constants, variables, sentential connectives, quantifiers, and parentheses. There are also two primitive constants - $\in$ for set membership, and 0 to represent the empty set. The identity symbol is equality $=$. Variables are the letters $a, b, c, ..$, the sentential connectives are the symbols $\neg, \&, \vee, \rightarrow, \leftrightarrow$. For quantifiers we use $\forall, \exists, E!$. Punctuation is given by left and right parentheses.

We would like to be able to quantify when a sentence written in this language is syntactically correct. This motivates the definition of a primitive atomic formula, so that we may define a primitive formula by induction.

Definition 28. (Primitive atomic formula). A sentence $(v \in w)$ or $(v = w)$, where v, w are general variables or 0, is a *primitive atomic formula*.

Definition 29. (Primitive formula). A primitive formula satisfies the following properties:

(i) Each primitive atomic formula is a primitive formula.

(ii) If P is a primitive formula then so is $\neg P$, its negation.

(iii) If P and Q are primitive formulae, then $(P \& Q)$, $(P \vee Q)$, $(P \to Q)$, and $(P \leftrightarrow Q)$ are primitive formulae.

(iv) If P is a primitive formula and v is a variable then $(\forall v)P$, $(\exists v)P$, $(E!v)P$ are primitive formulae.

No other expressions in ZF are primitive formulae unless they satisfy these properties.

As an immediate observation, essentially what the above definition is saying is that any primitive formula P should be closed under allowable logical operations in ZF.

We may now define the concept of an *axiom* within our language:

Definition 30. (Axiom). Suppose we have a primitive formula of the form

$$(\exists v)((\exists w_1)(w_1 \in v \vee v = 0)\&(\forall w)(w \in v \leftrightarrow w \in u\&\phi))$$

Then this is an axiom provided that v is not u or w_1, and that it is not separable from the primitive formula ϕ.

The first part of the statement essentially requires that, given v we do not have that v can play the role of 0 if it is not 0. The second part is more complex, and it is easy to see that the crucial information regarding our axiom is contained in ϕ.

Now, a general expression written in ZF may not be reduced, in that we might be able to eliminate certain variables and expressions. So consequently we would like allowable formulae to satisfy the property that when they are reduced they are primitive formulae.

We would also like to restrict to formulae that do not introduce new notation or concepts. In particular, if P is an expression, we will say that it satisfies the criterion of non-creativity if there is no primitive Q such that $P \to Q$ is derivable but Q is not. So in other words, we wish to restrict to expressions that can be expressed solely within the existing symbolic system.

Now we are ready to state the axioms of ZF set theory.

Axiom 2.1.5. *(Extensionality).* $(\forall x)(x \in A \leftrightarrow x \in B) \to A = B$.

In other words, if all elements of A are also in B, and vice versa, we should have equivalence between A and B.

Axiom 2.1.6. *(Schema of Separation).* $(\exists B)(\forall x)(x \in B \leftrightarrow x \in A \& \phi(x))$

Here ϕ plays the role of determining whether x has some particular property.

Axiom 2.1.7. *(Pairing Axiom).* $(\exists A)(\forall z)(z \in A \leftrightarrow z = x \vee z = y)$

This allows us, by induction, to form finite unions of objects (as we should be able to).

Axiom 2.1.8. *(Sum Axiom).* $(\exists C)(\forall x)(x \in C \leftrightarrow (\exists B)(x \in B \& B \in A))$

The sum axiom allows the extension of the idea of union of a set of sets. For instance, if $A = \{\{1,2\},\{2,3\},\{4\}\}$ then we want $\cup A = \{1,2,3,4\}$.

Axiom 2.1.9. *(Power Set Axiom).* $(\exists B)(\forall C)(C \in B \leftrightarrow C \subset A)$

ie, for any given set A, we have the existence of an associated powerset B consisting of all subsets C of A.

Axiom 2.1.10. *(Axiom of Regularity).* $A \neq 0 \to (\exists x)[x \in A \& (\forall y)(y \in x \to y \notin A)]$

This final axiom essentially states that for any nontrivial set A in ZF, there is an element x of A such that the intersection of A with x is empty. The motivation here is to try to circumvent the more obvious difficulties such as statements $A \in A$. But simply stating $A \notin A$ as an axiom does not forbid $A \in B$, and $B \in A$. Hence the formulation as above.

These Axioms provide the basis for the ZF formalism.

2.1.3 The definition of a category, and some examples

A category consists of components called "objects" and "arrows". Objects are abstract entities that exist in the category. Arrows $f : A \to B$ can be maps between objects in a category, but more generally are abstract entities associated to the ordering (A, B). Furthermore, if $f : A \to B$ and $g : B \to C$, then the category contains $g \circ f$ as an arrow from A to C. The composition of arrows must also be associative - $(g \circ f) \circ h = g \circ (f \circ h)$. Finally a category must have an identity arrow I_A associated to every object A, such that if $f : A \to B$, then $I_B \circ f = f = f \circ I_A$.

Examples of categories include for instance the category with the single object N, and arrows $m_+ : N \to N$ that are indexed by the natural numbers, such that the arrow $(m+n)_+ = n_+ \circ m_+ = m_+ \circ n_+$. This category can also be extended with arrows $m_\times : N \to N$ such that $(mn)_\times = n_\times \circ m_\times$.

Another example is $Mat(R)$, the category of matrices with real entries. Objects are the naturals $1, 2, 3, ..$ and arrows $m \to n$ are $n \times m$ matrices. We have a natural composition of such arrows $m \to n \to r$ as matrix multiplication of an $r \times n$ matrix with an $n \times m$ matrix.

For a simpler instance, Set is a category, with objects that are sets, and arrows that are maps between sets.

A category D is said to be a *subcategory* of a category C if all D-objects are C-objects and all arrows in D are arrows for the corresponding objects in C.

Given two categories C,D, the *product category* $C \times D$ is defined to have as objects (A, B), where A is a C-object and B is a D-object, and arrows $(a, b) \to (\alpha, \beta)$ given as the product map $(f, g) : (a, b) \to (\alpha, \beta)$, where f is an arrow from a to α in C and g is an arrow from b to β in D.

There are various ways of extending categories, in an analogous manner to considering the Set of Sets, or function spaces. One such example is the idea of an *arrow category*.

Consider again the category Set. We construct the associated arrow category $Set^{\to}$ as having as objects set maps $f : A \to B$, and having as arrows maps (γ, δ) between two objects $f : A \to B$ and $g : C \to D$, where $\gamma : A \to C$, $\delta : B \to D$, such that the following diagram commutes:

$$\begin{array}{ccc} A & \xrightarrow{\gamma} & C \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ B & \xrightarrow[\delta]{} & D \end{array}$$

It turns out that these arrows satisfy the required properties to make $Set^{\rightarrow}$ a category.

We also have an analogous idea to the construction of a power set. A functor F is an arrow between categories, such that it maps identity arrows to identity arrows, and such that $F(g \circ f) = F(g) \circ F(f)$, for arrows f, g in the original category. So, given a particular category C, we can construct the associated *power category* as being the subcategory in the category of categories that consists of all functors from C to categories D, and from these categories D to categories E, etc.

2.2 The Incompleteness Theorem

One of the most important results in the foundations of mathematics discovered in the early years of the 20th century was Gödel's Incompleteness Theorem. This essentially states that in any mathematical system of finite complexity subject to Peano arithmetic - that is, a structure with some form of multiplication and addition defined - it is possible to make statements that cannot be proven within the confines of the system. That is, any such construction is *incomplete*.

Corollary to this observation is that it is impossible to axiomatise mathematics using a finite number of axioms. This brings into new light some of the paradoxes plaguing the early attempts to formalise set theory.

In this section I will provide a sketch of the argument, following Gödel's wonderfully intuitive and coherent description [Gö].

The general intuition is as follows. Consider an axiomatic mathematical system, such as Zermelo-Frankel Set theory, or Principia Mathematica (PM). Within such a system one has certain symbols, such as variables, logical constants, and parentheses, which can be used to construct relations or formulae. We will also have within such a system that there are certain rules as to which formulae, or sentences of symbols are *correct* or allowable, and which are not. Then a *proof* is a sequence of such sentences.

It makes no difference if we write all variables and symbols as natural numbers since there are finitely many of them. So a sentence is clearly equivalent to a finite sequence of natural numbers, and a proof to a finite sequence of finite sequences of natural numbers.

Via this observation, it is then possible to prove, using the axioms of PM, that one can come up with a meta-mathematical statement $F(v)$ with one free variable v of type a sequence of natural numbers, such that $F(v)$ is a statement saying that v is a provable formula, ie there exists a proof for v in PM. (This is naturally obtained via consideration of a sequence of sequences of sequences of natural numbers, F_{ijk}).

The idea is to now construct a theorem A such that neither A nor its converse $\neg A$ is provable. Once this is demonstrated the proof of incompleteness of the axiomatic system PM (and any similar system) will be complete.

Now, a meta-mathematical statement with one free variable of type the natural numbers (ie, like F above, but only one natural number encoding v rather than a sequence) will be called a *class-sign*. It is possible to define a numbering of these by the natural numbers, as $N^{card(N)}$ is countable, call the nth one R_n. R is then a sentence of natural numbers defined within PM.

If α is an arbitrary class-sign, define $\alpha(n)$ to be the formula one obtains when one evaluates α at n. (Note, again, that n might be some abstract entity within the axiomatic system we are considering).

Construct then the class

$$K = \{n \in N | \neg provable(R_n(n))\}$$

by which we mean, the set of naturals n such that the theorem $R_n(n)$ is not provable. One might make the observation that this is beginning to look a bit both like Cantor's famous diagonalisation argument and Russell's paradox. But then K is expressible itself within PM, and so there is a class sign S such that $S(n)$ states that $n \in K$.

But since S is a class sign it is equivalent with R_q, for some $q \in N$. To complete our proof of the incompleteness of PM (or any similar logical system with finitely many symbols and axioms) we prove now that $R_q(q)$ is undecidable in PM.

But this is easy, since $R_q(q) \Leftrightarrow S(q) \Leftrightarrow q \in K \Leftrightarrow \neg provable R_q(q)$. In other words, the theorem $R_q(q)$ states that it is unprovable - a logical contradiction.

2.3 Key Constructions

2.3.1 Monic, Epic and Iso arrows

Roughly speaking, monic and epic arrows are generalisations of the notion of injective and surjective maps, respectively. Iso arrows are a generalisation of the concept of bijective maps.

It is possible to represent this information diagrammatically:

$$\begin{array}{ccc} C & \xrightarrow{g} & A \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle f} \\ A & \xrightarrow[f]{} & B \end{array}$$

We say that an arrow $f : A \to B$ is *monic* if, under the circumstance that the above diagram commutes, we have the implication $g = h$. This is an abstraction of the concept of injectivity; note that if $f \circ h = f \circ g$, and f is an injective map, then since the kernel of f is empty we must have that the restrictions of h and g to the domain of f are identical.

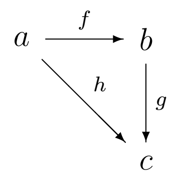
Now consider the diagram

$$\begin{array}{ccc} a & \xrightarrow{f} & b \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ b & \xrightarrow[h]{} & c \end{array}$$

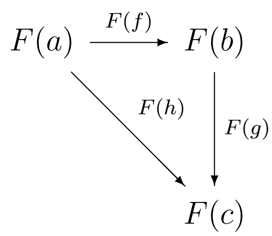
We say that an arrow $f : a \to b$ is *epic* if, under the circumstance that the above diagram commutes, we have the implication that $g = h$. Just like the above, this is an abstraction of the concept of surjectivity; if $g \circ f = h \circ f$, and f is surjective, then for each y in $dom(g) = b$ there is an x in dom(f) such that $f(x) = y$. But $dom(h) = b$ too. Consequently the expression $g \circ f = h \circ f$ is equivalent to $g = h$.

If an arrow is both monic and epic we will call it an *iso* arrow. This corresponds to the notion of bijectivity.

More generally, a *functor* is a map $F : \mathcal{A} \to \mathcal{B}$ between categories that associates to each arrow $f : a \to b$ in $\mathcal{A}$ an arrow $F(f) : F(a) \to F(b)$ in $\mathcal{B}$, such that the identity arrow in $\mathcal{A}$ is mapped to the identity arrow in $\mathcal{B}$, and $F(g \circ f) = F(g) \circ F(f)$. In other words we have that if the diagram



commutes, then



also commutes.

A functor is essentially an arrow between categories. In particular, we can define the notion of the *Power Category* associated to a base category $\mathcal{A}$ to essentially be the maximal category that contains all possible functors F mapping from $\mathcal{A}$ to some other category $\mathcal{B}$. Consequently the notion of monic, epic, and iso functors are well defined.

The Galois correspondence is actually a particular example of an *iso* arrow between categories.

2.3.2 Products and Coproducts

It is of some interest to ask how to characterise the product of a pair of categories. For sets, the notation is clear: we write

$$A \times B := \{(a, b) | a \in A, b \in B\}$$

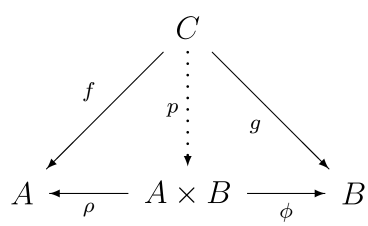
We also have two natural projection maps,

$$\rho : A \times B \to A, (a, b) \mapsto a$$

and

$$\phi : A \times B \to B, (a, b) \mapsto b$$

This allows us now to define the notion of the product more abstractly. Consider a third set C, with an associated pair of maps $f : C \to A$, $g : C \to B$. Define a map $p : C \to (A \times B)$ as $p(x) = (f(x), g(x))$. Then it is clear that the following diagram trivially commutes:



since $\phi \circ p = g$ and $\rho \circ p = f$ by definition. Conversely, p is the only arrow that can make the diagram commute. This leads us to define the notion of a *product* in a category.

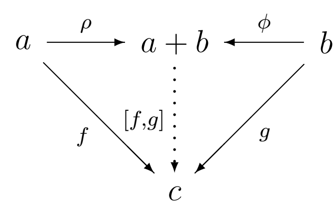
Definition 31. A *product* $A \times B$ of two objects A, B in a category $\mathcal{C}$ consists of projection arrows $\rho : A \times B \to A$, $\phi : A \times B \to B$ such that for any object C with arrows $f : C \to A$ and $g : C \to B$ there is a unique map $p : C \to A \times B$ such that the above diagram commutes.

The notion of *product category* $A \times B$ is very similar; it is constructed of all products of objects $a \in A$, $b \in B$, with the natural choice of projection functors - $\rho : A \times B \to A$, $(a, b) \mapsto a$, and $\phi : A \times B \to B$, $(a, b) \mapsto b$.

Before discussing the notion of *coproduct* we will need the idea of categorial duality. The dual category $\mathcal{C}^{op}$ to a category $\mathcal{C}$ is the category with the same objects but arrows reversed. So if $f : a \to b$ in $\mathcal{C}$, then there is a corresponding $f^{op} : b \to a$ in $\mathcal{C}^{op}$.

The advantage of this idea is that for any statement p which we prove in $\mathcal{C}$ there will be a corresponding statement p^{op} with arrows reversed in $\mathcal{C}^{op}$. So we immediately see that the following construction is well defined, from our previous considerations:

Definition 32. A *coproduct* or *sum* of objects a, b in a category $\mathcal{C}$ is an object $a+b$, together with "inclusion" maps $\rho : a \to a + b$, $\phi : b \to a + b$, such that for any pair of arrows $f : a \to c$ and $g : b \to c$, the following diagram commutes:



This of course is an extension of the set theoretic idea of disjoint union.

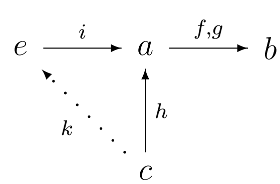
2.3.3 Limits and Colimits

Definition 33. Suppose $f, g : A \to B$ are maps, such that for $E \subset A$, f and g coincide. Then the inclusion map $i : E \to A$ is called the *equalizer* of f and g. In particular $f \circ i = g \circ i$.

Furthermore, if $h : C \to A$ is a map with $f \circ h = g \circ h$, it uniquely factors through i, ie there is a map $k : C \to E$ with $i \circ k = h$. To see this, note that since i is the inclusion we can trivially define $k = h$. Furthermore, this is well defined, since as $f \circ h(x) = g \circ h(x)$, $h(C) \subset E$.

Consequently we have the corresponding idea of an equalizer within a category:

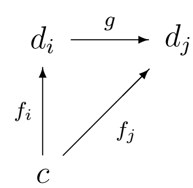
Definition 34. An arrow $i : e \to a$ is an equalizer for $f, g : a \to b$ if $f \circ i = g \circ i$, and if for any $h : c \to a$, there is a unique $k : c \to e$ such that $i \circ k = h$. In particular the following diagram commutes:



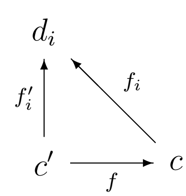
Note that all equalizers will be *monic*. This has the useful corollary that an epic equalizer will always be iso.

It is of some interest to ask as to whether we can abstract the idea of an equalizer. This is often of interest if we are trying to compare logical statements that require several objects and arrows to formulate.

Definition 35. A *diagram* D in a category $\mathcal{C}$ is a collection of objects d_i, together with some arrows $g : d_i \to d_j$. A *cone* for D is an object c together with maps f_i such that the following diagram commutes for every arrow g in D:

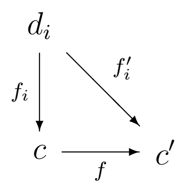


Definition 36. A *limit* for a diagram D is a D-cone $\{f_i, c\}$ such that if $\{f_i', c'\}$ is another D-cone, there is a unique arrow $f : c' \to c$ such that the following diagram commutes:



In other words, a limit is a generalisation of the concept of equalizer, from the consideration of arrows $f, g : a \to b$ to a diagram D of arrows. If a limit exists for a diagram D, it is unique.

Via duality we have the natural idea of a *co-limit*. This is a *co-cone* $\{f_i : d_i \to c\}$ from elements d_i in a diagram D, such that for every other *co-cone* $\{f_i' : d_i \to c'\}$, we have that there is exactly one arrow $f : c \to c'$ such that the following diagram commutes:

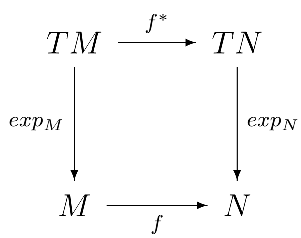


Remark. If D is an arrowless diagram consisting of objects a and b, then a limit is a product of a and b, and a colimit is a coproduct of a and b.

2.3.4 Pullbacks and Pushouts

It is useful to abstract the idea of a pullback map to category theory.

A pullback map is defined in the following way. Say we have maps $exp_M : TM \to M$, and $f : M \to N$. Then the pullback of f by exp_M, $exp_M^* : TM \to N$ is defined to be $f \circ exp$. We can furthermore lift f to a map $f^* : TM \to TN$ as the pullback of exp_N by f. Then we have a commutative diagram

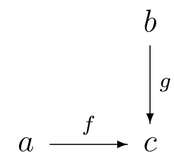


In particular we have that for arbitrary vectors v, w, that $\langle v, w \rangle_M = \langle f^* v, f^* w \rangle_N$. By duality we have a pushforward map $f_* : TN \to TM$, such that $\langle f_* v, f_* w \rangle_M = \langle v, w \rangle_N$.

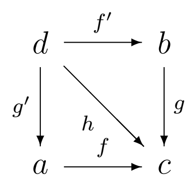
To abstract this, consider the following in some category $\mathcal{C}$:

$$a \xrightarrow{f} c \xleftarrow{g} b$$

Then a *pullback* is a limit in $\mathcal{C}$ for the above diagram, which I rewrite for clarity:

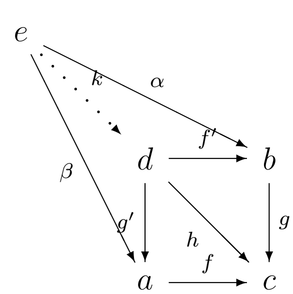


To see why this is similar to the idea of pullback from differential geometry, note that a cone on the above takes the following form:



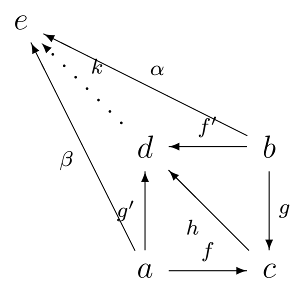
and we see immediately the similarity to the above. In particular we can remove the diagonal arrow WLOG as we are adjoining a cone.

Then a pullback in a category is a universal cone, in that if we have another cone on the pair $f : a \to c, g : b \to c$, then that cone factors through the limit. In particular, if our limit is the pair $g' : d \to a$, $f' : d \to b$ then there is exactly one arrow $k : e \to d$ such that



commutes, for a cone $\alpha : e \to b, \beta : e \to a$.

A pushforward, or pushout, is defined in an analogous way using co-cones, ie with arrows reversed, save for the original arrows $f : a \to c$ and $g : b \to c$. In particular we have that a pushforward is a colimit, such that there exists exactly one arrow $k : d \to e$ that makes the following diagram commute:



2.3.5 Completeness

Returning to our example of the category of Riemannian manifolds Mfd, I ask what it means to be complete. It is certainly clear that there is a limit for every diagram as above, via the use of exponential maps. So for any arrow $f : M \to N$ between Riemannian manifolds there is a pushforward $f^* : TM \to TN$. In this sense the category of Riemannian manifolds is complete.

(Note that TM, TN are themselves Riemannian, albeit with trivial metrics).

More generally, we say that a category $\mathcal{C}$ is *complete*, if given any arbitrary diagram D in $\mathcal{C}$ we have that a limit exists for D. A slightly less strong statement is that a limit exists for all finite D; in this case $\mathcal{C}$ is said to be *finitely complete*.

Now, we expect Mfd to be finitely complete, via virtue of the fact that exponential maps factor well through composition of arrows, so it is possible to extend the lifts f^*, g^* of maps $f : M \to N$, $g : N \to P$ to $(g \circ f)^* = g^* \circ f^*$.

There in fact is a theorem that states this is more generally true, which is mentioned in Goldblatt, but not proved. A proof can be found in [HS].

Theorem 2.3.1. *Suppose a category $\mathcal{C}$ has a terminal object, that is, an object from which there extend no arrows. If there is a pullback for each pair of arrows $f : a \to c$, $g : b \to c$, then $\mathcal{C}$ is finitely complete.*

In particular Mfd has as terminal object ϕ, the empty set, so the theorem applies in this instance.

2.3.6 Exponentiation

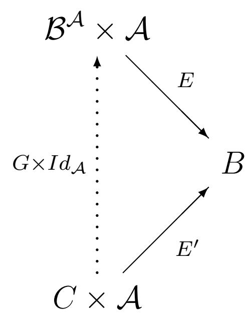
The idea of exponentiation is to consider two sets, $\mathcal{A}$ and $\mathcal{B}$, and then construct the function space $\mathcal{B}^{\mathcal{A}}$ consisting of maps from $\mathcal{A}$ to $\mathcal{B}$.

To abstract this to a category, note that there is an evaluation map

$$E : \mathcal{B}^{\mathcal{A}} \times \mathcal{A} \rightarrow \mathcal{B}$$

such that $E(f, x) := f(x)$.

It turns out in fact that among such maps, E is universal, in that if $E' : C \times \mathcal{A} \rightarrow \mathcal{B}$ is another map, then we have that there is a unique map $G : C \rightarrow \mathcal{B}^{\mathcal{A}}$ such that the following diagram commutes:



This universal property allows us to define exponentiation within a category.

We say that a category $\mathcal{C}$ has exponentiation if

(i) for any two objects A,B in $\mathcal{C}$ we can construct a product object $A \times B$.

(ii) for any two objects A,B there is an object B^A in $\mathcal{C}$ and an evaluation arrow $ev : B^A \times A \rightarrow B$, such that for any other arrow $g : C \times A \rightarrow B$ we have that ev factors through g as in the diagram above.

2.3.7 Topoi

I will now proceed to give a very abbreviated picture of what a Topos is. Broadly speaking, it is a space with

(i) An initial object,

(ii) A terminal object,

(iii) "Subobjects" that behave in a nice way,

(iv) It is finitely complete, that is, it has pullbacks.

A particular instance of the above - a very special case - is the category Mfd of differentiable manifolds and smooth maps between them. Then this has as initial object R^{∞}, terminal object ϕ, pullbacks via exponential maps, and subobjects defined in a natural way via the manifold topology.

However, Topoi are evidently more general than this. They also include more broadly categories such as "the category of sheaves of sets on a topological space" [W1]. In particular Topoi were developed as part of Grothendieck's work in Scheme theory. To be precise, Topoi can be thought of as a category equipped with a special topology, called the *Grothendieck Topology* that makes the objects behave like open sets in a standard topological space [W2].

I provide some additional background.

Definition 37. (Sheaf). A *sheaf* S associated to a topological space X is a structure that associates to every open set U in X an object $F(U)$ in a category $\mathcal{C}$. If $V \subset U$, then furthermore we have a morphism, or arrow, in $\mathcal{C}$ $res_{V,U} : F(U) \to F(V)$. We require that $res_{U,U}$ be the identity morphism on $F(U)$, and also that $res_{W,V} \circ res_{V,U} = res_{W,U}$, ie the morphisms respect the transitivity of set inclusion.

We also require that $\mathcal{C}$ be a concrete category. In other words, that it have a terminal object, such that $F(\phi)$ is this object. For every object in a concrete category is associated an underlying set - we then require that if $\{U_i\}$ is an open cover of U in X that if $s, t \in F(U)$ are such that $s_{U_i} = t_{U_i}$ for each U_i, then $s = t$. So S behaves well with respect to open covers.

Finally we would like the following condition for our sheaf. Let $\{U_i\}$ be an open cover once again of a set U in X. Suppose that we have a subset s_i of the set associated to $F(U_i)$ such that for every U_j with $U_i \cap U_j \neq \phi$ we have that $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$. Then there exists a subset s of the set associated to $F(U)$ such that $s_{U_i} = s_i$ for each i. In other words the gluing property in X lifts to the category $\mathcal{C}$.

The motivation then for a Topoi is to have a sheaf where the underlying space X is itself a category, but equipped with a special topology - the *Grothendieck Topology*, so that open sets and covers are well defined and well behaved. The methodology is roughly to construct a well behaved "pullback" from a standard topological space to a category. I will describe this now.

Definition 38. (Sieve). A *sieve* on an object A in a category $\mathcal{C}$ is a subfunctor of $Hom(\star, A) : Set \to \{$ morphisms from X to $\mathcal{C}|X \in Set\}$. If T is a sieve on $\mathcal{C}$, and $f : \mathcal{B} \to \mathcal{C}$ is a morphism, then we have a natural pullback f^*T over $\mathcal{B}$, such that for every object $b \in \mathcal{A}$, $f^*T(b) = \{g : b \to \mathcal{B}|fg \in T(b)\}$.

Definition 39. (Covering Sieve). Let U be an open set in a topological space X. Suppose $\{U_i\}$ is any open cover of U. Then for S to be a covering sieve of X we require that each $S(U_i)$ be nonempty, and that $S(U) \subset \cap_i(S(U_i \cap U))$.

Definition 40. (Grothendieck Topology). A Grothendieck Topology J on a category $\mathcal{C}$ satisfies these properties:

(i) Suppose S is a covering sieve on X, and $f : Y \to X$ a morphism. Then f^*S is a covering sieve on Y.

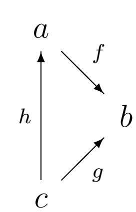
(ii) Let S be a covering sieve on X, and T is another sieve on X. Suppose that every object Y of the category $\mathcal{C}$, $f : Y \to X$ in $S(Y)$ induces a covering sieve f^*T on Y. Then T is a covering sieve on X.

(iii) $Hom(\star, X)$ is a covering sieve for X for every $X \in \mathcal{C}$

Definition 41. A *site* is a category $\mathcal{C}$ equipped with a Grothendieck Topology $\mathcal{J}$. Then, roughly it is possible to abstract the definition of a sheaf on a topological space to a sheaf on a site [W2]. The category of such sheaves for a given site is then a particular instance of a *topos*.

The treatment due to [RG] is not quite as strong as this, but might be perhaps easier to understand. In particular the notion of extending subsets to subobjects in a category turns out to characterise within a Topos, as requiring that a Topos be a category with a *subobject classifier*. This is in fact a consequence of Grothendieck's treatment.

Definition 42. (Subobject). A *subobject* of an object b in a category $\mathcal{C}$ is a monic $\mathcal{C}$-arrow $f : a \to b$. We will say that $f \subset g$ for subobjects $f : a \to b$, $g : c \to b$ if there exists $h : c \to a$ such that the following diagram commutes:

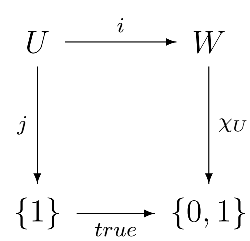


It turns out that this notion of inclusion is well defined - $f \subset f$ since $1_a : a \to a$ exists, and if $f \subset g$ and $g \subset h$ for subobjects f, g, h it is easy to demonstrate that $f \subset h$.

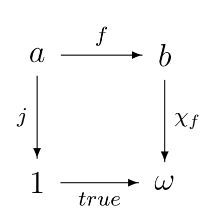
If $f \subset g$ and $g \subset f$ we will say that $f \equiv g$. The equivalence class of such subobjects will be notated as $[f]$.

This then forms a poset, $(Sub(b), \subset)$ of subobject equivalence classes of the object b in $\mathcal{C}$, ordered by inclusion. In particular, in *Set*, for any set B, $Sub(B)$ is the associated power set - the collection of all subsets of B.

Definition 43. (Subobject Classifier). A *subobject classifier* essentially captures the idea of either being in, or not in, a particular subset U of a larger set W. In *Set*, there is a natural characteristic map, $\chi_U : W \to \{0, 1\}$ such that $\chi(x) = 1$ if $x \in U$, and $\chi(x) = 0$ if $x \in W - U$. In particular, we have that the following diagram commutes:



where *true* is a map that sends 1 to 1, i is the set inclusion, and j is trivially defined. We then abstract this to a category in the following manner. Let $f : a \to b$ be a subobject in a category $\mathcal{C}$. Then a subobject classifier for f is an object ω in $\mathcal{C}$ and an arrow $true : 1 \to \omega$ such that there exists a unique arrow χ_f that makes the following diagram commute:



where j is the trivial map.

In other words, the existence of a subobject classifier is an abstraction of the idea of set complement to subobjects in categories.

Much of this might seem slightly unclear in motivation and necessity. However the picture to bear in mind is that Topoi are particular types of categories that admit an abstraction of the concept of topological structure. So in particular, these are the categories that are most likely to actually describe the underlying characteristics of mathematical structures modelling features that are observed in real situations.

In particular, this information will be useful later when I examine Grothendieck's proof of the Weil Conjectures. The associated mathematics actually turns out to be related to the abstract study of the theory of viscoplasticity, and it is possible that these ideas are more broadly applicable with respect to foundational studies of the other themes later developed in this treatise.

2.4 Introducing 2-Categories

This section is more of a remark than anything else. I motivate the development of a new type of object, a 2-category. This is not really a new concept, but corresponds to a rather large subfield of mathematics, and goes well beyond the scope of this project into territory that I will not require for the later developments in this dissertation. It does, however, relate possibly to the foundations of more general treatments. There are numerous sources available for the interested reader - the recent preprint by Daniel Mathews [Ms] is a good place to start.

Much as in set theory, category theory admits nonsense constructions. Consider, for instance, the category of categories that are not elements of themselves, ie categories that do not arise as objects in themselves. Then this has the same difficulty as the set of sets that are not elements of themselves in set theory.

This necessitates the consideration of a slightly more abstract approach to foundations, in order to circumvent this issue by simply showing that it is not constructible in this more general setting. Hence we have the idea of a *2-category.* A 2-category is a collection of things, called 2-objects, and things, called 2-arrows f associated to the *ordered triple* (a, b, c) of three 2-objects. We require furthermore that the 2-arrows have the following properties:

(i) If $f : (a, b, c)$,$g : (c, \beta, \gamma)$ are 2-arrows, then we have composition operators $\circ_1$,

$\circ_2$ such that $g \circ_1 f : (c, a, \gamma)$ is well defined if $a = \beta$, and $g \circ_2 f : (c, \beta, b)$ is well defined if $b = \gamma$. We also require that these operators be *transitive*.

(ii) For every 2-object A there is an identity $I_{1,A} : (A, A, A)$ associated to $\circ_1$ and an identity $I_{2,A} : (A, A, A)$ associated to $\circ_2$, such that for any 2-arrow $f : (A, B, C)$, $I_{1,A} \circ_1 f = f = f \circ_2 I_{1,C}$, and $I_{2,A} \circ_2 f = f = f \circ_1 I_{2,C}$.

As to concrete examples of 2-categories, this is slightly more difficult. The quickest one that comes to mind is a function space of function spaces, such as for instance $AutAut(Mfd)$, the collection of maps from the space of automorphisms back to itself over the collection of differentiable manifolds. But this seems like a silly observation - surely this space can be realised, naively at least, as a category? But it is possible that to examine this structure in full generality, standard category theory might not be sufficient, and to properly examine the foundations of the additional structure, deeper abstraction might be required.

Alternatively, it is possible that concrete examples of 2-categories are wilder than this. However I am hard-pressed to think of any further obvious examples.

Remark. It turns out in fact that 2-arrows are a special instance of another type of structure, that is, a *braid*. In particular, the set of generators for the permutation group of interchange operations on an n-braid - a braid with n-objects, has the interpretation of composition operations for an n-category. Then we can think of an n-category as a *braid*, or more generally a *link*, but possibly with "loose ends", ie n-objects α such that there is no β such that the jth permutation operation $\circ_j$ from α to β, or alternatively vice-versa, is defined.

This has the consequence that one might expect many of the insights from knot theory to be useful in the further development of foundations for abstract mathematical structures.

Remark. 2-categories have actually already been studied under another name, namely that of *spin networks*. In particular twistor theory [Pen] was motivated by the application of spin networks to perhaps providing a pre-geometric formulation of the study of particular forms of Hilbert spaces - operator spaces associated to the study of observables in standard (Schrödinger) quantum mechanics, though of course this can be considerably abstracted. Consequently the study of twistors can be viewed as a special instance of 2-category theory, where the underlying spaces have analytic structure - to perhaps attempt to do for 2-categories what Topoi do for 1-categories.

2.5 Topics in information theory

Here I provide a few results crucial to the discussion to follow.

2.5.1 Cramer-Rao for various geometric structures

When examining various forms of geometric structure, we often find that the structure of a proof is repeated. Consequently, we are merely interested in the *category* of the object. For instance, if we are looking at a tensor structure of some rank, we are interested in information inequalities for rank k tensors. If we are looking at precursor structures arising from a pairing of geometries, we are interested in information inequalities for these.

Consequently, I make the following observations, that will allow us to skip these formalities in the future.

Proposition 7. *Any statistical superstructure built on top of any geometric structure satisfies the Cramer-Rao inequality.*

Proof. Arbitrary geometric structures have as fundamental structural grounding some n-tensor Σ, together with an affine connection ∇_Σ. One can build an information from this by consideration of the density

$$F(m,a)(\partial ln(F(m,a)))^{\otimes n}$$

where $F(m,a)$ is a natural signal function for the induced statistical superstructure on top of Σ.

Hence it is sufficient to demonstrate that this density satisfies the Cramer-Rao inequality, and we will be done.

With a simple adaptation of Lemma 7.3.12 of [Go], we have

Lemma 2.5.1. $\int_M \Sigma \cdot \{cov(\theta^{i_1} \circ u, \theta^{i_2} \circ u, ..., \theta^{i_n} \circ u) - g^{i_1 \cdot ... \cdot i_n}\} \geq 0$

where u is a weakly unbiased estimator, that is $E(\theta \circ u) = E(\theta \circ Id)$, for θ a coordinate chart on M, and $g^{i_1 \cdots i_n}$ is the information density $F(m,a)(\partial ln(F(m,a)))^{\otimes n}$.

Then it is clear to see that theorem 7.3.14 of [Go] is applicable, albeit subject to slight generalisation, and shows that, in particular, that

$$\int_M \int_A \|\partial ln(F(m,a))\|^n F(m,a) dadm \geq 0$$

which is what we wanted to establish.

□

Proposition 8. *Any duply nested pregeometric structure will have a natural information density that satisfies a Cramer-Rao inequality.*

Proof. A duply nested pregeometric structure will be of the general form $T(f;g)$ for some positive operator $T : \{M \times A \to R\} \times \{M \times A \to R\} \to R$, and signal functions f, g over a base geometric structure with n-tensor κ. By the previous proposition, we have that κ satisfies a Cramer-Rao inequality for an information density ρ.

Construct $T(f;\rho(g))(\partial ln(T(f;\rho(g))))^{\otimes n}$. Then I claim that this density will satisfy a Cramer-Rao inequality.

In particular, if $n = 2$, the proof provided in section 9.2.2 of [Go] is sufficient, with the operator ∂^* replaced by T. Otherwise, for the more general case, it is reasonable to suppose that the proof of the above proposition can be used to extend the relevant discourse. □

Remark. Note that for all the pregeometric structures of interest to myself in this dissertation, the base geometric structure will usually be a Riemann-Cartan manifold. However these statements also extend to tensors of arbitrary rank.

2.5.2 The Cencov Representation Theorem and Incompleteness

In consideration of the Fisher Information as the optimal functional for a Riemannian geometry in [Go], we are led naturally to ask why this is the case. Certainly it satisfies the Cramer-Rao inequality, so is always positive. However there is no obvious reason why it should be preferable to consider it rather than any other nonnegative functional on a Riemannian manifold.

The key observation to make here is that there is the issue of incompleteness, or rather, finite complexity, of any particular geometric construction that one must deal with. In particular in more exotic geometries simpler notions of information are no longer sufficient.

Hence it is logical to frame the question in the following manner-

Question. *(Optimality). Given that one is working within a particular structural category $\mathcal{C}$, how can one prove that the associated Fisher Information for that category is optimal with respect to all other measures over $\mathcal{C}$?*

The answer to this question lies in an extension of an observation due to the Russian mathematician Cencov [Ce], where he was able to establish the uniqueness of the Fisher Information metric in Geometric Statistics on a discrete sample space containing at least three points.

This observation can be extended to prove what we need:

Theorem 2.5.2. *(Cencov Representation Theorem). Given a structural category $\mathcal{C}$, the Fisher Information $\rho(\mathcal{C})$ is optimal over all other positive functionals in the associated functional space.*

Remark. In other words, if we place a *sharp* distribution function over the space of positive measures over $\mathcal{C}$, then a variational argument will demonstrate that the meta-functional is critical iff the distribution function selects $\rho(\mathcal{C})$ at each point $m \in \mathcal{C}$.

Proof. We wish to demonstrate that, for all nonnegative functionals, ρ over $\mathcal{C}$ there exists a signal function f such that

$$\int_{\mathcal{C}} \rho \geq \int_{\mathcal{C}} \rho_f(\mathcal{C})$$

Assume wlog that we are working in the category of n-tensors. Then since ρ is positive, there will be a function f and an n-tensor Λ such that $\|\partial_\Lambda ln(f)\|_\Lambda^n = \rho$, where ∂_Λ is required to ensure the vector is timelike and therefore positive.

Then it is clear that there is a deformation Σ of Λ, such that for $\hat{f} = \delta(\Sigma(m) - a)$ and extension of the category $\mathcal{C}$ to $\hat{\mathcal{C}} := \mathcal{C} \times \{\mathcal{C} \times \mathcal{C} \to R\}$, which are equivalent categories under the assumption that we take a trivial section, then we have that within the category $\hat{\mathcal{C}}$, ρ is

$$\delta(\Sigma(m) - a)\|\partial_\Sigma ln(\delta(\Sigma(m) - a))\|_\Sigma^n$$

but this is just the Fisher information density $\hat{\rho}_\Sigma$ for the n-tensor Σ within $\hat{\mathcal{C}}$. Consequently we have via category equivalence (subject to our assumption of taking sharp sections of $\hat{\mathcal{C}}$) that the above inequality is satisfied; in fact it is an *equality.*

The proof for 2-tuple structures is quite similar. □

Chapter 3

K-theory and the index theorems

I will now direct my attention to K-theory. K-theory could be viewed as a continuation of the spirit of the theory of schemes, which were used by Grothendieck to tackle the Weil Conjectures. So it is natural to examine these structures and how they interrelate.

Key contributors to this area include Alexander Grothendieck, Jean-Pierre Serre, Daniel Quillen, and Michael Atiyah.

3.1 Preliminaries from algebraic geometry

Before I begin to motivate (and describe) the structures leading to K-theory, it will be necessary to flesh out a few elementary concepts from Scheme Theory and Algebraic Geometry. My primary reference here is Hartshorne [Ha].

3.1.1 Sheaves and Schemes

Recall from before that a sheaf is essentially a way of associating to each set of a particular topological space a mathematical object in a category, such as an abelian group. In particular, to reiterate the previous definition -

Definition 44. (Sheaf). A *sheaf* S associated to a topological space X is a structure that associates to every open set U in X an object $F(U)$ in a category $\mathcal{C}$. If $V \subset U$, then furthermore we have a morphism, or arrow, in $\mathcal{C}$ $res_{V,U} : F(U) \to F(V)$. We require that $res_{U,U}$ be the identity morphism on $F(U)$, and also that $res_{W,V} \circ res_{V,U} = res_{W,U}$, ie the morphisms respect the transitivity of set inclusion.

We also require that $\mathcal{C}$ be a concrete category. In other words, that it have a terminal object, such that $F(\phi)$ is this object. For every object in a concrete category is associated an underlying set - we then require that if $\{U_i\}$ is an open cover of U in X that if $s, t \in F(U)$ are such that $s_{U_i} = t_{U_i}$ for each U_i, then $s = t$. So S behaves well with respect to open covers.

Finally we would like the following condition for our sheaf. Let $\{U_i\}$ be an open cover once again of a set U in X. Suppose that we have a subset s_i of the set associated to $F(U_i)$ such that for every U_j with $U_i \cap U_j \neq \phi$ we have that $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$. Then there exists a subset s of the set associated to $F(U)$ such that $s_{U_i} = s_i$ for each i. In other words the gluing property in X lifts to the category $\mathcal{C}$.

Recall the notion of a Grothendieck topology -

Definition 45. (Grothendieck Topology). A Grothendieck Topology J on a category $\mathcal{C}$ satisfies these properties:

(i) Suppose S is a covering sieve on X, and $f : Y \to X$ a morphism. Then f^*S is a covering sieve on Y.

(ii) Let S be a covering sieve on X, and T is another sieve on X. Suppose that every object Y of the category $\mathcal{C}$, $f : Y \to X$ in $S(Y)$ induces a covering sieve f^*T on Y. Then T is a covering sieve on X.

(iii) $Hom(\star, X)$ is a covering sieve for X for every $X \in \mathcal{C}$

And also the definition of a site -

Definition 46. A *site* is a category $\mathcal{C}$ equipped with a Grothendieck Topology $\mathcal{J}$. Then, roughly it is possible to abstract the definition of a sheaf on a topological space to a sheaf on a site [W2]. The category of such sheaves for a given site is then a particular instance of a *topos*.

In the special case that the objects in $\mathcal{C}$ are abelian groups, then the associated sheaf (X, θ_X) over a topological space X - presupposing the natural Grothendieck topology admitted by the group structure in $\mathcal{C}$ (the so-called *Zariski topology*) - forms what is referred to as a *ringed space*. If, for each $x \in X$, $\theta_{X,x}$ - the group associated to x, is a local ring (which I will soon describe), we say that (X, θ_X) is a *locally ringed space.* An *affine scheme* then is a locally ringed (X, θ_X) which is isomorphic to the spectrum of some ring R (which I will also soon describe). Then a *scheme* is a (X, θ_X) such that each x has a neighbourhood U such that the associated sheaf $\theta_{X|U}$ is an affine scheme (in much the same way one defines the difference between trivial fibre bundles and more general versions).

Definition 47. (Local Ring). A ring R is *local* if it has a unique maximal left (or right) ideal.

The idea of Spectrum is that it is the space of subobjects for a given structure where inverses do not exist, ie, in a sense, they are the "zeroes" - much like in the beginnings, we have that an affine variety is the zero locus of a finite set of polynomial equations.

Definition 48. (Spectrum of a matrix). The spectrum of a matrix is the set of its eigenvalues.

Definition 49. (Spectrum of an operator). Suppose T is a bounded linear operator over a field k, or more generally, a Banach space X. Then if $\lambda I - T$ is not invertible for some $\lambda \in k$, we say that λ is in the Spectrum of T. The Spectrum for a linear operator T is a subspace of X, and the space of bounded linear operators forms a category $B(X)$ with a natural (Grothendieck) topology over X.

Definition 50. (Spectrum of multiple operators). Suppose $T_1, ..., T_k$ are bounded linear operators over a Banach space X. Then if $\lambda I - T_i$ is not invertible for each i and some $\lambda \in k$, we say that λ is in the Spectrum of $\cap_i T_i$.

Definition 51. (Spectrum of a Ring). Suppose R a ring with 1 acts as an R-module on a topological space X. Then we can play a similar game to above and say that if $\lambda 1 - r$ does not have a multiplicative inverse in R for each $r \in R$, then λ is in the Spectrum of R relative to X, $Spec_X(R)$. Naturally we expect most interesting examples to occur in rings which "generically" have only a finite number of ideals, though it is possible that for sufficiently sophisticated X one might be able to obtain non-trivial Spectra if R has a countable or uncountable infinity of ideals. Alternatively one might have a ring R with infinitely many ideals, but where for each component of the Spectrum there are only finitely many ideals I in R that are applicable - a "locally finite" ring.

The definition that will be of interest to us however is the restriction of this to where $X = R$; ie where one considers R to be a module of itself. I will then call the corresponding Spectrum $Spec(R)$. Clearly $Spec(R)$ will be contained in R - and in particular will enumerate the *prime*, or *critical ideals* of R - in particular one can view $R \mapsto Spec(R)$ as an "information preserving map".

One might ask as to why we are restricting ourselves to categories where the objects are abelian groups. The principal reason for this is that in the study of affine varieties, there is a natural correspondence between these and particular types of rings (finitely generated, and nilpotent-free "nondegenerate" rings). Since the Weil conjectures deal principally with varieties, this indeed should be sufficient for our purposes.

I will sketch this correspondence shortly.

Definition 52. (Zero sets and irreducibility). Consider $A := k[x_1, ..., x_n]$ as the polynomial ring of n variable functions over the field k. Then consider a subset B of A, generated say by finitely many functions $f_1, ..., f_m$. Since this is generated by functions it is said to be *algebraic*. The *zero set* $Z(B)$ of B is then defined to be the set of elements x in k such that $f_i(x) = 0$ for each i. If it is not possible to decompose B' into smaller pieces it is said to be *irreducible*.

Definition 53. (Variety and Zariski topology). Now, suppose we have a set $B \subset A$ that is not necessarily algebraic. If it is closed with respect to the *Zariski topology* - a topology defined on A by taking the open sets to be the complements of algebraic sets, *and* it is irreducible, that is, $Z(B)$ is connected in k, we call B an *affine variety*. A *variety* is the same but with the condition of irreducibility relaxed.

Remark. It is possible to show that the Zariski topology is indeed a topology. It is also possible to demonstrate that any variety can be broken up into affine varieties, provided that A is *noetherian*. Noetherian essentially means that if $Y_1 \supset Y_2 \supset ...$ is any sequence of closed subsets of A, then there exists r such that $Y_r = Y_{r+1} = ...$ This is indeed true in the case of curves defined say from R^n to R; then as there is a finite number of dimensions, and dimension of a variety must either remain the same or decrease, and furthermore as an irreducible variety cannot be reduced any further at the same dimension, it is evident that such objects are Noetherian.

As an example, but perhaps not a terribly elementary one, one might consider for A the set of maps from a manifold M with coordinates x_i that map to the real numbers. Naturally we expect that this should extend; in particular if the manifold itself has a field structure, perhaps induced by a four-tensor, we can take as the field k the manifold itself again; in this case we have a form of variety in terms of the automorphisms of M.

3.1.2 The variety - ring correspondence and its consequences

Now, to establish the connection between varieties and rings, I introduce the following result due to Hilbert, which I give here without proof, and I mirror Hartshorne in referencing either the proof in Lang's book [Lg], p.256, or that by Atiyah and Macdonald [AM], p.85.

Theorem 3.1.1. *(Hilbert's Nullstellensatz). Suppose k is an algebraically closed field, and a is an ideal in A, the polynomial ring in n-variables over k. Suppose $f \in A$ is a polynomial that vanishes on $Z(a)$. Then, for some integer $r > 0$, we must have that $f^r \in a$.*

A consequence of this is the correspondence result we need:

Corollary 3.1.2. *(The variety - ring correspondence). There are iso arrows mapping between the category of algebraic sets in A and the category of radical ideals in A, given by $Y \mapsto I(Y)$ and $a \mapsto Z(a)$. Also algebraic sets are irreducible iff the corresponding ideal is a prime ideal.*

Recall that the radical $Rad(I)$ of an ideal I is the set of elements x of the ring R containing the ideal such that $x^n \in I$ for some finite n. A prime ideal I is an ideal that has the property, that if A, B are ideals also in R and $AB \subset I$, then $A \subset I$ or $B \subset I$. This is essentially an analogue of the idea of factorisation extended to general rings.

So the variety - ring correspondence motivates the consideration of sheaves of abelian groups (of which ideals are a special instance) over a topological space X, where the sheaf is now taking the role of the polynomial ring A and X is taking the role of the n-variable space over a field k. In particular for general topological spaces defining varieties may not be possible, when it may be possible to construct many different sheaves. Consequently this is a natural starting point for the development of the deeper theory, where now X instead of being k^n could be a manifold or some more general object, such as the space of automorphisms over a given space M.

Then a *scheme* in this light is merely an abstraction of the idea of a *variety*, albeit for these more general objects. This is perhaps the most instructive way of considering the idea of scheme, and indeed varieties form a subclass of the class of schemes. But there are stranger examples of schemes as well. It will turn out that a proper geometric description of a scheme requires a special form of four-tensor construction; this will be discussed in the chapter on viscoplasticity. In particular, it could be argued that schemes can be viewed as "fractal" algebraic varieties; in other words, the twisting of one algebraic variety by another, with regards to some natural algebro-geometric operator. (For a simple example of a twisting construction, one has the idea of a warped product of Riemannian manifolds (O'Neill [ON], pp204-207).)

To make this more concrete, consider a topological space M with local coordinates $x_1, ..., x_n$, and corresponding local polynomial ring over k $A := k[x_1, .., x_n]$. Suppose $f_1, ..., f_m$ are functions from A to k, and $g_1, ..., g_m$ likewise. Then the f_i define a variety $Z(f)$ over M, and the g_i a variety $Z(g)$. Consider the pseudometrics defined on $Z(f)$, $Z(g)$ as $\sigma_{ij} := \nabla_M f_i \cdot \nabla_M f_j$, and $\tau_{kl} := \nabla_M g_k \cdot \nabla_M g_l$. These will be pseudometrics since we do not have guarantee of non-degeneracy in the bilinear form (and generically we except a fair degree of degeneracy).

Construct the corresponding signal functions $f(m, a) = \delta(\sigma(m) - a)$, $g(m, b) = \delta(\tau(m) - b)$. Then we can construct a *scheme* that in general will not be a variety by considering instead the new signal function $\star(f; g) := fg$. The corresponding geometric object for this "viscoplastic structure" is then the *scheme* associated to the twisted product of the varieties $Z(f)$ and $Z(g)$.

If one abstracts further one has the idea of a *general sheaf* (as in the previous chapter), wherein instead of considering groups as objects one considers more arbitrary objects in categories, with arrows that respect inclusions in the topological space X. This becomes useful in the extension of etale cohomology to development of the Grothendieck-Galois correspondence for topoi.

3.1.3 Morphisms

Suppose we have two varieties. How does one define a map between them? More generally, how does one define a map between schemes? This is the question that I shall answer in this section.

Definition 54. (Regular map). Let Y be a variety and k a field. Then $f : Y \to k$ is *regular* at $m \in Y$ if there is a subspace U of Y with $m \in U$, and polynomials g, h over k, ie in $A = k[x_1, ..., x_n]$ such that h is nonzero on U and $f = g/h$. In other words, f is *non-singular* - it has no singularities.

Definition 55. (Morphism of varieties). Suppose k is an algebraically closed field, and X, Y are varieties over k. Then a *morphism* $f : X \to Y$ is a map such that for every $V \subset Y$, and every $g : V \to k$ that is *regular*, $g \circ f : f^{-1}(V) \to k$ is also regular.

So a morphism is essentially a map that is not subject to collapse between varieties; ie it is a "lossless" transformation for information between varieties.

It is then possible to define the category of varieties with morphisms as the arrows. This is in fact a particular instance of a *topos*, as described in the previous chapter.

Definition 56. (Morphism of schemes, concrete version). Suppose k is an algebraically closed field, and X,Y are schemes over k, corresponding to the twisted product of varieties X_1, X_2, Y_1, Y_2 respectively. Then a *morphism* $f : X \to Y$ is a map such that for every $V \subset (Y, \Lambda)$, where Λ is the degenerate 4-tensor associated to the pushforward of the pseudometrics for X_1, X_2 under f, and every regular $g : V \to k$, we have that $g \circ f : f^{-1}(V) \to k$ is regular; ie f is a lossless transformation of geometric information.

Then, using these morphisms as arrows, it is possible to construct a category of schemes.

3.2 Etale cohomology and the Weil Conjectures

Etale cohomology is a special case of K theory, and was historically developed first by Grothendieck, to provide him with the means to approach certain conjectures within number theory due to Andre Weil. Naively stated, the statement by Weil was:

Conjecture. *(Weil). Consider an algebraic variety, defined by equations with integral coefficients. Then there exists a cohomology theory for the same that gives information about these equations; ie given knowledge of the cohomology, it should be possible to deduce part, or possibly all of the information regarding these equations.*

Indeed, this new architecture provided the means to resolve most of this question posed by Weil, and laid the foundation for the eventual resolution of the most difficult of the Weil conjectures- the analogue to the Riemann Hypothesis- in 1973 by Pierre Deligne [De].

In this section I will paint a picture of how Etale cohomology is defined. However I will stop short of describing how it is used to attack the Weil Conjectures, since I will not need the full machinery associated to the Grothendieck program in order to proceed to the K theory. I will follow Artin's lecture notes [Art] in my approach.

3.2.1 Cech Cohomology for presheaves, or "pseudo-varieties"

Definition 57. (Limits). Consider a functor $F : \mathcal{I} \to \mathcal{C}$ between categories $\mathcal{I}$ and $\mathcal{C}$. There is a constant functor associated to each $X \in \mathcal{C}$ that sends objects in $\mathcal{I}$ to X, and arrows in $\mathcal{I}$ to id_X. Call this functor c_X.

Then there is a covariant functor $Hom_{fun}(F, c_X) : \mathcal{C} \to Sets$, which will be notated the *right limit* of F at X, $\lim_{\to X} F$. Similarly $Hom_{fun}(c_X, F)$ is the *left limit* of F at X, $\lim_{X\leftarrow} F$.

If $\mathcal{C}$ is the category $Sets$, or the category Ab with objects that are abelian groups, we have that $\lim_{\to} F$ is *representable* as a functor from $\mathcal{I}$ to Ab, ie "the limit exists in Ab". This will again be called $\lim_{\to} F$.

Now, for each category $\mathcal{C}$, the dual category $\mathcal{C}^*$ is defined as morphisms from $\mathcal{C}$ to the category of sets, $Hom(\mathcal{C}, Sets)$ - in an analogous manner to homology.

Remark. To see this, if one considers the Homology of a manifold, $H_i(M)$, this will be generated by particular submanifolds which are geodesic which respect to an appropriate 4-tensor structure. In particular, the j_i generators of the ith homology will be defined as the zero loci of i distance functions $f_{j1}, ..., f_{ji}$ from M to the reals.

In fact, generically we do not expect the zero loci to actually be submanifolds themselves, though this is required for the homology to be well behaved - this motivates the definition of the Cech cohomology theory for varieties and schemes, to extend to zero loci admitting singularities.

Then the cohomology $H^i(M) := Hom(H_i(M), R)$, maps from each of the generators to the real numbers. In particular these can be represented as differential forms, eg $dxdy + 2dzdt$.

To construct the Cech cohomology, we now consider the category $P := P_{\mathcal{C}}$ of functors from $\mathcal{C}^*$ to Ab. Consider now a family of maps $\{U_\alpha \to V\}_{\alpha\in I}$ in P. It is then possible to construct a cohomology theory, known as the *Cech cohomology*, by construction of functors $H^q(\{U_\alpha \to V\},)$ from P to Ab.

In particular, these are constructed as follows. Consider the sequence $V \leftarrow_0 \{U_\alpha\} \leftarrow_1 \{U_\alpha \times_V U_\beta\} \leftarrow_2 ....$

If F is a presheaf on P then we obtain maps $\Pi_\alpha F(U_\alpha) \to \Pi_{\alpha,\beta} F(U_\alpha \times_V U_\beta) \to ...$ etc.

Then given such an F we can define d_n as an inverse to $\leftarrow_{n+1}$, in the following manner:

$$d_n : \Pi^n_{\alpha_i, i=0} F(U_{\alpha_0} \times_V ... \times_V U_{\alpha_n}) \to \Pi^{n+1}_{\alpha_i, i=0} F(U_{\alpha_0} \times_V ... \times_V U_{\alpha_{n+1}})$$

by $d_n = \Sigma^{n+1}_{i=0} (-1)^i F(U_{\alpha_0} \times_V ... \times \hat{U}_{\alpha_i} \times_V ... \times_V U_{\alpha_n})$.

In particular it can be checked that $d_{n+1} \circ d_n = 0$, and since P was in fact defined to be the category of presheaves on $\mathcal{C}$ this gives us a functor from $P \to \{$ cochain complexes $\}$ which is clearly exact. Consequently we obtain a well defined cohomology theory, in particular an exact sequence of cohomological functors $P \to Ab$, notated $H^q(\{U_\alpha \to V\},)$ as above.

Essentially, morally speaking, we now have a cohomology theory which provides us with deeper information regarding the structure of a manifold, or more generally of a variety, than the standard de-Rham cohomology theory provides. It allows us to consider generators in the corresponding homology which have a nontrivial zero locus; that is, a zero locus that admits collapse, or singularities.

Cohomology is important because it allows us to extract information regarding a category of spaces of interest. This is in particular related to functional analysis, Information Theory - to be precise, the Fisher Information corresponding to particular structures, and the study of Characteristic Classes.

3.2.2 Derivated Functors

I now digress and provide a similar treatment of the above, this time following Hartshorne, in this case the third chapter.

Definition 58. (Abelian category). A category $\mathcal{C}$ is abelian, if, for any two objects A, B within $\mathcal{C}$, $Hom(A, B)$ is an abelian group.

So for instance the category of abelian groups is abelian.

Definition 59. (Complex). A *complex* A in an abelian category $\mathcal{C}$ is a collection of objects A^i indexed by the naturals, and morphisms $d_i : A^i \to A^{i+1}$ such that $d_{i+1} \circ d_i = 0$ for each i. A *morphism* of complexes $f : A \to B$ is a set of morphisms $f^i : A^i \to B^i$ that commute with the *coboundary maps* d_i.

We define a cohomology theory then in the usual way, with $H^i(A)$, the ith cohomology object defined as the kernal of d_i mod the image of d_{i-1}. There is also a notion of homotopy, that is, deforming one morphism of complexes into another. If $f, g : A \to B$ are morphisms of complexes, then there are *homotopic* if there are maps $q^i : A^i \to B^{i-1}$ such that $f - g = d \circ k + k \circ d$. In particular a morphism

$f : A \to B$ induces a sequence of cohomology invariants in the category $\mathcal{D}$ containing B, and if f is homotopic to g these will be the same.

3.2.3 Cohomology for Sheaves and Schemes; The Leray Spectral Sequence

Definition 60. An *injective resolution* of an object A in a category $\mathcal{C}$ is a complex I (ie, a sequence $I_0 := \{U_\alpha\}$, $I_1 := \{U_\alpha, U_\beta\}$, ...) together with a morphism $\epsilon : A \to I_0$, such that $0 \to A \to_\epsilon I_0 \to I_1 \to ...$ is *exact*.

(This should look quite familiar to the treatment in Artin [Art].)

In order to build a cohomology theory for schemes, we will need to define invariants $E^{p,q}$, indexed by the square of the naturals. This fits into the intuition that schemes have more information than varieties, since they are 4-tensor pseudostructures, rather than 2-tensor pseudostructures.

Definition 61. (δ-functor). Suppose now $\mathcal{A}$,$\mathcal{B}$ are abelian categories. Then a δ-functor from $\mathcal{A}$ to $\mathcal{B}$ is a collection $T = (T^i)_{i \in N}$, such that there are morphisms $\delta_i : T^i(a') \to T^{i+1}(a)$ for any short exact sequence $0 \to a \to A \to a' \to 0$, such that

(i) The short exact sequence extends to a long exact sequence

$$0 \to T^0(a) \to T^0(A) \to T^0(a') \to_{\delta_0} T^1(a) \to ...$$

and

(ii) For each morphism $f : A \to B$, and for a corresponding short exact sequence $0 \to b \to B \to b' \to 0$, we have that δ commutes with f, in particular

$$\begin{array}{ccc} T^i(a') & \xrightarrow{\delta_i} & T^{i+1}(a) \\ \downarrow f & & \downarrow f \\ T^i(b') & \xrightarrow[\delta_i]{} & T^i(b) \end{array}$$

Definition 62. (Etale cohomology).

Now to define our cohomology theory for Schemes, note as before that a Scheme can be viewed locally as the twisted tuple of two varieties, A, B. (To emphasise, recall that there is a pseudometric σ for A, and τ for B; then the corresponding Scheme is the geometric object corresponding to $\sigma_{ij}\tau_{kl}$.) If we then represent A, B as *complexes* in the above sense (a generalisation of the representation of manifolds as simplicial complexes for homology), we have associated subobjects associated to the subvarieties of A, B - A_i, B_i. We then have morphisms induced by the twisting $f^i : A_i \to B_i$ between these subobjects.

In particular we have the existence of a map from subvarieties A_i to groups $T^i(A)$, and functors δ_i - much as for de-Rham homology - such that $\{T^i, \delta_i\}$ has the properties of the definition above. To be more specific, we have induced maps $\hat{f} : T^i(A) \to T^i(B)$.

Then we have that $\delta_{i+1} \circ \delta_i$ for the Scheme $(A, B)_f$ is zero for each i. This allows us to define cohomology invariants $E^{p,q}$, as $f(H^p(A); H^q(B))$; the relative δ-homology of A to the δ-homology of B rel the scheme product map f on A_σ, B_τ induced by the scheme operator $\star$, where $\star(\sigma_{ij}; \tau_{kl}) := \sigma_{ij}\tau_{kl}$. Note that we do not expect f to act in the same way as $\wedge$.

This is nothing other than the *Etale cohomology* of Grothendieck.

To be more concrete, suppose dy is a $n-p$ pseudoform on A and dz is a $m-q$ pseudoform on B. Then in certain situtations $\int_{A \times_{\hat{\wedge}} B} dydz$ defines a tuple of a p dimensional subvariety of A, and a q dimensional subvariety of B, twisted by the induced scheme map $\hat{\wedge}$. This correspondence, when it exists, actually turns out to be crucial in examination of the Weil conjectures.

Definition 63. (Etale cohomology, 2nd version). Consider algebraic varieties A, B, with Cech cohomology groups $H^p(A) = \langle G_p(A)|R_p(A)\rangle$, $H^q(B) = \langle G_q(B)|R_q(B)\rangle$ where $G_p(A), G_q(B)$ are the generators and $R_p(A), R_q(B)$ are the relations of the respective groups. Then the Etale cohomology of the *scheme* $A \star B$ induced by the twisting of A and B, is the free product of these groups; or $E^{pq} := \langle G_p(A), G_q(B)|R_p(A), R_q(B)\rangle = H^p(A) \star H^q(B)$.

However, we wish to consider more general situations, rather than just schemes induced by the twisting of two varieties, in much the same way a 4-tensor can be more general than the product of two 2-tensors.

Definition 64. (Etale cohomology for general schemes). Let M be a manifold, with a (possibly degenerate) four tensor Λ_{ijkl}, ie, a *scheme*. Then a (p, q) evaluation of Λ is a choice of bases (V, W), with V of rank p and W of rank q. We then have that $d_{\Lambda_{ijkl}}(V_{ij}, W_{kl})$ defines a degenerate volume form on M, where d_Λ is the induced $\star$-co-connection associated to Λ (see Chapter 6). The set of such forms up to homotopy defines the group $E^{pq}(\Lambda)$.

I will discuss these situations in more detail shortly. However first it is necessary to motivate the necessary developments by an examination once again of more elementary structures.

3.2.4 Serre Duality

Serre duality is an extension of Poincare duality but for Cech cohomology. In particular, there is a homology theory for so called projective schemes. This is particular is useful in proving certain results, such as the Riemann Roch theorem. I will provide enough information here to understand the general idea and intuition underlying these results.

Recall the statement of Poincare duality:

Theorem 3.2.1. *(Poincare-deRham Duality). If M is an oriented closed n-manifold, then for each k we have an isomorphism of homology and cohomology groups, given by $H^k(M) \cong H_{n-k}(M)$.*

Intuitively this should be clear, since there is a clear correspondence between k forms and $n-k$ dimensional subspaces of M, given by

$$f(x_{k+1}, ..., x_n) := \int_{x:=(x_1,...,x_n)\in M} dy_1(x)...dy_k(x)$$

ie the level sets of f define a fibering of M into $n-k$ dimensional spaces. What Poincare duality tells us is that if M is oriented and compact (ie without boundary), there will be a well defined representative - one of these level sets - that will be dual to the form $dy = dy_1...dy_k$.

The question posed by Cech-Serre is whether one can do this for varieties, and more generally, schemes. In other words, given a variety X, with a corresponding Cech cohomology, is it possible to define a homology theory (associated to the subvarieties of X), such that the kth elements of the Cech cohomology for X are associated with elements of the $n-k$th homology group. In other words, our "generalised" or rather degenerate k forms are to be associated with degenerate $n-k$ manifolds, or in other words algebraic varieties.

It turns out that it is possible to do this, but only for a limited class of varieties, known as *projective varieties*, that take the role of oriented closed manifolds in our theory.

Definition 65. (Projective Variety). A *projective variety* is an irreducible algebraic set V in projective n-space; that is, the set of points $(x_1, ..., x_n)$ equivalent up to scalar multiplication relative to an algebraically closed field, k.

Definition 66. (Projective Scheme). A *projective scheme* is the object associated to the product of the two pseudometrics corresponding to projective varieties within a polynomial ring $k[x_1, ..., x_n]$.

Consequently we expect this to have the property of being "compact", since the construction is independent of scalar multiplication by the field k. And it will be "oriented" since it is algebraic.

I now give a vague idea of the construction of the required homology theory.

Definition 67. (The Ext functor). Suppose (X, θ_X) is a *ringed space*. The prototyical example for instance, might be $Aut(X)$ - the ringed space that associates to each $U \subset X$ the group of automorphisms of U. Suppose now that $\mathcal{F}, \mathcal{G}$ are $Aut(X)$ modules. In other words, we might have that $\mathcal{F}$,$\mathcal{G}$ are superschemes of X; then $Aut(X)$ acts on these by extension of the local automorphism in X to $\mathcal{F}, \mathcal{G}$, in much the same way one extends a form $ds_1...ds_k$ on a submanifold $N^k \subset M^n$ to a new form $ds_1...ds_k dy_1...dy_{n-k}$ in M.

Such extensions in general will not be unique, and have the structure of a group.

Consequently we will have $Hom(\mathcal{F}, \mathcal{G})$ as the group of morphisms between superstructres induced by the automorphisms of X.

We can then build the *Ext functors* $Ext^i(\mathcal{F}, .)$ as the right derived functors of $Hom(\mathcal{F}, .)$. In other words, the functors dual to homomorphisms from the variety $\mathcal{F}$ to the underlying field k.

To build the Ext group associated to a particular embedding of X in a space $\mathcal{F}$, we need the idea of a dualizing sheaf.

Definition 68. (Dualizing sheaf). It is of interest to ask whether we can construct a dual structure ω_X to a superscheme $\mathcal{F}$ of a scheme X, in much the same way one can talk of the fiber of a bundle map $\mathcal{F} \to X$, or of a partition of a manifold M by a manifold N indexed by a dual submanifold K. Indeed this is sometimes possible. The map ω_X that associates to each superscheme $\mathcal{F}$ containing X a subscheme $\omega_X(\mathcal{F})$, if it exists, is referred to as the dualizing sheaf of X.

Now the importance of projective schemes suddenly becomes clear.

Lemma 3.2.2. *If X is a projective scheme, then it has a dualizing sheaf ω_X.*

There is one more idea that we will need before I can state the duality theorem.

Definition 69. (Cohen-Macaulay ring). The *Krull* dimension of a ring R is the length of a maximal chain of prime ideals contained in R. The *depth* of a ring is the maximal length of a *regular sequence* in R, that is a sequence $x_1, ..., x_k$ with the property that x_1 is not a zero divisor, and x_{i+1} is not a zero divisor of $R/(x_1, ..., x_i)R$. A local ring is *Cohen-Macaulay* if it is commutative, Noetherian, and has Krull dimension equal to its depth. A ring is Cohen-Macaulay if all its localisations at prime ideals are Cohen-Macaulay.

Essentially the geometric intuition here is that these rings are to correspond to nonsingular varieties - that is, varieties that are locally equidimensional. Certainly if the Krull dimension exceeded the depth we would have a situation where there were zero divisors in R, or "singular" elements. Here is perhaps the source of the difficulty with trying to prove the Weil conjectures purely using Serre Duality for algebraic varieties, as the correspondence turns out to require the Cohen-Macaulay property - ie it is only an isomorphism for nonsingular varieties - not for general algebraic structures that admit singular points. Consequently we need the idea of a *nonsingular*, or *Cohen-Macaulay scheme* - which actually will be generically a general algebraic variety - in order to get the resolution we require.

Definition 70. (Cohen-Macaulay scheme). A scheme is Cohen-Macaulay if all of its local rings are Cohen-Macaulay rings.

To codify my remark before, we have a correspondence principle that relates general algebraic varieties to Cohen-Macaulay schemes. In this viewpoint, to build an analogy with [Go], the latter takes the role of a *sharp* object, whereas the former a substructure with a general "statistical superstructure".

Lemma 3.2.3. *(Cohomology of Cohen-Macaulay schemes). There is a correspondence between Cohen-Macaulay schemes and general algebraic varieties. Consequently, the matrix of cohomology groups $E^{p,q} := f(H^p(X_\sigma); H^q(Y_\tau))$ is diagonal, ie these groups reduce to Cech cohomology groups $H^p(\hat{X})$, where $\hat{X}$ has the structure of a general algebraic variety. Also, all subschemes of a Cohen-Macaulay scheme are subvarieties of the associated variety.*

So in a sense we are doing a variational analysis within a space wherein which the class of varieties is naturally embedded, similar to the analysis of the class of Riemannian manifolds, via extension to statistical superstructures with the construction of the signal function $f(m, a) := \delta(\sigma(m) - a)$.

We are now ready to examine the Serre Duality theorem for schemes.

Theorem 3.2.4. *((Relative) Serre-Cech Duality). Suppose X is a projective scheme over an algebraically closed field k. Let $\mathcal{F}$ be a Cohen-Macaulay scheme containing X, and ω_X be the dualizing sheaf associated to X. Then there are natural functorial maps between associated Ext groups and Cech cohomology groups:*

$$\theta^i : Ext^i(\mathcal{F}, \omega_X(\mathcal{F})) \to H^{n-i}(X, \mathcal{F})$$

Furthermore, if X is also Cohen-Macaulay, these maps are isomorphisms.

In other words, we have that the relative Cech cohomology of X rel the supervariety $\mathcal{F}$ is dual to the relative Ext homology of the supervariety $\mathcal{F}$ with respect to the complementary subvariety to X in $\mathcal{F}$.

This duality thorem is the key result of the theory, and as a tool provides significant insight into the structure of algebraic varieties. I will not provide a survey of the results here, since they are really just mechanical calculations of the associated Cech cohomology groups, which are both extremely technical and, to my mind, not particularly enlightening.

A couple of the results in particular that can be proved are the celebrated Riemann-Roch theorem for curves and surfaces, and the Hurwitz theorem for representation of branched curves.

3.2.5 Concluding remarks

Essentially the etale cohomology theory now constructed is merely a way of providing us with stronger information than we would otherwise have regarding the geometry of algebraic varieties. In terms of geometric intuition, we could roughly think of this as computation of differential forms with respect to a four-tensor, in order to get control over the structure of a standard space where the geometry is determined by a standard Riemann-Cartan 2-tensor. Hopefully this remark will become somewhat clearer to the reader, in context of the developments later in this dissertation.

In particular, it is now possible to use this new cohomology theory to compute invariants of algebraic varieties with integer coefficients. By construction, the properties of these invariants are well understood, and provided Grothendieck with sufficient information to make progress on the motivating conjectures. However I have by no means provided a complete description of the constructions that he used to tackle these driving problems. To be more precise, he used "l-adic" Cohomology in order to obtain the resolution he needed. But I will not describe this here.

This completes my elementary description of the initial developments leading to K-theory.

3.3 Grothendieck-Galois theory

As an aside, I will now give an indication of a powerful generalisation of Galois theory, also due to Grothendieck. This allows one to characterise the category of continuous actions associated to a profinite topological group. This category moreover is actually a special type of category, as described earlier- it is a *Topos*. Consequently it is possible to apply much of the machinery from before.

I provide the results here mostly without proof; the interested reader is advised to read the wonderful paper by Dubac and de la Vega, [DV], which is my primary source for this material.

3.3.1 Transitive actions of a discrete group

The Galois-Grothendieck correspondence for transitive actions of a discrete group is described thusly. Suppose we have a category $\mathcal{C}$ that is somehow "nice" (in a way that I will make precise later). Consider an object in this category, A. Consider the space of arrows in $\mathcal{C}$ from A to some other object X, which I will denote by $[A, X]$. Define $G := [A, A]^{op}$, the opposite monoid of the monoid of endomorphisms of A; that is, $[A, A]$ but with arrows reversed.

Then there is an action of G on $[A, X]$, given by $\phi : G \times [A, X] \to [A, X]$, $\phi : (g, x) \mapsto gx : x \circ g$. Define $[A, X]$ equipped with this action ϕ as $[A, X]_G$.

Remark. Note that $g[A, X]$ will be a quotient of $[A, X]$ for "nice" categories (to be described below).

Recall now the definition of adjoint functor:

Definition 71. (Adjoint functor, [W3]). An *adjunction* between categories $\mathcal{C}$, $\mathcal{D}$ is a pair of functors $F : \mathcal{C} \leftarrow \mathcal{D}$, $G : \mathcal{C} \to \mathcal{D}$, such that

$$Hom_{\mathcal{C}}(FY, X) \cong Hom_{\mathcal{D}}(Y, GX)$$

We then say that F is *left adjoint* to G, and write $F \dashv G$.

Define now $H = Fix(x) = \{g \in G | gx = x\}$ for any arrow $x : A \to X$. But since the action of the group is transitive, we have that if $f : X \to Y$ in another arrow in $\mathcal{C}$ that $Fix(x) = Fix(f \circ x)$. Hence H is defined independent of x for a given object A.

I can now give the theorem for the Galois-Grothendieck correspondence for transitive actions of a discrete group:

Theorem 3.3.1. *(Galois-Grothendieck, v1).Suppose $\mathcal{C}$ is "nice" at an object $A \in \mathcal{C}$, that is*

(i) For all objects X there exists an epic *arrow $A \to X$,*

(ii) $A \to A/H$ for any subgroup $H \subset Aut(A)$ exists in $\mathcal{C}$, and

(iii) $[A, A] = Aut(A)$.

Then there is an equivalence of categories given by the adjunction $A \times_G (-) \dashv [A, -]$, with $[A, -] : \mathcal{C} \to tr^G$, $A \times_G (-) : tr^G \to \mathcal{C}$, where tr^G is the category of transitive G-sets. In particular, we have that the maps

(i) $[A, A]/H \to [A, A/H]$

(ii) $A/H \to X$

are isomorphisms.

For something in the way of interpretation of this result, we think of $[A, A/H]^{op}$ as being the analogue to the Galois group $G(A/H)$, where now the object A takes the role of a splitting field; $0_{\mathcal{C}}$ the base field, "k", and H the role of an intermediate field. Then the correspondence above essentially allows one to compute the intermediate objects in $\mathcal{C}$ between A and 0 given knowledge of $[A, A]^{op}$.

This is a significant generalisation of standard Galois theory because it is possible to have more than one object taking the role of a splitting field in a sufficiently complicated category. As we shall see in the later sections, it is actually possible to strengthen the theory further.

Ideally, we would like to somehow extend to the manifold case, where our objects are now points x in some differentiable structure, and we associate to each point arrows in such a way that the Galois theory extends. This is actually related to the idea of the self-referential calculus, which is discussed in chapter 8.

3.3.2 All continuous actions of a profinite group

I shall now develop the full version of the Grothendieck theory, following [DV].

Definition 72. (Profinite group). A *profinite group* is a Hausdorff, compact, and totally disconnected topological group.

Definition 73. (Finite object). An object A within a category $\mathcal{C}$ with Topos structure is *finite* if its size is finite with respect to the corresponding Grothendieck topology.

Let $\mathcal{C}$ be a category, and let $F : \mathcal{C} \to \mathit{Epi}$ be a functor mapping from $\mathcal{C}$ to the arrow category of epimorphisms. We will say that $X \in \mathcal{C}$ is finite if $F(X)$ is finite. Then suppose

(i) $\mathcal{C}$ has final object 1 and fiber products,

(ii) $\mathcal{C}$ has initial object 0, finite (and infinite) coproducts, and quotient of objects by a finite group,

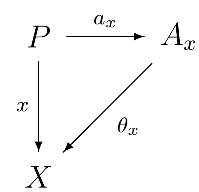
(iii) $\mathcal{C}$ has epi-mono factorisations, ie for any $f : X \to Y$, there is $X \to I \to Y$, where $e : X \to I$ is epi and $i : I \to Y$ is mono (ie, we have a weak cohomology theory vis a vis short exact sequences),

(iv) F is left exact (it preserves finite limits),

(v) F preserves the initial objects, finite and infinite sums, quotients by actions on finite groups, and strict epimorphisms are sent to surjections,

(vi) F reflects isomorphisms.

Consider $x : P \to X$, with $a_x : P \to A_x$ a factorisation as via item (iii):



Let $\pi := Aut(P)^{op}$. This has a natural continuous left action on $[P, X] =: F(X)$, since π is profinite, ie compact and Hausdorff. Write as before $[P, X]_\pi$ as $[P, X]$ equipped with this action.

Then

Theorem 3.3.2. *(Grothendieck Galois correspondence, v2). Suppose the above axioms hold for a pair $(\mathcal{C}, F)$ as above. Consider the functor $[P, -]_\pi : \mathcal{C} \to Epi^\pi$. Then this functor establishes an equivalence of categories.*

Corollary 3.3.3. *Let $f\mathcal{C}$ be the full subcategory of finite objects. Then $[P, -]_\pi$ restricts to a functor $[P, -]_\pi : f\mathcal{C} \to fEpi^\pi$ which is also an equivalence of categories.*

Roughly what this means is that these ideas do extend to the manifold case, provided that we equip our objects (sets in the manifold) with an appropriate Topos structure (say, integrand with respect to a volume form). Then, theoretically, one should be able to determine the information of a particular point in a manifold from that on various other points, with respect to the choice of profinite group action.

Naturally this is all extremely suggestive and might be related to interesting physics, as indeed it is.

3.4 Preliminaries of K theory

The K-theory of Atiyah and Hirzebruch grew out of a need to formalise Grothendieck's Etale cohomology program, and attempt to investigate its more general structures. This in fact led to the development of the K-theory as an overarching framework for investigating "extraordinary cohomology theories", of which the etale cohomology for schemes is one.

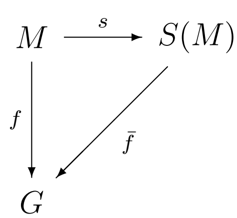
This is actually related roughly to the treatment in this dissertation in the chapters to follow; it turns out that there are actually *three* different ways to primitively extend de-Rham cohomology from the Riemannian case - these correspond to three different varieties of four-tensor construction, which are covered in exhaustive detail in chapters 5, 6, and 7. Of course the K-theory is more general and deals with more pathological constructs as well, but it is the extraordinary theories corresponding to 4-tensor structures that will be of most interest to us.

It turns out that a key result within this theory is the idea of Bott Periodicity, which has been adapted for various applications, such as the celebrated Atiyah-Singer index theorem [AS].

My primary source will be Max Karoubi's book [K], although as a secondary reference I will use lecture notes on the subject due to Atiyah [At].

3.4.1 Motivation and the definition of the Grothendieck group

Suppose we have a topological structure M with an abelian composition operation $+$ defined on it. Associate to M an abelian group $S(M)$ and a homomorphism $s : M \to S(M)$, such that for any abelian group G, and any homomorphism $f : M \to G$, we have that there is a unique $\bar{f}$ such that following diagram commutes:



In other words, $S(M)$ contains maximal information about M up to group structure.

Note that the motivation for doing this is that M, the structure of interest, may not have the properties of an abelian group, ie might not have inverses. Consequently in order to build a cohomology theory we need to extract information from this space into a new object $S(M)$, in a way as compatible with the structure of M as possible.

We are interested in this since this is related to the idea of generating groups for a space, similar to the idea of de-Rham cohomology, or etale cohomology. In particular, for a topological space X, one can build a simplicial representation. These simplices will generate a chain complex, which can be interpreted as an additive category M of vector bundles over X. Then there is a map s from this category to the de-Rham cohomology groups $S(M) := H(X)$.

If we were instead to view this more geometrically, suppose we endow X with a metric σ, and consider the set of forms $f(x)dx_1...dx_k$ such that these span k-dimensional geodesic submanifolds in X, up to isometry. Then the set of these forms for each k generates a chain complex M on X. Furthermore, we have a map $d(\sigma) : M \to H(M)$, such that $H(M)$ is given as the quotient of forms κ such that $d_\sigma \kappa = 0$, up to equivalence of boundaries, that is if $\gamma = \kappa + d_\sigma \lambda$, then $\gamma \sim \kappa$. Here $d_\sigma f = \partial_j(\sigma_{ij} f)$ is the gradient operator induced by σ.

To return to the development of our theory then, consider an additive category, $\mathcal{C}$; that is, a category where for any objects A, B there exists an object $A \oplus B$, and arrows $+_B : A \to A \oplus B$, $+_A : B \to A \oplus B$. To every $E \in \mathcal{C}$, let $\hat{E}$ be its isomorphism class; that is, the set of objects E_i such that the arrows $+_{E_i} : F \to E \oplus F$ map to the same object $E \oplus F$ for each F.

Then $M := \phi(\mathcal{C})$, the set of isomorphism classes under the additive structure of $\mathcal{C}$, becomes an abelian monoid if $\hat{E} + \hat{F}$ is defined to be $E \hat{\oplus} F$.

Then $S(M) = S(\phi(\mathcal{C}))$ is called the *Grothendieck group* of $\mathcal{C}$, and is equivalently written $K(\mathcal{C})$.

Suppose X is a compact manifold. Consider the additive category of vector bundles $\mathcal{C}$ over X. Then the Grothendieck group for $\mathcal{C}$ will be denoted $K(X)$, and, indeed in our example above, corresponds to the de-Rham cohomology if we are assuming a monoidal structure given by the algebra of forms over a Riemannian manifold.

3.4.2 Extension to functors

Suppose now we have a functor between categories $F : \mathcal{C} \to \mathcal{D}$. The motivation is that we are interested in examining the twisted product of two Riemannian structures on manifolds; in this case $\mathcal{C}$,$\mathcal{D}$ would correspond to the categories of vector bundles over a manifold M induced by the differing geodesic submanifolds for different metrics σ and τ. This will allow us to arrive at the etale cohomology of Grothendieck for schemes, or at least a generalised version that might be applicable to more exotic 4-tensor structures.

Definition 74. (Banach category). A *Banach structure* on an additive category $\mathcal{C}$ is given by a Banach space structure (ie limits exist) on $\mathcal{C}(E, F)$, and moreover the map

$$\mathcal{C}(E, F) \times \mathcal{C}(F, G) \to \mathcal{C}(E, G)$$

is bilinear and continuous. Then a *Banach category* is an additive category equipped with such a structure.

Essentially the key observation is to realise that a functor F between Banach categories $\mathcal{C}$ and $\mathcal{D}$ of vector bundles over compact spaces X and Y extends to a map between the corresponding groups $K(X) \to K(Y)$. Then $K(X, Y)$ is defined to be the group that makes the following sequence short exact:

$$0 \to K^i(X, Y) \to K^i(X) \to K^i(Y) \to 0$$

And this corresponds to the relative cohomology of a space, for $(X, X \cap Y)$, since short exactness preserves the information of the underlying categorical structure.

Of course we need not necessarily restrict ourselves to X, Y compact manifolds. These can also be Banach algebras, or groups, or other types of structures.

3.5 Bott Periodicity and some consequences

Theorem 3.5.1. *(Bott Periodicity). Suppose we are dealing with complex K theory. Then there is an isomorphism between the groups $K(X) \otimes K(S^2)$ and $K(X \times S^2)$ for all compact Hausdorff spaces X.*

Proof. (reference). A detailed proof of this theorem can be found in the thesis of David Tomairo, [To]. □

In other words, this states that the representatives between a compact space X and S^2, for a standard cohomology theory, are separable; that is if ω is a differential form in $X \times S^2$, representing a geodesic submanifold, then it factors into forms over X and over S^2. In other words, the geodesic submanifold N in $X \times S^2$ splits into a submanifold α of X and a submanifold β of S^2.

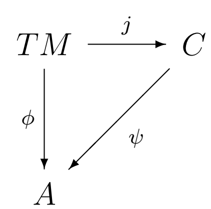
This obviously is of some interest, since this is related to the Hopf conjecture - whether there exist metrics of positive curvature on $S^2 \times S^2$. However it is not quite strong enough for such a purpose.

What it is useful for, however, is computation of the (higher) homotopy groups of spheres, which were not completely understood at the time that the Periodicity theorem was formulated by Raoul Bott.

It turns out to also be essential in the development of relative $K^{p,q}$ theory, which is its main important application.

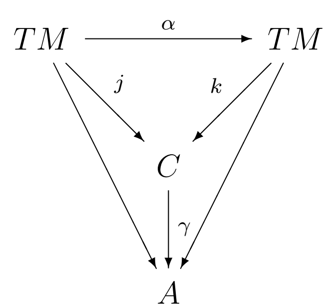
3.5.1 The groups K-pq

Before we can build the groups K^{pq} we need the idea of a Clifford algebra. Consider the tangent bundle TM (or more generally, the k-tangent bundle, with k a field) corresponding to a Riemannian manifold M with a quadratic form σ defined upon it. Let C be a R-algebra (or more generally, a k-algebra), and a map $j : TM \to C$ a homomorphism that encodes a maximal amount of information regarding TM, with $j(v)^2 = \sigma(v,v)\dot{I}d_C$; in other words we require that the following diagram commutes:



Then the pair (C, j), or $C(M, \sigma)$ is defined uniquely up to isomorphism. It will be called the Clifford algebra associated to the Riemannian manifold M.

More generically, suppose σ, τ are quadratic forms on TM, corresponding to different metrics on an underlying differentiable structure M. Let $\alpha : TM \to TM$ be the map such that $\langle v, w\rangle_\sigma = \langle \alpha(v), \alpha(w)\rangle_\tau$. Let C be a R-algebra, and suppose we have norm preserving maps $j, k : TM \to C$ that preserve the norms of σ, τ respectively, such that for any R-algebra A we have that the following diagram commutes:



Then $C(\sigma, \tau)$ will be unique up to isomorphism, and will be called the *generalised Clifford algebra* associated to $(M, \sigma) \times (M, \tau)$. This is a slight departure from Karoubi but is necessary; since schemes after all come from the twisting of two metrics, we need the necessary generality to proceed.

Regardless, this essentially allows us to provide the tangent bundle of a Riemannian manifold, or more generally, any vector space V, with the structure of an algebra, such that $(v \otimes w)_k = C_k^{ij} v_i w_j$. Consequently clifford algebras are often denoted $C^{ij}(M)$.

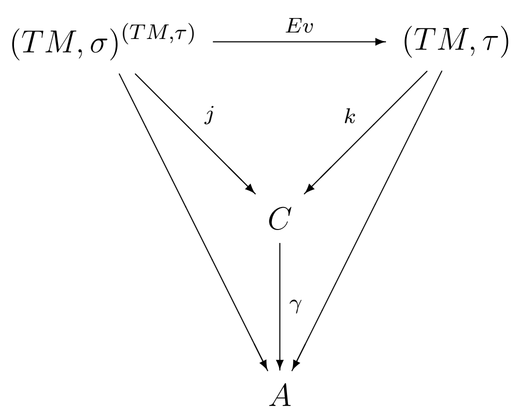
Suppose now that we are considering $\mathcal{C}(X)$ as the category of vector bundles over a compact base space X. Let $\mathcal{A}$ be a finite dimensional R-algebra. Then we can construct a new category $\mathcal{C}^{\mathcal{A}}(X)$ where the objects are (E, ρ), E an object in $\mathcal{C}(X)$ and $\rho : \mathcal{A} \to End(E)$ is an R-homomorphism that maps from the algebra to the space of Endomorphisms associated with the vector space E. Arrows, or morphisms in this space from (E, ρ) to (E', ρ') are $\mathcal{C}(X)$ morphisms $f : E \to E'$ such that $f \circ \rho(\lambda) = \rho(\lambda) \circ f$ for each $\lambda \in \mathcal{A}$.

In particular if $\mathcal{A}$ is the generalised Clifford algebra $C^{p,q}$ corresponding to (X, σ), then the corresponding category is denoted $\mathcal{C}^{p,q}$.

It is then possible to build functors $K^{pq}(X)$ in a manner analogous to the construction of the K^i; these essentially arise as the Grothendieck group of the category $\mathcal{C}^{p,q}$ of vector bundles equipped with the generalised Clifford action corresponding to a compact underlying space X.

But the K theory is more general than this; it is not only restricted to Clifford algebras for the tangent spaces of Riemannian manifolds. For one might consider instead multiplicative categories of vector bundles over compact spaces X and Y, where we have an operation on vector spaces defined as $V \times W := V^W$. In particular one builds a category $\mathcal{B}^{\mathcal{A}}$, using the categorical equivalent of set exponentiation.

Then define C to be the algebra such that the following diagram is commutative for any other algebra A:



where Ev is the evaluation map from the exponentiation of the tangent space of (M, σ) by that for (N, τ), back to (N, τ).

Then the K theory for the corresponding multiplicative functor will be different. Alternatively one might have fractal categories of vector bundles, with operation defined as $\partial(V; W) := \partial^W V$, where the differentiation is to be understood as differentiation with respect to a basis of the bundle corresponding to changes in position relative to the underlying differential forms - corresponding to a geodesic submanifold - in Y and X.

We consequently obtain different extraordinary cohomology theories corresponding to each such object, and I claim that these are the only extraordinary cohomology theories that can be obtained for pairs of spaces that have associated differential structure. The reason for this is not particularly deep - it is related to the fact that there are only three ways to build geometric theories for 4-tensor structures, ie to build affine-connections, as a consequence of the combinatorics. Note that there is only one way for 2-tensor structures; the de-Rham cohomology theory.

3.5.2 Relative K-pq theory

In an analogous manner to the standard K theory, it is possible to construct relative K^{pq} theory for pairs of compact space (X, Y), via defining the group $K^{pq}(X, Y)$ as that which makes the following sequence short exact:

$$0 \to K^{pq}(X, Y) \to K^{pq}(X) \to K^{pq}(Y) \to 0$$

In particular, if one is using a Clifford algebra, this corresponds to the relative etale cohomology for a scheme X rel Y, via the identity

$$E^{p,q}(X, Y) = K^{pq}(X, Y)$$

This is of interest in terms of Bott periodicity since K^{pq} theory can be used to understand real Bott periodicity. In particular it turns out there is really only one natural way to define a homomorphism

$$t : K^{p,q+1}(X, Y) \to K^{p,q}(X \times B^1, X \times S^0 \cup Y \times B^1)$$

Then we have that

Theorem 3.5.2. *(Fundamental Theorem). The map t is an isomorphism.*

Proof. The proof of this theorem is extremely technical, and is one of the more difficult parts of the theory to establish. The original proof by Raoul Bott [Bott] used Morse Theory [Mil]. Karoubi [K] pp161-174, follows the approach due to Atiyah and Bott [AB], which uses complex K-theory. □

From which we have the consequence of being able to compute the real K_R^i groups from the K_C^i, the latter of which are much easier to determine. This is the real reason Bott periodicity is so useful:

Theorem 3.5.3. *("K-duality", or computation of the integer real K groups). The following sequence is exact:*

$$\ldots \to K_R^{n-1}(X,Y) \to K_C^{n-1}(X,Y) \to K_R^{n+1}(X,Y) \to K_R^{n}(X,Y) \to K_C^{n}(X,Y) \to K_R^{n}(X,Y) \to \ldots$$

Proof. The t-isomorphism is required to make this proof work. Further details can be found in [K], pp154-156. □

This is an analogue in a way of Serre duality. Here, the K^{pq} groups play the role of the Etale cohomology to extend Poincare duality from the nondegenerate quadratic forms of de-Rham cohomology to the degenerate quadratic forms for elliptic curves in algebraic topology and Cech cohomology - the groups K^n.

3.5.3 Application to the Atiyah-Hirzebruch formulation of the Riemann-Roch Theorem

There are various applications of K-theory, which involve various forms of extraction of information of a geometric or algebraic structure. For instance, study of the Hopf invariant, and study of vector fields on spheres - in particular demonstrating that the only spheres that can be provided with H-space structures are S^1, S^3, and S^7 are within the province of the theory. (An H-space structure on a sphere is an operation $m : S^{n-1} \times S^{n-1} \to S^{n-1}$ such that the restriction maps $x \mapsto m(x, x_0)$, $y \mapsto m(x_0, y)$ are homotopic to the identity map of S^{n-1}; ie they are maps of degree 1. This is an abstraction of the idea of field structure on a sphere.) More generally, one can construct characteristic classes, which are generators of particular cohomology theories associated to a manifold, in order to study the properties of a geometric structure, using this theory.

However of all of these applications, it is probably the Atiyah-Hirzebruch, or Atiyah-Singer Index theorem, which is the most important. It is essentially the formulation of the Riemann-Roch theorem extended to the generalised cohomology of K-theory. Since I did not cover the Riemann-Roch theorem, which is a result of the classical theory, and its extension to etale cohomology earlier, I will do so now before discussing the formulation within K-theory.

The original statement of the Riemann Roch theorem is essentially an inequality computed for Riemann surfaces.

Theorem 3.5.4. *(Riemann Roch). Let M be a Riemann surface of genus g. Let D be a divisor of M - ie a codimension one subvariety. Then we have that*

$$I(D) - I(M - D) = deg(D) - (g - 1)$$

or more crudely,

$$I(D) + (g - 1) \geq 0$$

where $I(D)$ is the index *of D- ie the dimension of the vector space of functions h on M such that h rel D is non-negative- and $deg(D)$ is the* degree *of D, ie the degree of the field extension $[K(D) : K(M)]$ with respect to the natural embedding map $i : D \to M$.*

The latter statement looks like a very weak analogue of the Cramer-Rao inequality. The interpretation would be here that $I(D)$ represents the "transverse information" and the genus the "longitudinal information"; so the information of M splits into D and $M - D$ and we have that the Fisher Information roughly satisfies the equality above - although of course there are cross terms which are taken into account in the strong statement of the theorem. Indeed the idea of Index is very much a weak algebraic analogue of the idea of Fisher Information.

To be more precise, the idea of Index in K-theory is an abstraction of the idea of the Index of a metric. Recall that in pseudoriemannian geometry we refer to the idea of a geometric structure Λ associated to a manifold M as having index k if the dimension of its negative eigenspace is k. So if Λ is a p-tensor, we have that an *eigenvector* v with corresponding *eigenvalue* λ satisfies the equation $\Lambda(v, ..., \hat{v}, ..., v) = \lambda v$, for any removed entry $\hat{v}$ at the ith position, for $1 \leq i \leq p$.

One can construct various forms of K-theoretic structures via natural "twist" operators of which there are three natural types for 4-tensors - $\star$, $\wedge$, and ∂^*. Then it is natural to talk of the index of the 4-tensor constructed in this fashion as the index of the associated operator. It turns out that these operators are moreover *elliptic*, that is they admit a representation roughly of the form $a_{i_1 \cdots i_p} D_{i_1} \cdots D_{i_p}$, for some $p \geq 1$ an integer, such that a is a tensor with positive eigenvalues, or at least eigenvalues compatible with the causal signature of the domain of the operator. (To be more precise, it turns out for 4-tensors we only need to consider $p = 1$ or $p = 2$). This is the general idea of the *Atiyah-Singer index theorem*, to try to understand the index and hence causal properties of K-structures better.

To return to our original discussion, the Riemann Roch theorem is extended to Grothendieck's theory of schemes in the following manner.

Theorem 3.5.5. *(Grothendieck-Riemann-Roch, [W5]). Let Y be a scheme. Suppose that X is a subscheme divisor, ie there is a natural inclusion map $f : X \to Y$. Let $\kappa(f_\star) : K_0(X) \to K_0(Y)$ be the natural map in Cech Cohomology with respect to the Grothendieck group K_0, where $f_\star : A(X) \to A(Y)$ - the induced map between the algebras $A(X)$, $A(Y)$ of subschemes of X and Y. Let $ch : K_0(X) \to A(X)$ be the Chern character, that is, an equivalence between generalised forms in $K_0(X)$ and schemes in $A(X)$, where we are carrying over the tensor algebra structure in K_0 to A. Then the following identity holds:*

$$ch(\kappa\mathcal{F})td(Y) = f_*(ch(\mathcal{F})td(X))$$

where $\mathcal{F}$ is essentially a sheaf perpendicular to X in Y - if X was a submanifold of a manifold Y it would be a space transverse to X - and also $td(Y)$ is the Todd genus *of a scheme Y, which is essentially the extension of the idea of genus to schemes. Roughly, the Todd genus of a scheme Y is roughly the reciprocal of the Chern class of Y - it is a measure of the information perpendicular to that which the Chern class is measuring. Roughly it deals with the information of the* co-normal generalised bundle *associated to Y, rather than the* normal generalised bundle. *[W6]*

Of course this is rather difficult to digest. But essentially what this is saying is that, given a scheme Y and a subscheme X, the two different natural ways of computing the information of the part of the structure associated to X - either directly or through computation of the part of the structure normal to X - are equal. In other words, to put this into the terminology of Professor Frieden, we have that the channel information I and the bound information J are equivalent ;

ie that $I - J = 0$, if $I = ch(\kappa\mathcal{F})td(Y)$, and $J = f_*(ch(\mathcal{F})td(X))$. This suggests in turn that the structures we are examining must be critical in some way; I would hazard that the construction of the correspondence $ch : K_0(X) \to A(X)$ implies the criticality of X as a subscheme of Y.

We are now ready to begin to examine the Atiyah-Singer index theorem.

To reiterate, this essentially deals with the idea of an *elliptic operator*, or "twist map", that is, a map T roughly of the form $T\phi := a_{ij}D_iD_j\phi$, such that $a_{ij}(x)$ is a positive definite matrix for each x over the relevant domain. Broadly speaking, an operator defined over a vector bundle over a base space that is pseudoriemannian is elliptic if it is compatible with the signature of the base.

More generally, an operator T of the form $T\phi := a_{i_1...i_k}D_{i_1}...D_{i_k}\phi$ is said to be elliptic if $a_{i_1...i_k}(x)\zeta_{i_1}...\zeta_{i_k} > 0$ for all nonzero vectors ζ and all x in the relevant domain.

An important example of an elliptic operator - though possibly a slightly confusing one in this instance - is the Riemann curvature tensor R_{ijkl}. $R : TM \times TM \to \{TM \to TM\}$, or, equivalently, $R : TM \times TM \to TM \times TM$. Consequently it is sensible to consider elliptic operators $T : E \to F$ that map between vector bundles E, F over pseudoriemannian manifolds M, and more generally this is what the Atiyah-Singer index theorem is about.

To make this slightly more precise, I will now introduce a few definitions:

Definition 75. (Kernel and cokernel). The *kernel* $ker(S)$ of an operator $S : E \to F$ is the subspace of X that is mapped to 0 under S. The *cokernel* $coker(S)$ of an operator $S : E \to F$ is the quotient of F by the image of S in F.

Definition 76. (Fredholm operator). A *Fredholm operator* is a bounded linear operator between two Banach spaces whose *kernel* and *cokernel* are finite dimensional and whose range is closed. Equivalently, $S : E \to F$ is Fredholm if there exists a $T : F \to E$ such that $Id_E - T \circ S$ and $Id_F - S \circ T$ are compact operators, that is, they are invertible.

Definition 77. (Analytical Index). Suppose $D : E \to F$ is an elliptic operator. Then it has a pseudoinverse, that is, there exists a D' (which might not be unique) such that $D \circ D' - Id$ and $D' \circ D - Id$ are invertible. Consequently it is a *Fredholm operator*, and the kernel and cokernel are both finite dimensional. Hence we can define the *analytical index* of D to be

$$I(D) := dim(ker(D)) - dim(coker(D))$$

ie the dimension of the failure of D to be injective, minus the dimension of the failure of D to be surjective. Note that if D' is a pseudoinverse of D then $ker(D') = coker(D)$, so we see that essentially the index then measures the mismatch of injectivity between D and a pseudoinverse representative.

Definition 78. (Todd class). Recall the *Todd class* $td(X)$ associated to a space X, is essentially the extension of the idea of genus to objects more general than surfaces. It is the *characteristic class* associated to the co-normal bundle associated to X, as opposed to the Chern class of a space E/X, which is the characteristic class associated to the normal bundle associated to E/X. So for instance if E is a vector bundle over X, it is sensible to consider the information regarding how X embeds in E; consequently $td(X)$ will provide this information. Suppose now we consider an isotopy of embeddings of X in vector bundles induced by a map $D : E \to F$. Then we define $td_D(X)$ to be the characteric class associated to the embedding of X within that isotopy.

Definition 79. (Chern character of an elliptic operator). Suppose $F \subset E$. We can represent the information of the space normal to F in E within K theory by an elliptic operator $D : E \to F$, which in this case will be the projection mapping. Then the Chern character of D, $ch(D)$ is defined to be roughly the Chern class of the kernel of D. Note that in this case $coker(D) = \phi$.

Of course, if we reverse the roles of E and F, so that $E \subset F$, then D becomes the inclusion map, and $ker(D) = \phi$. Consequently it is sensible to define $ch(D)$ as the Chern class of the kernel of D, minus the Chern class of the cokernel of D. These ideas more generally extend to any pair of Banach spaces E, F linked by an elliptic operator $D : E \to F$, of course.

Remark. In practice $ch(D)$ is an operator that turns pseudodifferential forms into integers.

Definition 80. (Topological Index). Suppose E and F are smooth vector bundles on a compact manifold X, and D is an elliptic differential operator mapping between them. Then the *topological index* of D is defined to be

$$ch(D) \circ td_D(X)$$

Since $ch(D)$ is essentially $ch(ker(D)) - ch(coker(D))$, and since $td_D(X)$ is essentially the form relating the embedding of X into $D : E \to F$, it is natural to assume that the following result should follow:

Theorem 3.5.6. *(Atiyah-Singer Index Theorem, 1963, [W7]). Suppose T is an elliptic differential operator on a compact manifold M. Then the associated* topological index *is equal to the* analytical index.

This is stronger than Grothendieck-Riemann-Roch and has interesting consequences. For instance it is possible to prove certain intuitive results, such as for instance proving that ascending an infinite staircase to get back to where one starts is impossible - as might seem to be the case in various drawings by Escher.

More generally, this result has consequences in computing weak invariants for physical theories, as it is weakly related to the idea of bound information J and channel information I expounded by Frieden in his treatments of Fisher Information. There is also the strong connection in that the Fisher Information density for pseudoriemannian manifolds is encoded by an elliptic operator, although evidently one is not computing the Fisher information here, but rather the index of the associated map from the manifold back to itself, which forms (in the case of the Riemann curvature tensor) a four tensor which actually contains less information than the original metric.

In the case of the exotic geometries explored in the later chapters in this volume, it turns out that the generalised christoffel symbols often do become higher order in derivatives - so it is not quite so strange to consider the adoption of an elliptic operator to describe the transition from two Riemannian manifolds to something more abstract.

There are also many generalisations of the Atiyah-Singer index theorem that are possible, following the same style of abstractive process, but I will not go into these here.

3.6 L theory

It is possible, of course, to generalise further, and consider structures $K^{p_1,\dots,p_n}$, for any finite n. This might seem incredibly general, and it is, but it is not necessarily natural. For the foundations of K-theory lie firmly within the domain of 1-categories, with topoi as the structures of primary interest. To extend to triples $K^{p,q,r}$ intuition suggests that we are neglecting essential structure.

In fact, this is indeed the case, and it is necessary to build a new extraordinary cohomology theory grounded in the domain of 2-categories; in particular, one is interested in using twistors as the fundamental objects of study. This leads to the study of L-theory, [W4]. L-theory ultimately leads to the construction of groups $L^{p,q,r}$ which are useful in the study of surgery.

It is reasonable to assume that the duality theorem for L-theory should be useful in the study of extraordinary cohomology theories for degenerate 4-tensor Λ-structures on manifolds, ie the study of matrices of groups $K^{p,q}(M, \Lambda)$. However this goes significantly beyond the scope of this dissertation. The interested reader is advised to refer to the sources [Luck] and [Wall], if they wish to learn more about surgery and its connection to L-theory.

Chapter 4

Sieve Theory and Integrability

Within number theory, sieve theory is the study of methods whereby one can gain information about particular sets of numbers of interest. These are usually prime numbers. Integrability is the study of solutions of infinite iterative hierarchies of PDEs, which are related to various phenomena such as study of solitons and emergent solid state behaviours.

At first glance these areas might not seem to be related. But consider now instead the theory of n-categories, in the limit as n tends to infinity. Call the resultant objects limit-categories. It is self-evident that a full and finite description of limit categories is impossible; the self-referential calculus - the core of this dissertation - is an 8-tensor theory, with foundations in the study of 1-categories. Further extension of the ideas in this work to 2-categories requires a 128 (or possibly 192) tensor theory. Beyond this point I am fairly certain that the rate of increase in complexity is super-exponential.

There is however some interesting recent work on limit categories due to Jacob Lurie in his recent substantial dissertation [Lurie]. The general idea of his approach is not to treat these with full generality but to consider what he calls (∞, n)-categories, where all k-morphisms are invertible for $k > n$. This turns out to restrict the structure to an extent where it is possible to say something. In particular the core of his analysis is to consider $(\infty, 0)$-categories, however I will not pretend to understand his precise approach. Certainly however his results might well be applicable to the study of integrability and sieves.

An additional reference that might also be of some interest to the reader is the recent preprint by Carlos T. Simpson, [Simp], which is a comprehensive treatment of (∞, n)-categories.

Along these lines, I maintain that under certain circumstances it should be possible to extract partial information about limit structures. In the simplest case, this is nothing other than the construction of a sieve, or, more generally, the study of critical iterative hierarchies of PDEs.[1]

Further afield, it is possible that the study of the partial information of limit categories is tied to the foundations of *statistics* and probability theory. In fact it seems natural to view the discipline of statistics in this manner, as it is by nature an approximative and imprecise area, associated with the calculation of invariants providing partial information of structures whose true underlying behaviours are ultimately unknowable.

This is perhaps the main difference between the geometric and statistical points of view. The geometer starts from a clear idea of the objects of interest and investigates the resultant structures in a formal and deterministic manner. This is useful for making predictions and within the practice of engineering, since such models tend to deal with smooth data. The statistician starts assuming nothing about the objects of interest, but presumes that there are certain ways to extract information about them; hence they proceed in the fashion of a data analyst. Such an approach is necessarily inherently discrete in nature. Consequently this is why statistics is of such great value in the analysis of experimental data, as a means of testing hypotheses, usually in the form of geometric models, built by theoreticians.

4.1 Preliminaries

4.1.1 Basic concepts

Before I proceed I will need to describe a few preliminary results and basic concepts. In this I follow the wonderful book by Cojocaru, [Co].

Definition 81. (Möbius function). The *Möbius function*, $\mu : N \to \{0, 1\}$ is defined such that $\mu(1) = 1$, $\mu(p) = -1$ for each prime p, and $\mu(p^a) = 0$ for any integer $a > 1$.

Then this has the following property:

[1]Indeed, I suspect it should be possible to extend the idea of category in an appropriate fashion - from the idea of n-category, a fundamentally geometric concept, to its "statistical dual" as an n^*-category. Via abuse of procedure I will regard the structure of (∞, n)-categories and the structure of (as yet not precisely known by me) n^*-categories as one and the same; nonetheless the reader should bear in mind that it is possible some slight adaptation of the Simpson-Lurie approach might be required in order to find a sensible foundation for the analysis.

Lemma 4.1.1. *(Divisor property of the Möbius function).*

$$\Sigma_{d|n}\mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{otherwise} \end{cases}$$

This allows one to establish

Proposition 9. *(Inversion formula, v1). Suppose $f, g : N \to C$ are two complex valued functions, and that*

$$f(n) = \Sigma_{d|n} g(d)$$

then

$$g(n) = \Sigma_{d|n}\mu(d) f(n/d)$$

Proposition 10. *(Inversion formula, v2). Suppose $D \subset N$ is a divisor closed subset of the naturals. Let $f, g : N \to C$ be two complex valued functions as before. Then if*

$$f(n) = \Sigma_{n|d, d\in D} g(d),$$

we have that

$$g(n) = \Sigma_{n|d, d\in D}\mu(d/n) f(d)$$

An important result in the beginnings of Sieve theory is Chebycheff's theorem. In its strong form it allows one to conclude that there is always a prime between n and $2n$ for $n > 1$.

Theorem 4.1.2. *(Chebycheff). Define $\theta(x) = \Sigma_{p\leq x} log(p)$, where p is prime. Then there exist constants A and B such that*

$$Ax < \theta(x) < Bx$$

One then can use the technique of partial summation to demonstrate that $\pi(x)$, the number of primes up to x, is $O(x/log(x))$. To remind the reader of this technique, I provide the following:

Theorem 4.1.3. *(Partial summation). Suppose $c : N \to C$ is a sequence of complex numbers. Define $S(x) = \Sigma_{n \le x} c(n)$. Fix a natural number n_0. If $c(j) = 0$ for $j < n_0$ and $f : [n_0, \infty) \to C$ is smooth, then for any integer $x > n_0$, we have*

$$\Sigma_{n \le x} c(n) f(n) = S(x) f(x) - \int_{n_0}^{x} S(t) f'(t) dt$$

Proof. Essentially integration by parts. □

In fact, the well known prime number conjecture of Gauss (proved by Hadamard and Poussin) states that $\pi(x) \sim x/log(x)$, so it can already be seen that these elementary methods provide some insight into the distribution of the primes.

Proof. (of Chebycheff's theorem, due to Ramanujan. [Co] pp6-7).

Define $\psi(x) = \Sigma_{p \text{ prime}, \, ,a \in N, p^a \le x} log(p)$.

Let $T(x) = \Sigma_{n \le x} \psi(x/n)$.

Then $\Sigma_{n \le x} log(n) = \Sigma_{n \le x} (\Sigma_{p^a | n} log(p)) = \Sigma_{m \le x} \psi(x/m) = T(x)$.

It follows then by partial summation that $T(x) = xlog(x) - x + O(log(x))$. Consequently $T(x) - 2T(x/2) = log(2)x + O(log(x))$. However we also have by the definition of $T(x)$ that $T(x) - 2T(x/2) = \Sigma_{n \le x} (-1)^{n-1} \psi(x/n)$.

Now suppose $a_i \ge a_{i+1}$, $i \in N$ is a sequence of decreasing reals, such that $a_i \to 0$. Then it follows that

$a_0 - a_1 \le \Sigma_{n \in N} (-1)^n a_n \le a_0 - a_1 + a_2$.

If we take $a_n = \psi(x/n)$ this will be such a decreasing sequence, so we can now observe that, together with the above, we have

$$\psi(x) - \psi(x/2) + \psi(x/3) \ge log(2)x + O(log(x))$$

Similarly, we have

$$\psi(x) - \psi(x/2) \le log(2)x + O(log(x))$$

Suppose now we replace x with $x/2^k$, for $k = 0, 1, 2, ...$ Then we get a family of sequences $a_{nk} = \psi(x/(2^k n))$, which all satisfy similar inequalities. Following the logic through carefully, it is possible to conclude that

$$\psi(x) \le 2(log(2))x + O(log^2(x))$$

and we also obtain a lower bound for $\psi(x)$ in this manner which is linear in x.

□

Proof. (of Chebycheff's theorem, due to Chebycheff, [Co], pp8-9). Recall $\theta(x) := \Sigma_{p \text{ prime}, p \leq n} log(p)$.

Now, it is clear that $\Pi_{n<p\leq 2n, p \text{ prime}} p | \binom{2n}{n}$, and $\binom{2n}{n} \leq 2^{2n}$.

Consequently by passing to logarithms, we see that $\theta(2n) - \theta(n) \leq 2nlog(2)$. Then we have a sequence $\theta(2(n/2^k)) - \theta(n/2^k) \leq 2(n/2^k)log(2)$, for $k \in N$. Summation of this over k gives the inequality

$$\theta(2n) \leq 4nlog(2)$$

So we have demonstrated $\theta(x) = O(x)$. □

There are two more results due to Chebyshev that will be required in the arguments to follow.

Theorem 4.1.4. $\Sigma_{p \leq n, p \ prime} \frac{log(p)}{p} = log(n) + O(1)$

Proof. ([Co], p9). Note that $n!$ factors as $\Pi_{p\leq n} p^{e_p}$ for certain coefficients e_p. But note that since the number of multiples of p^k that are smaller than n is $[n/p^k]$, it follows that

$$e_p = \Sigma_k [n/p^k]$$

and this sequence will obviously terminate.

Then $log(n!) = \Sigma_{p\leq n, p\text{prime}} e_p log(p)$. But $log(n!) = \Sigma_{k\leq n} log(k) = nlog(n) - n + O(log(n))$, and

$$\Sigma_{p\leq n, p\text{prime}}([n/p^2] + [n/p^3] + \cdots]log(p) \leq n\Sigma_{p \text{ prime}} \frac{log(p)}{p(p-1)} << n$$

so $\Sigma_{p\leq n, p \text{ prime}}[n/p]log(p) = nlog(n) + O(n)$. □

Corollary 4.1.5. $\Sigma_{p\leq n, p \ prime} \frac{1}{p} = loglog(n) + O(1)$

4.1.2 A few elementary sieves

According to [Co], the Sieve problem is described in the following manner. Let H be a countable set of objects of finite measure, and P an index set, such that for each $p \in P$, there is an associated $H_p \subset H$. Then we wish to find upper and lower bounds for the size of

$$S(H, P) := H - \cup_{p \in P} H_p$$

To recast this into the terminology of the partial information of limit categories, suppose we have a countable set of objects, H, such that for each object $h_n \in H$, there are arrows and product operations $\phi_{nm} : h_n \to h_m$, $\phi_{n_1 n_2 n_3, 1} : h_{n_1} h_{n_2} \to h_{n_3}$, $\phi_{n_1 n_2 n_3, 2} : h_{n_1} \to h_{n_2} h_{n_3}$, third order arrows and product operators, fourth order arrows and product operators, etc. Let P be an index set, such that for each $p \in P$, we associate a subcategory (of finite complexity) H_p to H.

Then we wish to estimate the size of

$$S(H, P) := H - \cup_{p \in P} H_p$$

in terms of a measure of finite complexity on H, which will often be inherited from the structure of the subcategories H_p. Indeed, often we will want each H_p to have the same categorical structure, and rather have the difference between them to be in terms of the objects in H that they contain.

If H is a subset of the naturals, say $\leq x$ and H_p is a congruence class modulo p, say numbers such that $p|n$, then $S(H, P)$ will be the set of numbers $n \leq x$ coprime to P. This in fact arises inductively via the process of elimination in the Sieve of Erastothenes, where one has a sequence $P_k \subset P_{k+1}$ of index sets which are generated in terms of what is excluded at the kth iterate.

Definition 82. (von Mangoldt function). The following function -

$$\Lambda(n) := \begin{cases} log(p) & \text{if } n = p^a, \text{ some prime } p, \text{ some natural } a, \\ 0 & \text{otherwise} \end{cases}$$

is referred to as the von Mangoldt function.

Theorem 4.1.6. *(Gallagher's Larger Sieve). Instead of generating $B := S(H, P)$, one main approach to Sieve theory is to estimate it. For suppose B is a set of integers, and we know its image mod t for any $t \in T$, an index set of prime powers. Suppose furthermore there is a function u, such that $|B \ mod\ t| \leq u(t)$, where we are using the natural measure inherited from the categorical structure on B - usually the counting measure.*

Let X be the largest element in B. Then if $\Sigma_{t\in T}\frac{\Lambda(t)}{u(t)} - log(2X) > 0$, we can estimate the size of B to be

$$|B| \leq \frac{\Sigma_{t\in T}\Lambda(t) - log(2X)}{\Sigma_{t\in T}\frac{\Lambda(t)}{u(t)} - log(2X)}$$

Proof. (Sketch). The trick is to enumerate each residue class mod t for each $t \in T$ of B, calling them say $Y_t(B, r)$, and then observe that for each t the number of elements of B will be equal to the sum of the size of these classes. Then via use of the Cauchy-Schwarz inequality, using the fact we have a bound $u(t)$ on the size of each class, it is possible to observe that

$$|B| = \Sigma_r|Y_t(B, r)| \leq u(t)^{1/2}(\Sigma_r|Y_t(B, r)|^2)^{1/2}$$

Then after some rearrangement, and using the fact from before that $\Sigma_{t|n}\Lambda(t) = log(n)$, it is useful to multiply the inequality on both sides by $\Lambda(t)$ in order to complete the proof. □

Definition 83. (Jacobi symbol). The Jacobi symbol is defined to be the function such that

$$\left(\frac{a}{p}\right) := \begin{cases} 0 & \text{if } a \equiv 0\ , \\ 1 & \text{if } a \not\equiv 0 \text{ and } a \equiv x^2, \\ -1 & \text{if } a \not\equiv 0 \text{ and } a \text{ is squarefree} \end{cases}$$

where all congruences are taken modulo p.

Theorem 4.1.7. *(The square sieve). Suppose we have a set of naturals, A, and we wish to estimate the number of squares therein. Let P be a set of primes not including 2. Define $S(A) = |\{a \in A | a \text{ is square }\}|$. Then*

$$S(A) \leq \frac{|A|}{|P|} + max_{p,q\in P, p\neq q}|\Sigma_{a\in A}\left(\frac{\alpha}{pq}\right)| + E$$

where $\left(\frac{\alpha}{pq}\right)$ *is the Jacobi symbol, and E is an error term of order roughly* $\frac{|\cup_{a\in A}\{p|a,p\in P\}|}{|P|}$*, which is usually negligible.*

Proof. (sketch). In estimating the size of the number of squares we expect the appearance of the Jacobi symbol, since this is intimately connected with the computation of quadratic residues. In fact all that one needs is to observe that

$$\Sigma_{q\in P}\left(\frac{a}{q}\right) = |P| - |\cup_{a\in A}\{p|a, p\in P\}|$$

and the rest of the proof follows relatively mechanically. See [Co], pp21-22. □

Theorem 4.1.8. *(The Tauberian theorem for Dirichlet series). Consider a Dirichlet series*

$$\zeta(s) = \Sigma_n \frac{a_n}{n^s}$$

such that the a_n is a non-negative sequence in $l^\infty(N)$. It is well known then that ζ has an analytic continuation and will be analytic for $Re(s) = 1$ except at $s = 1$. Make the assumption that the singularity at $s = 1$ for $\zeta(s)$ is of the form

$$\zeta(s) = \frac{H(s)}{(s-1)^{1-\alpha}}$$

for some real α and $H(s)$ locally nonsingular, and analytic for $Re(s) \geq 1$. Then

$$\Sigma_{n\leq x} a_n \sim \frac{cx}{log(x)^\alpha}$$

with $c = \frac{H(1)}{\Gamma(1-\alpha)}$.

As a special case this allows us to establish an estimate on the number of primes less than a particular $n \in N$, if $a_n = 1$ for all n. So this is nothing really terribly new.

But note that this result is significantly stronger. For instance if we define

$$a_n = \begin{cases} 1 & \text{if } n \text{ is the sum of two squares,} \\ 0 & \text{otherwise} \end{cases}$$

then it is possible to demonstrate using properties of the associated L-function that

$$\zeta(s, \{a_n\}) = \frac{H(s)}{(s-1)^{1/2}}$$

about $s = 1$, and consequently by the above result that

Corollary 4.1.9. *The number of $n \leq x$ that can be written as the sum of two squares is $O(\frac{cx}{\sqrt{log(x)}})$.*

So we can already start to see that we can deduce interesting properties about sets of natural numbers using these elementary methods.

4.1.3 The normal order method

Continuing to follow [Co], I now give an abbreviated sketch of a technique due to Hardy and Ramanujan, and later expanded upon by Paul Turan.

Definition 84. (Radical). Let n be a natural number with prime factorisation $p_1^{a_1}...p_k^{a_k}$. Then the *radical* of n, $Rad(n) = \Pi_i p_i$, and $\nu(n) = k$ is the number of distinct prime factors.

Theorem 4.1.10. *(Hardy-Ramanujan). We have the following weak estimates on ν:*

$$\Sigma_{n \leq x} \nu(n) = xloglog(x) + O(x)$$

$$\Sigma_{n \leq x} \nu(n)^2 = x(loglog(x))^2 + O(xloglog(x))$$

Proof. (due to Turan, transcribed from [Co], pp33-4). First observe that

$$\begin{aligned}\Sigma_{n \leq x} \nu(n) &= \sigma_{p \leq x, \text{pprime}} \left[\frac{x}{p} \right] \\ &= x \Sigma_{p \leq x} \frac{1}{p} + O(x) \\ &= xloglog(x) + O(x) \text{ by a previous result}\end{aligned}$$

For the second part of the proof, we deduce

$$\begin{aligned}
\Sigma_{n\le x}\nu(n)^2 &= \Sigma_{n\le x}\Sigma_{p|n}\Sigma_{q|n}1 \text{ by definition}\\
&= \Sigma_{p,q\le x}\Sigma_{n\le x,p|n,q|n}1\\
&= \Sigma_{p,q\le x,p\ne q}\left[\frac{x}{pq}\right] + \Sigma_{p\le x}\left[\frac{x}{p}\right] \text{ breaking into cases } p=q \text{ and } p\ne q\\
&= \Sigma_{pq\le x}\left[\frac{x}{pq}\right] + O(xloglog(x)) \text{ using a previous estimate}\\
&= x\Sigma_{pq\le x}\frac{1}{pq} + O(xloglog(x))
\end{aligned}$$

Now $(\Sigma_{p\le\sqrt{x}}\frac{1}{p})^2 \le \Sigma_{pq\le x}\frac{1}{pq} \le (\Sigma_{p\le x}\frac{1}{p})^2$. But $\Sigma_{p\le\sqrt{x}}\frac{1}{p} = loglog\sqrt{x} + O(1) = loglog(x) + O(1)$, so we have that the lower and upper bounds in the above are the same, up to a term of order one. This completes the proof. □

The connection to statistics then becomes apparent when we view $loglog(x)$ as a "mean" for the number of distinct prime factors in naturals less than x. In fact, an easy consequence of the above is that the *variance* $\Sigma_{n\le x}(\nu(n) - E(x))^2$ is $O(xloglog(x))$. This is the theorem due to Turan.

A generalisation of this method is to consider instead of irreducible natural numbers (ie primes) to consider irreducible polynomials over the naturals, with image in the naturals. This has the interpretation of a self-referential map on N. In the general case when we consider N to be a manifold rather than the set of natural numbers, we have interesting consequences. Nonetheless we would like to know if the methods extend in the simpler case first.

In particular we would like to estimate $\nu(f(n))$ for an irreducible $f(x)$.

It turns out that one can do this, or at least estimate the mean, as

$$\Sigma_{n\le x}\nu(f(n)) = \Sigma_{p\le y}\left(\frac{x\rho_y(p)}{p} + O(\rho_f(p))\right) + O(x)$$

which turns out to be

$$\Sigma_{p\le y}\left(\frac{x\rho_y(p)}{p}\right) + O(x) + O(y)$$

as $\rho_f(p) \leq deg(f)$, where $\rho_f(p)$ is defined as the number of solutions mod p of $f(x) \equiv 0$.

In fact, it turns out via fairly deep but established classical results that the mean and the variance for irreducible maps $N \to N$ are of the same order as for the basic case.

4.2 Various well known sieves

4.2.1 The Sieve of Erastothenes

The theory of sieves has quite ancient origins. Indeed, it has its beginnings in an observation due to the Greek mathematician Erastothenes (276 BC - c. 195 BC) regarding how to brute force generate the set of prime numbers.

The general idea is quite simple. If one wishes to generate all primes less than a given number N, consider the sequence of numbers S_0 from 1 to N. Then strike out multiples of 2 from 1 to N, to generate a smaller set S_1. For the next largest number than 2 in S_1, this number must also be prime; strike out all multiples of this number in S_1, to generate a new set S_2. This process is repeated until one reaches a prime exceeding $\sqrt{N}$. We may then stop and have a set S of all primes from 1 to N.

This result, or sieving method, is known as the Sieve of Erastothenes.

There have since been minor amendments made to this general idea. Most notable amongst these are Brun's sieve, the Selberg sieve, and Rosser's sieve. I will give a brief description of these presently.

Furthermore, sieve theoretic methods have long been known to give surprisingly deep information. It is the view of the author that it might well be possible to extract optimal information from a sieve by recasting the theory in the language of meta-geometry (which will be described later). I shall defer this task for now, and concentrate instead on a survey of the existing state of the field.

4.2.2 Sieve invariants

In much the same way as characteristic classes represent the information of particular cohomology theories, as the generators of their homology (and higher homology) groups, we have that in geometric statistics, sieves represent the information of the dual analogue of cohomological structures. Instead of forms for *submanifolds* in the overall space M, in the case of sieve invariants one has the characterisation of random processes driving *events* in the overall state space A. To reinforce the analogy, random processes play a dual role to a state space of events that a geometric or differentiable structure plays as a means of implementation of structure on a manifold.

Events in the case of sieves for numbers might be "a is a square", or "a is the sum of two primes", where A is "the natural numbers up to and including N". I will proceed to give a brief sketch of the main sieves that have been studied or are the subject of current investigation, the underlying random processes, and discuss their general properties.

The sieve of erastothenes, the prototypical example of a sieving technique, could be thought of as the method associated to a generalised Markov process, and the associated 0^*, or $(\infty, 0)$-category on the real line. We can furthermore consider the "analytic extension" of the reals to the complex plane, and consequently we see that the "generators" of the Erastothenes sieve, when instead applied to consider tuples of natural numbers such as $(a_1, a_2, ...)$, and the event of interest modified to "$a_1, a_2, ...$ is a sequence of primes", then the associated Kähler structure taken, that these behave like Chern classes. In the case that the analytic extension is *not* taken, which in a sense is more generic, we arrive at the analogue of Stiefel-Whitney classes driving the underlying random process.

4.2.3 Brun's contribution

There is a modification of the original technique due to Brun, and later extended by Alte Selberg, who observed in the 1920s that one can use the so-called inclusion-exclusion principle into developing techniques to investigate more complicated events than merely "a is a prime".

Theorem 4.2.1. *(Inclusion-Exclusion Principle). Let $A_1, \cdots, A_n$ be finite sets. Then*

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i| - \sum_{i,j\,:\,1\le i<j\le n} |A_i \cap A_j| + \sum_{i,j,k\,:\,1\le i<j<k\le n} |A_i \cap A_j \cap A_k| - \ \cdots \ + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|$$

Proof. This is not too difficult to verify. Essentially one wishes to "correct" for overcounting and undercounting via an iterative procedure. Since the number of sets is finite this process will terminate. □

Remark. This is essentially what happens when sieving out a set of particular prime multiples, as with the Sieve of Erastothenes. Legendre observed that one can actually iteratively correct in the above fashion to estimate the number of elements that remain, which provided one of the first methods for estimating the number of primes less than a given finite property. The Möbius function moreover was inspired by the Inclusion-Exclusion principle as a book-keeping device, since one "alternates" in sign depending on the order of the correction.

Brun's sieve essentially deals with the idea of viewing the naturals via particular congruence classes; so instead of working over the integers, or more broadly, the reals as a field, one views the naturals via particular local rings F_p where p is prime.

The idea is to consider H as the set of naturals less than or equal to N, and take as index set P to be a set of primes. Then construct H_p to be the subset of H divisible by p, and H_d to be the intersection of the H_p for all primes p that divide d. Suppose $P(z)$ is the number of primes in P less than a number z. Then one is interested in estimating the size of the set

$$H - \cup_{p \in P(z)} H_p$$

If the sizes of the H_p can be estimated, then it is indeed possible to estimate the order of this quantity. I will not discuss this here - [Co] provides a thorough treatment. Rather I will discuss how this construction relates to the other associated structures in mathematics.

The key observation is that if P is the set of *all* primes less than or equal to N, then the above methods allow one to estimate the number of primes less than N, and in fact one recovers the observation that the size is $O(log(N))$. This is not new; what is new is that P does not need to be all the primes, but can rather be a special subset. Consequently it is possible to prove many additional results, through estimating the numbers of states less than N than satisfy particular generalised events.

Essentially the underlying philosophy is to consider a set of primes $P := \{p_1, ..., p_k\}$ and study the field $F_P := (F_{p_1}, ..., F_{p_k})$. Consequently one is really studying finite congruences over the entire set simultaneously, and the associated characteristic classes are those associated to random processes over k simultaneously running finite state machines.

The sieve of Rosser-Iwaniec [Iwan] is a further development of this approach but allows for sharper results, through the idea of "weighting" the sieve. The idea here is to define a function $a : P \rightarrow R$ that assigns an importance to each particular element p of the index set P. Through control or consideration of certain classes of such functions it is possible to establish additional consequences using this general technique.

4.2.4 The function field case; Turan's Sieve

It is possible using sieve methods to estimate the number of polynomials of particular rank, subject to bounds on their coefficients, over the integers, ie the number of elements of particular subsets of $Z[x]$. See, for instance, [Co], pp 51-53. There is a sieve which allows one to establish such estimates. This is known as Turan's sieve.

For a marvelous and comprehensive treatment of the theory underlying this, I greatly recommend the thesis by Yu-Ru Liu, [Liu]. This will be my primary source in this instance.

Turan's sieve is of particular interest since the underlying random process lies essentially in the $(\infty, 1)$, or 1^*-category. Consequently the generators correspond to particular classes of extraordinary cohomology theories, and hence are related to some of the fairly deep structures underlying Galois theory.

Suppose H is a finite set and I an index set. Then, we associated subsets H_i of H by writing $H_i := \{m \in H | m \text{ satisfies } \Omega(i)\}$. In the usual case, we have a map $I \rightarrow P$ from the index set to a set of primes, and condition $\Omega(i)$ for the ith prime $p(i)$ would be "m is congruent to 0 mod $p(i)$".

The rest of the procedure is fairly standard, since it essentially follows as an application of the inclusion-exclusion principle for the set H and subsets H_i.

To see why exactly this is a significant and useful abstraction, note that we are not working directly from a set of primes as the objects that generate our subsets H_i, but rather an indexing set. Consequently, the conditions $\Omega(i)$ can be more general.

For instance, we could specify H to be the set of polynomials in $Z[x]$ of degree less than N and with coefficients less than M. Hence it will be a set of some size a function of N and M, which we might like to estimate. We could have a map $I \rightarrow P$ such that the indexing set mapped to a subset of all possible irreducible polynomials in H. The conditions $\Omega(i)$ could then be that "polynomial f is divisible by g_i".

Why is this an improvement over specifying the set P directly? This set will still be finite. The key observation here is that it may well be finite, but certain crucial structure might be missed via the direct approach. Indeed, to appropriately specify an irreducible polynomial in $Z[x]$, it is more elegant to consider pairs of integers $i := (i_1, i_2)$ in the index set, choice of which will depend on the structure of the associated Galois extension field.

4.2.5 The Hardy-Littlewood Circle Method

I will conclude my discussion of Sieve theoretic techniques with a sketch of an important technique due to Hardy and Littlewood, known as the *circle method* ([Comp], p.346). The general idea is quite simple, and follows from the trivial observation

$$\int_0^1 e^{2i\pi nt} dt := \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise} \end{cases}$$

Consequently, if say one was interested in counting the number of solutions $\kappa(n)$ to a particular relation, such as $p + q = n$ for p, q primes, we obtain

$$\begin{aligned} \kappa(n) &= \Sigma_{p,q \leq n} \int_0^1 e^{2i\pi(p+q-n)t} dt \\ &= \int_0^1 e^{-2i\pi nt} \{\Sigma_{p \text{ prime } , p \leq n} e^{2i\pi pt}\}^2 dt \end{aligned}$$

It is in fact easier to estimate the integral on the right than the term on the left, and this general philosophy can be used to acquire weak information on numerous deep number theoretic problems.

Naturally one might be interested in asking the question as to what form of underlying structure there is behind this method. The integral of the exponential function, although simple, seems somewhat artificial, and it is of some interest to ask as to whether one can formulate the circle method in a slightly more abstract and aesthetic form.

Indeed, the circle method is really about deriving information from a compactification of the generators of a space. Note that the universal cover of S^1 is the reals, and S^1 is the space in which $e^{i\theta}$ lives. Consider then, more generally, the space R^n with a flat metric; this projects to S^n. Consequently, if we view R^n as the tangent space of a manifold M, with non-trivial metric, we see that more generally we are looking at the information associated to exponential families of generalised Markov processes.

Hence it is possible that the study of the critical properties of this information for Markov processes, or processes that are more abstract, might be useful in estimating the size of the space of solutions to particular number theoretic questions. Indeed, in the second last section of this chapter, on Fisher-Amari information theory, I discuss some of these generalities, albeit without a view to sieve theoretic applications.

4.3 Integrability

I refer here in parts to the Princeton Companion to Mathematics [Comp], in addition to my other sources on integrability - [PB], [Mik], [Das].

4.3.1 Solitons

The study of solitons was initiated by John Scott Russell, who observed in 1844 the curious phenomena of a wave, uninfluenced in shape passing via translation along a channel. This is one of the curious properties of solitons, and is characteristic of the fact that in a sense, they contain data of the category at infinity; since otherwise their shape would be subject to decay.

The motion of a wave in a shallow channel was first quantified by Korteweg and de Vries:

$$u_t + uu_x + \epsilon^2 u_{xxx} = 0$$

Here u is the "height" of the wave, as a function of spatial parameter x and temporal parameter t. ϵ is a small parameter that will be a function of the height of the channel, and of similar order.

Remark. In general, the equation could be canoically written as $u_t + auu_x + bu_{xxx}$. However, a transformation of x removes the occurence of a, and a rescaling of u removes the second coefficient. Consequently, we can ignore coefficients and simply study the equation $u_t + uu_x + u_{xxx} = 0$.

We can make the immediate observation that the solutions of such an equation are intuitively 3rd order, since we have the occurence of the third derivative for a phenomenon that is essentially unidirectional. As one might expect, solitons are phenemona associated with 3rd order models of dynamics. Riemannian geometry / Newtonian physics, for reference, is usually 1st order, if we ignore elasticity, viscosity etc. The principal focus of this treatise - the self-referential calculus - is 2nd order. So such solutions are sophisticated, and it is not surprising that they should have complex behaviours.

The connection with integrability and PDE hierarchies is established in the following manner. In particular I will demonstrate the construction of a hierarchy of iterated PDEs from the KdV equation; these follow from the fact that the KdV equation implies the existence of an infinite number of conserved quantities.

As a first observation, note that we can rewrite the equation as

$$u_t = \partial_x(\tfrac{1}{2}u^2 + u_{xx})$$

which is in the form of a continuity equation.

Suppose now that we have a quantity $Q(u)$, and that there is an overriding Hamiltonian driving the system H. Then Q is conserved if, for a solution u and $v := u_x$, where $K(u, u_x, t)$ is the density of H, we have that Q is constant along u and u_x, or

$$\begin{aligned} 0 = \frac{dQ}{dt} &= \frac{\partial Q}{\partial u}\frac{du}{dt} + \frac{\partial Q}{\partial u_x}\frac{du_x}{dt} \\ &= -\frac{\partial Q}{\partial u}\frac{dH}{du_x} + \frac{\partial Q}{\partial u_x}\frac{dH}{du} \\ &= \{Q, H\} \end{aligned}$$

where $\{Q, H\} := \Sigma_{i=1}^{N}\{\frac{\partial Q}{\partial u_{x,i}}\frac{\partial H}{\partial u_i} - \frac{\partial Q}{\partial u_i}\frac{\partial H}{\partial u_{x,i}}\}$ is the Poisson bracket.

Consequently if $Q(u) = \int \rho(u(x,t))dx$, we have the existence of a continuity equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial j(u)}{\partial x} = 0$$

for some function j.

Then, returning to the above, we see that we can identify $\rho_0(u) := u$, and $j_0(u) := -(\frac{1}{2}u^2 + u_{xx})$. So $\int u(x,t)dx$ is a conserved quantity.

Returning to the rescaled KdV equation

$$u_t = uu_x + u_{xxx}$$

write

$$uu_t = u^2u_x + uu_{xxx}$$

Then with very little work we see that it follows that

$$\partial_t(\tfrac{1}{2}u^2) = \partial_x(\tfrac{1}{3}u^3 - \tfrac{1}{2}(u_x)^2 + uu_{xx})$$

Then setting $\rho_1(u) := \frac{1}{2}u^2$, $j_1(u) := \frac{1}{3}u^3 - \frac{1}{2}(u_x)^2 + uu_{xx}$ we obtain another conservation equation for $\frac{1}{2}u^2$, ie

$$Q_1 := \int \rho_1(u)dx$$

is a constant of motion.

Furthermore, the Hamiltonian for the KdV equation

$$H = Q_2 = \int(\tfrac{1}{3!}u^3 - \tfrac{1}{2}(u_x)^2)dx$$

is also a constant of motion, since trivially

$$\tfrac{dH}{dt} = \{H, H\} = 0$$

It is possible to generate more conserved quantities than these. In fact, the total number possible is countably infinite. The technique to arrive at these systematically is through a technique known as the Miura transformation.

4.3.2 The Miura Transformation and Infinite Hierarchies

Consider the Riccati transformation $u(x,t) =: v^2(x,t) + i\sqrt(6)v_x$. Then this alters the KdV equation to the form

$$v_t = v^2v_x + v_{xxx}$$

This is referred to as the modified KdV, or MKdV equation. It does not have the same symmetries as KdV due to the nonlinearity of the transformation, but it is nonetheless possible to transform between the two in a fairly straightforward fashion, so this largely does not matter.

Furthermore, the MKdV equation has nice properties. In particular, we have "raising and lowering operators", or "morphisms between different levels of the categorical hierarchy", given by linear transformations of v and taking limits. If we map $v \mapsto \frac{\epsilon}{\sqrt{6}}v + \frac{\sqrt{6}}{2\epsilon}$, and then let ϵ limit to zero, then the MKdV equation becomes

$$v_t = vv_x + v_{xxx}$$

We then have an inverse transformation or morphism $v \mapsto \frac{\epsilon}{\sqrt{6}}v$, which when applied to the new equation, and taking the limit $\epsilon \to \infty$, transforms it back to the former

$$v_t = v^2 v_x + v_{xxx}$$

In a sense, this could be viewed as "perturbative analysis about limits".

Suppose now that v is a solution of the generalised equation

$$v_t = (\frac{\epsilon^2}{6}v^2 + v)v_x + v_{xxx} = \partial_x(\frac{\epsilon^2}{18}v^3 + \frac{1}{2}v^2 + v_{xx})$$

Then it gives a solution to KdV via the transformation $u(x,t) = \frac{\epsilon^2}{6}v^2 + v + i\epsilon v_x$.

The first observation to make is that, for similar reasons to before, we observe readily that $v(x,t)$ is the density of a conserved quantity Q_0.

Now expand $v(x,t)$ as a power series in ϵ

$$v(x,t) := \Sigma_{n \in N} \epsilon^n w_n(u(x,t))$$

Then each w_n must be a conserved quantity since each power of ϵ needs to satisfy a continuity equation. However it remains to show that the expansion of v in terms of powers of ϵ is not trivial, since otherwise we would not necessarily have infinitely many conserved quantities.

Suffice to say via careful argument it is possible to do this. The general idea is to consider once more the transformation

$$u = v + i\epsilon v_x + \frac{\epsilon^2}{6} v^2$$

and demonstrate that inversion of this relation, to get v as a function of u, would lead to pure polynomial terms in u; in particular, such terms cannot be written as total derivatives, and hence will not be trivial conserved quantities (pinching).

It is also possible to construct the quantities w_n iteratively, and consequently arrived at the conserved quantities for the KdV equation. However such a treatment is not terribly instructive, and I shall leave my discussion of solitons here.

Remark. In a way, one could view the study of the MKdV equation as really being in the dual 0-category, or 0^*-category (otherwise known as $(\infty, 0)$), and the transformation from v to u as passage to the dual of the dual, the 0^{**}-category. More on this later.

4.3.3 Instantons

Instantons were used by Donaldson [Don] in his celebrated result which established the uncountability of the total number of unique differentiable structures allowed on R^4. They can essentially be viewed as the "1-limit categorical extrapolation of solitons", where it is to be understood, naturally, that we are considering partial structure of the category at infinity.

In this section I will argue that these slightly more exotic beasts are related to the Toda equation, and the associated Toda hierarchy, [HP]. Instantons are also related to the self-dual Yang Mills equations [Gt1], [Gt2].

The $sl(\infty)$ formulation of the Toda equation in coordinates (x_i, t) is

$$\partial\bar{\partial}u + \partial_t^2 e^u = 0$$

Here $\bar{\partial}$ is the gradient operator rel the dual connection. It is defined to operate in terms of directional derivatives as $\bar{\nabla}_X u = \nabla_{\star X} u$, where $\star$ is the *Hodge star* operator. This essentially locally sends forms to their dual. ie if the ambient space is R^4, with coordinates x_i, then $\star(dx_1 \wedge dx_2) = dx_3 \wedge dx_4$.

According to [HP], the $sl(\infty)$ Toda equation is related to the self-dual Einstein equation with an appropriate choice of metric.

Returning to a rather standard treatment of the Toda Equation, the starting point is usually with the Toda Lattice - designed to describe the motion of N point masses on a line under the influence of an exponential interaction. Suppose these have coordinates u^i and momenta $v^i := u^i_x$. Then the Toda equations can be written as

$$u^0 = \infty$$
$$u^{n+1} = \infty$$
$$u^i_t = v^i$$
$$v^i_t = exp(-(u^i - u^{i-1})) - exp(-(u^{i+1} - u^i))$$

More abstractly one can view this as being a higher order process, since it involves the specification of a finite number of distinct points. Hence it has similarities with the travelling salesman problem, which can be attacked via the uses of the 1-categorical approach, using 8-tensors, as described in the second last chapter of this work.

Essentially it is possible to show that there are N conserved quantities for the Toda system. If N is set off to infinity, then, one gets an infinite number of conserved quantities, and I claim that there is an underlying process for this, the *instanton equation*. This is usually understood to be the self-dual Yang Mills equation. However, because I prefer a degree of simplicity and elegance to my studies, it is of some interest to try to understand how the KdV equation extends in this instance.

So consider the KdV equation, written in conservation form:

$$u_t = \partial_x(\tfrac{1}{2}u^2 + u_{xx})$$

Now abstract from u a position in M to a matrix U of positions in M determined by pairs of points in the ambient space $M \times R$. Abstract from coordinate x to coordinate vectors x and y, and from temporal variable t to temporal variables t and s. Then consider

$$\partial_t \partial_s U = \partial_x \partial_y(\tfrac{1}{2}U^2 + \Delta_x \Delta_y U)$$

The solutions $U : M^2 \times R^2 \to M$ to this equation have interpretation as instantons, since they essentially are self-referential solitons. To see why this is the case, note that $U_{ij}(x, t; y, s)$ can be seen as essentially a measure of the information of the connection between the ith and jth nodes of the underlying Toda lattice process linking the information of coordinates (x, t) and (y, s).

Just as with the KdV equation, I claim that it is possible to transform this equation to the general form

$$\partial_t \partial_s V = \partial_x \partial_y (\eta V^4 + \Delta_x \Delta_y V)$$

via the transformation $U =: V^2 + \zeta \Delta_x \Delta_y V$, for appropriate choice of constants η,ζ. Then via appropriately defined linear transformations of V one can build raising and lowering operators that act in the 1^*, or $(\infty, 1)$-category. These can be used then to construct an infinite hierarchy of invariants via power series expansion that are conserved via the original instanton equation. I believe this is consistent with the thesis of [HP], though my approach is probably slightly more direct.

4.3.4 Painleve Tests

Here my primary source is [Mik], chapter 7.

Recall in [Go] it was demonstrated that smooth functionals (that is, functionals defined over a smooth space, with C^∞ integrand) would always admit solutions to the associated PDE via requiring the first variation vanish. This indeed is the question that geometric measure theory was built to solve. However, the related question, as to whether all PDEs admit smooth solutions, is well known to have a negative answer. Not all PDEs over smooth domains with smooth coefficients admit smooth solutions.

In a similar fashion, it is of interest to know as to whether a particular nonlinear PDE is integrable - ie whether it admits solutions that conserve (possibly) a countably infinite number of quantities.[2] There are various ways of determining this, for various forms of PDE. One of the classes of better known techniques to this end are the so-called Painleve tests.

These involve analysis of the singularity structure of the general solution to a particular PDE. In the case that the differential equation in question is in one variable, it turns out that the following criterion is sufficient to establish integrability:

Definition 85. (Painleve property for ODE). An ODE has the *Painleve property* if all movable singularities- that is, singularities whose location is determined by initial data- are poles.

One can then proceed according to the following procedure in order to determine whether an ODE for a solution y in terms of variable x has the Painleve property.

[2]Note that this will differ slightly from the problem of GMT in that solutions may, and in fact will generically have singularities, and hence will not be smooth. The idea is rather to control the singularity structure to sufficient extent, so that the solutions are "well-behaved".

(i) Determine all singularities of the form $y \sim a(x-b)^\mu$.

(ii) If all exponents in step (i) are integers, find the resonances by perturbing around the singularities, plugging back into the original ODE and solving for the exponent of the expansion (the resonance). Eg by writing $y \sim (x-b)^\mu(1+\epsilon t(x-b)^r)$, ϵ small, t, b constants, and r the resonance associated to μ. Note that even for ODEs there may be more than one resonance associated to each singularity.

(iii) If all resonances are integers, check the resonance conditions, ie that the expansions in step (ii) can be matched to the initial data set.

If (i) to (iii) are satisfied, then the ODE in question has passed the associated Painleve test. This however does not guarantee that the ODE has the Painleve property; for this one needs to establish that all local expansions about singularities can be analytically continued to a global single-valued function. So the Painleve test is a necessary, but not sufficient condition for an ODE to be integrable.

4.3.5 Discussion

The analysis of the partial structure of limit categories has interesting consequences. It is of particular interest to examine such objects since, as opposed to the main treatment of this treatise - the development of a self-referential calculus, based on the theory of 1-categories - there is no natural limit to the speed of propagation of signals; since one is viewing the category at infinity, propagation of information is *instantaneous*. In other words, it might be possible to make progress on one of the greater intangibles of modern physics following this theoretical direction, namely, the construction and manipulation of "wormholes" [Wh].

However I am uncertain as to whether the study of solitons or instantons would really be enough to allow one to develop a knowledge sharp enough to enable physical objects, the bulk of whose information is grounded within the 0th and 1st information levels - such as people - to maintain integrity through wormhole passage. It is possible that one might need to build a theory based on the foundations of 2-limit categories, or $(\infty, 2)$ categories, in order to proceed in such a fashion. Additionally, it is of interest to ask as to the resistance to noise or distortion effects of information traversing such a construct, as a function of the distance between the apertures. Intuitively one might expect the need for more sophisticated control, based on more sophisticated theory, in order to increase this distance while maintaining acceptable error aggregation via transmission.

Regardless, a full description of what would be required in this instance goes well beyond the scope of this work, and doubtless would require considerable additional thought in order to make precise. So instead, in the next section I will consequently proceed to discuss the general principles required in the simplest instances of random processes. In other words, I will provide a basic description of the Cramer-Rao inequality for a generalised Markov Process, in the spirit of the monograph due to Amari [Am1], and his slightly more recent paper [Am2]. This will be tied off with a return to the roots of modern probability theory due to Kolmogorov, and mention of a rather nontrivial inequality, also due to Kolmogorov, through which preliminary understanding can be obtained on the interplay between consecutive levels of information complexity.

4.4 Information Statistics

4.4.1 Fisher-Amari information theory

I will provide a rough picture of how the Cramer-Rao inequality for a random process is established, following Amari [Am1]. My treatment will by no means be complete; the study of critical geometric statistical structures is a huge area. Consequently, I will only give a taste, by restricting myself to an intuitive consideration of the theory for generalised Markov processes only.

By generalised Markov process, I essentially mean the analogue in statistical structure on a state space as the idea of Riemannian metric is a geometric structure on a differentiable manifold. Recall that a Markov chain is a matrix that acts on a point a in a coordinate chart of state space A as Γa. Note that we have a natural correlation measure on A induced by Γ as $\langle a, b\rangle =: b\Gamma a$. Suppose now that for every element a of A we associate a Markov chain $\Gamma(a)$. This has the interpretation of a "sharp event structure" on A such that transition probabilities change according to position in state space. It is then necessary to consider $\Gamma(a)$ as acting on the "tangent space to A at a", or TA.

The map $\Gamma : A \to TA \times TA$ I shall from now on refer to as a *generalised Markov process.*

To be more precise, the space T_aA associated to a state a, for a particular Markov process Γ, has the interpretation as the "current sample space", and a as "the previous state". Furthermore, there is an associated map $exp_a(\Gamma) : T_aA \rightarrow A$, which has the interpretation as the measurement of a new event $exp^{-1}b$ rel a, and projection of b onto the differentiable structure of the underlying state space A. This is known as the *exponential family* associated to the Markov process Γ, since the structure of the probability distribution at b for a generalised Markov process will be different in general than at a.

Then, just as for the Riemannian case, we can build a *signal function* $f^*(a, m) := \delta(\Gamma(a) - m)$, $m \in M$, where M is to be thought of as "the natural universal event space associated to the state space A" and define the Amari-Fisher information functional as

$$I(f^*) = \int_A \int_M f^*(a, m)(\partial_A ln(f^*(a, m)))^2 dmda$$

Via similar techniques to that used in the dual case, it is possible to prove the Cramer-Rao inequality:

Theorem 4.4.1. *(Cramer-Rao-Amari).* $I(f^*) \geq 0$.

Consequently $I(f^*)$ will be zero iff $\delta I(f^*) = 0$, or, in other words, if the generalised Markov process is critical.

I now make the following claims.

Claim 4.4.2. *Setting* $\delta I(f^*) = 0$ *for a sharp generalised Markov process essentially gives the KdV equation, or equivalently the KdV hierarchy, which has soliton solutions as a consequence.*

It is possible to abstract the idea of Markov process to an *auto-correlated Markov process*, which is the dual analogue for statistical structures as a self-referentiable structure is for geometric structures. (I will discuss self-referential geometry in the penultimate chapter of this dissertation.)

Claim 4.4.3. *The fisher information density for a sharp auto-correlated Markov process leads to nothing other than the Toda equation, or equivalently the Toda hierarchy, which has instanton solutions.*

Remark. In order to actually derive objects of physical consequence from this theory, it is to be noted that we are, to an extent, viewing the solutions of the random structures as physical processes. So in a sense we are examining "the dual of the dual". Note from basic analysis that this will, in general, not always be the class of spaces that one starts off with; in fact, if P is a Banach space, and P^* is the dual of P, then we have the existence of an embedding $P \subset P^{**}$. We also expect this to be the case from the point of view of foundations and completeness. Consequently, a proper treatment might require one to examine these objects from the point of view of 2-categories, or at the most general level, if we call $(\infty, 2)$-categories 2^*-categories, then we might be interested in the study of 2^{**}-categories.

Remark. This probably also explains why the required degrees of freedom for a structural treatment of 2-categories - 192, as described in the conclusions of this work - is not a nice round number. I will hazard a guess and surmise that for 2^{**}-categories one has 256 or 2^8 natural degrees of freedom, and, for the associated quantum mechanical formulations, at least either 512 or 1024 dimensions are required for invariants to be nonzero.

4.4.2 Kolmogorov's foundations

I will conclude this chapter with a rough overview of the treatment to foundations of probability theory due to Andrei Kolmogorov (1903 - 1987), following the sources [Tao3] and [SV]. This will culminate in a discussion of the Kolmogorov inequality [W9], and its connections with the ideas due to Shun'ichi Amari.

Kolmogorov is widely noted as one of the principal founders of the modern theory of probability. The key observation that he made was to recognise that there was a strong correspondence between Lebesgue's theory of measure, and the classical notions of probability - events, state space, etc. In particular one of his major contributions was to rigorously establish that, to describe the classical theory and compute probabilities, nothing more than countable additivity, and standard notions of intersection and union of sets were needed.

In particular, he developed an axiomatic treatment of probability, which I shall now quickly describe.

Definition 86. (Weighted event space). Let A be a set of states, and F a set of subsets, of A, the latter of which we will call the event space. Suppose furthermore that F satisfies the following axioms:

(i) F is a field of sets, that is, it is closed under union and intersection.

(ii) F contains A.

(iii) There is a function $P : F \to R^+$. The number $P(k)$ is called the probability of the event k.

(iv) $P(A) = 1$.

(v) If x, y have empty intersection, then $P(x \cup y) = P(x) + P(y)$.

If F has infinite cardinality, we also require that if $A_1 \supset A_2 \supset \cdots$ is a decreasing sequence of elements in F, we $\cap_i A_i = \phi$, then $lim_n P(A_n) = 0$.

Then the pair (F, P) forms a *weighted event space* over the state space A.

This essentially is the core of Kolmogorov's *Grundbegriffe*. However there is another interesting result due to the fellow which uses this as a basis, the so called *Kolmogorov inequality.* Before I discuss this, however, we will need the Chebyshev inequality, named for Pafnuty Chebyshev (1821 - 1894).

Theorem 4.4.4. *(Chebyshev Inequality). Suppose (X, Σ, μ) is a measure space, that is, X is the underlying structure of states, Σ is a σ-algebra of events, and μ is a measure that assigns the "probability-density" for elements of X. Let f be a measurable function defined on X. Then, for any number $t > 0$ (usually small)*

$$\mu(\{x \in X | \|f(x)\| \geq t\}) \leq \tfrac{1}{t^2} \int_X f^2 d\mu$$

Proof. Let $g(t)$ be a nonnegative measurable function, nondecreasing with respect to the image of f. Then note that, if $A_t := \{x \in X | f(x) \geq t\}$, and 1_{A_t} is the function from X to $\{0, 1\}$ such that it is 1 for elements of A_t, and 0 otherwise, then

$$0 \leq g(t) 1_{A_t} \leq g \circ f 1_{A_t} \leq g \circ f$$

Consequently, passing to the measure

$$g(t)\mu(A_t) = \int_X g(t) 1_{A_t} d\mu \leq \int_{A_t} g \circ f d\mu \leq \int_X g \circ f d\mu$$

which establishes that

$$\mu(A_t) \leq \tfrac{1}{g(t)} \int_X g \circ f d\mu$$

But it is clear then that the Chebyshev inequality is corollary to this, if we define $g(t) = t^2$ for positive t, and 0 otherwise. ☐

Theorem 4.4.5. *(Extended Kolmogorov Inequality). Let $X_1, ..., X_n : A \to B$ be independent random variables with mean zero and finite variance, defined over a state space A, representing the probabilities of different events. More generally, suppose X_m are independent random variables for each $m \in M$, where M is the set of all possible events associated to A. Consider the set algebra associated to unions of events over M, or subsets U of M, which I will identify with B. Let $\lambda > 0$ be a number, and $Y(a, b; m, \lambda)$ be the random variable within the event space N over state space B, associated to the event in B that $sup_{U \subset M} \int_U \int_A X(a, m) da dm \geq \lambda$. Then, for every $\lambda > 0$, we have that*

$$\int_B \int_A \int_M Y(a, b; m, \lambda) dm da db \leq \frac{\int_A \int_M Var(X(a,m)) dm da}{\lambda^2}$$

Proof. (Sketch). The key is to realise that this is essentially an extended version of the Cramer-Rao-Amari inequality, but with additional information due to inclusion of the next level in the structural hierarchy. Note that if we take the limit $\lambda \to 0$, and assuming that contributions from the second structural level are finite and bounded, we have that the left hand side becomes zero and we are left with the relation

$$\int_A \int_M Var(X(a, m)) dm da \geq 0$$

If we then observe that the Fisher information is the variance of the score, and the score itself is just another random variable, it is clear to see that the above expression reduces to the standard Cramer-Rao-Amari inequality. So certainly the extended Kolmogorov inequality is not at odds with the more primitive results.

For a rough indication as to why the statement of the theorem is true, one needs to use the so-called Martingale methods. I adapt a proof of the discrete formulation of the Kolmogorov inequality due to Kareem Amin for the purposes of this exposition.

Let $U_1 \subset U_2 \subset \cdots$ be a countable sequence of sets converging to a set V as a subset of M, the total event space. Then, since the associated random variables are independent, we have by Doob's martingale inequality that such a sequence is a martingale. That is, the conditional expected value at step $n + 1$ in the chain only depends on that at step n.

Suppose we include $U_0 = \phi$, the empty set. Then any random variable associated to this set will be trivially zero. Now define $S_k = \int_{U_k} X(a, m) dm$. This will be a random variable associated to U_k for each k. From this, define a new sequence of random variables Z with $Z_0 = 0$, and

$$Z_{n+1} := \begin{cases} S_{n+1} & \text{if } max_{1\leq i\leq n} S_i < \lambda \ , \\ Z_n & \text{otherwise} \end{cases}$$

Then this too will be a martingale.

For the purposes of this analysis we will be interested in the set U, as the set for which $\int_A \int_U X(a,m)dmda$ is maximal in M. In particular we will construct a min-max argument, such that we will examine

$$inf_{V\supset U} sup_{U_i(V)} S_i$$

The idea will be that this will play the role of the left hand side in the theorem of question.

Now observe that $S_{n+1} - S_n$ are random variables that are independent with mean zero. Consequently it is not too difficult to demonstrate that

$$\Sigma_{i=1}^{n} E[(S_i - S_{i-1})^2] = \Sigma_{i=1}^{n+1} \int_A (S_i(a) - S_{i-1}(a))^2 da = \int_A S_n(a)^2 da = E(S_n^2)$$

and that the same relation consequently follows for the Z_i. Hence if $\int_B S(a,b;i,\lambda)db = Pr(sup_i S_i(a) \geq \lambda)$ is the expectation of the random variable defined on the meta-event space, we have that

$$\begin{aligned}
\int_B S(a,b;i,\lambda)db &= Pr(Z_n \geq \lambda) \\
&\leq \frac{1}{\lambda^2} E(Z_n^2) \text{ by Chebyshev's inequality} \\
&= \frac{1}{\lambda^2} \Sigma_{i=1}^{n} E[(Z_i - Z_{i-1})^2] \\
&\leq \frac{1}{\lambda^2} \Sigma_{i=1}^{n} E[(S_i - S_{i-1})^2] \\
&= \frac{1}{\lambda^2} E(S_n^2) = \frac{1}{\lambda^2} Var(S_n)
\end{aligned}$$

This is true for any chain $U_1 \subset U_2 \subset \cdots \subset V = U_\infty$. Consequently via a min-max construction as sketched above this applies for the set U in M that realises the maximal value of the quantity $\int_A \int_U X(a,m)dmda$. This completes the argument. □

Remark. The Kolmogorov inequality is interesting because it essentially provides a means of establishing a bound on the influence of higher order effects on a space subject to particular probabilistic assumptions. In particular, λ could be viewed as an "expansion parameter". Naturally if $\lambda = 0$ and the growth of 2nd order effects is not quicker than $\frac{1}{\lambda^2}$ for small λ, we can neglect such effects. This is one fascinating consequence of this analysis.

Chapter 5

Turbulent geometry

A recurring goal of this treatise will be to build and develop geometric structures corresponding to the resolutions of sequences of countably infinite geometric bifurcations.

Recall from my previous work on turbulent geometry that there are two types of turbulence that admit straightforward modelling - turbulence in the measure and turbulence in the stack. These results in actions of the form R^R and R_R respectively, where I am taking considerable liberties paraphrasing here. Similarly we can consider measure-measure turbulence, measure-stack turbulence, stack-measure turbulence, and stack-stack turbulence. These result in actions of the form R^{R^R}, R^{R_R}, R_{R^R}, and R_{R_R} respectively.

It is natural to then ask what happens in general. Well certainly at the nth iteration we will have 2^n different possibilities for an action. So complexity of our models if we wish to encompass all possibility grows exponentially with further attention to detail. This is evidently not desirable. In particular we would ideally like to know what happens if we push n off to infinity, to generate an infinite number of discrete geometric bifurcations in our models. Then it is readily seen that the number of possibilities is $2^{\aleph_0}$, or $\aleph_1$, the cardinality of the real numbers (I am taking a further liberty here - for those who wish to believe in intermediate infinities I adopt the convention that they might take a continuous range of values $\aleph_k$ where k is between 0 and 1).

So this leads one to ask, is there some formalism that would allow us to deal with an infinite number of discrete bifurcations? Needless to say, if this is to be doable, some novel new idea or way at looking at things is essential to make progress so that calculations do not become unmanageable. To cut things short, I believe that the answer is yes, and this is where the idea of geometric exponentiation enters the picture.

Set exponentiation is a fairly simple concept to understand. Consider two sets A and B. The product $A \times B$ may readily be formed, and is understood to be the set of tuples (a, b) where $a \in A$ and $b \in B$. So this is the product of two sets. How about raising one set to the power of another? What exactly should this mean? Obviously we expect the cardinality of A^B to be necessarily of quite a different order to that of $A \times B$. So this needs to follow from the definition.

Briefly, the exponential of A^B will be understood to mean a product $\times_{b \in B} A(b)$, where each $A(b)$ is a copy of A indexed by an element b of B. So if B is large, say infinite, as will often be the case, A^B will be very large indeed. In what is to follow, A and B will often be Riemannian manifolds. Then A^B will be a manifold itself of *uncountably infinite dimension*. However more information is required to specify and map out an appropriate amount of geometric structure for such spaces, and this leads directly to the notion of geometric exponention, which will be the focus of the next section.

5.1 Initial definitions

In revisiting this particular research focus it is necessary to remind ourselves of the foundations. Recall that, given two signal function f, g defined over a Riemann-Cartan manifold, the turbulent derivative acts in the following manner:

$$\partial^*(f(m,a); g(m,b)) = \int_A f(m,c) \partial_c^{g(m,b)} f(m,a) dc$$

It is then possible to develop an information theory for this, with associated information (for a sharp geometry) of the form

$$\int_M \partial^*(R_\sigma; R_\tau) = \int_M \partial_\sigma^{R_\tau} R_\sigma$$

Furthermore, it turns out that generalised notions of composition operators are associated with turbulent structures. For instance, consider the problem of finding an f such that $f \circ f = e^x$. Then I claim that $f = e^{\circ(x;1/2)}$.

In particular, note that we have in general the following:

Proposition 11. *Consider the case where f may not be an automorphism, but rather a signal function. Then we have that the composition operator $\circ$ naturally extends; in particular*

$$\circ(f;g) = exp(\partial^*(ln(f);g))$$

for another signal function g.

So to check in the case of our example, we write $f = exp(\partial^*(ln(e^x);1/2)) = exp(\partial^*(x;1/2))$. Then $f\circ f = exp(\partial^* exp(\partial^*(x;1/2));1/2) = exp(exp(\partial^*(x;1/2))\partial^*(x;1)) = exp(exp(\partial^*(x;1/2)))$. But it is easy to see that this is just e^x.

5.2 Tensor formulation of turbulent invariants

Let $x, y, p, q \in TM$ belong to the tangent space of a manifold M.

Then define $\langle (x,p),(y,q)\rangle_\Lambda = \Lambda_{ijkl}x_i p_k y_j q_l$. Then in this case Λ is a 4-tensor on M. This induces what I will call a turbulent structure if we associate a particular affine connection to it.

In Riemannian geometry, the scalar curvature can be written as

$$R_\sigma = -\sigma^{ij}\Gamma^\beta_{i\alpha}\Gamma^\alpha_{j\beta}$$

Here $\Gamma^k_{ij} := \langle \nabla_i X_j, X_k\rangle_\sigma$, where X_i are a local basis of TM are the Christoffel symbols for σ. These may be computed in terms of σ and its derivatives as so:

$$\Gamma^i_{kl} = \tfrac{1}{2}\sigma^{im}\left(\tfrac{\partial\sigma_{mk}}{\partial x^l} + \tfrac{\partial\sigma_{ml}}{\partial x^k} - \tfrac{\partial\sigma_{kl}}{\partial x^m}\right)$$

This may be generalised to turbulent geometry:

$$S_{\partial^*;\Lambda} = -\Lambda^{ijkl}\Gamma^{\gamma\delta}_{ik\alpha\beta}\Gamma^{\alpha\beta}_{jl\gamma\delta}$$

where $\Gamma^{\gamma\delta}_{ik\alpha\beta} = \langle\langle(\nabla_{X_i},\nabla_{Y_k})(X_\alpha,Y_\beta),(X_\gamma,Y_\delta)\rangle\rangle_\Lambda$ are the generalised Christoffel symbols.

We also require ∇ to be an affine connection such that

$$Y_i X_j \langle\langle (X_k, Y_l), (X_m, Y_n) \rangle\rangle =$$
$$\langle\langle (\nabla_{X_j} X_k, \nabla_{Y_j} Y_l), (X_m, Y_n) \rangle\rangle + \langle\langle (X_k, Y_l), (\nabla_{X_j} X_m, \nabla_{Y_j} Y_n) \rangle\rangle$$

It is essential that we choose this class of connections in order for ∇ to be uniquely defined for a given Λ. This follows from a suitable generalisation of the original fundamental theorem of Riemannian geometry.

The generalised Christoffel symbols can be computed wholly in terms of Λ and its derivatives, as so:

$$\Gamma^{i\alpha}_{k\gamma l\epsilon} = \frac{1}{2}\Lambda_{im\alpha\zeta}\left(\frac{\partial^2 \Lambda_{mk\zeta\gamma}}{\partial X^l \partial Y^\epsilon} + \frac{\partial^2 \Lambda_{ml\zeta\epsilon}}{\partial X^k \partial Y^\gamma} - \frac{\partial^2 \Lambda_{kl\gamma\epsilon}}{\partial X^m \partial Y^\zeta}\right)$$

So, to summarise - we have a generalisation of scalar curvature to turbulent structures, $S_{\partial^*;\Lambda}$, which can be written entirely in terms of Λ and its second order derivatives. For simplicity I will assume from now on that $M = N$ are the same manifold - certainly it would not really make much physical sense in this context if they were not the same space.

This scalar curvature forms a natural information density on a differentiable manifold.

5.3 Topics in geometric turbulence

5.3.1 Fluid dynamics

Perhaps one of the clearest applications of geometric turbulence is in the study of fluid dynamics. Recall from [Go], page 215, that the physical information of a sharp Riemannian structure σ on a differentiable manifold is

$$\int_M (R_\sigma - \|\phi\|^2 + \|\partial f\|^2)$$

where $\phi =: \rho v$ has the interpretation as the flow of matter, for ρ the density and v the velocity, and f has the interpretation of the frequency of electromagnetic radiation.

In the case that is usually studied for terrestial fluids, the curvature is of negligible order, so this expression simplifies to

$$\int_M (\|\partial f\|^2 - \|\phi\|^2)$$

We may make the further simplification that magnetic effects are also negligible. Consequently we obtain the underlying hamiltonian

$$\int_M \|\rho v\|^2$$

We then obtain the Navier-Stokes equations via requirement that this be critical:

$$\nabla\|\rho v\|^2 = 0$$

The aim will now be to obtain the analogous formulation of Navier-Stokes, assuming that we are not working with an underlying Riemannian geometry, but an underlying turbulent geometry. I will not give a full and careful derivation here, due to time constraints, but will indicate the rough idea of the argument.

Certainly we expect

$$\int_M (S_{\partial^*;\Lambda} - \|\partial^*(\phi,\psi)\|^2 + \|\partial^*(f;g)\|^2)$$

to be the relevant action, where Λ is a symmetric turbulent structure. The interesting part here are $\phi =: \rho v$ and $\psi =: \tau w$, which both have units as flow of matter in the space. ϕ is unimpeded flow in the absence of turbulence. ψ is a slightly more exotic beast, and has the interpretation as the local reaction of the flow of the fluid to movement via ϕ. f and g have the same form of relationship, in terms of self-reaction of radiation in the space.

Assuming that curvature and magnetic effects are negligible again, and once more requiring that the remaining information be critical, we obtain the Navier-Stokes equations subject to geometric turbulence:

$$\nabla\|\partial^*(\rho v;\tau w)\|^2 = 0$$

5.3.2 Mostow Rigidity

A slightly more surprising application of the theory is towards an enhanced understanding of the classic rigidity theorems.

Recall the statement of Mostow Rigidity:

Theorem 5.3.1. *(Mostow, 1968, [Mo]). Let M and N be complete finite volume hyperbolic n-manifolds - that is, manifolds with constant sectional curvature -1 - with $n > 2$. If there exists an isomorphism $f : \pi_1(M) \to \pi_1(N)$ then it is induced by a unique isometry from M to N.*

Essentially what this is saying is that if M and N have the same homogenous information density everywhere, then, subject to the condition that their topology be the same - as characterised by the isomorphism of their fundamental groups - then they are essentially the same object. So this seems relatively clear.

Nonetheless we would like to perhaps get a more general understanding of why it suffices to examine the first fundamental group of the spaces. In particular I will sketch how the fundamental group over a Riemannian manifold induces a turbulent geometric structure, and consequently demonstrate how the notion of fundamental group can be geometrised, and the Mostow rigidity theorem extended.

Recall the definition of fundamental group:

Definition 87. (Fundamental group). The fundamental group $\pi_1(M, x)$ defined at a point x in a topological space M is constructed as

$$\pi_1(M, x) := \{[\gamma] | \gamma : I \to M, \gamma(0) = \gamma(1) = x\}$$

where $[\gamma]$ is the homotopy class associated to the map γ.

Note that if M is connected, $\pi_1(M, x)$ is independent of choice of the base point x.

I claim that the non-local nature of this construction corresponds to a turbulent structure. To be more precise, suppose $\sigma : TM \times TM \to R$ is a metric on M. Then we can extend σ to a map $TM \times TM \to R^n$, via consideration of a vector of copies of the original metric $\sigma_1, ..., \sigma_n$. Define a metric structure on R^n, τ, and abstract R^n to a local chart of the original manifold M.

Then we have a 4-tensor $\Lambda : TM^4 \to R$. Furthermore $\vec{\sigma}$ defines a geometrised space of paths on M, and τ is a geometry on this space of paths. Consequently $\Lambda := \tau \otimes \sigma$ forms a turbulent structure on M, which geometrises the path space. (In the case that τ is again the original metric σ we recover the fundamental group; in this case I will write $\Lambda = \Lambda(\sigma)$.)

Hence we can characterise the information of $\pi_1(M)$ (supposing that M is connected) at each point $m \in M$ by computing $S_{\partial^*;\Lambda}(m)$. This observation leads us to guess the following result:

Conjecture. *(Sharpened Mostow Rigidity). Let* (M, σ_1),(N, σ_2) *be Riemannian manifolds with constant and equal information densities. Then if there is a bijection* $f : M \to N$ *such that* $S_{\partial^*;\Lambda(\sigma_1)}(m) = S_{\partial^*;\Lambda(\sigma_2)}(f(m))$ *for all* $m \in M$*,* f *is a Riemannian isometry between* M *and* N*.*

Chapter 6

Viscosity, Plasticity, and Schemes

The question I ask in this chapter is, is it possible, given any differentiable manifold, to endow it with a self-multiplication? To be precise, is there a way to sensibly associate with the manifold M a map $M \times M \to M$ that behaves like a multiplication?

It turns out that this is closely related to a different question, namely, can we generalise the notion of a tensor?

The broader question - of somehow giving a manifold an algebraic structure, or establishing a correspondence between algebra and geometry, is nothing new, and has been extensively explored in the literature. The most notable instance of this is the noncommutative geometry of Alain Connes, [Connes], towards which a more elementary and modern treatment is given in the book by Masoud Khalkhali [Kha]. Such considerations in turn were motivated by work due to I. Gelfand and M. Naimark [GN].

6.1 Motivation

6.1.1 "Rank" manifolds and generalised tensors

Recall that a metric is a tensor of rank two. An elastic structure consists of a tensor of rank four. It is natural then to ask, can we have a tensor of rank three? How about three and a half? Or even $(2, 3)$? In particular we are interested in somehow generalising the notion of rank of a tensor to somehow giving it a geometric structure.

Another way of considering a tensor of rank two is in the following way:

$$\sigma : M \times \mathcal{N} \times \mathcal{N} \to R$$

where $\mathcal{N}$ in this case is the set of natural numbers.

But there is no reason why we should not instead consider

$$\sigma : M \times (N, \tau) \to R$$

where now N is another copy of the underlying differentiable structure with metric τ. This should from now on be the definition I use for a hypermetric structure (σ, τ) on a manifold M.

In fact, we can represent this information as a signal function by writing

$$f(m, a, b) = \delta(\sigma(\delta(\tau(n) - a), m) - b)$$

where of course we have to take some care, since now σ is not purely a function any more via abuse of notation, but takes a functional as one of its inputs.

6.1.2 A more pedestrian approach?

Note that if we were to evaluate the information of the above we would get

$$R(R_\tau(n), m)_\sigma$$

Naively one might consider this a special case of considering merely the curvature of a space $M \times N$,

$$R_\sigma(m, n)$$

where now σ is a metric on $M \times N$, which is to have the interpretation of the component on the space N being a generalised rank for a standard metric structure on M.

However this fails to take into the account that the topmost expression is a higher order information, since it is of fourth order, whereas the lower is only second order. It is also more natural to consider the topmost expression since then $R_\tau(n)$ has the interpretation of the local *rank* of σ at m, so that this would be a sharp rank space, where to each point m a unique rank for σ was defined so as to vary smoothly with variation in the base point. And this is more or less what we are looking for.

So then $\sigma : M \times R \to (TM \times TM \to R)$, and $\tau : N \to (TN \times TN \to R)$.

We still have the issue of interpreting how to define a rank r tensor at a point m in our base space, and the associated generalised scalar curvature.

If r is a natural number, then we might consider an action on the tangent space of the following general form:

$$\sigma : TM \times ... \times TM \to R$$

where we have r copies of the tangent space TM. However then the associated curvature operator will not be of second order as we would like. So instead we consider

$$\sigma : TM^{r/2} \times TM^{r/2} \to R$$

where $TM^{r/2}$ is the composition of TM with itself $r/2$ times.

Then $R_\sigma(v, w) := R_{\bar{\sigma}}(v^{\circ r/2}, w^{\circ r/2})$ where $\bar{\sigma}$ is the associated inner product on the tangent space with itself, and $\circ$ is the generalised composition operator. The scalar curvature R_σ can then be defined in the usual way.

This definition allows us to generalise to the case where r is a real number, or even an element of R^n, and, following the treatment on viscous geometry as before, we may define an inner product on the composition space, which allows one in turn to define a meta-information on a new manifold N.

I then claim that the evaluation of the associated information will give the result mentioned at the beginning of this subsection.

Alternatively, consider the geometric inner product

$$\langle (v, \alpha), (w, \beta) \rangle_{\sigma,\tau} := \circ(\sigma_{ij}; \tau_{kl}) v_i w_j \alpha_k \beta_l$$

where v, w are tangent vectors in M and α, β are tangent vectors in N. This allows us to interpret $R_\sigma(m, R_\tau(n))$ as

$$R(\sigma(m)^{R(\tau(n))})$$

Furthermore we can play the same trick as in plastic geometry and define a 4-tensor Λ such that $\langle (v, \alpha), (w, \beta) \rangle_\Lambda := \Lambda_{ijkl}(m, n) v_i \alpha_k w_j \beta_l$.

Then I claim the above notation simplifies to

$$Q_\Lambda(m, n)$$

for some appropriate generalised curvature operator Q.

6.1.3 Viscosity, plasticity, and statistical structures

It turns out that all of the above can be expressed fairly simply in terms of statistical structures. In particular, these are special examples of $\star$-structures, where $\star(f;g) := f(m,a)g(m,b)$. However, of course in the above we are considering multiplication *in the stack*, so we have $exp(\star(ln(f);g)) = f(g(m,b)m,a)$ as the structural objects of interest.

In particular I refer to such structures as *viscous* since they are really about two statistical structures "rubbing against" one other. For slightly more direct physical intuition, $\star(f;g)$ provides geometric information to us about the resistance to flow locally in, say, a fluid of interest.

There however is also a connection to plasticity theory.

In particular recall the basics of elasticity theory:

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl}$$

where

(i) σ is the *stress tensor*, that is, where $\sigma_{ij}A_j$ is a force per unit area exerted on an infinitesimal component of the space given a vector A normal to it.

(ii) ϵ is the *strain tensor*, that is, the symmetrised gradient of displacement u from an undeformed state (where $\sigma = 0$).

(iii) C is the *elasticity tensor*. It is a four tensor representing the relation between the strain and the stress.

The above equation is approximately valid for small deformations u, but beyond this it becomes increasingly inaccurate. Nonlinear elastic deformation theory, or plasticity theory, deals with the behaviour of materials beyond the so called *yield point*, where deformations begin to become permanent.

Perhaps the most important thing to note is the occurrence of the four-tensor C. I will demonstrate shortly how this is closely interrelated with the idea of exponential geometry. In particular, we can give the exponential of a Riemannian manifold (M^n, g) by a Riemannian manifold (N^p, h) an induced structure, by stating that it acts on vector valued functions $v(s)$ and $w(t)$, $v : R^n \to R^p$ as

$$\langle\langle v(s), w(t)\rangle\rangle := g_{ij}h_{kl}v_i(s_k)w_j(t_l)$$

Note that $g \otimes h$ is a new 4-tensor naturally associated to this structure. We may abstract this construction and state that an exponential manifold (M, Λ) is a differentiable manifold with an inner product that has the following property

$$<< v(s), w(t) >> := \Lambda_{ijkl} v_i(s_k) w_j(t_l)$$

It is easy to see how it follows that we may write as an integral the overall stress with respect to internal variables (eg metrics) which will be a function on the manifold $\sigma : M \to TM \otimes TM$.

$$\sigma(m) = \int_A C(m,a)\dot{\epsilon}(m,a)da$$

But this looks very much like an information - for if we recall that ϵ behaves like ∇u, $\dot{\epsilon}$ will behave like a 2nd derivative, and in particular we expect it to behave like an information

$$\|\nabla_{g(m,a)} u\|^2 / \|u\|$$

Note that this will not be a scalar function, but a 2-tensor.

We may then take σ as a metric and take the information of this in turn. This gives us an elasto-plastic action of the form

$$\int_M R(C : Ric(g))dm$$

In particular I claim there is a natural 4-tensor Λ coming from the structure of a plastic manifold such that this is just

$$\int_M S_{\star^*,\Lambda}(m)dm$$

The toy model to keep in mind for a "sharp" Λ is not $\delta(\sigma(m) - a) \otimes \delta(\tau(m) - b)$ but rather

$$\delta(\delta(\tau(m) - b)\sigma(m) - a)$$

But this is just $exp(\star(ln(f); g)$ for appropriate sharp f and g. Hence we see immediately that whereas viscosity is the phenomenon that occurs in the measure for a $\star$-structure, plasticity is the associated phenomenon that occurs in the stack.

6.2 Schemes

The idea of scheme theory is to step away slightly from the standard structures in Algebraic Geometry, which are affine varieties, and to try to consider something slightly more fundamental and general. In this fashion scheme theory shares much of the spirit of geometric measure theory, which was developed to provide a more fundamental underpinning to the calculus of variations on manifolds. Also, in study of the theory, it is unclear initially as to which notions are purely technical, as opposed to those which could provide new and interesting perspectives.

To remind the reader, an affine variety is the zero locus of a finite set of polynomials.

Certainly all varieties are schemes. However it is interesting to ask as to what objects that are not varieties are also schemes, or as to what a scheme looks like in general. It is perhaps a mistake to merely consider schemes as a "regularisation" of the concept of variety, vis a vis the idea of submanifolds and rectifiable sets in geometric measure theory. In fact, schemes have additional structure, and intuitively can be roughly thought of as "(pre)geometric number fields".

I will not aim to provide here a complete survey of Grothendieck's theory, which is long and extensive [EGA]. Rather I will focus instead on focusing on important restricted cases of the theory, namely, geometric structures on differentiable manifolds.

Following Hartshorne [RH], I first introduce the notion of a *sheaf*.

Definition 88. (Sheaf). A sheaf is a topological space X, such that for all subsets U there exists a group $F(U)$ of functions from U to an algebraically closed field k. We also require $F(\phi) = 0$. Furthermore if $V \subset U$, there is a restriction map $\rho_{UV} : F(U) \to F(V)$. This map must satisfy the following properties:

(i) $\rho_{UU} = Id$,

(ii) If $W \subset V \subset U$, $\rho_{UW} = \rho_{VW} \circ \rho_{UV}$,

(iii) If U is open, and $\{V_i\}$ is an open cover, then if $s \in F(U)$ satisfies $s|_{V_i} = 0$ for each i, we must have $s = 0$, and

(iv) For the same U, $\{V_i\}$, if $s_i \in F(V_i)$ is a sequence with $s_i|_{V_i \cap V_j} = s_j|_{V_i \cap V_j}$, then there is an $s \in F(U)$ such that $s|_{V_i} = s_i$ for all i.

The particular example of a sheaf that will be of interest to us is that of the space of nondegenerate metrics over a differentiable manifold M:

$$\mathcal{T} = \{f | f : M \to \{\sigma : TM \to TM\}\}$$

Note that $\{\sigma : TM \to TM\}$ is an algebraically closed field as required.

Alternatively, consider the simpler, though perhaps slightly less natural example:

$$\hat{\mathcal{T}} = \{f | f : M \to \{g : M \to M\}\}$$

where now we are looking at the sheaf of automorphisms of a differentiable manifold. Of course this has something of a resonance with Galois theory! Note that $\{g : M \to M\}$ is just $Gal(M/R)$, the Galois group of M-automorphisms over the real numbers.

Now a scheme is a sheaf with additional geometric structure, in particular, algebraic structure.

Lemma 6.2.1. *(Structure sheaf for a ring, [RH]). Given a ring R, the associated structure sheaf is given by the pair $(Spec(R), \theta)$, where $Spec(R)$ is the set of all prime ideals of R, and θ is defined as a sheaf over $Spec(R)$ such that $\theta(U)$ is the set of functions x from U to the union of the localisations of R over U, with $x(p) \in R_p$ for every $p \in U$, and furthermore that x is locally a quotient of elements of R. The pair $(Spec(R), \theta)$ is otherwise known as the spectrium of R.*

The general intuition here is that we can convert the information in an algebraic ring R into a geometric structure with underlying space $Spec(R)$, and structural information of the ringed sheaf θ given by the nature of the maximal ideals in R. Since by definition rings are the most general objects in which one can define a multiplication, and we are interested in transition from such structures to geometric spaces, this does serve to motivate the above construction.

Definition 89. (Scheme, [RH]). A ringed space (X, θ) is a topological space X together with a sheaf of rings θ over X. An affine scheme is a locally ringed space that is isomorphic to the spectrum of some ring. A scheme is a locally ringed space (X, θ) in which every point in X has a neighbourhood U such that $(U, \theta|_U)$ is an affine scheme.

In practice what this means is that a scheme is essentially an underlying geometric space, M, with a superstructure N which has algebraic structure, ie self-multiplication, with properties determined by some structural tensor. This is not quite generally true, but in the case of examples that will be of interest to a differential geometer it is sufficient.

In particular, we will be interested in arithmetic schemes over the aforementioned sheafs of metrics, that is, the pairing $(\mathcal{T}, \Lambda)$, where Λ is a nondegenerate six tensor, that acts as a multiplication upon elements of $\mathcal{T}$ via the relation

$$(F \times_{\Lambda} G)_{ij} := \Lambda_{ij}^{abcd} F_{ab} G_{cd}$$

Of course we could define an alternative structure over the same sheaf, by considering

$$(F \times_{\Lambda} G)_{ij} := \Lambda_{ij}^{ad} F_{ab} G_{bd}$$

in which case Λ need only be a four-tensor.

For our simpler example once again we have that we can define an arithmetic scheme with the pairing $(\hat{\mathcal{T}}, \hat{\Lambda})$, where $\hat{\Lambda}$ is a non-degenerate four tensor, that acts in the following manner on $\hat{\mathcal{T}}$:

$$(F \times_{\hat{\Lambda}} G)_{ij} = \hat{\Lambda}_{ij}^{ab} (F \circ G)_{ab}$$

where now by abuse of notation I write f_{ij} for the mapping of the ith coordinate to the jth coordinate of M under f.

6.3 A fundamental theorem

It is interesting to ask if we can find the corresponding notion of Levi-Civita connection on such a space as in the previous class of examples, so as we are capable of doing calculus. In this section I will develop the required tools and prove the analogue of the Fundamental theorem of Riemannian geometry for this geometric structure.

One is faced with the difficulty of trying to determine which of the above three constructions contains the most natural example of the structures we would like to study. Namely, arithmetic structures on sheaves, or arithmetic schemes. In particular it turns out the the study of the sheaf of metrics over a manifold is actually related to the study of geometric viscosity, which is another topic altogether! So this might bias us towards the last construction.

For all intents and purposes though, the notion of multiplication of matrices somehow seems nicer than composition of functions. In particular I claim that the second and third constructions are in fact more or less equivalent (in fact, it is easy to demonstrate that composition of smooth automorphisms is the same as matrix multiplication in local coordinates). So it will be this that I study here. It is to be expected that this should contain no additional information, since it is merely a four-tensor structure.

However with the constructions above it is clear that they lack the necessary level of generality. In fact, it is necessary to take a step away from the intuition that we should be directly multiplying elements of the tangent sheaf together, to instead consider a four tensor structure $\Lambda : TM^4 \to R$. We can still induce a multiplication however - by taking necessary derivatives of Λ.

Consequently, we are looking at the pairing $(\mathcal{T}, \Lambda)$, with

$\mathcal{T} = \{f : M \to \{TM \to TM\}\}$, and Λ acts on $\mathcal{T}$ as a multiplication via the rule

$$(f \times_\Lambda g)_{ij} = \partial_{ij} \Lambda_{klpq} f_{kl} g_{pq}$$

Of course we will be primarily interested in the geometric structure coming from the structural tensor Λ. In particular we would like to arrange for an affine connection to obey the standard product rule for this structure:

$$X\langle Y, Z, A, B\rangle_\Lambda := \\ \langle \nabla_X Y, Z, A, B\rangle + \langle Y, \nabla_X Z, A, B\rangle + \langle Y, Z, \nabla_X A, B\rangle + \langle Y, Z, A, \nabla_X B\rangle$$

Then it is quite clear that the proof of the Levi-Civita theorem can be followed roughly to establish the uniqueness of this connection.

6.4 Generalised Christoffel Symbols and Local Information

In this section I will derive the generalised Christoffel symbols, and provide an expression for the local information. This latter result will be made more precise in a later section, when I deal with structures of slightly greater generality.

As mentioned earlier, we had established the uniqueness of our affine connection for the structure of interest. More precisely, we have that, if we assume symmetry of Λ, that

$$\begin{aligned}
&Y_0\langle Y_1, Y_2, Y_3, Y_4\rangle \\
&-Y_1\langle Y_0, Y_2, Y_3, Y_4\rangle \\
&+Y_2\langle Y_0, Y_1, Y_3, Y_4\rangle \\
&-Y_3\langle Y_0, Y_1, Y_2, Y_4\rangle \\
&+Y_4\langle Y_0, Y_1, Y_2, Y_3\rangle \\
&\qquad = 4\langle \nabla_{Y_0} Y_1, Y_2, Y_3, Y_4\rangle \\
&\qquad + \langle Y_1, [Y_0, Y_2], Y_3, Y_4\rangle + \langle Y_1, Y_2, [Y_0, Y_3], Y_4\rangle \\
&\qquad + \langle Y_1, Y_2, Y_3, [Y_0, Y_4]\rangle - \langle Y_0, [Y_1, Y_2], Y_3, Y_4\rangle \\
&\qquad - \langle Y_0, Y_2, [Y_1, Y_3], Y_4\rangle - \langle Y_0, Y_2, Y_3, [Y_1, Y_4]\rangle \\
&\qquad + \langle Y_0, Y_1, [Y_2, Y_3], Y_4\rangle + \langle Y_0, Y_1, Y_3, [Y_2, Y_4]\rangle \\
&\qquad - \langle Y_0, Y_1, Y_2, [Y_3, Y_4]\rangle
\end{aligned} \tag{6.1}$$

Note that this remains true even if we relax the assumption of symmetry.

If we use coordinates in which the Lie Brackets are trivial, such as orthonormal coordinates, then we have that this simplifies to

$$\begin{aligned}
&4\langle \nabla_{Y_0} Y_1, Y_2, Y_3, Y_4\rangle \\
&\qquad = Y_0\langle Y_1, Y_2, Y_3, Y_4\rangle - Y_1\langle Y_0, Y_2, Y_3, Y_4\rangle \\
&\qquad + Y_2\langle Y_0, Y_1, Y_3, Y_4\rangle - Y_3\langle Y_0, Y_1, Y_2, Y_4\rangle \\
&\qquad + Y_4\langle Y_0, Y_1, Y_2, Y_3\rangle
\end{aligned} \tag{6.2}$$

Then it follows that

$$\Gamma^{rs}_{kli} = \tfrac{1}{4}\Lambda^{rmsn}\{\partial_k\Lambda_{mlni} - \partial_m\Lambda_{klni} + \partial_l\Lambda_{mkni} - \partial_n\Lambda_{mkli} + \partial_i\Lambda_{mkln}\}$$

as the expression in normal coordinates for the generalised Christoffel symbols

$$\Gamma^{rs}_{kli} := \langle \nabla_{X_k} X_l \otimes X_i, X_r \otimes X_s \rangle$$

We then have an information theory with density (possibly up to subtle rearrangement of indices)

$$S_{\star,\Lambda} := -\Lambda_{abcd} \Gamma^{rs}_{apq} \Gamma^{pu}_{brt} \Gamma^{lq}_{cks} \Gamma^{kt}_{dlu}$$

where the $\star$ notation is to emphasise this is an S-curvature induced by multiplicative structure. Note that this is a *fourth order* geometric invariant, that is, it will become trivial ie zero on spaces of dimension less than four.

Finally, for something in the way of direct geometric interpretation of $S_{\star,\Lambda}$, we have the following result.

Lemma 6.4.1. *If we express Λ in normal coordinates about a point and expand in a δ-ball for small δ, we get*

$$\Lambda(x)_{ijkl} = \delta_{ijkl} + S_{\star,\Lambda,ijklmnop}(0) x_m x_n x_o x_p + o(\delta^8)$$

6.5 A functional principle for algebraic manifolds

In this section I will build and establish the appropriate form of the Cramer-Rao inequality for $\star$-structures on manifolds.

We build a statistical theory for these structures in the following fashion.

Definition 90. $\star(f; g) := fg$ for signal functions f, g over TM^2. In other words, the pointwise multiplication of the function f by g.

Definition 91. The *pregeometric fisher information density* for a $\star$-structure is defined to be $\star(f; \rho(g))\{\partial ln(\star(f; \rho(g))) \otimes \partial ln(\star(f; \rho(g)))\}$, where $\rho(g) = g\partial ln(g) \otimes \partial ln(g)$ is the standard fisher information density for a signal function g.

Definition 92. The *fisher information density* for a $\star$-structure is defined to be $F(\partial ln F)^{\otimes 4}$, where F is a signal function over TM^4.

From our earlier general considerations in the second chapter, we may conclude:

Corollary 6.5.1. *(Information inequality for a $\star$-structure) Let $F(m, a)$ be a statistical signal function over a star structure Λ. Then we have that*

$$\int_M \int_A F(m,a)\|\partial ln(F(m,a))\|_\Lambda^4 dadm \geq 0$$

To be more specific, if the signal function F is sharp, *then we have that*

$$\int_M S_{\star,\Lambda} dm \geq 0$$

and

Corollary 6.5.2. *(Information inequality for a pregeometric $\star$-structure). Let $f(m,a)$, $g(m,b)$ be statistical signal functions over a Riemann-Cartan manifold. Let $\rho(g)$ be the standard Fisher information density. Then we have an induced information inequality*

$$\int_M \int_A \int_B \star(f;\rho(g))\|\partial ln(\star(f;\rho(g)))\|^2 dbdadm \geq 0$$

In particular, if f and g are sharp, *and $f = \delta(\sigma(m) - a)$, $g = \delta(\tau(m) - b)$, then the above simplifies to the statement*

$$\int_M \star(R_\sigma; R_\tau) \geq 0$$

6.6 Topics in Algebraic Geometry

6.6.1 Some comments on the theory of Modular Forms

This comment will be based on a survey of the definitive work by Fred Diamond and Jerry Shurman [DS], wherein I shall seek to arrive at as quickly as possible the key result:

Theorem 6.6.1. *(Modularity Theorem). All elliptic curves are modular.*

It was this result in particular that enabled the resolution of Fermat's remaining infamous conjecture in the early 90s by Andrew Wiles. Since this is a treatise on information theory, and I have a preference for generality, I will be interested, if possible, in finding the key underlying geometric drivers for this result.

A certain function of interest which is connected with elliptic functions is the Weierstrass P-function

$$P(z;\omega_1,\omega_2) = \frac{1}{z^2} + \Sigma_{n+m\neq 0}\left(\frac{1}{(z-n\omega_1+m\omega_2)^2} - \frac{1}{(n\omega_1+m\omega_2)^2}\right)$$

This has the interesting symmetry

$$P(z;\omega_1,\omega_2) = P(z;1,\omega_2/\omega_1)$$

so it is really just a function of two (complex) variables, $P(z,\omega)$.

If we expand this for small z we get the Taylor series expansion

$$P(z) \approx z^{-2} + C_1 g_2 z^2 + C_2 g_3 z^4 + O(z^6)$$

where C_1, C_2 are constants chosen such that we have the identity

$$(P')^2 = 4P^3 - g_2 P - g_3$$

Here g_2, g_3 are numbers dependent on ω. These can in turn be used to describe the properties of an elliptic curve.

Now, inspired by the fact that we have fourth powers in our expansion, together with the nature of Wiles' proof, we expect that an almost sharp expansion in a metric over the space of variable z, together with a $\star$-structure over variable ω, should be sufficient to prove the modularity theorem.

In particular we are intuitively led to examine the functional

$$I(\epsilon,\Lambda) = \int_M S_{\star;\Lambda;(1)} + \epsilon S_{\star;\Lambda;(2)}$$

which I claim embodies the information of an elliptic curve. Then the modularity theorem should really only be a restatement of the Cramer-Rao inequality for this information.

6.6.2 K-theory and Index Theory

It is interesting to ask as to where K-theory fits into all of this, since there is a relationship between this and the theory of schemes. In addition to providing context, I will aim to provide an intuitive sketch of the Atiyah-Singer index theorem by the conclusion of this section.

In fact, I claim that an information theory of the above form is sufficient to establish the Atiyah-Singer index theorem - in particular that this follows from the Cramer-Rao inequality for the same. To establish this I will need to describe this particular result and convert the problem into a form wherein information theory may be applied.

Recall from before that the Atiyah Singer index theorem relates to the action of elliptic differential operators on Riemannian manifolds, which can essentially be viewed as generalised product structures. Then the theorem states that the analytical index of the space, that is, the dimension of the negative eigenspace of the associated 4th rank tensor - a local measure - is one and the same as the topological index of the space, which is essentially the dimension of the subspace of the manifold which is in the kernel of the generalised product map - a global measure.

But product maps, or product structures, are characterised by structures $\Lambda := \star(\sigma; \tau)$, where σ is the structure of the manifold, and τ is the structure of the action of the product. Consequently, we have that the local information induced by such an operator is $S_{\star;\Lambda}$, and the Cramer-Rao inequality states that

$$I(\Lambda) := \int_M S_{\star;\Lambda}(m)dm \geq 0$$

and consequently that it will be critical only if the information density is zero, or

$$S_{\star;\Lambda}(m) = 0$$

I suspect that the index theorem then follows from this observation.

6.6.3 The Weil Conjectures

Following a similar line of reasoning to the above, I will now briefly mention how this theory might be used to provide a quick, elegant proof of the Weil conjectures - which were largely resolved by 1974 by Grothendieck, Deligne and others. Since these conjectures deal with elliptic equations, and since elliptic equations arise from the study of correction terms within the Weierstrass P-function (of interest in analytic number theory), it makes sense to consider *almost sharp* $\star$-structures.

In other words, I claim that it should be possible to resolve the Weil conjectures, through examination of the perturbative action

$$\int_M (S_{\star,(1),\Lambda}(m) + \epsilon S_{\star,(2),\Lambda}(m) + \epsilon^2 S_{\star,(3),\Lambda}(m) + ...)dm$$

which is greater than or equal to zero via the appropriate analogue of the Cramer-Rao inequality.

Naturally in order to make this a convincing argument it is necessary to build a correspondence between the generating functions coming from such algebraic varieties and perturbative $\star$-structures. However, I do not intend to do so here.

6.6.4 C-star algebras

As a final demonstration of the inter-relation of these ideas to other areas of mathematics, I will describe how these tools can be used to establish the Gelfand-Naimark theorem for C^* algebras. These are of course particularly special forms of $\star$-structure, where all invariants and functions are analytic over some general complex manifold M.

References that might prove instructive to the reader in this instance are the book by Alain Connes [Cns], and the classic paper of Israel Gelfand and M. Naimark [GN].

Recall:

Definition 93. (Banach algebra). A *Banach algebra* $\mathcal{A}$ over the complex numbers is, roughly speaking, a Kähler manifold M with the property that there is a multiplication $\cdot$ such that $\|m \cdot n\|_M \leq \|m\|_M \|n\|_M$.

Definition 94. (C^* algebra). A C^*-algebra is a *Banach algebra* $\mathcal{A}$ over the complex numbers, with the additional existence of an *involution operator* $* : \mathcal{A} \to \mathcal{A}$ - which is best thought of as a generalisation of complex conjugate. This operator must satisfy the following properties:

(i) $(x + y)^* = x^* + y^*$ (linearity)

(ii) $(xy)^* = y^* x^*$ (conjugate property for algebraic multiplication)

(iii) $(\lambda x)^* = \bar{\lambda} x^*$ (conjugate property for scalars)

It can then be demonstrated that, under these conventions, the C^* identity holds for all $x \in \mathcal{A}$:

$$\|x^* x\| = \|x\| \|x^*\| = \|x x^*\|$$

It follows that $\|x^*x\| = \|x\|^2$ for all $x \in \mathcal{A}$.

This has the following interpretation. First, the product structure in a geometrised Banach algebra $\mathcal{A}$ can be extended to a metric for a Kähler manifold; in particular all Kähler manifolds admit such a product. To see why this is the case, note that it is well known that all complex submanifolds of a Kähler manifold are minimal and Kähler. Hence local forms extend to global Kähler submanifolds of the total space.

Consequently a Kähler manifold M has an induced algebra $\mathcal{A}$, with product for two points m and n constructed in the following way. Suppose we make the simplifying assumption that there is a universal chart for the total space. Associate to m and n then the vectors v, w under the preimage of the exponential map with respect to the metric σ. Then, in the tangent space for the chart take the cross product $v \times w$, and project back down to M. Then, we define $m \cdot n$ as $exp_\sigma(v \times w)$. Furthermore this operation is well defined via our previous observation that local forms extend globally in a Kähler space.

Hence the operation $\cdot$ can be associated to a Kähler metric σ.

I now claim that if we now introduce an involution operator $*$ to our space M, we can associate this to a secondary Kähler metric τ, which acts on σ via the standard notion of $\star$-structure as developed in this chapter, so that M is geometrised by the 4-tensor $\star(\sigma;\tau)$.

Generally speaking, suppose we have a vector $v \in C^k$. Then we can construct the trivial conjugate $\bar{v}$ of v by writing $v = v_r + iv_c$, for v_r, v_c real, and then mapping v to $v_r - iv_c$. An involution is then constructed in the following way. Let m once again be an element of a Kähler manifold. Let $n = exp_\sigma^{-1}(m)$ be its preimage. Let $v = exp_\tau^{-1}(n)$ be the preimage of n under a secondary Kähler metric. Then we can write $m^* := exp_\sigma exp_\tau(\bar{v}) = exp_{\star(\sigma;\tau)}(\bar{v})$. This establishes my claim.

But this is just another way of describing a triple $TO \rightarrow_\tau O \rightarrow_\sigma H$, where O is a trivial C^*-algebra of bounded operators over a Hilbert space (complete normed space) H. Hence this demonstrates, or at least leads us to suspect, why the following result should be true:

Theorem 6.6.2. *(Gelfand-Naimark). Every C^*-algebra $\mathcal{A}$ is isometrically *-isomorphic to a C^*-algebra of bounded operators on a Hilbert space.*

Chapter 7

Lattice processes and Transcendental Geometry

7.1 Transcendental Geometry and Reverberability

Recall from [Go] that a transcendental geometry is driven by the geometric precursor $\wedge(f;g) := f^g$. Also we have an information inequality:

$$\int_M \wedge(R_\sigma; R_\tau) dm \geq 0$$

in the case that f and g are sharp, and which extends to more general signal functions as well.

Similarly to before, one can have $\wedge$-structures in the stack or the measure - which once again leads to an infinite number of discrete geometric bifurcations.

This is the starting point for the study of *reverb*, or *transcendental structures*. These are 4-tensors that occur as the resolutions of the above sequence, playing a similar role to the more primitive composition operator as plastic structures played to turbulent geometries before.

Now, to understand reverb, we ask the natural question, is it possible to take a non-integral "composition" of a curvature operator, or, more generally, a signal function f? And if so, what does this mean physically?

First of all, note that we cannot directly define non-integral composition of a curvature operator or of a signal function, since these are not automorphisms. However, note that if we take the logarithm

$$log : R \to M \times A$$

then $log(f) : M \times A \to M \times A$ *is* an automorphism.

But this is not terribly natural either. However, borrowing this intuition, consider instead the standard logarithmic function $log : R \to R$.

Then $log(f) : M \times A \to R$. If we consider log-composition to be multiplication, then build $\wedge(log(f); g)$, where g is another signal function. Finally, to convert back to normal coordinates, apply the exponential mapping.

Hence we are interested in the *log-composition* operation

$$exp(\wedge(ln(f); g))$$

If f and g are delta functions, the result of this operation will be

$$\delta(\sigma(m)^{\delta(\tau(m)-b)} - a)$$

We expect the corresponding information then to be

$$\int_M R_{\wedge(\sigma(m);R(\tau)(m))} dm$$

7.2 The fundamental theorem of geometric reverb

Here I will develop the theory in a more formal manner, with the aim of building towards a proof of the existence of derivatives in transcendental manifolds.

Definition 95. Given a differentiable manifold M, a *reverb structure* Λ on M is a 4-tensor such that, for $x, y \in \{TM \to TM\}$, and $u, v \in TM$, we have

$$< x(u), y(w) >_\Lambda := \Lambda_{ijkl} x_i(u_k) y_j(w_l)$$

Note that, just as one can construct a local orthonormal basis $\{X_i\}$ for the tangent space of M, so can one construct a local orthonormal basis $\{X_{ij}\}$ for the space of functions from TM back to itself. These in turn form a kind of coordinates for our viscous structure.

Then one is interested in the idea of a connection on these coordinates. A connection ∇ will have the properties that

(i) $\nabla_{X_{ij}} f Y_{kl} = f\nabla_{X_{ij}} Y_{kl} + X_{ij}(f) Y_{kl}$ (product rule)

(ii) $\nabla_{fX} Y = f\nabla_X Y$ (linearity)

Then I claim in analogy to the Riemannian case that there is a unique connection, which I will call the plastic connection, such that

$$X_{ij} < Y_{kl}, Z_{mn} >=< \nabla_{X_{ij}} Y_{kl}, Z_{mn} > + < Y_{kl}, \nabla_{X_{ij}} Z_{mn} >$$

But this is trivially true since we can just view $1 \leq i, j \leq n$ as a basis for matrices, and then the result follows from the uniqueness theorem for the standard Riemannian connection.

This has some useful consequences.

Definition 96. The generalised Christoffel symbols associated to a viscous structure Λ take the form

$$\Gamma^{mn}_{ijkl} =< \nabla_{ij} X_{kl}, X_{mn} >$$

By the fundamental theorem, we can compute these terms explicitly in terms of Λ:

$$\Gamma^{i\alpha}_{k\gamma l\epsilon} = \tfrac{1}{2}\Lambda_{im\alpha\zeta}\left(\tfrac{\partial \Lambda_{mk\zeta\gamma}}{\partial X^{l\epsilon}} + \tfrac{\partial \Lambda_{ml\zeta\epsilon}}{\partial X^{k\gamma}} - \tfrac{\partial \Lambda_{kl\gamma\epsilon}}{\partial X^{m\zeta}}\right)$$

But observe that the generalised scalar curvature S_Λ takes the form

$$S_{\wedge;\Lambda} = -\Lambda^{ijkl}\Gamma^{\gamma\delta}_{ij\alpha\beta}\Gamma^{\alpha\beta}_{kl\gamma\delta}$$

So this can be computed wholly in terms of Λ and its derivatives.

Remark. Note that, unlike turbulent or viscous structures, this is a second order invariant, rather than fourth order. However, whereas the former were defined purely on the space M, these are defined on the space of tensors on M, $\{TM \rightarrow TM\}$, and are toy models for the study of self-referential structures, to be discussed in the sequel.

7.3 The Hodge Conjecture

I will now quickly indicate how one might attack this particular famous problem. The idea of my treatment will merely be to sketch and provide some intuition as to how one might proceed.

Essentially in spirit the Hodge conjecture asks whether, given a (possibly degenerate) bilinear form σ on a differentiable manifold M, if one constructs a generator of the associated Cech cohomology theory ω, does there exist a corresponding submanifold or subvariety $N(\omega)$ of the total space M?

There are essentially two difficulties here. One is the ambiguity between nondegenerate and degenerate. In order to treat submanifolds properly one needs to essentially work in the space of varieties. This was observed by the twentieth century algebraists, as might be gathered from my survey of K-theory. (Note also that I cheated a little in the earlier chapter, by assuming that one always has a correspondence between cohomology and submanifolds/subvarieties of spaces, in order to assist the development of my, and the reader's physical intuition regarding the theory. This is not always the case.) I will get around this difficulty by increasing the degree of abstraction and instead embedding nondegenerate Riemannian structures into nondegenerate Transcendental structures.

The second difficulty is the fact that forms are defined only locally, and submanifolds have global properties. So one needs a natural way to extend forms in a natural way that will produce submanifolds. The way I will address this is through transition from local to global via the formulation of an appropriate local information density, based on the Transcendental methods in this chapter.

Before I proceed I should stress that most generally, it is well known that there exist forms ω on manifolds M that do not extend to submanifolds N with Riemannian structure. However, I claim that it is possible to extend in general to submanifolds N with *reverb* structure, in which the class of Riemannian structures is naturally embedded. It is possible of course that even this may not be sufficient, and one needs the more general methods of the self-referential geometry in order to establish a correspondence. Nonetheless let us proceed; if nothing else we shall see some more of the shape of the problem.

Recall given a reverb structure Λ that we have an information density $S_{\wedge;\Lambda}$ defined on our manifold M.

Now suppose we instead consider a Riemannian metric σ on M, and also consider a form ω in the Tensor algebra $\mathcal{A}$ associated to M. I claim that we can place a natural

Riemannian structure τ on $\mathcal{A}$. Consider two forms α, β. We would like to define an invariant $\tau : \mathcal{A} \times \mathcal{A} \to R$. Note that $\mathcal{A}$ locally has the same dimension as the tangent space of M, and will also be a vector bundle over M.

To define this, if $\alpha_1 \wedge ... \wedge \alpha_n$ is one form and $\beta_1 \wedge ... \wedge \beta_n$ the other, possibly with degenerate entries, then we can define $\tau(\alpha, \beta) := \tau_{ij}\alpha_i\beta_j$ in the standard way.

We can then construct a reverb structure Λ on M as $\wedge(\sigma; \tau)$, and I claim that this provides us with what we need. In particular, if the information functional

$$I(\Lambda) = \int_M S_{\wedge;\Lambda}(m)dm$$

is critical then I claim that we have a correspondence between the structure of the tensor algebra $\mathcal{A}$ associated to M and its tangent bundle TM. In other words we have that forms ω are uniquely associated to appropriate submanifolds of M with reverb structures.

Obviously there is a fair bit to check here, and considerable further detail would be required before this was a convincing argument. Nonetheless I think this is suggestive and interesting.

Chapter 8

(First order) self-referential geometry

The goal of this chapter will be to draw upon the three main ideas, or "seals of existence" as so far described in this manuscript, and demonstrate how they form a cohesive and general description of a first order self-referential geometry on a given differentiable structure M.

In particular, the rough intuition is that we would like to describe how, for every coordinate component x_j of each point $m \in M$, there are natural information channels to every other point $n \in M$. In particular there is a natural geometric structure, viewed on the space of maps from TM to itself, or, naively, associated to $M \times M$, which can be used to describe such interrelations. Furthermore, I claim that it is possible to quantify the resultant dynamics.

There are in fact strong connections between these ideas and the attempt to build a theory of physics for 1-categories, or string theory. In fact it turns out that T-duality (see, for instance, [KV]) is essentially the statement of criticality for the Cramer-Rao inequality for the T-invariant of a symmetric 8-tensor representing the geometry, and the antisymmetric 8-tensor representing the distribution of matter. Of course, the treatment I provide here is more general, and I derive the T-invariant as the Fisher Information density for a nondegenerate 8-tensor with no assumption of symmetry.

This is also related to the work of Stephen Hawking, Robert Bartnik, and Gerhard Huisken on the quasi-local mass. This is an attempt to try to extend the idea of mass to general relativity, but, at a deeper level, it is related to the question of attempting to resolve the physics at singularities, or the black-hole solutions to the theory. Naturally one would expect this to be resolved by a more abstract mathematical underpinning to the theory, rather than using the theory of Riemannian manifolds, however the approach taken by relativists is rather less direct.

In attempts to obtain bounds on the behaviour about singularities, it turns out that the so-called Willmore functional, defined for surfaces $\Sigma \subset M$ as the integral of the mean curvature over Σ, is quite useful, as I learned in an interesting talk at the January 2010 Monash workshop on General Relativity and Geometric Analysis, given by Jan Metzger, one of Professor Huisken's students. This is not particularly surprising, since matrices of surfaces, glued at their boundary, with k marked points, actually occur in quite a natural way in the discussion of self-referential geodesics. Interestingly, this is related to structures on moduli space, in particular, Teichmuller space, and also the so-called "strings" of string theory have world sheets that are surfaces.

It can be demonstrated that one obtains quite natural 4th order nonlinear optimisation problems for trying to optimise the choice of Willmore functional, which hints at the need to develop an abstraction of Riemannian geometry to attack the singularity problem in general relativity, which is really just symptomatic of the incompleteness of the theory.

As a further comment, I have noticed since the completion of the bulk of this research that various other contemporary authors are interested in the study of self-referential dynamics. For instance, there is the 2000 paper by R. Cahill and CM Klinger, "Self-referential noise and the synthesis of three-dimensional space", `http://arxiv.org/pdf/gr-qc/9812083`, as well as R. Cahill's paper "Process Physics : Self-Referential Information and Experiential Reality", `http://processthought.info/publications/Articles/LSI05/Cahill-FinalPaper.pdf`. Other authors who have written on the subject include M Wyart and JP Bouchard, in their 2007 paper "Self-referential behaviour, overreaction and conventions in financial markets" at `http://arxiv.org/pdf/cond-mat/0303584`, and Richard Sutton and Brian Pinette, in their 1985 paper "The learning of world models by connectionist networks", `http://webdocs.cs.ualberta.ca/~sutton/papers/sutton-pinette-85.pdf`.

However philosophers have been interested in the idea of self-reference for significantly longer than this. In the article by Thomas Bolander, [Bol], it is described how the idea of an object that refers to itself can lead to mind-bending paradoxes - such as "this sentence is not true". This is an example of the so-called *Liar paradox*, which is generally attributed to Eubulides of Miletus- of the Megarian school- in the 4th century BC. Such examples demonstrate that not all self-referential constructions are logical or make sense. In fact, the realisation of the liar's paradox in mathematics (as Russell's paradox) was one of the key issues that frustrated the attempts of the early 20th century mathematicians in their development of axiomatic set theory (as mentioned in the second chapter of this work). It is therefore of interest to attempt to understand how to "tame" self-reference in a structured and formal fashion, and to do so to first order.

If the reader will forgive the culture reference, there is also of course the amusing and clever recent xkcd comic `http://xkcd.com/688/`. So certainly the idea of self reference is a meme that is currently "in the water", so to speak.

8.1 Key results

8.1.1 Preliminaries

Define

$$\Theta^{\epsilon\eta\zeta a}_{ijkl\alpha\beta\gamma\delta} = \langle \partial_{ij} X_{\alpha\beta}, \partial_{kl} X_{\gamma\delta}, X_{\epsilon\eta}, X_{\zeta a} \rangle$$

as the generalised Christoffel symbols, where $\partial_{ij} X_{kl} := \frac{\partial}{\partial X_{ij}} X_{kl}$, for X_{ij} a basis of the function space $\{TM \to TM\}$.

Require the affine connection ∂ to satisfy

$$\begin{gathered} X_{ij} X_{kl} \langle X_{\alpha\beta}, X_{\gamma\delta}, X_{\epsilon\eta}, X_{\zeta a} \rangle = \\ (\langle \partial_{ij} X_{\alpha\beta}, \partial_{kl} X_{\gamma\delta}, X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle \partial_{ij} X_{\alpha\beta}, X_{\gamma\delta}, \partial_{kl} X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle \partial_{ij} X_{\alpha\beta}, X_{\gamma\delta}, X_{\epsilon\eta}, \partial_{kl} X_{\zeta a} \rangle) \\ +(\langle \partial_{kl} X_{\alpha\beta}, \partial_{ij} X_{\gamma\delta}, X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle X_{\alpha\beta}, \partial_{ij} X_{\gamma\delta}, \partial_{kl} X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle X_{\alpha\beta}, \partial_{ij} X_{\gamma\delta}, X_{\epsilon\eta}, \partial_{kl} X_{\zeta a} \rangle) \\ +(\langle \partial_{kl} X_{\alpha\beta}, X_{\gamma\delta}, \partial_{ij} X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle X_{\alpha\beta}, \partial_{kl} X_{\gamma\delta}, \partial_{ij} X_{\epsilon\eta}, X_{\zeta a} \rangle + \langle X_{\alpha\beta}, X_{\gamma\delta}, \partial_{ij} X_{\epsilon\eta}, \partial_{kl} X_{\zeta a} \rangle) \\ +(\langle \partial_{kl} X_{\alpha\beta}, X_{\gamma\delta}, X_{\epsilon\eta}, \partial_{ij} X_{\zeta a} \rangle + \langle X_{\alpha\beta}, \partial_{kl} X_{\gamma\delta}, X_{\epsilon\eta}, \partial_{ij} X_{\zeta a} \rangle + \langle X_{\alpha\beta}, X_{\gamma\delta}, \partial_{kl} X_{\epsilon\eta}, \partial_{ij} X_{\zeta a} \rangle) \end{gathered}$$

8.1.2 The fundamental theorem of self-referential geometry

Via similar arguments to previous chapters, we conclude that

$$\begin{gathered} 16\Theta^{abcd}_{ijkl\alpha\beta\gamma\delta} := 16\langle \nabla_{ij} X_{\alpha\beta}, \nabla_{kl} X_{\gamma\delta}, X_{ab}, X_{cd} \rangle = \\ X_{ij}(X_{kl}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{cd} \rangle - X_{\alpha\beta}\langle X_{kl}, X_{\gamma\delta}, X_{ab}, X_{cd} \rangle + X_{\gamma\delta}\langle X_{\alpha\beta}, X_{kl}, X_{ab}, X_{cd} \rangle - \\ X_{ab}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{kl}, X_{cd} \rangle + X_{cd}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{kl} \rangle) \end{gathered}$$

$$-X_{kl}(X_{ij}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle - X_{\alpha\beta}\langle X_{ij}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle +$$
$$X_{\gamma\delta}\langle X_{\alpha\beta}, X_{ij}, X_{ab}, X_{cd}\rangle - X_{ab}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ij}, X_{cd}\rangle + X_{cd}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{ij}\rangle)$$
$$+X_{\alpha\beta}(X_{kl}\langle X_{ij}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle - X_{ij}\langle X_{kl}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle + X_{\gamma\delta}\langle X_{ij}, X_{kl}, X_{ab}, X_{cd}\rangle -$$
$$X_{ab}\langle X_{ij}, X_{\gamma\delta}, X_{kl}, X_{cd}\rangle + X_{cd}\langle X_{ij}, X_{\gamma\delta}, X_{ab}, X_{kl}\rangle)$$
$$-X_{\gamma\delta}(X_{kl}\langle X_{\alpha\beta}, X_{ij}, X_{ab}, X_{cd}\rangle - X_{\alpha\beta}\langle X_{kl}, X_{ij}, X_{ab}, X_{cd}\rangle + X_{ij}\langle X_{\alpha\beta}, X_{kl}, X_{ab}, X_{cd}\rangle -$$
$$X_{ab}\langle X_{\alpha\beta}, X_{ij}, X_{kl}, X_{cd}\rangle + X_{cd}\langle X_{\alpha\beta}, X_{ij}, X_{ab}, X_{kl}\rangle)$$
$$+X_{ab}(X_{kl}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ij}, X_{cd}\rangle - X_{\alpha\beta}\langle X_{kl}, X_{\gamma\delta}, X_{ij}, X_{cd}\rangle + X_{\gamma\delta}\langle X_{\alpha\beta}, X_{kl}, X_{ij}, X_{cd}\rangle -$$
$$X_{ij}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{kl}, X_{cd}\rangle + X_{cd}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ij}, X_{kl}\rangle)$$
$$-X_{cd}(X_{kl}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{ij}\rangle - X_{\alpha\beta}\langle X_{kl}, X_{\gamma\delta}, X_{ab}, X_{ij}\rangle + X_{\gamma\delta}\langle X_{\alpha\beta}, X_{kl}, X_{ab}, X_{ij}\rangle -$$
$$X_{ab}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{kl}, X_{ij}\rangle + X_{ij}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{kl}\rangle)$$

(This is assuming, WLOG, we have chosen coordinates so that the Lie Brackets are all trivial)

But derivatives commute, so a few terms cancel. Consequently the right hand side simplifies to

$$X_{ij}(-2X_{\alpha\beta}\langle X_{kl}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle - 2X_{ab}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{kl}, X_{cd}\rangle)$$
$$-X_{kl}(-2X_{\alpha\beta}\langle X_{ij}, X_{\gamma\delta}, X_{ab}, X_{cd}\rangle + 2X_{\gamma\delta}\langle X_{\alpha\beta}, X_{ij}, X_{ab}, X_{cd}\rangle -$$
$$2X_{ab}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ij}, X_{cd}\rangle + 2X_{cd}\langle X_{\alpha\beta}, X_{\gamma\delta}, X_{ab}, X_{ij}\rangle)$$
$$X_{\alpha\beta}(X_{\gamma\delta}\Lambda_{ijklabcd} - X_{ab}\Lambda_{ij\gamma\delta klcd} + X_{cd}\Lambda_{ij\gamma\delta abkl})$$
$$-X_{\gamma\delta}(-X_{\alpha\beta}\Lambda_{klijabcd} - X_{ab}\Lambda_{\alpha\beta ijklcd} + X_{cd}\Lambda_{\alpha\beta ijabkl}$$
$$+X_{ab}(-X_{\alpha\beta}\Lambda_{kl\gamma\delta ijcd} + X_{\gamma\delta}\Lambda_{\alpha\beta klijcd} + X_{cd}\Lambda_{\alpha\beta\gamma\delta ijkl})$$
$$-X_{cd}(-X_{\alpha\beta}\Lambda_{kl\gamma\delta abij} + X_{\gamma\delta}\Lambda_{\alpha\beta klabij} - X_{ab}\Lambda_{\alpha\beta\gamma\delta klij})$$

but this is just

$$2(-\partial_{ij\alpha\beta}\Lambda_{kl\gamma\delta abcd} - \partial_{ijab}\Lambda_{\alpha\beta\gamma\delta klcd} + \partial_{kl\alpha\beta}\Lambda_{ij\gamma\delta abcd} - \partial_{kl\gamma\delta}\Lambda_{\alpha\beta ijabcd} + \partial_{klab}\Lambda_{\alpha\beta\gamma\delta ijcd} -$$
$$\partial_{klcd}\Lambda_{\alpha\beta\gamma\delta abij}) + (\partial_{\alpha\beta\gamma\delta}\Lambda_{ijklabcd} - \partial_{\alpha\beta ab}\Lambda_{ij\gamma\delta klcd} + \partial_{\alpha\beta cd}\Lambda_{ij\gamma\delta abkl}) + (\partial_{\gamma\delta\alpha\beta}\Lambda_{klijabcd} +$$
$$\partial_{\gamma\delta ab}\Lambda_{\alpha\beta ijklcd} - \partial_{\gamma\delta cd}\Lambda_{\alpha\beta ijabkl}) + (-\partial_{ab\alpha\beta}\Lambda_{kl\gamma\delta ijcd} + \partial_{ab\gamma\delta}\Lambda_{\alpha\beta klijcd} + \partial_{abcd}\Lambda_{\alpha\beta\gamma\delta ijkl}) +$$
$$(\partial_{cd\alpha\beta}\Lambda_{kl\gamma\delta abij} - \partial_{cd\gamma\delta}\Lambda_{\alpha\beta klabij} + \partial_{cdab}\Lambda_{\alpha\beta\gamma\delta klij})$$

Consequently we may compute the generalised Christoffel symbols from the above expression, provided that we know Λ, where

$$\partial_{ijkl}f := \frac{\partial^2}{\partial X_{ij}\partial X_{kl}}f$$

This is naturally a synthesis of all the work that has been done prior to this point.

8.1.3 The information density and Cramer-Rao inequality

The information density for such a structure is given by something like the following relation

$$T_\Lambda = -\Lambda^{ijklmnpq}\Theta^{a_3b_2c_1d_3e_2}_{ija_0b_0c_0d_0e_0}\Theta^{a_0b_3c_2d_0e_3}_{kla_1b_1c_1d_1e_1}\Theta^{a_1b_0c_3d_1e_0}_{mna_2b_2c_2d_2e_2}\Theta^{a_2b_1c_0d_2e_1}_{pqa_3b_3c_3d_3e_3}$$

This is an eighth order geometric invariant.

For almost sharp, or perturbative dynamics, we immediately observe that the germ of the space $\{TM \to TM\}$ must need be at least 16 dimensional, since, $T_{\Lambda,(2)}$ is otherwise trivial. Consequently we have again that for such structures, an underlying four-dimensional space M is preferred. Via similar arguments to [Go] it is possible to establish too that the preferred index of M is one, via a dimensional analysis.

8.2 Self-referential structures in theoretical physics

8.2.1 Proof of preferred Lorentzian structure for the tensor product

As mentioned, we clearly require M be four dimensional at least for examination of quantum mechanical phenomena, under the umbrella of self-referential dynamics. Then, as before, it is not too hard to show that the number of dimensions preferred for a turbulent self-referential geometry is as small as possible. In particular

Theorem 8.2.1. *(Minimisation of dimension). Let M be a manifold with a self-referential structure. Then analysis on the dimension of M for purposes of stability requires that the dimension be as small as possible.*

Proof. (Idea). We consider variation on the dimension of M, given that it has a self-referential structure Λ. Set Σ as the associated self-referential structure for the dimension space. Then

$$\partial^*(T_\Lambda; T_\Sigma)$$

will be critical only if $\partial_\Lambda T_\Sigma$ and T_Σ are the same up to contraction of indices; then we must have that ∂_Λ is the identity operator, or that $\Lambda = 0$. □

Definition 97. An *eigenvector* of an n-tensor Λ is a vector v such that $\Lambda(v, ..., v, *)$, the evaluation of Λ in all entries save one is λv for some constant λ. Furthermore, if we were to omit a different entry, we require that the answer be the same. λ is then an *eigenvalue* of Λ. The set of eigenvectors forms the *eigenspace*. The *index* of a self-referential structure Λ is the dimension of its negative eigenspace, ie the set of vectors with negative eigenvalues.

Theorem 8.2.2. *(Index theorem for self-referential structures). Under this notion of index, the turbulent index* $\partial^*(ind(\Lambda); ind(\Sigma)) = ind(\Lambda)^{ind(\Sigma)}$.

Proof. The proof is similar to that if Λ and Σ are instead Riemannian metrics. See [Go]. □

Then, as in [Go], we have that for perturbative dynamics, our dimension must be bounded below by four. By the first theorem the dimension must also be as small as possible. By the second theorem in combination with the first, the index must be one. This completes our sketch as to why Lorentzian self-referential geometries are preferred.

8.2.2 AdS / CFT, and Geometric Langlands

I claim that one can establish the so called anti-de-Sitter / Conformal field theory correspondence [Mal] using these methods. In particular, I claim that the statement follows from the Cramer-Rao inequality applied to a sharp self-referential structure,

$$\int_M T_\Lambda(m)dm \geq 0$$

and the fact that the AdS / CFT and conformal structures correspond to the antisymmetric and symmetric parts of a critical sharp self-referential structure as above respectively.

I make the further claim that many of the key questions within the Geometric Langlands program also follow from the same consideration, albeit in perhaps the less general case where $M = C$, the complex numbers.

8.2.3 Consciousness and Artificial Intelligence

It is always very difficult dealing with issues such as the nature of conscious perception. In particular one encounters difficulties when using many of the standard tools, since they often underestimate the difficulty of the underlying problem. Many people have examined this problem in various levels of detail. Penrose has suggested [Pe] that some new form of quantum mechanics, of potentially quite different and deeper character than the classical theory may be what is required to gain some insight into this process.

Consequently, for a first order approximation to at least some idea of what is going on, I suspect a self-referential quantum mechanical treatment is required. This fits in with a large amount of intuition, since in much of AI research that has found practical utility, feedback loops of various description have been implicitly utilised in the associated engineering implementation. Various studies, I believe, also suggest that this is a reasonable model for how many processes in the brain operate. And of course (first-order) self-referential geometry is very strong in this regard.

So my proposed model is to consider an almost sharp self-referential action,

$$I(\Lambda, \epsilon) = \int_M (T_{(1)\Lambda}(m) + \epsilon(m) T_{(2)\Lambda}(m)) dm$$

(noting that higher order terms in the expansion are trivial in 4-dimensional space), where Λ is a self-referential structure- an 8-tensor- and ϵ is an expansion parameter that varies smoothly over M.

Then the Cramer-Rao inequality states that this is strictly non-negative. Evaluating the first-variation, and setting to zero, we examine critical dynamics, that one might be able to use to model conscious perception:

Variation with respect to Λ:

$$T_{(1)\Lambda}(m) + \epsilon T_{(2)\Lambda}(m) = 0$$

Variation with respect to ϵ:

$$\frac{\partial \Lambda}{\partial \epsilon} \frac{\partial T_{(1)\Lambda}}{\partial \Lambda} + T_{(2)\Lambda}(m) = 0$$

Naturally this is only a preliminary indication of how one might proceed with the analysis. The exciting thing, of course, is that this has immediate applications to the development of artificial intelligence in software agents.

An immediate consequence of the above is that we have, as with the Yang-Mills equations, eight stable families of solutions to these equations, with several eigenstates per family (this is essentially a more abstract consequence of the geometrisation conjecture). This is again in accordance with intuition. For instance, this is consistent with the multiple intelligences theory due to Howard Gardner (1983). (The orginal theory was advocated as a form of descriptively categorising different forms of human personality types.)

8.2.4 Superconductivity

I will make a few brief remarks here about what I believe may be required for the development of a theory of high temperature superconductivity. Recall that one of the main objects of interest in the standard theory is the London equation:

$$\epsilon j = -A$$

where j is the superconducting current (in normalised units), and ϵ is a small dimensionless parameter.

But $j = -curl(B)$, and $B = curl(A)$, so the London equation is really

$$A - \epsilon \nabla \times \nabla \times A = 0$$

We would like to abstract away from this, and in particular see if this arises naturally from an information. For then there is the possibility to look at deeper related information functionals and possibly find and/or derive equations describing subtler high-Tc physics.

Now first observe that $\nabla \times \nabla \times A = \epsilon_{ijk}\epsilon_{klm}\partial_{jl}A_m = (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl})\partial_{jl}A_m = \nabla(\nabla \cdot A) - \Delta A$.

So if we choose the Gauge so that $\nabla \cdot A = 0$ (which is the London gauge and is assumed for the initial equation to make sense), we obtain the following form for the London equation:

$$A + \epsilon \Delta_\sigma A = 0$$

which is starting to look quite promising indeed. Note that since A is antisymmetric we can generalise this into an equation in terms of the curvature of a generalised Cartan-Riemann metric σ:

$$\epsilon \Delta_\sigma R_\sigma = -R_\sigma$$

or more generally still

$$R_{(1)\sigma} + \epsilon R_{(2)\sigma} = 0$$

But when would this be the case?

Recall the functional for an almost sharp geometry takes the following approximate form:

$$\int_M K_\sigma(\epsilon) dm := \int_M (R_{(1)\sigma} + \epsilon R_{(2)\sigma} + \epsilon^2 R_{(3)\sigma} + ...) dm$$

So this is the underlying functional driving *normal* low temperature superconductivity.

It is well known that the High-Tc superconductors tend to have a significantly more complex crystalline structure than the Low-Tc superconducters. Hence this suggests that one use more elaborate geometric structures as the basis of a High-Tc theory, in order to obtain a better theoretical understanding.

It is therefore quite natural to suggest that an almost sharp self-referential structure may contain the information we need to describe the process of superconductivity in High-Tc materials.

$$\int_M (T_{(1)\Lambda} + \epsilon T_{(2)\Lambda} + \epsilon^2 T_{(3)\Lambda} + ...) dm$$

In other words, not only do we have Cooper pairs propagating through such materials giving rise to the superconducting state, but these Cooper pairs interact with one another in a particularly well-structured way, to first order. I claim this is the key difference between Low-Tc and High-Tc materials.

The engineering question here of course is whether this is immediately applicable to the development of room temperature (Room-Tc) superconductors. The short answer is no - in fact a fully general theory would require some even deeper structural results. Nonetheless, from a practical point of view, it is possible that full generality is not needed.

Recall that we have three geometric operators at our disposal from before - ∂^*, $\wedge$, and $\star$. It is natural in this instance to consider $\wedge$, or reverb, since it is closely tied with the study of solid-state lattice dynamics.

Consequently, we are interested in the modified action

$$I(\Lambda,\Sigma,\epsilon) = \int_M \wedge(T_{(1)\Lambda}(m) + \epsilon T_{(2)\Lambda}(m) + ...; T_{\Sigma}(m))dm$$

It might also be instructive to take into account viscoplastic effects, as these are also closely related with solid-state physics. In this case the action of interest will be

$$I(\Lambda,\Sigma,\epsilon) = \int_M \wedge(T_{(1)\Lambda_1}(m)+\epsilon_1 T_{(2)\Lambda_1}(m)+...; T_{\Sigma_1}(m)) \wedge (T_{(1)\Lambda_2}(m)+\epsilon_2 T_{(2)\Lambda_2}(m)+...; T_{\Sigma_2}(m))dm$$

8.3 A digression into economics

The aim of this section will be to introduce the notion of a financial derivative as a form of generalised currency, and ultimately seek to find the appropriate strategy for their use by a financial institution. This is evidently quite a current topic, given recent economic events.

Of course as a trained mathematician, rather than a trained economist, I am in something of the position of an outsider here. However, it is possible that a fresh perspective may be what is required, or at least be valuable. Certainly there is an interesting connection between Information theory - the mining of information from a system, subject to certain codified rules (ie, a *game*) - and the use of this information to determine an optimal strategy to play. Trading in financial securities is an example of a particular form of game, and is the one that I will focus on in this section.

It is argued by Nicolas Taleb [Ta1] that in certain types of statistical distributions, highly improbable events can carry a great deal of weight. He makes an analogy with such distributions (compared to the standard, well behaved ones studied in the academy) as being symptomatic of fractal behaviour [Ta2]. This is to a certain extent in accordance with intuition and observation, for instance the way that in capitalist economies the distribution of wealth tends to organise itself according to a scale free, or Pareto type distribution pattern.

Since I am naturally interested in fractal dynamics, I suspect that there may be something that I might be able to say here. To be more precise, I believe that the key observation is that an economic market responds holistically to the trades due to particular players. Normally, these movements are negligible and subject to arbitrage. However, in the event that feedback loops are not damped, such as in the instance of an economic bubble, I suspect that it might be instructive to consider self-referential geometric drivers for appropriate understanding of the underlying risk distribution. These will lead to inherently more complex behaviour than might be predicted by more naive models.

So I claim that self-referential geometry might be useful into taking into account feedback loops within the economic cycle, and hence indicate appropriate risk-management strategies for the instances where the market is not damped, but rather driven by the collective actions of its players.

There is another observation that one can make of course, which is that banking did perfectly well before derivatives were developed. Consequently in the first subsection I argue as to why derivatives are useful (if used properly). The key difficulty is that many of the current strategies used to valuate the exposure to risk for these generalised currencies may in fact lack the necessary generality. Corollary to this is increased susceptibility to fallout from the deflation of bubbles in the valuation of particular assets.

A final remark is that in practice, of course, full generality is not needed to make reasonably optimal decisions. However we would like the stripped down versions of our equations to give an appropriate intuition as to which strategies are safe, and when.

8.3.1 Derivatives as generalised currency

Derivatives have their roots in the paper by Kenneth Arrow and Gerard Debreu in their paper "Existence of an Equilibrium in a Competitive Economy" [AD]. In particular in proving the main result of their paper they demonstrate that there will always be a form of abstract currency, or a contract that will pay 1 unit, presupposing a particular state of an economy occurs at a particular future time, and 0 units otherwise.

In fact it is possible to demonstrate that the existence of such contracts theoretically should enable an economy to function with a greater degree of optimality than in their absence.

However it is necessary to restrict our consideration to only a particular class of generalised currencies. It is important not to introduce too great a deal of generality, otherwise even first-order self-referential geometries will lack the abstraction to cope. Fortunately most of the forms of contracts used by traders are not subject to such pathological generalisation, so this is sufficient in practice.

First, a reminder of a fairly standard notion:

Definition 98. A *unit of currency* is a contract that one may trade for goods at a certain agreed rate of exchange within an economy. It is also possible to directly exchange goods and services for such contracts.

The key advantages of currency over a pure barter system is that it helps facilitate the optimal usage of goods and services within an economy. For instance, if person A wants a service from person C, they do not need to barter their commodities directly with C, but rather can exchange for contracts of currency with B in order to obtain the services of C. C then can use these contracts to obtain goods from B. Hence the generative capacity of an economy is better utilised via such a universal unit of exchange, which, in general, improves the standard of living of all participants.

There are weaknesses to the currency system, naturally. These are the perils of inflation (during times of boom), deflation (during times of economic contraction), differing rates of exchange with different currencies of other societies, leading to *currency markets*. Boom and bust are subject to the psychology and general level of confidence of the market participants. However the advantages far outweigh the disadvantages, at least in the long run.

Definition 99. A *commodity market* is a valuation of a particular good or service dependent on its level of demand or excess within a society, in terms of a standard baseline *currency*. It is possible to also have *currency markets*, which represent the rate of exchange between two different societies of their contracts, which will in general be different according to the varieties of commodities produced by each society.

Definition 100. An *economic state* is a point, or range of points in valuation space, for one or multiple commodity markets.

Definition 101. A unit of *First-Order Generalised Currency* is a contract that will pay a unit of *currency* provided that at some predetermined point or period of time in the future a particular economic state occurs. Since an economic state is a valuation in terms of the baseline currency, in this sense such contracts are *derivatives* of the underlying unit of exchange.

So in this sense a derivative (in the form that I will use it) is a generalised form of currency - it is a contract built on the willingness for people to exchange goods and services for currency contracts within an economy. Since the existence of a notion of currency helped improve the use of the plant capacity of an economy, we would expect, in sufficiently complex economies, for the introduction of derivatives to further potentially improve the use of industrial and productive capacity. This in fact was the result of the Arrow-Debreu paper.

For a less formal argument of why we might expect this to be the case, consider again a person A. I make an analogy with the raising of capital on the basis of expected return. Suppose A has certain skills and resources to hand, and wishes to build a factory, but does not have the funds to construct it. Then he might offer contracts to directly finance the project, for which he will pay dependent on the future success of his business. For example, shares are one particular instance of such contracts. These again will be derivatives, since they are based on the state of the commodities of A at some point in the future, rather than directly related to what A has currently.

Consequently we have the notion of *investment*: an institution B with a store of currency contracts from a group of customers C might exchange these for derivative contracts from A on the basis of expectation of a future return. It is in the interest of A to offer such contracts at a rate that it is attractive for B to buy them. Hence, in the existence of such a generalised currency, A builds the factory, B takes a share of the future profits, and the customers C take a share of the profits of B. Hence in this situation, at least, it is clear that the existence of such contracts provides the means to improve the operation of an economy, as otherwise the factory would not have been constructed.

There are natural perils to the use of such a generalised currency, of course. An institution B does not want to extend credit too easily; conversely, if A is dependent on the supply of credit, a loss of confidence in derivatives markets might cause A to go out of business since credit might become too expensive to obtain. So once again, as with baseline currency, there is a vulnerability to market confidence, which depends in turn on the psychology of its participants.

8.3.2 A standard introduction to Black-Scholes

To recapitulate, I have introduced the notion of generalised currency, and why one might expect it to be useful. However there remains the problem as to how to price derivatives in terms of the baseline currency, given existing information as to the state of the relevant commodity markets.

There have been various methods used for such purposes over the years. Perhaps the most recent (and the most famous) is the Black-Scholes equation for options pricing.

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

where σ is the volatility of the underlying asset, V is the value of the derivative, t is time, r is the rate of interest, and S is the value of the underlying asset.

It is essentially a heat equation. There are various standard ad-hoc and intuitive arguments as to how to derive it in the literature; there is also a large (and growing) number of papers detailing its deficiencies. I will not focus on the latter, for the time being, but rather give a standard argument following the treatment of [Wt].

First, as a quick observation, it is easy to see why the value of an option will increase over time by rV, since this is the growth of the value of the derivative due to its defined interest rate (which presumably is assumed to be constant). So, ignoring the last term on the right hand side, I will focus on the remaining two.

Note that if we make the variable substitution $T = log(S)$, then $\frac{\partial}{\partial T} = S\frac{\partial}{\partial S}$. Consequently the two terms simplify to

$$\frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial T^2} + r\frac{\partial V}{\partial T}$$

In other words, it is the log-change of the value of the underlying asset that matters to the value of the derivative. Why this should be the case is intuitively not entirely obvious, but we can think of this geometrically in the following way. Suppose M is the space of variables which influences the value of particular commodities. Let $S : M \to R$ be the value of a particular commodity. Then suppose we wish to construct a derivative for the same commodity; this will be a function V from the tangent bundle TM, to R. Consequently to map S to the tangent bundle to express it in the same units, we must use the inverse of the exponential map, or the logarithmic function. Since the image of S lives in the reals, this will be the standard logarithmic function, QED.

Now, it is of interest to ask why first of all the increase in value of the derivative over time should decrease as $r\frac{\partial V}{\partial log(S)}$. The reason for this is intuitively as follows. One wishes to decrease the price of a derivative if the value of the underlying commodity undergoes an upward movement relative to the present value, since the consumer has the opportunity instead of selling the commodity direct and investing the capital in an interest fund or vehicle. Consequently the potential for arbitrage decreases if this term is positive, and the decrease will be proportional to the interest rate.

The final term, $\sigma^2 \frac{\partial^2 V}{\partial T^2}$, the volatility term, is perhaps the most interesting. Essentially this states that if the acceleration in price of the underlying asset is positive, the price of the corresponding offered derivative should be decreased proportional to the volatility in price of the underlying asset. This is a "risk" term; the underlying risk here is the price movement of the underlying asset, which will depend directly on its volatility - that is, the standard deviation of the statistical distribution of likely outcomes of time evolution of the value of the asset, which for the sake of argument I will assume is normally distributed (though this is a *gross* oversimplication in most instances, of course).

If the value of the asset is accelerating, the value of the derivative decreases, for the reason that there is the *risk* that the value will jump statistically higher than the average rate of increase. Consequently it is better in this instance for an investor to hold on to the underlying asset and sell at the end of the period, if the price of the derivative is not changed to compensate. The degree of the risk will depend on the volatility of the price of the asset, in particular, the variance; consequently the value of the derivative will decrease proportional to the variance by the acceleration.

Subject to the structural restrictions within which it is constructed, these are the only considerations that matter- this indeed is the genius of the Black-Scholes analysis. The fallacy in this instance is instead that of reification; the structural restrictions which have been applied do not actually represent the messy complexity of reality. I will not pretend that the techniques that I apply in the following sections are complete or the final word; what I do claim is that potentially they tell a more complete tale of the behaviours underlying these processes, in the context of 1-category theoretic terms.

8.3.3 Information theory and strategic pricing

Since, as mentioned before, trading is essentially a game played against the other players in the market - and not necessarily a zero-sum game, but even potentially a cooperative one - it stands to reason that there should be some basis for an optimal strategy based on the information that we have about the state of the system. Consequently, we might expect to be able to express data about the markets as a form of information functional, and consequently find the optimal strategy to proceed towards trying to extremise said information.

Following this vague intuition, I will now give an alternative derivation of Black-Scholes, grounded within information theory.

Recall the Black-Scholes equation:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0$$

Suppose now we have n assets as the basis of our option. The price of various assets will depend on the prices of all the others, since generally the cost of producing a commodity might depend on the resource available to either produce or consume it. Consequently, there is a metric T that acts on the space of commodities M such that $T_{ij}(S,t)$ is a correlation metric that represents the correlation between commodities i and j at time t and commodity vector price S.

In particular it is quite natural to expect the price of such options to behave according to the equation:

$$\frac{\partial V}{\partial t} + \frac{\partial_T^2 V}{\partial_T S^2} - rV = 0$$

where now $\partial_{T,i} = T_{ij}\partial_i$.

If we view interest itself as a commodity, we see that this lacks the necessary generality and will simplify to the expression

$$\frac{\partial V}{\partial t} + \frac{\partial_{\hat{T}}^2 V}{\partial_{\hat{T}} \hat{S}^2} = 0$$

Now if we extend $\hat{T}$ over parameter space $\{S,\text{interest }\}$ to a suitable metric σ including time, we see that the above equation simplifies again to

$$\Delta_\sigma V(m) = \frac{\partial_\sigma^2 V}{\partial_\sigma m^2} = 0$$

Subject to a deformation now of σ and V, by partially absorbing V into σ to leave a function $\bar{V}$, we can arrange now for $\bar{V}$ to be the volume functional for a new metric τ, such that

$$\Delta_\tau \bar{V}(m) = 0$$

but the operation of a metric Laplacian τ on its volume form is just R_τ; consequently we have that $R_\tau(m) = 0$, where m is the parameter space for the market, and τ represents the state information.

But now recall that the fisher information functional for a Riemannian manifold is just

$$\int_M R_\tau(m)dm = 0$$

which is critical precisely when $R_\tau = 0$. Consequently we have that the Black-Scholes pricing model is a consequence of assuming that the structure of the market is adequetely described through an appropriate commodity correlation metric τ over state space M, measuring the associated information, and requiring that it be critical.

In other words it is a means of optimal extraction of information, subject to certain structural assumptions. The key assumption is that a particular point in state space m does not directly affect points n which are not contiguous with respect to a standard distance function, such as induced by a Riemannian metric. This is the weak point of the Black-Scholes analysis.

8.3.4 Self-referential economic dynamics and application to valuation of derivative risk

Suppose as before that we have an option defined over an economic state space M, with correlation metric τ. But now suppose that τ_{ij} depends on τ_{kl}, ie that the correlations are correlated. Then naively we need a tensor Λ_{ijkl} to provide us the information of the nature of the correlation. However, as before, there are three different ways the data might self-correlate, so we need an eight tensor to specify in full the way τ self-correlates according to the three different possible varieties of structure.

Consequently to describe the economic indicators specifying the behaviour of our option appropriately we need an 8-tensor Σ defined over our state space M. We can then measure the associated information, and require this to be critical:

$$T_\Sigma(m) = 0$$

Then I claim using this as a basis for pricing an option, or first order generalised currency, will avoid the dangers of first order self-correlation within the markets.

It is of some interest to determine what this means in practice. In other words, we would like to strip the above expression down to a form that would be easy to use. The first step in the process of simplification is to break Σ up into four metrics - σ, the correlation metric for the economic state space, $\tau_\wedge$, $\tau_\star$, τ_{∂^*}.

The standard Black-Scholes equations can be written roughly as

$$\frac{\partial_\sigma^2 V}{\partial_\sigma S^2} = 0$$

Taking into account the $\star$ structure, we have

$$\frac{\partial_\sigma^2 \partial_{\tau_\star}^2}{\partial_\sigma S^4} V = 0$$

and adding the $\wedge$ dynamic:

$$\frac{\partial^2_{\wedge(\sigma;\frac{\partial^2_{\tau_\wedge}}{\partial S^2})} \partial^2_{\tau_\star}}{\partial_{\wedge(\sigma;\frac{\partial^2_{\tau_\wedge}}{\partial S^2})} S^4} V = 0$$

and finally the ∂^* structure:

$$\partial^*\left(\frac{\partial^2_{\wedge(\sigma;\frac{\partial^2_{\tau_\wedge}}{\partial S^2})} \partial^2_{\tau_\star}}{\partial_{\wedge(\sigma;\frac{\partial^2_{\tau_\wedge}}{\partial S^2})} S^4}; \frac{\partial^2_{\tau_{\partial^*}}}{\partial S^2}\right) V = 0$$

So essentially we have an eighth order differential operator acting on the pricing of our option. I will now attempt to provide some intuition for what the various forms of structure effectively mean economically.

It makes sense to view $\tau_\wedge$ as being a characteristic of the distribution of capital available to various players. In a normal economic system, the wealth distribution will not be uniform, but will follow some form of power law. From this viewpoint $\tau_\wedge$ details how the wealth distribution responds to various movements in the state space of the underlying assets.

For $\tau_\star$, this might well represent the resistance (or lack thereof) of movements within the state space of a market to further trading.

Finally, τ_{∂^*} could be viewed as an indicator of consumer confidence, or general volatility, given a particular state of the market. In particular situations where confidence is subject to high variability, this can lead to dynamics such as "market turbulence", wherein the pricing of an option might become highly volatile.

So this provides some indication of how better strategies for the valuation of options might be constructed, in the situation that feedback loops in the market are not damped. Note however that this is limited to first order generalised currencies, that is, derivatives built directly from an underlying asset, rather than derivatives built on top of generalised currencies. Furthermore, whereas in most situations the above considerations might be sufficient for appropriate risk valuation, these do not take into account higher order self-correlations in the market. Therefore a certain degree of caution must be taken in not mistaking this picture as representing the market reality.

However it is reasonable to suppose that for practical purposes, the above considerations are enough to construct appropriate strategies for options pricing, under a wider range of market conditions than standard Black-Scholes.

8.4 Applications to Pure Mathematics

8.4.1 Travelling Salesman Problem

The travelling salesman problem is a key unsolved problem in logistics. Primitively stated, it asks, given n points p_i in a space M, what the shortest sequence of geodesics γ_i that visits all the p_i in turn is.

It has been suggested that this problem might be amenable to variational methods, since if there is an elegant method to solve it, variational analysis would be a prime candidate. It is to be noted, though, that nobody has been able to find a means to make progress on this problem by such means to date.

However, we now have at our disposal a new marvelous machine of great generality; in particular, I claim that self-referential geometry may be of use here. I will now provide a sketch of how one might determine the appropriate sequence a_j such that nodes $i_1, ..., i_n$, satisfy the property that the geodesic sequence $i_{a_1} \mapsto i_{a_2} \mapsto ... \mapsto i_{a_n}$ is the shortest possible sequence.

Note that we may assume we have complete information of the matrix of geodesic paths γ between the nodes as a starting point for our analysis. Also, WLOG, we may assume that we know the metric σ on the space M wherein the nodes are embedded, and that it is Riemannian.

Consider

$$\Theta^{\alpha\beta\gamma\delta}_{ijklabcd} Y_{ijkl} Y_{abcd} Y_{\alpha\beta\gamma\delta} = 0$$

where $Y_{abcd}(t)$ is to be interpreted as $\partial_{ab}\gamma_{cd}(t)$, the ab component of the hessian of the cd path evaluated at time t. This in nothing other than the equation for a self-referential geodesic. Since we already know how the geodesics behave, we solve for Θ.

Now, for a critical self-referential system, we have that the information density

$$T_\Lambda = 0$$

Since we know Θ, we may solve for Λ.

But I claim that this is sufficient to solve our problem. In particular, if we again compute geodesics using Λ, we require them to solve

$$\nabla^\Lambda_{X_{abcd}} X_{ijkl} = 0$$

where $X_{abcd} = \partial_{ab}\gamma_{cd}$, where $\gamma_{cd} : I \times I \to M$ is a two parameter path, so that γ is a matrix of surfaces. Applying the constraints $\gamma_{00}(0,0) = i_1$, $\gamma_{ji}(1,s) = \gamma_{(j+1)i}(0,s)$, $\gamma_{ji}(t,1) = \gamma_{j(i+1)}(t,0)$, $\gamma_{nn}(1,1) = i_1$ provides us with enough information to determine a node sequence. This is done by diagonalising $\gamma(0,0)$, and reading off the vector $diag_i(0,0)$.[1]

Then I claim that this will be the sequence we are looking for.

Remark. The most non-trivial part of this analysis is the optimisation step, which allows us to solve for Λ. This relies on the deep structural theory developed in this dissertation. The rest is certainly quite difficult, but essentially quite mechanical.

Remark. In standard instances of this problem, which tend to be of greatest familiarity in practice, $Y_{abcd} = 0$ since the geometry is flat and hence geodesics are straight lines. Consequently $\Theta = 0$, so optimisation is satisfied "trivially", and we merely have that Λ is a homogeneous solution to a PDE of 2nd order with eighteen terms.

Remark. For something in the way of physical interpretation of the eight tensor Λ in this situation, $\Lambda_{ijklabcd}(p)$ can be interpreted to be the interaction strength between the ab derivative of the geodesic extension of γ_{ij} with the cd derivative of the geodesic extension of γ_{kl} to p via parallel translation with respect to the standard Riemannian metric σ on our base space M.

[1] The way to think of this is a way of extending the computation of single geodesic paths with respect to the metric σ (which we implicitly knew from the beginning) to the computation of a sequence of geodesic surfaces simultaneously, *such that not only the surfaces, but the sequence is geodesic*, with respect to our self-referential structure Λ (which we have laboriously computed). Note that we expect, and in fact need Λ to be *compatible* with σ, so that there exists a mesh for the geodesic surfaces γ_{ij} of Λ such that the paths of the mesh are in fact geodesic paths with respect to σ. Of course this is guaranteed by our original computation of the Christoffel symbols for Λ from the geodesics for σ.

8.4.2 The four colour theorem

A closely related problem to that just discussed, and also quite well known, is that of the four-colourability of planar graphs. It is easy to see that this is in fact equivalent to the standard description of the result, which is that every plane divided into regions is four-colourable.

The problem held the status of a conjecture from 1854 when it was first proposed, until it was resolved via the method of computer-assisted proof in the late 80s.

The question still remains, however, as to how the problem might be resolved in a more elegant fashion, without such brute force proof techniques. (The original proof of the conjecture ran to close to half a thousand pages of computer calculation).

I propose the means towards such a solution here, again using the self-referential geometry to indicate a possible approach. It is natural to consider the self-referential geometry, since we are considering each node p in a planar graph G rel its neighbouring nodes.

This is not quite as strong as travelling salesman, since we are only interested in the local planar graphs in G about each node being self-referential. Since we are considering only local structures, I claim that this is equivalent to placing an almost sharp self-referential structure on all of G, since we are subject to a degree of uncertainty regarding the colouring of the neighbourhood of each of the neighbouring nodes of a base node p, given a colouring about p - which in turn may necessitate a change in the colouring information at p. Consequently we have an information interplay between neighbouring local graphs in G, that might not necessarily be commutative. This leads one to consider quantum groups, and, consequently, almost sharp structures.

To be more specific as to why an almost sharp self-referential structure is sufficient, note that the operator $T_{(2)\Lambda}$ will be trivial for G iff, given a node p with neighbour q, with neighbouring nodes p_i, q_i respectively, there exists a p_i and a q_j that are neighbours with the same colour.

Then it is clear to see that we need at least four colours in order to have a non-trivial almost sharp self-referential structure, since otherwise $T_{(2)\Lambda} = 0$. Conversely, it is clear to see that, via calculation on the variation of dimension, that the number of colours will be as few as possible provided that the structures are still well defined, ie $T_{(2)\Lambda} \neq 0$. This completes the sketch of the conjecture.

8.4.3 The Goldbach Conjecture

This particular open question, whether any even number can be represented as the sum of two primes, is one of the oldest problems in mathematics. The first recorded mention of it in the literature is a letter from Christian Goldbach, written to Leonhard Euler in June, 1742, and so it is consequently usually attributed to the fellow. However it is a problem of such deceptive simplicity that it is certainly quite possible various philosophers were aware of this curious apparent property of the naturals prior to Christian's written observation.

The purpose of this section will be to indicate a potential line of attack towards the resolution of this problem, via the device of the self referential calculus.

To begin, one might expect the problem to be related to the twin prime conjecture - that for every odd prime p either $p-2$ or $p+2$ is also prime. Furthermore we expect the first to be induced by a functional principle on the measure, and the second to the corresponding functional principle on the stack. Consequently I will restrict myself to consideration of the former.

It is clear that, if true, this particular structural conjecture is a criticality result on 2-tuples of critical subsets of the plane, in an analogous manner to how the RH is induced by a criticality result on singletons within critical subsets of the line. Since primes are, by definition, the smallest set of integers which generates all the naturals by multiplication, they are critical - and the product of a critical set with itself will also be critical.

Consequently, since we are interested in the interaction of the set of primes with itself, we are naturally led to consider the application of our information theory for self-referential geometries, which will be represented by maps from $R \times R \to R$. In particular, we would like to consider a fully general self-referential signal function over the real plane, and view the analytic extension to $C \times C$.

In fact it is quite easy to demonstrate via similar proof techniques to those used in [Go] that we get

$$\int_{(Re(z+w)\geq 1)} \Sigma_{n,m=1}^{\infty} \frac{a_{nm}}{n^z m^w} dz dw = 0$$

where $a_{nm} \in l^{\infty}(N \times N)$.

Proof. (Sketch).

To prove this statement, we start by considering the information density

$$\partial_\Sigma(T_\Sigma; T_\Lambda)$$

and require it to be critical, for two separate self-referential structures Λ and Σ; by the correspondence principle this is equivalent to consideration of a general signal function over a self-referential statistical superstructure.

Set $\nabla := \partial_z + \partial_w + \partial_a + \partial_b + \partial_c + \partial_d$.

Then, by the Cramer-Rao inequality, we conclude that

$$T_\Sigma = \frac{1}{\Sigma^4} \star \wedge \partial^*(\nabla; \nabla; \nabla; \nabla)^2 ln(\Sigma) = 0,$$

and also that as $T_\Lambda = 0$ that the second statistical contribution is trivial, since Λ is constant.

Hence we obtain $\star \wedge \partial^*(\nabla; \nabla; \nabla; \nabla)^2 ln(\Sigma) = 0$, and consequently $\nabla^2 ln(\Sigma) = 0$. Then $(\partial_z + \partial_w + 2(\partial_a + \partial_b))^2 ln(\Sigma) = 0$ as Λ is constant, so that $\Sigma(z, w) = H(2(z + w) - (a + b))e^{A(a,b)z+B(a,b)w+C(a,b)}$. Using analyticity, we obtain the above result. □

Naturally it is also important to convert the statement of the problem into a criticality result that can be compared with the results of these methods. In particular we are interested in showing that it is equivalent to the above.

Now, it is well known that

$$\Sigma_{n=1}^{\infty} \frac{1}{n^z} = \Pi_{p \text{ prime}} \frac{1}{1-p^{-z}}$$

- the connection between the standard zeta function and the prime numbers.

Consequently we have

$$\Sigma_{n,m=1}^{\infty} \frac{1}{n^z m^w} = \Pi_{p,q \text{ prime}} \frac{1}{(1-p^{-z})(1-q^{-w})}$$

If we now rewrite the righthand side with the conditional that $p + q = 2k$, we have that for an arbitrary $k \in N$, there must exist a sequence a_{nm} such that the following equality holds:

$$\Sigma_{n,m=1}^{\infty} \frac{a_{nm}}{n^z m^w} = \Pi_{p,q \text{ prime}|p+q=2k} \frac{1}{(1-p^{-z})(1-q^{-w})}$$

But now via the criticality result for this sequence a_{nm} we have that the integral is zero; consequently

$$\int_{Re(z+w)\geq 1} \Pi_{p,q \text{ prime}|p+q=2k} \frac{1}{(1-p^{-z})(1-q^{-w})} dzdw = 0$$

Now suppose there does not exist a pair p, q such that $p + q = 2k$. Then the product is trivially the identity. But then we have that the above integral is infinite, which is in contradiction with the result that we have from criticality, which is that it must also be zero. Hence our assumption that such a pair p, q did not exist led us to a contradiction, and we have for each $k \in N$ that there exist primes p, q with $p + q = 2k$, which is what we wanted to demonstrate.

However, this argument does not work; it is circular. For it requires that each coefficient a_{nm} be finite, which depends on the existence of a pair p, q in the first place.

Consequently we need to use stronger information.

In particular we threw away some of the structure before in the statement of our first result. We would like to get more control over the coefficients a_{nm}, using this structure.

Recall

$$\star \wedge \partial^*(\nabla; \nabla; \nabla; \nabla)^2 ln(\Sigma) = 0$$

I claim that this is strong enough to establish the result

$$\int_{Aut(C^2)\times Aut(C^2), Re(\Sigma^2_{\alpha,\beta=1} X_{\alpha\beta})\geq 1} \Sigma^{\infty}_{i,j,k,l=1} \frac{a_{ijkl}}{n_{ij}^{X_{pq}} m_{kl}^{X_{rs}}} dX_{pq} dX_{rs} = 0$$

which I claim in turn has sufficient resolution for our purposes. Here $Aut(C^2) := \{C^2 \to C^2\}$ is the set of maps for the square of the complex numbers $C \times C$ back to itself, n, m are now indexed in turn by naturals i, j, k, l, and $\{X_{pq}\}$ is a basis for $Aut(C^2)$. It goes without saying that by the inequality $Re(\Sigma^2_{\alpha,\beta=1} X_{\alpha\beta}) \geq 1$ I am assuming that this holds for all values $X_{\alpha\beta}(z, w)$ of the functions $X_{\alpha\beta}$.

This leaves a two-fold task in hand - a careful derivation of the above identity, and its application to the question of interest. However, this is where I shall leave this discussion, at least for the time being.

Chapter 9

Conclusions

9.1 Combinatorics and Deeper Structure

Naturally the story progressively becomes more intricate beyond this point.

A deeper study of the area suggests that we examine the operators $\circ^*, \star, \wedge$ as before and also $\circ^{*2}, \star^2, \wedge^2, \circ^*\star, \circ^*\wedge, \star\circ^*, \star\wedge, \wedge\circ^*$, and $\wedge\star$. These operate in a natural manner on triples of signal functions. For instance

$$\star \wedge (f; g; h) := f^g h$$

There is also the new operator $\wedge_2$, which acts in the following manner on duples of signal functions:

$$\wedge_2(f; g) := f \wedge ... \wedge f$$

where g copies of $\wedge$ are taken.

So if the base geometry has a four tensor Λ, there are 12 possible ways to build structures using the 4 duple operators, leading to the necessity for an invariant of dimension 48 or greater. If the single duple operators are taken into account, this leads to necessity for an invariant of dimension greater than 60. Finally if the base is taken into account, this leads to an invariant of dimension 64; which happens to be the dimension of operators from $TM \to TM \to TM$ locally, if M is four dimensional.

But from before there are three different ways to build four tensors. So this means that an eight tensor base is required, leading to a 128 dimension invariant over a 64 dimensional geometry, which does not make much sense. Ultimately this leads one to the conclusion that maps from $\{TM \to TM\}$ to $\{TM \to TM\}$ are the natural ones to consider. This is a dimension 256 geometry, if M is 4 dimensional.

However again this lacks sufficient generality. For there is another new operator $\circ_2^*$, defined in the following way:

$$\circ_2^*(f;g) := \circ^*(f;\circ^*(f;...\circ^*(f;f)))$$

where g copies of $\circ^*$ are taken.

There is no new operator for $\star$, since $\wedge = \star_2$ by definition.

Consequently, we are led to 5 duple operators, or "meta-seals of existence", which act on triples of signal functions in 20 different ways; add the three single duple operators, together with the base, and we have 24 different aspects to take into consideration, leading to the necessity of a 192 dimensional invariant - and consequently, for quantum mechanics, a 384 dimensional geometry.

We obtain this increased dimensionality by observing that we previously did not take into account double composition in $Aut(TM)$, ie $TM \to_\phi TM \to_\psi TM$. If this is instanced together with the observation that this can occur either to the "left" or "right" of elements of $AutAut(TM)$, if M is four dimensional we obtain 128 additional degrees of freedom.

Even this, however, is not nearly enough. As observed in the fourth chapter of this work, it is necessary to consider 2^{**}-categories in order to obtain the necessary full generality, rather than 2-categories as above. In this case I suspect one has a 256 dimensional invariant for the theory, with 2^{**}-quantum mechanics requiring at least 512 or 1024 degrees of freedom.

To sketch roughly how the 2^{**}-categories (or enriched 2-categories) add addtional structure to the theory, I will need to make a brief digression to talk about Jet Bundles, attributed to Charles Ehresmann (1905-1979) [Ehr]. Recall that to every differentiable manifold M, one can associate a tangent bundle, TM. But TM is also a manifold; consequently it has a tangent bundle, $T^2M := TTM$. Iterating this process, we see that we can define objects T^rM for any natural number r.

Take the limit as $r \to \infty$. Then we have an object $\mathcal{T}M$, with coordinates $(m, v_1, v_2, v_3, ...)$, or equivalently (m, v), where v lives in the space of infinite matrices. This is known as the *Jet Bundle* associated to M.

We expect tensor structures on $\mathcal{T}M$ to have additional structure than with ones on TM, since they have an additional degree of freedom with which to act. Indeed, for full generality one requires tensors of double the rank to describe a similar level of resolution for physical processes.

Returning to our question, note that the left and right action of $Aut(M)$ on $AutAut(M)$ is intuitively a relationship between tangent spaces. Consequently for full generality, in order to deal with the surgery of tangent space to tangent space under the action, one needs to consider the Jet Bundle. Recall now that the original description of the right or left action required a rank 64 tensor. From our observations above we see that for full generality we need a tensor of rank 128. This provides us with the detail that we need.

In such a situation of generality it might seem next to impossible to compute anything. However, it is worth mentioning that the computation of the Christoffel symbols for a self-referential geometry can be viewed as the action of a form of finite discrete group with two generators, and two relations, on the indices of the relevant tensor. Consequently it makes sense to view the space of indices associated to high rank tensors as finite groups, and use representation theory to aid computation of the associated structural invariants.

The Haar integral might be relevant, though I am not certain about this.

It might then be useful to apply the considerable machinery of discrete mathematics, representation theory, lie algebras and geometric group theory in order to make further progress.

9.2 Potential Engineering Applications

Alternatively, instead of stepping into deeper realms of abstraction, one might choose instead to look back towards an applied physics / engineering direction. There are three particular applications of the self-referential theory that I can see might be interesting to pursue. These are, artificial intelligence (which needs no further explanation or motivation), possible application to the construction of a *Casimir pump* (a new type of battery), and possible application towards development of a "hyperdrive", ie a new means of propulsion for interstellar spacecraft.

A *Casimir pump*, for its namesake phenomenon the *Casimir effect*, is essentially a heat pump that pumps energy from the space of automorphisms associated to a spacetime manifold M (the "hyperspace") and directs it to M. The rationale for why self-referential calculus might be applicable here is that the Casimir effect is inherently quantum mechanical; consequently we expect the higher "self-referential" quantum mechanics to perhaps be enabling in this respect.

The picture to bear in mind is that of *activation energy*, commonly associated with biochemistry (see eg "Essential Cell Biology", by Alberts, Bray and Walter, p85). One might sometimes expect there to be a reservoir of energy sitting at a deeper information level, but inhibited from sifting through to the lower information level where one requires it by an activation energy "gap". In order to access it, one needs some sort of catalytic mechanism (to lower the gap), together with a certain amount of input, in order to overcome the activation energy, and release (some of) the heat in the reservoir to the lower information channel.

To paint an even more simplistic example, suppose one is in a cold house (physical space, M) and there is an "outside" ($Aut(M)$) that has been warmed by the morning sun. Then if one draws open the blinds covering a window, heat is free to flow through the glass and enter the house.

For the new methods of propulsion, I believe again the self-referential calculus will be useful - in particular the Maldacena action $\int_M T_\Lambda(m)dm$ associated to a self-referentiable 8-tensor structure Λ on M, where T is the natural information invariant, as described in Chapter 8. In particular it would be interesting and possibly worthwhile at this stage to attempt to sketch the principles required to propel a craft through $Aut(M)$ by exploring the consequences of the associated information dynamics - to the point where then applied physicists and engineers might then be able to turn these ideas into hardware.

I will paint some numbers to motivate the above. Suppose the speed of causal effects in $Aut(M)$ scales as the square of that in ordinary space. Suppose one furthermore is interested in travelling, say, 4 light years - the distance from Sol to the nearest star system. Then if the speed limit scales as the square, and if the amount of energy to reach a similar fraction of the limiting causal speed compared with that in ordinary space is of a similar order of magnitude, then we expect the equivalent distance to travelling in ordinary space to be $4/c$; or roughly 30000km. For reference, the distance from the earth to the moon is roughly 10 times this. Since the Apollo missions took roughly 5 days to traverse this distance, one might expect, with a similar amount of thrust traversing through the higher information channel, the time required to certainly be no more than 2 months, and possibly less than 10 days after reaching cruising velocity.

It is perhaps a sensible question to ask why I choose to sully what otherwise might be a perfectly good dissertation on matters pure and analytical with such speculative discussion. The answer is fairly simple - solutions are needed, particularly in the modern day and age, to pressing issues regarding the sustainability of modern living, and it is as much the duty of a scientist to look for solutions as it is for the scientist to guard their thinking against untempered thought.

To be more specific, I spoke to a distinguished gentleman (who shall remain nameless) a year or so ago and he pointed out that with the prevalence of television soap operas (and doubtless various other media), everybody everywhere was being encouraged to seek the New York city lifestyle. It goes without saying that, without a significant increase in tech level, this will be completely impossible to achieve, given resource limitations. Furthermore, it is possible that in the efforts of the growing world population to uniformly reach this goal, there may well be consequences that may not be totally desirable. Regardless, it is my assessment that if the tech level is advanced sufficiently beyond its current state, then it might be possible to avert, or at least mitigate, these eventualities.

Certainly a new source of (potentially) practically limitless energy would be extremely valuable; we would be able to switch to a zero carbon economy almost overnight. Artificial intelligence has interesting applications to cybernetic enhancement, and may ultimately universally enable potentially transformative effects in medicine and quality of life. Advanced propulsion technologies associated to automorphism spaces have the potential to extend the sphere of human influence beyond Sol, and to provide resources that our civilisation needs to grow and maintain our standard of living - such as from the asteroid belt.

It is possible that the ideas that I have outlined here, and my general approach, may be fundamentally flawed in various ways. Nonetheless, I am of the general belief that it is quite possible that the content of this dissertation might prove a beneficial source of inspiration to future scientists seeking an implementation of the aforementioned engineering solutions.

9.3 Concluding Remarks

In the first dissertation on these matters, I examined in some detail the structural details of particular exotic structures on differentiable manifolds. In the sequel I have studied the geometric resolution of the associated branching processes, culminating in the examination of first order self-referential geometry.

It is probably quite reasonable to posit that Riemannian geometry is a first order theory, whereas self-referential geometry is a second order theory. Consequently, it follows that the theory developed in these treatises only scratches the surface of what is possible in terms of structural results.

There are various areas of study that are related, and are certainly of no greater complexity than those studied by myself so far. In particular, there is a natural duality between many of these ideas and statistical theory, which I began to develop in the fourth chapter of this document. This might be worth investigating further. Again, as before, I mention the equivalence between Heisenberg's matrix mechanics - a fundamentally statistical construction, and Schrödinger's quantum mechanics - a fundamentally analytic construction.

As to additional questions which might motivate further enquiry I will suggest a few.

One is the three body problem, and chaos. Closely related is the area of mathematics known as catastrophe theory - dealing with abrupt changes in a system beyond a certain tipping point. For instance, to build an analogy with materials science, under stress a material behaves elastically up to a particular point. Beyond this it undergoes plastic deformation. However, if too much stress is applied, the material will fracture or break. So the mathematics of the associated processes might be interesting and worthwhile to study; in particular a third order approach might be useful here.

In graph theory, the Hadwiger conjecture [W8] might be amenable to third order methods - that is, methods based on the structural study of 2^{**}-categories.

Also, even though the second order approach might already be useful to a certain extent, a third order theory would certainly provide the enabling knowledge to design extremely rapid forms of spaceflight, and also provide the means to construct ultrafast communications networks. Furthermore, it is possible that control of even more exotic phenomena might follow via the understanding of critical third order structures.

There are also numerous problems in number theory that deal with particular criticality properties of three-tuples of natural numbers, such as the abc-conjecture. For this last problem, I will give an indication of what I think is required, in the form of a conjecture.

Conjecture. *(2-Aut Criticality). Let $Aut(X)$ be the set of automorphisms on X. Define $Aut^k(X)$ as the product $Aut(X) \times ... \times Aut(X)$, with k copies. Consider the arbitrary sequence $A_{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}$, with $\bar{\alpha} = \alpha_1 ... \alpha_4$, $\alpha_i = \alpha_{i_1 i_2 i_3 i_4}$ and similarly for β, γ, δ, such that the sum over all indices from 1 to ∞ is finite. Furthermore, consider arbitrary sequences $B_{1,\bar{\alpha}}$, $B_{2,\bar{\beta}}$, $B_{3,\bar{\gamma}}$, $B_{4,\bar{\delta}}$. Then I claim that the following equality is satisfied (subject to some uncertainty as to the appropriate bound for the domain of integration):*

$$\int_{Aut^4 Aut^4(C^4), Re(\Sigma_{\bar{i}=1}^{\infty} X_{\bar{i}}) \geq 3/2} \Sigma_{\bar{\alpha},\bar{\beta},\bar{\gamma},\bar{\delta}=1}^{\infty} \frac{A_{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}}{B_{1,\bar{\alpha}}^{X_{\bar{a}}} B_{2,\bar{\beta}}^{X_{\bar{b}}} B_{3,\bar{\gamma}}^{X_{\bar{c}}} B_{4,\bar{\delta}}^{X_{\bar{d}}}} dX_{\bar{a}} dX_{\bar{b}} dX_{\bar{c}} dX_{\bar{d}} = 0$$

where $X_{\bar{a}} := X_{a_1 a_2 a_3 a_4}$ *is an element of* $Aut^4 Aut^4(C^4)$*, such that* $a_i = a_{i_1 i_2 i_3 i_4}$ *are naturals indexed by the* i_j*,* $j = 1, .., 4$*, and similarly for* $b, c,$ *and* d*.*

Of course given my remarks towards the end of the first section of this chapter, it is clear that this statement does not encode the full generality of the dynamics of C^4 viewed as an Aut-Aut space. Consequently one might need to search for the appropriate formulation with the full degree of resolution associated to the underlying structure, possibly through first obtaining a better understanding of the foundations of the theory.

Bibliography

[1] **Galois theory and Group theory**

[DV] Eduardo J. Dubuc and Constanza S. de la Vega, "Grothendieck's Galois Theory", arxiv:math/0009145v1 [math.CT], 14 Sep 2000

[Ed] Harold Edwards, "Galois Theory", Graduate Texts in Mathematics, Springer-Verlag, 2nd Edition 1993

[Gal] Evariste Galois, "Memoires sur les conditions de resolubilite des equations par radicaux", Journal des mathematiques pures et appliquees, 1846

[Mill] Chuck Miller, 3rd year Algebra course notes, 2003

[Mill2] Chuck Miller, "Combinatorial Group Theory", 7 March 2004.

[W3] Wikipedia, "Adjoint functor", `http://en.wikipedia.org/wiki/Adjoint_functors`, revision 16:26, 28 December 2009

[2] **Modular Forms, Number Theory, and Graph Theory**

[DS] Fred Diamond and Jerry Shurman, "A First Course in Modular Forms", Graduate Texts in Mathematics, Springer, 2005

[F] Edward Frenkel, "Gauge Theory and Langlands Duality", arxiv:0906.2747v1 [math.RT]

[Tao] Terence Tao, "Week 9: Riemann-Zeta function; prime number theorem; functional equation.", lecture notes for Math 3228, Semester 2, 2003 at the Australian National University - "Complex Analysis for Applications".

[W8] Wikipedia, "Hadwiger conjecture", revision 21:09, 17 December 2009

[3] **Algebraic Geometry and Schemes**

[Art] M. Artin, "Grothendieck Topologies", Harvard University, Department of Mathematics - Spring 1962

[AM] Michael Atiyah and Ian MacDonald, "Introduction To Commutative Algebra", Westview Press (February 20, 1994)

[Cns] Alain Connes, "Non-commutative Geometry", http://www.alainconnes.org/docs/book94bigpdf.pdf

[De] Pierre Deligne, "La conjecture de Weil. I", Inst. Hautes tudes Sci. Publ. Math. No. 43 (1974), 273307. "La conjecture de Weil : II", Publications Mathmatiques de l'IHS, 52 (1980), p. 137252

[EGA] Alexandre Grothendieck and Jean Dieudonne (1971) (French). "Elements de geometrie algebrique" (2nd ed.). Berlin; New York: Springer-Verlag. ISBN 3540051139

[EH] David Eisenbud and Joe Harris, "The Geometry of Schemes", New York Springer 2000

[GN] I. Gelfand and M. Naimark, "On the imbedding of normed rings into the ring of operators on a Hilbert space", Math. Sbornik 12 (2), 1943: 197217.

[Gro] Alexander Grothendieck, "Seminaire de geometrie algebrique", 1960-1969,

[Mil] John Milnor, "Morse theory", Ann. of Math. Studies 51, Princeton University Press 1963

[Mu] Daniel Murfet, various mathematical notes at "The Rising Sea" weblog, http://therisingsea.org/?page_id=3

[Kam] Joel Kamnitzer, "Algebraic geometry without prime ideals", blogpost at Secret Blogging Seminar, http://sbseminar.wordpress.com/2009/08/06/algebraic-geometry-without-prime-ideals/

[Lg] Serge Lang, "Algebra", Springer; 3rd edition (June 21, 2005)

[Qu] Quillen, D. (1969), "Rational homotopy theory", Annals of Math 90: 205295

[RH] Robin Hartshorne, "Algebraic Geometry", Springer-Verlag, 1977

[S] Edoardo Sernesi, "Deformations of Algebraic Schemes", Springer, 2006

[4] **K Theory and Surgery**

[AB] M. Atiyah and R. Bott, "On the periodicity theorem for complex vector bundles", Acta. Math. 112, 229-247 (1964)

[AS] Atiyah, Michael F.; Singer, Isadore M. (1963), "The Index of Elliptic Operators on Compact Manifolds", Bull. Amer. Math. Soc. 69: 322433

[At] Michael Atiyah - "K-theory", Lecture notes 1964

[Bott] R. Bott, "The stable homotopy of the classical groups", Ann. of Math. 70, 313-337 (1959)

[K] Max Karoubi, "K-theory - an introduction", http://people.math.jussieu.fr/~karoubi/K.book/MK.book.pdf, Springer-Verlag New York 1978

[Luck] Wolfgang Lück, "A Basic Introduction to Surgery Theory", Fachbereich Mathematik und Informatik, Westfälische Wilhelms-Universität Münster, October 27, 2004

[To] David Manuel Murrugarra Tomairo, "Bott Periodicity", Masters Thesis, Virginia Polytechnic Institute and State University, http://scholar.lib.vt.edu/theses/available/etd-05112007-172831/unrestricted/ThesisMS2.pdf, May 8, 2007

[Wall] C.T.C. Wall, "Surgery on Compact Manifolds", 2nd Edition, November 1998. Edited by A.A. Ranicki

[W4] Wikipedia, "L theory", http://en.wikipedia.org/wiki/L-theory, revision 13:12, 4 December 2008

[W5] Wikipedia, "Grothendieck-Hirzebruch-Riemann-Roch Theorem", http://en.wikipedia.org/wiki/Grothendieck_riemann_roch, revision 23:46, 3 December 2009

[W6] Wikipedia, "Todd class", http://en.wikipedia.org/wiki/Todd_genus, revision 20:10, 30 December 2009

[W7] Wikipedia, "Atiyah-Singer index theorem", http://en.wikipedia.org/wiki/Atiyah-Singer_index_theorem, revision 02:52, 14 November 2009

[5] **Differential Geometry and Mathematical Physics**

[Am1] Shun'ichi Amari, Hiroshi Nagaoka, "Methods of Information Geometry", Translations of mathematical monographs; v. 191, American Mathematical Society, 2000

[Am2] Shun'ichi Amari, "Information geometry on hierarchy of probability distributions", IEEE Transactions on Information Theory, 2001

[C] Elie Cartan,(1923), (1924) and (1925). "Sur les vari'et'es 'a connexion affine et la th'eorie de la relativit'e g'en'eralis'ee", Part I: Ann. Ec. Norm. 40: 325412 and ibid. 41: 125; Part II: ibid. 42: 1788; English transl. by A Magnon and A Ashtekar, "On manifolds with an affine connection and the theory of general relativity", Napoli: Bibliopolis (1986)

[Ca] Manfredo Perdigao do Carmo, "Riemannian Geometry", Birkhäuser 1992

[Ce] N.N. Cencov, "Statistical Decision Rules and Optimal Inference", American Mathematical Society 1982

[Connes] Alain Connes, "Noncommutative Geometry", Academic Press, San Diego, CA, 1994

[DW] P. Deligne, P. Etinghof, D. Freed, L. Jeffrey, D. Kazhdan, D. Morrison, and E. Witten, "Quantum Fields and Strings: A Course for Mathematicians", American Mathematical Society (June 1999)

[Ehr] Charles Ehresmann, "Introduction a la theorie des structures infinitesimales et des pseudo-groupes de Lie." Geometrie Differentielle, Colloq. Inter. du Centre Nat. de la Recherche Scientifique, Strasbourg, 1953, 97-127.

[Fr] B. Roy Frieden, "Physics from Fisher Information - A Unification", 2nd Edition, Cambridge University Press 1999

[Go] C. Goddard, "A treatise on Information Geometry", `http://www.vixra.org/abs/0908.0073`

[KV] Alexander Kahle and Alessandro Valentino, "T-duality and Differential K-Theory", arXiv:0912.2516v2 [math.KT]

[Mal] Juan Maldacena, "The Large N Limit of Superconformal Field Theories and Supergravity", `http://arxiv.org/abs/hep-th/9711200`

[Mo] G. D. Mostow, "Quasi-conformal mappings in n-space and the rigidity of the hyperbolic space forms", Publ. Math. IHES 34 (1968) 53-104.

[ON] Barrett O'Neill, "Semi-Riemannian Geometry - with applications to relativity", Academic Press Elsevier 1983

[R] B. Riemann, "On the hypotheses which lie at the base of geometry", 1854. Translation by William Klingdon Clifford, Nature, Vol. VIII. Nos. 183, 184, pp. 14-17,36,37. `http://www.emis.de/classics/Riemann/WKCGeom.pdf`

[Wh] John Archibald Wheeler and Robert W. Fuller, "Causality and Multiply Connected Space-Time", Phys. Rev. 128(1962), 919-929

[6] **Sieves, Integrability, and Statistics**

[201] Kostya Borovkov, "Probability Notes 620201 Semester 1 2008", `http://www.ms.unimelb.edu.au/~s620201/`

[202] Ray Watson, "Statistics Notes 620202 Semester 2 2004", `http://www.ms.unimelb.edu.au/~rayw/ms202/`

[AH] Andrew R. Hodge and Motohico Mulase, "Hitchin integrable systems, deformations of spectral curves, and KP-type equations", University of California (Davis) 2008

[Co] Alina Carmen Cojocaru and M. Ram Morty, "An introduction to sieve methods and their applications", London Mathematical Society Student Texts, Cambridge University Press 2006

[Das] Ashok Das, "Integrable Models", World Scientific Lecture Notes in Physics, Vol. 30 (1989)

[Don] Simon Donaldson, "Self-dual connections and the topology of smooth 4-manifolds", Bull. Amer. Math. Soc. Vol 8, No. 1 (1983), 81-83

[Fe] William Feller, "An Introduction to Probability Theory and Its Applications, Vol. 2", Wiley; 2 edition (January 1, 1991)

[Ggr] P.X. Gallagher, "The large sieve and probabilistic Galois theory", Analytic Number Theory, Proc. Symp. Pure Math., 24, pp91-101

[Gr] George Greaves, "Sieves in Number Theory", A series of modern surveys in mathematics Volume 43, Springer-Verlag Berlin Heidelberg 2001

[Gt1] James E. Grant, "The ADHM Construction and non-local symmetries of the self-dual Yang-Mills equations", arxiv:0812.2005v2 [math-ph]

[Gt2] James E. Grant, "Reducible connections and non-local symmetries of the self-dual Yang-Mills equations", arxiv:0812.2147v2 [math-ph]

[Hi] Nigel Hitchin, "Stable bundles and integrable systems", Duke Math. J. Volume 54, Number 1 (1987), 91-114.

[Hm] Glyn Harman, "Prime Detecting Sieves", chapter 1 at `http://press.princeton.edu/chapters/s8585.pdf`, London Mathematical Society Monographs, Princeton University Press 2006

[HP] Jens Hoppe, Q-Han Park, "Infinite Charge Algebra of Gravitational Instantons", arXiv:hep-th/9312020v1

[Iwan] Henryk Iwaniec, "Rosser's sieve", ACTA Arithmetica XXXVI, (1980), 171-203

[Liu] Yu-Ru Liu, "The Turan sieve and some of its applications", masters thesis, Queen's University - Canada, submitted July 1998.

[Mik] A.V. Mikhailov, "Integrability", Berlin : Springer, c2009.

[PB] Mark Pinsky and Björn Birnir, "Probability, Geometry and Integrable Systems", Mathematical Sciences Research Institute Publications 55, Cambridge University Press 2008

[SV] Glenn Shafer and Vladimir Vovk, "The origins and legacy of Kolmogorov's Grundbegriffe", The Game-Theoretic Probability and Finance Project, October 3, 2005

[SW] Graeme Segal and George Wilson, "Loop groups and equations of KdV type", Publications mathematiques de l'I.H.E.S., tome 61 (1985), pp5-65.

[W9] Wikipedia, "Kolmogorov's Inequality", revision 17:57, 30 November 2009

[Tao2] Terence Tao, "Why are solitons stable?", arXiv:0802.2408v2 [math.AP]

[Tao3] Terence Tao, "245A, Notes 5: Free probability", What's New academic weblog, 10th February 2010

[7] **Materials Science**

[Ha] P. Haupt, "Continuum Mechanics and Theory of Materials", 2nd Edition, Springer, 2002

[HR] Weimin Han and B. Daya Reddy, "Plasticity: Mathematical Theory and Numerical Analysis", Interdisciplinary applied mathematics, Springer, 1999

[Hu] Kerson Huang, "Statistical Mechanics", Second Edition, Wiley 1987

[Kt] Charles Kittel, "Introduction to Solid State Physics", 7th Edition, Wiley 1996

[8] **Category Theory and Foundations**

[Bol] Thomas Bolander, "Self reference and Logic", 22nd August 2005, `http://www2.imm.dtu.dk/~tb/essay.pdf`

[Cn] Cohen, P. J. (1966), "Set theory and the continuum hypothesis", Addison-Wesley.

[CR] G.S.H. Cruttwell and Michael A. Shulman, "A unified framework for generalised multicategories", preprint January 2009

[Gö] Kurt Gödel, "On formally undecidable propositions of Principia Mathematica and related systems I", 1931, translation available from `http://www.research.ibm.com/people/h/hirzel/papers/canon00-goedel.pdf`

[HS] Horst Herrlich and George E. Strecker, "Category Theory", Allyn and Bacon, 1973

[Lurie] Jacob Lurie, "Higher Topos Theory", Annals of Mathematics Studies No. 170, revision October 8, 2009

[Ms] Daniel Mathews, "Chord diagrams, topological quantum field theory, and the sutured Floer homology of solid tori", `http://arxiv.org/abs/0903.1453`, arXiv:0903.1453v1 [math.GT], 8th March 2009

[Pen] Roger Penrose, "Twistor algebra", J. Math. Phys. 8: 345.

[RG] Robert Goldblatt, "Topoi - the categorical analysis of logic", Elsevier 2nd Edition 1984

[Simp] Carlos T. Simpson, "Homotopy theory of higher categories", http://hal.archives-ouvertes.fr/hal-00449826/fr/, preprint

[Sp] David Speyer, "The technical part of Gödel's proof", blogpost at Secret Blogging Seminar, http://sbseminar.wordpress.com/2009/12/07/the-technical-part-of-godels-proof/

[Su] Patrick Suppes, "Axiomatic Set Theory", Dover Publications Inc. New York, 1972

[W1] Wikipedia, "Topos", http://en.wikipedia.org/wiki/Topos, revision 21:02, 17 November 2009

[W2] Wikipedia, "Grothendieck Topology", http://en.wikipedia.org/wiki/Site_(mathematics), revision 21:04, 18 December 2009

[9] **Economics**

[AD] Kenneth Arrow and Gerard Debreu, "Existence of an Equilibrium in a Competitive Economy", Econometrica, Vol. 22, No.3 (Jul 1954),pp265-290, http://courses100.googlepages.com/ArrowDebreu1954ExistenceofanEquilibr.pdf

[Aw] Kenneth Arrow, "The role of securities in the optimal allocation of risk-bearing", Econometrie, 1953

[BS] Black, F. and Scholes, M., "The pricing of options and corporate liabilities", J. Polit. Econ., 1973, 81, 637654.

[Sh1] Steven E. Shreve, "Stochastic Calculus for Finance I - The Binomial Asset Pricing Model", Springer 2004

[Sh2] Steven E. Shreve, "Stochastic Calculus for Finance II - Continuous Time Models", Springer 2008

[Ta1] Nassim Nicolas Taleb, "The Black Swan: The impact of the highly improbable", Random House, April 2007

[Ta2] Nassim Nicolas Taleb, "The Fourth Quadrant", http://www.edge.org/3rd_culture/taleb08/taleb08_index.html

[TD] N. Taleb and Emanuel Derman, "The illusions of dynamic replication", Quantitative Finance, Vol. 5, No. 4, August 2005, 323326

[Wt] Paul Wilmott, Sam Howison, Jeff Dewynne, "The Mathematics of Financial Derivatives - A Student Introduction", Cambridge University Press 2009

[10] **Expository and Reference**

[ADH] Michael Atiyah, Robbert Dijkgraaf, and Nigel Hitchin, "Geometry and Physics", Phil. Trans. R. Soc. A, 13 March 2010, vol. 368, no. 1914, 913-926

[CC] Shiing-Chen Chern and Claude Chevalley, "Elie Cartan and his mathematical work", Bull. Amer. Math. Soc. 58 (1952), 217-250.

[Comp] Timothy Gowers, June Barrow-Green and Imre Leader (editors), "The Princeton Companion to Mathematics", Princeton University Press 2008

[Fdv] L. D. Faddeev, "Modern Mathematical Physics: what it should be", arXiv:math-ph/0002018v2, 10 Feb 2000

[Gb] Philip Gibbs, "Event Symmetric Space-Time", vixra:0911.0042 [Quantum Gravity and String Theory], 16 Nov 2009

[JK] Jay Jorgenson and Steven G. Krantz, "The Mathematical Contributions of Serge Lang", Notices of the American Mathematical Society, Vol 54 No. 4, pp476-497

[Kha] Masoud Khalkhali, "Basic Noncommutative Geometry" - Introduction, EMS Series of Lectures in Mathematics, December 2009

[Pe] Roger Penrose, "The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics", Oxford University Press, November 9 1989, ISBN 0-198-51973-7 (first edition)

[Wig] Eugene Wigner, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences", Communications in Pure and Applied Mathematics, vol. 13, No. I (Feb 1960).

[Witt] E. Witten, "Magic, Mystery and Matrix", Notices of the American Mathematical Society, Vol 45 No. 9, pp 1124-1129