Analysis notes

Aidan Backus

December 2019

Contents

Ι	\mathbf{Pr}	eliminaries	2
1	Fun	ctional analysis	3
	1.1	Locally convex spaces	3
	1.2	Hilbert spaces	6
	1.3	Bochner integration	6
	1.4	Duality	11
	1.5	Vector lattices	13
	1.6	Positive Radon measures	14
	1.7	Baire categories	17
2	Con	nplex analysis	20
	2.1	Cauchy-Green formula	20
	2.2	Conformal mappings	22
	2.3	Approximation by polynomials	24
	2.4	Sheaves	26
	2.5	Subharmonicity	28
	2.6	Operator theory	30
II	D	ynamical systems	31
3	Elei	nentary dynamical systems	32
	3.1	Types of dynamical systems	32
	3.2	Properties of the irrational rotation	34
4	Erg	odic theory	37
	4.1	The mean ergodic theorem	37
	4.2	Pointwise ergodic theorem	41
	4.3	Ergodic systems	45
	4.4	Properties of ergodic transformations	47
	4.5	Mixing transformations	48
	4.6	The Hopf argument	54

5	Flov	vs on manifolds	59		
	5.1	Ergodic theorems for flows	59		
	5.2	Geodesic flows in hyperbolic space	60		
6	6 Topological dynamics				
	6.1	Recurrence in topological dynamics	68		
7	Con	npletely integrable systems	71		
	7.1	Symplectic geometry	71		
	7.2	Hamiltonian flows on symplectic spaces	73		
	7.3	Hamiltonian flows on symplectic manifolds	75		
	7.4	Complete integrability	76		
	7.5	The Liouville-Arnold-Jost theorem	77		
	7.6	The Toda lattice	81		
	7.7	The QR algorithm	88		
8	Con	nplex dynamics	91		
Ũ	8.1	Siegel's KAM theorem	91		
9	Ent	ropy and information theory	96		
Ŭ	9.1	Shannon's axioms of entropy	96		
	9.2	Conditional entropy	97		
	9.3	The entropy of a group action	98		
	9.4	Topological entropy	.01		
Π	Ι	Operator algebras 10)6		
10	Ban	ach algebras 1	07		
	10.1	The spectrum	.07		
	10.2	Ideals	.09		
	10.3	The holomorphic functional calculus 1	10		
11 C^* -algebras					
	11.1	Weights	11		
	11.2	The GNS construction	12		
	11.3	The \mathcal{B}^p spaces	14		
	11.4	Representation theory of groups	15		
	11.5	Compact operators	16		
12	Gen	erators and relations 1	21		
	12.1	Construction of maximally free algebras	21		
	12.2	Tensor products of C^* -algebras \ldots 1	.22		

13 Representation theory of locally compact groups	124
13.1 Noncommutative dynamical systems	124
13.2 Group actions on locally compact spaces	129
13.3 Semidirect products of groups	134
13.4 The Heisenberg commutation relations	135
13.5 Projective representations	136
14 Noncommutative geometry	140
14.1 Quantum tori	140
14.2 The 2-torus	140
14.3 Noncommutative smooth manifolds	142
14.4 Noncommutative vector bundles	144
	1 4 17
IV Complex analysis	14(
15 Holomorphy in several complex variables	148
15.1 Cauchy-Riemann equations	148
15.2 Basic properties	150
15.3 Plurisubharmonicity and domains of holomorphy	152
15.4 Hormander L^2 estimates $\ldots \ldots \ldots$	161
16 Multivariable holomorphic functional calculus	164
16.1 The Gelfand transform	164
16.2 The holomorphic functional calculus	168
17 Algebraic geometry	170
17.1 Schemes and varieties	170
17.2 Formal power series	171
17.2 Forman kernels	172
17.6 Deiginan kernens	173
	110
18 Line bundles over complex varieties	179
18.1 Holomorphic line bundles	180
18.2 Integration on Hermitian line bundles	183
18.3 Asymptotics for the Bergman kernel	184
18.4 Morphisms into projective space	193
18.5 Zworski's conjecture	195
V. Harmania analysis	100
v narmonic analysis	тар
19 Rearrangement-invariant spaces	197
19.1 Log-convexity for L^p norms $\ldots \ldots \ldots$	198
19.2 Lorentz norms	199

20	Decoupling theory	202
	20.1 Square-root cancellation	202
	20.2 Decoupling estimates	204
21	Pseudodifferential calculus	211
	21.1 The Kohn-Nirenberg quantization	211
	21.2 The Calderon-Vallaincourt theorem	214
	21.3 Composition estimates	219
	21.4 The Gabor transform	221
V	I General relativity	222
22	Lorentzian geometry	223
	22.1 Axioms of special relativity	223
	22.2 Lorentz transformations	223
	22.3 Riemannian geometry	224
	22.4 Causality	226
23	The Cauchy problem in general relativity	227
	23.1 The Einstein equation	227
	23.2 Initial-data sets	229
	23.3 Well-posedness for the vacuum equation	230
24	General relativity in spherical symmetry	233
	24.1 Double-null pairs	233
	24.2 Local rigidity	235
	24.3 Einstein-Maxwell equations	237
25	Cosmic censorship	241
	25.1 Einstein-Maxwell-charged scalar field equations	242
	25.2 The structure of toy models	244
	25.3 Penrose inequalities	246
	25.4 Recent progress on strong censorship	248

Part I Preliminaries

Chapter 1

Functional analysis

Here we treat functional analysis in a high level of abstraction.

Throughout these notes, we mean by $f \leq g$ that there is a universal constant C > 0 such that $f \leq Cg$.

1.1 Locally convex spaces

Fix a vector space V.

Definition 1.1. V is said to be a *topological vector space* if it is equipped with a topology for which addition and multiplication are continuous.

Definition 1.2. V is said to be *locally convex* if V is equipped with a family of seminorms P_{α} and the initial topology with respect to the P_{α} .

This is the smallest topology containing the open sets $P_{\alpha}([0, \varepsilon))$ for each α and each $\varepsilon > 0$ and which is translation-invariant.

The most useful examples of locally convex spaces are Banach spaces.

Definition 1.3. V is said to be a *Banach space* if V is equipped with the topology arising from a complete norm.

Definition 1.4. If V is a topological vector space, then the *dual space* of V^* is the space of *continuous* linear maps $V \to \mathbb{C}$.

Definition 1.5. Let W be a Banach space and define a norm on Hom(V, W) by

$$||T|| = \sup_{||v|| \le 1} ||Tv||.$$

So V^* is a normed space, $V^* \subseteq \text{Hom}(V, \mathbb{C})$. In general it is very difficult to construct elements of V^* . In general we cannot guarantee constructively that V^* is nontrivial. On the other hand, it is often impossible to construct linear functions which are discontinuous (for example, any linear functional on a Banach space will be continuous if it was constructed without the axiom of choice). **Definition 1.6.** A function $f : V \to \mathbb{C}$ is said to be *sublinear* if it obeys the triangle inequality and if for each c > 0 and $x \in V$, f(cx) = cf(x).

Obviously seminorms are sublinear. Minkowski gauges are another useful example.

Definition 1.7. Let $K \subseteq V$. Then:

- 1. K is convex if for each $x, y \in K$, $c \in [0, 1]$, $cx + (1 c)y \in K$.
- 2. K is balanced if for each $c \in [0, 1], cK \subseteq K$.

If K is balanced and convex, then the *Minkowski gauge* of K is the functional

$$p_K(x) = \inf_{cK \ni x} c.$$

Notice that the balanced condition suggests that K needs to be close to the origin. Moreover, Minkowski gauges are sublinear.

Sublinear estimates allow us to construct functionals using the axiom of choice, while still guaranteeing that they are continuous.

Theorem 1.8 (Hanh-Banach). Assume that $p: V \to \mathbb{C}$ is sublinear, $U \subset V$ a subspace, and $f: U \to \mathbb{C}$ a linear functional. If f is dominated by p, i.e. for each $x \in U$, $|f(x)| \leq |p(x)|$, then f extends to V.

In general the extension of f will only be unique in case U is dense. So we have to use the axiom of choice to construct f.

Proof. The extension to the complex case is trivial so we replace \mathbb{C} with \mathbb{R} . Assume that f is defined on a space $W, U \subseteq W \subset V$. Choose $v \in V \setminus W$ and define f(v) such that for each $w \in W$ and $s, t \geq 0$,

$$\frac{p(w-sv)}{s} \le h(v) \le \frac{p(w+tv) - f(w)}{t}.$$

This is always possible because

$$f((t+s)w) \le p((t+s)w) = p((t+s)w + tsv - tsv) \le p(sw + stv) + p(tw - stv)$$

 \mathbf{SO}

$$\frac{f(w) - p(w - sv)}{s} \le \frac{p(w + tv) - f(w)}{t}.$$

Therefore for any W and v we can extend f to W + v. If W is the family of subspaces of V on which f is defined and $C \subset W$ is a chain, then C therefore has an upper bound. Since $U \in W$, Zorn's lemma implies that W has a maximal element, which is clearly V.

In case p is the norm of V, this implies that V^* is nontrivial. The Hanh-Banach theorem also has a useful geometric interpretation.

Theorem 1.9 (Hanh-Banach separation theorem). Let \mathbb{R} be the scalar field and $A, B \subset V$ be convex, nonempty, and disjoint. If A is open then there is a $\varphi \in V^*$ and $t \in \mathbb{R}$ such that for every $a \in A$ and $b \in B$,

$$\varphi(a) < t \le \varphi(b).$$

Proof. Choose $a_0 \in A$ and $b_0 \in B$, and let $C = A - B + b_0 - a_0$. Then $0 \in C$, C is convex, and $C = \bigcup_{b \in B} A - b + b_0 - a_0$, so C is open. If $x = b_0 - a_0$, then $x \notin C$.

By the Hanh-Banach theorem, choose a $\varphi \in V^*$ such that $\varphi(x_0) = 1$ and $\varphi < 1$ on C. Given $a \in A$ and $b \in B$ we have

$$\varphi(a) < \varphi(b) + \varphi(a_0) - \varphi(b_0) + 1 = \varphi(b).$$

If $t = \inf_B \varphi$, then this gives, for every a and b,

$$\varphi(a) \le t \le \varphi(b).$$

Since A is open, $\varphi(A)$ is open so the claim holds.

In particular, linear functionals separate points.

Let's consider more properties of convexity. Let the scalar field be \mathbb{R} and let $S \subset V$ be nonempty, compact, and convex.

Definition 1.10. A face of S is a nonempty, compact, convex set $K \subseteq S$ such that for each $x \in K$, if there are $y_1, y_2 \in S$ and $c \in (0, 1)$ such that $x = cy_1 + (1 - c)y_2$, then $y_1, y_2 \in K$. If |K| = 1, then K is called an *extreme point*.

Definition 1.11. If $X \subset V$ is a set, the *convex hull* of X is the smallest set containing X which is closed and convex.

For example, if S is a convex polygon, then the extreme points of S are its vertices, and S is the convex hull of its vertices. The Krein-Milman theorem says that this phenomenon happens even in infinite dimensions.

Theorem 1.12 (Krein-Milman). S is the convex hull of its extreme points.

Proof. Assume that S' is the convex hull of the extreme points of S. Then $S' \subseteq S$, so S' is compact. If there is an $x_0 \in S \setminus S'$, then since V^* separates points, there is a $\lambda \in V^*$ such that $\lambda(S') < \lambda(x_0)$. If $C = \max \lambda(x_0)$, then $\varphi^{-1}(C) \cap S$ contains no extreme points of S. We can contradict this by showing that every convex compact set has an extreme point.

Let \mathcal{F} be the set of all faces of S. Clearly $S \in \mathcal{F}$ so \mathcal{F} is nonempty. If $\mathcal{C} \subset \mathcal{F}$ is a chain, then $\bigcap \mathcal{C}$ is a face, so by Zorn's lemma \mathcal{F} has a minimal element S_0 .

Let $\varphi \in V^*$. Since S_0 is convex, it is connected, so $\varphi(S_0)$ is compact and connected. In particular, $\varphi(S_0) = [a, b]$ for some $a \leq b$. So $\varphi^{-1}(b) \cap S_0$ is nonempty, convex, and compact. If $x \in \varphi^{-1}(b) \cap S_0$,

$$x = ty_1 + (1 - t)y_2,$$

then $y_1, y_2 \in S_0$. Therefore $\varphi(x) = b = c\varphi(y_1) + (1-c)\varphi(y_2)$, so $\varphi(y_1), \varphi(y_2) \ge c$. Therefore $y_1, y_2 \in \varphi^{-1}(b) \cap S_0$, so $\varphi^{-1}(b) \cap S_0$ is a face and by minimality, $\varphi^{-1}(b) \cap S_0 = S_0$. So $\varphi(S_0) = b$. Since φ was arbitrary and V^* separates points, $|S_0| = 1$. So S has an extreme point.

1.2 Hilbert spaces

Definition 1.13. A *Hilbert space* V is a Banach space whose norm arises from an inner product.

The basic fact about Hilbert spaces V is that $V^* = V$.

Theorem 1.14 (Riesz representation theorem). The association

$$V \to V^*$$
$$v \mapsto (w \mapsto \langle v, w \rangle)$$

is a surjective isometry.

Proof. Evaluating $w \mapsto \langle v, w \rangle$ at v, we see $||v|| = ||v||_{op}$. So we just have to check surjectivity. Let $\varphi \in V^*$, and $F = \ker \varphi$. If F = 0 we're done; otherwise F^{\perp} is nonempty. Let $z \in F^{\perp}$ and $\alpha = \varphi(z)/||z||$. Then for any $x \in V$,

$$\langle x, \alpha z \rangle = \left\langle x - \frac{\varphi(x)}{\varphi(z)} z, \alpha z \right\rangle + \left\langle \frac{\varphi(x)}{\varphi(z)} z, \alpha z \right\rangle \\ = \left\langle \frac{\varphi(x)}{\varphi(z)} z, \alpha z \right\rangle = \frac{\varphi(x)}{\varphi(z)} \varphi(z) = \varphi(x).$$

-	-	-

1.3 Bochner integration

Now we fix a Banach space B and a measure space (X, Σ, μ) . Recall that the *Caratheodory* construction is the standard way of building (X, Σ, μ) : we define a semiring Σ_0 of sets (i.e. a family of sets closed under finite intersection and subsets of finite disjoint unions) and a countably additive function μ on Σ_0 , which then extends to an outer measure μ^* on the power set $\mathcal{P}(X)$. If $E \subseteq X$ satisfies the *Caratheodory criterion*, i.e. that for all $F \subseteq X$,

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \setminus E),$$

then we declare that E is measurable. The measurable sets form a σ -algebra Σ on which μ^* is outer measurable (note that μ^* did not have to be constructed from a semiring for this step to work; any outer measure will do) and we define the restriction μ of μ^* to Σ to be the desired outer measure.

Definition 1.15. A *B*-valued *integrable simple function* is a finite linear combination of functions

$$\begin{split} \chi^b_E : X &\to B \\ E &\ni x \mapsto b \\ E^c &\ni x \mapsto 0 \end{split}$$

where E is a measurable set with $||\mu(E)|| < \infty$, $b \in B$. The integral of a B-valued ISF $f = \sum_{n} \chi_{E_n}^{b_n}$ is

$$\int_X f \ d\mu = \sum_n b_n \mu(E_n)$$

and the L^1 -norm is $||f||_{L^1} = \int_X |f| \ d\mu$.

Then L^1 is naturally the Cauchy completion of the ISF.

Definition 1.16. A function $X \to B$ is a *B*-valued *integrable function* if it lies in L^1 .

Definition 1.17. For $p \in (1, \infty)$, the L^p norm of $f : X \to B$ is

$$||f||_{L^p} = \left(\int_X ||f(x)||^p \ d\mu(x)\right)^{1/p}$$

and the L^{∞} norm is $||f||_{L^{\infty}} = \lim_{p \to \infty} ||f||_{L^p} = \operatorname{ess sup} ||f||.$

The usual Lebesgue convergence theorems hold:

Theorem 1.18 (Lebesgue convergence theorems). Let $\{f_n\}$ be a pointwise convergent sequence of integrable functions. Then:

1. If each $f_n \leq f_{n+1}$,

$$\lim_{n} \int f_n = \int \lim_{n} f_n < \infty.$$

2. If there is an integrable function g > 0 such that every $|f_n| \leq g$,

$$\lim_{n} \int f_n = \int \lim_{n} f_n \le g.$$

3.

$$\int \liminf_{n} f_n \le \liminf_{n} \int f_n.$$

Now let's make some estimates which will actually prove that the L^p -norm is a norm, besides being useful later.

Theorem 1.19 (Jensen's inequality). Let $f : \mathbb{R} \to \mathbb{R}$ be convex and g an integrable function. Then

$$f\left(\int g\right) \leq \int f \circ g$$

Theorem 1.20 (Holder's inequality). Let

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then $||fg||_{L^1} \le ||f||_{L^p} ||g||_{L^q}$.

Proof. The mapping $x \mapsto x^p$ is convex so if $f, g \ge 0$,

$$\int fg = \left(\int g^q\right) \int fg^{1-q} \frac{g^q}{\int g^q} \le \left(\int g^q\right) \left(\int f^p g^{p(1-q)} \frac{g^q}{\int g^q}\right)^{1/p}$$
$$= \left(\int g^q\right) \left(\left(\int g^q\right) \left(\int f^p\right)\right)^{1/p} \le \left(\int f^p\right)^{1/p} \left(\int g^q\right)^{1/q}.$$

Notice that Holder's inequality implies that L^2 is a Hilbert space with inner product

$$\langle f,g\rangle = \int fg.$$

Theorem 1.21 (Minkowski's inequality). Let

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then

$$||f + g||_{L^p} \le ||f||_{L^p} + ||g||_{L^p}.$$

Proof. By Holder's inequality,

$$\int |f+g|^p = \int |f+g| |f+g|^{p-1} \le \int (|f|+|g|) |f+g|^{p-1}$$
$$\le \left(\left(\int |f|^p \right)^{1/p} + \left(\int |g|^p \right)^{1/p} \right) \left(\int |f+g|^{(p-1)\left(\frac{p}{p-1}\right)} \right)^{1-\frac{1}{p}}$$
$$= (||f||_{L^p} + ||g||_{L^p}) \frac{||f+g||_{L^p}}{||f+g||_{L^p}}.$$

г		٦	
L			
L			

Now we discuss change of variables.

Definition 1.22. Let ν be a measure. Then

- 1. ν is absolutely continuous with respect to μ if for every measurable set A, $\mu(A) = 0$ implies $\nu(A)$.
- 2. ν is singular with respect to μ if there are disjoint measurable sets A, B such that $X = A \cap B, \nu(A) = 0$ and $\mu(A) = 0$.
- 3. If there is a measurable function f such that for every measurable set A,

$$\nu(A) = \int_A f \ d\mu,$$

then f is the Radon-Nikodym derivative of ν , written

$$f = \frac{d\nu}{d\mu}.$$

Theorem 1.23 (Radon-Nikodym). Let μ be σ -finite and ν be a positive measure. Then there is a unique decomposition $\nu = \nu_a + \nu_s$ such that ν_a is absolutely continuous and ν_s is singular (with respect to μ). Moreover, ν_a has a Radon-Nikodym derivative.

In particular, if ν was already absolutely continuous, then $\nu_s = 0$ and ν has a Radon-Nikodym derivative.

Proof. Uniqueness is obvious. First assume $\mu(X) < \infty$. Then $\mu + \nu$ is finite, so $L^{\infty}(\mu + \nu) \subseteq L^{1}(\mu + \nu)$. So by the Cauchy-Schwarz inequality, if f is an ISF,

$$\left| \int f \, d\nu \right| \le ||f||_{L^{1}(\nu)} \le ||f||_{L^{1}(\mu+\nu)} \le ||f||_{L^{2}(\mu)}.$$

So $\int \cdot d\nu$ is L^2 -continuous on ISF, hence on $L^2(\mu + \nu)$. So by the Riesz representation theorem, there is a nonnegative $h \in L^1(\mu + \nu)$ such that

$$\int f \, d\nu = \int \int f h \, d(\mu + \nu)$$

for any $f \in L^2$. In particular, if A is measurable,

$$\int_A h \ d(\mu + \nu) = \nu(A) \le (\mu + \nu)(A).$$

Without loss of generality we assume $h \leq 1$. If $g \in L^{\infty}(\nu)$,

$$\int g \, d\nu = \int gh \, d\mu + \int gh \, d\nu$$

So if Y is the set of all y such that $0 \le h(y) < 1$, it follows that $\mu(Y) = \mu(X)$. By induction,

$$\int g \, d\nu = \int g(h + \dots + h^n) \, d\mu + \int gh^n \, d\nu.$$

Since $h \leq 1$, the dominated convergence theorem implies

$$\int gh^n \, d\nu \to \int_{X \setminus Y} g \, di$$

and if

$$f = \frac{h\chi_Y}{1-h}$$

we have

$$\int g \, d\nu = \int_Y g f \, d\mu + \int_{X \setminus Y} g \, d\nu$$

and take $\nu_s(A) = \nu(A \cap (X \setminus Y))$. Then we take

$$\nu_a(A) = \int_A f \ d\mu$$

so f is the Radon-Nikodym derivative of ν_a , $\nu_a + \nu_s = \nu$ by taking $g = \chi_A$.

To extend to the σ -finite case, break up X into countably many finite measure spaces and sum over them.

Next we discuss iterated integrals. Given measure spaces (X, S, μ) and (Y, T, ν) , we need a σ -algebra on $X \times Y$ and a measure defined on that σ -algebra. To do this, we use the Caratheodory construction.

Definition 1.24. If $E \in S$ and $F \in T$, then $E \times F$ is a measurable rectangle. Let $S \otimes T$ denote the smallest σ -algebra containing the measurable rectangles, and on for each measurable rectangle, define a countably additive function by

$$\mu \otimes \nu(E \times F) = \mu(E)\nu(F).$$

By the monotone convergence theorem $d\nu$, $\mu \otimes \nu$ is countably additive. So the Caratheodory construction gives rise to a measure $\mu \otimes \nu$ whose measurable sets include $S \otimes T$ (in fact, is the completion of $S \otimes T$).

Definition 1.25. The measure space $(X \times Y, S \otimes T, \mu \otimes \nu)$ is the *product measure space* of (X, S, μ) and (Y, T, ν) .

Straight from the definitions, we know that for every measurable rectangle $E \times F$,

$$\int \chi_{E\times F} \ d(\mu \otimes \nu) = \iint \chi_{E\times F} \ d\mu \ d\nu = \iint \chi_{E\times F} \ d\nu \ d\mu.$$

For a function f defined on $X \times Y$ we define $f^y(x) = f(x, y)$ and $f_x(y) = f(x, y)$. For a set $G \subseteq X \times Y$, we define $G^y = \{x \in X : (x, y) \in G\}$ and $G_x = \{y \in Y : (x, y) \in G\}$.

Theorem 1.26 (Fubini). Let $f \in L^1(\mu \otimes \nu)$ and assume $\mu \otimes \nu$ is σ -finite. Then for almost every $y, f^y \in L^1(\mu)$. Moreover, the function

$$F(y) = \int f^y \, d\mu$$

has $F \in L^1(\nu)$, and

$$\int f \ d(\mu \otimes \nu) = \iint f^y \ d\mu \ d\nu = \iint f_x \ d\nu \ d\mu.$$

The assumption of σ -finiteness is not optional here, and Fubini's theorem can fail for large cardinality measure spaces.

Definition 1.27. Let M be a family of subsets of X such that for every countable chain of A_n in M and $\bigcup_n A_n = A$ or $\bigcap_n A_n = A$, $A \in M$. Then we say M is a monotone class.

If R is a ring of sets, then the smallest monotone class M containing R is also a ring, and it is not hard to see that M is the smallest σ -algebra containing R.

Lemma 1.28. Let $G \in S \otimes T$. Then:

- 1. $G_x \in T$ and $G^y \in S$.
- 2. $x \mapsto \nu(G_x)$ and $y \mapsto \mu(G^y)$ are measurable.

3. One has

$$\mu \otimes \nu(G) = \int (x \mapsto \nu(G_x)) \ d\mu(x) = \int (y \mapsto \mu(G^y)) \ d\nu(y) = \iint \chi_G \ d\mu \ d\nu.$$

Proof. This is obvious if G is a measurable rectangle. We shall show that the algebra of sets on which this claim holds is a monotone class, hence a σ -algebra. Clearly if $\bigcup_n G_n = G$ then G has the property. Given $x \in X$, $\bigcup_n (G_n)_x = G_x$, so $G_x \in T$. Therefore the chain of functions $x \mapsto \nu(G_n)_x$ converges to $x \mapsto \nu(G_x)$ which is therefore measurable. So by the monotone convergence theorem,

$$\lim_{n} \mu \otimes \nu(G_n) = \lim_{n} \int (x \mapsto \nu((G_n)_x) \ d\mu(x) = \int (x \mapsto \nu(G_x) \ d\mu(x) = \mu \otimes \nu(G).$$

So this algebra is closed under ascending chains. The proof in the other direction is similar but you have to start by assuming that $\mu \otimes \nu(G_1) < \infty$.

Lemma 1.29. Let $f \ge 0$ be $S \otimes T$ -measurable. Then

$$\int f \ d\mu \otimes \nu = \iint f \ d\mu \ d\nu.$$

Proof. Let $\{f_n\}$ be a chain of ISFs. This claim is obvious for ISF, so the monotone convergence theorem on the f_n^y for each $y \in Y$.

Theorem 1.30 (Tonelli). If f is $S \otimes T$ -measurable, g(x) = ||f(x)||, $g^y \in L^1(\mu)$, and $(y \mapsto \int g^y d\mu) \in L^1(\nu)$, then $f \in L^1(\mu \otimes \nu)$.

Proof. Clear by the lemmata.

Proof of Fubini's theorem. Let g(x, y) = ||f(x, y)||. Then if $\{f_n\}$ is a sequence of ISF converging to f, g dominates the f_n . Apply the dominated convergence theorem twice, once for each integral.

1.4 Duality

Fix a normed space V. We consider properties of V^* . Since \mathbb{C} is complete, V^* is a Banach space; in particular, V^{**} is a Banach space. So we can always embed V in a Banach space by the mapping

$$V \to V^{**}$$
$$v \mapsto (\varphi \mapsto \varphi(v))$$

However, V^{**} is rarely the completion of V if V is infinite-dimensional. Moreover, the topology of V^* is a bit awkward to work with, since a convergence in operator norm is much stronger than convergence pointwise.

Definition 1.31. The weakstar topology of V^* is the initial topology such that every evaluation $\varphi \mapsto \varphi(v)$ is continuous.

In other words, the weakstar topology is the topology of pointwise convergence.

Theorem 1.32 (Banach-Alaoglu). Let B be the closed unit ball of V^* . Then B is weakstar compact.

Like the Hanh-Banach and Krein-Milman theorems, the proof of Banach-Alaoglu uses the axiom of choice. However, the Banach-Alaoglu theorem is not really nonconstructive, since if V is separable, we can use a diagonalization argument to prove it instead. Banach-Alaoglu generalizes to locally convex spaces.

Proof. Let

$$D_v = \{ z \in \mathbb{C} : |z| \le ||v|| \}$$

and $D = \prod_{v \in V} D_v$. By Tychonoff's theorem, D is compact. Moreover, there is a natural embedding

$$u: V^* \to D$$
$$f \mapsto \{f(v)\}_{v \in V}$$

Since the product topology is the topology of pointwise convergence, ι is a homeomorphism $V^* \to \iota(V^*)$. So we just need to show that $\iota(V^*)$ is closed. So let $\{\{f_\alpha(v)\}_{v\in V}\}_{\alpha\in A}$ be a net in D, which converges to a $\{\varphi_v\}_{v\in V}$. Then $f(v) = \varphi_v$ is a linear functional and $f_\alpha \to f$ pointwise so $\{\varphi_v\}_{v\in V} \in \iota(V^*)$.

Now we compute the duals of the main examples of Banach spaces we have presented so far.

Theorem 1.33. Let $p, q \in [1, \infty]$ and assume μ is σ -finite.

$$\frac{1}{p} + \frac{1}{q} = 1$$

Then $(L^{p}(\mu))^{*} = L^{q}(\mu)$.

Actually, this theorem is true without the σ -finiteness; however, it becomes much more difficult.

Proof. For $g \in L^q$, one has $||g||_{p^*} \leq ||g||_q$ by Holder's inequality and by taking larger and larger measurable sets E and considering $\int_E g$, we check $||g||_{p^*} \geq ||g||_q$. So we just need to show that the map $L^q \to L^{p^*}$ is surjective.

If $h \in L^p$ and X splits into finite measure spaces X_k we put $h_k = \chi_{X_k}h$, so $\sum_k h_k = h$ in L^p by the dominated convergence theorem. If $\varphi \in L^{p^*}$ then $\varphi(\sum_k h_k) = \sum_k \varphi(h_k)$ so we might as well assume $X = X_1$, viz. $\mu(X) < \infty$. Then $L^{\infty} \subseteq L^p$, so $\varphi \in (L^{\infty})^*$. We can define an absolutely continuous measure ν by $\nu(A) = \varphi(\chi_A)$, and by the Radon-Nikodym theorem, there is a Radon-Nikodym derivative f of ν .

Let $Y_n = \{x \in X : |f(x)| \le n\}$ and let $g = f/|f|^{q-2}$, where g(x) = 0 if f(x) = 0, and $g_n = \chi_{Y_n} g_n$. Then $|g|^p = |f|^q$ and

$$\int_{Y_n} |f|^q = \int g_n f = \varphi(g_n) \preceq ||g_n||_p \preceq ||f_n||_{L^p(Y_n)}.$$

So $||f||_{L^q(Y_n)} < \infty$, and by the monotone convergence theorem, $f \in L^q$.

1.5 Vector lattices

We now consider the natural order structure of a space.

Definition 1.34. A vector lattice is a vector space V equipped with a partial order \leq which is translation-invariant such that (V, \leq) is a lattice, and such that for each $c \geq 0$ and $x \leq y$, $cx \leq cy$.

Recall that a lattice is just a poset which is closed under finite joins \vee (suprema) and meets \wedge (infima). Actually, we just need to check that V is a semilattice, since multiplication by -1 implies that a semilattice is already a lattice.

If V is a vector lattice and $v \in V$, we define $f_{\pm} = \pm f \vee 0$. Then $f = f_{+} - f_{-}$ and we define the absolute value (or valuation) $|f| = f_{+} + f_{-}$.

Definition 1.35. A *Banach lattice* is a vector lattice V which is a Banach lattice, such that $|x| \leq |y|$ whenever $||x|| \leq ||y||$.

Example 1.36. A function space mapping into \mathbb{R} is usually a Banach lattice with the natural ordering, $f \leq g$ iff for every $x, f(x) \leq g(x)$. For example, C(X) is a lattice. Spaces of operators are Banach lattices as well, whose positive elements are precisely the positive operators; as are spaces of signed measures, where the positive measures are the positive elements.

Theorem 1.37. Let V be a Banach lattice. There is a natural ordering on V^* , such that $f \in V^*$ is positive iff for each positive $v \in V$, $f(v) \ge 0$, and such that $f \le g$ iff for every positive $v \in V$, $f(v) \le g(v)$.

Proof. Take the definition of positive functionals as in the statement of the theorem. If f and -f are both positive, each $v = v_+ - v_-$ has $f(v_+) \ge 0$ but $f(v_-) \le 0$. So f(v) = 0. Since v was arbitrary, f = 0.

Given $f \in V^*$, define

$$f^+(v) = \sup_{0 \le x \le v} f(x)$$

for $v \ge 0$. Then $f^+ \ge f$, and f^+ is finite because if $x \le v$, $|f(x)| \le ||f||||v||$. Moreover, if $v, w \ge 0$, it is easy to check $f^+(v+w) = f^+(v) + f^+(w)$. So f^+ is positive-linear, so extends to all of V and so $f^+ \in V^*$.

Clearly $f^+ - f \ge 0$. We need to show this is optimal, i.e. $f^+ = f \lor 0$. Assume $g \ge f \lor 0$. Then for $0 \le x \le v$, $f(x) \le g(x) \le g(v)$, so taking the sup over x we have $f^+(v) \le g(v)$. The other direction is similar. So $f^+ = f \lor 0$.

Fix a compact Hausdorff space X, $|X| \ge 2$ (so in particular, every set which separates points is nonempty). Let us now study the behavior of sublattices of $C(X) = C(X \to \mathbb{R})$.

Theorem 1.38 (Dini). Let L be a sublattice of C(X), and define $g(x) = \inf_{f \in L} f(x)$. For each $\varepsilon > 0$, there exists a $h \in L$ such that $g \leq h \leq g + \varepsilon$.

Proof. For each $f \in L$ let $U_f = \{x \in X : f(x) - g(x) \leq \varepsilon\}$. Then the U_f are an open cover of X, which has a finite subcover by functions f_1, \ldots, f_k . Take $h = \bigwedge_{j < k} f_j$. \Box

When can a lattice be used to approximate any function in C(X)? A necessary condition is that the lattice strongly separates points. This turns out to be sufficient as well.

Definition 1.39. A set $A \subseteq C(X)$ separates points if for each $x, y \in X$, there is an $f \in A$ such that $f(x) \neq f(y)$. If, in addition, the constant functions $\mathbb{R} \subseteq A$, then A strongly separates points.

Theorem 1.40 (Stone-Weierstrass). If $L \subseteq C(X)$ is a sub-vector lattice or a subalgebra which strongly separates points, then L is dense in C(X).

The lattice case is also called the *Kakutani-Krein theorem*.

Lemma 1.41. Let *L* be a sublattice of C(X) which separates points and is closed under multiplication and addition by elements of \mathbb{R} . Then if $B \subseteq X$ is compact, $p \in X \setminus B$, and $a, b \in \mathbb{R}$, there is a $g \in L$ such that $g \ge a$, g(p) = a and g > b on *B*.

Proof. For each $x \in B$ there exists $g_x \in L$ such that $g_x(p) = a$ and $g_x(x) = b + 1$. Let $U_x = \{y \in X : g_x(y) > b\}$. Since $x \in U_x$, the U_x are an open cover of B with finite subcover U_{x_1}, \ldots, U_{x_k} . Take $g = a \lor \bigvee_{j \le k} g_{x_k}$.

Lemma 1.42. Assume that L is a closed unital subalgebra of C(X). Then L is a lattice.

Proof. Choose $\varepsilon > 0$ and apply the classical Weierstrass theorem to [-1, 1] to find a polynomial P_{ε} which approximates $|\cdot|$ in L^{∞} -norm by ε . Then for each $f \in L$, we can approximate |f| by $P_{\varepsilon} \circ f$. Since L is unital, $P_{\varepsilon} \circ f \in L$. So $|f| \in L$, since L is closed. The lattice operations \lor and \land can be expressed in terms of algebra operations + and \cdot , and $|\cdot|$, so L is closed under lattice operations.

Proof of Stone-Weierstrass. First consider the case that L is a lattice. Given $f \in C(X)$, define $L_f = \{g \in L : g \geq f\}$. Then L_f is a sublattice of L. Given $x \in X$, $\delta > 0$, the set $B = \{y \in X : f(y) \geq f(x) + \delta\}$ is closed. Since X is compact, there is an M > 0 such that f < M. Apply Lemma 1.41 with $a = f(x) + \delta$ and b = M, so there is a $g \in L$ such that $g \geq f(x) + \delta$, $g(x) = f(x) + \delta$ and g > M on B. So $f \leq g \leq g + \delta$, so $f = \bigwedge L_f$. Therefore by Dini's theorem, there is an $h \in L$ with the desired properties.

For the algebra case, since L strongly separates points, L is unital. Therefore \overline{L} is a closed unital algebra, \overline{L} is a closed lattice whose closure is C(X), by Lemma 1.42 and the above case. So $\overline{L} = C(X)$.

This even extends to decaying functions on locally compact Hausdorff spaces, by taking the one-point compactification.

1.6 Positive Radon measures

The usual construction of measures by ISF is somewhat unnatural when we have a nice topology, since then we can define integration in terms of continuous functions. Clearly "nice" in this context implies locally compact Hausdorff; these conditions are also sufficient (though σ -compactness also helps). Throughout this section, we fix a locally compact Hausdorff space X and consider the space $C_c(X)$ of compactly supported continuous functions $X \to \mathbb{C}$. We have not given a topology on $C_c(X)$, so a functional is just an element of $\text{Hom}(C_c(X), \mathbb{C})$ for now.

Definition 1.43. A positive Radon measure on X is a functional on $C_c(X)$.

Let us prove that a positive Radon measure is actually a measure in a natural way. First, we put a topology on $C_c(X)$. We start by putting the L^{∞} -topology on $C_c(U)$ for each open set $U \subseteq X$ with compact closure.

Definition 1.44. The *inductive limit topology* of $C_c(X)$ is the final (i.e. strongest) topology on $C_c(X)$ such that $\varphi : C_c(X) \to Y$ is continuous provided that for each open set $U \subseteq X$ with compact closure, $\varphi|_{C_c(U)}$ is continuous.

In other words, the inductive limit topology is the final topology which makes the natural maps $C_c(U) \to C_c(X)$ continuous. A positive Radon measure is continuous for the inductive limit topology, as can be seen by taking an $h \in C_c(X)$ which is 1 on U, so $||\varphi||_{C_c(U)} \leq \varphi(h)$.

Now we need some general facts about locally compact Hausdorff spaces.

Definition 1.45. A (continuous) partition of unity on a subordinate to an open cover U_1, \ldots, U_n is a family of (continuous) functions f_1, \ldots, f_n which are compactly supported in U_i , such that $\sum_i f_i = 1$.

Theorem 1.46. For any finite open cover \mathcal{U} of a compact set, there is a partition of unity subordinate to \mathcal{U} .

Lemma 1.47. Let $K \subseteq X$ be compact. If U_1, \ldots, U_n is an open cover of K there are compact sets $K_1, \ldots, K_n, K_i \subseteq U_i$, which cover K.

Proof. For each $x \in K$ choose a j such that $U_j \ni x$ and an open set $V_x \ni x$ such that

$$V_x \subset \overline{V}_x \subset U_j.$$

Then the V_x are an open cover of K so they reduce to a finite subcover V_{x_1}, \ldots, V_{x_p} . For each $k \leq p$ choose a $j_k \leq n$ such that $V_{x_k} \subseteq U_{j_k}$ and let $W_j = \bigcup_{j_k=j} V_k \subseteq U_j$. Then $\overline{W}_j \subseteq U_j$ and the \overline{W}_j contain the V_x s, so are a compact cover of K.

Proof of Theorem 1.46. Fix a compact set K. By the lemma, we can choose $D_j \subseteq U_j$ a compact cover of K and g_j supported in U_j with $g_j \ge 1$ on D_j , and $h = \sum_j g_j$. Then $h \ge 1$ on C and put $k = h \lor 1 \ge 1$. So 1/k exists and we can put $f_j = g_j/k$, to force $\sum_j f_j = 1$. \Box

Definition 1.48. A *content* is a function defined on sets into $[0, \infty]$ which is monotone, countably subadditive, and finitely additive, and which carries compact sets to $[0, \infty)$. A content μ is said to be *inner regular* if for every open set U,

$$\mu(U) = \sup_{\substack{\overline{V} \subseteq U \\ V \text{ open} \\ \overline{V} \text{ compact}}} \mu(V).$$

Fix a positive Radon measure φ , and define an inner-regular content μ on the topology \mathcal{T} on X by

$$u: \mathcal{T} \to [0, \infty]$$
$$U \mapsto \sup_{\substack{f \in C_c(U)\\ 0 \le f \le 1}} \varphi(f)$$

Given a content ν , we can extend ν to an outer measure ν^* on the power set \mathcal{P} by

$$\nu^*(A) = \inf_{\substack{U \subseteq A \\ U \in \mathcal{T}}} \nu(U).$$

In turn, then, ν^* restricts to a measure, also called ν , on its measurable σ -algebra, by the Caratheodory construction. So, in particular, μ gives rise to a measure.

Definition 1.49. Let ν be a Borel measure. We say that ν is *outer regular* if for every Borel set E,

$$\mu(E) = \inf_{\substack{E \subseteq U \\ U \in \mathcal{T}}} \mu(U)$$

and inner regular if for every open set U,

$$\mu(U) = \sup_{\substack{\overline{V} \subseteq U \\ V \text{ open} \\ \overline{V} \text{ compact}}} \mu(V).$$

We state the main result.

Theorem 1.50 (Riesz-Markov representation theorem). μ is a positive Borel measure which is both inner and outer regular, and φ is the unique functional such that for every $f \in C_c(X)$,

$$\varphi(f) = \int f \ d\mu.$$

The proof of the Riesz-Markov representation theorem is quite long, so we only sketch it.

Proof sketch. Let ν^* be an outer measure which is finitely additive and inner regular on the topology of X, and let U be open. Then Caratheodory's criterion holds for U and ν^* on open sets. Approximating any subset of X by an open set, Caratheodory's criterion holds on the power set for U and ν^* . So U is ν^* -measurable, and ν^* restricts to a Borel measure ν . In particular, μ is a Borel measure.

If $f \in C_c(X)$, and $f \ge 1$ on an open set U, $\varphi(f) \ge \mu^*(U)$. Approximating any set Aby an open set, we see that $\varphi(f) \ge \mu^*(A)$ whenever $f \ge 1$ on A. On the other hand, if $f \le 1$ on A, a monotone convergence argument shows that $\mu^*(A) \ge \varphi(f)$. Since $C_c(X)$ is a Banach lattice, we can replace f with f^+ and by decomposing X into a chain of sets $X_n\{x \in X : f(x) \ge n\varepsilon\}$ and summing the $f|_{X_n} \setminus f|_{X_{n-1}}$ we prove

$$\varphi(f) = \int f \ d\mu.$$

Since μ was inner and outer regular as a content, approximation by open sets implies regularity on Borel sets. Moreover, if ψ is a positive Radon measure, define a content ν by

$$\nu(U) = \sup_{\substack{f \le \chi_U \\ f \in C_c(U)}} \int f \, d\nu$$

If $\nu = \mu$ it follows that $\psi = \varphi$.

Notice that on the other hand, a complex measure ν on $C_c(X)$ gives rise to a functional ψ by

$$\psi(f) = \int f \, d\nu.$$

The positive part of ψ is in fact the positive part of ν .

Now if S is a locally compact semigroup, we let M(S) be the set of all finite Radon measures on S. This is a convolution algebra, with

$$\mu * \nu(f) = \iint_S f(xy) \ d\mu(x) \ d\nu(y).$$

1.7 Baire categories

Now we look at a topological analogue of "measure zero."

Definition 1.51. Let X be a topological space. A set $S \subseteq X$ is nowhere dense if for every open set $U, S \cap U$ is not dense in U. A set $T \subseteq X$ is meager or of the first category if T is the countable union of nowhere dense sets. A set $W \subseteq X$ is of the second category if it is not of the first category, or comeager if it is the complement of a meager set.

Lemma 1.52. For a topological space, the following are equivalent:

- 1. Every countable union of closed sets with empty interior has empty interior.
- 2. Every countable intersection of open dense sets is dense.
- 3. Every nonempty open set is of the second category.

This is basically obvious.

Definition 1.53. A topological space is a *Baire space* if one (and all) of the above criteria hold.

Theorem 1.54 (Baire category theorem). Every completely pseudometrizable or locally compact Hausdorff space is Baire.

For example, a Banach space is Baire.

Proof. Let U_n be a sequence of open dense sets, and let W be open, in the space X. Then $W \cap U_1$ is nonempty and open, say $x_1 \in W \cap U_1$. If X is pseudometrizable, then there is a $\varepsilon_1 \in (0,1)$ such that the open ball $V_1 = B(x_1,\varepsilon_1)$ satisfies $K_1 = \overline{B}(x_1,\varepsilon_1) \subseteq W \cap U_1$; if X is locally compact Hausdorff, then there is a compact set with nonempty interior $V_1 \subseteq K_1 \subseteq W \cap U_1$. Iterate using the denseness of the U_n and the axiom of choice to construct a sequence $x_n \in V_n \subseteq K_n \subseteq V_{n-1} \cap U_n$. If X is pseudometrizable, then we can always choose $\varepsilon_n < 1/n$, so the x_n are a Cauchy sequence. Otherwise, $\bigcap_n K_n$ is nonempty anyways by the finite intersection property. Either way, we can find an $x \in \bigcap_n K_n \subseteq \bigcap_n U_n$ such that $x \in W$. So $\bigcap_n U_n$ is dense.

Actually, we didn't use the full axiom of choice. The Baire category theorem is equivalent over ZF to the following axiom.

Axiom 1.55 (Axiom of dependent choice). Let X be a nonempty set and R be a binary relation. If, for every $a \in X$, there is a $b \in X$ such that aRx, then there is a sequence of x_n such that x_nRx_{n+1} .

The axiom of dependent choice is not strong enough to prove the existence of nonmeasurable sets, for example. Moreover, if X is assumed to be separable, then the Baire category theorem just follows from induction, without even dependent choice.

Theorem 1.56 (uniform boundedness principle). Let X be a Banach space and Y a normed space, and F be a set of linear mappings $X \to Y$. If for every $x \in X$,

$$\sup_{T\in F} ||Tx|| < \infty,$$

then

$$\sup_{\substack{T \in F \\ |x||=1}} ||Tx|| = \sup_{T \in F} ||T||.$$

The uniform boundedness principle is also called the Banach-Steinhaus theorem. The proof is a standard application of the Baire category theorem: construct a chain of closed sets whose union is the entire space, which implies that one is not meager.

Proof. Let

$$X_n = \{x \in X : \sup_{T \in F} ||Tx|| \le n\}.$$

Then the X_n are a closed chain whose union is X. So by the Baire category theorem, there is an $x \in X$, m > 0, and $\varepsilon > 0$ such that $B(x, \varepsilon) \subset X_m$. So if ||u|| < 1 and $T \in F$,

$$||Tu|| = \varepsilon^{-1}||T(x+\varepsilon u) - Tx|| \le \varepsilon^{-1}||T(x+\varepsilon u)|| + \varepsilon^{-1}||Tx|| \le 2\frac{m}{\varepsilon}$$

Taking the sup over u of both sides,

$$\sup_{T \in F} ||T|| \le 2\frac{m}{\varepsilon} < \infty.$$

Theorem 1.57 (open mapping theorem). If $A : X \to Y$ is a surjective continuous linear mapping between Banach spaces, then A is open.

The open mapping theorem is also called the Banach-Schauder theorem.

Proof. We must show that if U is the open unit ball of X, then A(U) is open. Since $X = \bigcup_k kU, Y = \bigcup_k A(kU)$. By the Baire category theorem, there is a $k > 0, \varepsilon > 0$, and $y \in Y$ such that $B(y,\varepsilon) \subseteq \overline{A(kU)}$. If V is the unit ball of $Y, v \in V, y + \varepsilon v \in \overline{A(kU)}$ so

$$\varepsilon v \in \overline{A(kU)} + \overline{A(kU)} \subseteq \overline{A(2kU)}.$$

So if $L = 2k/\varepsilon$, $V \subseteq \overline{A(LU)}$.

In other words, for every $y \in Y$ and $\varepsilon > 0$ there is an $x \in X$ such that $||x|| \leq L||y||$ and $||y - Ax|| < \varepsilon$. In particular, given $y \in V$ we can choose x_1 such that $||x_1|| \leq L$ and $||x - Ax_1|| < 1/2$. Choose $||x_{n+1}|| \leq L2^{-n}$ such that

$$||y - A(x_1 + \dots + x_n) - Ax_{n+1}|| < 2^{-n-1},$$

by induction and the axiom of (dependent) choice. The sequence of partial sums is therefore Cauchy, so we can put $x = \sum_{n} x_{n}$, and Ax = y by the above estimates. Also

$$||x|| = \lim_{n \to \infty} \left| \left| \sum_{k \le n} x_k \right| \right| \le \sum_{n=1}^{\infty} x_n < 2L.$$

So $y \in A(2LU)$. Therefore $V \subseteq A(2LU)$ which was to be shown.

Theorem 1.58 (closed graph theorem). Let $A : X \to Y$ be a linear mapping between Banach spaces. If the graph of A is closed in $X \oplus Y$, then A is continuous.

Notice that while there isn't a canonical norm for $X \oplus Y$, any ℓ^p norm will do; since $X \oplus Y$ is a finite direct sum, all ℓ^p norms are equivalent. In particular, $X \oplus Y$ is a Banach space.

Proof. Let Γ be the graph of A, which is equipped with a natural (linear, bijective) projection $\pi_X : \Gamma \to X$. Since

$$||P(x, Ax)|| = ||x|| \le ||(x, Ax)||,$$

 $||P|| \leq 1 < \infty$. So by the open mapping theorem,

$$||Tx|| \leq ||P^{-1}x|| + ||x|| \leq ||x||.$$

Chapter 2

Complex analysis

Throughout, we identify $\mathbb{R}^2 = \mathbb{C}$ and write z = x + iy, dz = dx + idy, so that $d\overline{z} = dx - idy$. Then

$$dz \wedge d\overline{z} = 2idx \wedge dy = 2idA.$$

We thus write $2\partial f = \partial_x f - i\partial_y f$ and $2\overline{f} = \partial_x f + i\partial_y f$, so that

$$df = \partial f \, dz + \overline{\partial} f \, d\overline{z}.$$

We let K be a compact set in an Ω -precompact open set ω , where Ω is open in \mathbb{C} . So we have inclusions

$$K \subset \omega \subset \overline{\omega} \subset \Omega \subseteq \mathbb{C}.$$

We will always assume that $\partial \omega$ is a positively oriented, piecewise- C^1 Jordan curve.

2.1 Cauchy-Green formula

Making the change of variable $dA \mapsto dz \wedge d\overline{z}$ in Green's formula, we arrive at the following generalization of the Cauchy integal formula.

Theorem 2.1 (Cauchy-Green). Let $f \in C^{1}(\omega)$. For each $\zeta \in \omega$,

$$f(\zeta) = \frac{1}{2\pi i} \left(\int_{\partial \omega} \frac{f(z)}{z - \zeta} dz + \iint_{\omega} \frac{\overline{\partial} f(z)}{z - \zeta} dz \wedge d\overline{z} \right).$$

Definition 2.2. The Cauchy-Riemann equation is the equation

$$\partial f = 0.$$

If $f \in C^1(\omega)$ solves the Cauchy-Riemann equation, we say that f is a holomorphic function, written $f \in A(\omega)$.

So in case f is holomorphic, we recover the classical Cauchy integral formula from the Cauchy-Green theorem.

Theorem 2.3. Let μ be a finite Borel measure on \mathbb{C} with compact support K. Let

$$u(\zeta) = \int_{\mathbb{C}} \frac{d\mu(z)}{z-\zeta}.$$

Then u is holomorphic on K^c . If $\varphi \in C^k(\omega)$ and $2\pi i\mu = \varphi \, dz \wedge d\overline{z}$, then $u \in C^k(\omega)$ and $\overline{u} = \varphi$ on ω . In particular, if μ solves the Cauchy-Riemann equation in the distributional sense, then μ is holomorphic.

Corollary 2.4 (Cauchy). If $u \in A(\Omega)$ then

$$||\partial^j u||_{L^{\infty}(K)} \preceq_{K,j} ||u||_{L^1(\omega)}.$$

Proof. Differentiate under the integral sign in the Cauchy-Green formula, and then use the triangle inequality. \Box

Corollary 2.5. If we are given a sequence of $u_n \in A(\Omega)$ which converges locally uniformly to u, then $u \in A(\Omega)$.

Proof. This is obvious for uniform convergence, but holomorphy is a local property. \Box

Corollary 2.6 (Montel). If we are given a sequence in $A(\Omega)$ which is locally bounded, then there is a locally uniformly convergent subsequence, whose limit is in particular holomorphic.

Proof. Use Arzela-Ascoli on Cauchy's inequality. Then use locally uniform convergence. \Box

Corollary 2.7 (root test). If $u(z) = \sum_n a_n z^n$, then u is analytic on $D(0, \limsup_n |a_n|^{1/n})$.

Corollary 2.8 (Taylor). If $u \in A(D(0, R))$, then $u \in C^{\infty}(D(0, R))$ and for $z \in D(0, R)$ we have

$$u(z) = \sum_{n=0}^{\infty} \frac{\partial^n u(0)}{n!} z^n.$$

Proof. Differentiate under the integral sign in the Cauchy-Green formula, then use the root test. \Box

Corollary 2.9. If Ω is connected and $u \in A(\Omega)$, and there is a $z \in \Omega$ such that for all j, $\partial^j u(z) = 0$, then u = 0.

Proof. Taylor series propagate to connected components.

Corollary 2.10 (Weierstrass preparation theorem). If $u \in A(\Omega)$, $0 \in \Omega$, $u \neq 0$, and u(0) = 0 with order k, then there is a $v \in A(\Omega)$ so that $u(z) = z^k v(z)$.

Proof. Factor the z^k out of the Taylor series.

Corollary 2.11. If $u \in A(\omega) \cap C(\overline{\omega})$ then

$$||u||_{L^{\infty}(\omega)} = ||u||_{L^{\infty}(\partial\omega)}.$$

2.2 Conformal mappings

Definition 2.12. A function between open sets $f : U \to V$ is *biholomorphic*, *conformal*, *angle-preserving*, or a *complex diffeomorphism* if f is a bijection and f and f^{-1} are both holomorphic.

If such a conformal map exists then U and V are conformally equivalent or conformal.

Lemma 2.13. Let $F : \mathbb{D} \to \mathbb{D}$ be conformal and F(0) = 0. Then $\exists \omega \in S^1$ such that for each $z \in \mathbb{D}$,

$$F(z) = \omega z.$$

Proof. Apply Schwarz to F and F^{-1} so that

$$|z| \le |F(z)| \le |z|.$$

Theorem 2.14. $F : \mathbb{D} \to \mathbb{D}$ is conformal if and only if there exist unique $a \in \mathbb{D}$ and $\omega \in S^1$ such that

$$F(z) = \omega \frac{z-a}{1-\overline{a}z}$$

Moreover, a can be chosen so that F(a) = 0, and

$$F^{-1}(z) = \omega^{-1} \frac{z+a}{1+\overline{a}z}.$$

Proof. Suppose that F is conformal and let $G : \mathbb{D} \to \mathbb{D}$ be given by

$$G(z) = \frac{z-a}{1-\overline{a}z}$$

Then since $|\overline{a}z| < 1$, $G \circ F^{-1}$ is conformal and G(a) = 0, thus $G \circ F^{-1}(0) = 0$. So by 2.13, $G \circ F^{-1}(z) = \omega z$ for some $\omega \in S^1$. Existence of ω is immediate (how could one rotate the disk at two different speeds?) and since *a* is determined by the preimage of 0, *a* is unique as well.

On the other hand, if

$$F(z) = \omega \frac{z-a}{1-\overline{a}z},$$

then since $|\overline{a}z| < 1$, F is holomorphic, as is its inverse (that it actually is an inverse is immediate).

Theorem 2.15 (Riemann mapping theorem). If $U \subseteq \mathbb{C}$ is open and simply connected, then U is conformal with \mathbb{D} or \mathbb{C} .

Lemma 2.16. Suppose that $U \neq \mathbb{C}$ is simply connected and $z_0 \in U$. Then the space

 $\mathcal{F} = \{ f : U \to \mathbb{D} \mid f \text{ is holomorphic and injective, and } f(z_0) = 0 \}$

is a nonempty normal family.

Proof. For convenience, let V be the strip

$$V = \{ z \in \mathbb{C} : |\operatorname{Im} z| < \pi \}.$$

Let $a \notin U$ and observe that $z - a \neq 0$. Since exp has a period of $2\pi i$, exp is a conformal map $V \to \mathbb{C} \setminus \{0\}$. In particular, its inverse log is conformal $\mathbb{C} \setminus \{0\} \to V$. Now let $\ell(z) = \log(z - a); \ell$ is holomorphic and injective.

Suppose that there does not exist $\delta > 0$ such that for each $z \in U$,

$$|\ell(z) - \ell(z_0) - 2\pi i| < \delta.$$

Then there is a sequence z_n in U such that $\ell(z_n) \to \ell(z_0) + 2\pi i$. By continuity of exp, $z_n \to z_0 e^{2\pi i} = z_0$. So $\ell(z_n) \to \ell(z_0) \neq \ell(z_0) + 2\pi i$, which is a contradiction, so δ exists. Let

$$g(z) = \frac{1}{\ell(z) - \ell(z_0) - 2\pi i}$$

Then g is clearly injective and holomorphic and $|g| < \delta^{-1}$, so g is bounded, and

$$f(z) = \frac{g(z) - g(z_0)}{\delta^{-1} + |g(z_0)|}$$

satisfies $f \in \mathcal{F}$. Thus \mathcal{F} is nonempty.

Moreover, each $f \in \mathcal{F}$ satisfies |f| < 1, so \mathcal{F} is uniformly bounded by 1; therefore \mathcal{F} is normal by Montel's theorem.

Lemma 2.17. With \mathcal{F} and z_0 as in 2.16, there exists a conformal $F \in \mathcal{F}$ and if

$$\lambda = \sup_{f \in \mathcal{F}} |f'(z_0)|$$

then $|F'(z_0)| = \lambda$.

 \mathcal{F} is precompact in \mathcal{O} , but this doesn't mean that \mathcal{F} is compact! So, while we want to use compactness to prove the existence of a function F that satisfies the desired hypotheses, we'll have to prove that in fact $F \in \mathcal{F}$ rather than $F \in \partial \mathcal{F}$.

Proof of 2.17. Since elements of \mathcal{F} are injective, their derivatives are nonzero by the argument principle, so $\lambda > 0$, and there is a sequence f_n such that $|f'_n(z_0)| \to \lambda$. By Montel's theorem, it has a convergent subsequence with a limit $F \in \mathcal{O}$, such that $F(z_0) = 0$ and $F'(z_0) = \lambda$ by Weierstrass' theorem. Moreover, $|F| \leq 1$. F is nonconstant since $\lambda > 0$, so by the open mapping theorem, |F| < 1, and by Hurwitz' theorem, F is injective. Therefore $F \in \mathcal{F}$.

Moreover, if F is not surjective, then there exists $w \in \mathbb{D} \setminus F(U)$. Let

$$\psi(z) = \frac{w-z}{1-\overline{w}z}.$$

By 2.14, ψ is an automorphism of \mathbb{D} , $\psi(w) = 0$ and $|\psi \circ F| > 0$. Now define

$$g(z) = \exp\left(\frac{\log\psi(F(z))}{2}\right).$$

Then $g(z_0) = \sqrt{w}$. By 2.14 again,

$$\tilde{\psi}(z) = \frac{\sqrt{w-z}}{1-\sqrt{w}z}$$

is also an automorphism of \mathbb{D} and $\tilde{\psi}(\sqrt{w}) = 0$. Let $G = \tilde{\psi} \circ g$. Then $G \in \mathcal{F}$. Now let $s(z) = z^2$ and $\varphi = \psi^{-1} \circ s \circ \tilde{\psi}^{-1}$. Then $\varphi(0) = 0$ so by Schwarz $|\varphi(z)| \leq |z|$. Moreover, φ is not injective anywhere, so $|\varphi'| < 1$. But $F = \varphi \circ G$ so

$$\lambda = |F'(z_0)| = |\varphi'(0)G'(z_0)| < |G'(z_0)| \le \lambda$$

which is a contradiction.

So F is surjective. Therefore the inverse function theorem implies that F is conformal. \Box

Clearly this lemma implies the Riemann mapping theorem.

$\mathbf{2.3}$ Approximation by polynomials

Let $K \subset \Omega \subseteq \mathbb{C}$ be compact. If K is the compactification of a disc, then it is easy to uniformly approximate holomorphic functions on K by polynomials, just by truncating the Taylor series.

Example 2.18. Let K be the compactification of an annulus and assume u has a pole in the center of K. If $p_j \to u$ on K and the p_j are entire functions, then in particular $p_j \to u$ on ∂K , so they are a Cauchy sequence in A(K), so on a disk containing K, and in particular the pole of u. Therefore $p_i \to p$, an entire function, even though u has a pole. Notice that we can still approximate u by meromorphic functions, though.

Theorem 2.19 (Runge approximation theorem). The following are equivalent.

- 1. If u is holomorphic near K, then there are functions $u_j \in A(\Omega)$ such that $u_j \to u$ uniformly on K.
- 2. The complement $\Omega \setminus K$ has no Ω -precompact connected components.
- 3. For each $z \in \Omega \setminus K$ there is an $f \in A(\Omega)$ such that

$$|f(z)| > ||f||_{L^{\infty}(K)}.$$

Proof. Let us first prove that not-2 implies not-3 and not-1.

If not-2, then there is a K-precompact component O of $\Omega \setminus K$, so $\partial O \subseteq K$. If $f \in A(\Omega)$, then

$$||f||_{L^{\infty}(O)} = ||f||_{L^{\infty}(\partial O)} \le ||f||_{L^{\infty}(K)}$$

by the maximum principle, implying not-3. Moreover, if 1 were true, then for every fholomorphic near K, we could approximate f by $f_j \in A(\Omega)$ uniformly. We have

$$||f_j - f_k||_{L^{\infty}(\overline{O})} \le ||f_j - f_k||_{L^{\infty}(K)},$$

so the f_j form a Cauchy sequence in $L^{\infty}(K)$, which converge to a holomorphic function $F \in A(O)$. But if f has a pole in O, then $f \neq O$, a contradiction, so we have not-1.

Now we show 1 and 2 imply 3. Let $L = K \cup D(z, \varepsilon)$ where $\epsilon < d(z, K)$. Then every component of L is either a component of K or else $\overline{D(z, \varepsilon)}$, and L satisfies the hypotheses of 2, so in particular satisfies the hypotheses of 1. Let $u \in A(L)$ be given by u = 1 near zand u on K. By 1, u is uniformly approximable in $A(\Omega)$. Assume $f \in A(\Omega)$ is such that $||f - u||_{L^{\infty}(L)} < \delta$ for δ small enough; then f witnesses 3.

Finally we show 2 implies 1. Let X_2 be the set of restrictions of holomorphic functions to K, and let X_1 be the set of restrictions of holomorphic functions on Ω to K. Then $X_1 \subset X_2 \subset C(K)$, and 1 holds iff $\overline{X_1} = \overline{X_2}$. So we must show $\overline{X_2} \subseteq \overline{X_1}$. By the Hanh-Banach and Riesz-Markov theorems, this is equivalent to showing that for every finite Borel measure μ with support in K and every $f \in A(\Omega)$, $\int f d\mu = 0$. In particular, we can prove this with the addition assumption that f is only holomorphic near K. So fix such a f, μ .

Let φ be the holomorphic function given by μ . Since $\varphi = 0$ on $\mathbb{C} \setminus \Omega$, $\varphi = 0$ on any component of $\mathbb{C} \setminus K$. Moreover,

$$\frac{1}{z-\zeta}=-\sum_{j=0}^{\infty}\zeta^{-j-1}z^{j}$$

whenever the sum converges, i.e. for $|\zeta|$ large enough. Therefore

$$\varphi(\zeta) = -\sum_{j=0}^{\infty} \zeta^{-j-1} \int z^j \, d\mu(z) = 0$$

for $|\zeta|$ large enough. By 2, $\Omega \setminus K$ has no Ω -precompact components, so every component of $\Omega \setminus K$ touches $\partial \Omega$ or is unbounded. Therefore $\varphi = 0$ on $\mathbb{C} \setminus K$.

Let ψ be a cutoff which is 1 whenever f is holomorphic. Then, taking $\omega = \mathbb{C}$, we have

$$\psi(z) = \frac{1}{2\pi i} \iint_{\mathbb{C}} \frac{\overline{\partial}\psi(\zeta)}{\zeta - z} d\overline{\zeta} \wedge d\zeta.$$

For each $z \in K$, the function

$$\zeta \mapsto \frac{\overline{\partial}\psi(\zeta)f(\zeta)}{\zeta - z}$$

is smooth since $\zeta \notin K$. So we are entitled to use Fubini's theorem to prove

$$\int_{\mathbb{C}} f \ d\mu = \int_{\mathbb{C}} \psi f \ d\mu = \frac{1}{2\pi i} \iiint_{\mathbb{C}^2} \frac{\overline{\partial}\psi(\zeta)f(\zeta)}{\zeta - z} \ d\overline{\zeta} \wedge d\zeta \ d\mu(z)$$
$$= \frac{1}{2\pi i} \iint_{\mathbb{C}} f(\zeta)\overline{\partial}\psi(\zeta) \int_{\mathbb{C}} \frac{d\mu(z)}{\zeta - z} \ d\overline{\zeta} \wedge d\zeta = 0.$$

This proves 1.

Corollary 2.20. Let $K \subset \mathbb{C}$ be compact, such that $\mathbb{C} \setminus K$ is connected. Every function which is holomorphic near K can be approximated by polynomials uniformly on K.

Definition 2.21. The *holomorphically convex hull* of K in Ω , written \hat{K} , is the set of $z \in \Omega$ such that for every $f \in A(\Omega)$,

$$|f(z)| \le ||f||_{L^{\infty}(K)}.$$

If $K = \hat{K}$, we say that K is holomorphically convex.

Example 2.22. Let K be as in the previous example; then \hat{K} is the outer disc in the definition of K as an annulus.

It is easy to check that

$$d(K, \mathbb{C} \setminus \Omega) = d(K, \mathbb{C} \setminus \Omega).$$

Recall that ch K denotes the convex hull of K. If K is convex, then K is topologically $\overline{D(0,1)}$, which is clearly holomorphically convex by Runge's approximation theorem. But the connection between convexity and holomorphic convexity is stronger than that.

Proposition 2.23. For any $K, \hat{K} \subseteq \operatorname{ch} K$.

Proof. It is easy to check that ch K is the intersection of half-planes

$$H_{a,c} = \{ z \in \mathbb{C} : \operatorname{Re}(az) \le c \}.$$

Fix $a \in \mathbb{C}$, $c \in \mathbb{R}$; we will show $\hat{K} \subset H_{a,c}$. Let $z \in \hat{K}$. Then $|e^{az}| \leq \max_{w \in K} |e^{aw}|$. So $\operatorname{Re} e^{az} \leq \max_{w \in W} \operatorname{Re} |e^{aw}|$, implying

$$\operatorname{Re} az \le \max_{w \in K} aw \le c.$$

Therefore $z \in \hat{K}$.

Moreover, \hat{K} is the union of K with all \overline{O} , for each \mathbb{C} -precompact connected component O of $\Omega \setminus K$.

Definition 2.24. The polynomially convex hull of a compact set K is the set

$$\hat{K} = \{ z \in \mathbb{C} : |p(z)| \le \max_{w \in K} |p(w)| \text{for every polynomial} p \}.$$

If $K = \hat{K}$, then K is polynomially convex.

2.4 Sheaves

Let X be a topological space. By $\mathbf{Open}(X)$ we will denote the posetal category of open sets in X; that is, objects are open sets in X and morphisms are inclusions.

Definition 2.25. A *presheaf* on X is a functor $\mathbf{Open}(X)^{op} \to C$ for some concrete category C.

If $U \subseteq X$ is an open set, and $\mathcal{F} : \mathbf{Open}(X)^{op} \to C$ a presheaf on X, then elements of $\mathcal{F}(U)$ are called *sections* of \mathcal{F} at U. Sections of \mathcal{F} at X are called *global sections*.

To motivate this terminology, let us assume that we are given a fiber bundle $\pi : E \to X$ (the most trivial case of this is when $E = X \times Y$ for some space Y, and π is projection onto the first factor; as with any notion of bundle, the point is that E is locally a product space.) Then the sections of π , i.e. the sets $\pi^{-1}(U)$, are exactly a sheaf (to be defined later), and thus a presheaf, **Open**(X)^{op} \to **Set**.

Now notice that every morphism in $\mathbf{Open}(X)^{op}$ is epic, so it makes sense to define the restriction maps $f \mapsto f|_V$ for sections $f \in \mathcal{F}(U)$ and $V \subseteq U$.

Definition 2.26. Let \mathcal{F} : **Open** $(X)^{op} \to C$ be a presheaf on X. Assume that for every open set $U \in$ **Open**(X) and every open cover $\{U_j\}$ of U we have the following conditions:

- 1. For every pair of sections $f, g \in \mathcal{F}(U)$, if we have $f|_{U_i} = g|_{U_i}$ for every *i*, then f = g.
- 2. If for every *i* we have a section $f_i \in \mathcal{F}(U_i)$ such that on intersections, $f_i|_{U_i \cap U_j} = f_j|_{U_i \cap U_j}$, then there is a section $f \in \mathcal{F}(U)$ such that for every *i*, $f|_{U_i} = f_i$.

Then we say that \mathcal{F} is a *sheaf* on X.

Proposition 2.27. Let \mathcal{F} be a sheaf which is only defined on an open base, but otherwise satisfying all the conditions. Then \mathcal{F} uniquely determines a sheaf on the entire topology.

We now put sheaves into a category.

Definition 2.28. Let $\mathcal{F}, \mathcal{G} : \mathbf{Open}(X)^{op} \to C$ be presheaves. A morphism of presheaves over X is a natural transformation $\psi : \mathcal{F} \to \mathcal{G}$. If \mathcal{F} and \mathcal{G} are sheaves, then ψ is a morphism of sheaves.

That is, a morphism of sheaves $\psi : \mathcal{F} \to \mathcal{G}$ consists of, for each open set $U \subseteq X$, a morphism $\psi(U) \in \operatorname{Hom}(\mathcal{F}(U), \mathcal{G}(U))$ such that if $U \subseteq V$ for some open set $V \subseteq X$, then the diagram

$$\begin{array}{ccc}
\mathcal{F}(V) & \stackrel{\psi(V)}{\longrightarrow} & \mathcal{G}(V) \\
\downarrow & & \downarrow \\
\mathcal{F}(U) & \stackrel{\psi(U)}{\longrightarrow} & \mathcal{G}(U)
\end{array}$$

commutes.

In several complex variables, we are interested in holomorphic germs. The following family of definitions allows us to talk about germs algebraically.

Definition 2.29. Let C be a category such that for every directed set \mathcal{D} in C, a colimit exists at \mathcal{D} , and let $x \in X$. Let $\mathcal{F} : \mathbf{Open}(X)^{op} \to C$ be a sheaf. The *stalk* of \mathcal{F} at x is the colimit

$$\mathcal{F}_x = \lim_{U \in \mathcal{D}_x} \mathcal{F}(U)$$

where \mathcal{D}_x is the directed set of all open sets $U \ni x$. An element of \mathcal{F}_x is called a *germ* of \mathcal{F} at x.

2.5 Subharmonicity

Definition 2.30. Let X be a topological space. An *upper-semicontinuous function* on X is a function $u: X \to [-\infty, \infty)$ such that for each $s \in \mathbb{R}$, the preimage of the ray $[-\infty, s)$ is open in X.

Notice that there is a dual notion of lower-semicontinuity, by considering the rays $(s, \infty]$. A function which is both upper- and lower-semicontinuous is just continuous, since open rays form a base of the topology of \mathbb{R} .

Definition 2.31. Let $\Omega \subseteq \mathbb{R}^n$ be an open set. A subharmonic function on Ω is a uppersemicontinuous function $u: \Omega \to [-\infty, \infty)$ such that for every compact set $K \subseteq \Omega$ and every continuous function $h: K \to \mathbb{R}$ which is harmonic in K, if $h \leq u$ on ∂K , then $h \leq u$ on K.

This definition makes just as much sense in \mathbb{C}^n , or even on a Riemannian manifold; one just needs a Laplace-Beltrami operator Δ_g , so that we have a notion of harmonicity $\Delta_g h = 0$. By the maximum modulus principle, a harmonic function is already subharmonic (since it cannot attain its maximum on the boundary of a compact set).

Let u be subharmonic. Then if c > 0, cu is subharmonic (simply by replacing h with ch for each h in the definition of subharmonicity). Moreover, if A is a set of subharmonic functions and $u = \sup A$, then u is subharmonic provided that u is upper-semicontinuous (simply by considering the h such that $v \leq h$ for every $v \in A$, which exist since u is upper-semicontinuous and so finite).

Proposition 2.32. Let u_1, \ldots be a decreasing sequence of subharmonic functions. Then $u = \lim_{i \to j} u_i$ is subharmonic.

Proof. Note that

$$\{z \in \Omega : u(z) < s\} = \bigcup_{j} \{z \in \Omega : u_j(s)\}$$

is open so u is upper-semicontinuous. If h, K are as in the definition of subharmonicity and $\varepsilon > 0$ then the set

$$\{z \in \partial K : u_j(z) \ge h(z) + \varepsilon\}$$

is compact and decreasing as $j \to \infty$. The intersection of nonempty compact sets is nonempty, but the intersection is empty by definition of H, so there the sequence is eventually empty. Therefore $u_j \leq h + \varepsilon$ for j large enough. So $u \leq h$.

Now we consider equivalent definitions of subharmonicity.

Proposition 2.33. Let u be an upper-semicontinuous function on $\Omega \subseteq \mathbb{C}$. Let $\delta > 0$ and let $\Omega_{\delta} = \{z \in \Omega : d(z, \Omega^c) > \delta\}$. The following are equivalent:

- 1. u is subharmonic.
- 2. If $D \subseteq \Omega$ is a compact disk, and f is a polynomial such that $u \leq f$ on ∂D , then $u \leq f$ in D.

3. For each $z \in \Omega_{\delta}$,

$$2\pi u(z) \le \int_0^{2\pi} u(z + re^{i\theta}) \ d\theta.$$

4. For each positive measure μ on $[0, \delta]$ and each $z \in \Omega_{\delta}$,

$$2\pi\mu([0,\delta])u(z) \le \int_0^\delta \int_0^{2\pi} u(z+re^{i\theta}) \ d\theta \ d\mu(r).$$
(2.1)

5. For each $z \in \Omega_{\delta}$ there is a positive measure μ on $[0, \delta]$, such that Equation 2.1 holds and such that $\mu((0, \delta]) > 0$.

Proof. Obviously 1 implies 2 and 3 implies 4 implies 5.

Assume 2. To prove 3, let $z \in \Omega_{\delta}$ and $r \leq \delta$. Let D be the disk of all ζ such that $|\zeta - z| \leq r$. If $\varphi(\theta) = \sum_{k} a_{k} e^{ik\theta}$ is a trigonometric polynomial such that $u(z + re^{i\theta}) \leq \varphi(\theta)$ for every $\theta \in [0, 2\pi]$, then $f(\zeta) = a_0 + 2\sum_{k\geq 1} a_k(\zeta - z)^k/r^k$ has $u \leq \operatorname{Re} f$ on ∂D , so on D since f is a polynomial. Plugging in $\theta = 0$, we have

$$u(z) \le a_0 = \frac{1}{2\pi} \int_0^{2\pi} \varphi(\theta) \ d\theta.$$
(2.2)

Since the trigonometric polynomials are an algebra, they are dense in the space of continuous functions. Therefore Equation 2.2 holds for any continuous φ . This proves 3.

Assume 5 and let h, K be as in the definition of subharmonicity. If $M = \sup u - h > 0$ then u - h = M on some nonempty compact set K_0 by semicontinuity of u - h. Let $z_0 \in K_0$. Then

$$\int_0^{2\pi} \int_0^{\delta} (u-h)(z_0+re^{i\theta}) \ d\mu(r) \ d\theta < 2\pi(u-h)(z_0)\mu([0,\delta)).$$

This is a contradiction of 5, so u is subharmonic.

It follows that the class of subharmonic functions is closed under addition, and subharmonicity is a local property. Moreover, if f is holomorphic on Ω , it follows that $\log |f|$ is subharmonic: by the maximum modulus principle, |f| does not attain its maximum on a compact disk D.

Theorem 2.34. Assume $\Omega \subseteq \mathbb{C}$ is open and connected and u is subharmonic on Ω is not identically $-\infty$. Then $u \in L^1_{loc}(\Omega)$ and for any $v \in C^2_{comp}(\Omega)$, $v \ge 0$, $\langle u, \Delta v \rangle \ge 0$. If $u \in C^2(\Omega)$, then $\Delta u \ge 0$.

Theorem 2.35. Suppose $u \in L^1_{loc}(\Omega)$ and for every $v \in C^2_{comp}(\Omega)$, $v \ge 0$, $\langle u, \Delta v \rangle \ge 0$. Then, up to a null set, u is subharmonic. In particular, the mollification of u is subharmonic.

From the above theorem we see that for $u \in L^1_{loc}$, then u is subharmonic exactly when $\Delta u \ge 0$ in the weak sense.

2.6 Operator theory

Theorem 2.36 (Schur). Let X be a σ -finite measure space and let K be an integral operator. If there is a $p: X \to [0, \infty)$ and a $\lambda > 0$ such that

$$\int_X |K(x,y)| p(y) \, dy \le \lambda p(x)$$

and

$$\int_X |K(x,y)| p(x) \ dx \le \lambda p(y)$$

then

Proof. Let u be an integrable simple function. By the Cauchy-Schwarz inequality and Fubini's theorem,

 $||K||_{L^2 \to L^2} \le \lambda.$

$$\begin{split} ||Ku||_{L^{2}}^{2} &= \int_{X} \left| \int_{X} K(x,y)u(y) \ dy \right|^{2} \ dx \\ &\leq \int_{X} \left(\int_{X} |K(x,y)|p(y) \ dy \int_{X} \frac{|u(y)|^{2}}{p(y)} |K(x,y)| \ dy \right) \ dx \\ &\leq \lambda \int_{X} \frac{p(x)}{p(y)} |K(x,y)||u(y)|^{2} \ dy \ dx \\ &\leq \lambda^{2} |u(y)|^{2} \ dy = \lambda^{2} ||u||_{L^{2}}^{2}. \end{split}$$

Since ISF is dense in L^2 we're done.

As a corollary we can easily compute the ℓ^2 norm of a matrix, since every matrix is an integral operator for counting measure on $\{1, 2, ..., n\}$.

Theorem 2.37. Let A be a self-adjoint operator and

$$R_A(\lambda) = (A - \lambda)^{-1}$$

its resolvent. Then we have

$$||R_A(\lambda)|| = \frac{1}{d(\lambda, \sigma(A))}.$$

Notice the utility of this theorem: we do not assume that A is a bounded linear operator, so we cannot use the spectral radius theorem.

Proof. Since λ is in the resolvent set, $d(\lambda, \sigma(A)) > 0$. We can assume A is unbounded; therefore R_A is bounded, and its spectrum consists of 0 and the set of all $1/(\mu - \lambda)$, for $\mu \in \sigma(A)$. By the spectral radius theorem,

$$||(A - \lambda)^{-1}|| = \sup_{\mu \in \sigma(A)} |\lambda - \mu|^{-1}$$

which proves the claim.

Part II Dynamical systems

Chapter 3

Elementary dynamical systems

Definition 3.1. By a *discrete dynamical system* we mean a transformation $T: X \to X$.

We think of $T^n(X)$ as the state of X at time n.

Definition 3.2. By a *continuous dynamical system* we mean a family of transformations $\varphi_t : X \to X$ satisfying the homomorphism assumption $\varphi_{t+s} = \varphi_t \varphi_s$.

We can always build a continuous system from a discrete one and vice versa. Because discrete systems are much easier to study, we usually try to reduce the study of dynamical systems to the study of discrete dynamical systems.

3.1 Types of dynamical systems

Definition 3.3. A *periodic point* of a dynamical system T is a x such that there is a t with $T^t(x) = x$.

In dynamical systems, we want to know how many periodic points are there in a dynamical system. This is too specific so it is also common to look for recurrence: if a trajectory x starts at x_0 , how often does x_n approximate x_0 ?

Definition 3.4. An *invariant set* $Y \subseteq X$ of a dynamical system $T : X \to X$ is a set such that $T^{-1}(A) = A$.

Definition 3.5. An *invariant measure* μ (defined on a σ -algebra Σ) is one such that for every measurable set $A \in \Sigma$, $T^{-1}(A) \in \Sigma$ and $\mu(T^{-1}(A)) = \mu(A)$.

In dynamical systems, we want to study invariant sets and invariant measures. The reason why we study the pullback in the definition of invariant measure is that $A \mapsto \mu(T^{-1}(A))$ is always a measure, but if $X = \{0, 1\}$, T(0) = T(1) = 0, μ counting measure, then $A \mapsto \mu(T(A))$ is not a measure.

We also will study structural stability, i.e. when a small perturbation of T preserves the properties of T.

Example 3.6. KAM theory implies that the solar system is at least approximately structurally stable. Therefore the solar system will not collapse in our lifetimes.
Definition 3.7. Let $T: X \to X$ and $S: Y \to Y$ be dynamical systems. Then S is *semiconjugate* to T, or that T is a *factor* of S) if there is a surjective $\pi: Y \to X$ such that $T \circ \pi = \pi \circ S$. If π is actually invertible, then S and T are *conjugate*.

As far as set theory is concerned, conjugacy means that S and T are identical. If we may assume that π is measure-preserving, smooth, etc., then it will follow that S and T are identical in the appropriate categories.

Example 3.8. Let $T: S^1 \to S^1$ be the dynamical system which rotates the circle by $2\pi\alpha$ for some $\alpha \in \mathbb{R}$. Let $S: \mathbb{R}^2 \setminus 0 \to \mathbb{R}^2 \setminus 0$ to be the rotation of the punctured plane by $2\pi\alpha$. Then T is a factor of S, witnessed by the transformation $\pi: \mathbb{R}^2 \setminus 0 \to S^1$ which sends a point to its projection onto the circle.

We are mainly interested in T when α is irrational (in which case T is called the *irrational* rotation), in which case T has no periodic points. We will show that there are no interesting examples of invariant sets for T, and in fact if we restrict to Borel measures, there is exactly one invariant measure of T, namely the Lebesgue measure. There is no structural stability because the rational numbers are dense in \mathbb{R} .

We branch off into different subfields. Let $T : X \to X$ be a transformation. If X is a metric space and T is continuous, then we are studying *topological dynamics*. If X is a smooth manifold and T is smooth, then we are studying *smooth dynamics*. In this case, we have an auxiliary dynamical system defined by the differential form $df : TX \to TX$ which sends $T_xX \to T_{f(x)}X$. Finally, if X is a measure space and μ is an invariant measure, then we are studying *ergodic theory*. Ergodic theory will be one of the main themes of this course.

One can also study holomorphic dynamics, where $T : \mathbb{C} \to \mathbb{C}$ is a holomorphic function. One can "generalize" this to the study of rational functions from a variety to itself.

We now introduce Hamiltonian systems.

Definition 3.9. Let X be a smooth manifold and let ω be a nondegenerate 2-form such that $d\omega = 0$. Then we say that (X, ω) is a symplectic manifold.

Definition 3.10. Let X be a symplectic manifold. For a function $f : X \to \mathbb{R}$, let H_f be the vector field defined by the relation $\omega(\cdot, H_f) = df$. Then we define the Hamiltonian dynamical system on X by the ordinal differential equation

$$\dot{\rho}(t) = H_f(\rho(t))$$

where $\rho(0)$ is given.

In this case, the measure $\omega^n/n!$ is an invariant measure of ω .

Example 3.11. Let X be the cotangent space of \mathbb{R}^n and let $\omega = \sum_j d\xi_j \wedge dx_j$. Then (X, ω) is a symplectic manifold and

$$H_f = \sum_j \frac{\partial f}{\partial \xi_j} \partial x_j - \frac{\partial f}{\partial x_j} \partial_{\xi_j}$$

Example 3.12. We show that the irrational rotation is a Hamiltonian system. Let $(\mathbb{R}^2, dx \land dy)$ be our symplectic manifold, so

$$H_f = \frac{\partial f}{\partial x} \partial_y - \frac{\partial f}{\partial y} \partial_x$$

Now let

$$f(x,y) = \frac{x^2 + y^2 - 1}{2}$$

so $\dot{x} = -\partial_y f$ and $\dot{y} = \partial_x f$. Then $S = \varphi_{-2\pi\alpha}$.

The irrational rotation is an especially useful example because it is simultaneously a topological, smooth, holomorphic, ergodic-theoretic and Hamiltonian system.

3.2 Properties of the irrational rotation

Let T be the irrational rotation. For every $f: S^1 \to \mathbb{R}$, let

$$S_N f(\theta) = \frac{1}{N} \sum_{j=0}^{N-1} f(T^j(\theta)).$$

Let

$$\overline{f} = \frac{1}{2\pi} \int_0^{2\pi} f(\varphi) \, d\varphi.$$

Theorem 3.13. One has

$$\lim_{N \to \infty} S_N f = \overline{f},$$

pointwise if $f \in C(S^1)$ and in L^2 if $f \in L^2(S^1)$.

Proof. By the Stone-Weierstrass theorem, trigonometric polynomials are dense in $C(S^1)$ (in the L^{∞} topology), so to prove pointwise convergence we just need to check on trigonometric polynomials. It then suffices to check for the trigonometric basis $f(\theta) = e^{i\ell\theta}$. If $\ell = 0$ then this is obvious. If $\ell \neq 0$,

$$S_N f(\theta) = \frac{1}{N} \sum_{j=0}^{N-1} e^{i\ell(\theta + 2\pi j\alpha)} = \frac{e^{i\ell\theta}}{N} \frac{1 - e^{i\ell 2\pi N\alpha}}{1 - e^{i\ell 2\pi\alpha}}$$

Since α is irrational, the denominator is never 0, hence is bounded from below. The numerator is clearly bounded from above, so as $N \to \infty$, $S_N f \to 0$ uniformly. Meanwhile, $\overline{f} = 0$ since f is periodic of period $2\pi \ell$. This proves the pointwise claim.

One can prove L^2 -convergence by taking the Fourier series

$$f = \sum_{\ell} f_{\ell} e^{\ell}$$

where $e^{\ell}(\theta) = e^{i\ell\theta}$. Then $f_0 = \overline{f}$. We have $\langle S_N(e^{\ell}), S_N(e^k) \rangle = 0$ whenever $k \neq \ell$ by orthogonality.

$$\left\| \frac{1}{N} S_N\left(\sum_{\ell} f_{\ell} e^{\ell}\right) - f_0 \right\|_{L^2}^2 \leq \left\| \sum_{|\ell| \leq k} \frac{f_{\ell}}{N} S_N(e^{\ell}) \right\|_{L^2}^2 + \left\| \sum_{|\ell| > k} \frac{f_{\ell}}{N} S_N(e^{\ell}) - f_0 \right\|_{L^2}^2$$

and

$$\left\| \left| \sum_{|\ell| > k} \frac{f_{\ell}}{N} S_N(e^{\ell}) \right| \right\|_{L^2}^2 = \sum_{|\ell| > k} \left\| \left| \frac{f_{\ell}}{N} S_N(e^{\ell}) \right| \right\|_{L^2}^2 \le \sum_{|\ell| > k} |f_{\ell}|^2$$

and for any $\varepsilon > 0$ we may choose k so that the sum over $|\ell| > k$ is at most ε . The sum over $|\ell| \le k$ can also be chosen less than ε by taking N big enough so we're done.

Corollary 3.14. $(2\pi n\alpha)_{n\in\mathbb{N}}$ is dense in S^1 .

Proof. We have for every $f \in C(S^1)$,

$$\frac{1}{N}\sum_{n=0}^{N-1} f(2\pi n\alpha) = \frac{1}{2\pi} \int_0^{2\pi} f(\varphi) \, d\varphi$$

Suppose the claim fails. Then there is a open $U \subseteq S^1$ such that for every $n \in \mathbb{N}$, $n\alpha \notin U$. Let f be zero outside of U, such that $\int f \neq 0$. Then the left-hand side is 0 while the right-hand side is nonzero.

Corollary 3.15. The only invariant Borel probability measure of the irrational rotation is Lebesgue measure.

Proof. Suppose that μ is an invariant measure. Then

$$\int_{S^1} f(x+2\pi n\alpha) \ d\mu(x) = \int_{S^1} f(x) \ d\mu(x)$$

but the $2\pi n\alpha$ are dense, so in fact we have

$$\int_{S^1} f(x+y) \ d\mu(x) = \int_{S^1} f(x) \ d\mu(x)$$

whence μ is rotation-invariance, hence the Lebesgue measure.

Definition 3.16. A dynamical system T with fixed σ -algebra Σ is *uniquely ergodic* if there is a unique T-invariant probability measure on Σ .

So we have just proven that the irrational rotation is uniquely ergodic for the Borel σ -algebra. This is a very unusual property.

Corollary 3.17. Modulo null sets, there are no invariant proper subsets of the irrational rotation.

Proof. Let A be an invariant set and let $f = 1_A$, g = 1 - f. Then

$$\langle S_N f, g \rangle = \int_0^{2\pi} S_N f(\theta) \overline{g(\theta)} \ d\theta = \int_0^{2\pi} f(\theta) \overline{g(\theta)} \ d\theta = 0$$

but also

$$\langle f,g \rangle = |A|(2\pi - |A|)$$

so by the equality we have $|A|(2\pi - |A|) = 0$.

Chapter 4

Ergodic theory

4.1 The mean ergodic theorem

We now extend the above results on the irrational rotation to general dynamical systems.

Let (X, μ) be a probability space, and for a measurable function f and measure-preserving transformation T, let

$$S_n f(x) = \sum_{j=0}^{n-1} f(T^j(x)).$$

be n times the time average of f. For example if f is an indicator function for a set A then $S_n f$ counts the number of times that we visit f. Let

$$\overline{f} = \int_X f \ d\mu$$

be the space average of f. So if f is an indicator function then \overline{f} is the probability of A. For the irrational rotation we proved that $S_n/nf \to \overline{f}$ pointwise and in L^2 . This is a special case of the ergodic theorems.

Lemma 4.1. If $g \ge 0$ is an integrable function then

$$\int_X g \circ T \ d\mu = \int_X g \ d\mu.$$

Proof. For indicator functions this is obvious. By taking sums we extend to simple functions and then use monotone convergence. \Box

Definition 4.2. Let $Uf = f \circ T$, the Koopman operator of T.

By the lemma, U is an isometry on L^2 and we have

$$S_n f(x) = \sum_{j=0}^{n-1} U^j f(x).$$

Theorem 4.3. Let *H* be a Hilbert space and $U \in B(H)$ is an operator such that $||U|| \le 1$. Let

$$Inv = \{ f \in H : Uf = f \}.$$

Let $P: H \to \text{Inv}$ be the orthogonal projection. Then for every $f \in H$,

$$\sum_{1} n \sum_{j=0}^{n-1} U^j f \to P f$$

in L^2 .

The proof of this theorem uses the Banach-Alaoglu theorem and the following lemma:

Lemma 4.4. Ug = g iff $U^*g = g$.

Proof. Since the adjoint is an involution we just need to check one direction. Suppose Ug = g. Then

$$0 = ||U^*g - g||^2 = ||U^*g||^2 + ||g||^2 - 2\operatorname{Re}\langle U^*g, g\rangle = ||U^*g||^2 - ||g||^2 \le ||Ug||^2 - ||g||^2 = 0.$$

Proof of theorem. It suffices to prove the claim when $f \in \text{Inv}^{\perp}$, in which case Pf = 0. We expand

$$S_n f = \sum_{j=0}^{n-1} U^j f$$

as

$$||S_n f/n||^2 = \langle f, S_n^* S_n f/n^2 \rangle$$

and

$$||S_n f/n|| \le \frac{1}{n} \sum_{j=0}^{n-1} ||U^j f|| \le ||f||$$

which is bounded. Now $f \in \text{Inv}^{\perp}$ and $g_n = S_n^* S_n f/n^2$ is a bounded sequence, so g_n has a weak limit. Suppose that g is a weak limit; we claim that $g \in \text{Inv}$, so $\langle f, g \rangle = 0$. But

$$(1 - U^*)(S_n^*/n) = \frac{1}{n} \sum_{j=0}^{n-1} (1 - U^*)(U^*)^{j-1} = \frac{1 - (U^*)^n}{n}$$

whose operator norm is bounded by 2/n, so

$$(1-U^*)g_n \to 0.$$

This implies that $(1 - U^*)g = 0$, so by the lemma Ug = g and $g \in Inv$.

41

Corollary 4.5 (von Neumann mean ergodic theorem). Let (X, μ, T) be a measure-preserving system and $f \in L^2(\mu)$. Let Inv(T) be the space of Koopman-invariant functions for μ and let $P: L^2(\mu) \to Inv(T)$ be the orthogonal projection. Then

$$\frac{1}{n}\sum_{j=0}^{n-1}f(T^j(x)) \to Pf$$

in $L^2(\mu)$.

Note that the mean ergodic theorem does not assume that μ is a probability measure. However, when one applies the mean ergodic theorem he usually wants to assume that indicator functions are in L^2 , which is only possible when μ is a finite measure.

Lemma 4.6. For every $g \in \text{Inv}$ and $f \in L^2$,

$$\int_X (Pf)g \ d\mu = \int_X fg \ d\mu$$

Proof.

$$\langle Pf,g\rangle = \langle f,P^*g\rangle = \langle f,Pg\rangle = \langle f,g\rangle.$$

Corollary 4.7. If A is an invariant set and $\mu(A) < \infty$ then for every $f \in L^2$,

$$\int_A Pf = \int_A f.$$

Corollary 4.8. *P* is a positive-semidefinite operator.

Proof. Suppose $f \ge 0$. Then Pf(x) is the average of the Koopman iterates $f(T^j(x)) \ge 0$. \Box Corollary 4.9. If μ is a probability measure then

$$\int_X Pf \ d\mu = \int_X f \ d\mu$$

Corollary 4.10. If μ is a probability measure and f > 0 a.e. then Pf > 0 a.e. *Proof.* Note that $(Pf)^{-1}(0)$ is invariant since $Pf \in Inv$, and has finite measure, so

$$\int_{(Pf)^{-1}(0)} f = \int_{(Pf)^{-1}(0)} Pf = 0.$$

Since f is nonnegative this implies that Pf = 0 implies f = 0.

Example 4.11. Applying the mean ergodic theorem to the irrational rotation we have

$$Pf = \overline{f}.$$

Example 4.12. Let Σ_m^+ be the Cantor space of infinite sequences in the alphabet $\{0, \ldots, m-1\}$. Define the *shift map*

$$T(x_1x_2x_3\cdots)=x_2x_3x_4\cdots.$$

The dynamical system (Σ_m^+, T) has many invariant measures. If y is a word in $\{0, \ldots, m-1\}$, let C_y denote the cylinder of all sequences which begin with y. A Borel measure is defined uniquely (assuming it is well-defined at all) by assigning measures to each of the C_y . In fact choose a probability vector p, i.e. $p = (p_0, \ldots, p_{m-1})$ with $p_j \in [0, 1]$ such that $\sum_j p_j = 1$, and let

$$\mu_p(C_y) = \prod_j p_{y_j}.$$

Then μ_p is a Borel probability measure on Σ_m^+ , and

$$\mu_p(T^{-1}(C_y)) = \mu_p\left(\bigcup_j C_{j \cdot y}\right) = \sum_j \mu_p(C_{j \cdot y}) = \mu_p(C_y).$$

Therefore μ_p is an invariant measure. As we will prove, this gives an ergodic system that is not uniquely ergodic. In case $p = (1/10, \ldots, 1/10)$, and m = 10, this gives a construction of Lebesgue measure, which can be used to prove that almost every number is normal.

We now prove a recurrence theorem of Caratheodory, which confusingly is not named after Caratheodory.

Theorem 4.13 (Poincare recurrence). Suppose that (X, μ) is a probability space, $T : X \to X$ measure-preserving, and B is measurable. Then for μ -a.e. $x \in B$, there are infinitely many $n \in \mathbb{N}$ such that $T^n x \in B$.

Proof. Let $f = 1_B$. By the mean ergodic theorem,

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) \to_{L^2} Pf(x) > 0$$

a.e. in B. By the Riesz-Weyl theorem, there is a subsequence $(n_k)_k \in \mathbb{N}$ such that

$$\frac{1}{n_k}\sum_{j=0}^{n_k-1}f(T^j(x)) \to_{a.e.} Pf(x).$$

The claimed property is true a.e. for the n_k .

Example 4.14. Suppose we have two chambers connected, and a gas only in one chamber. By the second law of thermodynamics, the gas will spread throughout the two chambers. But by Poincare recurrence, the gas will almost surely return to the first chamber. However, the recurrence time of such a phenomenon may exceed the lifespan of the universe.

There is also a combinatorial proof of Poincare recurrence.

Proof of Poincare recurrence. If there were only finitely many such n we would be able to assume wlog that there were only zero such n, by time-translation. Then for every $x \in B$ and $n \in \mathbb{N}$, $T^n(x) \notin B$. So if $n \neq m$, and $x \in T^{-n}(B) \cap T^{-m}(B)$, then $T^{n-m}(T^m(x)) = T^n(x) \in B$, a contradiction. Therefore the $T^{-n}(B)$ are disjoint and have positive measure, yet there are infinitely many of them and μ is a probability measure, a contradiction. \Box

4.2 Pointwise ergodic theorem

To introduce the pointwise ergodic theorem, we need to define conditional expectation.

Definition 4.15. Let (X, Σ, μ) be a probability space and $J \subseteq \Sigma$ a σ -algebra. Given f a Σ -measurable function, $A \in \Sigma$, let

$$\mu_f(A) = \int_A f \ d\mu.$$

Then define the *conditional expectation*

$$E(f|J) = \frac{d(\mu_f|_J)}{d(\mu|_J)},$$

the Radon-Nikodym derivative of μ_f with respect to μ when restricted to the σ -algebra J.

Example 4.16. If J is the trivial σ -algebra then E(f|J) is the constant function given by the mean E(f).

Example 4.17. Let (X, Σ, μ) be [0, 1] and J be the σ -algebra generated by [0, 1/2]. Then E(f|J)(x) is the average of f on [0, 1/2] if $x \leq 1/2$ or is the average of f on (1/2, 1] for x > 1/2.

We actually have an equivalent definition of conditional expetation which uses the orthogonal projector from the mean ergodic theorem.

Lemma 4.18. Assume that P is a probability measure. Let P be the orthogonal projection $L^2(\mu) \to \operatorname{Inv} T$. Then $||P||_{L^1 \to L^1} \leq 1$, so P extends uniquely to a projection $L^1(\mu) \to \operatorname{Inv} T$.

Proof. Note that for any $f \in L^2(\mu)$, $\{x \in X : Pf(x) \ge 0\}$ its complement are T-invariant sets. So

$$\int_{X} |Pf| = \int_{Pf \ge 0} Pf - \int_{Pf < 0} Pf = \int_{Pf \ge 0} f - \int_{Pf < 0} f \le \int_{Pf \ge 0} |f| + \int_{Pf < 0} |f| = ||f||_{L^{1}(\mu)}.$$

Since μ is a probability measure, $L^2(\mu)$ is a dense subspace of $L^1(\mu)$, whence the claim. \Box

Corollary 4.19. Let J be the σ -algebra of T-invariant sets. Then for any $f \in L^1(\mu)$,

$$E(f|J) = Pf$$

Theorem 4.20 (Birkhoff pointwise ergodic theorem). Suppose that (X, Σ, μ) is a probability space, $T: X \to X$ measure-preserving, and $f \in L^1(\mu)$. Then for a.e. $x \in X$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^{j}(x)) = E(f|J)(x)$$

where J is the σ -algebra of invariant sets,

$$J = \{ A \in \Sigma : T^{-1}(A) = A \}.$$

Corollary 4.21. Let J, μ be as in the pointwise ergodic theorem, and P as in the mean ergodic theorem. If $f \in L^2(\mu)$ then E(f|J) = Pf.

Note that for any open set of reals $U, E(f|J)^{-1}(U) \in J$ iff $E(f|J) \circ T = E(f|J)$. In fact, for any invariant set A,

$$\int_{A} E(f|J) \ d\mu = \int_{A} f \ d\mu$$

Lemma 4.22. The limit in the pointwise ergodic theorem exists μ -a.e.

Proof. Let

$$\overline{f}(x) = \limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x))$$

and similarly for \underline{f} and liminf. Clearly $\overline{f} \geq \underline{f}$ and it suffices to show that $\overline{f} = \underline{f}$, μ -a.e. In fact it suffices to show that

$$\int_X \overline{f} \ d\mu \le \int_X f \ d\mu \le \int_X \underline{f} \ d\mu.$$

Replacing f with -f we see that in fact the seemingly weaker statement

$$\int_X \overline{f} \ d\mu \le \int_X f \ d\mu$$

is sufficient.

Fix M > 0, and let $\overline{f}_M = \min(\overline{f}, M)$. Then $\overline{f}_M \leq M$. Lemma 4.23. $\overline{f}_M > -\infty$, μ -a.e.

Proof of sublemma. For any function $g, g \ge -|g|$. So

$$\int_X \overline{f}_M \ d\mu \ge \int_X \limsup_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} -|f \circ T^j| \ d\mu.$$

By Fatou's lemma,

$$\int_X \overline{f}_M \ d\mu \ge \limsup_{n \to \infty} \frac{1}{n} \int_X \sum_{j=0}^{n-1} -|f \circ T^j| \ d\mu = -\int_X |f| \ d\mu > -\infty$$

since T is measure-preserving.

Fix $\varepsilon > 0$. If $\overline{f}(x) < \infty$ then there is an *n* such that

$$\frac{1}{n}\sum_{j=0}^{n-1}f(T^j(x)) \ge \overline{f}(x) - \varepsilon.$$

Let n(x) be the smallest such n witnessing this. If $\overline{f}(x) = \infty$ then there is a n such that

$$\frac{1}{n}\sum_{j=0}^{n-1}f(T^j(x)) \ge M$$

so we may let n(x) be the smallest such witness. Then

$$\overline{f}_M(x) \le \frac{1}{n(x)} \sum_{j=0}^{n-1} f(T^j(x)) + \varepsilon.$$

For each R > 0, let

$$A_R = \{x \in X : n(x) > R\}.$$

Then the A_R form a chain: if R' > R then $A_{R'} \subseteq A_R$. Moreover, $\bigcap_R A_R = \emptyset$ and μ is a probability measure, so $\mu(A_R) \to 0$ as $R \to \infty$. Since μ is a probability measure, the constants are in $L^1(\mu)$ so

$$\lim_{R \to \infty} \int_{A_R} (|f| + M) \ d\mu = 0.$$

Fix R > 1. Let $n_i(x)$ be defined inductively. Let $n_0(x) = 0$. If $T^{n_i(x)}(x) \notin A_R$, let

$$n_{i+1}(x) = n_i(x) + n(T^{n_i(x)}(x)).$$

Otherwise, $T^{n_i(x)}(x) \in A_R$ and let $n_{i+1}(x) = n_i(x) + 1$, and we have

$$1 \le n_{i+1} - n_i \le R$$

Suppose $T^{n_i(x)}(x) \notin A_R$. Then

$$n(T^{n_i(x)}(x))\overline{f}_M(T^{n_i(x)}(x)) \le \sum_{j=0}^{n(T^{n_i(x)}(x))} f(T^{k+n_i(x)})(x) + n(T^{n_i(x)}(x))\varepsilon.$$

Clearly the constants are invariant functions and \overline{f} is invariant under T. So

$$(n_{i+1} - n_i)(x)\overline{f}_M(x) \le \sum_{j=n_i(x)}^{n_i(x)+1} f(T^j(x)) + (n_{i+1} - n_i)(x)\varepsilon.$$

On the other hand, if $T^{n_i(x)}(x) \in A_R$, then we use the estimate $\overline{f}_M(x) \leq M$ and $n_{i+1}(x) - n(x) = 1$ to see that

$$(n_{i+1} - n_i)(x)\overline{f}_M(x) \le M$$

Let

$$\tilde{f}_M = f + (|f| + M)\mathbf{1}_{A_R}$$

Then in both cases,

$$(n_{i+1} - n_i)(x)\overline{f}_M(x) \le \sum_{j=n_i(x)}^{n_{i+1}(x)} \tilde{f}_M(T^j(x)) + (n_{i+1} - n_i)(x)\varepsilon.$$

Fix some large N; then for any x there is k such that $n_k(x) \leq N \leq n_{k+1}(x)$. Then the series in question telescope and

$$N\overline{f}_{M}(x) \leq \sum_{j=0}^{n_{k}(x)-1} \tilde{f}_{M}(T^{j}(x)) + (N - n_{k}(x))\overline{f}_{M}(x) + n_{k}(x)\varepsilon$$
$$\leq \sum_{j=0}^{N-1} \tilde{f}_{M}(T^{j}(x)) - \sum_{j=n_{k}(x)}^{N-1} \tilde{f}_{M}(T^{j}(x)) + RM + N\varepsilon$$
$$\leq \sum_{j=0}^{N-1} \tilde{f}_{M}(T^{j}(x)) - \sum_{j=N-R}^{N-1} |f(T^{j}(x))| + RM + N\varepsilon$$

since $-|f| \leq |\tilde{f}|$. Dividing both sides by N,

$$\overline{f}_M(x) \le \frac{1}{N} \sum_{j=0}^{N-1} \widetilde{f}_M(T^j(x)) + \frac{1}{N} \sum_{j=N-R}^N |f(T^j(x))| + \varepsilon + \frac{RM}{N}$$

We integrate both sides $d\mu(x)$. Then since $\mu(A_R) \leq \mu(X) = 1$,

$$\begin{split} \int_X \overline{f}_M \ d\mu &\leq \int_X \widetilde{f}_M + \frac{R}{N} \int_X |f| + \varepsilon + \frac{RM}{N} \ d\mu \\ &\leq \int_X f \ d\mu + \int_{A_R} (|f| + M) \ d\mu + \frac{R}{N} ||f||_{L^1(\mu)} + \varepsilon + \frac{RM}{N} \end{split}$$

Taking $N \to \infty$, we have

$$\int_X \overline{f}_M \ d\mu \le \int_X f \ d\mu + \int_{A_R} (|f| + M) \ d\mu + \varepsilon.$$

Taking $R \to \infty$ and $\varepsilon \to 0$,

$$\int_X \overline{f}_M \ d\mu \le \int_X f \ d\mu$$

Let $g_M = \overline{f}_M - \overline{f}_0$. Since the \overline{f}_M are an increasing sequence, $g_M \ge 0$ and the g_M form an increasing sequence. Moreover, we already proved that

$$\int_X \overline{f}_0 \ d\mu > -\infty$$

Since $\overline{f}_0 \leq 0$, $\overline{f}_0 \in L^1(\mu)$. So the monotone convergence theorem implies that $g_M \to \overline{f} - \overline{f}_0$ in $L^1(\mu)$. This implies that

$$\int_X \overline{f} - \overline{f}_0 \le \int_X f - \overline{f}_0 \ d\mu$$

as desired.

We are finally ready to prove the pointwise ergodic theorem.

Proof of pointwise ergodic theorem. It remains to show that the limit is E(f|J). This follows from the mean ergodic theorem if $f \in L^2(\mu)$, since then E(f|J) = Pf. Otherwise, $f \notin L^2(\mu)$ but there is a $g \in L^2(\mu)$ such that $S_ng/n \to Pg$ a.e. and $||f - g|| < \varepsilon$. Moreover, we can find an n such that

$$||S_n f/n - Pf||_1 \le ||S_n (f - g)/n||_1 + ||S_n g/n - Pg||_1 + ||Pf - Pg||_1 \le ||f - g||_1 + ||S_n g/n - Pg||_2 \le ||f - g||_1 < 3\varepsilon.$$

Therefore $S_n f/n \to Pf = E(f|J)$.

Example 4.24. We apply the pointwise ergodic theorem to the shift map. Fix a probability vector p. As a black box, we will assume that the shift map is ergodic; i.e. if $A \subseteq \Sigma_m^+$ is a Borel set and $T^{-1}(A) = A$ then $\mu_p(A)(1 - \mu_p(A)) = 0$. Let ℓ be a letter and $f = 1_{C_\ell}$ the indicator function of its cylinder. Then the Birkhoff ergodic average for n iterates on a sequence x is 1/n times the number of j such that $x_j = \ell$, for $j \leq n$, since

$$T^j(x) = x_{j+1} x_{j+2} \cdots .$$

By the pointwise ergodic theorem, the Birkhoff average converges to p_{ℓ} , μ_p -a.e. As a corollary, we have proven the *law of large numbers*.

4.3 Ergodic systems

Definition 4.25. Let (X, Σ, μ) be a measure space and $T : X \to X$ a μ -invariant transformation. We say that T is an *ergodic transformation* for μ , or that μ is an *ergodic measure* for T, if for every $A \in \text{Inv } T$, $\mu(A) = 0$ or $\mu(A^c) = 0$.

Definition 4.26. Let (X, Σ, μ, T, d) be an ergodic system equipped with a metric d and assume that Σ is the Borel σ -algebra for d. We say that T is *uniquely ergodic* if the choice of μ is unique on Σ .

We have already proven that the irrational rotation is uniquely ergodic, and since we have multiple choices of probability vector, the shift is not uniquely ergodic but is ergodic.

Example 4.27. We consider the multiplication map by m on the circle, namely, $T: S^1 \to S^1$ has $T(x) = mx \mod 1$ for some m > 0. Then T is ergodic for Lebesgue measure μ , in a particularly strong form, namely that

$$\lim_{n \to \infty} \mu(A \cap T^{-n}(B)) = \mu(A)\mu(B).$$

If we put $B = A^c$ this obviously implies that μ is a *T*-ergodic measure. To prove this claim we use this lemma.

Lemma 4.28. Let T be the multiplication map, $f, g \in L^2(\mu)$. Then

$$\int_0^1 f \circ T^n(x)g(x) \, dx = \int_0^1 f \int_0^1 g.$$

Proof. It suffices to check on an orthonormal basis of $L^2(\mu)$, namely $e_{\ell}(x) = e^{2\pi i x \ell}$. Then

$$\int_0^1 e_\ell \circ T^n(x) e_k(x) \ dx = \int_0^1 e^{2\pi i (\ell m^n + k)x} \ dx$$

As $n \to \infty$ and either $\ell \neq 0$ or $k \neq 0$ then this integral goes to 0. Otherwise, if $\ell = k = 0$, the integral is 1. This proves the lemma on an orthonormal basis.

The multiplication map has many invariant measures. In fact, T has many periodic points and fixed points. The fixed points are intersections of the graph of T with the identity map, of which there are a positive but finite set, which grows as $m \to \infty$. Replacing T with T^n replaces m with m^n which clearly $\to \infty$ as n does. Fixed points of T^n are periodic points, so the periodic points are dense.

Example 4.29. We show that the shift map on Σ_m^+ are ergodic. We identify $\overline{x} \in \Sigma_m^+$ with the number $x \in [0, 1]$ of which \overline{x} is a *m*-ary expansion. The map $\pi(\overline{x}) = x$ is not injective because *m*-ary expansions are not unique. If we let $p = (1/m, \ldots, 1/m)$ then $\pi_*\mu_p$ is Lebesgue measure and *T* is conjugate up to a null set to the multiplication map by *m*. If m = 3 and p = (1/2, 0, 1/2) then $\pi_*\mu_p$ is supported on the standard null Cantor set, which is its Hausdorff measure (the *Cantor-Lebesgue measure*).

For any cylinder C_y generated by the word y,

$$\mu_p(T^{-1}(C_y)) = \mu_p \bigcup_{\ell=0}^{m-1} c_{\ell \cdot y} = \sum_{\ell=0}^{m-1} \mu_p(C_{\ell \cdot y}) = \sum_{\ell=0}^{m-1} p_\ell p_y = \mu_p(C_y).$$

So μ_p is T-invariant. Moreover, if $N > \ell > 1$ then

$$\mu_p(T^{-N}(C_y \cap C_z)) = \mu_p \bigcup_{|x|=N} C_{x \cdot y} \cap C_z = \mu_p \bigcup_{|x|=N-\ell} C_{z \cdot x \cdot y} = \mu_p(C_z)\mu_p(C_y)$$

where |x| denotes the length of the word x. This proves that μ_p is an ergodic measure, and since p is very far from unique, we have lots of ergodic measures for T. Thus $\pi_*\mu_p$ is an ergodic measure for the multiplication map. In fact these measures are often mutually singular:

Lemma 4.30. Let p, p' be probability vectors. Then $\mu_p \perp \mu_{p'}$.

Proof. We prove this in the case m = 2. Let p = (P, 1 - P) and p' = (P', 1 - P'). For any P, let

$$F_P = \{ x \in \Sigma_m^+ : \lim_{n \to \infty} \frac{\operatorname{card}\{x_i = 0 : i \le n\}}{n} = P \}.$$

Then if $P \neq P'$, $F_P \cap F_{P'}$ is empty. Let f(x) = 1 if $x_1 = 0$ and f(x) = 0 otherwise. Then the ergodic theorem implies that

$$\frac{\operatorname{card}\{x_i : i \le n\}}{n} = \frac{S_n f(x)}{n} \to \int f = P,$$

 μ_p -a.e. Therefore $\mu_p(F_P) = \mu_p(\Sigma_m^+) = 1$. So $\mu_p(F_{P'}) = 0$ whence $\mu_p \perp \mu_{p'}$.

This is very shocking because in the weakstar topology, μ_p can be approximated arbitrarily well by the $\mu_{p'}$ as $p' \to p$.

4.4 **Properties of ergodic transformations**

Theorem 4.31. For (X, μ) a measure space, $T : X \to X$ is ergodic iff for every measurable $f : X \to \mathbb{R}$, $f \circ T = f$ a.e. implies that f is constant a.e.

Proof. Let c_1 be the supremum of all t such that $\mu(f^{-1}((-\infty, t))) = 0$ and c_2 the infimum of all t such that $\mu(f^{-1}([t, \infty])) = 0$. Then the union of these two preimages is X so $c_1 = c_2$, and $f = c_1$ a.e.

We now restrict to the case that X is a compact metric space.

Definition 4.32. Let X be a compact metric space. Let $\mathcal{M}(X)$ denote the set of all Borel probability measures on X. If $T: X \to X$ is continuous, let $\mathcal{M}(X,T)$ be the set of all T-invariant Borel probability measures and $\mathcal{M}_e(X,T)$ the set of ergodic measures in $\mathcal{M}(X,T)$.

Example 4.33. Ergodic transformations are not dense in a typical topology for mappings $X \to X$. In fact, if it was, then a slight perturbation of the solar system would make the orbit of the earth dense in the solar system, which would be quite bad.

Theorem 4.34 (Krylov-Bogolibabov). The set $\mathcal{M}(X,T)$ is nonempty.

That is, every transformation has an invariant measure. In fact, this measure can be taken to be a Radon measure.

Proof. Let $S_n f(x) = \sum_{j=0}^{n-1} f(T^j(x))$ be the ergodic average of f, as usual. Since X is compact, C(X) is separable and we may choose $(\varphi_m)_m$ to witness this. Fix (x,m) and consider $S_n(\varphi^m(x))/m$. This sequence is bounded in \mathbb{R} , so has a congergent subsequence, and by Cantor's diagonal argument we may choose the subsequence to converge for every m, say to $J(\varphi_m)$. Then $|J(\varphi_m)| \leq ||\varphi^m||_{\infty}$, and J is a positive functional, so J extends to a Radon probability measure μ on X. Moreover,

$$\frac{S_{n_k}}{n_k}\varphi_m(T(x)) = \frac{S_{n_k}}{n_k}\varphi_m(x) + \frac{\varphi_m(T^{n_k})(x) - \varphi_m(x)}{n_k}$$

which as the subsequence $n_k \to \infty$ converges to

$$J(\varphi_m \circ T) = J(\varphi_m) + 0$$

whence μ is *T*-invariant.

It is easy to see that $\mathcal{M}(X,T)$ is weakstar closed, and that it is convex. Since $\mathcal{M}(X)$ is bounded, the Banach-Alaoglu theorem implies that $\mathcal{M}(X,T)$ is actually weakstar compact.

Theorem 4.35. The set $\mathcal{M}_e(X,T)$ is the set of extreme points of $\mathcal{M}(X,T)$.

Proof. Assume μ is not ergodic. Then there is an invariant set A such that $0 < \mu(A) < 1$. Let $\mu|B$ denote the restriction of μ to the measurable set B. We have

$$\mu = \mu(A) \frac{\mu|A}{\mu(A)} + \mu(A^c) \frac{\mu|A^c}{\mu(A^c)}.$$

Then $\mu |A/\mu(A)$ and $\mu |A^c/\mu(A^c)$ are invariant Borel probability measures, so μ is not an extreme point.

Now suppose μ is ergodic and suppose $\mu = t\mu_1 + (1-t)\mu_2$ and 0 < t < 1, where μ_1, μ_2 are invariant measures. We claim $\mu_1 = \mu_2$, which implies that μ is an extreme point. In fact, μ_1, μ_2 are absolutely μ -continuous. Let $\rho_1 \in L^1(\mu)$ be the Radon-Nikodym derivative of μ_1 with respect to μ . Then $\rho_1 \ge 0$. Let $F = \{x \in X : \rho_1(x) < 1\}$. Then $\mu(T^{-1}F) = \mu(F)$ and some set-theoretic computations prove

$$\mu(F \setminus T^{-1}F) = \mu(T^{-1}F \setminus F).$$

Suppose $\mu(F \setminus T^{-1}F)$ is nonzero. Then

$$\mu(F \setminus T^{-1}F) > \int_{F \setminus T^{-1}F} \rho_1 \ d\mu = \int_{T^{-1}F \setminus F} \rho_1 \ d\mu \ge \mu(T^{-1}F \setminus F)$$

which is a contradiction. Therefore $\mu(F \setminus T^{-1}F) = 0$, so F is almost T-invariant.

Since μ is ergodic, either $\mu(F) = 0$ or $\mu(F) = 1$. If $\mu(F) = 1$ then

$$1 = \mu_1(X) = \int_F \rho_1 \ d\mu < \mu(F) = 1$$

which is a contradiction. Therefore F is μ -null. Something similar happens on the set of x such that $\rho_1(x) > 1$. Therefore $\rho_1 = 1$ in $L^1(\mu)$. So $\mu = \mu_1$.

Corollary 4.36. Every transformation has an ergodic measure.

Proof. Use the Krein-Milman theorem.

Corollary 4.37. If $\mu, \nu \in \mathcal{M}_e(X, T)$ then either $\mu = \nu$ or $\mu \perp \nu$.

Proof. By the Radon-Nikodym-Lebesgue decomposition, we can write

 $\mu = t\nu_1 + (1-t)\nu_2$

where $\nu_1, \nu_2 \in \mathcal{M}(X, T)$, ν_1 is ν -a.c. and $\nu_2 \perp \nu$. It follows that t = 0.

4.5 Mixing transformations

Definition 4.38. Let (X, μ) be a probability space. We say $T : X \to X$ is *mixing* if for every measurable sets A, B,

$$\lim_{n \to \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

Lemma 4.39. A mixing transformation is ergodic.

Proof. For any invariant set A,

$$0 = \lim_{n \to \infty} \mu(A \cap T^{-n}(A^c)) = \mu(A)\mu(A^c)$$

so $\mu(A) = 0$ or $\mu(A^c) = 0$.

51

Example 4.40. Let $X = \{0, 1\}$ and $\mu(\{0\}) = \mu(\{1\}) = 1/2$. Let T be the only nontrivial bijection $X \to X$. Then the invariant sets are X, \emptyset so (X, μ, T) is ergodic. But then

$$\mu(T^{-n}(\{0\}) \cap \{0\})$$

oscillates between 1/2 and 0 so the system is not mixing. Here the problem is periodicity.

Lemma 4.41. A system (X, μ, T) is mixing iff there is a $D \subset L^2(\mu)$ such that the span of D is dense in $L^2(\mu)$ and for every $f, g \in D$,

$$\lim_{n \to \infty} \int_X f(T^n(x))g(x) \ d\mu(x) = \int_X f \ d\mu \int_X g \ d\mu$$

Proof. The set of all indicator functions is an example of such a D.

In other words, any two random variables become "approximately independent" as time goes on.

Example 4.42. The multiplication map is mixing; here D is the standard basis of the trigonometric polynomials.

Example 4.43. The irrational rotation is not mixing. Again this can be checked on the standard basis of trigonometric polynomials. This is because the irrational rotation is approximable arbitrarily well by rational rotations, which are periodic.

Let $M \in GL(n, \mathbb{Z})$. Since M extends to an element of $GL(n, \mathbb{R})$ which preserves \mathbb{Z}^n , M drops to an automorphism of the torus \mathbb{T}^n . Moreover, Lebesgue measure drops to a measure on \mathbb{T}^n , so \mathbb{T}^n is a probability space.

Theorem 4.44. Let $M \in GL(n, \mathbb{Z})$ viewed as an automorphism of \mathbb{T}^n . If det $M \neq 0$ then M is measure-preserving. Moreover, if Spec M does not contain any roots of unity, then M is mixing.

Proof. We note that M is measure-preserving if and only if for every $f \in L^2(\mathbb{T}^n)$,

$$\int_{\mathbb{T}^n} f \circ M = \int_{\mathbb{T}^n} f$$

and in fact it suffices to check this on a dense subset D of $L^2(\mathbb{T}^n)$. In fact we let $D = \{e_\ell : \ell \in \mathbb{Z}^n\}$ where

$$e_{\ell}(x) = e^{2\pi i \langle \ell, x \rangle}.$$

Then D is dense in $L^2(\mathbb{T}^n)$ because every function in $L^2(\mathbb{T}^n)$ has a multivariate Fourier series. Moreover,

$$\int_{\mathbb{T}^n} e_\ell \circ M = \int_{\mathbb{T}^n} e^{2\pi i \langle Mx,\ell \rangle} \, dx = \int_{\mathbb{T}^n} e^{2\pi i \langle x,M^t\ell \rangle} \, dx$$

Since det $M^t \neq 0$, $M^t \ell \neq 0$ if $\ell \neq 0$ and $M^t 0 = 0$. But then

$$\int_{\mathbb{T}^n} e^{2\pi i \langle x, M^t \ell \rangle} \, dx = \int_{\mathbb{T}^n} e_\ell.$$

For mixing, note that

$$\int_{\mathbb{T}^n} e_\ell \circ M^N \overline{e_k} = \int_{\mathbb{T}^n} e^{2\pi i \langle x, (M^N)^t \ell - k} \, dx.$$

In fact the integral is zero unless $(M^N)^t \ell - k = 0$, in which case the integral is 1. If $\ell \neq 0$ or $k \neq 0$ then the integral is 0 unless $(M^N)^t \ell = k$. Therefore the limit is nonzero if there are $N_1 < N_2$ such that $(M^{N_1})^t \ell = (M^{N_2})^t \ell = k$. Solving for k this happens iff $(M^{N_2-N_1})^t \ell = \ell$. But this means that there is an eigenvalue λ such that $\lambda^{N_2-N_1} = 1$, which contradicts our hypothesis.

Example 4.45. Suppose $M = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. Then det M = 1 and M sends \mathbb{T}^2 to itself, in fact is mixing and invertible. In particular M is ergodic for Lebesgue measure.

We show that M is not uniquely ergodic by looking for periodic points, i.e. x such that there exists n with $M^n x = x$. For n = 1, 0 is periodic (i.e. a fixed point). In fact it is the unique fixed point. Thus we let ν be the point mass centered at 0. Clearly ν is a Borel measure which is invariant, and in fact ergodic.

Stronger, periodic points are dense in \mathbb{T}^2 , namely rational points are periodic. We count the number of points of period n. Let $Fix(n) = \{x \in \mathbb{T}^2 : M^n x = x\}$. Viewing $[0, 1)^2$ as the fundamental domain of \mathbb{T}^2 we have

$$Fix(n) = \{ x \in [0,1)^2 : (M^n - 1)x \in \mathbb{Z}^2 \}.$$

We compute the integer points in the parallogram $(M^n - 1)[0, 1)^2$ using Pick's theorem. This is the area of $(M^n - 1)[0, 1)^2$, i.e.

$$\det(M^{n} - 1) = (\lambda^{n} - 1)(\lambda^{-n} - 1) = \lambda^{n} + \lambda^{-n} - 2$$

where λ is the largest eigenvalue of M (so λ^{-1} is its smallest eigenvalue.) Thus

card Fix
$$(n) = \lambda^n + \lambda^{-n} - 2.$$

In particular, M is far from uniquely ergodic.

We now introduce the Artin-Mazur zeta function defined by

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{z^n \operatorname{card} \operatorname{Fix}(n)}{n}\right).$$

For any C^{∞} diffeomorphism of a compact manifold this zeta function has a positive radius of convergence. We check this for M. In fact

$$\zeta(z) = \exp\left(\sum_{n=1}^{\infty} \frac{(\lambda^n + \lambda^{-n} - 2)z^n}{n}\right) = \exp\left(\sum_{n=1}^{\infty} \frac{(\lambda z)^n}{n} + \frac{(\lambda^{-1}z)^n}{n} - \frac{2z^n}{n}\right)$$

and since

$$\log(1-z) = -\sum_{n=1}^{\infty} \frac{z^n}{n}$$

it follows that

$$\zeta(z) = \frac{(1-z)^2}{(1-\lambda z)(1-\lambda^{-1}z)}$$

This is a rational function whose smallest pole is $-\lambda^{-1}$. This is related to the fact that the rate of growth of Fix(n) is asymptotically λ^n .

Theorem 4.46. Suppose that $M = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, viewed as an automorphism of \mathbb{T}^2 . Then there is a $\varepsilon > 0$ such that for every $g = M + \delta g$, where $\delta g : \mathbb{T}^2 \to \mathbb{T}^2$ satisfies $||\delta g||_{C^1} < \varepsilon$, there is a homeomorphism $h : \mathbb{T}^2 \to \mathbb{T}^2$ such that $h \circ g = M \circ h$.

To prove this we need to use *Neumann series*; namely, if ||A|| < 1 is a linear operator then

$$(1-A)^{-1} = \sum_{k=0}^{\infty} A^k$$

Proof. We construct h to be of the form $h = id + \delta h$ where $||\delta h||_{C^1}$ is small. If this is to hold then

$$(\mathrm{id} + \delta h) \circ (M + \delta g) = M \circ (\mathrm{id} + \delta h)$$

which simplifies to

$$M^{-1} \circ (\mathrm{id} + \delta h) \circ (M + \delta g) = \mathrm{id} + \delta h$$

i.e.

$$M^{-1} \circ \delta g + M^{-1} \circ \delta h \circ h = \delta h.$$

Thus δh is a fixed point of the mapping

$$f \mapsto M^{-1} \circ \delta g + M^{-1} \circ f \circ g.$$

Unfortunately we cannot use the contraction fixed point theorem because M^{-1} is not contracting (since $\lambda > 1$).

Let $\{e_+, e_-\}$ be the normalized eigenbasis of M. Decompose δh as

$$\delta h = h_+ e_+ + h_- e_-$$

and similarly for δg . Then the fixed-point equation is

$$\lambda^{-1}g_+ + \lambda^{-1}h_+ \circ g = h_+,$$

$$\lambda g_- + \lambda h_- \circ g = h_-.$$

Let $X = C(\mathbb{T}^2 \to \mathbb{R})$. Let $F_+f = \lambda^{-1}g_+ + \lambda^{-1}fg$ and $F_-f = -g_- \circ g^{-1} + \lambda^{-1}f \circ g^{-1}$. Then F_+, F_- send X to itself. Since $||M - g||_{C^1} = ||\delta g||_{C^1}$ is small and M is invertible, g is invertible by the inverse function theorem.

Then F_+, F_- are contractions in L^{∞} , so by the contraction fixed point theorem, there are h_{\pm} which are fixed points of F_{\pm} . So

$$h = \mathrm{id} + h_+ e_+ + h_- e_-$$

is the desired function. But h may not be a homeomorphism. To fix this, we construct $\tilde{h} = \mathrm{id} + \delta \tilde{h}$ such that $\delta \tilde{h}$ is periodic and $||\delta \tilde{h}||_{C^1}$ is small such that $g \circ \tilde{h} = \tilde{h} \circ M$. Then $h^{-1} = \tilde{h}$.

Note that the fixed point f of the contraction fixed-point theorem in a Banach space is given by the telescoping sum

$$f = f_0 + \sum_{n=0}^{\infty} f_{n+1} - f_n$$

where f_0 is the initial datum of the iteration and $f_{n+1} = Ff_n$, F the given contraction. We assume that $f_0 = 0$ and θ is the constant of contraction for F, then

$$||f|| \le \sum_{n=0}^{\infty} ||f_{n+1} - f_n|| \le \frac{\theta}{1-\theta} ||f_1||.$$

In this case,

$$||h_{+}||_{L^{\infty}} \le \frac{1}{\lambda - 1}||g_{+}||_{L^{\infty}}$$

and

$$||h_{-}||_{L^{\infty}} \le \frac{1}{1-\lambda^{-1}}||g_{-}||_{L^{\infty}}.$$

Now \tilde{h} has the required properties iff

$$\delta \tilde{h} \circ M - M - \delta \tilde{h} = \delta g \circ (\mathrm{id} + \delta \tilde{h})$$

So we define a linear operator L by

$$Lf = f \circ M - M \circ f.$$

We now write $\delta \tilde{h}$ in the eigenbasis $\{e_{-}, e_{+}\}$, so

$$L(\tilde{h}_{-}e_{-} + \tilde{h}_{+}e_{+}) = e_{-}\tilde{h}_{-} \circ M + e_{+}\tilde{h}_{+} \circ M - \tilde{h}_{-}\lambda^{-1}e_{-} - \tilde{h}_{+}\lambda e_{+}.$$

The equation decouples so we have

$$L_{\pm}\tilde{h}_{\pm} = \tilde{h}_{\pm} \circ M - \lambda^{\pm 1}\tilde{h}_{\pm}.$$

We invert the operators L_{\pm} using a Neumann series. In fact,

$$L_{-}\tilde{h}_{-} = (\tilde{h}_{-} - \lambda^{-1}\tilde{h}_{-} \circ M^{-1}) \circ M$$

Therefore

$$L_{-}^{1}H = \sum_{n=0}^{\infty} \lambda^{-n}H \circ M^{-n-1}.$$

One can then show

$$L_{+}^{-1}H = -\sum_{n=0}^{\infty} \lambda^{-1-n}H \circ M^{n}.$$

Then

$$L^{-1}(H_{+}e_{+} + H_{-}e_{-}) = L_{+}^{-1}H_{+}e_{+} + L_{-}^{-1}H_{-}e_{-}$$

So it suffices to construct $\delta \tilde{h}$ such that

$$\delta \tilde{h} = L^{-1}(\delta g \circ (\mathrm{id} + \delta \tilde{h}))$$

Let $G(f) = L^{-1}(\delta g \circ (\mathrm{id} + f))$. We must show G is a contraction. In fact,

$$||G(f_1 - f_2)||_{C^0} \le ||L^{-1}||_{C^0 \to C^0} ||\delta g||_{C^1} ||f_1 - f_2||_{C^0} < \varepsilon ||L^{-1}||_{C^0 \to C^0} ||f_1 - f_2||_{C^0}.$$

So if $||L^{-1}||_{C^0 \to C^0} > 1/\varepsilon$ it follows that G is a contraction of C^0 .

Therefore h exists so

$$h \circ g \circ h = h \circ M$$

whence M commutes with $h \circ \tilde{h} = id + f$ for some periodic function f. In particular f commutes with M, and writing f in the eigenbasis,

$$\lambda f_{+}e_{+} + \lambda^{-1}f_{-}e_{-} = (f_{+} \circ M)e_{+} + (f_{-} \circ M)e_{-}.$$

Decoupling again,

$$f_{\pm} = \lambda^{-n} f_{\pm} M^{\pm n}$$

and iterating f_{\pm} sends it to 0, so $f_{\pm} = 0$.

We consider when the above stability properties generalize. Let $S \in C^{\infty}(\mathbb{T}^2 \to \mathbb{T}^2)$ be a diffeomorphism and consider the mapping of the tangent space

$$dS(x): T_x \mathbb{T}^2 \to T_{S(x)} \mathbb{T}^2.$$

Suppose that each tangent space splits

$$T_x \mathbb{T}^2 = E_+(x) \oplus E_-(x)$$

where dS(x) sends $E_{\pm}(x)$ to $E_{\pm}(S(x))$ and $x \mapsto E_{\pm}(x)$ is a continuous mapping into the Grassmannian. In the above proof we used the existence of a $\theta \in (0, 1)$ such that

$$||dT^n(x)v|| \le C\theta^{|n|}||v||$$

where $|| \cdot ||$ is the norm on $T_x \mathbb{T}^2$ induced by the Riemannian metric of \mathbb{T}^2 and if $v \in E_-(x)$ then the inequality ranges over $n \ge 0$, if $v \in E_+(x)$ then $n \le 0$. Such a diffeomorphism is called an *Anosov diffeomorphism*. Very few compact Riemannian manifolds admit Anosov diffeomorphisms, and they have quite remarkable properties. For example, if S is Anosov and measure-preserving then S is mixing, but it is not even known if an Anosov diffeomorphism is necessarily measure-preserving.

More concretely, let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be the action of a matrix $M \in \mathrm{SL}(2,\mathbb{Z})$. Suppose M has eigenvalues $\lambda^{\pm 1}$ and $m = \lambda + \lambda^{-1}$ is the trace of M, so $m \in \mathbb{Z}$. If $m^2 > 4$ then we say that M is a hyperbolic matrix and in this case T acts just like the cat map, and in fact is mixing. If $m^2 = 4$ then M is said to be a parabolic matrix, is conjugate to a shear matrix, and T is known as a Dehn twist. If $m^2 < 4$ then m = 0 (in which case $\lambda = \pm i$) or $m^2 = 1$ (in which case $\lambda = \pm e^{i\pi/3}$). Either way the matrix is idempotent and hence is a rational rotation. We say that M is an elliptic matrix, and note that T is not ergodic.

4.6 The Hopf argument

We introduce a technique to show that certain transformations of Riemannian manifolds are ergodic.

Definition 4.47. Let X be a metric space and $T : X \to X$ be continuous. The *stable* manifold of $x \in X$ is the set $W^s(x)$ of $y \in X$ such that

$$\lim_{n \to \infty} d(T^n(x), T^n(y)) = 0.$$

If T is a homeomorphism then we also define the *unstable manifold* $W^{u}(x)$ to be the stable manifold for T^{-1} .

That is, $W^{s}(x)$ is the set of y such that forward orbits of x and y approach each other.

Example 4.48. Suppose T is the cat acting on \mathbb{T}^2 . (Henceforth we consider the action of any hyperbolic matrix to be a cat map, since by a number-theoretic argument the eigenvalues of a hyperbolic matrix cannot be rational.) Then $W^s(x) = x + e_-\mathbb{R}$. Since e_- is not a rational point, $e_-\mathbb{R}$ is dense in \mathbb{T}^2 for the same reason that orbits of the irrational rotation are dense. So $W^s(x)$ is dense in \mathbb{T}^2 , as is $W^u(x) = x + e_+\mathbb{R}$.

Theorem 4.49 (Banach-Saks). Let H be a Hilbert space and suppose that $f_n \to f$ in the weakstar topology of H. Then there is a subsequence $(f_{n_k})_k$ such that the averages

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f_{n_k} = f$$

in the norm of H.

Proof. Replacing f_n with $f_n - f$ we may assume that f = 0. We will choose the subsequence so

$$\sum_{i < j} |\langle f_{n_i}, f_{n_j} \rangle| \le 2.$$

This is proven by induction. Let $n_1 = 1$, and suppose we have chosen n_1, \ldots, n_k so that whenever $i < j \leq k$,

$$|\langle f_{n_i}, f_{n_j} \rangle| < 2^{-j}.$$

To choose n_{k+1} , we use the fact that for every fixed $i \leq k$,

$$\lim_{n \to \infty} \langle f_{n_i}, f_n \rangle = 0$$

by definition of the weak star topology. Therefore there is a \boldsymbol{n}_{k+1} such that

$$|\langle f_{n_i}, f_{n_{k+1}}\rangle| < \frac{1}{2^{k+1}}$$

This completes the induction, and

$$\sum_{i < j} |\langle f_{n_i}, f_{n_j} \rangle| \le \sum_{j=1}^{\infty} j 2^{-j} \le 2.$$

Thus

$$\left\| \frac{1}{n} \sum_{k=1}^{n} f_{n_{k}} \right\|^{2} \leq \frac{1}{n^{2}} \sum_{k=0}^{n} ||f_{n_{k}}||^{2}$$
$$= \frac{2}{n^{2}} \sum_{1 \leq i < j \leq n} |\langle f_{n_{i}}, f_{n_{j}} \rangle|$$
$$\leq \frac{1}{n} \sup_{n} ||f_{n}||^{2} + \frac{4}{n^{2}} = O(n^{-1})$$

since the $f_n \to 0$ weakly and hence are bounded.

Definition 4.50. Let (X, d, μ) be a metric Borel probability space, $T : X \to X$ measurable. A function $f : X \to \mathbb{C}$ is W^s -invariant if there is a X_0 such that $\mu(X_0) = 1$ and for every $x, y \in X_0$, such that $y \in W^s(x)$, f(x) = f(y).

Theorem 4.51. Let (X, d, μ) be a metric Radon probability space and $f \in L^2(\mu)$. If $T : X \to X$ is continuous then every weakstar accumulation point of $\{f \circ T^n\}_n$ is W^s -invariant. If T is also invertible then every weakstar accumulation point is W^u -invariant.

We use the fact that μ is a Radon probability measure to apply the below lemma.

Lemma 4.52. If μ is a Radon probability measure then the Lipschitz functions are dense in $L^2(\mu)$.

Proof. We need only to show that the indicator functions can be approximated by Lipschitz functions (here we are using that μ is a probability measure). If A is measurable then there is an open set U such that $A \subseteq U$, $\mu(U \setminus A) < \varepsilon$, and in particular

$$||1_A - 1_U|| < \sqrt{\varepsilon}.$$

Therefore we only need to approximate 1_U by Lipschitz functions. Let

$$u_k(x) = \min(1, kd(x, X \setminus U)).$$

Then the u_k are Lipschitz and tend to 1_U as $k \to \infty$.

Proof of Theorem 4.51. Suppose g is a weakstar accumulation point of $\{f \circ T^n\}_n$; passing to a subsequence we may assume that $\lim_n f \circ T^n = g$ in the weakstar topology.

If f is Lipschitz (and M its Lipschitz seminorm) then by applying the Banach-Saks theorem and passing to a subsequence we may assume that

$$\lim_{n \to \infty} \Psi_n(x) = \lim_{n \to \infty} \sum_{k=1}^n f \circ T^n = g$$

in the norm topology, where Ψ_n is the avergae of the first *n* terms. By the Riesz-Fisher theorem and passing to a subsequence again, we may assume that $\Psi_n \to g$ a.e. Since *f* is Lipschitz,

$$|\Psi_{\ell}(x) - \Psi_{\ell}(y)| \le \frac{1}{\ell} \sum_{k=1}^{\ell} |f \circ T^{k}(x) - f \circ T^{k}(y)| \le \frac{M}{\ell} \sum_{k=1}^{\ell} d(T^{k}(x), T^{k}(y)) < \varepsilon$$

if ℓ is large and $y \in W^s(x)$. But $||g(x) - g(y)| - |\Psi_\ell(x) - \Psi_\ell(y)|| < \varepsilon$ if ℓ is large so g is W^s -invariant.

By Lemma 4.52, if $f \in L^2(\mu)$, we can find a Lipschitz function f' such that $||f - f'|| < \varepsilon$ and then $||f' \circ T^n||_{L^2} < ||f||_{L^2} + \varepsilon$ whence we can pass to a subsequence and assume $f' \circ T^n \to g'$ for some g'. Then

$$||g - g'|| \le \liminf_{n \to \infty} ||(f - f') \circ T^n||_{L^2} \le ||f - f'|| < \varepsilon.$$

Taking $\varepsilon = 1/k$ we let g'_k be the given g'. Then $g'_k \to g$ in L^2 and by the Riesz-Fischer theorem again we may again pass to a subsequence and assume $g'_k \to g$ a.e., which implies that g is W^s -invariant.

Now let I be the space of functions in $L^2(\mu)$ which are W^u -invariant. Then I is L^2 -closed, since any sequence in I has an a.e. convergent subsequence by the Riesz-Fischer theorem, and the limit of such a subsequence is W^u -invariant. In addition the Koopman operator of T preserves I.

We will prove that if $f \in I^{\perp}$ then $f \circ T^n \to 0$ in the weakstar topology. Applying the above argument to T^{-1} that if $f \circ T^{-n} \to g_0$ in the weakstar topology then g_0 is W^u -invariant. Suppose $f \circ T^{n_k} \to g$ in the weakstar topology. Since T is measure-preserving we have

$$\int_X (f \circ T)g \ d\mu = \int_X f(T(x))g(T(T^{-1}(x))) \ d\mu(x) = \int f(x)g(T^{-1}(x)) \ d\mu(x)$$

whence $T^* = T^{-1}$. Therefore

$$\langle g,g\rangle = \lim_{n \to \infty} \langle f \circ T^n,g\rangle = \lim_{n \to \infty} \langle f,g \circ T^{-n}\rangle = \langle f,g_0\rangle = 0$$

since $g_0 \in I$ and $f \in I^{\perp}$.

So if f is just any function in $L^2(\mu)$, write $f = f_1 + f_2$ where $f_1 \in I$ and $f_2 \in I^{\perp}$. If $f \circ T^n \to g$ in the weakstar topology then we want to show that g is W^u -invariant. Now $f^2 \circ T^n \to 0$ weakly, so we just need to check this for f_1 . Since $f_1 \in I$ it follows that $f_1 \circ T^n \in I$ and so $g \in I$.

Note that we have proven something stronger. If f is Lipschitz, then we do not need to assume that μ is a probability measure. However, the Hopf argument for ergodicity requires that f be an arbitrary L^2 function, since we must consider simple functions (which are far from Lipschitz) so this technique for proving that T is ergodic does not work in infinite measure.

Theorem 4.53 (Hopf argument). Let (X, d, μ) be a metric Radon probability space and $f \in L^2(\mu)$. If $f = f \circ T$ then f is W^s -invariant.

Proof. Every weak accumulation point of $\{f \circ T^n\}_n$ is just f, which is hence W^s -invariant. \Box

Now suppose T is invertible and $f = f \circ T$. Then f is W^s -invariant and also W^u -invariant. Modulo measure zero, we expect that f is simply constant, since we can propagate any value of f along the stable and unstable manifolds of T. Since f was arbitrary it should follow that T is ergodic. However, this argument does not work in general; if X is a Riemannian manifold, for example, we need to show that the stable and unstable manifolds form a coordinate system for X, and that Lebesgue measure disintegrates into a tensor product of measures on the stable and unstable manifolds.

As an example we prove the following theorem again.

Definition 4.54. A hyperbolic matrix is a matrix M such that Spec M does not meet the unit circle.

Theorem 4.55. Let $M \in SL(n, \mathbb{Z})$ be a hyperbolic matrix and T the action of M on \mathbb{T}^n . Then T is mixing.

We already proved this already using Fourier analysis. But we can also prove this using the Hopf argument, a version of the Jordan canonical form, and a Fubini disintegration lemma.

Lemma 4.56. For every $M \in \mathbb{C}^{n \times n}$ there are projections $P_j : \mathbb{C}^n \to \mathbb{C}^n$, eigenvalues $\lambda_j \in \mathbb{C}$, and nilpotent matrices N_j , such that $P_i P_j = \delta_i^j P_j$, $M P_j = P_j M$, $\sum_i P_j = 1$, and

$$M = \sum_{j=1}^{J} (\lambda_j + N_j) P_j.$$

To prove this form of the Jordan form we note that we can write the resolvent as

$$(M - \lambda)^{-1} = \frac{M(\lambda)}{\det(M - \lambda)},$$

so $(M - \lambda^{-1})$ is a rational function which has poles at zeroes of det $(M - \lambda)$. This uses the fact that λ is a zero of det $(M - \lambda)$ iff λ is an eigenvalue. Each pole is surrounded by a small disc D_j containing no other poles, and we let

$$P_j = \frac{1}{2\pi i} \int_{\partial D_j} (\lambda - M)^{-1} \, d\lambda.$$

This decomposes M into generalized eigenspaces using residue calculus.

As a consequence of this Jordan form, we note that if n_j is the least power such that $N_i^{n_j=0}$ and $L > n_j$ for all j, then

$$M^{L} = \sum_{j=1}^{J} \lambda_{j}^{L} + L\lambda_{j}^{L_{1}}N_{j} + \dots + \binom{L}{n_{j}-1}\lambda_{j}^{L-n_{j}+1}N_{j}^{n_{j}-1}P_{j}$$

so, if M is a hyperbolic matrix, we let

$$P_s = \sum_{|\lambda_j| < 1} P_j$$

and $P_u = 1 - P_s$, and then $P_u M^{-L} \to 0$ and $P_s M^L \to 0$ as $L \to \infty$.

If M is real, then every eigenvalue of M is either real or has a complex conjugate which is also an eigenvalue. In particular, $P_u \mathbb{C}^n$ and $P_s \mathbb{C}^n$ admit real bases, and so restrict to subspaces of \mathbb{R}^n , say E_u and E_s . **Lemma 4.57.** Let (W, λ) and (Y, ν) be probability spaces. If $g \in L^2(\lambda \otimes \nu)$ and there is a $Z \subset W \times Y$ such that $\lambda \otimes \nu(Z) = 1$, and there are measurable functions $\varphi_W : W \to \mathbb{R}$ and $\varphi_Y : Y \to \mathbb{R}$ such that for every $(w, y) \in Z$, $g(w, y) = \varphi_1(w)$ and $g(w, y) = \varphi_2(y)$, then g is constant a.e.

Proof. By Fubini's theorem,

$$1 = (\lambda \otimes \nu)(Z) = \int_{W} \int_{Y} 1_{Z}(w, y) \, d\nu(y) \, d\lambda(w).$$

Therefore there is a $w_0 \in W$ and $Y_0 \subseteq Y$ such that $\nu(Y_0) = 1$ and $\{w_0\} \times Y_0 \subseteq Z$. Suppose $(w, y) \in Z \cap (W \times Y_0)$, which is true for almost every $(w, y) \in W \times Y$. Since $(w_0, y) \in Z$ it follows that

$$\varphi_1(w_0) = g(w_0, y) = \varphi_2(y) = g(w, y).$$

Proof that T is mixing. Let E_u be the span of eigenvectors whose eigenvalues λ satisfy $|\lambda| > 1$ and E_s the span of eigenvectors whose eigenvalues are $|\lambda| < 1$. This is possible as a consequence of our discussion of the Jordan form of M, and in particular $E_u \oplus E_s = \mathbb{R}^n$.

We will prove that $W^u(x)$ is the projection of $x + E_u$ into the torus and similarly for E_s . In fact, $E_{s,u}$ is *M*-invariant; E_s is contracted by *M* and E_u is expanded by *M*.

If U is a small open neighborhood of $x_0 \in \mathbb{T}^n$, and using the manifold structure to view U as a subset of \mathbb{R}^n , then locally it is true that $W_u \cong \mathbb{R}^{n_u}$ and $W_s \cong \mathbb{R}^{n_s}$, where $n_u + n_s = n$. Thus for $x \in U$ we have a decomposition x = t + s, $t \in \mathbb{R}^{n_u}$ and $s \in \mathbb{R}^{n_s}$. Then $W^s(t_0, s_0) \cap U \supseteq \{(t_0, s) : s \in \mathbb{R}^{n_s}\} \cap U$, and similarly for W^u .

We now use the Hopf argument. Suppose g is a weakstar limit of a subsequence of $f \circ T^n$, for some $f \in L^2(\mathbb{T}^n)$. Then g is W^s -invariant and W^u -invariant. After normalization we see that U can be viewed as a subset of $\mathbb{R}^{n_u} \times \mathbb{R}^{n_s}$, but then the Fubini disintegration lemma implies that g is constant a.e. in U.

So g is locally constant a.e., hence constant a.e. In fact, $g = \int_{\mathbb{T}^n} f$. So if $h \in L^2(\mathbb{T}^n)$,

$$\lim_{k \to \infty} \int_{\mathbb{T}^n} (f \circ T^n) h = \iint_{\mathbb{T}^n \times \mathbb{T}^n} f h = \int_{\mathbb{T}^n} f \int_{\mathbb{T}^n} h.$$

We are also interested in mixing time, i.e. how many times we need to mix before two random variables f, h become "within ε of being independent." If we want this to happen at an exponential rate, we will need to assume $f, h \in C^{\infty}(\mathbb{T}^n)$, or at least some amount of Sobolev or Hoelder regularity. In particular, indicator functions are rarely exponentially mixing.

Chapter 5

Flows on manifolds

Definition 5.1. A *flow* on a metric space X is a continuous action of \mathbb{R} on X.

By a continuous action we mean that the map $\mathbb{R} \to C(X \to X)$ is continuous.

Example 5.2. Let $\dot{x} = Mx$ by an ODE on \mathbb{R}^n , $M \in \mathbb{R}^{n \times n}$. The solution of this flow is of course

$$\varphi_t(x) = e^{tM} x.$$

We say that this flow is hyperbolic if Spec M does not meet the imaginary axis, i.e. e^M is a hyperbolic matrix.

This readily generalizes to when $f : \mathbb{R}^n \to \mathbb{R}^n$ is a bounded Lipschitz function, in which case the Picard-Lindelof theorem guarantees that $\dot{x} = f(x)$ has a solution.

Now let μ be a Borel measure on X. Then μ is φ -invariant if for all A,

$$\mu(\varphi_t(A)) = \mu(A)$$

like usual. This is of course equivalent to the assumption that for every $f \in L^2(\mu)$,

$$\int_X f \circ \varphi_t \ d\mu = \int_X f \ d\mu.$$

Example 5.3. Let $\dot{x} = Mx$ like usual and suppose M is traceless. Then the action of M preserves Lebesgue measure. Similarly, $\dot{x} = f(x)$ is measure-preserving iff $\nabla \dot{f} = 0$.

5.1 Ergodic theorems for flows

We now generalize the ergodic theorem to actions of \mathbb{R} .

Theorem 5.4. Let (X, μ) be a probability space, φ an action of \mathbb{R} on X, and μ is φ -invariant. Then if $f \in L^1(\mu)$,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\varphi_t(x)) \, dt = E(f|J)(x)$$

where J is the σ -algebra generated by the φ -invariant measurable sets. In fact, if $f \in L^2(\mu)$, then

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\varphi_t(x)) \, dt = Pf(x)$$

where $P: L^2(\mu) \to \operatorname{Inv} T$ is the orthogonal projection.

Proof. By reduction to the discrete ergodic theorems we just need to show that the limit of those integrals exists. Let $\Omega = \mathbb{R}^{\mathbb{Z}}$, and define $\Gamma : X \to \Omega$ by

$$\Gamma(x)_j = \int_j^{j+1} f(\varphi_t(x)) \, dt.$$

Now let $T: \Omega \to \Omega$ be the shift $(T\omega)_j = \omega_{j+1}$. Then $\Gamma \circ \varphi_1 = T \circ \Gamma$.

Define $\tilde{\mu}(A) = \mu(\gamma^{-1}(A))$. This is defined for some σ -algebra in Ω , and is a measure since $\tilde{\mu} = \gamma_* \mu$. Moreover $(\Omega, \tilde{\mu}, T)$ is a measure-preserving system:

$$\tilde{\mu}(T^{-1}A) = \mu(\Gamma^{-1} \circ T^{-1}A) = \mu(\varphi_1^{-1} \circ \Gamma^{-1}A) = \mu(\Gamma^{-1}A) = \tilde{\mu}(A).$$

Now apply Birkhoff's ergodic theorem to the function

$$F(\omega) = \omega_0.$$

Then

$$\frac{1}{n}\sum_{j=0}^{n-1}F(T^{j}(\omega)) = \frac{1}{n}\sum_{j=0}^{n-1}\omega_{j}.$$

 So

$$\tilde{\mu}\{\omega : \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \omega_j \text{ exists}\} = 1.$$

But on the other hand this is

$$\mu\{x: \lim_{n \to \infty} \int_0^n f \circ \varphi_t(x) \, dt \text{ exists}\} = 1.$$

Up to a small error term we may replace a large n in this limit by any sufficiently large positive real number.

5.2 Geodesic flows in hyperbolic space

Let \mathcal{H}^2 be the upper-half plane, viewed as a Riemann surface. We define a Hermitian metric on \mathcal{H}^2 . This means that for every tangent vector $a + ib \in T_z \mathcal{H}^2$ we assign a length

$$|a+ib|_z = \frac{\sqrt{a^2+b^2}}{\operatorname{Im} z}.$$

If we were to translate the vector a+ib along a curve γ then a+ib becomes longer as Im $\gamma(s)$ decreases. In fact the distance element is given by

$$ds^2 = \frac{dx^2 + dy^2}{y^2};$$

i.e. the length of a curve γ is given by

$$\ell(\gamma) = \int_{\gamma} ds = \int_{a}^{b} |z'(s)|_{z(s)} ds = \int_{a}^{b} \frac{|z'(s)|}{\operatorname{Im} z(s)} ds.$$

We also have an area form

$$dA = \frac{dx \ dy}{y^2}$$

or in other words, the area of a set Ω is given by

$$A(\Omega) = \iint_{\Omega} \frac{dx \, dy}{y^2}.$$

Now consider the group

$$\operatorname{PSL}(2,\mathbb{R}) = \frac{\operatorname{SL}(2,\mathbb{R})}{\pm 1}.$$

Then $\mathrm{PSL}(2,\mathbb{R})$ acts on $\mathbb{P}^1_{\mathbb{C}}$ by linear fractional transformations; in fact

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} z = \frac{az+b}{cz+d}$$

and it is easy to check that this action preserves \mathbb{R} and orientation, so restricts to an action on \mathcal{H}^2 ; namely,

$$\operatorname{Im} \frac{az+b}{cz+d} = \frac{\operatorname{Im} z}{|cz+d|^2}$$

The action of $PSL(2, \mathbb{R})$ is transitive; i.e. the orbit of every element is \mathcal{H}^2 . It suffices to check this on a single point, say *i*; clearly any element in the upper-half plane can be written in terms of *i*.

Note that the stabilizer of i is the set of matrices such that

$$\frac{ai+b}{ci+d} = i$$

i.e. those matrices such that a = d, c = -b, $a^2 + b^2 = 1$, or in other words the rotation matrices. Thus the stabilizer of *i* is isomorphic to $\mathbb{P}^1_{\mathbb{R}}$, and so the action of $\mathrm{PSL}(2,\mathbb{R})$ on \mathcal{H}^2 is not free.

Since the action of $PSL(2, \mathbb{R})$ has no poles in \mathcal{H}^2 and it has nonzero derivatives, $PSL(2, \mathbb{R})$ is a subgroup of the group of conformal automorphisms of \mathcal{H}^2 . Thus $PSL(2, \mathbb{R})$ also acts on the tangent bundle $T\mathcal{H}^2$; namely, if $f \in PSL(2, \mathbb{R})$, $(z, \xi) \in T\mathcal{H}^2$, then

$$f(z,\xi) = (f(z), f'(z)\xi).$$

Moreover, if $f = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $f'(z) = (cz+d)^2$.

Lemma 5.5. The action of $PSL(2, \mathbb{R})$ preserves the lengths of tangent vectors in $T\mathcal{H}^2$. *Proof.* Let $\xi \in T_z \mathcal{H}^2$; we must show

$$|\xi|_z = \left|\frac{\xi}{(cz+d)^2}\right|_{f(z)}$$

where $f = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Now $|\xi|_z \operatorname{Im} z = |\xi|$ so

$$\frac{\xi}{(cz+d)^2}\Big|_{f(z)} = \frac{\left|\frac{\xi}{(cz+d)^2}\right|}{\operatorname{Im}\frac{az+b}{cz+d}}$$
$$= \frac{\frac{|\xi|}{|cz+d|^2}}{\frac{|mz}{|cz+d|^2}} = |\xi|_z.$$

Thus $PSL(2, \mathbb{R})$ acts on \mathcal{H}^2 by isometries. In particular, $PSL(2, \mathbb{R})$ preserves the area element dA.

Thus to study $PSL(2, \mathbb{R})$ we do not actually need to consider the entire tangent bundle $T\mathcal{H}^2$ but rather the unit sphere bundle

$$S\mathcal{H}^2 = \{(z,\xi) \in T\mathcal{H}^2 : |\xi|_z = 1\}.$$

Thus $PSL(2,\mathbb{R})$ preserves $S\mathcal{H}^2$. Having carried out all this setup, we are now interested in dynamical systems with phase space $S\mathcal{H}^2$. This is a 3-dimensional real manifold with parametrization

$$(x, y, \theta) \mapsto (x + iy, ye^{i\theta})$$

and so is diffeomorphic to $\mathbb{R} \times (0, \infty) \times \mathbb{T}^1$. Moreover $S\mathcal{H}^2$ has an invariant metric

$$\frac{dx^2 + dy^2}{y^2} + d\theta^2.$$

In fact we have

$$d((z, v), (z', v')) = d(g(z, v), g(z', v'))$$

where $g \in PSL(2, \mathbb{R})$. So it has a volume form

$$dV = \frac{dx \ dy \ d\theta}{y^2}.$$

The volume form dV is left-invariant and right-invariant for the action of $PSL(2, \mathbb{R})$. Therefore dV is the Haar measure on the unimodular group $PSL(2, \mathbb{R})$.

Lemma 5.6. The map

$$\Phi : \mathrm{PSL}(2, \mathbb{R}) \to S\mathcal{H}^2$$
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \left(\frac{ai+b}{ci+d}, \frac{i}{(ci+d)^2}\right)$$

is a diffeomorphism.

Proof. It suffices to show that this map is a bijection. In fact, the stabilizer of i is sent to

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \mapsto (i, e^{2i\theta}i).$$

So the first two variables of the parametrization of $S\mathcal{H}^2$ are controlled by $PSL(2,\mathbb{R})$ modulo the stabilizer and the final variable is controlled by the stabilizer.

We now introduce a very important flow on $S\mathcal{H}^2 = PSL(2,\mathbb{R})$. To do this, we note that we want to define a distance function \mathcal{H}^2 by

$$d(z,w) = \inf_{\gamma} \frac{|\gamma'(s)|}{\operatorname{Im} \gamma(s)} \, ds$$

where the inf is taken over all curves γ from z to w. Now we can find a group element $g \in \text{PSL}(2, \mathbb{R})$ and a a > 0 such that z = g(i) and w = g(ai). Thus

$$d(z,w) = d(i,ai)$$

and so we are only interested in curves

$$\gamma(s) = x(s) + iy(s)$$

such that x(0) = x(1) = 0, y(0) = 1 and y(0) = a. We have

$$\int_0^1 \frac{\sqrt{x'(s)^2 + y'(s)^2}}{y(s)} \, ds \ge \int_0^1 \frac{|y'(s)|}{y(s)} \, ds \ge \left| \int_0^1 \frac{y'(s)}{y(s)} \, ds \right|$$
$$= \left| (\log y(s))_{s=0}^1 \right| = \left| \log a \right|$$

and so

$$d(z, w) = |\log a|$$

and the curve which witnesses this is the image of the curve which witnesses that $d(i, ai) = |\log a|$. In fact this curve γ is a curve on the imaginary axis, which is sent by $PSL(2, \mathbb{R})$ to a circle on which $g \circ \gamma$ is an arc. Moreover, i was the tangent vector to γ at $\gamma(0) = i$, so $i/(ci+d)^2$ is the tangent vector to $g \circ \gamma$ at $g \circ \gamma(0) = g(i)$.

We now introduce the *geodesic flow* which carries (z, ξ) along its respective great circle. Namely,

$$\varphi_t(i,i) = (e^t i, e^t i)$$

and if $g(i,i) = \Phi(g)$ is an element of $S\mathcal{H}^2$ then it we define

$$\varphi_t(g(i,i)) = g(\varphi_t(i,i)) = g(e^t i, e^t i).$$

In other words, if $G_t = \begin{bmatrix} e^{t/2} & 0\\ 0 & e^{-t/2} \end{bmatrix}$ then we have

$$\varphi_t(\Phi(g)) = \Phi(gG_t)$$

Now $G_t \in \text{PSL}(2, \mathbb{R})$ so dV is left invariant by the geodesic flow. So in fact φ is a measurepreserving action of \mathbb{R} on $(S\mathcal{H}^2, V)$.

Note that there is a standard linear fractional transformation $\mathcal{H}^2 \to \mathbb{D}$ given by

$$\zeta\mapsto \frac{\zeta-i}{\zeta+i}.$$

This gives all of the above structure to \mathbb{D} and hence $S\mathbb{D}$. In this space a geodesic is not a great circle, but a circle which is perpendicular to the unit circle. So in fact an annulus in \mathbb{D} centered at the origin is preserved by φ so φ is not an ergodic flow.

We are interested in the stable and unstable manifolds of φ . We have

$$W^{s}(z,\xi) = \{(w,\zeta) \in S\mathcal{H}^{2} : \lim_{t \to \infty} d(\varphi_{t}(w,\zeta),\varphi_{t}(z,\zeta)) = 0\}$$

and similarly for W^{u} . Using the action of $PSL(2,\mathbb{R})$ it suffices to compute $W^{s}(i,i)$, and

$$W^{s}(i,i) = \{(i+t,i) : t \in \mathbb{R}\}.$$

This consists of the horizontal line through i, which is being translated upwards. As the imaginary part increases the distances will vanish. It is reasonable to view $W^{s}(i,i)$ as the action of the 1-parameter group $h_{t}^{s}: t \in \mathbb{R}$ on (i,i) where

$$h_t^s = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

Similarly

$$W^{u}(i,i) = \{-(i+t)^{-1}, i(i+t)^{-2} : t \in \mathbb{R}\}$$

which is the action of the 1-parameter group $h_t^u : t \in \mathbb{R}$ on (i, i) where

$$h_t^u = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix}.$$

The actions of h_t^s and h_t^u are called the *horocycle flows* determined by the geodesic flow on $S\mathcal{H}^2$, and we have

$$\varphi_t \circ h_s^s = h_{se^{-t}}^s \circ \varphi_t$$

and similarly

$$\varphi_t \circ h^u_s = h^u_{se^t} \circ \varphi_t.$$

Henceforth we change notation to W^+ for the stable manifold and W^- for the unstable, and identify Φ with the identity map. We write H for the horocycle flow and G for the geodesic flow Thus

$$W^{\pm}(g) = \{g' \in \mathrm{PSL}(2,\mathbb{R}) : \lim_{t \to \pm \infty} d(gG_t, g'G_t) = 0\}.$$

Then H_s^+ is the matrix $\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$ and H_s^- is its inverse $\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}$, and $G_t = \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix}$.

Unfortunately $S\mathcal{H}^2 = \mathrm{PSL}(2,\mathbb{R})$ is not a compact Lie group; in fact, it has infinite volume. So we cannot hope to use the ergodicity theory to study the horocycle flow. Therefore we fix a discrete subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$. For example, the modular group $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ is sufficient. Famously, $\mathrm{PSL}(2,\mathbb{Z})$ is generated by $\begin{bmatrix} 1\\-1 \end{bmatrix}, \begin{bmatrix} 1&1\\1 \end{bmatrix}$. Now Γ may not be a normal subgroup so the set of left cosets $\Gamma \setminus \mathcal{H}^2 = \{\Gamma z : z \in \mathcal{H}^2\}$ is not a group. But it does have a Riemannian manifold structure and so it makes sense to talk about the unit tangent bundle

$$S(\Gamma \setminus \mathcal{H}^2) = \{ \Gamma g : g \in PSL(2, \mathbb{R}) \}.$$

In fact the standard fundamental domain M of the modular surface $\Gamma \setminus \mathcal{H}^2$ is contained in the complex projective line \mathbb{P}^1 . It is \mathcal{H}^2 modulo the action of $z \mapsto z + 1$ and $z \mapsto$ -1/z. The modular surface is a Riemann surface, and the volume of the unit sphere bundle $\Gamma \setminus \mathrm{PSL}(2,\mathbb{R}) = S(\Gamma \setminus \mathcal{H}^2)$ is

$$V(\Gamma \setminus \mathrm{PSL}(2,\mathbb{R})) = \iint_{M \times S^1} \frac{dx \, dy \, d\theta}{y^2} \le \int_0^{2\pi} \int_{-1/2}^{1/2} \int_{1/2}^\infty \frac{dy}{y^2} \, dx \, d\theta < \infty.$$

Renormalizing we assume $V(\Gamma \setminus PSL(2, \mathbb{R})) = 1$ as necessary. Then the geodesic flow and the horocycle flows drop to an action on $\Gamma \setminus PSL(2, \mathbb{R})$.

Having discussed the motivating example of $PSL(2, \mathbb{R})$ we consider discrete subgroups Γ more generally. Notice that Γ is discrete iff there is a $g \in PSL(2, \mathbb{R})$ such that

$$d(g, \Gamma g \setminus \{g\}) > 0.$$

As another example, suppose that we have an 8-gon in the Poincare disk model of hyperbolic space (so \mathcal{H}^2 is mapped into the unit disk by the Cayley transform). By identifying sides we end up with a compact genus-2 Riemann surface. In particular it is a probability space. We are interested in when the 8-gon is the fundamental domain of Γ .

So suppose that Γ is a discrete subgroup of $PSL(2, \mathbb{R})$ such that

$$S(\Gamma \setminus \mathcal{H}^2) = \Gamma \setminus \mathrm{PSL}(2,\mathbb{R})$$

is a probability space. Then

$$\varphi_t(\Gamma g) = \Gamma(gG_t)$$

is the action of the geodesic flow on $S(\Gamma \setminus \mathcal{H}^2)$. Similarly we have horocycle flows

$$h_s^{\pm}(\Gamma g) = \Gamma(gH_s^{\pm}).$$

These parametrize the stable and unstable manifolds as

$$W^{\pm}(\rho) = \{h_s^{\pm}(\rho) : s \in \mathbb{R}\}.$$

We have the commutation relation

$$\varphi_t \circ h_s^{\pm} = h_{se^{\mp t}}^{\pm} \circ \varphi_t.$$

We now prove ergodicity of the geodesic flow on $\Gamma \setminus \text{PSL}(2, \mathbb{R})$ using the Hopf argument. We consider the space of weakstar limits of $\{f \circ \varphi_t : t \in \mathbb{R}\}$ whenever $f \in L^2(\Gamma \setminus \text{PSL}(2, \mathbb{R}))$. If $f \circ \varphi_t = f$, we will show that f is invariant under the horocycle flows. Therefore f will be constant. In fact, we will then show that the horocycle flows are ergodic, which implies that

$$\lim_{t \to \pm \infty} \langle f \circ \varphi_t, g \rangle = E(f)E(g)$$

so the geodesic flow is mixing. Let ω be the invariant volume form

$$\omega = \frac{dx \ dy \ d\theta}{y^2}.$$

Example 5.7. In zero curvature all this theory is kind of trivial. Let \mathbb{T}^2 be the flat torus and let φ be the geodesic flow on \mathbb{T}^2 , so

$$\varphi_t(x,v) = (x+tv,v).$$

Then $W^{\pm}(x, v) = \{(x, v)\}$. This corresponds to \mathbb{T}^2 being flat, so that there is "no gravity", i.e. no two points are attracted to each other.

Theorem 5.8. Suppose that $\Gamma \setminus PSL(2, \mathbb{R})$ has finite volume. Then the geodesic flow on $\Gamma \setminus PSL(2, \mathbb{R})$ is ergodic.

Proof. We have

$$h_{u'}^- \circ h_s^+ \circ \varphi_t = h_{s'}^+ \circ \varphi_{t'} \circ h_u^-$$

where $u' = u(1 - e^t s u)^{-1}$, $s' = s(1 - e^t s u)$, and $t' = t - 2\log(1 - e^t s u)$. This is true whenver u, s, t are close to 0.

Fix $g \in \Gamma \setminus PSL(2, \mathbb{R})$ and define local coordinates

$$\Psi(s, t, u) = h_{u'}^- \circ h_s^+ \circ \varphi_t(g).$$

Then Ψ is a diffeomorphism close to 0. Let $U \ni g$ be open. If $\tilde{g} \in U$ then

$$U \cap W^{-}(\tilde{g}) = \{\Psi(s, t, \xi)\}$$

where $\xi \in \mathbb{R}$ is sufficiently close to the $u \in \mathbb{R}$ such that $\Psi(s, t, u) = \tilde{g}$. In fact,

$$W^{-}(\tilde{g}) = \{h_{\eta}^{-}(\tilde{g}) : \eta \in \mathbb{R}\} = \{h_{u'+\eta}^{-} \circ h_{s}^{+} \circ \varphi_{t}(g) : \eta \in \mathbb{R}\},\$$

at least when s, t, u, η are close to 0.

We now prove that

$$U \cap \bigcup_{T} W^{+}(\varphi_{T}(\tilde{g})) = \{\Psi(\sigma, \tau, u)\}$$

where (σ, τ) is sufficiently close to (s, t). In fact, this is the set of all

$$h_{\sigma'}^+ \circ \varphi_{t'} \circ h_{u'}^- \circ h_s^+ \circ \varphi_t(g)$$

or in other words the set of

$$h_{u''}^{-} \circ h_{\sigma}^{+} \circ h_{\sigma}^{+} \circ \varphi_{t} a u \circ h_{s}^{+} \circ \varphi_{t}(g)$$

or in other words

$$h_{u''}^- \circ h_{\sigma+se^{-t}}^- \circ \varphi_{\tau+t}(g)$$

or in other words

$$\Psi(\sigma + se^{-\tau}, t + \tau, u)$$

or in other words

$$u(1 - e^{t+\tau}(e^{-\tau}s + \sigma)u)^{-1}.$$

This proves the claim.

Therefore we have written the space as a product of a two-dimensional space $\{(s, t)\}$ and one-dimensional space $\{u\}$, and can easily split ω into forms on each of these two space, and can therefore use the Hopf argument.

Chapter 6

Topological dynamics

6.1 Recurrence in topological dynamics

Definition 6.1. Let X be a topological space, $T: X \to X$. We say that T is topologically transitive if for every pair of nonempty open sets $U, V \subseteq X$, there is a sequence of $n_i \to \infty$ such that $T^{-n_i}U \cap V$ is nonempty.

In other words, there is a $x \in U$ such that $T^n x \in V$ for infinitely many n.

Definition 6.2. Let X be a topological space, $T : X \to X$. We say that T is topologically mixing if for every pair of nonempty open sets $U, V \subseteq X$ and every n large enough, $T^{-n}U \cap V$ is nonempty.

Example 6.3. Consider rotations of \mathbb{T}^1 by α . If α is rational then T is not transitive. But if α is irrational then T is transitive. But T is not mixing because if U, V are tiny intervals, then they cannot be trapped in each other since T is an isometry.

We now show that transitive maps are analogously to ergodic maps.

Theorem 6.4. If T is a Borel map and preserves an ergodic probability measure μ with full support, then T is topologically transitive. In fact, if (X, T, μ) is mixing, then T is topologically mixing.

Proof. Let U, V be given. For almost every $x \in U$ the fraction of the time that the forward orbit of x lands in V is $\mu(V)$. Since μ has full support it follows that $\mu(V) > 0$. Therefore T is transitive.

If (X, T, μ) is mixing, then $\lim_n \mu(T^{-n}U \cap V) \to \mu(U)\mu(V) > 0$, so for large enough n, $\mu(T^{-n}U \cap V) > 0$.

Definition 6.5. For any $x \in X$, let $\omega(x)$ be the set of accumulation points of $T^n x$ as $n \to \infty$.

Thus $\omega(x)$ is closed, and is the set of all subsequential limits of $T^n x$ as $n \to \infty$.

Theorem 6.6. Let (X, μ, T) be an ergodic system, X a metric space, and μ a probability measure. For almost every $x \in X$, supp $\mu \subseteq \omega(x)$.
Proof. Note that $\operatorname{supp} \mu$ is separable, since μ is a Borel probability measure. Let $\{x_i\}_i$ be a dense subset of $\operatorname{supp} \mu$ and let $r \in \mathbb{Q}$. By ergodicity there is a set $\Omega_{i,r}$ such that $\mu(\Omega_{i,r}) = 1$ and for every $x \in \Omega_{i,r}$, there are infinitely many n such that $T^n x \in B(x_i, r)$. If $x \in \bigcap_{i,r} \Omega_{i,r}$, which is a set of full measure, then for every i, r there are infinitely many n such that $T^n x \in B(x_i, r)$. If such that $T^n x \in B(x_i, r)$. Then $x_i \in \omega(x)$, so $\{x_i : i \in \mathbb{N}\} \subseteq \omega(x)$. Since $\omega(x)$ is closed, $\operatorname{supp} \mu \subseteq \omega(x)$.

Corollary 6.7. Let (X, μ, T) be an ergodic system, X a metric space, μ a probability measure, and suppose supp $\mu = X$. Then for almost every $x \in X$, supp $\mu \subseteq \omega(x)$.

Definition 6.8. A topological space X is topologically complete if X admits a complete metric. If X additionally is separable, then we say that X is a Polish space.

Every Polish space admits a countable basis. The Baire category theorem also holds in Polish spaces (even topologically complete spaces); namely, a countable intersection of open dense sets is still dense.

Theorem 6.9. If X is a Polish space and $T: X \to X$ is continuous, then the following are equivalent:

- 1. T is transitive.
- 2. The set of $x \in X$ such that $\omega(x) = X$ is dense in X.
- 3. There is an $x \in X$ such that $\omega(x) = X$.

Proof. Assume that $\omega(x) = x$ and let U, V be given. Then the forward orbit of x hits U, V infinitely many times. Suppose $T^n x \in U$; then $T^n x$ meets V infinitely many times, so T is transitive.

If T is transitive, let $Y = \{x \in X : \omega(x) = X\}$. Then

$$Y = \bigcap_{U} \bigcap_{N \ge 0} \bigcup_{n \ge N} T^{-n} U.$$

Here the first intersection is taken over nonempty open sets. Since X is Polish we can replace the first intersection with an intersection over a countable basis for the topology of X. Since T is transitive, $\bigcup_n T^{-n}U$ is dense, and is open since T is continuous. So Y is a countable intersection of open dense sets, so is dense by the Baire category theorem.

Definition 6.10. A point $x \in X$ is recurrent if $x \in \omega(x)$.

If X is a metric space then the set of recurrent points is equal to

$$\bigcap_{k>0}\bigcap_{N\geq 0}\bigcup_{n\geq N}\{x\in X: d(x,T^nx)<\frac{1}{k}\}.$$

Theorem 6.11. Let (X, μ, T) be a measure-preserving system, X a metrizable space, and μ a probability measure. Then almost every $x \in \text{supp } \mu$ is recurrent.

Proof. Let D be a countable basis for the topology of $\operatorname{supp} \mu$. Then by Poincare recurrence, for every $U \in D$, the set U' of x such that $x \notin U$ or there are infinitely many k such that $T^k x \in U$ has full measure. Now $\bigcap_{U \in D} U'$ has full measure, and consists of recurrent points.

Definition 6.12. A point $x \in X$ is *nonwandering* if for every open $U \ni x$, there are infinitely many n such that $T^{-n}U \neq U$ is nonempty. If x is nonwandering then we write $x \in \Omega$.

If $x \in X$ then $\omega(x) \subseteq \Omega$. In particular, if x is recurrent, then x is nonwandering. The converse is not true.

Example 6.13. Let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be the cat map. Let E^s be the stable eigenline of T. If $x \in E^s$ then x approaches 0 exponentially fast. Therefore x is not recurrent.

But if v has rational coefficients, then v is recurrent, and we can approximate any point on \mathbb{T}^2 arbitrarily well by such v. Therefore T has no wandering points.

Theorem 6.14. Let X be a metric space, $T : X \to X$ continuous. Then $x \in X$ is a wandering point iff there is an open $U \ni x$ such that for all $n, T^{-n}U \cap U$ is empty.

Proof. One direction is clear. For the converse, suppose x wanders, so there is an open $U \ni x$ such that for all n large enough, $T^{-n}U \cap U$ is empty. We need to remove the clause "large enough." Let N be the minimal n. Now x is not periodic, so there is an open $V \ni x$ such that for every $i \leq N$, $T^{-i}V \cap V$ is empty and $T^NV \subseteq U$. Therefore for every $i, T^{-i}V \cap V$ is empty.

Chapter 7

Completely integrable systems

Ergodic systems are intended to be understood as "chaotic"; therefore they can often be not useful in practice. We now consider more predictable dynamical systems.

7.1 Symplectic geometry

Definition 7.1. Let S be a (finite-dimensional, real) vector space. Then a symplectic form on S is an antisymmetric, nondegenerate bilinear form. A symplectic space is a vector space equipped with a symplectic form.

Example 7.2. The canonical example of a symplectic space is \mathbb{R}^{2n} , which it will be convenient to view as $T^*\mathbb{R}^n$, the cotangent bundle of \mathbb{R}^n . Let I denote the identity matrix on \mathbb{R}^n and let $J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$. Then $\sigma(z, w) = \langle Jz, w \rangle$ is the standard symplectic form on \mathbb{R}^{2n} . In fact, if $(x_1, \ldots, x_n, \xi_1, \ldots, \xi_n)$ are coordinates on \mathbb{R}^{2n} , and $\omega = \sum_j \xi_j dx_j$, then $\sigma = d\omega$. So σ is a closed (hence exact) 2-form.

Since symplectic spaces are vector spaces, we may always view them as having the standard differential structure of \mathbb{R}^{2n} . In particular, it makes sense to ask for the linearization of a C^1 -map $\kappa : S_1 \to S_2$, say $\partial \kappa$, and then if ω is a multilinear form on S_2 , we can define the pullback κ of ω , a multilinear form on S_1 , by

$$\kappa^*\omega(v_1,...,v_n) = \omega(\partial\kappa(v_1),...,\partial\kappa(v_n)).$$

Definition 7.3. Let (S_1, σ_1) and (S_2, σ_2) be symplectic spaces. A symplectomorphism is a map $\kappa : S_1 \to S_2$ such that $\sigma_1 = \kappa^* \sigma_2$.

Example 7.4. Let $\kappa : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a linear map, and let σ, J be standard. Let $\kappa(x,\xi) = (Ax + B\xi, Cx + D\xi)$; then $\kappa^*\sigma = \sigma$ iff the matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ satisfies $M^t J M = J$. This example motivates the definition of a symplectic matrix.

Definition 7.5. Let J be standard for \mathbb{R}^{2n} ; a symplectic matrix is a matrix $M \in \mathbb{R}^{2n \times 2n}$ such that $M^t J M = J$. The symplectic group is the group of all symplectic matrices.

Example 7.6. There exist nonlinear symplectomorphisms. Let $\varphi : \mathbb{R}^{2n} \to \mathbb{R}$ be a smooth function such that the Hessian determinant of φ at (x_0, y_0) does not vanish, i.e.

$$\det \partial_x \partial_y \varphi(x_0, y_0) \neq 0.$$

For example, let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix, and consider the bilinear form φ defined by A.

Let $\xi_0 = \partial_x \varphi(x_0, y_0)$ and $\eta_0 = -\partial_y \varphi(x_0, y_0)$. Then the implicit function theorem defines a mapping close to $(x_0, \xi_0) \in T^* \mathbb{R}^n$ by

$$\kappa(x,\partial_x\varphi(x,y)) = (y,-\partial_y\varphi(x,y)).$$

To see that κ is a symplectomorphism, we note that

$$\kappa^*(d\eta \wedge dy) = d(-d_y\varphi) \wedge dy = (-\partial_y^2\varphi \, dy) \wedge dy + (-\partial_x\partial_y\varphi \, dx) \wedge dy$$
$$= -\partial_x\partial_y\varphi \, dx \wedge dy = \partial_x\partial_y\varphi \, dy \wedge dx = d\xi \wedge dx.$$

Therefore κ defines a symplectomorphism from a neighborhood of (x_0, ξ_0) to a neighborhood of (y_0, η_0) .

Theorem 7.7. Let (S, σ) be a symplectic space. Then there is an invertible linear symplectomorphism $S \to \mathbb{R}^{2n}$.

Proof. We prove this by the induction on the dimension 2n of S. Since σ is antisymmetric and nondegenerate, $2n \ge 2$. Assume 2n = 2. Then there are $e, f \in S$ such that $\sigma(e, f) = 1$, and $\{e, f\}$ is a basis of S. Define a linear map κ by $\kappa(e) = (0, 1)$ and $\kappa(f) = (1, 0)$; then κ is an invertible symplectomorphism.

Now if 2n > 2, there are e_1, f_1 , linearly independent, such that $\sigma(e_1, f_1) = 1$. Let S_1 be the span of $\{e_1, f_1\}$, and let S_0 be the sum of the kernels of $\sigma(\cdot, w)$ for $w \in S_1$. Then the codimension of S_0 is 2, and if $z = xe_1 + yf_1 \in S_0$, $0 = \sigma(z, e_1) = -y$ and $0 = \sigma(z, f_1) = x$, so z = 0. Therefore $S_1 \cap S_0 = 0$. Therefore $S_0 \oplus S_1 = S$. For every $z \in S_0$, if $\sigma(z, S_0) = 0$ then $\sigma(z, S) = 0$ so z = 0. Therefore $\sigma|_{S_0}$ is nondegenerate. So by the inductive hypothesis, we have an invertible linear symplectomorphism $\kappa_0 : S_0 \to \mathbb{R}^{2n-2}$. Let $\kappa : S \to \mathbb{R}^{2n}$ extend κ_0 by sending e_1, f_1 to $\mathbb{R}^{2n}/\mathbb{R}^{2n-2}$ as in the base case. \Box

We now consider manifolds whose tangent bundles are bundles of symplectic spaces.

Definition 7.8. Let S be a 2n-dimensional manifold equipped with a closed, nondegenerate 2-form σ . Then (S, σ) is called a *symplectic manifold*.

In other words, on every tangent space, σ defines an antisymmetric nondegenerate bilinear form which varies smoothly, and $d\sigma = 0$.

Example 7.9. Let \mathcal{H}^2 be the hyperbolic plane; then its tangent bundle $T\mathcal{H}^2$ can be turned into a symplectic manifold by setting coordinates $(x + iy, \xi + i\eta), y > 0$, and letting

$$\sigma = d\xi \wedge dx + d\eta \wedge dy.$$

In fact the tangent bundle of $T\mathcal{H}^2$ is diffeomorphic to the half-space of \mathbb{R}^4 , on which σ is the standard symplectic form.

We now consider differential forms on symplectic manifolds. For κ a symplectomorphism and X a vector field, let $\kappa_* X$ denote the pushforward of X along κ . That is, $\kappa_* X = \partial \kappa \circ X$. Recall that the pullback κ^* is defined dually to the pushfoward, i.e. if η is a 1-form and X is a vector field, we have

$$\kappa^*\eta(X) = \eta(\kappa_*X).$$

We extend this definition to *m*-forms by declaring that κ^* preserves \wedge .

Definition 7.10. If η is an m + 1-form and X is a vector field, define for vector fields Y_1, \ldots, Y_m

$$(X \lrcorner \eta)(Y_1, \ldots, Y_m) = \eta(X, Y_1, \ldots, Y_m).$$

In particular, $X \lrcorner \eta$ is a *m*-form. One can check that if ν is a *k*-form and η is an *m*-form then

$$X \lrcorner (\nu \land \eta) = (X \lrcorner \nu) \land \eta + (-1)^k \nu \land (X \lrcorner \eta).$$

None of the above used that κ was a symplectomorphism but we will mainly be interested in the case that κ is a symplectomorphism.

Example 7.11. Let $g: \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism and w is a smooth function. Then

$$\kappa(x,\xi) = (g(x), (\partial g(x)^t)^{-1}(\xi + \nabla w(x)))$$

is a symplectomorphism of \mathbb{R}^{2n} and these are the only symplectomorphisms which lift from g.

7.2 Hamiltonian flows on symplectic spaces

We treat Hamiltonian flows first on symplectic spaces, revisiting symplectic manifolds later. So fix a symplectic space (S, σ) and coordinates (x, ξ) so that $\sigma = d\xi \wedge dx$.

Definition 7.12. For any function $f \in C^{\infty}(S)$, we let H_f , the Hamiltonian vector field of f, be defined by the relation $\sigma(\cdot, H_f) = df$.

In other words,

$$H_f = \sum_j \frac{\partial f}{\partial \xi_j} \partial x_j - \frac{\partial f}{\partial x_j} \partial \xi_j$$

where $\{\partial x_i, \partial \xi_i\}$ define a basis for the tangent space.

Any vector field X determines a flow by the ODE $\cdot x = X(x)$. This can also be expressed by writing

$$x(t) = \exp(tX)(x(0)).$$

Thus $t \mapsto \exp(tX)$ is a one-parameter group and a diffeomorphism. So in particular we have defined a *Hamiltonian flow*.

Example 7.13. The notion of a symplectic manifold, and of a Hamiltonian flow, is motivated by classical mechanics. View x as the position and ξ as the momentum of a particle; let V(x) be its potential energy and m its mass. Then the energy of the particle is given by

$$f(x,\xi) = \frac{|\xi|^2}{2m} + V(x).$$

The Hamiltonian vector field is given by

$$H_f = \frac{1}{m} \langle \xi, \partial_x \rangle - \langle \nabla V(x), \partial_\xi \rangle$$

Therefore the particle moves with velocity ξ/m , and its change in momentum is given by $-\nabla V(x)$. In particular, its acceleration A satisfies Newton's second law of motion,

$$mA + \nabla V(x) = 0.$$

To continue further we will need to introduce a Lie algebra structure on the space $C^{\infty}(S)$.

Definition 7.14. The *Poisson bracket* is defined by

$$\{f,g\} = \sigma(\nabla f, \nabla g)$$

Thus the Poisson bracket also satisfies $\{f, g\} = H_f g$ and

$$\{f,g\} = \sum_{j} \frac{\partial f}{\partial \xi_j} \frac{\partial g}{\partial x_j} - \frac{\partial f}{\partial x_j} \frac{\partial g}{\partial \xi_j}.$$

Since vector fields act on functions as linear maps, it makes sense to talk about the commutator of vector fields, and then we have

$$H_{\{f,g\}} = [H_f, H_g].$$

One can check also that we have the Jacobi identity $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$, so $\{\cdot, \cdot\}$ is a Lie bracket. Moreover, the flow

$$\varphi_t = \exp(tH_f)$$

is a symplectomorphism.

While we defined the Poisson bracket on functions, it will be useful to talk about a Lie algebra structure on any symplectic space (S, σ) , so given vectors $v, w \in S$, let

$$\{v,w\} = \sigma(v,w).$$

We first prove the linear case of Darboux's theorem. To motivate this theorem, note that if $\{e_1, \ldots, e_n, f_1, \ldots, f_n\}$ is a basis of (S, σ) , it is useful that $\{e_i, e_j\} = 0$, $\{f_i, f_j\} = 0$, and $\{e_i, f_j\} = \delta_{ij}$. This is a similar condition to requiring that a basis of a Hilbert space is orthonormal. We say that such a basis is a *symplectic basis*. Darboux's theorem says that a symplectic linearly independent set can always be extended to a symplectic basis. **Theorem 7.15** (linear Darboux theorem). Let S be a symplectic space of dimension 2n and $A, B \subset \{1, \ldots, n\}$. Let $\{e_j\}_{j \in A}$ and $\{f_i\}_{i \in B}$ be linearly independent sets and suppose that $\{e_i, e_j\} = 0, \{f_i, f_j\} = 0, \text{ and } \{e_i, f_j\} = \delta_{ij}$. Then there are $\{e_j\}_{j \notin A}$ and $\{f_i\}_{i \notin B}$ such that $\{e_i, e_j\} = 0, \{f_i, f_j\} = 0, \text{ and } \{e_i, f_j\} = \delta_{ij}$.

Proof. We first show that we may assume that A = B. In fact, suppose that the symmetric difference of these two sets is nonempty, say $J \in B \setminus A$. Then there is an e_J such that for every $j \in A$, $\{e_J, e_j\} = 0$ and every $k \in B$, $\{e_J, f_k\} = \delta_{Jk}$. Here we used nondegeneracy of σ and the linear independence hypothesis. One can check that e_J is linearly independent of e_j and f_k , so we can add J to A without contradicting the hypothesis.

Suppose that $A = B \neq \{1, ..., n\}$, and let S' be the span of $\{e_j\}_{j \in A}$ and $\{f_j\}_{j \in A}$. Let S_0 be the symplectic complement of S', i.e. the set of $z \in S$ such that for every vector $w \in S'$ we have $\sigma(z, w) = 0$ Then (S_0, σ) is a symplectic space. Choose a symplectic basis for S_0 ; then $S = S_0 \oplus S'$ and we're done.

7.3 Hamiltonian flows on symplectic manifolds

We now revisit Hamiltonian flows but for symplectic manifolds (S, σ) . If $f \in C^{\infty}(S)$ we define the Hamiltonian vector field H_f for every $(p, z) \in TS$ by

$$\sigma_p(z, H_f) = df_p(z).$$

Here, if $dx_1, \ldots, dx_n, d\xi_1, \ldots, d\xi_n$ is a basis for T_p^*S , the differential df_p is the covector

$$df_p = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(p) \ dx_j + \frac{\partial f}{\partial \xi_j}(p) \ d\xi_j$$

Writing $\partial_{x_1}, \ldots, \partial_{x_n}$ for the dual basis of $T_p S$, $z = \sum_j c_j \partial_{x_j} + d_j \partial_{\xi_J}$, we have $dx_j(\partial_{\xi_k}) = \delta_{jk}$ and similarly for ξ_j , so

$$df_p(\partial_{x_j}) = \frac{\partial f}{\partial \xi_k}(p)$$

and similarly for ξ_j . We then define $\{f, g\} = H_f g$ as before, so

$$H_f = \sum_{j=1}^n \frac{\partial f}{\partial \xi_j} \partial_{x_j} - \frac{\partial f}{\partial x_j} \partial_{\xi_j}.$$

The notion of Poisson bracket still makes sense, as does the Jacobi identity. The proof of the Jacobi identity uses $d\sigma = 0$.

Definition 7.16. The *Hamiltonian flow* is defined by

$$\varphi_t = \exp(tH_f).$$

The Hamiltonian flow is again a symplectomorphism, $\varphi_t^* \sigma = \sigma$. In the classical mechanics interpretation of symplectic geometry, $f(x,\xi)$ is the energy of a particle with position x and momentum ξ , and is known as the Hamiltonian observable of the dynamical system (S, σ, φ) . **Example 7.17.** Take $f(x,\xi) = \xi^4 - x^4$. This is the Hamiltonian of the system $\cdot x = 4\xi^3$ and $\cdot \xi = 4x^3$. This can be solved as

$$x(t) = \frac{x(0)}{\sqrt{1 - 8tx(0)^2}}$$

which blows up in finite time. So we need a Hamiltonian which meets the hypotheses of Picard's theorem to be able to solve for the Hamiltonian flow. Here the problem is that energy is not conserved.

In fact, if $|\partial^2 f| \leq C$ then φ_t exists for all time, by Picard's theorem. This is in particular true if f is smooth and M is compact.

It is useful to know that Hamiltonian flows are always measure-preserving where the measure is given by the symplectic form. So we can use the tools of ergodic theory to study Hamiltonian flows.

7.4 Complete integrability

Definition 7.18. Let D be a symplectic manifold of dimension 2n and $f \in C^{\infty}(D)$. Then H_f is said to be a *completely integrable flow* if there are f_1, \ldots, f_n , known as *integrals of motion*, such that df_1, \ldots, df_n are linearly independent, $\{f, f_j\} = 0$ and $\{f_i, f_j\} = 0$.

Note that the Hamiltonian f is an integral of motion; i.e. $H_f f = 0$, or in other words $\varphi_t^* f = f$. An integrable flow is one with maximally many independent integrals of motion. Intuitively an integral of motion is a conserved quantity. It should not be a surprise that energy is conserved.

Example 7.19. Take f to only depend on momentum, and let $f_j(x,\xi) = \xi_j$. Then $H_f = \sum_j \partial_{\xi_j} f(\xi) \partial_{x_j}$ and $H_{f_j} = \partial_{x_j}$ so $\{f, f_j\} = 0$. In fact,

$$\varphi_t(x,\xi) = (x + t\nabla f(\xi), \xi).$$

Thus this is a completely integrable system.

Completely integrable systems are in some sense the opposite of ergodic systems: they have as many invariant sets as they possibly can!

For example, let L be a set where all the integrals of motion f_j are constant. Since the df_j are independent, L is an *n*-dimensional submanifold. This follows from the implicit function theorem. Moreover the H_{f_j} are a basis for the tangent space of L.

Recall the following result.

Theorem 7.20. A vector field X is tangent to a submanifold $L = \{g_1 = g_2 = \cdots = g_n = 0\}$ iff $Xg_i|_L = 0$.

Example 7.21 (uncoupled harmonic oscillators). Let $(q, p) \in \mathbb{R}^{2n}$ be a coordinate system and let $\sigma = dp \wedge dq$. Let $\omega_j > 0$ be constants and put

$$f(q,p) = \frac{1}{2}\sum_{j}p_j^2 + \omega_j^2 q_j^2$$

be the Hamiltonian of a family of n uncoupled harmonic oscillators. Let

$$f_j(q,p) = \frac{1}{2} \sum_j p_j^2 + \omega_j^2 q_j^2$$

be the Hamiltonian of the *j*th oscillator. Then the f_j s are uncoupled so $\{f_j, f_k\} = 0$. Let D be the open submanifold $f_j > 0$. We need to restrict to D because on the tangent spaces of $q_j = 0$ or $p_k = 0$, the df_j are not linearly independent. We have

$$H_f = \sum_j p_j \partial_{q_j} - \omega_j^2 q_j \partial_{p_j}$$

Moreover

$$(q_j(t), p_j(t)) = (q_j(0)\cos(\omega_j t) + p_j(0)\omega_j^{-1}\sin(\omega_j t), p_j(0)\cos(\omega_j t) - \omega_j q_j(0)\sin(\omega_j t)).$$

Interestingly, if we choose coordinates $(x,\xi) \in \mathbb{T}^n \times \mathbb{R}^n_+$ where

$$q_j = \sqrt{\frac{2\xi_j}{\omega_j}}\cos(x_j)$$

and $p_j = -\sqrt{2\xi_j\omega_j}\sin(x_j)$, then $f(q,p) = \sum_j \cos(\xi_j)$ and $dp \wedge dq = d\xi \wedge dx$. So the state space of this harmonic oscillator is identical to the positive tangent space of a torus. In these coordinates,

$$\varphi_t(x,\xi) = (x+t\omega,\xi).$$

7.5 The Liouville-Arnold-Jost theorem

Example 7.22. Take $M = \mathcal{H}^2 \times \mathbb{R}^2$ and let

$$f(x, y, \xi, \eta) = y^2(\xi^2 + \eta^2).$$

Then $f_1 = f$ and $f_2 = \xi$ are integrals of motion. We should look for coordinates on which the flow is linear. Doing this is far from obvious, but a theorem guarantees that such coordinates exist for any completely integrable system.

Theorem 7.23 (Liouville-Arnold-Jost). Suppose that $f = (f_1, \ldots, f_n)$ are such that the df_j are linearly independent, $\{f_i, f_j\} = 0$. Let $N = f^{-1}(0)$. If N is compact and connected, then N is diffeomorphic to \mathbb{T}^n , and there is an open neighborhood U of N which admits coordinates $(x, \xi) \in \mathbb{T}^n \times D_1$, where $0 \in D_1 \subset \mathbb{R}^n$ is open, given by a map $\psi : \mathbb{T}^n \times D^1 \to U$, and a local diffeomorphism μ near 0 of \mathbb{R}^n such that

$$\mu \circ f \circ \psi(\xi) = \xi$$

and ψ is a symplectomorphism.

In the coordinates (x,ξ) , the action of the Hamiltonian is linear. Unfortunately the construction of (x,ξ) is far from easy in general; the Liouville-Arnold-Jost theorem is far from constructive.

Corollary 7.24. If H_f is completely integrable then there are coordinates (x, ξ) such that the Hamiltonian f only depends on ξ .

In particular, if the Hessian of f is nondegenerate in (x, ξ) then we can solve for ξ using the implicit function theorem, say $\nabla_{\xi} f = \omega$. In this case there is a vector j orthogonal to ω , and quotienting out by the lattice generated by j, we end up with an irrational rotation.

We first prove a weak form of the Liouville-Arnold-Jost theorem, which intuitively says that symplectic geometry is locally trivial.

Theorem 7.25 (Darboux). Let (M, σ) be a 2*n*-dimensional symplectic manifold. Let $A, B \subseteq \{1, \ldots, n\}$. Let $q_j, p_k \in C^{\infty}$ near $p_0 \in M$, $j \in A$, $k \in B$, and $\{q_i, q_j\} = 0$, $\{p_k, p_\ell\} = 0$, $\{p_k, q_j\} = \delta_{kj}$. Assume that the dp_k, dq_j are linearly independent. Then there is a symplectomorphism κ from a neighborhood of $0 \in \mathbb{R}^{2n}$ to a neighborhood of p_0 , such that $\kappa^* q_j = x_j, \kappa^* p_k = \xi_k$.

To prove Darboux's theorem we need another theorem of differential geometry.

Theorem 7.26 (Frobenius). Let V_1, \ldots, V_r be vector fields on \mathbb{R}^n and $r \leq n$. Suppose that the $V_j(0)$ are linearly independent and $[V_j, V_k] = \sum_{\ell} c_{jk\ell} V_{\ell}$, where the $c_{jk\ell} \in C^{\infty}$ near 0.

If S is a submanifold of \mathbb{R}^n of codimension r such that $T_0S + \operatorname{span} V_j(0) = \mathbb{R}^n$, then the system of equations

$$V_j u = f_j$$
$$u|_S = u_0$$

has a unique solution near 0 iff

$$V_j f_i - V_i f_j = \sum_k c_{ijk} f_k.$$

In addition, there are local coordinates (y_1, \ldots, y_n) near 0 and a $r \times r$ invertible matrix $B = (b_{ij})$ of smooth functions near 0 such that

$$\partial_{y_i} = \sum_{j=1}^r b_{ij} V_j.$$

Proof. We first show that the third paragraph implies the second paragraph. The condition $V_j f_i - V_i f_j = \sum_k c_{ijk} f_k$ is invariant under linear combinations so it suffices to check when $V_j = \partial_{x_j}$, where S is given by $(x_1, \ldots, x_r) = h(x_{r+1}, \ldots, x_n)$. In this case the $c_{ijk} = 0$ and we are being asked to check

$$\partial_{x_i} f_j - \partial_{x_j} f_i = 0$$

or in other words $\omega = \sum_{j=1}^{r} f_j dx_j$ is a closed form. Locally then it is exact, say $du = \omega$. Then u solves the system of equations. We prove the third paragraph using induction on the dimension, say m. If m = 1 let $V_1 = a(x)\partial_x$ so $a(0) \neq 0$. Change coordinates to y so that dy = dx/a(x) and y(0) = 0.

Now if $m \ge 2$, we may change variables so that $V_1 = \partial_{x_1}$. This is always possible for any nonvanishing vector field. Without loss of generality we may assume that if $j \ge 2$,

$$V_j = \sum_{\ell=2}^m b_{j\ell} \partial_{x_\ell}$$

by subtracting off $b_{j1}V_1$ from V_1 . Let $x' = (x_2, \ldots, x_m)$ and change variables in x' using induction so that if $x_1 = 0$,

$$V_j = \sum_{\ell=2}^{\prime} b_{j\ell} \partial_{x_\ell}.$$

Thus we have discarded (x_{m+1}, \ldots, x_r) . Then

$$\frac{\partial b_{j\ell}}{\partial x_1} = V_1 V_j x_\ell = [V_1, V_\ell] x_\ell = \sum_{k=2}^r c_{1jk} V_k x_\ell = \sum_{k=2}^r c_{1jk} b_{k\ell}.$$

By uniqueness of ODE, it follows that $b_{j\ell}(x) = 0$ for $\ell > r$.

Proof of Darboux's theorem. We can assume $p_0 = 0$. Set $x_j = q_j$ and $\xi_k = p_k$ where $j \in A$ and $k \in B$. By the linear Darboux theorem, we can extend (x, ξ) to a full set of coordinates such that at 0, $\{x_i, x_j\} = \{\xi_i, \xi_j\} = 0$ and $\{\xi_k, x_\ell\} = \delta_{k\ell}$. In particular, $H_{q_j}(0) = -\partial_{\xi_j}$ and $H_{p_k}(0) = \partial_{x_k}$.

Suppose $J \notin A$ (the case for B is similar). We must find q_J such that $\{q_J, q_j\} = 0$ and $\{q_J, p_k\} = \delta_{Jk}$. We thus try to solve the equation

$$H_{q_i}q_J = 0, \quad H_{p_k}q_J = \delta_{kJ}, \quad q_J(x,\xi) = x_J$$

when $\xi_j = 0$ and $x_k = 0$. The solution to this equation exists by Frobenius' theorem. \Box

An important special case to Darboux's theorem is when $A = \emptyset$ and $B = \{1, \ldots, n\}$, i.e. we are given coordinates $\xi_1, \ldots, \xi_n \in C^{\infty}$ such that $d\xi_j(p_0)$ are linearly independent and $\{\xi_i, \xi_j\} = 0$. One then can find coordinates x_1, \ldots, x_n such that $\sum_j d\xi_j \wedge dx_j$ is indistinguishable from the standard system of symplectic coordinates in \mathbb{R}^{2n} .

Proof of the Liouville-Arnold-Jost theorem. Let $\Phi : \mathbb{R}^n/\Gamma \to f^{-1}(0)$ be our torus diffeomorphism, and ψ_0 be a local diffeomorphism from close to 0 to close to $\rho \in f^{-1}(0)$, such that $\psi_0^* f_j = \xi_j$ and $\psi_0^* \sigma$ is the standard symplectic form. Let Φ_t be the Hamiltonian flow.

If |t| is small,

$$\Phi_t \circ \psi_0(x,\xi) = \psi_0(x+t,\xi).$$

We want to use this relation to define ψ , an extension of ψ_0 . That would mean that for all R there is a neighborhood of 0, $D_2(R)$, such that $\Phi_t(\psi(0,\xi))$ is well-defined for $t \in B(0,R)$ and $\xi \in D_2(R)$. If not then there is an R such that for all r, there are $|t| \leq R$, and $|\xi| \leq r$, such that $\Phi_t(\psi(0,\xi))$ is not well-defined. Then $\Phi_t(\psi_0(0,\xi)) \notin M$ which is impossible.

Let γ_i be the generators of Γ .

Lemma 7.27. The map $\theta: B(0,R) \times D_2(R) \to M, \ \theta(x,\xi) = \Phi_x(\psi(0,\xi))$ satisfies

$$\theta^* \sigma = \sum_j d\xi_j \wedge dx_j$$

Proof of lemma. If (x, ξ) is small,

$$\theta(x,\xi) = \Phi_x(\psi_0(0,\xi)) = \psi_0(x,\xi)$$

and the claim follows from the fact that ψ_0 is a symplectomorphism.

Now put $\theta_s(x,\xi) = \theta(x-s,\xi)$, which is defined on $B(s,R) \times D_2(R)$. Then

$$\theta^*\sigma = (\Phi_s \circ \theta_s)^*\sigma = \theta^*_{-s}\Phi^*_s\sigma = \theta^*_{-s}\sigma$$

 \mathbf{SO}

$$\theta^* \sigma = \sum_j d\xi_j \wedge (dx_j - ds_j)$$

but the standard symplectic form is invariant under translation.

We want to find $\gamma_j(\xi)$ such that

$$\theta(\gamma_j(\xi),\xi) = \theta(0,\xi).$$

This mimics the fact that $\theta(\gamma_j, 0) = \theta(0, 0)$. Now $\psi_0^{-1}(\theta(\gamma_k, 0)) = (0, 0)$ so if (y, η) is small, the definition

$$\rho(y,\eta) = \psi_0^{-1}(\theta(\gamma_k + y,\eta))$$

makes sense (we put $\rho(0) = 0$). Then

$$\xi_j = \psi_0^* f_j(x,\xi) = f_j(\theta(\gamma_k + y,\eta)) = f_j(\psi_0(y,\eta)) = \eta_j$$

but $\rho = \psi_0^{-1} \circ \theta_{\gamma_k}$ and $\psi_0, \theta_{\gamma_k}$ are symplectomorphisms, so ρ is a symplectomorphism. Then

$$\rho^* \sum_j d\xi_j \wedge dx_j = \sum_j d\eta_j \wedge dy_j$$

This follows by the lemma. So

$$\rho(y,\eta) = (y + \frac{\partial Q_k}{\partial \eta}(\eta), \eta)$$

for some function Q such that

$$\frac{\partial Q_k}{\partial \eta}(0,0) = 0.$$

In particular if we put $(x,\xi) = \rho(y,\eta)$ we have $\xi = \eta$. Therefore, since ψ_0 is injective,

$$\psi_0(y + \partial_\xi Q_k(\xi), \xi) = \psi_0(0, \xi)$$

implies that we have the differential equation

$$y + \frac{\partial Q_k}{\partial \xi} = 0.$$

Thus y is determined by ξ . This uniquely determines the function γ_k . Now, by the group property of Φ_t , we can translate γ_k for all times, using the periodicity

$$\theta(x + \gamma_k(\xi), \xi) = \theta(x, \xi).$$

Thus $\theta : \mathbb{R}^n \times D_2 \to M$ is well-defined.

Now forget that k was the index used to describe the generators of the lattice Γ , and use it as a sequential index.

If we define $\Gamma(\xi) = \{\sum_{j} m_j \gamma_j(\xi)\}$, then we have a map

$$\theta_{\xi}: \frac{\mathbb{R}^n}{\Gamma(\xi)} \to f^{-1}(\xi) \cap U.$$

Now this map is surjective, but to check that it is injective, we notice that we just need to check when ξ is close to 0. If (x_k, ξ_k) and (x'_k, ξ'_k) are sequences of points (necessarily not in Γ) such that $\gamma_k \to 0$, $x_k - x'_k \notin \Gamma(\xi_k)$, and $\theta(x_k, \xi_k) = \theta(x'_k, \xi'_k)$. Passing to a subsequence, we choose x, x' such that $x_k \to x$ and $x'_k \to x'$. Then $\theta(x, 0) = \theta(x', 0)$, so $x - x' \in \Gamma(0)$, so there is a $\gamma_k \in \Gamma(\xi_k)$, such that $d(x_k - x'_k, \gamma_k)$ is small, but since Γ is discrete it follows that γ_k is a constant sequence. Translating by γ_k we have $x_k - x'_k = 0$. Thus close to any point θ_{ξ} is injective, hence θ_{ξ} is injective.

Finally we must "normalize periods", i.e. find symplectomorphisms to replace $\Gamma(\xi)$ with Γ . This is accomplished by a map σ such that

$$\sigma^* \sum_j d\xi_j \wedge dx_j = \sum_j d\eta_j \wedge dy_j.$$

7.6 The Toda lattice

We now consider the first example of a completely integrable system which is important in practice.

Suppose we have n particles in \mathbb{R}^N , interacting according to the Hamiltonian

$$H(q,p) = \sum_{j=1}^{n} \frac{p_j^2}{2m_j} + \sum_{i < j} V_{ij}(q_i - q_j).$$

Here q_i is the position of the *i*th particle and p_i is its momentum.

Example 7.28. The planetary system is given by

$$V_{ij}(q) = \frac{c_{ij}}{|q|}$$

for some $c_{ij} = c(m_i, m_j)$. Here

$$\dot{q}_j = \frac{m_j}{p_j}$$

and

$$\dot{p}_j = -\partial_{q_j} U(q)$$

where

$$-\partial_{q_j} U(q) = m_j \ddot{q}_j.$$

Trying to understand this system (using 7th-order perturbation theory by hand) lead to the discovery of Neptune.

Definition 7.29. Let $W_j(x) = e^x$ be the potential, and N = 1. Here

$$H(q,p) = \frac{1}{2} \sum_{j=1}^{n} p_j^2 + \sum_{j=1}^{n-1} e^{q_j - q_{j+1}},$$

where the particles are lined up so $-\infty = q_0 < q_1 < \cdots < q_n < q_{n+1} = \infty$. This system is called the *Toda lattice*.

We observe that the total momentum

$$P = \sum_{j} p_{j}$$

of the Toda lattice is conserved. Since $\{P, H\} = 0$, the Toda lattice is completely integrable when n = 2. But a miracle happens, and the Toda lattice is actually completely integrable for all n. But the state space of the Toda lattice is not compact so we cannot use the Liouville-Arnold-Jost theorem.

Theorem 7.30. There are integrals of motion F_j for the Toda lattice, which are polynomials in p_k and $e^{q_k-q_{k+1}}$.

To see this, we introduce Flaschka's coordinates. Here

$$a_j = \frac{e^{(q_j - q_{j+1})/2}}{2}$$

where j < n and

$$b_j = -\frac{p_j}{2}$$

where $j \leq n$. Then

$$\dot{a}_j = a_j(b_{j+1} - b_j)$$

and

$$\dot{b}_j = 2(a_j^2 - a_{j-1}^2).$$

Now this is missing a dimension but note that $s \mapsto (q + (s, \ldots, s), p)$ is an action of \mathbb{R} on \mathbb{R}^{2n} which preserves the system, namely translating all the q_j by t, so we just need coordinates on $\mathbb{R}^{2n-1} = \mathbb{R}^{2n}/\mathbb{R}$.

Now let L be the matrix with b_i on the diagonal and a_i on the subdiagonal and superdiagonal. Let B be the matrix with a_i on the superdiagonal and $-a_i$ on the subdiagonal. Then

$$\frac{dL}{dt} = [B, L]$$

if and only if (q, p) is a solution of the Hamiltonian flow.

Definition 7.31. A pair of matrices (B, L) such that dL = [B, L] dt if and only if the entries of B, L satisfy a Hamiltonian flow H are called a *Lax pair* for H.

The fact that (B, L) is a Lax pair for the Toda lattice is a "miracle." It completely trivializes the study of the Toda lattice.

Example 7.32. The KdV equation is a PDE

$$u_t = 6uu_x - u_{xxx}$$

which admits a Lax pair. Here $L = -\partial_x^2 + u$ and $B = -4\partial_x^3 + 6u\partial_x + 3u_x$ are operators acting on C^{∞} .

Note that L is symmetric, so it is diagonalizable and its eigenvalues are real. The $a_j > 0$ since they are exponentials, and a computation shows that every eigenvalue has a 1dimensional eigenspace. In fact, if φ is an eigenvector for the eigenvalue λ and we normalize so $\varphi_n = 1$, $a_{k-1}\varphi_{k-1} + b_k\varphi_k = \lambda\varphi_k$, so φ is uniquely defined by the normalization and a backwards recursion. Therefore L only has simple eigenvalues. Therefore $\lambda_j(a, b)$, the *j*th eigenvalue of L(a, b), determines a smooth function on $\mathbb{R}^{n-1}_+ \times \mathbb{R}^n$. This can be proven by using the implicit function theorem on the characteristic polynomial of L when (a, b) is close to 0.

In fact λ_i is constant on any flow.

Definition 7.33. Two matrices are *isospectral* if their spectra are identical.

We must show that the L(a(t), b(t)) are isospectral in t. To see this we solve the equation

$$\partial_t U(t) = B(a(t), b(t))U(t)$$

where U(0) = 1. This is a system of ODE, so U exists. Now U(t) is an orthogonal matrix, since U(0) is and

$$\frac{d}{dt}(U(t)^T U(t)) = 0$$

so $U(t)^T U(t) = U(0)^T U(0) = 1$. In other words, we have unitary of Schrodinger propagation. We also can show that L(t)U(t) = U(t)L(0). By unitarity it suffices to show that

$$\frac{d}{dt}U(t)^T L(t)U(t) = 0$$

and this follows by a straight computation:

$$\dot{U}^T L U + U^T L \dot{U} + U^T \dot{L} U = U^T B^T L U = U^T L B U = U^T [B, L] U$$
$$= U^T [L, B] U + U^T [B, L] U = 0.$$

Therefore L(a(t), b(t)) is isospectral. So

$$F_k(q,p) = \operatorname{tr}(L(a,b)^k)$$

is a preserved quantity. In fact $F_1 = \operatorname{tr} L = -P/2$, while $F_2 = \operatorname{tr} L^2 = H/2$.

As we require, the dF_k are linearly independent. It suffices to show that the determinant

$$\det\left(\frac{\partial F_k}{\partial \lambda_j}\right) \neq 0$$

and we have $F_k = \sum_j \lambda_j^k$. So

$$\det\left(\frac{\partial F_k}{\partial \lambda_j}\right) = n! \begin{vmatrix} 1 & \cdots & 1\\ \lambda_i & \cdots & \lambda_i\\ \lambda_i^2 & \cdots & \lambda_i^2\\ \vdots\\ \lambda_i^{n-1} & \cdots & \lambda_i^{n-1} \end{vmatrix} = n! \prod_{i < j} \lambda_j - \lambda_i$$

by Vandermonde's determinant formula, which is nonzero since the eigenvalues are simple.

We must show that $\{F_j, F_k\} = 0$ and it suffices to show $\{\lambda_j, \lambda_k\} = 0$. Let $\lambda = \lambda_j, \mu = \lambda_k$, $L\varphi = \lambda\varphi$, and $L\psi = \mu\psi$. Assume $||\varphi|| = ||\psi|| = 1$ and $\langle \varphi, \psi \rangle = 0$. If f, g are smooth then

$$\Psi^*\{f,g\}_* = \{\Psi^*f,\Psi^*g\},\$$

and the pushforward is along the change-of-coordinates $\Psi : (p,q) \mapsto (a,b)$. So we just want to show

$$\{\lambda,\mu\}_* = 0.$$

Theorem 7.34 (Feynman-Hellman). Suppose L is a differentiable one-parameter family of operators, and $L\varphi = \lambda\varphi$, where λ is a simple eigenvalue. If $||\varphi|| = 1$ then

$$\cdot \lambda = \langle (L\varphi)', \cdot \varphi \rangle.$$

Proof. Differentiating $L\varphi$ using the Leibniz rule we have

$$(L\varphi)' = (\lambda\varphi)' + \lambda \cdot \varphi.$$

Now take the inner product with φ on both sides and use the fact that $(||\varphi||)' = 1$.

We have

$$\partial_{b_j} L = e_j \otimes e_j$$

$$\partial_{a_j} L = e_j \otimes e_{j+1} + e_{j+1} \otimes e_j$$

so by the Feynman-Hellman theorem,

$$\partial_{b_j} \lambda = \varphi_j^2$$

$$\partial_{a_j} \lambda = 2\varphi_{j+1}\varphi_j.$$

Moreover

$$4\{\lambda,\mu\}_* = \sum_j (a_j\partial_{a_j}\lambda - a_{j-1}\partial_{a_{j-1}}\lambda)\partial_{b_j}\mu - (a_j\partial_{a_j}\mu - a_{j-1}\partial_{a_{j-1}}\mu)\partial_{b_j}\lambda$$

and plugging in the Feynman-Hellman computation,

$$2\{\lambda,\mu\}_{*} = \sum_{j} (a_{j}\varphi_{j}\varphi_{j+1} - a_{j}\varphi_{j-1}\varphi_{j})\psi_{j}^{2} - (a_{j}\psi_{j}\psi_{j+1} - a_{j}\psi_{j-1}\psi_{j})\varphi_{j}^{2}.$$

We take the convention $a_0 = a_n = 0$. We introduce the Wronskian,

$$W_j = a_j(\varphi_{j+1}\psi_j - \varphi_j\psi_{j+1}).$$

Here we are thinking of the ψ, φ as functions on finite sets, and viewing $W_j - W_{j-1}$ as the "derivative" of W_j . In this sense W_j is the Wronskian of φ, ψ .

In fact,

$$W_j - W_{j-1} = (\lambda - \mu)\varphi_j \psi_j.$$

This can be verified by computing

$$L\varphi_j = \lambda \varphi_j \psi_j$$
$$L\psi_j = \mu \psi_j \varphi_j$$

and subtracting the two lines.

It follows that

$$W_{j}^{2} - W_{j-1}^{2} = (\lambda - \mu)(a_{j}\varphi_{j+1}\varphi_{j} - a_{j-1}\varphi_{j-1}\varphi_{j})\psi_{j}^{2} - A)$$

where A is the same as the previous term but with φ replaced with ψ . Then

$$2\{\lambda,\mu\}_* = \frac{1}{\lambda-\mu}\sum_j W_j^2 - W_{j-1}^2 = (\lambda-\mu)^{-1}(W_n^2 - W_0^2)$$

since the Wronskians telescope, and by our convention we have $W_n = W_0 = 0$. So $\{\lambda, \mu\}_* = 0$.

Now define

$$x_j = |\langle \varphi_j, e_n \rangle|.$$

We will study the evolution of x. Let

$$f(z) = \langle (z - L)^{-1} e_n, e_n \rangle = \sum_j \frac{x_j^2}{z - \lambda_j}.$$

Since the $a_j > 0$ it is no loss of information to replace a_j with a_j^2 . Define

$$\chi : \mathbb{R}^{n-1}_{-} \times \mathbb{R}^n \to \mathbb{R}^n \times \{ x \in \mathbb{R}^n_+ : ||x|| = 1 \}$$
$$(a^2, b) \mapsto (\lambda, x).$$

If we can invert χ we will know how (a, b) and hence (q, p) evolves explicitly in terms of the eigenvalues λ and scalars x.

Lemma 7.35. We have

$$-\frac{\dot{x}_j}{x_j} = -\lambda_j + \sum_k \lambda_k x_k^2.$$

Moreover, this flow preserves the set $\{x \in \mathbb{R}^n_+ : ||x|| = 1\}$.

Proof. To see that ||x|| = 1 is preserved, we simply expand out the definition of ||x||, differentiate it, and notice that ||x||' = 0. Moreover, if $\alpha_j = \log x_j$, then $\dot{\alpha}$ is Lipschitz and so exists for all time by a Picard-type theorem.

As for the main claim,

$$L(t)U(t) = U(t)L(0)$$

 \mathbf{SO}

$$\varphi_k(t) = U(t)\varphi_k(0)$$

Since $\cdot U = BU$,

$$\varphi_k(t) = B(t)\varphi_k(t).$$

Then

$$\langle \cdot \varphi_k(t), e_n \rangle = a_{n-1} \langle \varphi_n, e_{n-1} \rangle + b_n x_k$$

 \mathbf{SO}

$$\cdot x_k = -\lambda_k x_k + b_n x_k$$

Since ||x|| = 1, (||x||)' = 0, so differentiating,

$$0 = -\sum_{k} \lambda_k x_k^2 + b_n$$

н

The system given by the lemma is easy to solve. Set

$$y_j(t) = y_j(0)e^{-\lambda_j t}$$

and set

$$x_j = \frac{y_j}{||y||}.$$

Using this normalization,

$$\frac{x_j(t)^2}{x_j(0)^2} = \frac{e^{-2\lambda_j t}}{\sum_k x_k(0)^2 e^{-2\lambda_k t}}.$$

We can therefore view x as a linear dynamical system. Thus, if we can invert χ , we will explicitly have solved the system (q, p).

We can view the situation as a commutative diagram

$$\begin{array}{cccc} (q(0),p(0)) & \longrightarrow & L(0) & \longrightarrow & (\lambda,x(0)) \\ & & & & \downarrow \\ & & & & \downarrow \\ (q(t),p(t)) & \longleftarrow & L(t) & \longleftarrow & (\lambda,x(t)) \end{array}$$

We are looking for the arrow $(\lambda, x(t)) \to L(t)$.

Put

$$f(z) = \langle (z - L)^{-1} e_n, e_n \rangle = \sum_j \frac{x_j^2}{z - \lambda_j}$$

Theorem 7.36 (Stjeltes). We have

$$f(z) = \frac{1}{z - b_n - \frac{a_{n-1}^2}{z - b_{n-1} - \frac{a_{n-2}^2}{z - b_{n-2} - \cdots}}}.$$

Proof. Let L_j be the matrix containing the first j rows and columns of L. Let $\Delta_j(z) = \det(z - L_j)$. By Cramer's rule,

$$f(z) = \frac{\Delta_{n-1}(z)}{\Delta_n(z)}.$$

Moreover,

$$\Delta_j(z) = (z - b_j)\Delta_{j-1}(z) - a_{j-1}^{-1}\Delta_{j-2}(z)$$

by the recurrence definition of the determinant. Here we take $\Delta_0 = \Delta_{-1} = 0$.

Let $\rho_j = \Delta_j / \Delta_{j-1}$, so

$$\rho_j(z) = z - b_j - \frac{a_{j-1}^2}{\rho_j - 1}.$$

Here we take $\rho_1(z) = z - b_1$. Then $\rho_n = f^{-1}$ gives the continued fraction expansion.

Now $f = P_{n-1}/Q_n$ where

$$P_{n-1}(z) = \sum_{j=1}^{n-1} x_j^2 \prod_{i \neq j} z - \lambda_i$$

and

$$Q_n(z) = \prod_{j=1}^n z - \lambda_j$$

A straight computation gives

$$\frac{Q_n(z)}{P_{n-1}(z)} = z - \beta - \alpha \frac{Q_{n-2}(z)}{P_{n-1}(z)}.$$

Thus

$$f(z) = \frac{1}{z - \beta - \alpha g(z)}$$

where $g = Q_{n-2}/P_{n-1}$ and α, β are uniquely determined by x, λ , and g is of the same form as f, say

$$g(z) = \sum_{j=1}^{n-1} \frac{\sigma_j}{z - \mu_j}$$

so σ_j is the residue of μ_j at z. Thus

$$\sigma_j = -\frac{1}{\alpha f'(\mu_j)} > 0.$$

We now show that the Toda lattice admits scattering, i.e. we can determine its behavior at $t = +\infty$ from its behavior at $t = -\infty$.

Theorem 7.37. There is a $\delta > 0$ such that $q_j(t) = \alpha_j^+ t + \beta_j^+ + O(e^{-\delta t})$ as $t \to +\infty$ and $q_j(t) = \alpha_j^- t + \beta_j^- + O(e^{\delta t})$ as $t \to -\infty$, where $\lambda_i < \lambda_{i+1}, \alpha_j^+ = -2\lambda_{n-j+1}$, and $\alpha_j^- = -2\lambda_j$.

Proof. $a_j(t)^2$ and $b_j(t)$ are rational functions of $e^{-\lambda_j t}$. Thus they have asymptotic expansions $c_1^{\pm}e^{\eta_1^{\pm}t} + c_2^{\pm}e^{\eta_2^{\pm}t} + \cdots$ as $t \to \pm \infty$ and $\eta_j^{\pm} = \sum_k m_{jk}^{\pm}\lambda_j$ where $m_{jk}^{\pm} \in \mathbb{Z}$.

Now

$$\int_{t_1}^{t_2} a_j^2 - a_{j-1}^2 = b_j(t_2) - b_j(t_1)$$

telescopes, and since $2\sum_j b_j^2 + 4\sum_j a_j^2$ is a conserved quantity a sum of b_j must be bounded. Therefore $a_j^2 \in L^1(dt)$. Since a_j has an asymptotic expansion in exponentials, it follows that $a_j \to 0$. Thus $L(t) \to L(\pm \infty)$ which is a diagonal matrix. Solving $\dot{a}_j/a_j = b_{j+1} - b_j$ we have $b_j(+\infty) = \lambda_{n-j+1}$ and $b_{-j}(-\infty) = \lambda_j$.

So the velocities are not just permuted, but are reversed in order.

7.7 The QR algorithm

We now apply the theory of completely integrable systems to study a diagonalization algorithm from numerical linear algebra. It was voted among the top ten algorithms of the last century, along with the fast Fourier transform, which confusingly was discovered by Gauss long before the last century.

Theorem 7.38. For every invertible real matrix A there is a unique factorization A = QR such that $R_{ii} > 0$, Q is orthogonal, R is upper triangular.

Proof. Let $A = [a_1 \cdots a_n]$. Since A is invertible the a_i are linearly independent, so apply the Gram-Schmidt algorithm to A to obtain an orthogonal matrix Q. Let

$$R = \begin{bmatrix} \langle e_1, a_1 \rangle & \langle e_1, a_2 \rangle & \cdots \\ 0 & \langle e_2, a_2 \rangle & \cdots \\ \vdots & \vdots & \end{bmatrix}$$

then A = QR. If $Q_1R_1 = Q_2R_2$ then $Q_1^{-1}Q_2 = R_2R_1^{-1}$ so $R_2R_1^{-1}$ is an upper triangular matrix which is positive on the diagonal and yet has determinant 1, so $R_2 = R_1$.

Given A = QR, let T(A) = RQ. Note that

$$T(A) = Q^* Q R Q = Q^{-1} A Q.$$

Thus T(A) and A are isospectral, and T preserves the space B_{sa} of self-adjoint matrices.

Theorem 7.39 (Kublanovskya-Francis, 1961). If A is positive-definite and self-adjoint with simple eigenvalues then

$$\lim_{n \to \infty} T^n(A) = \operatorname{diag}(\lambda_i)$$

where λ_i are the eigenvalues of A, $\lambda_i > \lambda_{i+1}$.

That is, T is a discrete isospectral system which converges to the diagonalization. This is analogous to the Toda lattice, where we had a continuous isospectral system which converged to the diagonalization. We proved that the convergence of L is exponential, namely that

$$||L(t) - L(\infty)|| = O(\exp(-2t\min_{j}(\lambda_{j} - \lambda_{j-1}))).$$

Let J be the space of tridiagonal matrices such that the subdiagonal and superdiagonal are positive and identical. An algebraic computation shows that T sends J to tridiagonal matrices. Recall from Toda lattice theory that there is a diffeomorphism $\chi : J \to \{(\lambda, x)\}$ where $\lambda_i > \lambda_{i+1} > 0$, λ is the spectrum of the preimage L, $x_j > 0$, and ||x|| = 1. Here x is defined by the property that if $L = U\Lambda U^*$ then $x = U^*e_1$.

Lemma 7.40. Let

$$(\lambda, x^k) = \chi(T^k(\chi^{-1}(\lambda, x_0))).$$

Then

$$||\Lambda^k x^0||x^k = \Lambda^k x^0.$$

Proof. It suffices to check this when k = 1. This can be checked by noting that the unitary matrix U_1 obtained by diagonalizing $T(\Lambda)$ is given by the Gram-Schmidt algorithm. \Box

Recall that the Toda flow is given by

$$x(t) = \frac{e^{At}x(0)}{||e^{At}x(0)||}.$$

Similarly we here have

$$x^{k} = \frac{e^{Ak}x(0)}{||e^{Ak}x(0)||}$$

where $A = \log \Lambda$. Thus in the coordinates (x, λ) the QR algorithm is a discretized Toda flow, provided that our initial conditions were tridiagonal.

We now construct a discrete Lax pair for the QR algorithm. For any matrix A, we may uniquely decompose

$$A = \pi_S A + \pi_L A$$

where $\pi_S A$ is anti-self-adjoint and $\pi_L A$ is lower triangular. In fact, $\pi_S A$ is given by taking the part of A above the diagonal and reflecting it over the diagonal. Given a continuous function $G : \mathbb{R}_+ \to \mathbb{R}$ we extend it to self-adjoint matrices L by

$$G(L) = UG(\Lambda)U^{*}$$

where $LU = U\Lambda$ is the diagonalization of L given by the spectral theorem and G acts on diagonal matrices pointwise. This is a functional calculus for self-adjoint operators. Let $B_G = \pi_S \circ G$. Then the map $L \mapsto [B_G(L), L]$ preserves tridiagonality. This can be easily checked for $G(\lambda) = \lambda^k$ and then may use a variant of the Stone-Weierstrass theorem. One may then check that if we define

$$\cdot x(t) = G(\Lambda)x(t) - \langle G(\Lambda)x(t), x(t) \rangle x(t)$$

then $\cdot L = [B_G(L), L]$. For appropriate G this gives a Lax pair for the QR algorithm.

Chapter 8

Complex dynamics

8.1 Siegel's KAM theorem

Let U, V be neighborhoods of 0 in \mathbb{C} . Assume $f : U \to V$ is holomorphic, f(0) = 0, $f'(0) = e^{2\pi i \theta}$ where θ is an irrational angle. We would like to find a conformal transformation φ defined locally near 0 which conjugates f to rotation by $e^{2\pi i \theta}$. If this is the case, we will have lots of invariant sets for f, namely the φ -image of circles centered around 0. Whether this is possible depends strongly on number-theoretic properties of θ .

One can view the rotation $e^{2\pi i\theta}$ as a simple Copernican model of the solar system; then the real solar system is a slight perturbation, and we would like to know that it is quasiperiodic. That is, a slight perturbation of f preserves the dynamics.

Theorem 8.1 (Siegel). Suppose that θ is *diophantine* in the sense that there is a C > 0 and μ large enough that for all $p, q \in \mathbb{Q}$,

$$\left|\theta - \frac{p}{q}\right| \ge \frac{C}{q^{\mu}}$$

Then there is a conformal transformation φ such that

$$f(\varphi(z)) = \varphi(e^{2\pi i\theta}z).$$

We first show that there are diophantine numbers. Let $\mu > 2$. Let E be the set of all θ such that for infinitely many $p, q, |\theta - p/q| < q^{-\mu}$. Then for every n,

$$E \subseteq \bigcup_{q \ge n} \bigcup_{p=0}^{q} [p/q - q^{-\mu}, p/q + q^{-\mu}]$$

 \mathbf{SO}

$$|E| \le \sum_{q \ge n} \sum_{p=1}^{q} 2q^{1-\mu} = O(n^{2-\mu})$$

so E has measure 0, hence most numbers are diophantine.

So fix a diophantine number θ and let $\lambda = e^{2\pi i \theta}$.

We now observe that if

$$f(\varphi(z)) = \varphi(\lambda z),$$

then φ is automatically injective, hence conformal by the inverse function theorem. In fact if $\varphi(z) = \varphi(w)$ then

$$\varphi(\lambda z) = f(\varphi(z)) = f(\varphi(w)) = \varphi(\lambda w)$$

and iterating we see that

$$\varphi(\lambda^n z) = \varphi(\lambda^n w)$$

but the $\lambda^n z$ are a set with an accumulation point, and if two holomorphic functions agree on a set with an accumulation point then they agree everywhere.

Lemma 8.2. There is a φ iff there is a $\delta > 0$ such that $f^n(z)$ is a bounded sequence if $|z| < \delta$.

Proof. If φ exists then

$$f^n(z) = \varphi(\lambda^n \varphi^{-1}(z))$$

which is obviously bounded.

Conversely, assume a bound M exists and let

$$h_n(z) = \frac{1}{n} \sum_{m=0}^{n-1} \lambda^{-m} f^m(z)$$

be the ergodic average. By assumption $|h_n(z)| \leq M$. By the Cauchy estimate for the disc $|z| < \delta/2$,

$$|h'_n(z)| \le \frac{2M}{\delta}$$

so by the Ascoli theorem, we can pass to a Cauchy subsequence and assume $\lim_n h_n = h$ in C^0 . By Morera's theorem h is holomorphic, and

$$h'_n(0) = \frac{1}{n} \sum_{m=0}^{n-1} \lambda^{-m} \lambda^m = 1$$

so h'(0) = 1 and by the inverse function theorem h has a conformal inverse φ .

We now show that some hypothesis on θ is necessary.

Theorem 8.3 (Pfeiffer). There is a θ such that for any polynomial f of degree ≥ 2 , no φ exists.

Proof. Assume

$$f(z) = z^d + \dots + \lambda z.$$

Assume φ exists and is defined if $|z| < \delta$. Let z_1, \ldots, z_{d^n-1} be the nonzero fixed points of f^n , which exist since $f'(0) = \lambda \neq 0$. Then they are nonzero zeroes of

$$f^n(z) - z = z^{d^n} + \dots + (\lambda^n - 1)z.$$

By assumption, $|z_j| \ge \delta$; since $|z| < \delta$ is conjugated to rotation. Thus

 $\delta^{d^n} < \delta^{d^n - 1} \le \prod_j |z_j| = |\lambda^n - 1|.$

The last equality is true because the z_j are zeroes of a polynomial whose constant term is $\lambda^n - 1$. So we must find a θ such that for every d, δ we can find an n such that

For some sequence $q_k \in \mathbb{N}$, to be determined, let

 $\theta = \sum_{k=1}^{\infty} 2^{-q_k}.$

Then

Thus

Now let

Let

which is a contradiction.

Assume $\delta^{2^{q_k}} < |\lambda^{2^{q_k}} - 1|$, so

We now prove Siegel's theorem. Siegel used an extremely complicated number-theoretic proof, but we will give Moser's proof, which easily follows from KAM theory.

Proof of Sieqel's theorem. Our goal is to construct φ with the desired properties, which in particular imply

Thus we define $\varphi(z) = z + \hat{\varphi}(z)$ so $\hat{\varphi}(z) = O(z^2)$, and write $f(z) = \lambda z + \hat{f}(z)$. As in the case of the cat map, we want to solve the equation

 $\hat{\varphi}(\lambda z) - \lambda \hat{\varphi}(z) = \hat{f}(z + \hat{\varphi}(z))$

for $\hat{\varphi}$. Unfortunately the cat map was hyperbolic, while the dynamics here are rotational.

 $\hat{f}(z) = \sum_{n=1}^{\infty} b_n z^n.$

Linearize the equation we want to solve, we get a function

 $\hat{\psi}(z) = \sum_{n=1}^{\infty} c_n z^n,$

$$\varphi(z) = z + O(z^2).$$

$$\delta^{d^n} \ge |\lambda^n - 1|.$$

$$\delta^{d^{2^{q_k}}} = O(2^{q_k - q_{k+1}}).$$

 $|1 - \lambda^{2^{q_k}}| \sim 2^{q_k - q_{k+1}}.$

 $q_{k+1} = O_{\delta}(d^{2^{q_k}}).$

 $\log q_{k+1} > k e^{q_k}$

such that

$$\hat{\psi}(\lambda z) - \lambda \hat{\psi}(z) = \hat{f}(z).$$

Solving the Taylor series we have

$$c_n = \frac{b_n}{\lambda^n - \lambda}.$$

Since θ is diophantine, there are C, μ such that

$$|\lambda^n - 1|^{-1} \le C \frac{n^\mu}{\mu!}.$$

Moreover $|\hat{f}'(z)| < \delta$ if $|z| < \varepsilon$ for some $\varepsilon > 0$.

In fact, Cauchy estimates give us

$$|b_j| \le \frac{\delta}{j\varepsilon^{j-1}}.$$

Assume $\eta < 1/5$ and η is much smaller than ε . Then if $|z| < (1 - \eta)\varepsilon$,

$$|\psi'| \le \sum_{j=2}^{\infty} \frac{jb_j}{|\lambda^j - \lambda|} \varepsilon^{j-1} (1-\eta)^{j-1} < \frac{C\delta}{\mu!} \sum_{j=1}^{\infty} j(j+1) \cdots (j+\mu-1)(1-\eta)^{j-1} = \frac{C\delta}{\eta^{\mu+1}}$$

where we used the Newton binomial formula. Thus if $C\delta < \eta^{\mu+2}$, which is possible if ε and hence δ are small enough, if $|z| < (1 - \eta)\varepsilon$ then

 $|\hat{\psi}'(z)| \le \eta.$

Thus ψ carries $D(0, \varepsilon(1-4\eta)) \to D(0, \varepsilon(1-3\eta))$ and similarly if $|z| = \varepsilon(1-\eta)$,

$$|\psi(z)| = |z + \hat{\psi}(z)| \ge \varepsilon (1 - 2\eta).$$

Assume $|w| < \varepsilon(1-2\eta)$ and $|z| = \varepsilon(1-\eta)$. Then

$$|w| = |\psi(z) - (\psi(z) - w)| < |\psi(z)|.$$

By Rouche's theorem, ψ and $\psi - w$ have the same number of zeroes in $D(0, \varepsilon(1-\eta))$. Then ψ is a bijection $D(0, \varepsilon(1-\eta)) \to D(0, \varepsilon(1-2\eta))$. We thus have a diagram

$$D(0,\varepsilon(1-4\eta)) \xrightarrow{\psi} D(0,\varepsilon(1-3\eta)) \xrightarrow{f} D(0,\varepsilon(1-2\eta)) \xrightarrow{\psi^{-1}} D(0,\varepsilon(1-\eta)).$$

Let $\psi \circ g = f \circ \psi$, so g carries $D(0, \varepsilon(1-4\eta)) \to D(0, \varepsilon(1-\eta))$.

We now estimate \hat{g} , where $g(z) = \lambda z + \hat{g}(z)$. We have $\psi \circ g = f \circ \psi$. Thus

$$f(\psi(z)) = \lambda z = \hat{g}(z) + \hat{\psi}(g(z)) = \lambda(z + \hat{\psi}(z) + \hat{f}(z + \hat{\psi}(z))).$$

Therefore

$$\hat{g}(z) + \hat{\psi}(\lambda z + \hat{g}(z)) = \lambda \hat{\psi}(z) + \hat{f}(z + \hat{\psi}(z)).$$

Since $\hat{f}(z) = \hat{\psi}(\lambda z) - \lambda \hat{\psi}(z)$,

$$\hat{g}(z) = \hat{\psi}(\lambda z) - \hat{\psi}(\lambda z + \hat{g}(z)) + \hat{f}(z + \hat{\psi}(z)) - \hat{f}(z).$$

Let

$$C = \max_{D(0,(1-4\eta)\varepsilon)} |\hat{g}|.$$

Then

$$C \le \sup |\hat{\psi}'|C + \sup |f'| \sup |\hat{\psi}| \le C\eta + \delta \sup |\hat{\psi}|.$$

We apply Schwarz' lemma to $\hat{\psi}'$ to see

$$|\hat{\psi}'(z)| \le c_0 \frac{\delta|z|}{\varepsilon(1-\eta)\eta^{1+\mu}}$$

Integrating both sides, we bound $|\hat{\psi}|$ and see that

$$C \le C\eta + \frac{c_0 \delta^2}{2\eta^{\mu+1}} (1-\eta)\varepsilon$$

Solving for C,

$$\max_{D(0,(1-4\eta)\varepsilon)} |\hat{g}| \le \frac{c_0 \delta^2}{2} \eta^{-\mu-1} \varepsilon.$$

By a Cauchy estimate,

$$|\hat{g}'(z)| \le \frac{c_0 \delta^2}{2\eta^{\mu+2}}$$

if $|z| < (1 - 5\eta)\varepsilon$.

We now iterate: let $g_0 = f$, and let $\psi_{n+1} \circ g_{n+1} = g_n \circ \psi_{n+1}$. Let $\eta_k = \alpha^k \eta_0$, $\alpha < 1$, and

$$R = \varepsilon \prod_{k} 1 - 5\eta_k > 0.$$

Then $|\hat{g}'_n| \leq \delta_n \leq \eta_k^{\mu+2} \to 0$. So $g_n \to \lambda$ and

$$f \circ \varphi_n = \varphi_n g_n.$$

A calculation using a Cauchy estimate shows that $|\varphi_{n+1} - \varphi_n| \leq \delta_n C \to 0$, so the φ_n are a Cauchy sequence, say $\varphi_n \to \varphi$, and $f \circ \varphi = \varphi \circ \lambda$.

Chapter 9

Entropy and information theory

We now consider a way to measure the complexity of a dynamical system. Though this was originally motivated by thermodynamics and statistical mechanics, it did not come to full form until work of Shannon in 1948.

9.1 Shannon's axioms of entropy

Fix a probability space (X, Σ, μ) and a partition $\mathcal{P} = (P_i)_i$ of X (so $P_i \cap P_j$ is empty and $\bigcup_i P_i = X$). Let $p_i = \mu(P_i)$, so p_i is the "probability of finding a particle in P_i ." Entropy is supposed to measure the "uncertainty" of the box P_i that we find the particle in.

Equivalently, we seek motivation from ergodic theory. Let $(X_j, \Sigma_j, \mu_j, T_j)_j$ be a family of measure-preserving systems. Recall that $(X_1, T_1), (X_2, T_2)$ are isomorphic if there is an isomorphism of measurable spaces $\varphi : X_1 \to X_2$ such that $\varphi \circ T_1 = T_2 \circ \varphi$. (Note that φ is not assumed to be an isomorphism of measure spaces.) We want an invariant which allows us to distinguish two measure-preserving systems. (This is analogous to algebraic topology.)

We want a function H defined on probability vectors $\vec{p} \in \Delta_n$, where $p_i = \mu(P_i)$ is obtained from the partition \mathcal{P} , and Δ_n is the *n*-simplex, and $H : \Delta_n \to \mathbb{R}$. In this case, we let $H(\mathcal{P}) = H(\vec{p})$. We want $H(\vec{p}, 0) = H(\vec{p})$ (so adding an empty set to the partition does not affect the complexity) and $k \mapsto H(1/k, \ldots, 1/k)$ is increasing (so complexity increases as the partition gets finer).

If \mathcal{Q} is also a partition, we let $\mathcal{P} \lor \mathcal{Q}$ be the set of all intersections of entries in \mathcal{P} and in \mathcal{Q} . We will demand

$$H(\mathcal{P} \lor \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q}|\mathcal{P})$$

where

$$H(\mathcal{Q}|\mathcal{P}) = \sum_{P \in \mathcal{P}} \mu(P) \mu\left(\frac{\mu(Q_1 \cap P)}{\mu(P)}, \cdots, \frac{\mu(Q_k \cap P)}{\mu(P)}\right)$$

is the "conditional complexity" of \mathcal{Q} given that we know \mathcal{P} . These conditions on H are known as *Shannon's axioms*.

Theorem 9.1 (Shannon 1948). Suppose H satisfies Shannon's axioms; then there is a constant $K \ge 0$ such that

$$H(\vec{p}) = -K \sum_{j} p_j \log p_j.$$

Notice that this implies that H is (up to a choice of Boltzmann constant, which amounts to a choice of physical units) the thermodynamic entropy, where j ranges over a set of microstates, as defined by Boltzmann.

Proof. Define

$$h(n) = H(1/n, \dots, 1/n).$$

Assume $n = k\ell$, say $p_j = 1/k$, $q_i = 1/\ell$, where \vec{p}, \vec{q} arise from partitions \mathcal{P}, \mathcal{Q} respectively. Then

$$h(n) = h(k\ell) = H(\mathcal{P} \lor \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q}|\mathcal{P}) = h(k) + \frac{1}{k} \sum_{j=1}^{k} h(\ell) = h(k) + h(\ell)$$

Therefore h is a multiplicative-to-additive isomorphism. The only such isomorphism defined on integers which is increasing is log. In fact, we can use the monotonicity to show that for any n, if $k \ge \ell$,

$$\left|\frac{h(k)}{h(2)} - \frac{\log k}{\log 2}\right| < \frac{1}{n}$$

Then if \vec{p} is a rational point in Δ_n , one can use this to show that $H(\vec{p}) = c \sum_j p_j \log p_j$. Continuity extends this to all of Δ_n .

Definition 9.2. If \vec{p} is the probability vector corresponding to the partition \mathcal{P} of a fixed measure space, we define the *Shannon entropy* by

$$H(\mathcal{P}) = H(\vec{p}) = -\sum_{j} p_j \log p_j.$$

9.2 Conditional entropy

We now give a more abstract setup for Shannon entropy. Given $\xi = \{A_j\}_j$ a partition of (X, Σ, μ) , let $\xi(x) = A_j$ where $x \in A_j$. Then the join $\bigvee_i \xi_i$ is defined by

$$\left(\bigvee_{i}\xi_{i}\right)(x)=\bigcap_{i}\xi_{i}(x).$$

Definition 9.3. The *information function* $I(\xi)$ of ξ is defined by

$$I(\xi, x) = -\log \mu(\xi(x)).$$

Then entropy satisfies

$$H(\xi) = \int_X I(\xi, x) \ d\mu(x).$$

Indeed, $I(\xi, \cdot)$ is constant on any A_j , so it follows that

$$\int_X I(\xi, x) \ d\mu(x) = -\sum_j \mu(A_j) \log \mu(A_j).$$

Definition 9.4. Let $\mathcal{A} \subseteq \Sigma$ be a sub- σ -algebra. The *conditional entropy* of ξ with respect to \mathcal{A} is defined by

$$H(\xi|\mathcal{A}) = -\sum_{j} \int_{X} E(A_{j}|\mathcal{A}) \log(E(A_{j}|\mathcal{A})) \ d\mu$$

Here

$$E(A|\mathcal{A}) = E(1_A|\mathcal{A})$$

is the conditional expectation. If \mathcal{A} is generated by a partition η ,

$$E(f|\mathcal{A})(x) = \frac{1}{\mu(\eta(x))} \int_{\eta(x)} f \ d\mu,$$

and we write $E(f|\eta) = E(f|\mathcal{A})$. In particular,

$$E(A|\eta)(x) = \frac{\mu(A \cap \eta(x))}{\mu(\eta(x))}.$$

Unravelling the definitions, $H(\xi|\eta)$ agrees with the previous definition for conditional entropy that we gave.

We think of $H(\xi|\eta)$ as the entropy given that we know everything about η . In particular, $H(\xi|\Sigma) = 0$ (where Σ is the σ -algebra). To see this, note that $E(f|\Sigma)$ is constant for any f, so $E(A|\mathcal{A}) \in \{0,1\}$ and hence $E(A|\mathcal{A}) \log E(A|\mathcal{A}) = 0$. On the other hand, $H(\xi) =$ $H(\xi|\{0,X\})$. More generally if $\mathcal{A} \subseteq \mathcal{B}$ then

$$H(\xi) \ge H(\xi|\mathcal{A}) \ge H(\xi|\mathcal{B}) \ge 0.$$

If T is measure-preserving then

$$H(\xi|\mathcal{A}) = H(T^{-1}\xi|T^{-1}\mathcal{A}).$$

In particular $H(\xi) = H(^{-1}\xi)$. We also have $H(\xi \lor \eta) \le H(\xi) + H(\eta)$.

9.3 The entropy of a group action

Let (X, Σ, μ, T) be a measure-preserving system.

Lemma 9.5. For any partition ξ , the limit

$$\lim_{N \to \infty} \frac{1}{N} H\left(\bigvee_{j=0}^{N-1} T^{-j} \xi\right) = \inf_{N} H\left(\bigvee_{j=0}^{N-1} T^{-j} \xi\right) > -\infty.$$

Proof. Let

$$a_N = H\left(\bigvee_{j=0}^{N-1} T^{-j}\xi\right).$$

Since $H(\xi \lor \eta) \le H(\xi) + H(\eta)$, we have

$$0 \le a_{N+M} \le a_N + a_M,$$

Therefore

$$\lim_{N \to \infty} \frac{a_N}{N} = \inf_{N \to \infty} \frac{a_N}{N} > -\infty.$$

п	-	

Definition 9.6. Let

$$h(T,\xi,\mu) = \lim_{N \to \infty} \frac{1}{N} H\left(\bigvee_{j=0}^{N-1} T^{-j}\xi\right)$$

Then we may define the *entropy* of the transformation T by

$$h(T,\mu) = \sup_{\xi} h(T,\xi,\mu),$$

 ξ ranging over all finite partitions of (X, Σ, μ) into measurable sets.

We have $h(T, \eta) \leq h(T, \xi) + H(\eta|\xi)$. Moreover,

$$h(T,\xi) = \lim_{N \to \infty} H(\xi) \bigvee_{i \le N} T^{-i}\xi.$$

Trivially, the σ -algebra generated by $T^{-1} \bigvee_{j \leq k-1} T^{-j} \xi$ is contained in the σ -algebra generated by $T^{-1} \bigvee_{j \leq k} T^{-j} \xi$, so if we let

$$b_k = H(\xi | T^{-1} \bigvee_{j \le k} T^{-j} \xi)$$

then the b_k form a decreasing sequence, and is bounded from below, so say $b_k \to b$. In particular, b_k converges in Cesaro mean to b, so

$$\lim_{N \to \infty} H(\bigvee_{j \le k} T^{-j} \xi) = \lim_{N \to \infty} H(\xi | T^{-1} \bigvee_{j \le k} T^{-j} \xi).$$

Example 9.7. Take the partition ξ , $A_0 = \{x : x > 0\}$, $A_1 = \{x : x \leq 0\}$ of the circle $S^1 = \{(x, y) : x^2 + y^2 = 1\}$. Let T be the irrational rotation, so at each state $\bigvee_{j \leq k} T^{-j} \xi$ has 2 more sets than the previous stage. Thus

$$H\bigvee_{j=0}^{N-1}T^{-j}\xi \le \log 2N$$

and dividing both sides by N and taking $N \to \infty$ we see that

$$h(T,\xi) = 0$$

On the other hand, for any transformation, $h(T^q) = qh(T)$, so if T is a rational rotation of denominator q, $T^q = 1$ and h(1) = 0. Therefore h(T) = 0. More generally, this implies that any periodic transformation has 0 entropy. The definition of h(T) is slightly useless because one cannot easily take a supremum over partitions. However, Sinai, a student of Kolmogorov, found a less useless equivalent definition.

Definition 9.8. Let (X, Σ, μ, T) be a measure-preserving system. A partition ξ is said to be a *one-sided generator* of Σ if Σ is the σ -algebra generated by

$$\{A: \exists j (A \in T^{-j}\xi)\}$$

modulo μ -null sets.

Theorem 9.9 (Sinai-Kolmogorov theorem). If (X, Σ, μ, T) is a measure-preserving system and ξ is a one-sided generator of Σ , then $h(T, \xi) = h(T)$.

Example 9.10. For (S^1, Σ, μ, T) an irrational rotation and ξ the above partition, ξ is a one-sided generator, so $h(T, \xi) = 0$.

Theorem 9.11 (martingale convergence theorem). Let $\{\xi_n\}$ be a sequence of σ -algebra such that $\xi_{n+1}(x) \subseteq \xi_n(x)$. Let \mathcal{A} the σ -algebra generated by $\bigcup_n \xi_n$. Then for any event A,

$$\lim_{n \to \infty} E(A|\xi_n) = E(A|\mathcal{A})$$

in L^2 .

We take this as a black box.

Proof of Sinai-Kolmogorov theorem. By the martingale convergence theorem,

$$h(T,\xi) = \lim_{n \to \infty} H(T,\xi_n)$$

if ξ_n refine to ξ . In fact, if ξ is a one-sided generator and we put

$$\xi_n = \bigvee_{j=1}^n T^{-j} \xi,$$

this means

$$h(T,\xi) = H(\xi|T^{-1}\Sigma).$$

Now take

$$\xi_n = \bigvee_{j=0}^n T^{-j} \xi,$$

which actually generates all of Σ . Thus for any partition η ,

$$\lim_{n \to \infty} H(\eta | \xi_n) = H(\eta | \Sigma) = 0$$

Also

$$h(T,\eta) \le h(T,\xi_n) + H(\eta|\xi_n)$$

so taking the limit,

$$h(T,\eta) \le h(T,\xi)$$

and taking the supremum of both sides over η we get the claim.

Corollary 9.12. Suppose that we are in the situation of the Sinai-Kolmogorov theorem but

$$\bigvee_{j=-n}^{n} T^{j} \xi$$

generates Σ , instead of $\bigvee_{j=0}^{n} T^{-j} \xi$. Then $h(T,\xi) = h(T)$.

Proof. The proof is the same.

Example 9.13. Kolmogorov formulated the definition of entropy to try to understand a question of von Neumann: which asked if there was a measure-theoretic sense in which rolling a die is more complex than flipping a coin. In fact their entropies are not the same, and so as measure-preserving systems they are not isomorphic.

To see this, take the one-sided Bernoulli shift T on k letters with probability vector \vec{p} . Let C_{σ} be the cylinder set of Cantor space k^{ω} generated by the string σ . Let ξ be the partition generated by all C_{σ} s where σ has length 1.

Then

$$h(T) = h(T,\xi) = \lim_{n \to \infty} \frac{1}{n} H(T^{-n+1}\xi) = \lim_{n \to \infty} \frac{1}{-n} \sum_{|\sigma|=n} p_{\sigma_0} \cdots p_{\sigma_n} \log p_{\sigma_0} \cdots p_{\sigma_n}$$

Thus

$$h(T) = -\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} p_{\sigma_j} \log p_{\sigma_j}.$$

Thus if we have uniform probability vectors $(1/n, \ldots, 1/n)$ of length n then $h(T) = \log n$.

Example 9.14. The entropy of the cat with eigenvalue λ is log λ . This is a tedious computation that we omit.

Example 9.15. Let $E_m(x) = x \mod 1$ be the circle multiplication map with winding number $m \in (0, 1)$. Let

$$\xi = \{(\ell/m, \ell + 1/m) : \ell\}$$

be a partition, then

$$T^{-j}\xi = \{(\ell/m^{j+1}, \ell+1/m^{j+1}):\ell\}$$

which clearly generates the Borel σ -algebra Σ as $n \to \infty$. Then

$$h(T) = h(T,\xi) = \lim_{n \to \infty} \frac{1}{n} H(T^{-n+1}\xi) = \lim_{n \to \infty} \frac{1}{n} \sum_{\ell=0}^{m^n-1} \frac{1}{m^n} = \log m.$$

9.4 Topological entropy

We now relate the entropy defined above (which we now call *metric entropy*) to *topological* entropy defined in the sense of metric spaces (confusingly enough.)

Fix a compact metrizable space X and a continuous transformation $T: X \to X$. Any compact, second-countable Hausdorff space will do, by the Urysohn metrization theorem.

Definition 9.16. Given an open cover \mathcal{U} of X, we define the *covering entropy*

 $H(\mathcal{U}) = \log\min_{\mathcal{V}} \operatorname{card} \mathcal{V}$

where \mathcal{V} ranges over all finite subcovers of \mathcal{U} .

So an extremely complicated cover with lots of barely overlapping parts has very high entropy.

Definition 9.17. The topological entropy of (X, T) is

$$h_{top}(T, \mathcal{U}) = \lim_{n \to \infty} \frac{1}{n} H(\bigvee_{j=0}^{n-1} T^{-j} \mathcal{U}).$$

We define

$$h_{top}(T) = \sup_{\mathcal{U}} h_{top}(T, \mathcal{U}).$$

The entropy $h_{top}(T)$ is supposed to measure the exponential growth of the orbits of T. In fact we can give counting laws for the number of periodic points in terms of h_{top} . But we will not do that today.

Henceforth let $\mathcal{M}(T)$ be the space of *T*-invariant Borel probability measures on *X*. We proved previously the Krylov-Bogolibabov theorem, which guarantees that $\mathcal{M}(T)$ is nonempty, convex, and compact, and its extreme points are the ergodic measures of *T*.

Theorem 9.18. We have

$$h_{top}(T) = \max_{\mu \in \mathcal{M}(T)} h(\mu).$$

Here the maximum is achieved because $\mathcal{M}(T)$ is compact.

Example 9.19. The Cantor space m^{ω} is metrizable as witnessed by

$$d(x,y) = \theta^{\min_{x_i \neq y_i} i}$$

where $\theta \in (0, 1)$ is fixed. This is actually an ultrametric; i.e. all triangles are isoceles.

The topological entropy of any shift is

$$h_{top}(T) = \log m,$$

and this is attained by the cover by open cylinders; any other cover would add inefficiency. We were interested in Bernoulli probability measures on m^{ω} and they had metric entropy $\log m$ if the probability was distributed correctly. So topological and metric entropies agree here.

Proof that $h \leq h_{top}$. Given a Borel partition ξ and a Borel measure μ we will find a cover \mathcal{U} such that

$$h_{\mu}(T,\xi) \le h_{top}(T,\mathcal{U}) + 1 + \log 2$$

and hence

$$h_{\mu}(T^{k},\xi) \leq h_{top}(T^{k},\mathcal{U}) + 1 + \log 2$$

which implies

$$h_{\mu}(T,\xi) \le h_{top}(T,\mathcal{U}) + \frac{1+\log 2}{k}$$

and taking the limit as $k \to \infty$ the claim follows.

Let $\xi = \{A_1, \ldots, A_k\}$ be a Borel partition. Since μ is Borel, there are C_j so that $\mu(A_j \setminus C_j) < 1/k$. This gives a new partition

$$\eta = \{C_1, \dots, C_k, X \setminus_{j=1}^k C_j\}$$

by closed sets, plus an open set, such that

$$h_{\mu}(T,\xi) \le h_{\mu}(T,\eta) + 1$$

Let

$$U_j = X \setminus \bigcup_{i \neq j} C_i;$$

then the U_j form an open cover \mathcal{U} of X.

Since ξ was a partition, for any U_j there is an $x \in U_j$ such that $x \notin U_i$ if $i \neq j$. In particular, \mathcal{U} is a *coarse cover* in the sense that it has no proper subcover, and

$$H(\mathcal{U}) = \log \operatorname{card} \mathcal{U} = \log k.$$

Moreover,

$$\operatorname{card} \bigvee_{j=0}^{n-1} T^{-j} \eta \leq 2^n \operatorname{card} \bigvee_{j=0}^{n-1} T^{-j} \mathcal{U}.$$

Here we are using the inequality

$$(k+1)^n \le 2^n k^n,$$

the fact that card $\eta = k + 1$, that \mathcal{U} is coarse, and that card $\mathcal{U} = k$.

Thus

$$H(\bigvee_{j=0}^{n-1} T^{-j}\eta) \le \exp H(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{U}) + \log 2,$$

 \mathbf{SO}

$$h_{\mu}(T,\xi) \le h_{\mu}(T,\eta) + 1 \le h_{top}(\mathcal{U}) + 1 + \log 2.$$

For the converse, we need to use the metric properties of X.

Definition 9.20. The Bowen distance of T is given by

$$d_n(x,y) = \max_j d(T^j(x), T^j(y)).$$

The Bowen ball is defined by

$$B(x,\varepsilon,n) = \{ y \in X : d_n(x,y) < \varepsilon \}.$$

If $d_n(x, y) < \varepsilon$, then until time *n*, the orbits of *x* and *y* must remain within ε of each other. For hyperbolic systems with positive Lyapunov exponent, this requires that either *n* be very small, or d(x, y) be very small.

Definition 9.21. We say that $A \subseteq X$ is (n, ε) -spanning if

$$\bigcup_{a \in A} B(a,\varepsilon,n) = X$$

The spanning number of (n, ε) is

$$\operatorname{Span}(n,\varepsilon) = \min_{A} \operatorname{card} A$$

where A ranges over (n, ε) -spanning sets.

Such sets must exist since X is compact.

Definition 9.22. The separating number of (n, ε) is

$$\operatorname{Sep}(n,\varepsilon) = \max_{B} \operatorname{card} B$$

where B ranges over sets such that whenever $x, y \in B$ either x = y or $d_n(x, y) \ge \varepsilon$. The covering number of (n, ε) is

$$\operatorname{Cov}(n,\varepsilon) = \min_{\mathcal{U}} \operatorname{card} \mathcal{U}$$

where \mathcal{U} is an open cover by sets whose Bowen *n*-diameter is at most ε .

Then

$$\operatorname{Cov}(n, 2\varepsilon) \leq \operatorname{Span}(n, \varepsilon) \leq \operatorname{Sep}(n, \varepsilon) \leq \operatorname{Cov}(n, \varepsilon).$$

Definition 9.23. The *Bowen entropy* is defined by

$$H_{\varepsilon}(T) = \lim_{n \to \infty} \frac{\log \operatorname{Cov}(n, \varepsilon)}{n}$$

Lemma 9.24. We have

$$h_{top}(T) = \sup_{\varepsilon > 0} H_{\varepsilon}(T).$$

Proof. Fix \mathcal{U} . Let δ be the Lebesgue number of \mathcal{U} . Then

$$\exp H(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{U}) \le \operatorname{Span}(n, \delta/2)$$

 \mathbf{SO}

$$h_{top}(T) \ge \sup_{\varepsilon > 0} H_{\varepsilon}(T).$$

Conversely, if the diameter of every set in \mathcal{U} is $< \varepsilon$ then

$$\operatorname{Sep}(n,\varepsilon) \le \exp H(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{U})$$

which proves the converse.
Proof that $h_{\top} \leq \sup_{\mu} h_{\mu}$. For every *n* choose $E_n \subset X$ to be (n, ε) -separated and maximal. Let ν_n be the uniform probability measure on E_n , which extends to an atomic measure on X (just put delta functions at each point in E_n).

Let

$$\mu_n = \frac{1}{n} \sum_{j=0}^{n-1} T^j_* \nu_n$$

where S_* denotes pushforward by S. Here the set $F_n = \bigcup_{j=0}^{n-1} T^j E_n$ is $(0, \varepsilon)$ -separated and μ_n is a uniform measure on F_n .

Then

$$H_{\varepsilon}(T) = \limsup_{n \to \infty} \frac{\log \operatorname{card} E_n}{n},$$

and we pass to a subsequence where this sequence converges. We then pass to a subsequence again, using compactness of $\mathcal{M}(T)$, to find a limit point μ of the μ_n .

We then construct a Borel partition $\xi = \{A_0, \ldots, A_{k-1}\}$ of X such that $\mu(A_j) < \varepsilon$ and $\mu(\partial A_j) = 0$. Then

$$\lim_{n \to \infty} H(\mu_n, \bigvee_{j=0}^{n-1} T^{-j}\xi) = H(\mu, \bigvee_{j=0}^{n-1} T^{-j}\xi),$$

as a computation will check, and so

$$\lim_{n \to \infty} \frac{1}{n} \log \operatorname{card} E_n = \lim_{n \to \infty} \frac{1}{n} H(\nu_n, \bigvee_{j=0}^{n-1} T^{-j} \xi) \le \frac{1}{q} H(\mu, \bigvee_{j=0}^{q-1} T^{-j} \xi)$$

and we now take $q \to \infty$.

Part III Operator algebras

Chapter 10

Banach algebras

Definition 10.1. A *Banach algebra* is a Banach space equipped with a bilinear, associative multiplication such that

$$||xy|| \le ||x||||y||.$$

If * is a linear involution on \mathcal{A} such that $(xy)^* = y^*x^*$ and $1^* = 1$ if \mathcal{A} is unital. then we say that \mathcal{A} is a *-algebra.

Definition 10.2. Let \mathcal{A} be a *-algebra. If one has the C*-*identity*

$$||x^*x|| = ||x||^2,$$

then we say that \mathcal{A} is a C^* -algebra.

For example, if \mathcal{H} is a Hilbert space, then $\mathcal{B}(\mathcal{H})$ is a C^* -algebra. Later we will learn that sub-*-algebras of $\mathcal{B}(\mathcal{H})$ are the only examples of C^* -algebras.

Often the norm topology is too strong, so we introduce a new topology which is weaker on $\mathcal{B}(\mathcal{H})$.

Definition 10.3. The strong operator topology is the locally convex topology on $\mathcal{B}(\mathcal{H})$ defined by the seminorms

$$P_{\xi}(T) = ||T\xi||.$$

In other words, a sequence converges in the strong operator topology $T_n \to T$ iff for each $\xi \in \mathcal{H}$, $||(T_n - T)\xi|| \to 0$. So the strong operator topology is the topology of pointwise convergence.

Definition 10.4. A von Neumann algebra \mathcal{A} is a sub-*-algebra of $\mathcal{B}(\mathcal{H})$ which is closed in the strong operator topology.

10.1 The spectrum

Fix a Banach algebra A.

Definition 10.5. Let $a \in A$. The spectrum $\sigma(a)$ is the set of $z \in \mathbb{C}$ such that $\sigma(a) - z$ is not invertible. The resolvent $\rho(a)$ is the complement of $\sigma(a)$.

Lemma 10.6. Let $a \in A$. If ||a|| < 1 then 1 - a is invertible with inverse

$$(1-a)^{-1} = \sum_{n=0}^{\infty} a^n.$$

Proof. The partial sums converge since ||a|| < 1. Therefore

$$(1-a)\sum_{n=0}^{\infty} a^n = (1-a)\lim_{n \to \infty} \sum_{k=0}^n a^k = \lim_{n \to \infty} \sum_{k=0}^n a^k - a^{k-1} = 1$$

since the summands telescope.

In particular, if ||1 - a|| < 1 then a is invertible.

Definition 10.7. The general linear group of A is GL(A), the group of invertible elements of A.

By the above lemma, there is a ball B around 1 contained in GL(A). By continuity of translation, we can carry B to be centered at any point of GL(A). Therefore GL(A) is an open set.

Proposition 10.8. The function $z \mapsto (z-a)^{-1}$ is holomorphic on $\rho(a) \cup \infty$.

In this case, holomorphy is indicated by local existence of a convergent power series.

Proof. We have

$$(a-z)^{-1} = \sum_{n=0}^{\infty} (a-z_0)^{-n-1} (z-z_0)^n$$

for each $z_0 \in \rho(a)$ and z close enough to z_0 that the power series converges. To see that the function is still holomorphic at ∞ , notice that

$$(a - z^{-1})^{-1} = z(1 - az)^{-1}$$

which vanishes as $z \to 0$. Replacing z by z^{-1} , we see that the function is bounded close to infinity, and continuous, so holomorphic there.

We now observe that the usual proofs of Cauchy's integral formula and its friends such as Cauchy's estimate and Liouville's theorem go through even in case of holomorphic functions $U \to A, U \subseteq \mathbb{C}$ open.

We now come to the famous Gelfand-Mazur theorem, which can be thought of as a "restatement of the fundamental theorem of algebra" for our purposes. For the notation, recall that the map $z \mapsto z1$ is an embedding of \mathbb{C} in any Banach algebra.

Theorem 10.9 (Gelfand-Mazur). If $A = GL(A) \cup 0$, then $A = \mathbb{C}$.

Proof. Let $a \in GL(A)$ and assume towards contradiction that $a \notin \mathbb{C}$. Then the resolvent $z \mapsto (a-z)^{-1}$ is a holomorphic function defined on the Riemann sphere, so constant. Taking $z = \infty$, the resolvent is identically 0, but also identically a^{-1} (taking z = 0). This is a contradiction.

Notice that this fails over \mathbb{R} , as witnessed by \mathbb{C} as well as the quaternions \mathbb{H} . This is why we study Banach algebras over \mathbb{C} .

10.2 Ideals

Let I be an ideal of A. It is immediate that the norm-closure \overline{I} is an ideal. Moreover, since $GL(A) \ni 1$ is open, if I is a proper ideal, then I does not meet GL(A) and so \overline{I} does not contain 1, so $\overline{\cdot}$ preserves propriety. Therefore maximal ideals are closed. Moreover, for continuous morphisms, kernels are closed, so we might as well only study closed ideals.

If I is a (left, right) ideal then A/I is a (left, right) module over A, equipped with the seminorm

$$||a|| = \inf_{d \in I} ||a - d||.$$

In case I is closed, this seminorm is actually a norm, and complete since A is complete. So we end up with a Banach space.

Definition 10.10. A *Banach module* over A is an A-module M which is a Banach space, such that

$$||am||_M \le ||a||_A ||m||_M.$$

It is not very hard to check that M = A/I is a Banach module. In fact, for $b, c \in I$, we have

$$||am||_M \le ||(a-c)(m-d)||_A \le ||a-c||_A ||m-d||_A.$$

Taking the inf over c, d of both sides, we have

$$||am||_M \le ||a||_A ||m||_M.$$

In case I is two-sided, M is a Banach (A, A)-bimodule, or in other words, a Banach algebra.

In what follows we use Hom(A, B) to mean the K-algebra of morphisms of K-algebras $A \to B$ over some field K (which is usually \mathbb{C}).

If I is a maximal ideal, therefore, A/I is a field, and so $A/I = \mathbb{C}$ by the Gelfand-Mazur theorem. But a maximal ideal gives a epimorphism $A \to \mathbb{C}$, and conversely, the kernel of a such an epimorphism is a maximal ideal. This gives a bijection between the maximal spectrum of A and Hom $(A, \mathbb{C}) \setminus 0$, which we call \hat{A} .

Lemma 10.11. Let K be a field and A a unital K-algebra. Let $\varphi \in \text{Hom}(A, K)$. Then if $a \in A, \varphi(a) \in \sigma(a)$.

Proof. We have
$$\varphi(a - \varphi(a)) = 0.$$

Lemma 10.12. If $\varphi : A \to \mathbb{C}$ is a nonzero morphism, then $||\varphi|| \leq 1$.

Proof.
$$\varphi(a) \in \sigma(a)$$
 so $||\varphi(a)|| \le ||a||$.

Therefore \hat{A} is contained in the unit ball A'_1 of the dual A'. Since nets in \hat{A} act continuously on A, their pointwise convergence preserves operations of A. So \hat{A} is closed. In particular, the Banach-Alaoglu theorem implies that \hat{A} is a weakstar compact Hausdorff space.

Definition 10.13. Let $a \in A$. The *Gelfand transform* \hat{a} is the function

$$\hat{a}(\varphi) = \varphi(a),$$

for $\varphi \in \hat{A}$.

Notice that $||\hat{a}||_{L^{\infty}(\hat{A})} \leq ||a||_A$ and $\hat{a}(\hat{A}) \subseteq \sigma_A(a)$. Conversely, let $\lambda \in \sigma_A(a)$. Then $a - \lambda$ is not invertible, so there is a maximal ideal $I \supseteq (a - \lambda)$ and an epimorphism φ such that ker $\varphi = I$. Thus $\lambda \in \hat{a}(\hat{A})$. Therefore $\hat{A} = \sigma_A(a)$, but the proof of this is highly nonconstructive.

Example 10.14. Recall that $c_0(\mathbb{N})$, the set of $x \in \ell^{\infty}(\mathbb{N})$ such that $x_n \to 0$ as $n \to \infty$, is a closed ideal of $A = \ell^{\infty}(\mathbb{N})$. Therefore there is a φ such that $\varphi(c_0(\mathbb{N})) = 0$. But, in fact, $\hat{A} = \beta \mathbb{N}$, where β is the Stone-Cech functor. It follows that it is consistent with ZF without the axiom of choice that φ does not exist.

Definition 10.15. Let $a \in A$. The spectral radius of a is

$$r(a) = \max_{\lambda \in \sigma(a)} |\lambda|$$

Equivalently, $r(a) = ||\hat{a}||_{L^{\infty}(\hat{A})}$. Therefore we have $r(ab) \leq r(a)r(b)$.

10.3 The holomorphic functional calculus

As usual, let A be a commutative Banach algebra.

Definition 10.16. Let $a \in A$ and let f be a holomorphic function on $D(0, ||a|| + \varepsilon)$. Put

$$f(z) = \sum_{n=0}^{\infty} \alpha_n z^n.$$

The holomorphic functional calculus is the morphism $f \mapsto f(a)$ defined by

$$f(a) = \sum_{n=0}^{\infty} \alpha_n a^n.$$

The Taylor series of f converges uniformly absolutely on D(0, ||a||), so the partial sums of f(a) form a Cauchy sequence in A. Therefore f(a) is a well-defined element of A, and we can think of f as a mapping $U \to A$, where U consists of elements of A that are small enough. If f is entire, then f lifts to a function $A \to A$.

Theorem 10.17 (spectral mapping theorem). If $\lambda \in \sigma(a)$ then $f(\lambda) \in \sigma(f(a))$.

Proof. We have

$$f(a) - f(\lambda) = \sum_{n=0}^{\infty} \alpha_n (a^n - \lambda^n) = \sum_{n=0}^{\infty} \alpha_n (a - \lambda) (a^{n-1} + a^{n-2}\lambda + \dots + \lambda^{n-1})$$
$$= (a - \lambda)b$$

for some b, if we can show that the partial sums are a Cauchy sequence. In fact

$$||a^{n-1} + \dots + \lambda^{n-1}|| \le n||a||^{n-1}$$

which is the right-hand side of f'(||a||) (which clearly converges, so partial sums are Cauchy). Therefore $f(a) - f(\lambda) = (a - \lambda)b$. So if $f(a) - f(\lambda)$ is invertible, then so is $a - \lambda$.

Chapter 11

C^* -algebras

11.1 Weights

Definition 11.1. Let \mathcal{A} be a sub-*-algebra of $\mathcal{B}(\mathcal{H})$. A map $\omega : \mathcal{A}^+ \to [0, \infty]$ is a *weight* if ω is additive and if $\omega(ta) = t\omega(a)$ whenever $t \ge 0$.

Fix a weight ω . By m_{ω} we mean the span of the set of positive a such that $\omega(a) < \infty$, and by m_{ω}^{sa} we mean the closure of the set of positive a such that $\omega(a) < \infty$ under subtraction. Clearly ω extends uniquely to m_{ω}^{sa} by $\omega(b-c) = \omega(b) - \omega(c)$. So ω extends to a positive linear functional on m_{ω} in the obvious way. On the other hand, if φ is any positive linear functional on $\mathcal{B}(\mathcal{H})$, then φ is a weight such that $m_{\varphi} = \mathcal{B}(\mathcal{H})$.

Now we define n_{ω} to be the set of $a \in \mathcal{A}$ such that $\omega(a^*a) < \infty$, which is clearly a subspace of \mathcal{A} .

Lemma 11.2. n_{ω} is a left ideal of \mathcal{A} .

Proof. If $T \ge 0$ then

$$\langle S^*TS\xi, \xi \rangle = \langle TS\xi, S\xi \rangle \le ||T|| ||S\xi||^2 = ||T|| \langle S\xi, S\xi \rangle = ||T|| \langle S^*S\xi, \xi \rangle.$$

So if $d \in \mathcal{A}$ and $a \in n_{\omega}$ then

$$(da)^*da = a^*d^*da \le ||d^*d||a^*a = ||d||^2a^*a$$

whence

$$||\omega((da)^*(da)) \le ||d||^2 \omega(a^*a) < \infty.$$

Definition 11.3. If $\omega(a^*a) = \omega(aa^*)$ then ω is tracial.

Clearly if ω is tracial then n_{ω} is a two-sided ideal. For example, if ω is actually the trace,

$$\omega(x) = \sum_{j} \langle x^* x e_j, e_j \rangle$$

for $\{e_j\}_j$ an orthonormal basis of the separable Hilbert space \mathcal{H} , then ω is tracial and n_{ω} is just the space of trace-class operators and ω is tracial.

Recall the polarization identity:

$$4b^*a = \sum_{k=0}^{3} i^k (a+i^k b)^* (a+i^k b).$$

From this we are justified in defining, on n_{ω} ,

$$\langle a, b \rangle_{\omega} = \omega(b^*a).$$

This would be an inner product if $N_{\omega} = \{a \in \mathcal{A} : \omega(a^*a) = 0\}$ were trivial. Clearly N_{ω} is a subspace, so we can take the completion of n_{ω}/N_{ω} and recover a Hilbert space.

11.2 The GNS construction

Definition 11.4. The completion of n_{ω}/N_{ω} is denoted $L^2(\mathcal{A}, \omega)$.

If $b \in n_{\omega}$, then

$$\langle ab, ab \rangle_{\omega} = \omega(b^*a^*ab) \le ||a^*a||\omega(b^*b) = ||a||^2 ||b||_{\omega}^2.$$

So if $a \in \mathcal{A}$ then $\xi \mapsto a\xi$ is a well-defined, bounded operator on n_{ω}/N_{ω} and so extends to $L^{2}(\omega)$.

Definition 11.5. A *-representation is a morphism of *-algebras (i.e. a morphism of algebras preserving *) into $\mathcal{B}(\mathcal{H})$.

If $a \in \mathcal{A}, b, c \in n_{\omega}$, then

$$\langle ab, c \rangle_{\omega} = \omega(c^*ab) = \omega((a^*c)^*b) = \langle b, a^*c \rangle_{\omega},$$

which descends to $L^2(\omega)$. So we can define a *-representation

$$L: \mathcal{A} \to \mathcal{B}(L^2(\omega))$$
$$a \mapsto (\xi \mapsto a\xi).$$

Definition 11.6. The map L is called the *GNS construction* (for Gelfand-Neimark-Segal) of \mathcal{A} , or the *left regular representation* of \mathcal{A} .

The GNS construction allows us to assume that \mathcal{A} is actually acting on a Hilbert space, namely $L^2(\omega)$. So a C^* -algebra is always an operator algebra.

We can also define a right regular representation,

$$R: \mathcal{A} \to \mathcal{B}(L^2(\omega))$$
$$a \mapsto (\xi \mapsto \xi a).$$

Notice that R is an antihomomorphism.

Lemma 11.7. Assume that for each positive $a \in \mathcal{A}$, \sqrt{a} exists. Then $m_{\omega} \subseteq n_{\omega}$, and in particular m_{ω} is a sub-*-algebra of \mathcal{A} .

Proof. If $a \in m_{\omega}$ is positive,

$$\omega(\sqrt{a}^2) = \omega(\sqrt{a}^*\sqrt{a}) = \omega(a)$$

so $\sqrt{a} \in n_{\omega}$. Since n_{ω} is a left ideal, $a \in n_{\omega}$.

Example 11.8. Let X be a measure space and $K \in L^2(X \times X)$. Then the integral operator $T_K : L^2(X) \to L^2(X)$ has $||T_K||_{\mathcal{B}^2} = ||K||_{L^2}$. Indeed, if $\{\xi_n\}_n$ is a Hilbert basis for $L^2(X)$ then

$$\sum_{n} ||T_k \xi_n||^2 = \sum_{m,n} |\langle T_k \xi_m, \xi_n \rangle|^2 = \sum_{m,n} \left| \iint_{X \times X} K(x, y) \xi_n(y) \xi_m(x) \, dx \, dy \right|^2$$
$$= \sum_{m,n} |\langle K, \xi_m \otimes \xi_n \rangle|^2 = ||K||_{L^2}$$

since the $\xi_m \otimes \xi_n$ form a Hilbert basis for $L^2(X \times X) = L^2(X) \otimes L^2(X)$.

Example 11.9. If A = C([0, 1]) and

$$\omega(f) = \int_0^1 f(t) \, dt$$

then ω is a tracial weight on \mathcal{A} such that $n_{\omega} = \mathcal{A}$. But of course n_{ω} is a Banach space when given the $\mathcal{B}^2 = L^2$ norm. Its completion is $L^2([0, 1])$.

Proposition 11.10. Let ω be a tracial weight and \mathcal{A} be a sub-*-algebra of $\mathcal{B}_0(\mathcal{H})$. If $b \geq 0$ and $b \in m_{\omega}$, and $a \in \mathcal{B}(\mathcal{H})$ then

$$|\omega(ab)| \le ||a||_{op} |\omega(b)|.$$

Proof. We have

$$\begin{split} |\omega(ab)|^2 &= |\omega(a\sqrt{b}\sqrt{b})|^2 = |\omega(\sqrt{b}a\sqrt{b})|^2 = |\langle a\sqrt{b},\sqrt{b}\rangle|^2 \\ &\leq \langle a\sqrt{b},a\sqrt{b}\rangle\langle\sqrt{b},\sqrt{b}\rangle = \omega(\sqrt{b}a^*a\sqrt{b})\omega(b) \leq ||a||^2\omega(b)^2. \end{split}$$

We summarize the GNS construction, and the Gelfand transform, in the following theorem. We define $C_{\infty}(X)$ to be a subspace of the space of continuous functions C(X). If X is compact we define $C_{\infty}(X) = C(X)$. Otherwise, we let $C_{\infty}(X)$ be those functions in C(X) which vanish at the point at infinity given by the one-point compactification of X. For example, $C_{\infty}(\mathbb{R})$ consists of functions on \mathbb{R} which go to zero as $|x| \to \infty$.

Theorem 11.11 (Gelfand-Naimark). For every C^* -algebra A, there is a faithful *-representation of A. Moreover, if A is commutative, then there is a locally compact Hausdorff space X such that $A = C_{\infty}(X)$ (which gives a representation of A on $L^2(X)$). If A is also unital, then Xis compact and is naturally in bijection with the maximal ideal space of A. Moreover, the map $A \mapsto X$ is a contravariant equivalence of categories between compact Hausdorff spaces and commutative, unital C^* -algebras.

11.3 The \mathcal{B}^p spaces

Definition 11.12. If $T \in \mathcal{B}^2(\mathcal{H})$ then T is called a *Hilbert-Schmidt operator*.

Example 11.13. Let \mathcal{H} be the separable Hilbert space. Take $\mathcal{A} = \mathcal{B}_0(\mathcal{H})$ and ω to be the trace. Since $\mathcal{B}_0(\mathcal{H})$ has square roots and ω is tracial, we can apply the above result to prove that n_{ω} and m_{ω} are two-sided ideals and hence sub-*-algebras.

If we write $|T| = \sqrt{T}^2$, and let $\mathcal{B}^p(\mathcal{H})$ be the space of $T \in \mathcal{B}_0(\mathcal{H})$ such that $\omega(|T|^p) < \infty$, then $\mathcal{B}^1(\mathcal{H}) = m_\omega(\mathcal{H})$ and $\mathcal{B}^2(\mathcal{H}) = n_\omega(\mathcal{H})$.

We think of $\mathcal{B}^p(\mathcal{H})$ as the noncommutative analogue of ℓ^p .

Let's check that that example actually makes sense.

Theorem 11.14. $\mathcal{B}^2(\mathcal{H})$ is a Banach space.

Proof. First observe that $||T||_{op} \leq ||T||_2$. To do this, compute the trace of T by using an orthonormal basis containing a ξ such that $||T\xi|| \geq ||T||_{op} - \varepsilon$. As this is possible for any $\varepsilon > 0$ the claim holds.

Now assume that $\{T_n\}_n$ is 2-Cauchy, so in particular *op*-Cauchy. So there is a $T \in \mathcal{B}_0(\mathcal{H})$ such that $T_n \to^{op} T$.

If P is a finite-rank projection then $(T - T_n)P$ is a finite-rank operator, hence $\in \mathcal{B}^2(\mathcal{H})$. So

$$||(T - T_n)P||_2^2 = trP(T - T_n)^*(T - T_n)P = tr(T - T_n)P(T - T_n)^* = \lim_{k \to \infty} tr(T_k - T_n)P(T_k - T_n)^*$$

$$\leq \limsup_{k \to \infty} (T_k - T_n)(T_k - T_n)^* = \limsup_{k \to \infty} ||T_k - T_n||_2^2.$$

Let $C_n = \limsup_{k \to \infty} ||T_k - T_n||_2^2$. Then $C_n \to 0$ and

$$||(T - T_n)P||_2^2 \le C_n$$

regardless of the choice of n and P. Since $T - T_n$ is a compact operator, we can approximate it arbitrarily well by $(T - T_n)P$ by choosing P. So $||T - T_n||_2^2 \to 0$.

Recall the *polar decomposition* of $T \in \mathcal{B}_0(\mathcal{H})$ is the factorization

$$T = V|T|$$

where $|T| = \sqrt{T}^2$ and V is a *partial isometry*, i.e. an isometry on its cokernel.

Lemma 11.15. If $T \in \mathcal{B}^1(\mathcal{H})$ and $A \in \mathcal{B}(\mathcal{H})$ then

$$|tr(AT)| \le ||A||_{op} tr|T|.$$

Proof. Write T = V|T|. Then

$$|tr(AT)| = |tr(AV|T|)| \le ||AV||_{op}tr|T| \le ||A||_{op}tr|T|.$$

г		٦
L		

Lemma 11.16. $||\cdot||_1 = tr|\cdot|$ is a norm on $\mathcal{B}^1(\mathcal{H})$.

Proof. Let
$$S, T \in \mathcal{B}^1(\mathcal{H})$$
 and $S + T = W|S + T|$. Then
 $tr|S + T| = trW^*(S + T) = tr(W^*S) + tr(W^*T) \le |trW^*S| + |trW^*T| \le tr|S| + tr|T|$. □

Theorem 11.17. $\mathcal{B}^1(\mathcal{H})$ is a Banach algebra.

Proof. Since $||T||_{op} \leq ||T||_1$ the proof is basically the same as for Hilbert-Schmidt operators.

Theorem 11.18. $\mathcal{B}^1(\mathcal{H})^* = \mathcal{B}(H).$

Proof. If $A \in \mathcal{B}(\mathcal{H})$, let $\Psi_A(T) = tr(AT)$. Then

$$|\Psi_A(T)|| \le ||A||||T||_1.$$

So $A \mapsto \Psi_A$ is an isometry and so $\mathcal{B}(\mathcal{H}) \subseteq \mathcal{B}^1(\mathcal{H})^*$.

Let $\Psi \in \mathcal{B}^1(\mathcal{H})^*$ and $\xi, \eta \in \mathcal{H}$. Define a bounded operator $\langle \xi, \eta \rangle_O$ by

$$\langle \xi, \eta \rangle_O \zeta = \xi \langle \eta, \zeta \rangle.$$

(So $\langle \cdot, \cdot \rangle_O$ is an operator-valued pseudo-inner product (the pseudo- here means that it could be zero). Define a semilinear form

$$B_{\Psi}(xi,\eta) = \Psi\langle\xi,\eta\rangle_O.$$

So $|B_{\Psi}(\xi,\eta)| \leq ||\Psi|| ||\xi|| ||\eta||$. Therefore by the Riesz representation theorem, there is an operator A such that $B_{\Psi}(\xi,\eta) = \langle A\xi,\eta \rangle$. Therefore $||A|| = ||\Psi||$ and $\Psi = \Psi_A$. So $\mathcal{B}^1(\mathcal{H})^* \subseteq \mathcal{B}(\mathcal{H})$.

11.4 Representation theory of groups

Let G be a group with a good topology (so G admits a Haar measure).

Definition 11.19. A unitary representation of G is a continuous morphism of groups $G \rightarrow U(H)$. It is *irreducible* if the only G-invariant subspaces are trivial.

For π a unitary representation, we have $\pi(x)^* = \pi(x)^{-1}$.

Example 11.20. Let $G = SL(3, \mathbb{Z})$. Then the "obvious" map $G \to SL(3, \mathbb{C})$ is not a unitary representation. In fact G has very few finite-dimensional unitary representations, because G is not compact.

Definition 11.21. The *left regular representation* of G is the map $G \to U(L^2(G))$ given by

$$\pi(x)(\xi)(y) = \xi(x^{-1}y).$$

It is natural to want to study the subalgebra of B(H) generated by $\pi(G)$ for π a unitary representation. This will be given by linear combinations of the $\pi(x)$ s as $x \in G$, which we identify with the space $C_c(G)$ of compactly supported continuous functions on G. Namely, for $f \in G$ we define

$$\pi(f) = \int_G f(x)\pi(x) \ dx$$

Definition 11.22. The norm-closure of $\pi(C_c(G))$ is the reduced C^{*}-algebra of G.

Now an easy computation shows

$$\pi(f)\pi(g) = \pi(f * g)$$

and of course $\pi(f)^* = \pi(f^*)$ where we define $f^*(x) = \overline{f}(x^{-1})$. Finally, we observe that

 $||\pi(f)|| \le ||f||_{L^1(G)}$

so π is a *-Banach algebra morphism which extends to a map

$$\pi: L^1(G) \to B(L^2(G)).$$

This leads to the abstract theory of Fourier transform.

11.5 Compact operators

Let $B_0(H)$ denote the algebra of compact operators in H. This is a closed ideal of H, hence a C^* algebra (proof: it is the closure of the ideal $B_f(H)$ of finite rank operators in H.) It will be one of our main examples of a noncommutative, nonunital C^* algebra.

We now study the representation theory of $B_0(H)$.

Like any C^* algebra, $B_0(H)$ has a normalized approximate identity, sequential if H is separable. Decompose H by transfinite recursion as

$$H = \bigoplus_{\alpha < \kappa} \mathbb{C}$$

where κ is some cardinal ($\kappa = \aleph_0$ if H separable) and the biproduct is in the category of Hilbert spaces. For $\lambda < \kappa$, let $H_{\lambda} = \bigoplus_{\alpha < \lambda} H_{\lambda}$ so H is the injective limit of the H_{λ} ; then let e_{λ} be the natural projection $H \to H_{\lambda}$. The e_{λ} form a net with respect to the natural ordering on κ and are obviously an approximate identity.

Recall that if we fix a representation $\pi : A \to B(H)$, we can view H as a module over A by defining $a\xi = \pi(a)(\xi)$. Recall also that a representation is said to be nondegenerate if HA is dense in A.

In fact, any representation of $B_0(H)$ is faithful. Since representations are continuous, and $B_0(H)$ has no closed ideals (since $B_f(H)$ contains all proper ideals of $B_0(H)$, and is dense in H), any representation of $B_0(H)$ is faithful.

There is a natural *-representation $B_0(H) \to B(H)$ given by the inclusion map. Since $e_\lambda \xi \to \xi$, this representation is nondegenerate. In some sense this is the only such representation.

Lemma 11.23. A nondegenerate *-representation of $B_0(H)$ is isomorphic to a direct sum to copies of the representation $B_0(H) \to B(H)$. In particular, the only irreducible such representation is the representation $B_0(H) \to B(H)$.

Proof. Let $\langle \xi, \eta \rangle_0$ be the $B_0(H)$ -valued inner product

$$\langle \xi, \eta \rangle_0 \zeta = \xi \langle \eta, \zeta \rangle.$$

In fact such an inner product has values in rank-1 operators since $\xi \langle \eta, \zeta \rangle$ lies in the span of ξ .

For $T \in B(H)$, $T\langle \xi, \eta \rangle_0 \zeta = (T\xi)\langle \eta, \zeta \rangle$ so $T\langle \xi, \eta \rangle_0 = \langle T\xi, \eta \rangle_0$, and $\langle \xi, \eta \rangle_0 T = \langle \xi, T^*\eta \rangle_0$.

Let $\pi : B_0(H) \to B(V)$ be a nondegenerate *-representation, $\xi \in H$ a unit vector. Then $\langle \xi, \xi \rangle_0$ is a rank-1 projection. Since π is faithful, $\pi(\langle \xi, \xi \rangle_0)$ is a nonzero projection. Let v be a unit vector of $\langle \xi, \xi \rangle_0(V)$ and define $Q : H \to V$ by $Q\eta = \langle \eta, \xi \rangle_0 v$. Then by a tedious computation, Q is an isometry.

We now show that Q commutes the representations. Let $T = \langle \omega, \zeta \rangle_0$. Any operator in $B_0(H)$ can be written as an infinite linear combination of rank-1 operators so it suffices to show that QT = TQ. In fact,

$$Q(T\eta) = \langle T\eta, \zeta \rangle_0 v = TQ(\eta).$$

Also, $Q(H)^{\perp}$ is π -invariant, so we repeat the argument on $Q(H)^{\perp}$ to see that we have

$$V = Q(H) \oplus Q(H)^{\perp}$$

as $B_0(H)$ -modules. Now run Zorn's lemma to keep decomposing $Q(H)^{\perp}$ until we hit an irreducible representation.

This is a very remarkable property of $B_0(H)$. To see why, we need something stronger than ZFC.

Definition 11.24. A \Diamond -sequence is a net of sets $\alpha \mapsto A_{\alpha}$, for $\alpha < \aleph_1$, such that for any $A \subseteq \aleph_1$,

$$\hat{A} = \{ \alpha < \aleph_1 : A \cap \alpha = A_\alpha \}$$

is stationary in \aleph_1 .

In other words, for every closed and unbounded ("club") set $C \subseteq \aleph_1, C \cap \hat{A}$ is nonempty. The existence of a \diamond -sequence implies that V = L, in particular implying GCH.

Naimark conjectured that if A was a C^{*}-algebra with only one irreducible representation, then $A = B_0(H)$. This is true if A is separable.

Theorem 11.25 (Ackemann-Weaver). If there is a \Diamond -sequence, then there is a C^* -algebra A which has only one irreducible representation such that $A \neq B_0(H)$.

In this lemma, you should read A = B(H) and $I = B_0(H)$.

Lemma 11.26. Let A be a *-normed algebra and I a *-ideal of A with normalized approximate unit. Then every nondegenerate *-representation of I extends uniquely to A.

Proof. Let $\pi: I \to B(H)$ be such a representation. Define

$$\tilde{\pi}(a)\sum_{\alpha}\pi(d_{\alpha})\xi_{\alpha}=\sum_{\alpha}\pi(ad_{\alpha})\xi_{\alpha}$$

where the ξ_{α} are a Hilbert basis of H. Then $\tilde{\pi}$ is a well-defined function since if $\sum_{\alpha} \pi(d_{\alpha})\xi_{\alpha} = 0$, then

$$\sum_{\alpha} \pi(ad_{\alpha})\xi_{\alpha} = \lim_{\lambda} \sum_{\alpha} \pi(ae_{\lambda}d_{\alpha})\xi_{\alpha} = \lim_{\lambda} \pi(ae_{\lambda}) \sum_{\alpha} \pi(d_{\alpha})\xi_{\alpha} = 0$$

since the e_{λ} are a normalized approximate unit. This is unique because

$$\tilde{\pi}(a)\pi(d)\xi = \pi(ad)\xi$$

and the $\pi(ad)\xi$ are dense in H.

Lemma 11.27. Let A, I be as above and let π be an irreducible representation of A. Then either $I \subseteq \ker \pi$ or π is an irreducible representation of I.

Proof. Assume I is not contained in ker π . Then \overline{IH} is nonzero and A-invariant. Since π is irreducible, $\overline{IH} = H$. Therefore π is a nondegenerate representation of I.

We have $e_{\lambda}\xi \to \xi$ for any ξ since π is nondegenerate. Let $K \subseteq H$ be nondegenerate and nonzero. Using e_{λ} , $\overline{IK} = K$ so \overline{IK} is A-invariant. Therefore since π is irreducible, K = H.

Lemma 11.28. Let A, I be as above. Let $\pi : A \to B(H)$ and $\rho : A \to B(K)$ be irreducible representations. If $\pi \cong \rho$ as representations of I, then $\pi \cong \rho$ as representations of A.

Proof. Let $U: H \to K$ be an isomorphism of *I*-modules. For $d \in I$,

$$U(\pi(a)\pi(d)\xi) = U(\pi(ad)\xi) = \rho(ad)U\xi = \rho(a)U(\pi(d)\xi).$$

So $U\pi = \rho U$.

Theorem 11.29 (Burnside). Assume $H \neq \mathbb{C}$. Let $A \subseteq B_0(H)$ be a C^* -algebra. If A acts on H irreducibly, then $A = B_0(H)$.

Proof. By assumption on $H, A \neq 0$. Let $T \neq 0$. Then $T^*T \in A$ is nonzero, so we can assume without loss of generality that T is self-adjoint. Moreover, $C^*(T, 1) = C(\sigma(T))$ acts on H as an abelian monoid.

Let $\lambda \in C(\sigma(T))$. Taking bump functions centered on λ we can find a sequence of $\xi_n \in H$ such that $(T - \lambda)\xi_n \to 0$. Since T is comapct, T sends the unit ball of H to a precompact set. So the $T\xi_n$ have a weak limit η . Thus $T\eta = \lambda \eta$. Therefore T has an eigenvector for λ . So the only limit point of $\sigma(T)$ is 0, because the other eigenvectors are all orthogonal.

Let *P* be a projection of minimal rank in *A*. We claim that *P* is a rank-1 projection. In fact, *PTP* is a self-adjoint operator on the finite-dimensional space *PH*, so has spectral projections in *A* whose rank is the same as that of *P* by minimality. Thus there is a unique such spectral projection; i.e. there is an $s \in \mathbb{R}$ such that PTP = sP. Moreover, if $\xi, \eta \in PH$ are orthonormal, $R \in A$, then $\langle R\xi, \eta \rangle = s \langle \xi, \eta \rangle = 0$. Therefore $\langle A\xi, \eta \rangle = 0$, yet *A* acts irreducibly, which is a contradiction. Therefore *P* is a rank-1 projection.

Now $R, S \in A$ implies that RPS is a rank-1 projection. Thus $\overline{APAH} = H$. So \overline{APA} is the set of rank-1 operators. Any compact operator can be written as an infinite linear combination of those.

Corollary 11.30. Let A be a C^{*}-algebra and $\pi : A \to B(H)$ an irreducible representation. If $\pi(A)$ contains a nonzero compact operator then $\pi(A)$ contains $B_0(H)$.

<u>Proof.</u> Let $I = \pi(A) \cap B_0(H)$. Then I acts irreducibly on H so by Burnside's theorem, $\overline{\pi(I)} = B_0(H)$. But π takes $I/\ker \pi$ to $B_0(H)$ isometrically (since it is injective), so $\pi(I)$ is closed since I is complete. Thus $\pi(I) = B_0(H)$.

We write A to denote the set of all isomorphism classes of irreducible representations of A.

Definition 11.31. Let $\pi : A \to B(H)$ range over \hat{A} . We say that A is *liminal* or *CCR* if for every π , $\pi(A) = B_0(H)$. We say A is *postliminal* or *GCR* if for every π , $B_0(H) \cap \pi(A) \neq 0$. We say A is *antiliminal* or *NCR* if for every π , $B_0(H) \cap \pi(A) = 0$.

If A is a von Neumann algebra, we say that A is type-I if A is postliminal. We say that A is non-type-I if A is not postliminal.

We will prove later that if G is a semisimple Lie group or a nilpotent Lie group, then $C^*(G)$ (by which we really mean $C^*(L^1(G))$) is CCR. But if G is solvable we cannot even prove that $C^*(G)$ is GCR.

Example 11.32. Let α be the Lie action of \mathbb{R} on \mathbb{C}^2 by $\alpha(t)(z, w) = (e^{2\pi i t} z, e^{2\pi i \mu t} w)$ where μ is irrational. Then we take the outer semidirect product $G = \mathbb{C}^2 \times_{\alpha} \mathbb{R}$. Then $C^*(G)$ is not GCR.

Theorem 11.33. Let $\pi : A \to B(H)$ be an irreducible representation and let $I = \ker \pi$. If $\pi(A) \cap B_0(H) \neq 0$ then for every irreducible representation ρ of A such that ker $\rho = I$, $\rho \cong \pi$.

Proof. Let $J = \pi^{-1}(B_0(H))$, so $I \subseteq J$. Then π is an irreducible representation of J with kernel I, so drops to an irreducible representation of J/I. By Burnside's theorem, π is an isomorphism $J/I \to B_0(H)$.

But by assumption on ρ , ρ is an isomorphism $J/I \cong B_0(H)$. Since there can be only one irreducible representation of $B_0(H)$, $\rho \cong \pi$ as representations of J. But ρ, π extend uniquely to A, so $\rho \cong \pi$ as representations of A.

Corollary 11.34. If A is a GCR algebra, then every representation of A is uniquely determined by its kernel.

Definition 11.35. A *primitive ideal* is a kernel of some irreducible representation.

So if A is GCR, then there is a bijection between \hat{A} and the set of primitive ideals of A. On the other hand, NCR algebras are very bad:

Theorem 11.36. Let $\pi : A \to B(H)$ be an irreducible representation and assume $\pi(A) \cap B_0(H) = 0$. Then there are uncountably many irreducible representations ρ of A such that $\ker \rho = \ker \pi$.

In fact Mackey showed that in some sense the set of representations sharing a kernel with π is "unclassifiable."

Example 11.37. Let A be a unital, infinite-dimensional, simple C^* -algebra and $\pi : A \to B(H)$ an irreducible representation. Since A is simple, it has no proper ideals; yet $\pi^{-1}(B_0(H))$ is an ideal. If $\pi^{-1}(B_0(H)) = A$ then A is not unital, a contradiction. So $\pi(A) \cap B_0(H) = 0$. Therefore A is an NCR algebra.

Theorem 11.38. Every primitive ideal is prime.

Proof. Let I be a primitive ideal of A, say $I = \ker \pi$. Let J_1, J_2 be ideals of \underline{A} such that $J_1J_2 \subseteq I$. If $J_1 \subseteq I$ there is nothing to prove. Otherwise, $\pi(J_1) \neq 0$, so $\overline{\pi(J_1)H} \neq 0$ is π -invariant. By irreducibility, $\overline{\pi(J_1)H} = H$, so $\overline{\pi(J_2)H} = \overline{\pi(J_2)\pi(J_1)H} = \overline{\pi(J_1J_2)H} \subseteq \overline{\pi(I)H} = 0$. So $J_2 \subseteq I$.

Recall that the prime spectrum Spec R of a ring R is defined by the Zariski (or Jacobson, or hull-kernel) topology is given by declaring that $S \subset \text{Spec } R$ is closed if there is an ideal Isuch that $S = \{J \in \text{Spec } R : J \subseteq I\}$. By the above theorem, if R is a C^* -algebra, then the Zariski topology drops to a topology on the set of primitive ideals Prim R. This topology is far from Hausdorff in general, but is at least locally compact. If R is a commutative C^* -algebra, then Prim R is naturally the maximal ideal space of R (since every primitive ideal is maximal in that case) and this just generalizes the Gelfand-Naimark theorem in that case.

In fact, if A is separable, then every closed prime ideal of A is primitive. This follows from applying the Baire category theorem to Prim A. This does not work in general: Weaver used transfinite induction to find a counterexample if A is not separable. (It may be undecidable whether there is a counterexample in ZF alone.)

Chapter 12

Generators and relations

We now study C^* -algebras determined by generators and relations.

12.1 Construction of maximally free algebras

Let $\{a_i\}$ be a set of generators, and take the free *-algebra F over \mathbb{C} generated by the a_i . This is the set of noncommutative polynomials in the a_i and a_i^* (where a_i^* is just a formal symbol for now).

Given a set R of relations, we can view R as noncommutative polynomial equations. So we take the ideal (R, R^*) , which is the *-ideal generated by R. Then we let $A = F/(R, R^*)$, which is a *-algebra still. We consider the set Π of all *-representations of A. Then for $a \in A$ we set

$$||a|| = \sup_{\pi \in \Pi} ||\pi(a)||.$$

A priori we have $||a|| = \infty$. This happens if F is the free *-algebra on one generator. So we need R to force $||a_i|| < \infty$ for each generator a_i .

Assume that R forces $||a_i|| < \infty$ (for example, if the generators are sent to unitary operators). Since the image of every $\pi \in \Pi$ is a C^* -algebra, $||\cdot||$ is a seminorm satisfying the C^* -identity $||a^*a|| = ||a||^2$. So we take the completion with respect to $||\cdot||$; i.e. we annihilate the kernel of the seminorm and then complete. What is left over is a C^* -algebra.

Just because a set of generators and relations gives a valid C^* -algebra of A does not mean that we necessarily can find a natural, faithful representation. Moreover, even if we have a natural, faithful representation of A, the norm arising from that representation is not necessarily the norm given by taking the supremum over Π .

Example 12.1. Let G be a discrete group, which we view as a set of generators. We impose relations corresponding to each true relation in G. (This can be viewed as taking the first-order theory of G.) We also take the relations $x^* = x^{-1}$ for $x \in G$ so the resulting maximally free operator algebra A naturally represents the elements of G as unitary operators. Therefore for $x \in G$ we have ||x|| = 1; so A is actually a C*-algebra. In fact, it is easy to see that A is the completion of $C_c(G)$, i.e. finitely supported functions on G. In other words, $A = C_r^*(G)$ is the reduced group C*-algebra.

Moreover, A acts on $\ell^2(G)$ by left translation, since G does. This representation is faithful, so A is unusual amongst the free C^* -algebras in that it has a natural faithful representation. In general, the norm in $\ell^2(G)$ does not always agree with the norm on A; this happens if and only if G is an *amenable group*. (It turns out that since G is discrete, G is amenable if and only if there is a finitely additive, left-invariant probability measure on G. For example, this fails if G is the free group on n letters, $n \geq 2$.)

12.2 Tensor products of C^* -algebras

An important application of generators and relations is the ability to define the tensor product of C^* -algebras.

Definition 12.2. Let A, B be unital C^* -algebras. Their *tensor product* $A \otimes B$ is the C^* -algebra with generators $A \cup B$ and relations consisting of all true relations in A, all true relations in B,

$$\forall a \in A \; \forall b \in B \; ab = ba,$$

and $\forall x \ (1_A 1_B) x = x$.

The relation ab = ba is the "tensor product relation", so we can reasonably think of ab as $a \otimes b$. Similarly, the relation $(1_A 1_B)x = x$ requires that $1_{A \otimes B} = 1_A 1_B$, so $A \otimes B$ is unital. We have embeddings $A \to A \otimes B$, $B \to A \otimes B$ given by $a \mapsto a \otimes 1_B$ and $b \mapsto 1_A \otimes b$.

To see that this is actually well-defined (i.e. has finite norm), let π be a *-representation of $A \otimes B$. Then π restricts to a *-representation of A (and similarly to B) along the mapping $A \to A \otimes B$. Therefore for any $a \in A$,

$$||\pi(a\otimes 1_B)|| \le ||a||.$$

Similarly for B; so there is an upper bound on the norm of any generator. Therefore $A \otimes B$ is a C^* -algebra.

Example 12.3. Let X, Y be compact Hausdorff spaces. Then we have $C(X \times Y) = C(X) \otimes C(Y)$.

We now consider the representation theory of tensor products.

Definition 12.4. Let $\pi : A \to B(H)$ and $\rho : B \to B(K)$ be representations. The tensor product of representations $\pi \otimes \rho : A \otimes B \to B(H \otimes K)$ is defined by

$$\pi \otimes \rho(a \otimes b)(\xi \otimes \eta) = \pi(a)\xi \otimes \rho(b)\eta.$$

The norm induced by the tensor product of representations is also a C^* -norm, so we have two reasonable C^* -norms on $A \otimes B$. If we use the norm obtained by taking the supremum over representations, we emphasize this by writing $A \otimes_{max} B$. If we are using the norm obtained by the tensor product of Gelfand-Naimark-Segal representations, we write $A \otimes_{min} B$. **Example 12.5** (Takesaki, 1959). Let G be the free group on 2 generators, λ its left regular representation, ρ its right regular representation. Consider the tensor product $C_r^*(G) \otimes C_r^*(G)$, represented by $\lambda \otimes \rho$. Then the two norms given above are the "minimum" (i.e. tensor product $\lambda \otimes \rho$) and "maximum" norms respectively, but there are many intermediate norms between the two that have been studied in recent years. In particular, the tensor product does not have a unique norm.

Definition 12.6. A nuclear C^* -algebra A is a C^* -algebra such that for every C^* -algebra B, the norm on $A \otimes B$ is unique.

Example 12.7. $B_0(H)$ is nuclear, since it uniquely embeds in B(H). It follows that any GCR algebra is nuclear. But if G is a discrete group, then $C^*(G)$ is nuclear if and only if G is amenable. Thus the free group on 2 generators is not nuclear.

Definition 12.8. An *exact* C^* -algebra A is a C^* -algebra such that $A \otimes_{min} \cdot$ is an exact functor.

Note that $A \otimes_{max} \cdot$ is an exact functor for any A. We now treat the coproduct in the category of C^* -algebras.

Definition 12.9. The *free product* of C^* -algebras A, B, A * B, is defined to have generators $A \cup B$ and relations induced from A, B as well as $1_A = 1_B$.

The norm is defined as in the case of tensor products. A representation of A * B consists of a pair of representations of A and B on the same Hilbert space, by the universal property.

Chapter 13

Representation theory of locally compact groups

13.1 Noncommutative dynamical systems

Definition 13.1. A noncommutative dynamical system is an action of a group G on a C^* -algebra A.

Often it is convenient that a noncommutative dynamical system is taking place inside a Hilbert space H.

Definition 13.2. Let $\alpha : G \to \operatorname{Aut}(A)$ be a noncommutative dynamical system. Let $\pi : A \to B(H)$ a *-representation, $U : G \to U(H)$ a unitary representation. If the dynamical system satisfies the *covariance relation*

$$\alpha(x)a = U_x \pi(a) U_x^*,$$

then (π, U) is called a *covariant representation* of α .

we now define *-operations on $C_c(G \to A)$. If $f, g \in C_c(G \to A)$ then

$$\left(\sum_{x} f(x)x\right) \left(\sum_{y} g(y)y\right) = \sum_{x,y} f(x)xg(y)y$$
$$= \sum_{x,y} f(x)\alpha(x)(g(y))xy$$
$$= \sum_{x,y} f(x)\alpha(x)(g(x^{-1}y))y$$

This motivates the following definition.

Definition 13.3. If α is a noncommutative dynamical system, then for $f, g \in C_c(G \to A)$, we define

$$(f * g)(y) = \sum_{x} f(x)\alpha(x)(g(x^{-1}y)),$$

the twisted convolution of f, g by α .

Definition 13.4. If α is a noncommutative dynamical system, $f \in C_c(G \to A)$, then

$$f^*(x) = \alpha(x)(f(x^{-1})^*),$$

the twisted involution of f by α .

With twisted convolution and involution, $C_c(G \to A)$ is a *-algebra. Moreover, we have an injective mapping $A \to C(G \to A)$ give by $a \mapsto a\delta_1$, 1 the identity of the group. We also have an injective mapping $G \to C(G \to M(A))$ with $x \mapsto 1x$ (here 1 is the identity of the noncommutative Stone-Cech compactification M(A); if A is unital then A = M(A).

Definition 13.5. Given any covariant representation (π, U) of $\alpha : G \to \operatorname{Aut}(A)$ we call the *integrated form* σ , a *-representation $\sigma : C_c(G, A) \to B(H)$ by

$$\sigma(f) = \sum_{x} \pi(f(x)) U_x.$$

The set of all possible integrated forms is bounded. Therefore we can make the following definition by generators and relations:

Definition 13.6. The C^{*}-algebra generated by $C_c(G, A)$ and A is called $C^*(G, A)$ or $A \rtimes_{\alpha} G$. It is called the *crossed product* C^{*}-algebra or the *covariance* C^{*}-algebra.

The generators of $A \rtimes_{\alpha} G$ will consist of all elements of G and of A, the relations will be all relations in G and A as well as

$$xa = \alpha(x)ax$$

since G acts on A by convolution (since α has a covariant representation).

To see that that there are, in fact, covariant representations, we give an explicit one.

Definition 13.7. Let $\rho: A \to B(H_0)$ be a *-representation, and define

$$H = \ell^2(G \to H_0) = \ell^2(G) \otimes H_0.$$

Now define a unitary representation of G on H by (for $x, y \in G, \xi \in H$)

$$U(x)(\xi)(y) = \xi(x^{-1}y).$$

We then define $\pi: A \to B(H)$ by

$$\pi(a)(\xi)(x) = \rho(\alpha(x^{-1})(a))(\xi)(x),$$

which is a *-representation. It is called the *induced covariant representation* of α by ρ .

Lemma 13.8. The induced covariant representation is covariant.

Proof. We have

$$U(x)(\pi(a))(\xi)(y) = (\pi(a)(\xi))(x^{-1}y) = \rho(\alpha((x^{-1}y)^{-1}))(U(x)(\xi))(y)$$

= $\rho(\alpha(y^{-1}x)(a)) = \rho(\alpha(y^{-1})(\alpha(x))(a))(U(x)(\xi))(y)$
= $(\pi(\alpha(x)(a))U(x)(\xi))(y).$

$$U(x)(\pi(a))(\xi) = \pi(\alpha(x)(a)U(x))(\xi)$$

which implies

$$U(x)(\pi(a)) = \pi(\alpha(x)(a))U(x)$$

so that (π, U) is a covariant representation.

If K is a subgroup of G, (ρ, V) a covariant representation of $\alpha|_K$ on H_0 , then the induced representation of (ρ, V) is a covariant representation of α . Here $H = \ell^2(G/H) \otimes H_0$.

Definition 13.9. The reduced group C^* -algebra of the representation α , $C^*_r(A, G, \alpha)$, is the C^* -algebra generated by $C_c(G \to A)$ with ||f|| defined to be the supremum of $||\sigma(f)||$ for σ ranging over the integrated forms of induced covariant representations.

Definition 13.10. If $C^*(A, G, \alpha) = C^*_r(A, G, \alpha)$, then α is said to be an *amenable action*.

If α is trivial, then

$$C^*(A, G, \alpha) = A \otimes C^*(G)$$

so the theory of covariant representations includes the theory of unitary representations.

Though all the above theory was developed for discrete groups, it works fine for locally compact groups, as we now describe in detail. In fact, if G is a locally compact group, then we consider the left Haar measure μ , which is a Radon measure that is unique up to positive scalars (but need not be two-sided). Moreover, if $y \in G$, if we let

$$\nu(f) = \int_G f(xy) \ d\mu(x)$$

then ν is a left Haar measure, so we can find a $\Delta(y)$ such that $\nu = \Delta(y)\mu$.

Definition 13.11. The function $\Delta : G \to \mathbb{R}^+$ is called the *modular function* of G.

The modular function is a continuous morphism of groups.

Definition 13.12. A group is unimodular if $\Delta = 1$.

For unimodular groups, the left and right Haar measures coincide. So any abelian group is unimodular. Moreover, \mathbb{R}^+ has no compact subgroups, so any compact group is unimodular. Semisimple Lie groups and nilpotent Lie groups can also be shown to be unimodular. Solvable Lie groups are often not unimodular.

Example 13.13. Let (M, ω) be a Poisson manifold (e.g. ω is a symplectic form on M). We let h be a semiclassical parameter for a family of noncommutative ring structures on $C_c^{\infty}(M \to \mathbb{C})$, so we can view (M, ω, h) as a family of operator algebras. There is a notion of modular function for M.

The trouble with locally compact groups is that the maps $G \to U(H)$ are typically not norm-continuous. In fact, this is already apparent for the left regular representation $\mathbb{R} \to U(L^2(\mathbb{R}))$, where we take a function with compact support and translate it by far outside its support.

Definition 13.14. An action α of G on a Banach space A is *strongly continuous* if for every $a \in A, x \mapsto \alpha(x)(a)$ is continuous.

This is the correct definition of the continuity of a representation of a locally compact group.

Definition 13.15. If G is a locally compact group, a C^* -dynamical system for G is a strongly continuous action of G on a C^* -algebra. A *covariant representation* of the C^* -dynamical system is one which is also strongly continuous.

The integrated form of a covariant representation (π, U) of a C^{*}-dynamical system is given by

$$\sigma_f(\xi) = \int_G \pi(f(x))U(x)(\xi) \ dx.$$

Using Bochner integration, we can define the integrated form for any $f \in C_c(G \to A)$. Now

$$||\sigma_f(\xi)|| \le \int_G ||f(x)|| \ dx \ ||\xi||$$

so it follows that $||\sigma_f|| \leq ||f||_{L^1(G)}$. So σ extends to $L^1(G \to A)$. Though the elements of $C_c(G \to A)$ are functions, it makes sense to think of them as Radon-Nikodym derivatives of A-valued measures on G with respect to Haar measure. As in the theory of discrete groups, $\sigma_f \sigma_g = \sigma_{f*\alpha g}$ where the twisted convolution is defined by

$$f *_{\alpha} g(x) = \int_{G} f(y) \alpha(y)(g(y^{-1}x))(y) \, dy$$

Therefore $||f *_{\alpha} g||_{L^{1}(G)} \leq ||f||_{L^{1}(G)} ||g||_{L^{1}(G)}$.

Another complication comes in the form of groups that are not unimodular. This happens because

$$\sigma_f^* = \int_G \pi(f(x))U_x \, dx = \int_G U_x^* \pi(f(x))^* \, dx$$
$$= \int_G U_{x^{-1}} \pi(f(x))^* \, dx = \Delta(x) \int_G U_x \pi(f(x^{-1})^*) \, dx$$

Therefore we must define the twisted involution

$$f^*(x) = \Delta(x)\alpha(x)(f(x^{-1})^*)$$

if we want $\sigma_f^* = \sigma_{f^*}$.

Example 13.16. Let G be the group of affine transformations of \mathbb{R} (the "ax + b group"). This group is far from unimodular, and its action on \mathbb{R} is important in the theory of wavelets. So the modular function is important in signal processing.

We write $C^*(A, G, \alpha) = A \rtimes_{\alpha} G$ for the completion of $C_c(G \to A)$ with respect to the norm given by taking the supremum over all integrated forms. The reduced algebra $A \rtimes_{\alpha}^r G$ is given by taking the supremum over all integrated forms arising from covariant representations. Now $L^1(G)$ does not have an identity since it does not have a delta function if G is not discrete. But we could always take an approximate delta function. Specifically, we let Λ denote the filter of all open sets containing the identity 1 of G. (A filter-base also suffices.) Then given a neighborhood λ of 1, let $f_{\lambda} \in C_c(G \to \mathbb{R}^+)$ be supported in λ with $||f_{\lambda}||_{L^1(G)} = 1$. We view f_{λ} as a probability measure carried by λ . Obviously the $(f_{\lambda})_{\lambda}$ are an approximate delta function in $L^1(G)$.

Let α be an action of G on a C^{*}-algebra A with approximate identity $(e_{\mu})_{\mu}$. We then define

$$h_{\mu,\lambda}(x) = f_{\lambda}(x)e_{\mu}$$

to obtain an approximate identity for $L^1(G \to A)$.

Theorem 13.17. There is a natural bijection between nondegenerate *-representations of $A \rtimes_{\alpha} G$ and covariant representations of α .

Proof sketch. Let σ be a nondegenerate *-representation of $A \rtimes_{\alpha} G$ and consider its multiplier algebra $M(A \rtimes_{\alpha} G)$. We have an injection $G \to M(A \rtimes_{\alpha} G)$ by $x \mapsto \delta_x$, using the fact that $L^1(G)$ is a 2-sided ideal in the multiplier algebra M(G), where we think of double centralizers as finite Radon measures (this is true up to natural isomorphism). We also have an injection $A \to M(A \rtimes_{\alpha} G)$, by $a \mapsto a\delta_{1_G}$. We can therefore obtain a covariant representation (π, U) of α obtained by restricting σ to A, G. It follows that σ is the integrated form of (π, U) . \Box

In particular, it makes sense to talk about the hull-kernel topology on the set of covariant representations of α .

Definition 13.18. For G a locally compact group, we let (A, α) and (B, β) be C^{*}-dynamical systems. Let $\varphi : A \to B$ be a *-morphism. Then φ is a *equivariant morphism* with respect to α, β if for every $a \in A$,

$$\varphi(\alpha(x)(a)) = \beta(x)(\varphi(a)).$$

The category of C^* -dynamical systems over G has equivariant morphisms as its morphisms by definition. Equivariant morphisms $\varphi : A \to B$ give rise to maps $C_c(G \to A) \to C_c(G \to B)$ given by

$$\varphi(f)(x) = \varphi(f(x)).$$

This naturally extends to the group C^* -algebras, so gives rise to a morphism $\varphi : A \rtimes_{\alpha} G \to B \rtimes_{\alpha} G$. Therefore the following theorem holds.

Theorem 13.19. The map $\alpha \mapsto A \rtimes_{\alpha} G$ is a functor from the category of C^* -dynamical systems over G to the category of C^* -algebras.

Using category theory, we can obtain the following theorem.

Theorem 13.20. Let (A, α) be a C^* -dynamical system. Let I be an α -invariant ideal of A. Then the natural action of α on A/I is a C^* -dynamical system, and the natural arrows

$$0 \to I \rtimes_{\alpha} G \to A \rtimes_{\alpha} G \to (A/I) \rtimes_{\alpha} G \to 0$$

form a short exact sequence.

In other words, the functor $\cdot \rtimes_{\alpha} G$ is exact on α -invariant ideals. This is not true for the reduced product.

Proof. Consider the short exact sequence

$$0 \longrightarrow I \stackrel{i}{\longrightarrow} A \stackrel{p}{\longrightarrow} A/I \longrightarrow 0 \ .$$

Straight from the definitions, the induced map p^* is onto and $p^*(i^*(C_c(G, A))) = 0$. So i^* maps into the kernel of p^* .

We now claim that the induced map i^* is injective. Let σ be a faithful representation of $I \rtimes_{\alpha} G$. Let (π, U) the covariant representation of σ . Then if σ is nondegenerate, so is (π, U) . Then π extends to a representation $\tilde{\pi}$ of A. Since π is nondegenerate, we can restrict to the image of $\pi(I)$ without any loss of generality when proving that $(U, \tilde{\pi})$ is covariant. In fact,

$$U(x)\tilde{\pi}(a)(\pi(d)\xi) = \pi(\alpha(x)(ad))U(x)\xi$$

which proves covariance of $(U, \tilde{\pi})$. So let $\tilde{\sigma}$ be a representation of $A \rtimes_{\alpha} G$ for which $(U, \tilde{\pi})$ is a covariant representation. Then

$$\tilde{\sigma}|_{i^*(I\rtimes_{\alpha}G)} = \sigma_i$$

so $\tilde{\sigma} \circ i^*$ is faithful on $I \rtimes_{\alpha} G$. Therefore ker $i^* = 0$. In particular $I \rtimes_{\alpha} G$ is isomorphic to an ideal of $A \rtimes_{\alpha} G$, and without loss of generality we assume that they are equal (i.e. i^* is the identity).

Finally we show exactness at $A \rtimes_{\alpha} G$; i.e. ker $p^* \subseteq I \rtimes_{\alpha} G$. Since $I \rtimes_{\alpha} G$ is a C^* -algebra, $A \rtimes_{\alpha} G/I \rtimes_{\alpha} G$ exists, and has a faithful representation σ . Pulling σ back along the quotient map $A \rtimes_{\alpha} G \to A \rtimes_{\alpha} / I \rtimes_{\alpha} G$, we obtain a representation of $A \rtimes_{\alpha} G$. Let (π, U) be the corresponding covariant representation of $A \rtimes_{\alpha} G$.

Let $d \in I$, $h \in C_c(G)$. Let $f \in C_c(G \to I) \subseteq I \rtimes_{\alpha} G$ be defined by

$$f(x) = h(x)d$$

so $\sigma(f) = 0$. Thus

$$0 = \int_{G} \pi(d) f(x) U(x) \, dx = \pi(d) \int_{G} f(x) U(x) \, dx$$

Since the integral on the right does not have to be zero, $\pi(d) = 0$. So $I \subseteq \ker \pi$. Therefore π drops to a representation τ of $A/I \rtimes_{\alpha} G$. It is routine to prove that $\sigma = \tau \circ p^*$. So $\ker \sigma = \ker p^* \subseteq I \rtimes_{\alpha} G$.

Note that the representation τ may not exist on the reduced product, which explains why the theorem fails there.

13.2 Group actions on locally compact spaces

Let A be a commutative C^{*}-algebra and let G be a locally compact group which acts on A by α . Then we can find a locally compact Hausdorff space M such that $A = C_{\infty}(M)$. We have an action α of G on M by homeomorphisms, and $G \times M$ is locally compact. We will assume that α is *jointly continuous*, i.e. the map

$$G \times M \to M$$
$$(x,m) \mapsto \alpha(x)m$$

is continuous. Thus the action

$$\alpha(x)(f)(y) = f(\alpha(x)^{-1}y)$$

of G on $C_c(M) \subseteq A$ is isometric, in particularly, strongly continuous in L^{∞} -norm. Moreover, α is continuous in the inductive limit topology of $C_c(M)$. Therefore if μ is an α -invariant Radon measure on M, α acts strongly continuously on $L^p(\mu)$.

We now consider $A \rtimes_{\alpha} G$, which contains $C_c(G \to A)$. Since

$$(f *_{\alpha} g)(y) = \int_G f(x)\alpha(x)(g(x^{-1}y)) dx,$$

it follows that

$$(f *_{\alpha} g)(y)(m) = \int_{G} f(x)(m)g(x^{-1}y)(\alpha(x)^{-1}m) \ dx.$$

Theorem 13.21. Let M be a second-countable locally compact Hausdorff space, $A = C_{\infty}(M) \rtimes_{\alpha} G$. Let σ be an irreducible nondegenerate representation of A and let (π, U) be the covariant representation of σ , $I = \ker \pi$. Let $Z \subseteq M$ be the hull of I. Then there is an α -orbit whose closure is Z.

Notice that I is α -invariant, hence an ideal of A. Expanding out the definitions, I is the set of f whose supports are disjoint from Z. In particular, Z is closed and α -invariant and $A/I = C_{\infty}(Z)$. The theorem says that there is a $m_0 \in M$ such that

$$Z = \overline{\{\alpha(x)m_0 : x \in G\}}.$$

Example 13.22. For the irrational rotation, every orbit-closure is the entire circle, so for every ideal, the hull is the entire space. This generalizes to various ergodic actions.

Proof of theorem. Since M is second-countable, so is Z. Let $\{B_n\}_n$ be open subsets of M such that the $B_n \cap Z$ form a countable base for the topology of Z, $B_n \cap Z$ nonempty. Let

$$O_n = \bigcup_{x \in G} \alpha(x)(B_n).$$

Then the O_n are open, α -invariant, and meet Z.

Let $J_n = C_{\infty}(O_n) \subseteq C_{\infty}(M)$. Since $O_n \cap Z$ is nonempty, there is a $f \in J_n$ which is not identically zero on Z, by Urysohn's lemma. So $\pi(J_n)$ is nonzero. Since σ is nondegenerate, so is π , and $\pi(J_n)H$ generates a nonzero closed σ -invariant subspace. Since σ is irreducible, $\pi(J_n)H$ generates H. Let ξ be a unit vector of the representation space H. Define a Radon probability measure μ on M by

$$\int_M f \ d\mu = \langle \pi(f)\xi,\xi\rangle$$

If $f \in I$, f = 0 μ -almost everywhere. Therefore Z contains the support of μ . Since M is second-countable, J_n is a separable C^* -algebra and we can find a countable normalized approximate unit $\{e_{n,m}\}_m$ of J_n . We can assume that the $e_{n,m}$ are compactly supported in O_n . Since $\pi|_{J_n}$ is nondegenerate,

$$\lim_{m \to \infty} \pi(e_{n,m})\xi = \xi.$$

The $e_{n,m}$ are supported on O_n , so O_n contains the support of μ . Therefore

$$\operatorname{supp} \mu \subseteq \bigcap_{n=1}^{\infty} O_n \cap Z.$$

(Here we are using the cardinality assumption; the complements must be μ -null and there are only countably many of them.) Since μ is a probability measure, supp μ is nonempty.

Let $m_0 \in \operatorname{supp} \mu$. Each of the O_n is α -invariant, so $\alpha_G(m_0) \subseteq \operatorname{supp} \mu$. So for each n, $\alpha_G(m_0) \subseteq O_n \cap Z$. Since $\alpha_G(m_0)$ is contained in every element of an open base of Z, $\alpha_G(m_0)$ is dense in Z.

Example 13.23. Let M be the two-point compactification of \mathbb{R} . Then the action of \mathbb{R} on M by translation is jointly continuous, and \mathbb{R} is a dense orbit, but the boundary points $\pm \infty$ are fixed points. So not every point has a dense orbit. We will study $C(M) \rtimes \mathbb{R}$ soon.

Let α be an action of G on M. For $m \in M$, let G_m be the stabilizer of m. By the orbit-stabilizer theorem, the map

$$G/G_m \to \alpha_G(m)$$

 $x \mapsto \alpha_{xG_m}(x)$

is a bijection (where xG_m is the coset of G_m by x). Now G_m is a closed normal subgroup so G/G_m is a locally compact group. In general the orbit-stabilizer map $G/G_m \to \alpha_G(m)$ is not a homeomorphism. It is favorable that the orbit of x is open in its closure, in which case the orbit is locally compact.

Theorem 13.24. Let $m_0 \in M$. If G is a second countable group which acts on M by α , and $\alpha_G(m_0)$ is locally compact, then $G/G_{m_0} \to \alpha_G(m_0)$ is a homeomorphism.

Proof. Use the Baire category theorem on the locally compact space $\alpha_G(m_0)$.

Let H be a closed normal subgroup and let M = G/H. Then G acts on M by left translation, and $A = C_{\infty}(M) \rtimes G$ is a C^* -algebra. If H = G, then $A = C^*(G)$. If H = 0, then $A = C_{\infty}(G) \rtimes G$.

In case H = 0, we study the covariant representation on $L^2(G)$ given by U the left regular representation, π the representation by pointwise representation; i.e.

$$\pi(f)(\xi)(x) = f(x)\xi(x).$$

To see covariance, we compute

$$U(x)\pi(f)(\xi)(y) = \pi(f)(\xi)(x^{-1}y) = f(x^{-1}y)\xi(x^{-1}y) = \pi(\alpha(x)(f))(U(x)(\xi))(y).$$

Definition 13.25. The Schrodinger representation of a group G is the covariant representation (π, U) of $C_{\infty}(G) \rtimes_G$ on $L^2(G)$ given above.

We compute the integrated form σ of the Schrödinger representation by realizing that

$$C_c(G \to A) = C_c(G \to C_\infty(G))$$

is generated by $C_c(G \times G)$. Given $F \in C_c(G \times G)$ we have

$$\sigma(F)(\xi)(x) = \left(\int_G \pi(F(y))U(y)(\xi) \, dy\right)(x)$$
$$= \int_G F(y,x)\xi(y^{-1}x) \, dy.$$

Now $f, g \in C_c(G)$ can be viewed as elements of $L^2(G)$, which has a rank-1-operator-valued inner product $\langle \cdot, \rangle_{0}$. In fact,

$$\begin{split} \langle f,g \rangle_0(\xi)(x) &= f(x) \langle g,\xi \rangle(x) \\ &= f(x) \int_G \overline{g(y)} \xi(y) \ dy \\ &= \int_G f(x) \overline{g(y^{-1})} \xi(y^{-1}) \Delta(y^{-1}) \ dy \\ &= \int_G f(x) g(y^{-1}x) \xi(y^{-1}x) \Delta(y^{-1}x) \ dy. \end{split}$$

We define

$$\langle f, g \rangle_E(x, y) = f(x) \overline{g(y^{-1}x)} \Delta(y^{-1}x).$$

Then this is an inner product which has values in $C_c(G \times G)$. Let E be the (algebraic) span of

$$\{\langle f,g\rangle_E: f,g\in C_c(G)\}$$

Thus

$$\langle f,g\rangle_E * \langle h,k\rangle_E = \langle \langle f,g\rangle_E h,k\rangle_E$$

and

$$\pi(\langle f,g\rangle_E)\pi(\langle h,k\rangle_E) = \langle f,g\rangle_0\langle f,g\rangle_0 = \langle\langle f,g\rangle_0h,k\rangle_0$$
$$= \langle g,h\rangle\langle f,k\rangle_0 = \pi(\langle g,h\rangle\langle f,k\rangle_E)$$

allows us to define

$$\langle f,g\rangle_E * \langle h,k\rangle_E = \langle g,h\rangle\langle f,k\rangle_E.$$

This defines a convolution on E which is compatible with the convolution on $C^*(G, C_{\infty}(G))$. Therefore E is a subalgebra of $C^*(G, C_{\infty}(G))$. Clearly E is a *-algebra since

$$\langle f,g\rangle_E = \langle g,f\rangle_E.$$

Moreover, $C_c(G)$ is dense in $L^2(G)$, so $\pi(E)$ is dense in $B_0(H)$.

We claim that E is also closed under pointwise multiplication. In fact,

$$\langle f, g \rangle_E \langle h, k \rangle_E(x, y) = f(x) \overline{(\Delta g)} (y^{-1}x) h(x) \overline{(\Delta k)} (y^{-1}x)$$

= $f(x) g(x) \overline{\Delta g} k(y^{-1}x) \Delta(y^{-1}x).$

Clearly E is closed under complex conjugation and separates points of $G \times G$ from zero. Thus we can apply the Stone-Weierstrass theorem, but this is not very interesting because we actually want to prove that E is dense in $C_c(G \times G)$ for the inductive limit topology. In fact, if O is an open, precompact subset of $G \times G$, we can find $V \times W \subseteq G \times G$, where V, Ware open, precompact subsets of G and consider the algebraic span of $\langle C_c(V), C_c(W) \rangle_E$. By the Stone-Weierstrass theorem, this is L^{∞} -dense in $C_c(V \times W)$. One can then check that Eis dense in $L^1(G \to C_{\infty}(G))$ and hence dense in $C^*(G, C_{\infty}(G))$.

We claim that E has the same operator norm as $C^*(G, C_{\infty}(G))$. In fact if f_1, \ldots, f_n are an L^2 -orthonormal set in $C_c(G)$ then the $\langle f_j, f_k \rangle_0$ span the C^* -algebra $\mathbb{C}^{n \times n}$ once we choose a basis. On C^* -algebras the operator norm is uniquely determined, so E agrees with $C^*(G, C_{\infty}(G))$ in operator norm on any finite-dimensional subalgebra. Such matrix algebras can be used to approximate $C^*(G, C_{\infty}(G))$ so we have proven the claim. We consider that we have proven the following theorem.

Theorem 13.26. $C_{\infty}(G) \rtimes G = B_0(L^2(G)).$

Since $B_0(L^2(G))$ has no proper ideals, one also has $C_\infty(G) \rtimes^r G = B_0(L^2(G))$. Therefore the translation action is amenable.

Example 13.27. If G is not an amenable group, then G still admits an amenable action by translation.

Now if G acts on X by α , and O is an orbit of α , then if O is an orbit which is open in its closure, $C_{\infty}(O) \subseteq C(\overline{O})$. Moreover, $C_{\infty}(O) \rtimes_{\alpha} G = B_0(L^2(G))$.

Example 13.28. Let \mathbb{R} act on its two-point compactification X by translation. Then $C_{\infty}(X) = C(X)$ contains $C_{\infty}(\mathbb{R})$. So

$$C(X) \rtimes \mathbb{R} \supset C_{\infty}(\mathbb{R}) \rtimes \mathbb{R} = B_0(L^2(\mathbb{R}))$$

which gives a GCR representation of C(X) on $L^2(\mathbb{R})$. It is not CCR because C(X) is unital.

If we instead look at the orbits of $\pm \infty$, we see that $C^*(\mathbb{R}) = C(\pm \infty) \rtimes_{\alpha} \mathbb{R}$. By the Fourier transform, $\widehat{C^*(\mathbb{R})} = \mathbb{R}$. (More generally, if G is a locally compact abelian group, then $C^*(G)$ is a commutative C^* -algebra, $\widehat{C^*(G)}$ consists of one-dimensional representations of G, which are exactly the continuous morphisms $G \to S^1$.)

Definition 13.29. Let H be a closed normal subgroup of G. For simplicity we assume that the Haar measure on G/H is G-invariant. Let $V : H \to U(K)$ be a unitary representation. We define a Hilbert space by taking all functions $\xi : G \to K$ such that for all $x \in G$, $s \in H$, $\xi(xs) = V(s)^*\xi(x)$ where we define $\langle \xi, \eta \rangle(x) = \langle \xi(x), \eta(x) \rangle$. Since $\langle \xi, \eta \rangle$ is constant on cosets, it drops to a function on G/H such that

$$\langle \xi, \eta \rangle(x) = \int_{G/H} \langle \xi, \eta \rangle(x) \, dx.$$

We take the Hilbert space to be all ξ such that $\langle \xi, \xi \rangle < \infty$, which G acts on by left translation. This action of G is called the *induced representation* of G from V, Ind V.

If H, V are as above, α the action of G on $C_{\infty}(G/H)$ by left translation, then we obtain a covariant representation of α on the induced representation space by letting $C_{\infty}(G/H)$ act by left translation and λ be the left action of G on G/H. Then $(C_{\infty}(G/H), \lambda)$ is a covariant representation of α .

Let G be a unimodular group (though the same argument goes through without too much trouble otherwise). Recall that $C_c(G/H \to G) \subseteq C(G/H) \rtimes_{\alpha} G$. We define for $f, g \in C^*(H)$,

$$\langle f, g \rangle_{C^*(H)} = f^* * g|_H.$$

Here we are using continuity; if H is a Haar null set then the restriction map is not defined for measurable functions in general. Let $B = C_c(G/H \to G)$. Then we can define

$$\langle f, g \rangle_B h = f \langle g, h \rangle_{C^*(H)}.$$

One can then prove $C(G/H) \rtimes_{\alpha} G$ is strongly Morita equivalent to $C^*(H)$. This theory generalizes to when G is merely a groupoid rather than a group.

13.3 Semidirect products of groups

Definition 13.30. If N and Q are locally compact groups, and $\alpha : Q \to \operatorname{Aut}(N)$ a jointly continuous action, then we define $N \rtimes_{\alpha} Q$ as follows. As a Hausdorff space, $N \rtimes_{\alpha} Q$ is the product of topological spaces $N \times Q$. The group operation is defined by

$$(n_1, q_1)(n_2, q_2) = (n_1 \alpha(q_1)(n_2), q_1 q_2).$$

Then $N \rtimes_{\alpha} Q$ is the *semidirect product* of locally compact groups.

We remember the group operation on $N \rtimes_{\alpha} Q$ by recalling that "whenever we want to commute an n and a q, the q must act on the n."

Let $G = N \rtimes_{\alpha} Q$. Then Q and N embed in G in the obvious way and we have a split exact sequence

$$0 \to N \to G \to Q \to 0.$$

Therefore any representation of G restricts to representations of N and Q. Moreover, Q acts on N by inner automorphisms, i.e.

$$qnq^{-1} = \alpha(q)(n).$$

The action of q does not preserve Haar measure, but it does send Haar measure to a translation-invariant measure; i.e. it multiplies Haar measure by a scalar, say $\sigma(q)$.

We now define an action of Q on $C^*(N)$. In fact, for $f \in C_c(N)$, $n \in N$, $q \in Q$,

$$\alpha(q)(f)(n) = \sigma(q)f(\alpha(q)(n)).$$

Then $\alpha(q)$ is an isometry in L^1 -norm, so Q acts on $L^1(N)$ by isometries. This action α immediately extends to $C^*(N)$.

Let U be a unitary representation of G. Restricting, we obtain representations of $C^*(N)$ and Q, and U is a covariant representation of $\alpha : Q \to \operatorname{Aut}(C^*(N))$. Conversely, a representation of $C^*(N) \rtimes_{\alpha} Q$ gives rise to a covariant representation of α . Then

$$C^*(N) \rtimes_{\alpha} Q = C^*(G).$$

This can be remembered as $C^*(N) \rtimes_{\alpha} Q = C^*(N \rtimes_{\alpha} Q).$

Let N be an abelian group. Then $C^*(N) = C_{\infty}(\hat{N})$ (where \hat{N} denotes the Fourier transform, $\hat{N} = \text{Hom}(N, S^1)$ in the category of locally compact groups). Since Q acts on N and hence $C^*(N)$, Q also acts on \hat{N} . In fact if $\varphi \in \hat{N}$, then

$$\alpha(q)(\varphi)(n) = \varphi(\alpha(q^{-1})(n)).$$

Then $C^*(N \rtimes_{\alpha} Q) = C_{\infty}(\hat{N}) \rtimes_{\alpha} Q$. Thus we are back in the original situation of a locally compact group acting on a locally compact space.

Example 13.31 (Wigner 1939). Let L be the Lorentz group, the automorphism group of Minkowski spacetime (linear automorphisms that preserve the Lorentzian metric $g(x, y) = -x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3$.) Then L acts on \mathbb{R}^4 , so we have a semidirect product $\mathbb{R}^4 \rtimes L$, the *Poincare group*. The unitary representations of the Poincare group are important in relativistic quantum mechanics. Elementary particles "should be" completely determined by their symmetries, so correspond to representations of certain stabilizers of $\mathbb{R}^4 \rtimes L$. This paper led to the discovery that electrons have spin. In principle one could use the representation theory of $\mathbb{R}^4 \rtimes L$ to rederive the periodic table of elements.

13.4 The Heisenberg commutation relations

We now look at an algebra with "invalid" generators and relations.

In quantum physics, the position q and momentum p observables act on certain dense subspaces of tensor powers of $L^2(\mathbb{R})$ with

$$[p,q] = i\hbar$$

They must be unbounded operators, since their commutator is a scalar. So q, p are not elements of a C^* -algebra. This relation is called the *Heisenberg commutation relation*.

We want to be able to form the holomorphic functional calculus for an unbounded operator T. In particular, we would like to define a one-parameter unitary group by the group morphism $t \mapsto e^{itT}$.

Example 13.32. The Schrödinger equation is the PDE that says that if H is the Hamiltonian, the action of its one-parameter unitary group $t \mapsto e^{itH}$ is the time-advance map.

Reasoning just formally about how the holomorphic functional calculus should behave, we let $U(s) = e^{isP}$ and $V(t) = e^{itQ}$. Weyl observed that

$$U(s)V(t)U(s)^* = e^{itU(s)QU(s)^*}$$

so we let

$$\varphi(s) = U(s)QU(-s).$$

Then

$$\varphi'(s) = iU(s)(PQ - QP)U(-s) = -\hbar$$

So $\varphi(s) = Q - s\hbar$ whence

$$U(s)V(t) = e^{itQ}e^{-i\hbar ts}U(s) = e^{-i\hbar st}V(t)U(s).$$

Recall that $\hat{\mathbb{R}} \cong \mathbb{R}$ (noncanonically). If we let $\langle \cdot, \cdot \rangle$ be the pairing of \mathbb{R} and $\hat{\mathbb{R}}$, then we have just proved

$$U(s)V(t) = \langle s, t \rangle V(t)U(s).$$

Here the choice of isomorphism is induced by some normalization of the Fourier transform.

Definition 13.33. Let G be an locally compact abelian group. By a representation for the Heisenberg commutation relations of G we mean a pair of unitary representations (U, V), $U: G \to \operatorname{Aut}(H), V: \hat{G} \to \operatorname{Aut}(H)$, such that

$$U(s)V(t) = \langle s, t \rangle V(t)U(s).$$

Any unitary representation V of \hat{G} lifts to a representation π of the commutative C^* algebra $C^*(\hat{G})$, which is C_{∞} of the double dual of G. By the Pontryagin duality theorem, the double dual of any locally compact abelian group is itself, so $C^*(\hat{G}) = C_{\infty}(G)$. Let $f = \hat{h} \in C_{\infty}(G)$ for some $h \in L^1(G)$. Then

$$\pi(f) = \int_{\hat{G}} h(t) V(t) \ dt.$$

Therefore

$$U(s)\pi(f)U(s)^* = \int_{\hat{G}} h(t)U(s)V(t)U(s)^* dt = \int_{\hat{G}} h(t)\langle s,t\rangle V(t) dt$$
$$= \pi(\alpha_s(f))$$

where α is the action of G on $C_{\infty}(G)$ by left translation. So (π, U) is a covariant representation of α and hence a representation of $C_{\infty}(G) \rtimes_{\alpha} G = B_0(L^2(G))$. But $B_0(L^2(G))$ only has one irreducible representation, which turns out to be the Schrödinger representation. This shows that the Heisenberg picture and the Schrödinger picture are equivalent. This is a theorem of von Neumann which was important to the foundations of physics.

13.5 **Projective representations**

Let $W: G \times \hat{G} \to U(H)$ be defined by

$$W(s,t) = U(s)V(t),$$

where (U, V) is a representation for the Heisenberg commutation relations. Then

$$W(s,t)W(s',t') = U(s)V(t)U(s')V(t') = -U(s+s')V(t)\langle s',t\rangle V(t')$$
$$= \langle s',t\rangle U(s+s')V(t+t') = \langle s',t\rangle W(s+s',t+t').$$

Definition 13.34. Let G be a group. A projective representation of G is a continuous function $W: G \to U(H)$ defined by

$$W(x)W(y) = c(x, y)W(xy)$$

for some $c: G^2 \to S^1$.

Here we are thinking of S^1 as the unit circle subgroup of \mathbb{C} . It is a morphism up to a harmless constant. In fact, if $\mathbb{P}H$ is the projective space of some Hilbert space, then every projective representation drops to a morphism of groups $G \to \operatorname{Aut}(\mathbb{P}H)$, since it permutes the one-dimensional subspaces. Wigner proved that every automorphism of $\mathbb{P}H$ is given by a unitary or antiunitary (i.e. conjugate linear) operator. If P is a rank-1 projection, then Pis sent to UPU^{-1} by any such automorphism, for U a unitary or antiunitary operator.

Example 13.35. Charge-conjugation, parity, and time-reversal are examples of antiunitary operators in quantum field theory.

So we have constructed a projective representation of $G \times \hat{G}$. Up to a normalization we may assume $U_1 = 1$. Given $d: G \to S^1$, set V(x) = d(x)U(x). Then

$$V(x)V(y) = d(x)d(y)U(x)U(y) = d(x)d(y)d(xy)V(xy).$$

Associativity of $\operatorname{Aut}(\mathbb{P}H)$ manifests as

$$c(xy, z)c(x, y) = c(x, yz)c(y, z).$$

We say that c is a 2-cocycle for G valued in S^1 .

Example 13.36. Let C_k be the set of all functions $G^k \to S^1$. The boundary operator for the homology of groups with values in S^1 is defined by $\partial : C_1 \to C_2$ by

$$\partial d(x, y) = d(x)d(y)d(xy),$$

and $\partial: C_2 \to C_3$ by

$$\partial d(x, y, z) = c(xy, z)c(x, y)\overline{c(x, yz)c(y, z)}.$$

This extends to a homology theory for all k. Here S^1 can be replaced by any abelian group. If $c' = (\partial d)c$, then c' and c are homologous.

Assume that for all $\xi \in H$, $x \mapsto U(x)(\xi)$ is measurable. Then for every $f \in L^1(G)$, we define

$$U(f)(\xi) = \int_G f(x)U(x)(\xi) \, dx$$

Then $||U(f)|| = ||f||_{L^1}$ and $U(f)U(g) = U(f *_c g)$ where $*_c$ is the twisted convolution defined by

$$f *_{c} g(x) = \int_{G} f(y)g(y^{-1}x)c(y,y^{-1}x) \, dy.$$

So $L^1(G, c)$ (which is L^1 with the twisted convolution $*_c$) is a Banach algebra, which is not commutative even if G is abelian. If c is homologous to c' then we have an isomorphism $L^1(G, c) \to L^1(G, c')$.

Given a 2-cocycle c we can consider all projective representations of G where the 2-cocycle is homologous to c. In these cases, the isomorphism of Banach algebras above implies that we can assume that the 2-cocycle is actually c. For $f \in L^1(G, c)$, this defines the C^* -norm by

$$||f||_{C^*(G,c)} = \sup_U ||U(f)||$$

where U ranges over all projective representations whose cocycles are homologous to c. Here we have a twisted adjoint, which for unimodular groups can be explicitly expressed as

$$f^*(x) = \overline{f(x)c(x,x^{-1})}.$$

We have a left regular representation $L: L^1(G, c) \to L^2(G)$ defined by

$$L(f)(\xi)(x) = \int_G f(y)\xi(y^{-1}x)c(y,y^{-1}x) \, dy.$$

This gives rise to the reduced C^* -algebra $C^*_r(G, c)$.

Example 13.37. Let $G = \mathbb{R}^n \times \widehat{\mathbb{R}^n}$. We define the cocycle $c((x, s), (y, t)) = \langle (x, s), (y, t) \rangle$. We already saw that $C^*(G, c) \cong B_0(L^2(G))$ in an unnatural way, by uniqueness of the Heisenberg commutation relation. Now G is abelian, but $B_0(L^2(G))$ is far from commutative.

Example 13.38. Let $G = \mathbb{Z}^n$. Then every cocycle is homologous to a *bicharacter*, a cocycle *c* of the form

$$c(m,\ell) = e^{\langle im,\Theta\ell\rangle}$$

where $\Theta \in \mathbb{R}^{n \times n}$.

When we study G we will assume without loss of generality that c is a bicharacter. Henceforth we will mainly be interested in discrete groups, but really we are actually studying \mathbb{Z}^n .

If G is a discrete group with a cocycle c, we can define a faithful tracial state τ on $\ell^1(G,c)$ (hence on $C^*(G,c)$) by $\tau(f) = f(e)$. Moreover, δ_1 is the identity of $\ell^1(G,c)$, and $\tau(f *_c f^*) = \sum_{y \in G} |f(y)|^2$. From this it follows that the GNS construction for τ gives a faithful representation $\ell^1(G,c) \to \ell^2(G)$, so extends to a representation $C^*(G,c) \to \ell^2(G)$.

If G is discrete and abelian, then \hat{G} is compact, and we have an action $\hat{\alpha} : G \to \operatorname{Aut}(C^*(G,c)),$

$$\hat{\alpha}(t)(f)(x) = \langle x, t \rangle f(x).$$

Then

$$\hat{\alpha}(t)(f \ast_c g)(x) = \hat{\alpha}(t)(f) \ast_c \hat{\alpha}(t)(g)(x).$$

To study the properties of this action $\hat{\alpha}$, G be a compact group with its Haar probability measure, and $\alpha : G \to \operatorname{Aut} A$ an action. Then we define $P : A \to A$,

$$P(a) = \int_G \alpha(x)(a) \ dx.$$

Then P is α -invariant, $\alpha(y)(P(a)) = P(a)$. In particular, if we let A^G be the algebra of all fixed points of α , P carries A into A^G . Conversely, if a is actually a fixed point, then P(a) = a. So P is the projection map $A \to A^G$.

Definition 13.39. Assume $B \subseteq A$, and $P : A \to B$ is a projection. If for every $b \in B$, $a \in A$, P(ab) = P(a)b and P(ba) = bP(a), we say that P is a *conditional expectation*.

It is easy to check that the projection $P: A \to A^G$ is a conditional expectation. Moreover, if a > 0 then P(a) > 0.

If G is a compact abelian group (in applications, G is usually a torus), then \hat{G} is discrete (in applications, \mathbb{Z}^n). We let α be an action of G on A. For each $t \in \hat{G}$, set

$$a_t = \int_G \langle x, t \rangle \alpha_x(a) \ dx.$$

So the a_t are the generalized Fourier coefficients of a. We have

$$\alpha(y)(a_t) = \alpha(y) \left(\int_G \overline{\langle x, t \rangle} \alpha(x)(a) \ dx \right) = \int_G \overline{\langle y^{-1x}, t \rangle} \alpha(x)(a) \ dx = \langle y, t \rangle a_t$$

We now set $A_t = \{a \in A : \forall y \ \alpha(y)(a) = \langle y, t \rangle a\}$. Then $a_t \in A_t$ and A_t is a closed subspace, hence a C^* -algebra. For $a \in A_t$, $b \in A_s$, we have

$$\alpha(y)(ab) = \alpha(y)(a)\alpha(y)(b) = \langle y, t \rangle \langle y, s \rangle ab = \langle y, ts \rangle ab$$

so $ab \in A_{ts}$.

Chapter 14

Noncommutative geometry

14.1 Quantum tori

Fix $\Theta \in \mathbb{R}^{d \times d}$ and let $c_{\Theta}(m, n) = e^{2\pi i m \cdot \Theta n}$, for $(m, n) \in \mathbb{Z}^{d+d}$. Then c_{Θ} is a cocycle for the duality of \mathbb{Z}^d and the torus T^d . In fact the pairing is given by

$$c_{\Theta} = \langle m, \Theta n \rangle.$$

Definition 14.1. The C^* -algebra $A_{\Theta} = C^*(\mathbb{Z}^d, c_{\Theta})$ is called the algebra of functions on the *noncommutative torus* or *quantum torus* of dimension *d*.

Now

$$\delta_m * \delta_n * \delta_m^* = \langle n, (\Theta - \Theta^t) m \rangle \delta_n.$$

So we can reasonably define

$$\rho_{\Theta}(m) = (\Theta - \Theta^t)m \in T^d.$$

We now define

$$H_{\Theta} = \overline{\{\rho_{\Theta}(m) \in T^d : m \in \mathbb{Z}^d\}}.$$

Then H_{Θ} is an subgroup of T^d , so $\rho_{\Theta} : \mathbb{Z} \to T^d$ is a morphism of groups. It gives rise to an action α of H_{Θ} on A defined by

$$\delta_m a \delta_m^* = \alpha(\rho_\Theta(m))(a).$$

14.2 The 2-torus

Theorem 14.2. If $H_{\Theta} = T^d$ then A_{Θ} is simple.

Proof. If I is a closed ideal of A_{Θ} then I is closed under conjugation, hence under the action of H_{Θ} . Now A_{Θ} is a space of noncommutative functions on T^d so we're done.

Example 14.3. If
$$d = 2$$
, we take $\Theta = \begin{bmatrix} \theta \\ 0 \end{bmatrix}$. Then $\Theta - \Theta^t = \begin{bmatrix} \theta \\ -\theta \end{bmatrix}$. So $\rho_{\Theta}(m_1, m_2) = (\theta m_2, -\theta m_1).$
If θ is irrational then $\{\theta m_2 : m_2 \in \mathbb{Z}\}$ is dense in S^1 . Therefore ρ_{Θ} acts densely on T^2 . So if Θ is irrational then A_{Θ} is simple.

Let $M = S^1$, $\theta \in \mathbb{R}$, α the action of \mathbb{Z} on M by rotation by θ . Then $C(M) \rtimes_{\alpha} \mathbb{Z}$ is a rotation algebra, i.e. it is the universal C^* -algebra generated by a unitary, namely $U = e^{2\pi i t}$. If V is the unitary acting on C(M) by $Vf = \alpha(1)(f)$ (so translation by θ), then $VU = e^{2\pi i \theta} UV$. One can then show that

$$C(M) \rtimes_{\alpha} \mathbb{Z} = C^*(\mathbb{Z}^2, c_{\Theta})$$

where $\Theta = \begin{bmatrix} \theta \\ -\theta \end{bmatrix}$. So this is another construction of A_{Θ} .

If θ is irrational, then α is a free action (i.e. all stabilizers are trivial). We now define a morphism $C(M) \to C_b(\mathbb{Z})$ by $\tilde{f}(n) = f(\alpha(n)(t_0))$ for some fixed $t_0 \in M$. Now $C_b(\mathbb{Z})$ acts on $\ell^2(\mathbb{Z})$ by multiplication and \mathbb{Z} acts on $\ell^2(\mathbb{Z})$ by translation. This gives a covariant representation of α on $\ell^2(\mathbb{Z})$. One can then show using certain commutation relations that the covariant representation is irreducible, hence gives an irreducible representation of A_{Θ} . This depends on the orbit of t_0 , so we construct uncountably many irreducible representations of A_{Θ} , all of which have kernel 0 since A is simple.

Thus we have constructed a C^* -algebra with lots of irreducible representations that have the same primitive ideal but are not unitarily equivalent. There are even more irreducible representations that we have not treated.

If θ is rational then every orbit is finite, and $C(M) \rtimes_{\alpha} \mathbb{Z}$ is a continuous field of $d \times d$ matrix algebras, which is not isomorphic to $C(T^2 \to M^d)$.

Example 14.4. Let M be a compact Hausdorff space, G a finite group, α a free action of G on M. Then M/α is a compact Hausdorff space and we have a Morita equivalence $C(M) \rtimes_{\alpha} G \to C(M/\alpha)$. "Most" of noncommutative algebraic topology is only defined up to Morita equivalence, so from the point of view of an algebraic topologist, $C(M) \rtimes_{\alpha} F =$ $C(M/\alpha)$. This is a very unusual situation!

If G is an infinite group instead, then M/α may not be Hausdorff (for example, if G is a Lie group which foliates M badly). Then $C(M/\alpha)$ may not be a C^{*}-algebra, so we have no way of studying its algebraic topology. We can still find topological invariants of the dynamical system α by instead studying the topology of $C(M) \rtimes_{\alpha} G$.

Example 14.5. Let $H = U + U^* + r(V + V^*)$ where U, V are the generating unitaries of the 2-dimensional quantum torus. In physics, $U + U^*$ is the potential energy, $r(V + V^*)$ is the kinetic energy, r "electron coupling", and H the Hamiltonian. Hofstadter (of Godel-Escher-Bach fame) showed that if θ is rational but with large denominators, then the spectrum of H approximates a Cantor set. So he conjectured that if θ is irrational then the spectrum is Cantor space. Katz offered 10 martinis for anyone who could prove this, which was known as the *ten martinis conjecture*. Avila et al. proved the ten martinis conjecture.

If $\delta_m \in Z(A_{\Theta})$ then $\alpha(\rho_{\Theta}(m))$ is the identity. So for all n,

$$1 = \langle n(\Theta - \Theta^t)m \rangle = \langle (\Theta^t - \Theta)n, m \rangle$$

so *m* lies in the dual group H_{Θ}^{\perp} of H_{Θ} . One can then show that $Z(A_{\Theta}) = C^*(H_{\Theta}^{\perp}) = C(\widehat{H_{\Theta}^{\perp}})$. Then we can express A_{Θ} as a continuous field over $C(\widehat{H_{\Theta}^{\perp}})$.

14.3 Noncommutative smooth manifolds

We now take the theory of Lie groups and smooth manifolds and turn it into a noncommutative theory.

Let G be a Lie group. We can always assume that G is a closed, connected subgroup of $\operatorname{GL}(\mathbb{R}^n)$. In fact $\operatorname{GL}(\mathbb{R}^n)$ can be obtained by applying the exponential map to $\mathbb{R}^{n \times n}$; i.e. the exponential of a matrix is an invertible matrix. We therefore define Lie $G = \{X \in \mathbb{R}^{n \times n} : \forall t \in \mathbb{R} \ e^{tX} \in G\}$. Then Lie G is a Lie algebra and Lie is the functor that sends a Lie group to its Lie algebra. Besides, exp : Lie $G \to G$ is the exponential map (in the sense of Riemannian geometry), so is close to the identity of G a homeomorphism.

Given $X \in \text{Lie } G, t \mapsto e^{tX}$ is a morphism of groups $\mathbb{R} \to G$; i.e. a smooth one-parameter subgroup of G. In fact every one-parameter subgroup is of this form, though we note that $t \mapsto e^{tX}$ may not be injective. (For example S^1 is a one-parameter subgroup which is periodic.)

Example 14.6. Lie $T^d = \mathbb{R}^d$.

Let α be a strongly continuous action of \mathbb{R} by isometries on a Banach space B. Let $b \in B$. Then we have a one-parameter semigroup $r \mapsto \alpha(r)(b)$.

Definition 14.7. Let G be a Lie group and α a strongly continuous action of G by isometries on a Banach space B. Given $X \in \text{Lie } G, b \in B, X \mapsto \alpha(X)(b)$, the *directional derivative* is

$$D_X b = \lim_{r \to 0} \frac{\alpha(\exp(rX))(b) - b}{r}.$$

We let B^{∞} be those $b \in B$ such that every higher directional derivative $D_{X_1} \cdots D_{X_n} b$ exists.

Theorem 14.8 (Garding). Let $f \in C^{\infty}_{comp}(G)$, f supported in a small enough neighborhood of the identity. Given $b \in B$, then the integrated form $\alpha(f)(b) \in B^{\infty}$.

Proof. Let $X \in \text{Lie} G$. Then

$$\begin{aligned} D_X(\alpha(f)(b)) &= \lim_{t \to 0} \frac{\alpha(\exp(tX))(\alpha(f)(b)) - \alpha(f)(b)}{t} \\ &= \lim_{t \to 0} \frac{1}{t} \left(\alpha(\exp(tX)) \int_G f(x)\alpha(x)(b) \ dx - \int_G f(x)\alpha(x)(b) \ dx \right) \\ &= \lim_{t \to 0} \frac{1}{t} \left(\int_G f(\exp(-tX)x)\alpha(x)(b) \ dx - \int_G f(x)\alpha(x)(b) \ dx \right) \\ &= \lim_{t \to 0} \int_G \frac{f(\exp(-tX)x) - f(x)}{t} - \alpha(x)(b) \ dx \\ &= \int_G D_{-X} f(x)\alpha(x)(b) \ dx. \end{aligned}$$

So $\alpha(f)(b)$ is once differentiable. Now use the fact that

$$D_Y D_X \alpha(f)(b) = \alpha(D_Y D_X f)(b)$$

to see that $\alpha(f)(b)$ twice differentiable and induct.

Corollary 14.9. B^{∞} is dense in B.

Proof. Choose an action α and let $f_n \in C^{\infty}_{comp}(G)$ be an approximate identity for $L^1(G)$. Then the sequence of $\alpha(f_n)(b)$ approximates b arbitrarily well.

Let A be a Banach algebra. If α is a strongly continuous action by algebra homomorphisms of the Lie group G, then for $a, b \in A^{\infty}$, $X \in \text{Lie } G$, the directional derivative D_X is a derivation of A. It is reasonable to think of the space of derivations of A as "vector fields on the noncommutative space \hat{A} ," assuming that the space of derivations has the structure of a A-module. But in general, it is only a Z(A)-module. Therefore, in general, we cannot define the tangent bundle of a noncommutative smooth manifold.

Example 14.10. Let $G = T^d$ so $\hat{G} = \mathbb{Z}^d$ and Lie $G = \mathbb{R}^d$, and let α be an action of G on a Banach space B. Let $b \in B^{\infty}$. Then

$$\alpha(f)(D_X b) = \lim_{t \to 0} \int_G f(x)\alpha(x) \left(\frac{\alpha(\exp(tX))(b) - b}{t}\right) dx$$
$$= \lim_{t \to 0} \int_G \frac{(f(x\exp(-tX)) - f(x))\alpha(x)(b)}{t} dx = \alpha(D_X f)(b)$$

where we used the fact that G is abelian, hence unimodular. (This formula is therefore true for any unimodular group.) The Fourier transform of b is given by

$$(D_X b)_n = \alpha(e_n)(D_X b) = -\alpha(D_X e_n)(b) = 2\pi i n X b_n$$

where

$$e_n(t) = e^{2\pi i n t},$$

and the multiplication of \mathbb{Z}^d and \mathbb{R}^d is given by the dot product.

Let the Laplacian Δ act on B^{∞} by

$$(\Delta b)_n = \sum_j (2\pi)^2 (nE_j)^2 b_n$$

where the E_j form a basis for \mathbb{R}^d . Then for $k \in \mathbb{N}$,

$$((1+\Delta)^k b)_n = \left(1 + (2\pi)^2 \sum_j (nE_j)^2\right)^k b_n$$

 \mathbf{SO}

$$b_n = \frac{((1+\Delta)^k b)_n}{(1+(2\pi)^2 \sum_j (nE_j)^2)^k}$$

whence

$$||b_n|| \le \frac{||(1+\Delta)^k b||}{(1+(2\pi)^2 \sum_j (nE_j)^2)^k}$$

Therefore if p is a polynomial on \mathbb{Z}^d ,

$$p(n)||b_n|| \le \frac{|p(n)| \cdot ||(1+\Delta)^k b||}{(1+(2\pi)^2 \sum_j (nE_j)^2)^k}$$

and if k is large enough, it follows that $n \mapsto |p(n)|||b_n||$ is bounded (since $(1 + \Delta)^k b$ is independent of n). Since p can grow arbitrarily fast, the function $n \mapsto ||b_n||$ lies in $C_{\infty}(\mathbb{Z}^d)$. In fact it lies in the Schwartz space of $C_{\infty}(\mathbb{Z}^d)$.

Theorem 14.11. Let $b \in B$. Then $b \in B^{\infty}$ if and only if $n \mapsto ||b_n||$ is a Schwartz function on \mathbb{Z}^d .

Example 14.12. Recall that $A_{\Theta} = C^*(\mathbb{Z}^d, c_{\Theta})$. Then $a_n \in \mathbb{C}$ and c_{Θ} is an action of T^d , so A_{Θ}^{∞} is isomorphic to the Schwartz space of \mathbb{Z}^d .

We now introduce noncommutative differential forms. Given $a \in A^{\infty}$, let $da : \text{Lie } G \to A$ be given by

$$da(X) = \alpha(X)(a).$$

Then d is a derivation. We let $\tilde{\Omega}$ be the space of linear maps Lie $G \to A$, viewed as a (A, A)-bimodule. Then let Ω be the submodule generated by d; i.e. linear combinations of elements of the form a db, i.e. 1-forms on A.

Definition 14.13. Let A be a C^{*}-algebra. The (A, A)-bimodule Ω is known as the noncommutative cotangent bundle of A.

From this it is not difficult to define the higher exterior power Ω^k and define the boundary map

$$d: \Omega^k \to \Omega^{k+1}.$$

14.4 Noncommutative vector bundles

Let X be a compact Hausdorff space, E a vector bundle over X, and let $\Gamma(E)$ be the vector space of continuous sections of E. Given $\xi \in \Gamma(E)$ and $f \in C(X)$, $(f\xi)(x) = f(x)\xi(x)$ by scalar multiplication so $\Gamma(E)$ is a C(X)-module.

In this section we will assume all C(X)-modules are finitely generated.

Definition 14.14. Let R be a unital ring. A *projective module* over R is a R-module V which is isomorphic to a direct summand of a free R-module.

In other words, if V is free then there is a R-module W and a free R-module F such that $V \oplus W \cong F$.

Theorem 14.15 (Swan). Let X be a compact Hausdorff space. A C(X)-module V is projective if and only if there is a vector bundle E such that $V \cong \Gamma(E)$. Moreover, we have an isomorphism of vector bundles $E \cong F$ if and only if $\Gamma(E) \cong \Gamma(F)$.

So we have an equivalence of categories relating projective C(X)-modules and vector bundles over X. Though Swan proved this result in 1962, by this time Grothendieck had already started identifying projective modules with vector bundles over algebraic varieties.

Let R be a ring. If we view R^n as a right R-module, then $\operatorname{End}_R(R^n) \cong M_n(R)$, where $M_n(R)$ is viewed as acting on R^n from the left. So we usually will view R^n as a right R-module. Henceforth we assume that every ring acts on its modules from the right.

If V is a projective module which appears as a direct summand in \mathbb{R}^n , then there is a projection $P \in \operatorname{End}_{\mathbb{R}}(\mathbb{R}^n)$ such that $P(\mathbb{R}^n) \cong V$. This is not a bijection between projections and projective modules, but it is often useful. Indeed, for any projection P, $P(\mathbb{R}^n)$ is a projective \mathbb{R} -module.

Example 14.16. Let R = C(X), $P \in M_n(R)$ a projection. Then P acts on $\mathbb{R}^n = C(X \to \mathbb{C}^n)$. So $P(\mathbb{R}^n)$ is a projective module, and we can find the vector bundle from Swan's theorem by looking at its localizations.

If R is a unital ring, let S(R) be the space of isomorphism classes of projective R-modules. For V, W isomorphism classes, define $V + W = V \oplus W$. Then S(R) is an abelian monoid, and S is a functor from unital rings to abelian monoids. But S(R) is badly behaved because $V \oplus W \cong V' \oplus W$ does not imply $V \cong V'$.

Example 14.17. Let T be the circle, $A = C(T \to \mathbb{R})$. So A consists of periodic functions $\mathbb{R} \to \mathbb{R}$ which are continuous of period 1. Let Ξ_n^{\pm} be the set of continuous $\xi : \mathbb{R} \to \mathbb{R}$ such that $\xi(t+n) = \pm \xi(t)$. Then

$$S(R) = \{ \Xi_n^{\pm} : n \in \mathbb{Z} \}.$$

From the space Ξ_1^- we can recover the Moebius strip.

Example 14.18. Let $A = C(T^2)$, viewed as continuous functions $f : \mathbb{R}^2 \to \mathbb{C}$ which are periodic of period 1 in both variables. Now let $\Xi_{m,q}$ be the space of $\xi \in C(\mathbb{R}^2 \to \mathbb{C})$ such that $\xi(s+1,t) = \xi(s,t)$ and

$$\xi(s, t+m) = e^{2\pi i q s} \xi(s, t).$$

Together with the free A-modules, we recover all projective modules over A, i.e. all \mathbb{C} -vector bundles over T^2 . So we have classified vector bundles on a torus.

Example 14.19. Let S^2 be the 2-sphere, viewed as the unit sphere of \mathbb{R}^3 , and let $A = C(S^2 \to \mathbb{R})$. Let Ξ be the A-module of continuous sections of the tangent bundle TS^2 of S^2 . This can be viewed as the set of $\xi \in C(S^2 \to \mathbb{R}^3)$ such that for all $x \in S^2$, $\langle \xi(x), x \rangle = 0$. By the hairy ball theorem, TS^2 is a nontrivial bundle. But the normal bundle NS^2 , whose continuous sections consist of $\xi : S^2 \to \mathbb{R}^3$ such that $\xi(x) \in \mathbb{R}x$, is isomorphic to $S^2 \times \mathbb{R}$, so is a trivial bundle. Therefore, if S(A) was a cancellative monoid, then

$$\Xi \oplus A \cong \Xi \oplus NS^2 \cong A^3 \cong A^2 \cong A$$

so we could conclue that $\Xi \cong A^2$ and hence TS^2 is trivial, a contradiction. Therefore S(A) is noncancellative, and constructing its Grothendieck group will be quite difficult.

Let C(R) be the universal cancellative abelian monoid for S(R). In other words, for every V, V' for which there exists W with $V \oplus W \cong V' \oplus W$, we impose the relation V = V'. We then take the universal abelian group containing C(R), say $K_0(R)$, i.e. the Grothendieck group¹ of S(R). (Constructively, elements of $K_0(R)$ are pairs (V, W) where V, W are elements of C(R), and we are thinking of (V, W) as meaning V - W.) We define the positive elements of $K_0(R)$ to be those in C(R).

¹According to Rieffel, Grothendieck does not deserve to have such a trivial construction named after him.

In fact one constructs K-theoretic groups $K_n(R)$ for every $n \in \mathbb{N}$. The first groups $n \in \{0, 1, 2\}$ were well-known previously, but Quillen introduced K-theoretic groups for any ring and any natural number. But we are not interested in Quillen's K-theoretic groups. We introduce, for R a Banach algebra, the topological K-theoretic group $K_1^{top}(R)$. If one tries to define $K_2^{top}(R)$ over \mathbb{C} we find $K_2^{top}(R) = K_0(R)$; similarly over R we have $K_8^{top}(R) = K_0(R)$. This is the Bott periodicity theorem.

Example 14.20. Let $A_{\theta} = C^*(\mathbb{Z}^2, c_{\theta})$ be the quantum 2-torus. Then $A_0 = C(T^2)$. In case of the 2-torus, $A_{\theta} = C(T) \rtimes_{\alpha_{\theta}} \mathbb{Z}$ acts on $L^2(T)$. We have projections on A_{θ} provided that $\theta \in (0, 1)$, namely

$$P = U_{-1}M_h + M_f + M_q U_1$$

for certain multiplication operators M_f, M_g, M_h . If $\varepsilon > 0$ and $\theta + \varepsilon < 1$, then the function f is supported on $[0, \theta + \varepsilon]$ and identically 1 on $[\varepsilon, \theta]$ and the trace t(P) of the projection is given by

$$t(P) = \int_0^1 f = \theta.$$

If there is a unitary equivalence $P \cong P'$, then t(P') = t(P). By a theorem on the homework, there are only countably many projections in a separable C^* -algebra up to unitary equivalence, so A_{θ} contains countably many traces, and its set of traces is determined by θ . Yet there are uncountably many choices of θ . So *Rieffel's theorem* says that there are uncountably many quantum tori up to C^* -isomorphism. In fact $(\mathbb{Z} + \mathbb{Z}\theta) \cap [0, 1]$ indexes the quantum tori that embed in A_{θ} . But when Rieffel showed this result to Voicolescu, he proved *Voicolescu's theorem*, which shows that

$$K_0(A_\theta) \cong \mathbb{Z}^2$$

where $C(A_{\theta}) = (Z + Z\theta) \cap [0, \infty)$. So the K-theoretic group is not a complete invariant of the quantum tori, but the positive elements of the K-theoretic group give more information.

Part IV Complex analysis

Chapter 15

Holomorphy in several complex variables

This chapter follows Hormander's SCV book, Chapter II, and Zworski's lectures on SCV.

15.1 Cauchy-Riemann equations

Let us generalize the Cauchy-Riemann equations to higher dimensions.

Definition 15.1. Let $f : \mathbb{C}^n \to \mathbb{C}$ be a function. We write z = x + iy and define the partial derivatives

$$\frac{\partial f}{\partial z_j} = \frac{1}{2} \left(\frac{\partial f}{\partial x_j} - i \frac{\partial f}{\partial y_j} \right)$$

and

$$\frac{\partial f}{\partial \overline{z}_j} = \frac{1}{2} \left(\frac{\partial f}{\partial x_j} + i \frac{\partial f}{\partial y_j} \right).$$

We define the Wirtinger differential of f by $\partial f = \sum_j \partial_{z_j} f dz_j$ and $\overline{\partial} f = \sum_j \partial_{\overline{z}_j} f d\overline{z}_j$. Finally, we define the total differential $df = \partial f + \overline{\partial} f$.

For ease of notation we frequently make the decomposition $dz_j = dx_j + idy_j$ and $d\overline{z}_j = dx_j - idy_j$. Then $df = \sum_j \partial_{x_j} f \, dx_j + \partial_{y_j} f \, dy_j$, as it should be.

Notice that if n = 1 and f is holomorphic, then $\overline{\partial} f = 0$. Indeed, f = u + iv solves the Cauchy-Riemann equations, so

$$\frac{\partial f}{\partial z} = \frac{1}{2} \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} + \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) = 0.$$

This motivates the general definition of holomorphy.

Definition 15.2. The *Cauchy-Riemann equation* is the equation

$$\overline{\partial}f = 0$$

If $f : \mathbb{C}^n \to \mathbb{C}$ solves the Cauchy-Riemann equation, then f is said to be a holomorphic function of several complex variables. If $f = (f_1, \ldots, f_m)$ is a function $\mathbb{C}^n \to \mathbb{C}^m$ such that each f_j is holomorphic, then f itself is said to be holomorphic.

It is easy to check that the composite of holomorphic functions is holomorphic.

Recall that the implicit function theorem guarantees that a C^r relation between \mathbb{R}^n and \mathbb{R}^m that "passes the vertical hyperplane test" is actually the graph of a C^r function $\mathbb{R}^n \to \mathbb{R}^m$. In particular, this holds if $r = \infty$, but demanding holomorphy of the function is actually a much stronger condition, so we must check that it holds.

Theorem 15.3 (implicit function theorem). Let U be a neighborhood of $(w_0, z_0) \in \mathbb{C}^m \times \mathbb{C}^n$, and let $f: U \to \mathbb{C}^m$ be holomorphic. Suppose that $f(w_0, z_0) = 0$ and $\det(\partial f_j / \partial w_k)_{j,k=1}^m \neq 0$. Then there is a unique holomorphic function $g: \mathbb{C}^n \to \mathbb{C}^m$ such that f(g(z), z) = 0 and $g(z_0) = w_0$.

Proof. By replacing \mathbb{C} with \mathbb{R}^2 , we can apply the classical implicit function theorem. To do this, we write f = u + iv and consider the Jacobian

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{bmatrix} \operatorname{Re} \partial f & -\operatorname{Im} \partial f \\ \operatorname{Im} \partial f & \operatorname{Re} \partial f. \end{bmatrix}$$

One easily checks that the determinant of this matrix is $|\det(\partial f_j/\partial w_k)|_{j,k=1}^m$ which is nonzero. This gives a function g with f(g(z), z) = 0.

To prove holomorphy, we apply ∂ . This is

$$\overline{\partial}_k f_j(g(z), z) = \sum_{\ell} \partial_{w_\ell} f_j \overline{\partial} g_\ell(z)$$

by the chain rule, using that $\overline{\partial} f = 0$. We have $\overline{\partial}_k f_j(g(z), z) = 0$, so by linear algebra, $\overline{\partial} g_\ell = 0$. So g is holomorphic.

Corollary 15.4 (inverse function theorem). Let $z_0 \in \mathbb{C}^m$ and $f : \mathbb{C}^m \to \mathbb{C}^m$ be a holomorphic function whose Jacobian does not vanish at z_0 . Then there is a neighborhood $U \ni z_0$ such that f is a holomorphic diffeomorphism of U into its image.

Recall that if $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multiindex, then the differential form dx_α is given by

$$dx_{\alpha} = \bigwedge_{j=1}^{n} dx_{\alpha_j}$$

Definition 15.5. A differential form ω is type (p,q) if it can be written

$$\omega = \sum_{|\alpha|=p} \sum_{|\beta|=q} f_{\alpha,\beta} dz_{\alpha} d\overline{z}_{\beta}$$

For each function space \mathcal{F} , we let $\mathcal{F}_{p,q}$ denote the space of differential forms of type (p,q) over \mathcal{F} .

15.2 Basic properties

Theorem 15.6 (Hartogs). Let $u : \Omega \to \mathbb{C}$ be holomorphic in each variable alone; then u is holomorphic.

Allegedly this theorem is useless (though it comes up in dynamical systems). Its proof is very difficult, using the Baire category theorem and the Schwarz lemma. However, if we are allowed to assume that u is continuous, then the proof is almost trivial.

Lemma 15.7. Assume that we are given a sequence u_j of uniformly bounded subharmonic functions, such that $\limsup_j u_j$ is bounded from above. Then the u_j are locally uniformly bounded from above by $\sup \limsup_j u_j + \varepsilon$.

Proof. Without loss of generality, we may assume that the $u_j \leq 0$. Let K be a compact set with $d(K, \Omega^c) \geq 3r$. Let $z \in K$, $\varepsilon > 0$. Then we can find a n_0 such that for any $n > n_0$,

$$\int_{|z'-z| < r} u_n(z') \, dz' \le (C + \frac{\varepsilon}{2})\pi r^2$$

by Fatou's lemma. If δ is small enough and $|w - z| < \delta$, then

$$\pi(r+\delta)^2 u_n(w) \le \int_{|z'-z| \le r+\delta} u_n(z') \, dz' \le \int_{|z'-z| < \delta} u_n(z') \, dz' \le (C+\varepsilon/2)\pi r^2$$

for n large enough. Therefore

$$u_n(w) \le (C + \varepsilon/2) \left(\frac{r+\delta}{r}\right)^2 \le C + \varepsilon$$

for δ small enough. This does not depend on z, w.

Cover K by discs $D(z, \delta)$, so reduce to a finite subcover to find a uniform δ .

Definition 15.8. A *polydisk* in \mathbb{C}^n is a set D of the form

$$D = \prod_{j=1}^{n} D_j$$

where each $D_j \subset \mathbb{C}$ is an open disk. The *distinguished boundary* $\partial_0 D$ is the set

$$\partial_0 D = \prod_{j=1}^n \partial D_j.$$

If $u : \overline{D} \to \mathbb{C}^m$ is a continuous function which is holomorphic on D, we simply say u is holomorphic on \overline{D} .

Notice that $\partial_0 D$ is in general a torus (since it is a product of circles) and a very small subset of ∂D .

By induction on n, one easily proves the following.

Theorem 15.9 (Cauchy's integral formula in a polydisk). Let D be a polydisk in \mathbb{C}^n and let $u: \overline{D} \to \mathbb{C}$ is holomorphic in each variable separately and continuous on \overline{D} . Then one has

$$u(z) = \frac{1}{(2\pi i)^n} \int_{\partial_0 D} \frac{u(\zeta) \, d\zeta}{(\zeta_1 - z_1) \cdots (\zeta_n - z_n)}$$

Corollary 15.10. Let $U \subseteq \mathbb{C}^n$ be an open set and $f: U \to \mathbb{C}^m$ be a holomorphic function. Then $f \in C^{\infty}(U)$, and for every compact set $K \subset U$, every multiindex α , and every open neighborhood V of K, we have

$$||\partial^{\alpha}u||_{L^{\infty}(K)} \preceq_{K,\alpha} ||u||_{L^{1}(V)}.$$

Proof. Since K is compact, it can be covered by finitely many sets contained in polydiscs contained in V. Now use Cauchy's integral formula (the implicit constant arising from the denominator of the formula, the measures of the distinguished boundary, and the obligatory factors of 2π).

Corollary 15.11 (Montel). Let $(u_k)_k$ be a sequence of holomorphic functions on an open set $U \subseteq \mathbb{C}^n$. If $u_k \to u$ locally uniformly, then u is holomorphic. On the other hand, if one has

$$||u_k||_{L^{\infty}(K)} \preceq_K 1$$

for $K \subset U$ compact, then $(u_k)_k$ has a locally uniformly convergent subsequence (which, in particular, has a holomorphic limit).

The proof is the same as in one variable.

We view $\overline{\partial}$ as an exterior derivative. For (0, 1)-forms f we have

$$\overline{\partial}f = \sum_{j < k} (\overline{\partial}_j f_k - \overline{\partial}_k f_j) d\overline{z}_j \wedge d\overline{z}_k$$

and if $\overline{\partial} u = f$ for some (0, 0)-form $u, \overline{\partial} f = 0$.

Theorem 15.12. Let $\Omega \subset \mathbb{C}^n$ be bounded, n > 1, and assume that $\mathbb{C}^n \setminus \Omega$ is connected. If there is a $\rho \in C^4(\mathbb{C}^n \to \mathbb{R})$ such that $\partial\Omega$ is the zero set of ρ and $d\rho|_{\partial\Omega} \neq 0$, and there is a $u \in C^4(\overline{\Omega})$ such that $\overline{\partial}u \wedge \overline{\partial}\rho = 0$ on $\partial\Omega$, then there is a $U \in A(\Omega) \cap C^1(\overline{\Omega})$ such that U = uon $\partial\Omega$.

Note that if U exists, then $U - u = \rho h$ for some $h \in C^1(\overline{\Omega})$, with $-\overline{\partial}u = \overline{\partial}\rho h$ on $\partial\Omega$. This is what we mean by $\overline{\partial}u \wedge \overline{\partial}\rho = 0$ on $\partial\Omega$, which is hence a necessary condition as well as sufficient.

Definition 15.13. For ρ as above, the equation $\overline{\partial}u \wedge \overline{\partial}\rho = 0$ is called the *tangential Cauchy-Riemann equation*.

Clearly if u is holomorphic then it solves the tangential Cauchy-Riemann equations. Otherwise, what this condition is saying is that $\overline{\partial}u$ and $\overline{\partial}\rho$ are proportional. The intuition is that if $\sum_j t_j \overline{\partial}_j \rho = 0$ then $\sum_j t_j \overline{\partial}_j u = 0$, and the hypothesis here is that the vector field $\sum_j t_j \overline{\partial}_j$ is tangent (in the sense of the complexified tangent bundle) to $\partial\Omega$. Thus, what the tangential Cauchy-Riemann equations say is that any section V of the tangent bundle of $\partial\Omega$ which only contain antiholomorphic coordinates has Vu = 0, so annihilates u. **Example 15.14.** Take the unit ball B of \mathbb{C}^2 . Here $\rho(z) = |z|^2 - 1$. If $V = t_1\overline{\partial}_1 + t_2\overline{\partial}_2$ then $V\rho(z) = 2t \cdot z$. Thus the antiholomorphic vector fields which are tangent to ∂B are elements of the ideal generated by $z_2\overline{\partial}_1 - z_1\overline{\partial}_2$.

Proof of theorem. We construct $U_0 \in C^1(\overline{\Omega})$ such that $U_0 = u$ on $\partial\Omega$ and $\overline{\partial}U_0 = O(\rho^2)$ on $\partial\Omega$. By the tangential Cauchy-Riemann equations, we have a $h_0 \in C^3$ such that $\overline{\partial}u = h_0\overline{\partial}\rho$ on $\partial\Omega$. Thus $\overline{\partial}u = h_0\overline{\partial}\rho + \rho h_1$ for some $h_1 \in C^3_{0,1}$. Then $\overline{\partial}(u - h_0\rho) = \overline{\partial}u - h_0\overline{\partial}\rho - \overline{\partial}h_0\rho = \rho h_1 - \overline{\partial}h_0\rho = \rho(h_1 - \overline{\partial}h_0) = \overline{\partial}\rho h_2$ for $h_2 \in C^2_{0,1}$, and $0 = \overline{\partial}(\rho h_2) = \overline{\partial}\rho \wedge h_2 + \rho\overline{\partial}h_2$. Now $\overline{\partial}h_2 = 0$ on $\partial\Omega$ so $\overline{\partial}\rho \wedge h_2 = 0$ on $\partial\Omega$, i.e. $h_2 = h_3\overline{\partial}\rho + \rho h_4$ for some $h_3 \in C^2$ and $h_4 \in C^2_{0,1}$. Now let $h_5 = -h_3/2$, then $2\overline{\partial}\rho h_5 = -h_2 + O(\rho)$, so

$$\overline{\partial}(u-h_0\rho+h_5\rho^2) = \rho(h_3\overline{\partial}\rho+\rho h_4) - \rho^2\overline{\partial}h_3/2 + 2\rho\overline{\partial}\rho(-h_3/2) = \rho^2h_4 - \rho^2\overline{\partial}h_3/2 = O(\rho^2).$$

Thus let $U_0 = u - h_0 \rho + h_5 \rho^2$. So there is an $f \in C^1$ such that $\overline{\partial} U_0 = \rho^2 f$.

Now let $F = \rho^2 f$ on $\overline{\Omega}$ and 0 away from Ω . Then $F \in C^1(\mathbb{C})$. So there is a $v \in C^1_{comp}$ such that $\overline{\partial}v = F$. By continuity, v = 0 on $\partial\Omega$. Let $U = U_0 - v$.

15.3 Plurisubharmonicity and domains of holomorphy

Definition 15.15. Let $\Omega \subseteq \mathbb{C}^n$ be open, and $u : \Omega \to [-\infty, \infty)$ be a upper-semicontinuous function. We say u is a *plurisubharmonic function* or *plush function* if for every $a, b \in \mathbb{C}^n$, the function $z \mapsto u(az + b)$ is subharmonic whenever it is defined.

Lemma 15.16. If u is C^2 , then u is plush iff the Hessian matrix

$$\lambda_{ij} = \partial_i \overline{\partial}_j f$$

is positive-semidefinite everywhere.

Proof. We have $0 \leq \Delta_{\tau} u(z + \tau w) = 4 \partial_{\tau} \overline{\partial}_{\tau} u(z + \tau w)$. By the chain rule,

$$0 \le 4\partial_\tau \sum_j \overline{\partial}_j u(z+\tau w) \overline{w}_j = 4 \sum_{k,j} \partial_k \overline{\partial}_j u(z+\tau w) \overline{w}_j w_k.$$

Now divide both sides by 4.

We let $P(\Omega)$ denote the set of plush functions on Ω . It is easy to see from this characterization that for any holomorphic function f, $\log |f| \in P(\Omega)$. A tensor product of subharmonic functions, $u \otimes v(z, w) = u(z)v(w)$ is also plush. Any convex function is plush.

Recall that the decreasing sequence of a subharmonic functions converges to a subharmonic function. Restricting to a line we see:

Corollary 15.17. The limit of a decreasing sequence of plush functions is plush.

Thus we could define a plush function to be a decreasing limit of smooth plush functions (where a smooth plush function is one satisfying the Hessian characterization.)

Corollary 15.18. Let φ be a standard mollifier and u plush. Let u_{ε} denote the ε -mollification of u. Then u_{ε} decreases to u as $\varepsilon \to 0$ and $u_{\varepsilon} \in C^{\infty}(\Omega_{\varepsilon}) \cap P(\Omega_{\varepsilon})$ for $\Omega_{\varepsilon} = \{z \in \Omega : d(z, \partial\Omega) > \varepsilon\}$.

The proof is the same as in one dimension (i.e. for subharmonic functions).

Corollary 15.19. Let $\Omega \subseteq \mathbb{C}^n$, $\Omega' \subseteq \mathbb{C}^m$, $f : \Omega \to \Omega'$ a holomorphic mapping, and $u \in P(\Omega')$. Then the pullback $f^*u \in P(\Omega)$.

Proof. Without loss of generality we can assume that u is smooth (since the pullback of a decreasing sequence is decreasing). By the Hessian characterization, the pullback is plush.

Definition 15.20. An open set $\Omega \subseteq \mathbb{C}^n$ is a *domain of holomorphy* if there are no $\Omega_1, \Omega_2 \subset \mathbb{C}^n$, Ω_1 is nonempty, $\Omega_1 \subset \Omega_2 \cap \Omega$, Ω_2 is not contained in Ω_1 , and for every $u \in A(\Omega)$ there is a $\tilde{u} \in A(\Omega_2)$ such that $u|_{\Omega_1} = u_1|_{\Omega_1}$.

For example, a polydisk $D = \prod_j D_j$ is a domain of holomorphy because we could always find $f_j \in A(D_j)$ which cannot extend to any open set beyond D_j , and then $f(z) = \sum_j f_j(z_j)$ cannot be extended to any open set. More generally, a product of domains of holomorphy is a domain of holomorphy. If n = 1 then every set is a domain of holomorphy, so this is the distinction between one and many variables.

We choose a function $\delta : \mathbb{C}^n \to [0, \infty)$ such that $\delta(z) > 0$ for $z \neq 0$, $\delta(tz) = |t|\delta(z)$ for $\delta \in \mathbb{C}$. For example, $\delta(z) = |z|$, or $\delta(z) = \max_j |z_j| r_j^{-1}$ for $r_j > 0$. Put

$$\delta(z,\Omega) = \inf_{w \in \Omega} \delta(z-w).$$

Recall that for $K \subset \Omega$ a compact set, the holomorphic hull \hat{K}_{Ω} is defined by the set of $z \in \Omega$ such that for every $f \in A(\Omega), |f(z)| \leq ||f||_{L^{\infty}(K)}$.

Lemma 15.21. Suppose Ω is a domain of holomorphy, $f \in A(\Omega)$, and K compactly contained in Ω . If for every $z \in K$,

$$|f(z)| \le \delta(z, \mathbb{C}^n \setminus \Omega),$$

then this estimate is also true for $z \in \hat{K}_{\Omega}$, the holomorphic hull of K. In particular,

$$f(z) = \inf_{\substack{w \in \mathbb{C} \setminus \Omega \\ z \in K}} \delta(z - w) = \inf_{\substack{w \in \mathbb{C} \setminus \Omega \\ z \in \hat{K}_{\omega}}} \delta(z - w).$$

Proof. For $D = \{|z_j| < r_j\}$ a polydisc, let

$$\Delta_{\Omega}^{D}(z) = \sup_{z+rD \subseteq \Omega} r = \delta(z, \mathbb{C}^{n} \setminus \Omega)$$

where $\delta(z) = \max_j |z_j| r_j^{-1}$.

Suppose $f \in A(\Omega)$ and $|f(z)| \leq \Delta_{\Omega}^{D}(z)$ for $z \in K$. We claim that for each $\zeta \in \hat{K}_{\Omega}$ and $u \in A(\Omega)$,

$$u(z) = \sum_{\alpha} \frac{(z-\zeta)^{\alpha}}{\alpha!} \partial^{\alpha} u(\zeta)$$

on $\zeta + |f(\zeta)|D$. Since Ω is a domain of holomorphy, this Taylor series cannot converge on a large enough polydisc, so $|f(\zeta)| \leq \Delta_{\Omega}^{D}(z)$, which proves the claim in case Ω itself is a polydisc.

To prove the claim, let $L_t = \{z \in \Omega : \exists w \in K | z_j - w_j | \leq tr_j | f(w) |\}$. If t < 1, then L_t is compact. In fact, if $\{z^k\}_j$ is a sequence in L_t then there is a sequence of w^k witnessing that $z^k \in L_t$. Since K is compact we can choose a limit w of the w^k . Thus for k large,

$$|z_j^k - w_j| \le tr_j |f(w)| + \varepsilon$$

Taking ε small, we see that the z^k lie in a compact set in K since $|z_j^k - w_j^k| \leq \Delta_{\Omega}^D(w)$, so have a limit $z \in K$.

Let $M_t = ||u||_{L^{\infty}(L_t)}$. By the Cauchy inequality applied to the polydisc $\zeta + |f(\zeta)|D$,

$$|\partial^{\alpha} u(w)|t^{|\alpha|} \frac{r^{\alpha}}{\alpha!} |f(w)|^{|\alpha|} \le M_t$$

Let $F(\zeta) = \partial^{\alpha} u(\zeta) t^{|\alpha|} r^{\alpha} (\alpha!)^{-1} f(\zeta)^{|\alpha|}$. Then F is holomorphic and $|F(\zeta)| \leq M_t$ for $\zeta \in K$. So by the definition of \hat{K}_{Ω} , $||F||_{L^{\infty}(\hat{K}_{\Omega})} \leq M_t$. Therefore

$$\frac{|\partial^{\alpha} u(\zeta)|}{\alpha!} \le M_t t^{|\alpha|} r^{\alpha} |f(\zeta)|^{|\alpha|}$$

so if $|z_j - \zeta_j| \leq tr_j |f(\zeta)|$ then the Taylor series converges.

The above argument proves the lemma when $\delta(z) = \max_j |z_j| r_j^{-1}$. The general δ satisfies

$$\delta(z, \mathbb{C}^n \setminus \Omega) = \sup\{r \in R : \forall w \in \mathbb{C} \ \delta(w) \le 1 \implies \forall a \in B(0, r) \ z + aw \in \Omega\}.$$

Thus $\delta(z, \mathbb{C}^n \setminus \Omega) = \inf_{\delta(w) \leq 1} \delta_w(z, \mathbb{C}^n \setminus \Omega)$ where $\delta_w(z, \mathbb{C}^n \setminus \Omega) = \sup_{\forall a \ z \in aw \in \Omega} r$ where a ranges over B(0, r). Take $w = (1, 0, \dots, 0)$ and $D_k = \{z : |z_1| < 1, j \neq 1 \implies |z_j| < 1/k\}$. Thus $\Delta_{\Omega}^{D_k}(z) \to \delta_w(z, \mathbb{C}^n \setminus \Omega)$ as $k \to \infty$. Moreover $\Delta_{\Omega}^{D_{k+1}}(z) \geq \Delta_{\Omega}^{D_k}(z)$ so by Dini's theorem the convergence of the $\Delta_{\Omega}^{D_k}$ is uniform.

If $|f(z)| \leq \delta_w(z, \mathbb{C}^n \setminus \tilde{\Omega})$ then $|f(z)| \leq (1 + \varepsilon)\Delta_{\Omega}^{D_k}(z)$ for k large enough and $z \in K$. By the lemma, $|f(z)| \leq (1 + \varepsilon)\Delta_{\Omega}^{D_k}(z)$ and $z \in \hat{K}_{\Omega}$. By Dini's theorem again one has

$$|f(z)| \le \delta_w(z, \mathbb{C}^n \setminus \Omega).$$

Take the infimum over w of both sides. So $|f(z)| \leq \delta(z, \mathbb{C}^n \setminus \Omega)$ even for "weird" choices of δ .

Theorem 15.22. The following are equivalent:

- 1. Ω is a domain of holomorphy.
- 2. For every compact subset K of Ω , the holomorphic hull \hat{K}_{Ω} is compactly contained in Ω , and for every $f \in A(\Omega)$,

$$\sup_{K} \frac{|f(z)|}{\delta(z, \mathbb{C}^n \setminus \Omega)} = \sup_{\hat{K}_{\Omega}} \frac{|f(z)|}{\delta(z, \mathbb{C}^n \setminus \Omega)}$$

- 3. For every compact subset K of Ω , the holomorphic hull K_{Ω} is compactly contained in Ω .
- 4. There is a $f \in A(\Omega)$ which does not extend to any larger set.

Proof. Obviously 2 implies 3 and 4 implies 1. By the lemma above, 1 implies 2. So we just need to show 3 implies 4.

Let $\{K_j\}$ a compact exhaustion of Ω (so for any compact $L \subset \Omega$ and every j large enough $L \subseteq K_j$). Define $D_{\zeta} = \{\zeta\} + rD$ for $D = \{|z_j| < r_j\}$ and $r = \sup_{\{\zeta\} + \rho D \subset \Omega} \rho$.

Let M be the set of rational points of Ω ; then we choose a sequence of $\zeta_j \in M$ so that every element of M appears infinitely often. We also choose a sequence of $z_j \in D_{\zeta_j}$ such that $z_j \notin K_j$; this is possible because D_{ζ_j} touches $\partial \Omega$ but K_j does not because it is compact.

By (3), there is a $f_j \in A(\Omega)$ such that $f_j(z_j) = 1$ and $||f_j||_{L^{\infty}(K_j)} < 1$. Replacing f_j by a power of f_j , we can assume without loss of generality that $||f_j|_{L^{\infty}(K_j)} < 2^{-j}$.

Now let

$$f(z) = \prod_{j} (1 - f_j(z))^j.$$

To see that this product converges we just have to show convergence in any compact set L, and we can assume without loss of generality that there is a J such that $L = K_J$. For j > Jwe have $|f_j(z)| < 2^{-j}$ and since we only care about the tail we can assume J = 1. Then

$$\log f(z) = \sum_{j} j |\log(1 - f_j(z))| \le \sum_{j} j |f_j(z)| \le \sum_{j} j 2^{-j} < \infty.$$

Therefore convergence is locally uniform so $f \in A(\Omega)$.

For every D_{ζ_j} there is a w_N such that for every $|\alpha| \leq N$, $\partial^{\alpha} f(z_N) = 0$. Therefore f has a zero of order N in D_{ζ_j} . In particular, the zeroes of f have higher and higher order as we approach $\partial\Omega$. Therefore if f is defined at a point z of $\partial\Omega$ then z is an infinite-order zero of f. So f = 0. But f is nonzero so this is a contradiction. \Box

Example 15.23. As a counterexample, notice that if $\Omega = B(0,3) \setminus \overline{B(0,1)}$ and $K = \partial B(0,2)$ then $\hat{K}_{\Omega} = \overline{B(0,2)} \setminus \overline{B(0,1)}$ by Hartogs' theorem and the maximum principle for $n \ge 2$. This is not a compact subset of Ω , so Ω is not a domain of holomorphy.

Corollary 15.24. If Ω is convex then Ω is a domain of holomorphy.

Proof. Recall that \hat{K}_{Ω} is contained in the convex hull of K, which is a compact subset of Ω .

Corollary 15.25. If Ω_{α} are domains of holomorphy then the interior of $\bigcap_{\alpha} \Omega_{\alpha}$ is a domain of holomorphy.

Proof. Let $K \subset \Omega$ be a compact set. Then \hat{K}_{Ω} is a compact subset of $\hat{K}_{\Omega_{\alpha}}$ for every α , in particular of the compact set $\bigcap_{\alpha} \hat{K}_{\Omega_{\alpha}}$, which is a compact subset of Ω .

Corollary 15.26. Let Ω be a domain of holomorphy, $f_1, \ldots, f_N \in A(\Omega)$. Then

$$\Omega_f = \{ z \in \Omega : |f_j(z)| < 1 \}$$

is a domain of holomorphy.

Proof. Let K be a compact subset of Ω_f . Then by compactness, there is a r < 1 such that $K \subset \{z \in \Omega : |f_j(z)| < r\}$. Thus for any $z \in \hat{K}_{\Omega_f}, |f_j(z)| \le r$. Moreover

$$\hat{K}_{\Omega_f} \subseteq \hat{K}_{\Omega} \subset \{|f_j| \le r\} \subseteq \Omega_f$$

Therefore K_{Ω_f} is a compact subset of Ω_f .

Corollary 15.27. Let $u: \Omega \to \mathbb{C}^m$ be a holomorphic transformation, $\Omega' \subseteq \mathbb{C}^m$ a domain of holomorphy. Then the pullback $u^{-1}(\Omega')$ is a domain of holomorphy.

We now relate domains of holomorphy to plurisubharmonicity.

Theorem 15.28. If Ω is a domain of holomorphy and δ as above, then $z \mapsto -\log \delta(z, \mathbb{C}^n \setminus \Omega)$ is a continuous plurisubharmonic function.

Proof. For $z_0 \in \Omega$, $w \in \mathbb{C}^n$, take $D = \{z_0 + \tau w : \tau \in \mathbb{C}, |\tau| \leq r\}$. If r is small enough then $D \subseteq \Omega$. Let f be a polynomial in τ and $-\log \delta(z_0 + \tau w, \mathbb{C}^n \Omega) \leq \operatorname{Re} f(\tau)$ for every $|\tau| = r$. We must show this is true for $|\tau| < r$ as well.

Let F be a polynomial on \mathbb{C}^n such that $F(z_0 + \tau w) = f(\tau)$. Then

$$|e^{-F(z)}| \le \delta(z, \mathbb{C}^n \setminus \Omega)$$

for $z \in \partial D$, hence for $z \in \widehat{\partial D}_{\Omega}$, in particular for $z \in D$. Thus the claim holds for $|\tau| < r$. \Box

In fact the converse of this result holds, using Hormander L^2 -estimates with plush weights on Ω .

Definition 15.29. Let $K \subset \Omega$ be a compact set. The *plurisubharmonic hull*

$$\hat{K}^p_{\Omega} = \{ z \in \Omega : \forall u \in P(\Omega) \ u(z) \le \sup_K u. \}$$

If $K = \hat{K}^p_{\Omega}$, we say that K is plurisubharmonically convex.

Now if $f \in A(\Omega)$ we have $\log |f| \in P(\Omega)$, so we are testing by fewer functions that in the case of the analytic hull \hat{K}_{Ω} . Thus $\hat{K}^p_{\Omega} \subseteq \hat{K}_{\Omega}$.

Theorem 15.30. The following are equivalent:

- 1. $z \mapsto -\log \delta(z, \mathbb{C}^n \setminus \Omega)$ is plush.
- 2. There is a $u \in P(\Omega)$ such that for every $c \in \mathbb{R}$, $\Omega_c = \{z \in \Omega : u(z) < c\}$ is Ω -precompact.
- 3. For every $K \subset \Omega$ compact, \hat{K}^p_{Ω} is compact in Ω .

Definition 15.31. Ω is *pseudoconvex* if one (and hence) all of the above conditions hold. The function *u* appearing in (2) is called a *plurisubharmonic exhaustion function*.

If Ω is a domain of holomorphy, then Ω is pseudoconvex.

Proof of theorem. To see 1 implies 2, let $u(z) = -\log \delta(z, \mathbb{C}^n \setminus \Omega) + |z|^2$. This is clearly plush and blows up at the boundary. So it is a plush exhaustion function.

To see 2 implies 3, let $c = \sup_{K} u$. Then $\{z \in \Omega : u(z) \leq \sup_{K} u\}$ is clearly compact and contains \hat{K}^{p}_{Ω} .

So we just need to prove 3 implies 1. Take $z_0 \in \Omega$ and $w \in \mathbb{C}^n \setminus 0$. We need to show that for every $|\tau| = r$, if $-\log(z_0 + \tau w, \mathbb{C}^n \setminus \Omega) \leq \operatorname{Re} f(\tau)$ then for every $|\tau| \leq r$ we have $-\log(z_0 + \tau w, \mathbb{C}^n \setminus \Omega)$.

If $|\tau| = r$ then $\delta(z_0 + \tau w, \mathbb{C}^n \setminus \Omega) \ge |e^{-f(\tau)}|$. Let $a \in \mathbb{C}^n$ be such that $\delta(a) < 1$. Let $F_{\lambda}(\tau) = z_0 + \tau w + \lambda z e^{-f(t)}$ and let $D_{\lambda} = F_{\lambda}(D(0,r))$ and $\Lambda = \{\lambda \in [0,1] : D_{\lambda} \subseteq \Omega\}$. We must show $\Lambda = [0,1]$ by showing that Λ is clopen and nonempty.

If $\lambda \in \Lambda$ and we perturb λ , then we do not move D_{λ} by much, so it remains in the open set Ω . Therefore Λ is open. Moreover, $0 \in \Lambda$ by assumption 3, so Λ is nonempty.

Let $K = \{z_0 + \tau w + \lambda a e^{-f(\tau)} : |\tau| = r, \lambda \in [0,1]\}$. Then K is Ω -compact since by assumption 3, $\delta(z_0 + \tau w, \mathbb{C}^n \setminus \Omega) \ge |e^{-f(\tau)}|$. Since |a| < 1 we have $|a\lambda e^{-f(\tau)}| < 1$ whence $\delta(z_0 + \tau w + \lambda a e^{-f(t)}) > 0$. Thus the function

$$\tau \mapsto u(z_0 + \tau w + a\lambda e^{-f(\tau)})$$

is subharmonic near $|\tau| \leq r$. Thus

$$u(z_0 + \tau w + \lambda a e^{-f(\tau)}) \le \sup_K u$$

for $|\tau| \leq r$, by the maximum principle, since K contains the boundary of \hat{K}^p_{Ω} . Thus $D_{\lambda} \subseteq \hat{K}^p_{\Omega}$.

If we have a sequence of $\lambda_j \in \Lambda$, say $\lambda_j \to \lambda_0 \in [0,1]$, then the $D_{\lambda_j} \subseteq \hat{K}^p_{\Omega}$, giving a continuous family of closed sets which converge to a closed set D_{λ_0} . So $D_{\lambda_0} \subseteq \hat{K}^p_{\Omega}$, so $\lambda_0 \in \Lambda$. So Λ is closed, which proves the theorem.

Corollary 15.32. If $(\Omega_{\alpha})_{\alpha}$ is a family of pseudoconvex domains then the interior Ω of $\bigcap_{\alpha} \Omega_{\alpha}$ is pseudoconvex.

Proof. One has

$$\delta(z, \mathbb{C}^n \setminus \Omega) = \delta\left(z, \mathbb{C}^n \setminus \bigcap_{\alpha} \Omega_{\alpha}\right) \inf_{\alpha} \delta(z, \mathbb{C}^n \setminus_{\alpha}).$$

Taking $-\log$ of both sides we arrive at

$$-\log \delta(z, \mathbb{C}^n \setminus \Omega) = \sup_{\alpha} -\log \delta(z, \mathbb{C}^n \setminus \Omega_{\alpha})$$

and the right-hand side is plush since the supremum of plush functions is plush. \Box

Corollary 15.33. Ω is pseudoconvex if and only if for every $z \in \hat{\Omega}$ there is a neighborhood $\omega \ni z$ such that $\Omega \cap \omega$ is pseudoconvex.

Proof. If Ω is pseudoconvex, let ω be a convex neighborhood of z. So ω is pseudoconvex; use the previous corollary.

For the converse, notice that this is trivial if $z \in \Omega$ by (3) of the above theorem. If $z_0 \in \partial\Omega$ and $\omega \ni z_0$ we notice that $\delta(z, \mathbb{C}^n \setminus \Omega) = \delta(z, \mathbb{C}^n \setminus (\Omega \cap \omega))$ if $|z - z_0|$ is small.

Thus the function $z \mapsto -\log \delta(z, \mathbb{C}^n \setminus \Omega)$ is plush near z_0 . But plurisubharmonicity is a local property so the function is plush on a neighborhood of $\partial\Omega$, i.e. there is a closed set $F \subset \Omega$ such that $z \mapsto -\log \delta(z, \mathbb{C}^n \setminus \Omega)$ is plush on $\Omega \setminus C$.

Let

$$\Phi(r) = \max_{\substack{|\zeta| \le r\\\zeta \in F}} -\log \delta(\zeta, \mathbb{C}^n \setminus \Omega)$$

so Φ is increasing. Now let Φ_1 be a convex increasing function such that $\Phi_1 \ge \Phi$. So we define $\varphi(z) = \Phi_1(|z|)$. So φ is a plush function and we can put

$$u(z) = \max(\varphi(z), -\log \delta(z, \mathbb{C}^n \setminus \Omega))$$

which is a supremum of plush functions, hence plush. Clearly u satisfies (2).

So pseudoconvexity is a local property.

Theorem 15.34. Let $\rho \in C^2(\mathbb{C}^n)$ with $d\rho|_{\rho=0} = 0$ and let $\Omega = \{z \in \mathbb{C} : \rho(z) < 0\}$. Then Ω is pseudoconvex if and only if for every $z \in \partial\Omega$, $w \in \mathbb{C}^n$ such that $\sum_j \partial_j \rho(z) w_j = 0$ we have

$$\sum_{j,k=1}^n \partial_j \overline{\partial}_k \rho(z) w_j \overline{w}_k \ge 0.$$

Definition 15.35. If Ω satisfies the hypotheses of Theorem 15.34 and is pseudoconvex then Ω satisfies the *Levi condition*.

Example 15.36. Let $\Omega = \{z \in \mathbb{C}^2 : |z_1|^2 + 2 \operatorname{Im} z_2 < 0\}$. Then

$$\rho(z) = |z_1|^2 + 2 \operatorname{Im} z_1.$$

Calculating, we see that

$$\partial \overline{\partial} \rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

so Ω satisfies the *strict* Levi condition (where \geq is replaced with >). In this case, for any $z_0 \in \partial \Omega$ there is a $U \in A(\Omega)$ which has a singularity at z_0 , so U does not extend beyond Ω . In fact, we put

$$U(z) = (z_1\overline{a}) - iz_2 - |a|^2/2 + ib)^{-1}$$

where $a \in \mathbb{C}$, $b \in \mathbb{R}$. We put $z_0 = (a, b - i|a|^2/2)$, then U has a singularity at z_0 . However, as we will prove in a later theorem, any function on $\mathbb{C}^2 \setminus \Omega$ admits an analytic continuation to Ω .

For this example, the tangential Cauchy-Riemann equation is $(\overline{\partial}_1 + iz_1\overline{\partial}_2)u(z) = 0$. That tangential Cauchy-Riemann operator, viewed as an operator on a 3-dimensional manifold (since $\partial\Omega$ is a 3-dimensional manifold) was used by Levi to disprove the version of the Cauchy-Kovaleskai theorem for smooth (rather than analytic) functions because for the generic $f \in C^{\infty}(\partial\Omega)$ we do not have Pu = f.

Proof of Theorem 15.34. First, a one-line lemma: Let $\rho_1 = h\rho$, $\rho > 0$, $h \in C^2$ with h > 0. Then if ρ satisfies the Levi condition then ρ_1 does as well. So we can replace the ρ in the hypotheses in theorem with any ρ satisfying the same conditions, and we will.

To prove that if Ω is pseudoconvex with a C^2 boundary then the Levi condition holds, we let $\rho(z) = -\inf_{w \notin \Omega} |z - w|$ for $z \in \Omega$ and $\rho(z) = \inf_{w \in \Omega} |z - w|$ for $z \notin \Omega$. Thus $\rho(z)$ is the "signed distance" from z to $\partial\Omega$. If $z \notin \partial\Omega$, we have

$$z = w + \rho(z)n(w)$$

for a minimizer $w \in \partial \Omega$ (which exists since $\partial \Omega$ is closed) and *n* the unit normal. Since Ω has a C^2 -boundary, *n* is a C^1 function of the element w' of \mathbb{R}^k where *k* is the dimension of $\partial \Omega$ as a real manifold. Let f(w') = w. Then

$$F(z, w', \rho(z)) = z - (f(w'), w') - \rho(z) \frac{(\nabla f(w'), 1)}{\sqrt{|\nabla f(w')|^2 + 1}}$$

and F is C^1 . By rotating the manifold $\partial\Omega$ so that the tangent plane near w is horizontal and translate so w = 0, i.e. w' = 0, f(0) = 0, $\nabla f(0) = 0$. So

$$\frac{\partial F}{\partial(w',\rho(z))}(z_0,0,\rho(z_0)) = \begin{bmatrix} \frac{\partial F}{\partial w'} & \frac{\partial F}{\partial \rho} \end{bmatrix} = \begin{bmatrix} I + O(\rho(z_0)) & 0\\ O(\rho(z_0)) & -1 \end{bmatrix}$$

which is invertible if $\rho(z_0)$ is small. Therefore we can use the implicit function theorem to see that ρ is well-defined and C^1 . We now implicitly differentiate $x = y + \rho n$ to see that

$$e_j = \partial_j (y'(x) + \rho(x) \left(y'(x) + \frac{-\nabla f(y'(x))}{\sqrt{1 + |\nabla f(y'(x))|}}, f(y') + \rho(1 + O(y')^2) \right)$$

which evaluates at $x = x_0, y' = 0$ to show that

$$\delta_{Nj} = \partial_j \rho(x_0).$$

Rotating back to the original coordinate frame,

$$\nabla \rho(x) = n(y(x))$$

whenever x is close to $\partial\Omega$. Therefore $\nabla\rho \in C^1$ so $\rho \in C^2$.

Now for $z \in \Omega$ and the standard δ (namely $\delta(z, w) = |z - w|$) we have $\rho(z) = -\delta(z, \mathbb{C}^n \setminus \Omega)$. Then $-\delta = \rho$ so $\delta \in C^2$ whence

$$-\partial_j \overline{\partial}_k \log \delta = \delta^{-2} \partial_j \delta \overline{\partial}_k \delta - \delta^{-1} \partial_j \overline{\partial}_k \delta$$

so it follows that

$$\sum_{j,k} \delta^{-1} \partial_j \delta w_j \overline{\partial}_k \delta \overline{w}_k - \partial_j \overline{\partial}_k \delta w_j \overline{w}_k \ge 0.$$

We know that $\partial_j \rho(z_0) w_j = 0$ for $z_0 \in \partial \Omega$ (hypothesis of the Levi condition) so

$$\sum_{j} \partial_{j} \rho(z) w_{j} = O(|z - z_{0}|).$$

Taking $z \to z_0$,

$$\sum_{j,k} \partial_j \overline{\partial}_k \rho w_j \overline{w}_k \ge 0$$

(conclusion of the Levi condition).

We prove the converse by contradiction. Assume that $-\log \delta(z, \mathbb{C}^n \setminus \Omega)$ is not plush in z, in any neighborhood of $\partial\Omega$. Then $\partial\overline{\partial}\log(z + \tau w, \mathbb{C}^n \setminus \Omega) > 0$ for some z close to $\partial\Omega$. We will expand this function in a Taylor series in τ . In fact,

$$\log \delta(z + \tau w, \mathbb{C}^n \setminus \Omega) = \log \delta(z, \mathbb{C}^n \setminus \Omega) + \operatorname{Re}(A\tau) + \operatorname{Re}(B\tau^2) + C|\tau|^2 + o(|\tau|^2)$$

for some $A, B \in \mathbb{C}$ and C > 0.

Now let $z(\tau) = z + \tau w + ae^{A\tau + B\tau^2}$ for some $a \in \mathbb{C}$. Then

$$\delta(z(\tau), \mathbb{C}^n \setminus \Omega) \ge \delta(z + \tau w, \mathbb{C}^n \setminus \Omega) - \delta(a) |e^{A\tau + B\tau^2}|$$

$$\ge \delta(a) (e^{C|\tau|^2/2} - 1) |e^{A\tau + B\tau^2}| \sim |\tau|^2$$

for $|\tau|$ small enough. Choose a so that $\delta(a) = d(z, \mathbb{C}^n \setminus \Omega)$. Then $z(0) = z + a \in \partial \Omega$. Since the function looks like $|\tau|^2$ we have

$$\partial_{\tau} \delta(z(\tau), \mathbb{C}^n \setminus \Omega)|_{\tau=0} = 0$$

$$\partial_{\tau}^2 \delta(z(\tau), \mathbb{C}^n \setminus \Omega)|_{\tau=0} > 0.$$

By the chain rule,

$$0 = \partial_{\tau} \delta(z(\tau), \mathbb{C}^n \setminus \Omega)|_{\tau=0} = -\sum_j \partial_j \rho z(0) z'_j(0)$$
$$0 < \partial_{\tau}^2 \delta(z(\tau), \mathbb{C}^n \setminus \Omega)|_{\tau=0} = -\sum_{j,k} \partial_j \overline{\partial}_k \rho z'_j(0) \overline{z}'_k(0)$$

Since $\rho < 0$, this contradicts the Levi condition.

We will not bother to prove the following theorem, but it is true. The proof uses the theory of Hormander L^2 estimates on complex manifolds with boundary, which is technically complicated but not very interesting.

Theorem 15.37 (Levi problem). If Ω is pseudoconvex then Ω is a domain of holomorphy.

Theorem 15.38. Assume that ω is a neighborhood of z_0 and $\rho \in C^4(\omega)$, $\rho(z_0) = 0$, $d\rho(z_0) = 0$. Suppose that there is a $w \in \mathbb{C}^n$ such that

$$\sum_{j,k} \partial_j \overline{\partial}_k \rho(z_0) w_j \overline{w}_k < 0$$

and

$$\sum_{j} \partial_{j} \rho(z_{0}) w_{j} = 0.$$

Then there is an $\omega' \subseteq \omega$, $z_0 \in \omega'$, such that for every $u \in C^4(\omega')$ such that for every $z \in \omega'$, $\rho(z) = 0$ implies

$$\overline{\partial}u \wedge \overline{\partial}\rho(z) = 0.$$

Let $\omega'_{+} = \{z \in \omega' : \rho(z) > 0\}$. Then there is a function U defined on ω' such that if $\rho(z) = 0$ then U(z) = u(z), and such that $U \in A(\omega'_{+})$.

Proof. Write $z = (z_1, z', z_n)$. There is an affine change of coordinates such that

$$\rho(z) = \operatorname{Im} z_n + A_{11}|z_1|^2 + O(|z_1|^3) + O(|z'|^2)$$

for some $A_{11} < 0$. In particular, we can find $\delta, \varepsilon > 0$ such that

 $\omega' = \{ z \in \mathbb{C}^n : |z_1| < \delta, \ |z'| + |z_n| < \varepsilon \} \subseteq \omega$

satisfies $\rho(z) < 0$ if $|z_1| = \delta$, $z \in \omega'$, and also such that

$$\partial_1 \overline{\partial}_1 \rho < 0$$

on ω' . Now the set of $z_1 \in \mathbb{C}$ such that $|z'| + |z_n| < \varepsilon$ implies $\rho(z) < 0$ is connected: otherwise, ρ would have a local minimum in the second connected component, yet $\Delta \rho < 0$ so this is impossible.

Lemma 15.39. For every $f \in C_{(0,1)}^k(\omega')$ such that $f|_{\omega'\setminus\omega'_+} = 0, k \ge 1, \overline{\partial}f = 0$, there is a $v \in C^k(\omega')$ such that $\overline{\partial}v = f$ and $v|_{\omega'\setminus\omega'_+} = 0$.

Proof. We define

$$v(z) = \frac{1}{2\pi i} \int_{|z_1| < \delta} \frac{f_1(\tau, z')}{\tau - z_1} d\tau \wedge d\overline{\tau}$$

so $\overline{\partial}v = f$. In particular v is analytic whenever $\rho \leq 0$. We claim v = 0 on $\omega' \setminus \omega'_+$. Near the boundary (except for the top) of ω' , $\overline{\partial}v = 0$, and v = 0 at the bottom, so v = 0 near the bottom. The set of points where $\rho < 0$ is connected, so v = 0 there.

Let v be as in the lemma. Let $U_0 \in C^2(\omega')$ be such that $\overline{\partial}U_0 = O(\rho^2)$ and $U_0|_{\rho=0} = u|_{\rho=0}$. Then let $U = U_0 + v$. So $\overline{\partial}U = \overline{\partial}U_0 + \overline{\partial}v$, $\overline{\partial}v = -\overline{\partial}U_0$, and $v|_{\omega'\setminus\omega'_+} = 0$ so we're done. \Box

15.4 Hormander L^2 estimates

We want to use the method of a priori estimates to show that $\overline{\partial} u = f$ has a solution, but the Hilbert space $L^2(\mathbb{C}^n)$ contains no holomorphic functions except 0. Therefore we must weight the inner product to apply Hilbert space theory. Fix $\Omega \subseteq \mathbb{C}^n$ open. We let λ denote Lebesgue measure.

Definition 15.40. Let $\varphi \in C^2(\Omega)$. We say that φ is a strictly plurisubharmonic function or simply that φ is strictly plush if

$$\inf_{t\in\mathbb{C}^n\backslash 0}\frac{\sum_j\sum_k\partial_j\overline{\partial}_k\varphi(z)t_j\overline{t_k}}{\sum_j|t_j|^2}>0.$$

Definition 15.41. Let φ be a strictly plush function on Ω . We define the weighted inner product with strictly plush weight φ to be the inner product $(\cdot, \cdot)_{\varphi}$ of the Hilbert space $L^2(\varphi)$ corresponding to the Borel measure $e^{-\varphi} d\lambda$ on Ω .

In other words,

$$(f,g)_{\varphi} = \int_{\Omega} f(z)g(z)e^{-\varphi(z)} dz$$

To motivate this definition, notice that

$$\int |u|^2 e^{-\varphi} = \int e^{2\log|u|-\varphi}$$

so we must have $\varphi > 2 \log |u|$ at infinity. But $\log |u|$ is subharmonic, so the point is that φ must have an especially strong form of subharmonicity for this to work on as many holomorphic functions as possible. This leads us to consider φ as a plush function.

Suppose that we have solved the equation $\overline{\partial} u = f$, for $f: \Omega \to \mathbb{C}^n$ a good function. Then for f to be the "gradient" of u with respect to $\overline{\partial}$, it must be the case that

$$\partial_j f_k = \partial_k f_j$$

since u is smooth and so has equality of mixed partials. We call this condition the Cauchy-Riemann constraint equation. One could view it as the statement that the 1-form du = $\sum_{k} f_j dx_j$ is closed. Of course, du is an exact form, hence closed.

Theorem 15.42 (Hormander's estimates). Let φ be strictly plush and let

$$\kappa(z) = \inf_{t \in \mathbb{C}^n \setminus 0} \frac{\sum_j \sum_k \partial_j \overline{\partial}_k \varphi(z) t_j \overline{t_k}}{\sum_j |t_j|^2}$$

witness that φ is strictly plush. Assume that $f \in L^2(\varphi + \log \kappa, \Omega \to \mathbb{C}^n)$ satisfies the Cauchy-Riemann constraints $\overline{\partial}_j f_k = \overline{\partial}_k f_j$. Then there is a $u \in L^2_{\varphi}$ such that $\overline{\partial}_j u = f_j$ and

$$||u||_{L^2(\varphi,\Omega\to\mathbb{C})} \le ||f||_{L^2(\varphi+\log\kappa,\Omega\to\mathbb{C}^n)}.$$

Before proving the theorem, we need the notion of weak solution for the operator $\overline{\partial}$. We introduce the differential operators

$$\delta_j = \partial_j - (\partial_j \varphi).$$

If u is a smooth solution to the equation $\overline{\partial} u = f$ and g is smooth, then

$$(\delta_j g, u)_{\varphi} = \int_{\Omega} \partial g \overline{u} e^{-\varphi} - \int_{\Omega} \partial_j \varphi g \overline{u} e^{-\varphi} = (g, f_j)_{\varphi}.$$

In other words, $\overline{\partial}_j^* = \delta_j$. For functions $\Omega \to \mathbb{C}^n$ we define

$$(g,h)_{\varphi} = \sum_{j} (g_j,h_j)_{\varphi}.$$

Proof of the Hormander estimates. Fix a convex function $\phi \geq 0$ such that

$$\phi(z) \ge |z| \log \kappa(z),$$

which is possible because $\kappa > 0$. Then for any $\varepsilon > 0$, we have $\log \kappa \leq \varepsilon \phi$ on $|z| > 1/\varepsilon$. Thus

$$\int_{|z|>1/\varepsilon} |f|^2 e^{-\varphi-\varepsilon\phi} \le \int_{|z|>1/\varepsilon} |f|^2 e^{-\varphi-\log\kappa} < \infty,$$

and the set $\{|z| \leq 1/\varepsilon\}$ is no problem for integrability, so $f \in L^2_{\varphi+\varepsilon\phi}$. Since ϕ is convex, it is plush, so $\varphi_{\varepsilon} = \varphi + \varepsilon \phi$ is strictly plush. Taking weakstar limits as $\varepsilon \to 0$, we can assume that $f \in L^2(\varphi)$.

Fix $g \in C_c^{\infty}(\Omega \to \mathbb{C}^n)$. Then

$$\begin{aligned} (\delta_j g_j, \delta_k g_k)_{\varphi} &= -(\partial_j \delta_k g_j, g_k)_{\varphi} \\ &= -(\delta_k \partial_{\overline{z}_j} g_j, g_k)_{\varphi} - ([\overline{\partial}_j, \delta_k] g_j, g_k)_{\varphi} \\ &= (\overline{\partial}_j g_j, \overline{\partial}_k g_k)_{\varphi} + (\partial_j \overline{\partial}_k \varphi g_j, g_k)_{\varphi} \end{aligned}$$

so, summing both sides over j, k,

$$(\delta g, \delta g)_{\varphi} + \frac{1}{2} \sum_{j \neq k} ||\overline{\partial}_j g_k - \overline{\partial}_k g_j||_{\varphi}^2 = \sum_j ||\overline{\partial}_j g||_{\varphi}^2 + \sum_{jk} (g_j \partial_j \overline{\partial}_k \varphi, g_k).$$

By the Cauchy-Schwartz inequality,

$$(g,f)_{\varphi}^{2} = (g\kappa,f)_{\varphi+\log\kappa}^{2} \le ||g\kappa||_{\varphi+\log\kappa} ||f||_{\varphi+\log\kappa}$$

and

$$\begin{split} ||g\kappa||_{\varphi+\log\kappa} &= \inf_{t\neq 0} ||t||^{-2} \sum_{jk} \int_{\Omega} g\overline{g} e^{-\varphi} \partial_j \overline{\partial}_k \varphi t_j \overline{t_k} \le \sum_{jk} \int_{\Omega} g\overline{g} e^{-\varphi} \partial_j \overline{\partial}_k \varphi \\ &\le \sum_j ||\overline{\partial}_j g||_{\varphi}^2 + \sum_{jk} (g_j \partial_j \overline{\partial}_k \varphi, g_k) \end{split}$$

In conclusion,

$$(g,f)_{\varphi}^{2} \leq ||\delta g||_{\varphi}^{2} ||f||_{\varphi+\log\kappa}^{2} + \frac{1}{2} \sum_{j \neq k} ||\overline{\partial}_{j}g_{k} - \overline{\partial}_{k}g_{j}||_{\varphi}^{2} ||f||_{\varphi+\log\kappa}^{2}$$

Let N be the subspace of $L^2(\varphi, \Omega \to \mathbb{C}^n)$ consisting of g which satisfy the Cauchy-Riemann constraints. If $h \in N^{\perp}$ (with respect to $(\cdot, \cdot)_{\varphi}$) and ψ is a test function, then $\overline{\partial}\psi \in N$, so $0 = (h, \overline{\partial}\psi)_{\varphi} = (\delta h, \psi)_{\varphi}$. But ψ was arbitrary, so $\delta h = 0$.

Let $P: L^2(\varphi, \Omega \to \mathbb{C}^n) \to N$ be the canonical projection. Then

$$|(g,f)_{\varphi}| = |(Pg,f)_{\varphi}| \le ||f||_{\varphi+\log\kappa}^2 ||\delta g||_{\varphi}.$$

Let us define the space D of elements of $L^2(\varphi, \Omega \to \mathbb{C}^n)$ of the form δg for some $g \in L^2(\varphi)$, and define $T \in D^*$ by $T(\delta g) = (g, f)_{\varphi}$. Then $||T|| \leq ||f||_{\varphi + \log \kappa} < \infty$. So by the Hanh-Banach theorem, T admits a linear extension \tilde{T} to $L^2(\varphi, \Omega \to \mathbb{C}^n)$. Then \tilde{T} has a Riesz representation, say u, which completes the proof. \Box

Chapter 16

Multivariable holomorphic functional calculus

This chapter follows Hormander's SCV book, Chapter III.

16.1 The Gelfand transform

Let B be an abelian, unital Banach algebra. One of the fundamental problems of the theory of such algebras is to consider to what extent that B can be approximated by algebras of the form C(K), for K a compact Hausdorff space.

Definition 16.1. A Banach algebra representation of B on K is a continuous morphism of algebras $B \to C(K)$.

To classify representations, we consider the space of characters on B.

Definition 16.2. A character or multiplicative functional on B is a continuous morphism of algebras $m: B \to \mathbb{C}$ which is not identically 0. The space of all characters on B is denoted M_B .

For each $f \in B$, the function \hat{f} defined on characters by $\hat{f}(m) = m(f)$ is the *Gelfand* transform of f. We give the space M_B the weakstar topology, namely the weakest topology such that for each $f \in B$, the Gelfand transform \hat{f} is continuous. The resulting map $B \to C(M_B)$ is called the *Gelfand* representation of f.

By the Banach-Alaoglu theorem, M_B is a compact Hausdorff space.

The Gelfand representation is universal among representations of B.

Lemma 16.3. If $T: B \to C(K)$ is a representation of B, then there is a map φ so that for every $f \in B$,

$$Tf = \hat{f} \circ \varphi.$$

Proof. Recall that Te is idempotent, so its image (Te)(K) consists only of 0 or 1. Let K_0 be the kernel of Te and K_1 be its complement in K. Then $\{K_0, K_1\}$ is a partition of K into compact, open sets. But then for any $f \in B$, Tf = 0 on K_0 , so we might as well assume $K = K_1$. Under this assumption, for each $x \in K$, the map $f \mapsto Tf(x)$ is a character, which we denote $\varphi(k)$.

Definition 16.4. Let $f \in B$. The resolvent function of f is the meromorphic function $R_f: U \to B$ given by

$$R_f(\lambda) = \frac{1}{f - \lambda e}.$$

The domain U of the resolvent function of f is called the *resolvent set* of f. The complement in B of U of f is called the *spectrum* of f, denoted $\sigma(f)$.

In case B is a space of matrices, then the spectrum of f is exactly the set of eigenvalues of f viewed as a linear operator, since then f solves the eigenvalue equation

$$f(x) = \lambda x$$

Theorem 16.5 (spectral radius theorem). For each $f \in B$, $\sigma(f) = \{\hat{f}(m) : m \in M_B\}$. Moreover,

$$\sup_{m \in M_B} |\hat{f}(m)| = \lim_{n \to \infty} ||f^n||^{1/n}.$$

The proof of this theorem uses some complex analysis, which we now consider.

Lemma 16.6. Let $g \in B$. If g is invertible, then the mapping

$$\lambda \mapsto (g - \lambda h)^{-1}$$

is continuous on the disk D of all λ such that

$$|\lambda| < \frac{1}{||g^{-1}h||}.$$

Assume $\omega \subseteq D$ is bounded by a finite number of C^1 arcs. If φ is holomorphic on ω and C^1 on $\overline{\omega}$, then

$$\int_{\partial \omega} (g - \lambda h)^{-1} \varphi(\lambda) d\lambda = 0$$

Proof. Let $H = g^{-1}h$ and

$$I(\lambda) = g^{-1} \sum_{n=0}^{\infty} \lambda^n H^n.$$

This series converges locally uniformly on D, in fact by definition of D. Also,

$$I(\lambda)(g - \lambda h) = I(\lambda)g(e - \lambda H) = e.$$

Therefore we can integrate term by term after multiplying by $\varphi(\lambda)$, and each term is holomorphic.

Lemma 16.7. If $(e - \lambda f)^{-1}$ exists for every $|\lambda| \leq R$, then for each $n \geq 0$,

$$R^{n}||f^{n}|| \leq \sup_{|\lambda|=R} ||(e - \lambda f)^{-1}||.$$

Proof. By homotopy invariance and the above lemma, the integral

$$\frac{1}{2\pi i} \int_{|\lambda|=r} (e - \lambda f)^{-1} \lambda^{-n-1} \, d\lambda$$

is independent of r if $r \leq R$, and if r||f|| < 1, then the integral is equal to f^n .

Lemma 16.8. For each $f \in B$, $\sigma(f)$ is nonempty.

Proof. Assume $\sigma(f)$ is empty. Then $(e - \lambda f)^{-1}$ exists for every $\lambda \in \mathbb{C}$, and the holomorphic function

$$||(e - \lambda f)^{-1}|| \le \frac{||f^{-1}||}{|\lambda|}||(e - \lambda^{-1}f^{-1})^{-1}||$$

is bounded as $\lambda \to \infty$, contradicting Liouville's theorem.

Corollary 16.9. If *B* is a field, then $B = \mathbb{C}$.

Proof. By the lemma, for every $f \in B$ we can find $\lambda \in \mathbb{C}$ such that $f - \lambda e$ is not invertible, but since B is a field, it follows that $f = \lambda e$.

Lemma 16.10. If I is a proper ideal of B then there is a $m \in M_B$ such that m(f) = 0 for every $f \in I$.

Proof. By Zorn's lemma, we can find a maximal ideal $\mathfrak{m} \supseteq I$. The natural map $B \to B/\mathfrak{m} = \mathbb{C}$ is a character, call it m.

Finally we are ready to prove the spectral radius theorem.

Proof of spectral radius theorem. We first claim that $\{\hat{f}(m) : m \in M_B\} \subseteq \sigma(f)$: if $\lambda \notin \sigma(f)$, then there is a $g \in B$ so that $g(f - \lambda e) = e$, so $\hat{g}(\hat{f} - \lambda) = 1$. Therefore $\hat{f}(m) \neq \lambda$ for any m. Now let

Now let

$$1/R \le \sup_{z \in \sigma(f)} |z|.$$

So if $|\lambda| \leq R$, $(e - \lambda f)^{-1}$ exists by the above lemmata. So

$$\left(\limsup_{n \to \infty} R^n ||f^n||\right)^{1/n} = R \limsup_{n \to \infty} ||f^n||^{1/n} \le 1$$

and it follows that

$$\limsup_{n \to \infty} ||f^n||^{1/n} \le \frac{1}{R} \le \sup_{z \in \sigma(f)} |z|.$$

Third, if $\lambda \in \sigma(f)$, then $f - \lambda e$ is not a unit, so the ideal $I = (f - \lambda e)$ is proper. Therefore we can find a $m \in M_B$ which is annihilated by I. Then $\lambda = m(f)$, so $\sigma(f) \subseteq \{\hat{f}(m) : m \in M_B\}$. This proves the first assertion of the spectral radius theorem.

Since the Gelfand representation is continuous, there is a $C \ge 1$ such that

$$\sup_{m \in M_B} |f(m)| \le C||f||.$$

 So

$$\sup_{m \in M_B} |\hat{f}(m)| \le C^{1/n} ||f^n||^{1/n} \le C^{1/n} ||f||$$

implying that

$$\sup_{n \in M_B} |\hat{f}(m)| \le \liminf_{n \to \infty} ||f^n||^{1/n}.$$

Therefore by the lemmata

$$\sup_{m \in M_B} |\hat{f}(m)| \le \liminf_{n \to \infty} ||f^n||^{1/n} \le \limsup_{n \to \infty} ||f^n||^{1/n} \le \sup_{z \in \sigma(f)} |z| = \sup_{m \in M_B} |\hat{f}(m)|.$$

This proves the second assertion.

Now we generalize the notion of a spectrum to several complex variables. A version of the spectral radius theorem holds still.

Definition 16.11. The *joint spectrum* $\sigma(f_1, \ldots, f_n)$ is the set of all $\lambda \in \mathbb{C}^n$ such that the ideal

$$(f_1 - \lambda_1 e, \ldots, f_n - \lambda_n e)$$

is proper.

Corollary 16.12. For $f_1, \ldots, f_n \in B$,

$$\sigma(f_1, \ldots, f_n) = \{ (\hat{f}_1(m), \ldots, \hat{f}_n(m) : m \in M_B \}.$$

The proof is essentially the same. This generalization of the spectral radius theorem will allow us to classify the Gelfand representations of finitely generated Banach algebras.

Theorem 16.13. Let B be the Banach algebra generated by f_1, \ldots, f_n . Then the mapping

$$\varphi: M_B \to \sigma(f_1, \dots, f_n)$$
$$m \mapsto (\hat{f}_1(m), \dots, \hat{f}_n(m))$$

is a homeomorphism. Moreover, $\sigma(f_1, \ldots, f_n)$ is polynomially convex, and for each $f \in B$, one can approximate $\hat{f} \circ \varphi^{-1}$ uniformly by polynomials on $\sigma(f_1, \ldots, f_n)$.

Proof. Since M_B carries the weakstar topology, φ is continuous, and injective since if p is a polynomial,

$$m(p(f_1,\ldots,f_n)) = p(\hat{f}_1(m),\ldots,\hat{f}_n(m))$$

(and polynomials in the f_j are dense in B, by definition of B). By the corollary of the spectral radius theorem, φ is surjective.

To prove the second statement, let $K = \sigma(f_1, \ldots, f_n)$, $z \in \hat{K}$, and define a map on the generators by $f_j \mapsto z_j$. This will extend to a character on all of B if it is continuous; and indeed

$$|p(z)| \le \sup_{w \in K} |p(w)| \le \sup_{m \in M_B} |m(p(f))| \le |p(f)|.$$

Therefore $z \in K$, proving the second claim. The final claim follows because polynomials are dense in B.

16.2 The holomorphic functional calculus

We now show that holomorphic functions can be extended to an (abelian, unital) Banach algebra B, even in several complex variables.

Theorem 16.14 (holomorphic functional calculus in several complex variables). Let $f_1, \ldots, f_n \in B$ and let φ be a holomorphic function on a neighborhood of the joint spectrum $\sigma(f_1, \ldots, f_n)$. Then there is a $g \in B$ such that

$$\hat{g} = \varphi(\hat{f}_1, \dots, \hat{f}_n).$$

One generally writes $g = \varphi(f_1, \ldots, f_n)$, so we think of φ as a function $B^n \to B$.

Lemma 16.15. Let $\Omega \subseteq \mathbb{C}^n$ be open and contain $\sigma(f_1, \ldots, f_n)$. Then there is a finitely generated Banach subalgebra B' of B such that $f_1, \ldots, f_n \in B'$ and $\sigma_{B'}(f_1, \ldots, f_n) \subseteq \Omega$.

Proof. As B' increases, $\sigma_{B'}(f_1, \ldots, f_n)$ decreases. So we show that for every $z \notin \sigma(f_1, \ldots, f_n)$, we can find B' so that $z \notin \sigma_{B'}(f_1, \ldots, f_n)$. Indeed, we can find $f_{j+n} \in B$ so that

$$e = \sum_{j=1}^{n} f_{j+n}(f_j - z_j e).$$

Now let B' be the Banach algebra generated by the f_j .

Lemma 16.16. Let B' be as above. Let f_1, \ldots, f_v be the generators of B', and let $\pi : \mathbb{C}^v \to \mathbb{C}^n$ be the projection which annihilates $(0, \ldots, 0, z_{n+1}, \ldots, z_v)$. Then there are polynomials p_k such that for each $z \in \mathbb{C}^v$ such that for each $j, |z_j| \leq ||f_j||$, if

$$|p_k(z)| \le ||p_k(f_1, \dots, f_v)||,$$

then $\pi(z) \in \Omega$.

Proof. Assume $\pi(z) \notin \sigma_{B'}(f_1, \ldots, f_n)$. Then the map $f_j \mapsto z_j$ cannot extend to a character on B', so is discontinuous if we were to try to extend it; i.e. there is a p so that

$$|p(z_1, \ldots, z_v)| \ge ||p(f_1, \ldots, f_v)||.$$

This is still true close to z, so use compactness of the closed polydisk

$$\{z \in \mathbb{C}^v : |z_j| \le ||f_j||\}.$$

Now we come to the theorem that we will use to prove the holomorphic functional calculus.

Theorem 16.17. Let $f_1, \ldots, f_n \in B$ and let φ be holomorphic in a neighborhood of $\sigma(f_1, \ldots, f_n)$. Then there are f_{n+1}, \ldots, f_N and a holomorphic function Φ on a neighborhood of the polydisk

$$\{z \in \mathbb{C}^N : |z_j| \le ||f_j||\}$$

such that $\varphi(f_1,\ldots,f_n) = \Phi(f_1,\ldots,f_N).$

_		

Proof. Let Ω be a neighborhood of $\sigma(f_1, \ldots, f_n)$. By the lemma, we can find f_{n+1}, \ldots, f_v and p_1, \ldots, p_μ satisfying certain conditions. Let $N = v + \mu$ and $f_{v+k} = p_k(f_1, \ldots, f_v)$. The function $\varphi \circ \pi$ is holomorphic in a neighborhood of the compact set of all $z \in \mathbb{C}^v$ such that $|z_j| \leq ||f_j||$ and $|p_k(z)| \leq ||f_{k+v}||$. Therefore by results in several complex variables, we can find the desired Φ .

Proof of holomorphic functional calculus. Let Φ and f_{n+1}, \ldots, f_N be as above. By holomorphy, we can write

$$\Phi(z) = \sum_{\alpha} a_{\alpha} z^{\alpha}$$

such that

$$\sum_{\alpha} |a_{\alpha}| R^{\alpha} < \infty$$

where $R = (||f_1||, \ldots, ||f_N||)$ and $z = (z_1, \ldots, z_N)$. Therefore the series

$$g = \sum_{\alpha} a_{\alpha} f^{\alpha}$$

norm-converges. Moreover,

$$\hat{g} = \sum_{\alpha} a_{\alpha} \hat{f}^{\alpha} = \Phi(\hat{f}_1, \dots, \hat{f}_N) = \varphi(\hat{f}_1, \dots, \hat{f}_n).$$

-		
Е		
L		
L		
L		

Chapter 17 Algebraic geometry

Throughout this chapter, we assume that all rings are commutative and unital.

17.1 Schemes and varieties

Definition 17.1. A ringed space $X = (X, \mathcal{F})$ consists of a sheaf of rings \mathcal{F} on X. If the stalks of \mathcal{F} are all local rings, then we say that X is a *locally ringed space*.

A morphism of ringed spaces $\psi : (X, \mathcal{F}) \to (Y, \mathcal{G})$ consists of a continuous map $\psi : X \to Y$ and for each $U \in \mathbf{Open}(Y)$, a morphism of rings $\psi_U : \mathcal{G}(U) \to \mathcal{F}(\psi^{-1}(U))$ such that for every open set $V \subseteq U$, the diagram

$$\begin{array}{cccc}
\mathcal{G}(V) & \stackrel{\psi_V}{\longrightarrow} & \mathcal{F}(\psi^{-1}(V)) \\
& & & \downarrow \\
\mathcal{G}(U) & \stackrel{\psi_U}{\longrightarrow} & \mathcal{F}(\psi^{-1}(U))
\end{array}$$

commutes.

Let (X, \mathcal{F}) and (Y, \mathcal{G}) be locally ringed spaces. A morphism of locally ringed spaces $\psi : (X, \mathcal{F}) \to (Y, \mathcal{G})$ is a morphism of ringed spaces such that for every $x \in X$, the maximal ideal \mathfrak{m} of the stalk \mathcal{F}_x is given by $\mathfrak{m} = \psi_x(\mathfrak{n})$ where \mathfrak{n} is the maximal ideal of the stalk $\mathcal{G}_{\psi(x)}$ and ψ_x is the colimit of morphisms ψ_U as U ranges over the directed set $\mathcal{D}_{\psi(x)}$ of all open sets $U \ni \psi(x)$.

So a morphism of locally ringed spaces is a morphism of ringed spaces, whose domain and codomain are locally ringed, which preserves the maximal ideals at each stalk.

Definition 17.2. Let R be a ring. Let X be the spectrum of R, equipped with the Zariski topology. If U = D(f) is a distinguished open set, let $\mathcal{F}(U)$ be the localization of R at $R \setminus D(f)$. Let \mathcal{F} be the induced sheaf. Then Spec $R = (X, \mathcal{F})$ is called the *affine scheme* associated to R.

Proposition 17.3. Let R be a ring. The affine scheme Spec R is a locally ringed space, and R is the ring of global sections of Spec R.

Definition 17.4. A *scheme* is a locally ringed space which is locally an affine scheme. A *morphism of schemes* is a morphism of locally ringed spaces.

If X is a scheme, then there is a unique morphism $X \to \operatorname{Spec} \mathbb{Z}$, which is the categorical dual of the morphisms $\mathbb{Z} \to R$ for each ring R appearing in the definition of X.

Definition 17.5. Let S be a scheme. A scheme over S, X, is a scheme X such that there is a morphism of schemes $\pi : X \to S$.

Let X and Y be schemes over S, witnessed by morphisms $\pi : X \to S$ and $\varphi : Y \to S$. A morphism of schemes over S, $\psi : X \to Y$, is a morphism of schemes $\psi : X \to Y$ such that $\pi = \varphi \circ \psi$.

In case $S = \operatorname{Spec} \mathbb{C}$, we say that X is a *complex scheme*.

Notice that if $X = (X, \mathcal{F})$ is a complex scheme and $U \subseteq X$ is an affine subscheme, then $\mathcal{F}(U)$ admits a morphism of rings $\mathbb{C} \to \mathcal{F}(U)$. This gives rise to a complex algebra structure on $\mathcal{F}(U)$.

Definition 17.6. If X is a complex scheme and every algebra $\mathcal{F}(U)$ is finitely generated over \mathbb{C} , we say that X is a *complex scheme of finite type*.

Definition 17.7. A reduced scheme $X = (X, \mathcal{F})$ is a scheme such that for every open set $U \subseteq X$, the ring $\mathcal{F}(U)$ has no nilpotents.

Definition 17.8. Let $\pi : X \to Y$ be a morphism of schemes. The *diagonal morphism* $\delta_{\pi} : X \to X \times_Y X$ is the fiber product of the identity $X \to X$ with itself induced by π . If $\delta_{\pi}(X)$ is closed in $X \times_Y X$, we say that π is a *separated morphism*.

If the unique morphism $X \to \operatorname{Spec} \mathbb{Z}$ is separated, we say that X is a separated scheme.

Definition 17.9. A *variety* is a reduced, separated complex scheme of finite type.

17.2 Formal power series

Let $\sum_{\alpha} a_{\alpha} z^{\alpha}$ be a formal power series on \mathbb{C}^n with domain of (absolute) convergence D. Let B be the set of z such that $|a_{\alpha} z^{\alpha}|$ is uniformly bounded in α . Clearly $D \subseteq B$.

Lemma 17.10. Assume $w \in B$ and $U = \{z \in \mathbb{C}^n : |z_j| < |w_j|\}$. Then $U \subseteq D$.

Theorem 17.11. $D^* = \{\xi \in \mathbb{R}^n : (e^{\xi_1}, \dots, e^{\xi_n}) \in D\}$ is an open, convex set. If $\xi \in D^*$ and $|\eta_j| \leq |\xi_j|$ then $\eta \in D^*$. Moreover, $z \in D$ if and only if $|z_j| \leq e^{\xi_j}$ for some $\xi \in D^*$.

Proof. D^* is the interior of B^* . We will show B^* is convex. There is an M such that $\xi, \eta \in B^*$ if and only if

$$|a_{\alpha}e^{\alpha\xi}| \le M$$

and similarly for $e^{\alpha\eta}$. This remains true when we raise both sides to the t or 1-t power. Then

$$|a_{\alpha}|e^{\alpha(t\xi+(1-t)\eta)} \le M.$$

Thus $t\xi + (1-t)\eta \in B^*$.

The other claims follow from the definition or are obvious from the lemma.

Definition 17.12. A *Reinhardt domain* is a set $\Omega \subset \mathbb{C}^n$ such that for every $z \in \Omega$ and every $\theta \in \mathbb{R}^n$, $(e^{i\theta_1}z_1, \ldots, e^{i\theta_n}z_n) \in \Omega$.

17.3 Bergman kernels

Suppose Φ is a strictly plush quadratic form on \mathbb{C}^n . That is,

$$\Phi(z) = \operatorname{Re}\langle Az, z \rangle + \langle Cz, \overline{z} \rangle$$

where $A \in \mathbb{C}^{n \times n}$ and C is a positive matrix. (The inner product is antilinear in \overline{z} , hence why we needed complex conjugation). Then we define

$$L^2_{\Phi}(\mathbb{C}^n) = \left\{ u : \int_{\mathbb{C}^n} |u(z)|^2 e^{-2\Phi(z)} \, dz < \infty \right\}.$$

Thus we can define $H_{\Phi}(\mathbb{C}^n) = L^2_{\Phi}(\mathbb{C}^n) \cap A(\mathbb{C}^n)$. Taking the holomorphic part of a compactly supported function (which is always possible for smooth functions by the Hormander L^2 estimates), we see that $H_{\Phi}(\mathbb{C}^n)$ is nonempty. $H_{\Phi}(\mathbb{C}^n)$ is closed, hence a Hilbert space.

It is often useful to have a semiclassical parameter, so we put

$$||u||_{\Phi}^2 = \int_{\mathbb{C}^n} |u|^2 e^{-2\Phi/h} dm$$

Putting

$$u_{\alpha}(z) = e^{-\langle Az, z \rangle/h} z^{\alpha}$$

we recover an orthonormal basis

$$f_{\alpha} = C(h^{n+|\alpha|}\alpha!)^{-1/2}u_{\alpha}$$

of H_{Φ} .

Let

$$\Pi_{\Phi}: L^2_{\Phi} \to H_{\Phi}$$

be the orthogonal projection. Then one can check that

$$\Pi_{\Phi} u(z) = C \int_{\mathbb{C}^n} e^{2\Psi(z,\overline{w})/h - 2\Phi(w)/h} u(w) \ dm(w),$$

where $\Psi: \mathbb{C}^{2n} \to \mathbb{C}$ is the unique analytic quadratic function such that

$$\Psi(z,\overline{z}) = \Phi(z)$$

Example 17.13. If $\Phi(z) = |z|^2/2$ then $\Psi(z, w) = \langle z, w \rangle/2$ and

$$\Pi u(z) = \frac{1}{(\pi h)^n} \int_{\mathbb{C}^n} e^{-\langle z, \overline{w} \rangle / h - |w|^2 / h} u(w) \ dm(w).$$

In quantum mechanics, H is the space of wavefunctions. If $g\in L^\infty$ is a classical observable, we quantize g by

$$T_g = \Pi g \Pi$$

In fact, g itself cannot be holomorphic by Liouville's theorem, but T_g preserves A.

Definition 17.14. The projection T_g is called the *Toeplitz operator* of g.

Example 17.15. Let $g(z) = \overline{z}_j$, $u \in H_{\Phi+\varepsilon|z|^2}$, where $\Phi(z) = |z|^2/2$. Then T_g carries H to itself, and

$$T_g u(z) = \frac{1}{(h\pi)^n} \int_{\mathbb{C}^n} e^{-\langle z, \overline{w} \rangle / h - |w|^2 / h} \overline{w}_j u(w) \ dm(w).$$

We first see that

$$\overline{w}_j e^{-\langle z, \overline{w} \rangle - |w|^2/h} = -h \partial_{w_j} e^{-\langle z, \overline{w} \rangle - |w|^2/h}$$

and since the integrals converge, we integrate by parts to see that

$$T_g u(z) = \frac{1}{(h\pi)^n} e^{-\langle z, \overline{w} \rangle / h - |w|^2 / h} h \partial_{w_j} u(w) \ dm(w) = \Pi(h \partial_j u(z)) = h \partial_j u(z).$$

That is, multiplication by \overline{z}_j is the same as differentiating in z_j . So this is, in fact, a quantization.

17.4 Quillen's theorem

Theorem 17.16 (Catlin-d'Angelo-Quillen). Let

$$f(z,\overline{z}) = \sum_{|\alpha| = |\beta| = m} c_{\alpha\beta} z^{\alpha} \overline{z}^{\beta}$$

be a bihomogeneous quadratic form on \mathbb{C}^n such that $f(z,\overline{z}) > 0$ for every $z \neq 0$. Then there is a $N \in \mathbb{N}$ and a polynomial

$$P_j(z) = \sum_{|\alpha|=m} p_{\alpha}^j z^{\alpha}$$

such that

$$|z|^{2N} f(z, \overline{z}) = \sum_{j=1}^{J} |P_j(z)|^2.$$

Quillen developed this theorem to prove the complex Nullstellensatz. It was also used by Polya for the following theorem.

Theorem 17.17 (Polya). Suppose p is a real homogeneous polynomial on \mathbb{R}^n and p(x) > 0 if $\forall j \ x_j \ge 0$ and $\sum_j x_j = 1$. Then there is a $N \in \mathbb{N}$ such that $(x_1 + \cdots + x_n)^N p(x)$ has positive coefficients.

Proof. Write $x_j = z_j \overline{z_j}$ and put $c_{\alpha\beta} = 0$ for $\alpha \neq \beta$, then use Quillen's theorem.

To prove Quillen's theorem, we develop the theory of real quadratic forms. Let $q: \mathbb{C}^n \to \mathbb{R}$ be a quadratic form.

Lemma 17.18. There is a decomposition $q = h + \ell$ such that h(iz) = -ih(z) and $\ell(z) = \ell(iz)$.

Definition 17.19. ℓ is the *Levi form* of q.

Proof. Define Jq(z) = q(iz). Then $J^2 = 1$. Put h = (q - Jq)/2, $\ell = (q + Jq)/2$. Then h is *pluriharmonic* in the sense that for any j in any coordinate system, $\partial_j \overline{\partial}_j h(iz)$ is constant because h is a quadratic form. Moreover,

$$\partial_j \overline{\partial}_j h(iz) = -\partial_j (\overline{\partial}_j h(iz)) = \partial_j \overline{\partial}_j h(iz)$$

which proves the claim for h.

To prove Quillen's theorem, let

$$P_f(z) = \sum_{|\alpha| = |\beta| = m} c_{\alpha\beta} z^{\alpha} (h\partial)^{\beta}$$

be a quantization of f. Then if h = 1/N, P_f carries the space \mathcal{P}_{m+N} of homogeneous polynomials of degree m+N to itself, as we will prove, and this will prove Quillen's theorem.

Lemma 17.20. There is a polynomial P such that

$$f(z,\overline{z}) = \sum_{j=1}^{J} |P_j(z)|^2$$

if and only if $A = (c_{\alpha\beta})_{\alpha\beta}$ is a positive-definite matrix.

Note that this is not trivial because A acts on the vector space \mathbb{C}^K , where K is the set of partitions of m, which by some combinatorics can be very large!

Proof. For any symmetric matrix $A \in \mathbb{C}^{K \times K}$ we can find $w_k \in \mathbb{C}^K$, λ_j , $j, k \in K$, such that

$$A = \sum_{j,k} \lambda_j w_k w_k^*$$

by the spectral theorem. Here f is real-valued, so $c_{\alpha\beta} = \overline{c_{\beta\alpha}}$ whence A is symmetric. Thus we can use that decomposition.

Let $Z = (z^{\alpha})_{\alpha \in K}$. Then

$$f(z,\overline{z}) = Z^* A Z = \sum_{\alpha \in K} \lambda_\alpha Z^* w_\alpha w_\alpha^* Z$$

so $f(z,\overline{z})$ is the sum over α of the sign of λ_{α} times $|P_{\alpha}(z)|^2$, where

$$P_{\alpha}(z) = |\lambda_{\alpha}|^{1/2} w_{\alpha}^* Z = |\lambda_{\alpha}|^{1/2} \sum_{\beta \in K} \overline{w}_{\alpha}^{\beta} z^{\beta}.$$

So if A is positive then the $\lambda_{\alpha} \geq 0$.

For the converse, just run the same argument in reverse.

Example 17.21. Let $z \in \mathbb{C}^2$,

$$f(z,\overline{z}) = |z_1|^4 + |z_2|^4 + c|z_1|^2|z_2|^2$$

for some $|c| \in (0,2)$. Then f is a positive quadratic form. There are only 3 partitions of 2, so the matrix A acts on \mathbb{C}^3 by

$$A = \begin{bmatrix} 1 & & \\ & c & \\ & & 1 \end{bmatrix}.$$

Then A is positive-definite if and only if c > 0, so that is exactly when f is a positive quadratic form.

Let \mathcal{P}_{n+M} be the space of homogeneous polynomials of degree n + M. Let $\Phi(z) = |z|^2/2$ as in the theory of Bergman kernels and T be the quantization operator.

Lemma 17.22. Let

$$|z|^{2N}f(z,\overline{z}) = \sum_{|\alpha|=|\beta|=N+m} c^N_{\alpha\beta} z^{\alpha} \overline{z}^{\beta}$$

be a bihomogeneous form. Let

$$P_f(z) = \sum_{|\alpha| = |\beta| = m} c_{\alpha\beta} z^{\alpha} T_{\overline{z}^{\beta}}.$$

Then $A = (c_{\alpha\beta})$ is positive if and only if there is a c > 0 such that for every $u \in \mathcal{P}_{n+M}$ we have

$$\langle P_f u, u \rangle_{\Phi} \ge c ||u||_{\Phi}$$

Proof. The orthonormal projection $\Pi: L^2 \to H$ is self-adjoint so

$$T_g^* = T_{\overline{g}}$$

Moreover, it is easy to check that P_f is self-adjoint, and $P_f = \sum_{\alpha,\beta} c_{\alpha\beta} z^{\alpha} (h\partial)^{\beta}$. Then

$$\begin{split} P_f g(z) &= \sum_{\alpha,\beta} c_{\alpha\beta} z^{\alpha} \Pi(\overline{z}^{\beta} g(z)) \\ &= \sum_{L=0}^{\infty} \frac{1}{(\pi h)^n} \int_{\mathbb{C}^n} \frac{\langle z, \overline{w} \rangle^L}{h^L L!} f(z, \overline{w}) g(w) e^{-|w|^2/h} \ dm(w) \\ &= \sum_L (h\pi)^{-n} \int_{\mathbb{C}^n} \sum_{|\mu|=L} \frac{z^{\mu} \overline{w}^{\mu}}{\mu! h^{|\mu|}} f(z, \overline{w}) g(w) e^{-|w|^2/2h} \ dm(w). \end{split}$$

Assume $g \in \mathcal{P}_{m+N}$. By homogeneity one can only get a nonzero contribution to the integral when L = N. So

$$P_f g(z) = (\pi h)^{-n} \int_{\mathbb{C}^n} \sum_{|\mu|=N} \frac{z^{\mu} \overline{w}^{\mu}}{\mu! h^{|\mu|}} f(z, \overline{w}) g(w) e^{-|w|^2/h} \, dm(w).$$

Let $u \in \mathcal{P}_{m+N}$ expand as

$$u(z) = \sum_{|\gamma|=m+N} u_{\gamma}(z) z^{\gamma}.$$

Then

$$\langle P_f u, u \rangle = h^{N+2m} \sum_{\substack{|\alpha| = |\beta| = m \\ |\mu| = N}} c_{\alpha\beta} \frac{(\alpha + \mu)!(\beta + \mu)!}{\mu!} u_{\beta + \alpha} \overline{u}_{\alpha + \mu}$$
$$= (\pi h)^{-n} h^{N+2m} \sum_{|\gamma| = |\rho| = N+m} c_{\rho\gamma}^N \rho! \gamma! u_\rho \overline{u}_{\gamma}$$

and $(c^N_{\rho\gamma}\rho!\gamma!)$ is positive iff A is positive. Since

$$\frac{\langle z,\overline{z}\rangle^N}{N!} \sum_{|\alpha|=|\beta|=m} c_{\alpha\beta} z^{\alpha} z^{\beta} = \sum_{\substack{|\alpha|=|\beta|=m\\|\mu|=N}} \frac{c_{\alpha\beta}}{\mu!} z^{\alpha+\mu} \overline{z}^{\beta+\mu}$$

we plug this back into $\langle Pfu, u \rangle$.

Lemma 17.23. For every $\delta > 0$ and m > 1 there is a $\varepsilon > 0$ such that if $|Mh - 1| < \varepsilon$, then for every $u \in \mathcal{P}_M$,

$$|| |z|^m u||_{L^2_{\Phi}(|z| \le 1-\delta)} \le O(h^{\infty})||u||_{\Phi}.$$

Proof. Let

$$\Pi_M: L^2_\Phi \to \mathcal{P}_M$$

be the orthogonal projection. Now

$$||\Pi_M|z|^m \mathbf{1}_{|z| \le 1/2} \Pi_M||_{L^2_{\Phi} \to L^2_{\Phi}} = O(h^{\infty})$$

if hM is close to 1. In fact, it is no trouble to reduce to when m = 0 since $|z|^m < 2^{-m}$ and then

$$||u||_{L^2_{\Phi}(|z|<1/2)}^2 = \langle 1_{|z|<1/2}u, u \rangle_{\Phi} = \langle \Pi_M 1_{|z|<1/2}\Pi_M u, u \rangle_{\Phi}.$$

We now use Schur's criterion with $p(w) = |w|^M$. In fact

$$\Pi_M u(z) = (\pi h)^n \int_{\mathbb{C}^n} \langle z, \overline{w} \rangle^M \frac{u(w)}{M! h^M} e^{-|w|^2/h} dw$$

so if K is the integral kernel defined by

$$\Pi_M 1_{|z| \le 1/2} \Pi_M = \int_M K(z, w) u(w) e^{-|w|^2/h} \, dw$$

we have

$$K(z,w) = (h\pi)^{-2n} \int_{\mathbb{C}^n} \frac{\langle z,\overline{\zeta} \rangle^M}{M!h^M} 1_{|\zeta| < 1/2} \frac{\langle \zeta,\overline{w} \rangle^M}{M!h^M} e^{-|\zeta|^2/h} d\zeta$$
since by Fubini's theorem the integral kernel of a product of integral operators is the product of the integral kernels. Thus

$$|K(z,w)| \le (\pi h)^{-2n} \int_{|\zeta| < 1/2} \frac{|z|^M |\zeta|^{2M} |w|^M}{M! h^M} e^{-|\zeta|^2/h} \, d\zeta$$

implies

$$\int_{\mathbb{C}^n} |K(z,w)| |w|^M e^{-|w|^2/h} \, dw \le \frac{|z|^M}{(\pi h)^{2M}} \int_{\mathbb{C}^n} \int_{|\zeta| < 1/2} \frac{|\zeta|^{2M}}{M!h^M} \frac{|w|^{2M}}{M!h^M} e^{-|\zeta|^2/h} e^{-|w|^2/h} \, d\zeta \, dw$$
$$=: I_1 I_2$$

where we are thinking of the $e^{-|\cdot|^2/h}$ as a Radon-Nikodym derivative of a certain measure (so we can use Schur's criterion on that measure). Using polar coordinates,

$$I_{1} \leq \frac{C}{M!h^{M+n}} \int_{0}^{\infty} r^{2M+2n-1} e^{-r^{2}/h} dr$$

= $O\left(\frac{1}{M!}\right) \int_{0}^{\infty} t^{M+n-1} e^{-t} dt$
 $\leq O\left(\frac{1}{M!}\right) (M+n-1)! \leq O((M+n)^{n})$

Since Mh is close to 1, $e^{(M+m+n-1)t}$ looks like $e^{t/h}$, and if $1 - \rho = Mh$ then ρ is small. We recall that $\log 1/4 - 1/4 \approx -1.6 < 1$. So if we take $r^2 = t$,

$$\begin{split} I_2 &\leq C \int_0^{1/2} \frac{r^{2M+2n-1}}{M!h^{M+n}} e^{-r^2/h} dr \\ &= C \int_0^{1/4} \frac{t^{M+m+n-1}}{M!h^{M+n}} e^{-t/h} dt = \frac{Ce^{\delta/4h}}{M!h^{M+n}} \int_0^{1/4} (te^{-t})^{1/h} dt \\ &= \frac{Ce^{\delta/4h}}{M!h^{M+n}} \int_0^{1/4} e^{(\log t-t)/h} dt \leq C \frac{e^{\delta/4h} - 1.5/h}{M!h^{M+n}} \\ &\leq C \frac{e^{-1.2/h}}{M!(1-\rho)^{-(n+M)}M^{-(n+M)}} \leq C \frac{(1+\rho)^{n+M}e^{-1.2/h}}{M^{M+1/2}M^{-M}M^{-n}} \\ &\leq O(M^n(1+\rho)^{n+M})e^{-(1-\rho)M} \leq O(e^{-CM}) = O(e^{-C/h}) \end{split}$$

by Stirling's formula.

Lemma 17.24. Let $|\alpha| = |\beta| = m$. There are $p_j \in \mathcal{P}_{m-j}$ such that if

$$q_{\alpha\beta}(z,\overline{z}) = z^{\alpha}\overline{z}^{\beta} + \sum_{j=1}^{n} h^{j} p_{j}^{\alpha\beta}(z,\overline{z})$$

then $z^{\alpha}T_{\overline{z}^{\beta}} = T_{q_{\alpha\beta}}$.

Proof. This is obvious if m = 0. Otherwise, assume that it is true for m, and use the fact that

$$T_{z^{\alpha}\overline{z}^{\beta}} = (h\partial_z)^{\beta} z^{\alpha} = z^{\alpha} (h\partial_z)^{\beta} \sum_{j=1}^m h^j \sum_{|\mu|=m-j} p_j^{\alpha\beta} z^{\mu} (h\partial_z)^{\mu}$$

to prove the lemma by induction.

Proof of Quillen's theorem. By a previous lemma, we have $f = T_q$ where

$$q(z,\overline{z}) = f(z,\overline{z}) + \sum_{j=1}^{m} h^{j} p_{j}(z,\overline{z})$$

for some $p_j \in \mathcal{P}_{m-j}$ which depend on our choice of h. We claim that for some N > 1, h < 1and every $u \in \mathcal{P}_{m+N}$, there is a c > 0 such that

$$\langle T_q u, u \rangle_\Phi \ge c ||u||_\Phi^2$$

In fact we take c so that $f(z, \overline{z}) > 2c$ for 2|z| > 1, which is possible by homogeneity of f. So if h > 0 is small enough, then

$$q(z,\overline{z}) > c$$

for every z with 2|z| > 1. In particular we have the luxury to choose h so that we can find N such that N = 1/h. Now Π is self-adjoint and $u \in H$, so

$$\langle T_q u, u \rangle_{\Phi} = \langle \Pi q \Pi u, u \rangle_{\Phi} = \langle q \Pi u, \Pi u \rangle_{\Phi}$$

$$= \langle q u, u \rangle_{\Phi} \ge c \langle 1_{|z| \ge 1/2} u, u \rangle_{\Phi} - c ||u||_{L^2_{\Phi}(|z| < 1/2)}$$

$$= c ||1_{|z| \ge 1/2} u||^2_{L^2_{\Phi}(|z| \ge 1/2)} - O(h^{\infty})||u||_{\Phi}$$

$$\ge c \frac{||u||^2_{\Phi}}{2}.$$

	-

Г	-	-	
L			L
L	_	_	

Chapter 18

Line bundles over complex varieties

Definition 18.1. An *immersion* is an injective continuous function.

Definition 18.2. An holomorphic atlas on a topological space M is an open cover $(U_{\alpha})_{\alpha}$ equipped with open immersions $\tau_{\alpha} : U_{\alpha} \to \mathbb{C}^n$ such that whenever $U_{\alpha} \cap U_{\beta}$ is nonempty, the mapping

$$\tau_{\alpha} \circ \tau_{\beta}^{-1} : \tau_{\beta}(U_{\alpha} \cap U_{\beta}) \to \tau_{\alpha}(U_{\alpha} \cap U_{\beta})$$

is holomorphic.

Definition 18.3. A *complex manifold* M is a Hausdorff space equipped with a holomorphic atlas.

Definition 18.4. A function $f: M \to \mathbb{C}^m$ is a holomorphic function if for every α , $f \circ \tau^{-1}$: $\tau_{\alpha}(U_{\alpha}) \to \mathbb{C}^{\gamma}$ is holomorphic.

Example 18.5. Let

$$M = \{ z \in \mathbb{C}^{n+1} : z_0^2 + \dots + z_n^2 = 1 \}.$$

Then there is a j such that $z_j \neq 0$. By the implicit function theorem, we can show that M is a complex manifold. But it is not compact, and in fact has complex codimension 1.

Definition 18.6. Let V be a complex vector space, which the multiplicative group \mathbb{C}^* acts on by scalar multiplication. Let $\mathbb{P}(V) = (V \setminus 0)/\mathbb{C}^*$, the *complex projective space*. Define for $x \in \mathbb{P}(V)$ the equivalence class $[x] \in \mathbb{P}(V)$. Define $\Omega_j = \{[x] \in \mathbb{P}(V) : x_j \neq 0\}$. Then define

$$\tau_j^{-1}([x]) = (x_0 x_j^{-1}, \dots, x_{j-1} x_j^{-1}, x_{j+1} x_j^{-1}, \dots, x_n x_j^{-1}).$$

Then for $z \in \tau_i(\Omega_i \cap \Omega_j)$,

$$\tau_i \circ \tau_j^{-1}(z) = (z_0 z_i^{-1}, \dots, z_{j-1} z_i^{-1}, z_i^{-1}, z_{j+1} z_i^{-1}, \dots, z_{i-1} z_i^{-1}, z_{i+1} z_i^{-1}, \dots, z_n z_i^{-1}).$$

Then if $V = \mathbb{C}^{n+1}$, $\mathbb{P}(V) = S^{2n-1}/S^1$ where S^1 acts on S^{2n-1} by scalars. So $\mathbb{P}(V)$ is a compact complex manifold.

Example 18.7. Let $M = \{x \in \mathbb{C}^{n+1} : x_0^2 + \cdots + x_{n+1}^2 = 0\}/\mathbb{C}^*$, where \mathbb{C}^* acts on \mathbb{C}^{n+1} by scalars. By a similar argument as with projective space, M is a compact complex manifold.

Theorem 18.8. The only holomorphic functions on a compact connected complex manifold are constant.

Proof. A function would have to attain a maximum since the manifold is compact, and by the maximum principle if a function attains its maximum on the interior of a connected set then it is constant there. \Box

18.1 Holomorphic line bundles

Definition 18.9. A holomorphic line bundle is a holomorphic projection of complex manifolds $\pi : E \to M$, such that for every $x \in M$, $E_x = \{\pi^{-1}(x)\}$ is isomorphic to \mathbb{C} , and that there is an open set $U \ni x$ and a holomorphic function $\Theta : \pi^{-1}(U) \to U \times \mathbb{C}$ such that $\Theta|_{E_x} : E_x \to \{x\} \times \mathbb{C}$ is an isomorphism of vector spaces.

We define $\Theta_{j\ell} = \Theta_j \cap \Theta_{\ell}^{-1}$ whenever $U_j \cap U_{\ell}$ is nonempty. Then $\Theta_{j\ell}(x, \cdot), x \in M$, is a linear map $\mathbb{C} \to \mathbb{C}$, so there is a scalar $g_{j\ell}(x)$ such that $\Theta_{j\ell}(x,t) = g_{j\ell}(x)t$.

Definition 18.10. The functions $g_{j\ell} : \Theta_j \cap \Theta_\ell^{-1} \to \mathbb{C}$ are called *transition functions* for the holomorphic line bundle.

We have $g_{j\ell}g_{\ell j} = 1$. In fact, on $U_j \cap U_k \cap U_\ell$, then $g_{j\ell}g_{\ell k}g_{kj} = 1$. On the other hand, we have a holomorphic line bundle for any family of functions with these compatibility conditions. Let J be the index set, $(j, x, t) \sim (j', x, t')$ whenever $t' = g_{j'j}t$. Then $J \times M \times \mathbb{C}$ projects by \sim to a holomorphic line bundle E.

Example 18.11. The trivial line bundle is $E = M \times \mathbb{C}$.

Example 18.12. The *tautological line bundle* O(-1) over \mathbb{P}^n , the moduli space of all lines through $0 \in \mathbb{C}^{n+1}$, sends the line corresponding to each point of \mathbb{P}^n to itself. It is defined by

$$O(-1) = \{ ([x], \xi) \in \mathbb{P}^n \times \mathbb{C}^{n+1} : \xi \in [x] \}.$$

(Here we allow $0 \in [x]$.) We define $\Theta_j([x], \xi) = \xi_j$. Then $\Theta_j \circ \Theta_\ell^{-1}([x], \xi) = x_j \xi / x_\ell$. That is,

$$g_{j\ell}(x) = \frac{x_j}{x_\ell}.$$

Definition 18.13. Let $\pi : E \to M$ be a holomorphic line bundle. A *section* is a right inverse of π .

So a section carries $x \in E$ to the complex line $\mathbb{C} \times \{x\}$. We denote by $C^{\infty}(M, E)$ the space of smooth sections of π . The space of holomorphic sections is denoted by $H^0(M, E)$, for sheaf-theoretic reasons.

Example 18.14. For the tautological line bundle, let $x \in U_j$. We define a vector space isomorphism $e_j : E_x \to \mathbb{C}$ that sends a point in [x] to its *j*th coordinate. This allows us to express as section *s* as $s(x) = s_j(x)e_j(x)$ whenever $x \in U_j$. If $x \in U_j \cap U_\ell$, then $e_\ell(x) = g_{j\ell}(x)e_j(x)$ and $s_j(x)e_j(x) = s_\ell(x)e_\ell(x)$. So $s_j(x)e_j(x) = s_j(x)g_{\ell j}(x)e_\ell(x) = s_\ell(x)e_\ell(x)$ whence $s_j(x)/x_j = s_\ell(x)/x_\ell$. (Clearly this does not depend on the choice of $x \in [x]$.) A section s of O(-1) defines a function on $\mathbb{C}^{n+1} \setminus 0$, \tilde{s} , by, for $[x] \in U_j$,

$$\tilde{s}(x) = \frac{s([x])}{x_j}.$$

Then \tilde{s} is well-defined, does not depend on the choice of j, and is homogeneous of degree -1. So $C^{\infty}(\mathbb{P}^n, O(-1))$ is infinite-dimensional while $H^0(\mathbb{P}^n, O(-1)) = 0$.

We now define operations on the category of line bundles. The goal is to construct line bundles that have holomorphic sections.

Definition 18.15. Let M be a complex manifold. If E is a holomorphic line bundle, we define the *dual line bundle* E^* by requiring that the fibers E_x^* are dual vector spaces to the fibers E_x . Given g_{ij} the transition maps for E, we define the dual transition maps $g_{ij}^* = g_{ij}^{-1}$.

Then E^* is a holomorphic line bundle, and given $x \in U_j \cap U_k \subset E$, $\xi \in E_x$, $f \in E_x^*$, $f(\xi)$ does not depend on whether we compute $f(\xi)$ in U_k or in U_j .

Example 18.16. The dual of the tautological line bundle is by definition $O(1) = O(-1)^*$. It has the transition maps

$$g_{ij}^*([x]) = \frac{x_j}{x_i}.$$

As above, the sections are functions on $\mathbb{C}^{n+1} \setminus 0$ which are homogeneous of degree 1. So $H^0(\mathbb{P}^n, O(1))$ consists of linear forms on \mathbb{C}^{n+1} (since every holomorphic function extends over 0 on \mathbb{C}^{n+1} .)

Definition 18.17. Let F, E be holomorphic line bundles over M. We define the *tensor* product of line bundles by $(F \otimes E)_x = F_x \otimes E_x$ on fibers and

$$g_{ij}^{F\otimes E} = g_{ij}^F \otimes g_{ij}^E.$$

Here the tensor product of holomorphic functions is defined by pointwise multiplication.

Example 18.18. Let $O(k) = O(1)^{\otimes k}$ (for $k \geq 0$; for k < 0 we have $O(k) = O(-1)^{\otimes -k}$). Then we have transition maps $g_{ij}^k([x]) = x_j^k x_i^{-k}$ so $H^0(\mathbb{P}^n, O(k))$ consists of kth degree forms on \mathbb{C}^{n+1} . The dimension of $H^0(\mathbb{P}^n, O(k))$ is (n+k)!/n!.

Definition 18.19. The *Picard group* is the group $\{O(k) : k \in \mathbb{Z}\}$ of line bundles on \mathbb{P}^n . with tensor product as multiplication and duality as inversion.

So the Picard group is isomorphic to \mathbb{Z} . The identity of the Picard group O(0) is the trivial bundle. The sections on O(0) are exactly those functions $\mathbb{P}^n \to \mathbb{C}$.

Definition 18.20. A Hermitian metric on a holomorphic line bundle $h = (h_j)_j$ is a vector of smooth $h_j : U_j \to [0, \infty)$ satisfying the compatibility condition $h_j(x) = |g_{ij}(x)|^2 h_i(x)$, such that with $e_j(x) = \Theta_j^{-1}(x, 1)$ defined for $x \in U_j$, we have norms $|| \cdot ||$ on each fiber defined by

$$||e_j(x)||_h^2 = h_j(x).$$

The norm is well-defined because any element of a fiber E_x is a scalar multiple of $e_j(x)$. For a section s we have

$$||s(x)||_{h}^{2} = h_{j}(x)|s_{j}(x)|^{2}$$

in U_j .

Definition 18.21. The *Fubini-Study metric* is the Hermitian metric on O(1) given by

$$h_j([x]) = \frac{|x_j|^2}{|x|^2}$$

for $[x] \in \mathbb{P}^n$.

Example 18.22. Let

$$f(x,\overline{x}) = \sum_{|\alpha| = |\beta| = k} c_{\alpha\beta} x^{\alpha} \overline{x}^{\beta}$$

be such that $f(x, \overline{x}) > 0$. Then we can define

$$h_j(x) = \frac{|x_j|^{2k}}{f(x,\overline{x})}.$$

By Quillen's theorem, if k is large enough then there is a polynomial such that the metric we have just defined is the pullback of the Fubini-Study metric.

Definition 18.23. A *Hermitian line bundle* is a holomorphic line bundle equipped with a Hermitian metric.

Definition 18.24. A positive line bundle is a Hermitian line bundle (E, h) such that $h_j = e^{-\varphi_j}$ for some strictly plush functions φ_j .

On a positive line bundle, there will be lots of holomorphic sections.

Example 18.25. Take $U_0 \subset \mathbb{P}^n$ to be the set of [x] such that $x_0 \neq 0$. Then $h_0([x]) = |x_0|^2 |x|^{-2}$ for the Fubini-Study metric. The Fubini-Study metric makes O(1) into a positive line bundle by

$$\varphi_0(z) = -\log(|x_0|^2|x|^{-2}) = \log(1+|z|^2).$$

In fact,

$$\partial_j \overline{\partial}_k \varphi_0(z) = \frac{\delta_{jk}}{1+|z|^2} - \frac{z_k \overline{z}_j}{(1+|z|^2)^2}.$$

It is easy to check that this matrix is positive-semidefinite, so (O(1), h) is a positive line bundle.

18.2 Integration on Hermitian line bundles

Let h be the Hermitian line bundle defined by $h_{\alpha} = e^{-\varphi_{\alpha}}$ for some strictly plurisubharmonic function φ_{α} .

Definition 18.26. The Levi form of h is given by

$$\omega_{\alpha} = \frac{1}{2i} \partial \overline{\partial} \partial_{\alpha}.$$

The Levi form is a (1, 1)-form, given by

$$\omega_{\alpha} = \frac{1}{2i} \sum_{j,k} \partial_j \overline{\partial}_k \varphi_{\alpha} dx_j \wedge d\overline{x}_k.$$

Then $\varphi_{\alpha} = 2 \log |g_{\alpha\beta}| + \varphi_{\beta}$ for transition functions φ . It follows that $\omega_{\alpha} = \omega_{\beta}$ on overlapping patches so ω is a real 2-form.

Example 18.27. Let h be the Fubini-Study metric. Then the Levi form of h is

$$\omega_{\ell} = \frac{i}{2} \sum_{j} \frac{dx_{j} \wedge d\overline{x}_{j}}{1 + |x|^{2}} - \sum_{\substack{j \neq \ell \\ k \neq \ell}} \frac{\overline{x}_{j} x_{k} \, dx_{j} \wedge d\overline{x}_{k}}{(1 + |x|^{2})^{2}}$$
$$= \frac{i}{2} \left(\frac{d(x \, d\overline{x})}{(1 + |x|^{2})} - \frac{(\overline{x} \, dx) \wedge (x \, d\overline{x})}{(1 + |x|^{2})^{2}} \right).$$

Given a Levi form ω on an *n*-dimensional manifold, we obtain a volume form $dV = \omega^n/n!$ (where ω^n is the exterior power).

Example 18.28. The Fubini-Study Levi form is invariant under the unitary group. We can use this diagonalize the Fubini-Study metric so that the double sum is only a single sum, and prove that

$$dV = \frac{i^n}{2^n} \frac{dx_1 \wedge d\overline{x}_1 \wedge \dots \wedge dx_n \wedge d\overline{x}_n}{(1+|x|^2)^{n+1}}.$$

Because we have a volume form, integration makes sense.

Definition 18.29. The space $L^2(M, E, h)$ is the completion of the space of $s \in C^{\infty}(M, E)$ for which the norm defined by the inner product

$$\langle s, s' \rangle_h = \int_M \langle s(x), s'(x) \rangle_h \, dV$$

Here the inner product in the integral is the inner product induced on each line appearing in the line bundle induced by the Hermitian metric h.

We thus introduce the Bergman projection $L^2(M, E) \to H^0(M, E)$, which is orthogonal. To construct it, we compute an orthonormal basis of $H^0(M, E)$. **Example 18.30.** For the Fubini-Study metric on O(k), O the Picard group, consider the basis s^k_{α} such that on the patch Ω_0 , $s^k_{\alpha}(x) = x_1^{\alpha_1} \dots x_n^{\alpha_n}$, $|\alpha| = k$. Then

$$\langle s_{\alpha}, s_{\beta} \rangle_{FS} = \int_{\mathbb{P}^n} x^{\alpha} \overline{x}^{\beta} h_{FS,0}(x) \ dV_{FS} = \frac{i}{2} \int_{\mathbb{C}^n} \frac{x_1^{\alpha_1} \cdots x_n^{\alpha_n} \overline{x}_1^{\beta_1} \cdots \overline{x}_n^{\beta_n}}{(1+|x|^2)^{k+n+1}} \ dx_1 \wedge d\overline{x}_1 \wedge \cdots \wedge dx_n \wedge d\overline{x}_n = \delta_{\alpha\beta} \pi^n \frac{\alpha!}{(n+k)!}$$

where we used polar coordinates. Therefore we normalize by dividing $s_{\alpha,0}^k$ by $\sqrt{(k+n)!/(\alpha!\pi^n)}$ to get an orthonormal basis. The Bergman projection is given by

$$\Pi_k u(x) = \sum_{|\alpha|=k} s^k_{\alpha}(x) \langle u, s^k_{\alpha} \rangle_k$$

The Bergman kernel B_k is defined by

$$\Pi_k u(x) = \int_{\mathbb{P}^n} \langle u(y), B_k(y, x) \rangle_h \, dV$$

So we have

$$B_k(y,x) = \sum_{\alpha} s_{\alpha}^k(y) \overline{s_{\alpha}^k(x)}$$

Therefore B_k is holomorphic in the first variable and antiholomorphic in the second variable. We conclude that B_k can be given by the explicit formula

$$B_k(x,y) = \sum_{|\alpha|=k} \frac{(k+n)!}{\alpha!\pi^n} z^{\alpha} \overline{w}^{\alpha} = \frac{(k+n)!}{k!\pi^n} \langle z, w \rangle^k$$

where x = [z], y = [w]. In particular, the Bergman kernel is given by the Hermitian metric:

$$B_k(x,x) = e^{-2k\varphi_0(x)} = \frac{(k+n)!}{k!\pi^n}.$$

18.3 Asymptotics for the Bergman kernel

Here we let, for L a line bundle, L^k denote the kth tensor power of L with itself.

Theorem 18.31 (Fefferman, Boutet, de Monvel, Sjostrand, Zelditch, Catlin, Berman, Bernsom). Let $L \to M$ be a positive Hermitian line bundle and let s_j^k be an orthonormal basis of $H^0(M, L^k)$. Then there is an asymptotic expansion

$$\sum_{j} |s_{j}^{k}(x)|_{h}^{2} = \frac{k^{n}}{\pi^{n}} + \sum_{\ell=1}^{\infty} a_{\ell}(x)k^{n-\ell}.$$

Towards this theorem, we work locally. We identify a coordinate patch with $B_{\mathbb{C}^n}(0,1)$ and fix a strictly plurisubharmonic weight φ on $B_{\mathbb{C}^n}(0,1)$ and a semiclassical parameter $h \in (0,1)$. This gives a (1,1)-form

$$\omega = -i\partial\overline{\partial}\varphi = \sum_{jk} \partial_j\overline{\partial}_k\varphi \ dx_j \wedge d\overline{x}_k.$$

We then have a volume form defined by $n! dV = \omega^k$, the right hand side being an exterior power. Then

$$||u||_{\varphi}^{2} = \int_{B_{\mathbb{C}^{n}}(0,1)} |u(x)|^{2} e^{-2\varphi(x)/h} \, dV(x)$$

is a L^2 -norm and we denote its inner product by $(\cdot, \cdot)_{\varphi}$. We will obtain a local reproducing kernel K given some $\varepsilon > 0$ and χ a $B_{\mathbb{C}^n}(0, \varepsilon)$ -cutoff function, i.e.

$$u(x) = \int_{B_{\mathbb{C}^n}(0,1)} \chi(y) u(y) \overline{K(y,x)} e^{-2\varphi(y)/h} \, dV(y).$$

Definition 18.32. The reproducing kernel K is called a *Bergman kernel* if $K(x, \cdot)$ is holomorphic.

Example 18.33. Reproducing kernels also have applications in machine learning.

To construct the Bergman kernel, we obtain an approximate reproducing kernel, which is correct up to an error of order h^{∞} .

Recall that for $A: L^2_{\varphi} \to L^2_{\varphi}$ an integral operator, $e^-\varphi A e^{\varphi}$ sends L^2 to itself an its integral kernel is given by $e^{-\varphi(x)/h} K_A(x, y) e^{-\varphi(y)/h}$.

Definition 18.34. Let $\Lambda \subseteq B_{\mathbb{C}^n}(0,1) \times \mathbb{C}^n$, $\Lambda = \{(y,\xi)\}$ where ξ is a function of (x,y). Then Λ is a *good contour* if

$$\operatorname{Im}\langle\xi, x - y\rangle \ge \delta |x - y|^2 + \varphi(y) - \varphi(x).$$

Example 18.35. If $\varphi(x) = |x|^2/2$ then

$$\overline{K(y,x)} = (\pi h)^{-n} e^{\langle x,\overline{y} \rangle/h}$$

and this Bergman kernel is valid on \mathbb{C}^n . Here $\omega = i/2 \sum_j dz_j \wedge d\overline{z}_j$ and dV is the Euclidean volume form. We obtain a good contour by $\xi(x, y) = i\overline{y}$. Then

$$2 \operatorname{Im} \langle i\overline{y}, x - y \rangle = |y|^2 + |x|^2 + |x - y|^2.$$

Taylor expanding,

$$\varphi(x) = \varphi(y) - \operatorname{Im} Q(x, y)(x - y) + (\partial \overline{\partial} \varphi(x)(x - y))\overline{x - y} + O(|x - y|^3)$$

where Q is some holomorphic function and

$$2\delta|x-y|^2 \le (\partial\overline{\partial}\varphi(x)(x-y))\overline{x-y}.$$

We now define a good contour by $\xi = Q$.

Lemma 18.36. Suppose Λ is a good contour for φ . Then there is a $\varepsilon > 0$ and a $B(0, \varepsilon)$ -cutoff χ such that for every $u \in A(B(0, 1))$,

$$u(x) = i^{-n^2} (2\pi h)^{-n} \int_{\Lambda} e^{i\langle\xi, x-y\rangle/h} u(y)\chi(y) \ d\xi \wedge dy + O(e^{(\varphi(x)-\delta)/h})||u||_{\varphi}.$$

This lemma can be intuited by deforming Λ into a flat space and then applying the Fourier inversion formula. That is not a proof, however. We put $c_n = i^{-n^2}$ for simplicity.

Proof. Define the form

$$\eta = c_n (2\pi h)^{-n} e^{i\langle \xi, x-y \rangle/h} u(y) \chi(y) \ d\xi \wedge dy.$$

Define

$$I_s = \int_{\Lambda_s} \eta$$

where Λ_s is the form defined by $\xi_s(x, y) = \xi(x, y) + i\overline{s(x - y)}$.

Lemma 18.37. As $s \to \infty$, $I_s \to u(x)$ for $x \in B(0, 1/2)$.

Proof. We have the expression

$$e^{i\langle\xi,x-y\rangle/h}e^{-s|x-y|^2/h}u(y)\chi(y)(c_n\ d_{\overline{y}}\xi(x,y)\wedge dy+s^n\ dm(y)).$$

By the dominated convergence theorem $e^{-s|x-y|^2/h}c_n \ d_{\overline{y}}\xi(x,y) \wedge dy \to 0$ as $s \to \infty$. As a distribution,

$$s^{n}(2\pi h)^{-n}e^{-s|x-y|^{2}/h} \to \delta_{0}(x-y).$$

Now $\Omega = \mathbb{C}^n \times [0, s]$ and $\partial \Omega = \mathbb{C}^n \times s - \mathbb{C}^n \times 0$ (as a chain). We introduce the homotopy $h(y, t) = \langle y, \xi \rangle + it\overline{x - y}$, so $h : \mathbb{C}^n \to [0, \infty) \to \mathbb{C}^{2n}$ and put $\omega = h^*\eta$. Since ω is compactly supported (since η has a factor of χ) we can apply Stokes' theorem to see that

$$\int_{\Omega} d\omega = \int_{\partial \Omega} \omega$$

Since η has a factor of $d\xi \wedge dy$, $\partial \eta = 0$. So

$$d\eta = \overline{\partial}\eta = c_n (2\pi h)^{-n} e^{-\langle \xi, x-y \rangle/h} u(y) \overline{\partial}\chi \wedge dy \wedge d\xi.$$

Then

$$I_s - I_0 = \int_{\mathbb{C}^n \times s} h^* \eta = \int_{\mathbb{C}^n \times 0} h^* \eta = \int_{\mathbb{C}^n \times [0,s]} h^* \, d\eta$$

Since $I_s \to u(x)$ we can estimate $d\eta$. If |y| < 1/2 then $d\chi = 0$, and

$$\operatorname{Re}(i\langle x-y,\xi\rangle/h) \le -\delta|x-y|^2 + \varphi(x) - \varphi(y)$$

since Λ is good. Since $|x| < \varepsilon$ and we can take $\varepsilon < 1/4$ we only need to consider |y| > 1/2, in which case

$$\operatorname{Re}(i\langle x-y,\xi\rangle)/h \le -\delta/C - \varphi(y) + \varphi(x).$$

Now

$$\begin{aligned} \left| \int_{\mathbb{C}^n \times [0,s]} h^*(d\eta) \right| &\leq C h^{-n} e^{-\delta/16h + \varphi(x)/h} \int_{|y| \geq 1/2} |\chi'(y)| |u(y)| e^{-\varphi(y)/h} \ dm(y) \int_0^s (1+t^n) e^{-t/16h} \ dt \\ &\leq C e^{-\delta/32h + \varphi(x)/h} ||u||_{\varphi}. \end{aligned}$$

We conclude that

$$I_s - I_0 = O(e^{-\delta/h} e^{\varphi(x)/h}) ||u||_{\varphi}$$

The right hand side does not depend on s so we take $s \to \infty$ to see that we have constructed an approximate kernel.

Unfortunately the approximate kernel we constructed here may not be holomorphic. So we consider for which a do we have

$$u(x) = c_n (2\pi h)^{-n} \int_{\Lambda} e^{i\langle\xi, x-y\rangle} u(y)(1+a)\chi(y) \ d\xi \wedge dy + O(\varphi(x)/h - \delta/h) ||u||_{\varphi}.$$

If

$$ae^{i\langle\xi,x-y\rangle/h} d\xi = d\xi (e^{i\langle\xi,x-y\rangle}A)$$

where $A = \sum_j A_j(x, y, \xi, h) \ d\hat{\xi}_j$ and $d\hat{\xi}_j$ is the unique (n-1, 0)-form such that $\xi_j \wedge d\hat{\xi}_j = d\xi$ then

$$\begin{split} \int_{\Lambda} e^{i\langle\xi, x-y\rangle/h} u(y)\chi(y)a(x,y,\xi,h) \ d\xi \wedge dy &= \int_{\Lambda} u(y)\chi(y)d(e^{i\langle\xi, x-y\rangle/h}A) \wedge dy \\ &= \int_{\Lambda} \chi(y)d(u(y)e^{i\langle\xi, x-y\rangle/h}A) \wedge dy \\ &= -\int_{\Lambda} d\chi \wedge u(y)e^{i\langle\xi, x-y\rangle/h}A \wedge dy \\ &= O(e^{-\delta/h}e^{\varphi(x)/h})||u||_{\varphi} \end{split}$$

using that Λ is a good contour. Intuitively we are replacing a pseudodifferential symbol that depends on (x, y, ξ) with a pseudodifferential operator that depends on (x, ξ) .

Definition 18.38. A symbol $a(x, y, \xi, h)$ is a *negligible symbol* if $a = \sum_j \mathcal{D}_{\xi_j} A_j + (x - y)A/h$ for some A.

We introduce the differential operator $\tilde{\nabla} = \partial_{\xi} + i(x-y)/h$. Then $ia = \tilde{\nabla}A$ and we absorb the *i* into the *A*.

In this case the error term introduced by adding a is $O(e^{-\delta/h}e^{\varphi(x)/h})||u||_{\varphi}$. We are specifically interested in when A has an asymptotic sum $A = \sum_{m} A_{m}h^{m}$. What this means is that for every N,

$$A - \sum_{m \le N} A_m h^m = O(h^{N+1})$$

or in other words the sum "converges" modulo $O(h^{\infty})$. We write $A^{(N)}$ for the Nth partial sum. Then $a^{(N)} = \tilde{\nabla}A^{(N)} + O(h^{N+1})$.

We now introduce the pseudodifferential operator

$$S = e^{ih\langle D_y, D_\xi \rangle} = \sum_{m=0}^{\infty} \frac{(ih)^m}{m!} D_y \cdot D_\xi.$$

A priori this makes no sense, but it does make sense modulo $O(h^{\infty})$. In fact if $a = \sum_{m} a_{m}h^{m}$, $b = \sum_{m} b_{m}h^{m}$ we say Sa = b if and only if

$$b_j = \sum_{m+p=j} \frac{i^m}{m!} (D_{\xi} \cdot D_y)^m a_p$$

where the dot product of differential operators is defined by $D_{\xi} \cdot D_y = \sum_j D_{\xi_j} D_{y_j}$. Moreover $S^{-1} = e^{-ihD_y \cdot D_{\xi}}$. Then $S^{-1}Sa = a$ up to $O(h^{\infty})$.

Lemma 18.39. *a* is a negligible symbol if and only if $Sa|_{x=y} = 0$.

Proof. $S\partial_{\xi} = \partial_{\xi}S$ and, thinking of y_j as the multiplication operator, $Sy_j = y_jS + hD_{\xi_j}S$. Therefore

$$S((x-y)A) = (x-y)SA - hD_{\xi_j}SA.$$

If a is negligible, say $a = \tilde{\nabla} A$, then

$$Sa = S\tilde{\nabla}A = \partial_{\xi_j}SA + i(x-y)SA/h - \partial_{\xi_0}SA = i(x-y)SA/h$$

which restricts to 0 on x = y. Conversely, if $Sa|_{x=y} = 0$ we can write Sa = (x - y)B. Now $S\tilde{\nabla} = i(x - y)S/h$ so

$$\tilde{\nabla}S^{-1} = iS^{-1}(x-y)/h.$$

Therefore

$$\tilde{\nabla}S^{-1}B = iS^{-1}(x-y)B/h.$$

So $ia/h = \tilde{\nabla}(S^{-1}B)$.

Definition 18.40. Let $Y \subset \mathbb{C}^N$ be a \mathbb{R} -linear subspace. Then Y is a *totally real space* if $Y \cap iY = 0$.

Assume φ is analytic as a function on $\mathbb{R}^n \times \mathbb{R}^n$. We extend φ to a function on the maximal totally real subspace $\{(x, \overline{x}) \in \mathbb{C}^n \times \mathbb{C}^n\}$. To do this, let $\iota(x, y) = (x + y, -i(x - y))$. Then let $\iota^* \tilde{\varphi} = \varphi$. We then can find a ψ which is holomorphic near (0, 0) such that $\psi(z, \overline{z}) = \varphi(z)$. Then $\psi(x, \overline{y}) = \psi(y, \overline{x})$ with

$$2\operatorname{Re}\psi(x,\overline{y}) - \varphi(x) - \varphi(y) \le -\delta|x-y|^2.$$

We solve for ξ in the equation

$$i\langle\xi, x-y\rangle = 2(\psi(x,z) - \psi(y,z)).$$

This is possible by the lemma on negligible symbols and ξ is a function of (x, y, z). By Taylor's theorem,

$$\xi(x,y,z) = -2i \int_0^1 \partial_x \psi(tx + (1-t)y,z) dt$$

so $\xi(x, x, z) = -2i\partial_x \psi(x, z)$. In fact $(x, y, z) \mapsto (x, y, \xi)$ is a biholomorphic function close to 0 on \mathbb{C}^{3n} by the analytic implicit function theorem. To see this we must show that the quadratic form given by the matrix $\partial_z \xi(0)$ is nondegenerate. Now

$$\partial_z \xi(0) = -2i\partial_x \partial_z \psi(0) = -2i\partial_x \overline{\partial}_x \varphi(0)$$

which is nondegenerate since φ is strictly plush. By our previous estimate on ψ , the contour

$$\Lambda = \{(y,\xi) : \xi = \xi(x,y,\overline{y})\}$$

is good.

Lemma 18.41. We have

$$u(x) = (\pi h)^{-n} \int_{B(0,1)} \chi(y) e^{i(2\psi(x,\overline{y}) - 2\psi(y,\overline{y}))/h} \frac{i}{2} \det \overline{\partial}_y \xi(1 + a^{(N)}) u(y) \, d\overline{y} \wedge dy + O(e^{\varphi(x)/h} h^{N+1}) ||u||_{\varphi}$$

Definition 18.42. The *Bergman function* is defined on a compact complex manifold equipped with a positive line bundle by

$$B(x) = ||K(x, x)|| = K(x, x)e^{-2\varphi(x)}.$$

Lemma 18.43. We have

$$B(x) = \sup_{\substack{s \in H^0(M \to L) \\ ||s||_{L^2} \le 1}} ||s(x)||^2.$$

Proof. We have

$$K(x,y) = \sum_{\alpha} u_{\alpha}(x) \overline{u_{\alpha}(y)},$$

the sum ranging over an orthonormal basis u_{α} of $H^0(M \to L)$. It follows that

$$B(x) = \sum_{\alpha} ||u_{\alpha}(x)||^2$$

and now we use the Cauchy-Schwarz inequality.

From this we can easily compute

$$\int_M B \ dV = \dim H^0(M \to L).$$

Given $L \to M$ fixed, we let B_k denote the Bergman function determined by L^k .

Lemma 18.44. The Bergman function satisfies

$$|B_k(x)| \le Ck^n.$$

Therefore dim $H^0(M \to L^k)$ is finite and grows like k^n . Morally this lemma is a restatement of the uncertainty principle because it describes how many independent states can be on a compact set, with h = 1/k. The uncertainty principle says we cannot localize past $\sqrt{1/h} = \sqrt{k}$. Here we cannot localize too much because we only have so many holomorphic sections.

Proof. Let $s \in H^0(M \to L^k)$. We want to show

$$||s(x)||^2 \le Ck^n ||s||_{L^2_{\mu}}.$$

Then the claim follows from the previous lemma. Since $||s||_{L_k^2} \leq ||s||_{L_k^2(\Omega_j)}$ we might as well prove the claim locally.

Choose an atlas so $\theta_j(x_0) = 0$ and write s(x) = (x, u(x)) where u is holomorphic near $x_0 = 0$. Now φ_j is strictly plurisubharmonic so up to a linear change of variables

$$\varphi_j(x) = \sum_{j=1}^n \lambda_j |x_j|^2 + \operatorname{Re} Q(x) + \varphi(0) + O(|x|^3)$$

where $Q(x) = \langle Ax, x \rangle + \langle a, x \rangle$ is holomorphic. Now $\varphi(0)$ does not affect our computation so we might as well take $\varphi(0) = 0$. By absorbing the holomorphic term Q into u we may assume Q = 0. Close to 0 we may assume $O(|x|^3) = 0$. Now

$$\frac{||u(0)||^2}{||u|_{L^2_k(\Omega_j)}} \le \frac{||u(0)||^2}{||u||_{L^2_k(B(0,R_k))}}$$

and we now make the change of variables

$$f(w) = u(w/\sqrt{k})$$

and

$$k\varphi(w) = \varphi_0(w) = \frac{1}{2} \sum_j \lambda_j |w_j|^2 + O(|w|^3 k^{-1/2}).$$

By our assumptions on φ , u(0) = f(0). Here dm and dV are absolutely continuous to each other so we can replace dV with dm up to a constant error. Hence

$$\frac{||u(0)||^2}{||u||_{L^2_k(\Omega_j)}} \le \frac{Ck^n ||f(0)||^2}{\int_{|w| \le B_k(\sqrt{k})} |f(w)|^2 e^{-2\varphi_0(w)} dm(w)} \\ \le \frac{Ck^n ||f(0)||^2}{\int_{|w| \le B_k(\sqrt{k})} |f(w)|^2 e^{-\sum_j \lambda_j |w_j|^2} dm(w)}.$$

Now $|f|^2$ is strictly plush so by the mean value theorem for measures,

$$\int_{|w| \le B_k(\sqrt{k})} e^{-\sum_j \lambda_j |w_j|^2} |f(0)|^2 \, dm(w) \le \int_{|w| \le B_k(\sqrt{k})} |f(w)|^2 e^{-\sum_j \lambda_j |w_j|^2} \, dm(w).$$

Pulling out $|f(0)|^2$ and taking k big enough that the integral converges we see our lemma. \Box

Actually Mike Christ showed that

$$|K(x,y)| \le Ce^{-\sqrt{k\log k}} d(x,y).$$

But we have only proved $|K(x,y)| = O(k^{-\infty})$.

Theorem 18.45. Assume φ is real analytic. If $d(x, y) < \varepsilon$ then in a suitable trivialization,

$$\partial_{x,y}^{\alpha}(K(x,y) - K^{(N)}(x,y)) = O(k^{-N-1+n+|\alpha|} e^{k(\varphi(x) + k\varphi(y))} ||u||_k.$$

To prove the theorem we rephrase the Hormander L^2 -estimates to work for manifolds.

Lemma 18.46. For every (0,1)-form $f \in C_{0,1}^{\infty}(M \to L^k)$ such that $\overline{\partial} f = 0$ there is a $u \in C^{\infty}(M \to L^k)$ such that $\overline{\partial} u = f$ and $||u||_{L^2_k} \leq C||f||_{L^2_k}$, and C can be chosen independently of k.

Lemma 18.47. One has

$$K(y,x) = (xK_x, K_y^{(N)})_{L^2_k} + O(k^{-N-1+n})e^{k\varphi(x) + k\varphi(y)}||K_x||_{L^2_k}.$$

Proof. We use the fact that

$$u(y) = (\chi_y u, K^N(\cdot, y))_{L^2_k} + O(k^{-N-1+n} e^{k\varphi(y)}) ||u||_{\varphi}$$

on the function $u(y) = K_x(y)$. Then

$$K(y,x) = (\chi K_x, K_y^{(N)})_{L_k^2} + O(k^{N-1})e^{k\varphi(y)}||u||_{L_k^2}$$

Here

$$||K_x||_{L^2_k}^2 = B_k(x)e^{2k\varphi(x)}$$

 \mathbf{SO}

$$||K_x||_{L^2_k} \le Ck^{n/2}e^{k\varphi(x)}.$$

Now use

$$e^{k\varphi(y)}||u||_{L^2_k}^2 \le e^{k\varphi(x)+k\varphi(y)}k^{n/2}$$

to prove the lemma.

Recall also that if $P: H \to \ker A$ is an orthogonal projection and Pu = v, w = u - Pu, then A(u - Pu) = Au and

$$||w|| = \min_{A\tilde{w} = Au} ||\tilde{w}||.$$

Proof of theorem. We may assume that χ is real-valued. Then

$$(\chi K_x, K_y^{(N)})_{L_k^2} = (K_x, \chi K_y^{(N)})_{L_k^2} = P_k(\chi K_x^{(N)})(y).$$

Then we define

$$u_y(x) = K_y^{(N)}(x) - (\chi K_y^{(N)}, K_x)_{L_k^2} = K_y^{(N)}(x) - P_k(\chi K_y^{(N)})(x).$$

By the linear algebra above, u_y is the L_k^2 -minimal solution to the problem

$$\overline{\partial} u_y = \overline{\partial} \chi K_y^{(N)} + \chi \overline{\partial} K_y^{(N)} = \overline{\partial} \chi K_y^{(N)}$$

because φ was assumed analytic. Moreover, $\overline{\partial}\chi = 0$ away from the diagonal. So

$$\overline{\partial}u_y(x) = O(e^{-\delta k}e^{k\varphi(x) + k\varphi(y)}).$$

Since u_y is the minimal solution, $||u_y||_{L^2_k} \leq ||\tilde{u}||_{L^2_k}$ where \tilde{u} is the solution to the $\overline{\partial}$ -problem yielded by the Hormander estimate. Therefore the Hormander estimate gives

$$||u_y||_{L^2_k} \le C e^{-\delta k} e^{k\varphi(y)}.$$

By the Cauchy-Green formula we have, for any compactly supported ψ which is identically 1 near 0,

$$|u(0)| \le \sup_{D} e^{k\varphi(y)} O(e^{-\delta/k} e^{k\varphi(y)})$$

for any open neighborhood D of 0. If we take $D \subset B(0, k^{-1})$ and translate appropriately we have

$$|u_y(x)| \le e^{-\delta k} e^{k\varphi(y)} e^{k\varphi(x)}.$$

Therefore we can actually approximate the Bergman kernel by its asymptotic expansion. We use this theorem to find an asymptotic expansion for the Bergman function.

Corollary 18.48. One has

$$B_k(x) = \left(\frac{k}{\pi}\right)^n \left(1 + k^{-1}b_1(x,\overline{x}) + k^{-2}b_2(x,\overline{x}) + \dots\right)$$

where the sum is meant in the asymptotic sense.

Corollary 18.49. One has

dim
$$H^0(M \to L^k) = \frac{k^n}{\pi^n} (1 + O(k^{-1})) \int_M dV$$

All that remains is to extend from the real-analytic case to the C^{∞} case. To do this we use a technique introduced by Hormander and Nirenberg.

Lemma 18.50. Suppose $f \in C^{\infty}_{comp}(\mathbb{R}^m)$. Then there is a $\tilde{f} \in C^{\infty}(\mathbb{C}^m)$ such that $\tilde{f}|_{\mathbb{R}^m} = f$ and $\overline{\partial}\tilde{f}(z) = O(|\operatorname{Im} z|^{\infty})$.

Proof. By a Paley-Weiner type theorem, we can define

$$\tilde{f}(x+iy) = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} e^{ix\xi} \chi(|\xi||y|) e^{-y\xi} \hat{f}(\xi) \ d\xi$$

for some compactly supported function χ which is identically 1 at 0. Obviously \tilde{f} is smooth, and $\overline{\partial}$ only falls on the χ term. Then we end up with χ' , which is not supported at 0, and a factor of $|\xi|y_j|y|^{-1}$. We can then add factors of $|\xi||y|^{-1}$ with impunity since $\chi^{(N)}$ is 0 close to 0. Iterating we get the decay we need.

Definition 18.51. The function \tilde{f} is called the *almost analytic extension* of f.

Using the theory of almost analytic extensions, one can prove the following lemma.

Lemma 18.52. Assume φ is smooth. There is a function ψ which is C^{∞} in a neighborhood of $0 \in \mathbb{C}^{n+n}$ such that $\psi(x, \overline{x}) = \varphi(x)$ and $\overline{\partial}_x \psi(x, \overline{y})$ and $\partial_y \psi(x, \overline{y}) = O(|x - y|^{\infty})$.

Now notice that we can replace any assumption that $\overline{\partial}\varphi = 0$ with $\overline{\partial}\varphi$ with the property in the above lemma anywhere in the above construction of the approximate Bergman kernel, since the function was only an approximation away from the diagonal anyways. This completes the proof of the theorem of Fefferman, Boutet, de Monvel, etc.

18.4 Morphisms into projective space

Throughout this section, let M be a compact complex manifold, $L \to M$ a line bundle. Let ω_{FS} be the Fubini-Study metric on O(1), the dual of the tautological line bundle, of \mathbb{P}^n for some n. We let $(s_j^k)_j$ be an orthonormal basis of $H^0(M \to L^k)$. Define the morphism

$$\varphi_k : M \to \mathbb{P}^{d_k - 1}$$
$$x \mapsto [s_0^k(x), \dots, s_{d_k}^k(x)].$$

Theorem 18.53 (Catlin-Tian-Yau-Zelditch asymptotics). Let ω be a positive Hermitian metric on L. Then there is an asymptotic expansion

$$\omega = \frac{1}{k}\varphi_k^*\omega_{FS} + \frac{\omega_2}{k^2} + \frac{\omega_3}{k^3} + \dots$$

for some ω_i which are (1, 1)-forms.

Proof. Write $s_j^k(z) = f_j^k(z)e_k$ for some holomorphic functions f_j^k . Then

$$\varphi_k^* \omega_{FS} = i \partial \overline{\partial} \left(\log \sum_{j=1}^{d_k} |f_j^k(z)|^2 \right).$$

On the other hand, the Bergman projector is given by

$$B_k(z) = e^{-2k\varphi(z)} \sum_{j=1}^{d_k} |f_j^k(z)|^2.$$

Therefore

$$i\partial\partial\log B_k(z) = -k\omega + \varphi_k^*\omega_{FS}.$$

So

$$\omega = \frac{1}{k} \varphi_k^* \omega_{FS} + \frac{1}{ik} \partial \overline{\partial} \log B_k(z)$$
$$= \frac{1}{k} \varphi_k^* \omega_{FS} + \frac{1}{ik} \partial \overline{\partial} \frac{a_1(x)}{k} + \dots$$

as $k \to \infty$.

Definition 18.54. A Hodge metric is a positive (1, 1)-form ω such that $[\omega] \in H^2(M, \mathbb{Q})$.

In other words, ω is in the second rational-valued cohomology class of M.

Theorem 18.55 (Kodiara embedding theorem). The following are equivalent:

- 1. If n is large enough, then M can be embedded in \mathbb{P}^n .
- 2. M admits a positive Hermitian line bundle.
- 3. *M* admits a Hodge metric.

In the below proof we use φ to mean the weight and φ_k to mean the embedding, oops.

Proof. Obviously if we have an embedding in projective space we can just pull back $(O(1), \omega_{FS})$ to M. The proof that positive Hermitian line bundles are equivalent to Hodge metrics uses Chern classes.

If M has a positive Hermitian line bundle, then φ_k is an immersion $M \to \mathbb{P}^{d_k}$. Suppose that φ_k is never injective. Then there are sequences x_k, y_k such that $x_k \neq y_k$ and $\varphi_k(x_k) = \varphi_k(y_k)$.

First suppose $d(x_k, y_k)\sqrt{k} \to \infty$. Since $\varphi_k(x_k) = \varphi_k(y_k)$, for all x, the Bergman kernel K has $K(x, x_k) = K(x, y_k)$. Then

$$\int_{B(x_k, r_{k/2})} |K_k(x, x_k)|^2 e^{-2k\varphi(x)} \, dV(x) \sim e^{2k\varphi(x_k)} \frac{k^n}{\pi^n}$$

yet

$$\int_{B(x_k, r_{k/2})} |K_k(x, x_k)|^2 e^{-2k\varphi(x)} \, dV(x) = \int_{B(y_k, r_{k/2})} |K_k(x, y_k)|^2 e^{-2k\varphi(x)} \, dV(x).$$

Now $d(x_k, y_k) \gg k^{-1/2}$ so the balls can be taken to be disjoint. Then

$$K_k(x_k, x_k) = \int_M |K_k(x, x_k)|^2 e^{-2k\varphi(x)} \, dV(x) \ge \frac{k^n}{\pi^n} (e^{2k\varphi(x_k)} + e^{2k\varphi(y_k)}).$$

But

$$K_k(x_k, y_k) \sim \frac{k^n}{\pi^n} e^{2k\varphi(x_k)}.$$

Without loss of generality we can assume $\varphi(y_k) \ge \varphi(x_k)$. But

$$e^{2k\varphi(x_k)} \gtrsim e^{2k\varphi(x_k)} + e^{2k\varphi(y_k)}$$

which is a contradiction.

Otherwise we can assume $d(x_k, y_k) \leq Ck^{-1/2}$. By changing coordinates we may assume that $x_k = 0, y_k = w_k k^{-1/2}, |w_k| \leq C$, and $x_k \neq y_k$. Now put

$$f_k(t) = \frac{|K_k(0, tw_k k^{-1/2})|^2}{K_k(0, 0)K_k(tw_k k^{-1/2}, tw_k k^{-1/2})}.$$

Then f_k is smooth on [0, 1] and by the Cauchy-Schwarz inequality $0 \le f_k(t) \le 1$. In fact $f_k(0) = 1$ and $f_k(1) = 1$. So there is a $t_k \in [0, 1]$ such that $f''_k(t_k) \ge 0$. But

$$f_k(t) = \exp(-2k\psi(0, tw_k k^{-1/2} - \varphi(0) - \varphi(tw_k k^{-1/2})))(1 + r(tw_k)k^{-1} + \dots)$$

= $\exp(-t^2 \langle Aw_k, \overline{w}_k \rangle + O(t^3 |w_k|^3))(1 + r_k(t_w k)k^{-1} + \dots)$

which implies $|w_k| = O(k^{-1/2})$. So

$$f_k''(t) = -2\langle Aw_k, \overline{w}_k \rangle O(|w_k|^3 k^{-1/2} + |w_k|^2 k^{-1}) \le -C|w_k|^2 < 0$$

which is a contradiction.

Theorem 18.56 (Chow). If M admits a positive Hermitian line bundle, then M is a projective variety.

Proof. A compact submanifold of \mathbb{P}^n is a projective variety.

18.5 Zworski's conjecture

Conjecture 18.57 (Zworski). Let $\Pi_N : L^2(\mathbb{P}^n \to O(N)) \to H^0(\mathbb{P}^n \to O(N))$ be the Bergman projector, and consider the differential equation

$$iu_t = \prod_N (|u|_N^2 u)$$

on the finite-dimensional space $H^0(\mathbb{P}^n \to O(N))$. Then this differential equation has a Lax pair and hence is a completely integrable system.

Part V Harmonic analysis

Chapter 19

Rearrangement-invariant spaces

We are interested in studying integral transforms. One often only defines these transforms on test functions and extends them using a continuity argument. For example, the *Hilbert* transform is defined by

$$Hf(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(y)}{x - y} \, dy.$$

This makes no sense if $f(0) \neq 0$, for example, but one can fix this by taking

$$Hf(x) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \left(\int_{-\infty}^{\varepsilon} + \int_{\varepsilon}^{\infty} \right) \frac{f(y)}{x - y} \, dy$$

Then Hf converges if f is a Schwartz function. One can then check that for every $p \in (1, \infty)$ there is a constant C_p such that for every f for which Hf is defined,

$$||Hf||_{L^p} \le C_p ||f||_{L^p}.$$

So H extends uniquely to a bounded linear operator on L^p , since such f are dense in H.

We therefore consider certain generalizations of L^p spaces. Fix a σ -finite measure space (X, Σ, μ) .

Definition 19.1. A norm $||\cdot||$ is *rearrangement-invariant* if for every isomorphism of measure spaces $\Phi : X \to X$ and every measurable function f we have

$$||f|| = ||f \circ \Phi||.$$

Definition 19.2. A norm $||\cdot||$ is *monotone* if for all measure functions f, g such that $|f| \le |g|$ a.e. we have

$$||f|| \le ||g||.$$

A rearrangement-invariant, monotone norm ignores the sign of a function, and is only interested in how "tall" a function is versus how "wide" the function is. Clearly L^p norms are rearrangement-invariant and monotone.

19.1 Log-convexity for L^p norms

Recall that L^p norms are *log-convex*, i.e. if $\theta \in (0, 1)$ and

$$\frac{1}{r} = \frac{1-\theta}{p} + \frac{\theta}{q},$$

we have

$$||f||_{L^r} \le ||f||_{L^p}^{1-\theta} ||f||_{L^q}^{\theta}$$

This is essentially equivalent to the Hölder inequality. We give another proof of log-convexity.

Theorem 19.3 (three-lines lemma). Let f be holomorphic on the strip $0 \leq \text{Re } z \leq 1$ of at most double-exponential growth. Suppose that $|f(z)| \leq A$ when Re z = 0 and $|f(z)| \leq B$ when Re z = 1. Then

$$|f(z)| \le A^{1 - \operatorname{Re} z} B^{\operatorname{Re} z}.$$

Proof. The hypothesis and conclusion are invariant under multiplying f by Ce^{cz} where $C, c \in \mathbb{R}$, as long as we appropriately rescale A, B. Thus we may assume that A = B = 1, and want to show $||f||_{L^{\infty}} \leq 1$.

Assume f goes to 0 at infinity, so there is an N such that if $\text{Im } z \ge N$, we have $|f(z)| \le 1$. Then by the maximum principle, $||f||_{L^{\infty}} = 1$.

Otherwise, note that

$$f(z) = \lim_{\varepsilon \to 0} f(z) \exp(\varepsilon e^{i((\pi - \varepsilon)z + \varepsilon/2)}),$$

and by assumption on f, it follows that the function in the limit goes to 0, so the right-hand side is ≤ 1 .

Proof of log-convexity. Let f be a simple function and suppose $\mu(X) < \infty$. Let

$$g(z) = \int_X |f|^z.$$

Then $|g(z)| = O(e^{|z|})$ and $|g(s+it)| = ||f||_{L^s}^s$. The claim follows for f by rescaling the three-lines lemma.

The claim is invariant under monotone convergence and μ is σ -finite, so the assumptions that f is simple and μ is finite can be dropped.

Log-convexity allows us to prove that certain L^p spaces embed in others.

Corollary 19.4. If $\mu(\operatorname{supp} f) < \infty$ and $p \leq q$ then

$$||f||_{L^p} \le \mu(\operatorname{supp} f)^{1/p-1/q} ||f||_{L^q}.$$

In particular, if $\mu(X) < \infty$, then $L^q(\mu) \subseteq L^p(\mu)$.

Proof. This is obvious if $q = \infty$ and otherwise follows by log-convexity.

Definition 19.5. A granular measure is a measure ν such that there is a $\delta > 0$ such that for every set A which is not ν -null, $\nu(A) > \delta$.

 \square

Corollary 19.6. If μ is δ -granular and $p \leq q$ then

$$||f||_{L^q} \leq \delta^{1/q-1/p} ||f||_{L^p}.$$

In particular, $L^p(\mu) \subseteq L^q(\mu)$.

The proof is the same. In particular, ℓ^p spaces satisfy this estimate, and if card supp $f = N, f \in \ell^p$, then we have

$$||f||_{\ell^q} \le ||f||_{\ell^p} \le N^{1/p-1/q} ||f||_{\ell^q}$$

if $q \ge p$. Though Lebesgue measure is not granular, if a function is constant on boxes which tile \mathbb{R}^n , one can replace the Borel algebra with the σ -algebra generated by the boxes, and restrict Lebesgue measure to this σ -algebra; then it becomes granular.

Now note that by log-convexity with $q = \infty$, if $p \leq q$ we have

$$||f||_{L^q} \leq C^{1-p/q} ||f||_{L^p}^{p/q}$$

if $|f| \leq C$. On the other hand, if $|f| \geq c$ on supp f then replacing f by 1/f in the above estimate we have

$$||f||_{L^p} \le c^{1-q/p} ||f||_{L^q}^{q/p}.$$

19.2 Lorentz norms

Still fix a measure space (X, μ) .

Definition 19.7. The weak L^p norm is defined by

$$||f||_{L^{p,\infty}} = \sup_{\lambda>0} \lambda \mu (\{|f| \ge \lambda\})^{1/p}.$$

Theorem 19.8 (Chebyshev inequality). One has

 $||f||_{L^{p,\infty}} \le ||f||_{L^p}.$

Proof. Take \sup_{λ} of both sides of the inequality

$$\lambda^p \mu(\{|f| \ge \lambda\}) \le \int_{|f| \ge \lambda} \lambda^p \ d\mu \le \int_X |f|^p \ d\mu = ||f||_{L^p}^p.$$

So $L^p \subseteq L^{p,\infty}$. This motivates the terminology "weak L^p ".

Let λ be the coordinate function on \mathbb{R}^+ . Recall that $d\lambda/\lambda$ is the Haar measure on \mathbb{R}^+ , the group of positive reals under multiplication. Now $L^{\infty}(d\lambda/\lambda) = L^{\infty}(\mathbb{R}^+)$ since Haar and Lebesgue measures are mutually absolutely continuous. Thus we have

$$||f||_{L^{p,\infty}} = p^0 ||\lambda \mu(\{|f| \ge \lambda\})^{1/p}||_{L^{\infty}(d\lambda/\lambda)}.$$

This seems a little silly until we reason by metaphor with the following lemma.

Lemma 19.9. One has

$$||f||_{L^p} = p^{1/p} ||\lambda \mu(\{|f| \ge \lambda\})^{1/p}||_{L^p(d\lambda/\lambda)}$$

Proof. One has

$$p \int_0^{|f(x)|} \lambda^p \ \frac{d\lambda}{\lambda} = \int_0^{|f(x)|} p\lambda^{p-1} \ d\lambda = |f(x)|^p$$

by the fundamental theorem. Integrating both sides and using Fubini's theorem, we have

$$\begin{split} ||f||_{L^{p}}^{p} &= p \int_{X} \int_{0}^{|f(x)|} \lambda^{p} \frac{d\lambda}{\lambda} dx \\ &= p \int_{0}^{\infty} \lambda^{p} \int_{X} \mathbf{1}_{|f(x)| \ge \lambda} dx \frac{d\lambda}{\lambda} \\ &= p \int_{0}^{\infty} \lambda^{p} \mu(\{|f| \ge \lambda\}) \frac{d\lambda}{\lambda}. \end{split}$$

The claim follows by taking pth roots.

Thus we make the following definition.

Definition 19.10. The (p,q)th Lorentz norm is defined by

$$||f||_{L^{p,q}} = p^{1/q} ||\lambda \mu(\{|f| \ge \lambda\})^{1/p}||_{L^q(d\lambda/\lambda)}.$$

Thus the definition of $L^{p,\infty}$ is consistent with that of the weak L^p norm and we have $L^{p,p} = L^p$.

We now consider certain approximate step functions that will allow us to determine what the Lorentz norms actually measure.

Definition 19.11. A substep function of height H and width W is a function f such that $||f||_{L^{\infty}} \leq H$ and $\mu(\operatorname{supp} f) \leq W$.

A δ -quasistep function of height H and width W is a function f such that $\delta H \leq |f| \leq H/\delta$ and $\delta W \leq \operatorname{supp}(f) \leq H/\delta$.

Usually the exact value of δ is not important, and we suppress the δ by writing $|f| \sim H$. However, one is always implicitly fixed when working with a family of quasistep functions; for example, every function is trivially 0-quasistep, which is not very useful.

We can use substep and quasistep functions to understand the behavior of Lorentz norms.

Theorem 19.12. Let f be a measurable function and let p, q, A be given. Then the following are equivalent:

- 1. $||f||_{L^{p,q}} \leq_{p,q} A.$
- 2. There is a decomposition $f = \sum_{m=-\infty}^{\infty} f_m$ where f_m is quasistep of height 2^m and width W_m such that the supp (f_m) are disjoint and

$$\left(\sum_{m=-\infty}^{\infty} 2^{mq} W_m^{q/p}\right)^{1/q} \lesssim_{p,q} A.$$

203

3. There is a pointwise estimate $|f| \leq \sum_{m=-\infty}^{\infty} 2^m \mathbf{1}_{E_m}$ where

$$\left(\sum_{m=-\infty}^{\infty} 2^{mq} \mu(E_m)^{q/p}\right)^{1/q} \lesssim_{p,q} A.$$

4. There is a decomposition $f = \sum_{n=-\infty}^{\infty} f_n$ where f_n is substep of width 2^n and height H_n such that the supp (f_n) are disjoint, H_n are nonincreasing, $H_{n+1} \leq |f_n| \leq H_n$ on supp f_n , and

$$\left(\sum_{n=-\infty}^{\infty} H_n^q 2^{nq/p}\right)^{1/q} \lesssim_{p,q} A.$$

5. There is a pointwise estimate $|f| \leq \sum_{n=-\infty}^{\infty} H_n \mathbf{1}_{E_n}$ where $\mu(E_n) \lesssim_{p,q} 2^n$ and

$$\left(\sum_{n=-\infty}^{\infty} H_n^q 2^{nq/p}\right)^{1/q} \lesssim_{p,q} A.$$

Intuitively, if f is quasistep of height H, W then $||f||_{L^{p,q}} \approx HW^{1/p}$, and this estimate propagates in *sums* of quasistep functions in ℓ^q norm.

Proof. We may assume that A = 1. The implications $2 \implies 3$ and $4 \implies 5$ are obvious. Assume 1. Let $E_m = \{2^{m-1} < |f| \le 2^m\}$, and carry out the dyadic decomposition $W_m = \mu(E_m), f_m = f \mathbb{1}_{E_m}$. Clearly

$$2^m W_m^{1/p} \lesssim_{p,q} ||\lambda \mu(\{|f| \ge \lambda\})^{1/p}||_{L^q([2^{m-2}, 2^{m-1}], d\lambda/\lambda)}$$

and summing both sides over m in $\ell^q,\,2$ holds.

Assume 3. TODO

Chapter 20 Decoupling theory

This chapter is based on Terry Tao's online course on harmonic analysis.

20.1 Square-root cancellation

If z_i are "random, independent" complex numbers, one intuitively expects

$$\left|\sum_{j} z_{j}\right| \approx \sqrt{\sum_{j} |z_{j}|^{2}}.$$

This is true for the expected value where the z_j are given random, uniformly selected signs, up to a constant. This is *Khinchine's inequality*.

Another way one can make the above claim rigorous is using Littlewood-Paley theory.

Theorem 20.1 (Littlewood-Paley estimates). Let $p \in (1, \infty)$, A_j be the annuli in frequency space

$$A_j = \{\xi \in \mathbb{R}^d : 2^j \le |\xi| \le 2^{j+1}\}.$$

Assume that f_j are functions such that \hat{f}_j are supported on disjoint annuli A_{k_j} . Then

$$||\sum_j f_j||_{L^p} \lesssim_{p,d} ||\sqrt{\sum_j |f_j|^2}||_{L^p}.$$

If the annuli are separated then we can replace $\leq_{p,d}$ with $\sim_{p,d}$.

We will prove this later on. We note however that it is trivial when p = 2. In fact, since the \hat{f}_j have disjoint support, the f_j are orthogonal by Plancherel's theorem. So by the Pythagorean theorem,

$$||\sum_{j} f_{j}||_{L^{2}} = ||\sqrt{\sum_{j} |f_{j}|^{2}}||_{L^{2}}.$$

If we even just have "almost-orthogonality" we still get decent estimates in L^2 . One can manipulate the L^2 estimate to get estimates in L^{2^k} , $k \ge 2$, but one needs something stronger than orthogonality here, and as $k \to \infty$ this becomes extremely complicated. To generalize to p > 2 where $p \neq 2^k$, we want to study *decoupling estimates*, those of the form

$$||\sum_{j} f_{j}||_{L^{p}} \lesssim_{p,d} \sqrt{\sum_{j} ||f_{j}||_{L^{p}}^{2}}$$

This is weaker than a square root cancellation estimate, but it is much easier to prove. If p = 2 it follows easily by noting that

$$||\sum_{j} |f_{j}|^{2}||_{L^{1}} \leq \sum_{j} |||f_{j}|^{2}||_{L^{1}}$$

by the triangle inequality, and then taking the square root of both sides. One can therefore think of decoupling estimates as vast generalizations of the Pythagorean theorem. The reason these estimates are called decoupling estimates is that we start with the left-hand sides, where we take a norm of $\sum_j f_j$, and then on the right-hand side we take norms of each of the f_j s separately, and they "decouple" in the sense that their norms no longer influence each other.

One can iterate decoupling estimates in the sense that

$$||\sum_{j_1,\dots,j_n} f_{j_1,\dots,j_n}||_{L^p} \lesssim_{p,d,n} \sqrt{\sum_{j_1,\dots,j_n} ||f_{j_1,\dots,j_n}||_{L^p}^2},$$

simply by induction on n. Thus decoupling plays nicely with induction, though notice that the constant possibly depends on n. An important concept in the study of decoupling estimates is to control the dependency of the constant on n.

Proof of Littlewood-Paley estimates. We first prove that if $|k_j - k_{j'}| \ge 2$ then

$$|| \sqrt{\sum_{j} |f_j|^2} ||_{L^p} \lesssim_{p,d} || \sum_{j} f_j ||_{L^p}.$$

Let ψ be a bump function on $\{\xi : 1 - \varepsilon < |\xi| < 2 + \varepsilon\}$ which is identically 1 on $\{\xi : 1 < |\xi| < 2\}$. Let

$$\widehat{P_{k_j}f}(\xi) = \psi(2^{-k_j}\xi)\widehat{f}(\xi)$$

so P_{k_j} is a Fourier multiplier (actually a Littlewood-Paley projector), which is bounded $L^p \to L^p$. Let $f = \sum_j f_j$, so $f_j = P_{k_j} f_j$. Let $\epsilon_1, \ldots, \epsilon_n$ be random signs, drawn uniformly and independently, and let

$$P = \sum_{j} \epsilon_{j} P_{k_{j}}$$

be a random Fourier multiplier, which again is bounded $L^p \to L^p$ uniformly in the choice of signs. (Here we use p > 1 and Calderon-Zygmund theory, or more accurately the Hormander-Mikhlin theorem.) Thus

$$||\sum_{j} \epsilon_{j} f_{j}||_{L^{p}} \lesssim_{p,d} ||\sum_{j} f_{j}||_{L^{p}}.$$

The claim then follows from Khinchine's inequality.

To prove $\gtrsim_{p,d}$, use the triangle inequality to split up into two sums, one where the k_j are odd and one where the k_j are even. This trick is called *sparsification*. We work with each term separately so that $|k_j - k_{j'}| > 2$ and we meet the hypotheses of $\leq_{p,d}$. Now

$$f_j = \epsilon_j P_{k_j} \sum_{\ell} \epsilon_{\ell} f_{\ell},$$

 \mathbf{SO}

$$f = \sum_{j} \epsilon_{j} P_{k_{j}} \sum_{\ell} \epsilon_{\ell} f_{\ell}.$$

Again by Calderon-Zygmund theory,

$$||\sum_{j} f_{j}||_{L^{p}} \lesssim_{p,d} ||\sum_{j} \epsilon_{j} f_{j}||_{L^{1}}$$

and we use Khinchine's inequality again.

Lemma 20.2. Let f_1, \ldots, f_n have Fourier supports $\Omega_1, \ldots, \Omega_n$. If any ξ lies in at most A_2 of the Ω_i ("the Ω_i have overlap A_2 ") then

$$||\sum_{j} f_{j}||_{L^{2}} \leq \sqrt{A_{2}}|| \sqrt{\sum_{j} |f_{j}|^{2}} ||_{L^{2}}.$$

If $\Omega_i + \Omega_j$ have overlap A_4 then

$$||\sum_{j} f_{j}||_{L^{4}} \leq \sqrt[4]{A_{4}}||_{\sqrt{\sum_{j}}} |f_{j}|^{2}||_{L^{4}}.$$

Proof. The first claim follows by using Plancherel to pass to Fourier space and then using Cauchy-Schwarz.

For the second claim, note that

$$||\sum_{j} f_{j}||_{L^{4}}^{2} = \sum_{ij} f_{i}f_{j}||_{L^{2}}$$

and since $f_i f_j$ has Fourier support $\Omega_i + \Omega_j$ this reduces the claim to the first claim.

20.2 Decoupling estimates

Definition 20.3. Let $\mathcal{U} = \{U_1, \ldots, U_n\}$ be a finite multiset of nonempty open sets and let $p \in [1, \infty]$. The *decoupling constant* $\text{Dec}_p(\mathcal{U})$ is the optimal constant in the inequality

$$||\sum_{j} f_{j}||_{L^{p}} \leq \operatorname{Dec}_{p}(\mathcal{U}) \sqrt{\sum_{j} ||f_{j}||^{2}}$$

whenever f_j has Fourier support in U_j .

If all but one of the f_j vanish we see that $\text{Dec}_p(\mathcal{U}) \geq 1$, and by Cauchy-Schwarz we have $\text{Dec}_p(\mathcal{U}) \leq \sqrt{\text{card}\,\mathcal{U}}$. Intuitively $\text{Dec}_p(\mathcal{U})$ measures how spread out \mathcal{U} is: if they are "more disjoint" then $\text{Dec}_p(\mathcal{U}) \to 1$.

One can check that if $U_i \subseteq U'_i$ then

$$\operatorname{Dec}_p(U_1,\ldots,U_n) \leq \operatorname{Dec}_p(U'_1,\ldots,U'_n).$$

We also have a triangle inequality

$$\operatorname{Dec}_p(\mathcal{U} \cup \mathcal{V}) \leq \sqrt{\operatorname{Dec}_p(\mathcal{U})^2 + \operatorname{Dec}_p(\mathcal{V})^2}.$$

If L is an affine transformation then L preserves Dec_p , so Dec_p is dimensionless.

They also interpolate: if $p_0 \leq p \leq p_1$ and

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1},$$

and

$$||P_{U_j}f||_{L^{p_i}} \lesssim_{p_i,d} ||f||_{L^{p_i}},$$

where P_{U_j} is the Littlewood-Paley projector, then

$$\operatorname{Dec}_p(\mathcal{U}) \lesssim_{p_0,p_1,p,d} \operatorname{Dec}_{p_1}(\mathcal{U})^{1-\theta} \operatorname{Dec}_{p_0}(\mathcal{U})^{\theta}.$$

If we have partitions $U_j = \bigoplus_i U_{j,i}$ into open sets up to measure zero, then

$$\operatorname{Dec}_p(\{U_{j,i}:j,i\}) \leq \operatorname{Dec}_p(\mathcal{U}) \sup_j \operatorname{Dec}_p(\{U_{j,i}:i\}).$$

We can also add dummy dimensions: if $d' \geq 1$ then

$$\operatorname{Dec}_p(\mathcal{U}) = \operatorname{Dec}_p(\{U_i \times \mathbb{R}^{d'} : i\}).$$

If p = 2, decoupling is easy: by a previous lemma, if the sets have overlap A_2 , then

$$\operatorname{Dec}_p(\mathcal{U}) \leq \sqrt{A_2}.$$

We now prove L^6 decoupling for the parabola.

Theorem 20.4. Let $\delta > 0$, let $\phi(\xi) = (\xi, \xi^2)$ parametrize the parabola in \mathbb{R}^2 , let Σ be a δ -separated subset of [-1, 1] and $\Omega = \phi(\Sigma)$, and take the parallelograms $R(\omega, \delta)$ centered on points ω of Ω with width 2δ , height $2\delta^2$, and slope equal to the slope of the parabola at ω . Then for any $\varepsilon > 0$,

$$\operatorname{Dec}_6(\{R(\omega, \delta) : \omega \in \Omega\}) \lesssim_{\varepsilon} \delta^{\varepsilon}.$$

This theorem was proven in the early 2010s by Bourgain and Demeter. We will give an argument of Li published in 2018. If p > 6 then the above decoupling estimate fails; we omit the proof of this.

To prove the above theorem, if $\delta > 0$ let

$$D(\delta) = \sup_{\Sigma} \operatorname{Dec}_6(R(\omega, \delta) : \omega \in \phi(\Sigma))$$

where Σ ranges over all δ -separated subsets of [-1, 1]. Then

$$1 \le D(\delta) \lesssim \delta^{-1/2}.$$

This is because in this case we would be decoupling $\sim \delta^{-1}$ many sets. We must show

$$D(\delta) \lesssim_{\varepsilon} \delta^{-\varepsilon}.$$

Note that D is not necessarily increasing in δ . If δ increases to $\delta + \varepsilon$, we cannot refine the partition Σ_{δ} to a partition $\Sigma_{\delta+\varepsilon}$ if ε is too small – so we end up with a completely different partition!

We first show that D is stable.

Lemma 20.5. If $\delta \sim \delta'$ then $D(\delta) \sim D(\delta')$.

Proof. This claim is transitive, so we may actually prove it when δ and δ' are not symmetrical, say

$$10\delta \leq \delta' \lesssim \delta.$$

We first show $D(\delta) \lesssim D(\delta')$, i.e. if Σ is δ -separated then

$$\operatorname{Dec}_6(R(\omega, \delta) : \omega \in \phi(\Sigma)) \lesssim D(\delta').$$

Using sparsification and the triangle inequality for Dec we may split up Σ into finitely many pieces, each of which is δ' separated, and then assume that Σ is δ' -separated. Thus

$$\operatorname{Dec}_6(R(\omega,\delta):\omega\in\phi(\Sigma))\lesssim\operatorname{Dec}_6(R(\omega,\delta'):\omega\in\phi(\Sigma))\lesssim D(\delta').$$

Conversely, we must show that if Σ' is δ' -separated then

$$\operatorname{Dec}_6(R(\omega', \delta') : \omega' \in \phi(\Sigma')) \lesssim D(\delta).$$

That is,

$$||\sum_{\omega'} f_{\omega'}||_{L^6} \lesssim D(\delta) \sqrt{\sum_{\omega'} ||f_{\omega'}||_{L^6}^2},$$

for any $f_{\omega'} \in L^6(\mathbb{R}^2)$ whose Fourier support is in $R(\omega', \delta')$. By breaking up the $R(\omega', \delta')$ into parallelograms $R(\phi(\xi), \delta) + (0, jc\delta^2)$ where j = O(1) is an index and c is a small absolute constant, and applying Littlewood-Paley projectors, we see the claim.

Lemma 20.6. Let $\delta \leq \delta_0$. Let Σ be δ -separated and contained in an interval I of length $2\delta_0$. Then

$$\operatorname{Dec}_6(R(\omega, \delta) : \omega \in \phi(\Sigma)) \leq D(\delta/\delta_0).$$

Proof. If $\delta_0 \sim 1$ then this follows from the stability lemma. Otherwise, choose an absolute constant c > 0 and $\delta_0 < c$. Acting on the parabola by the Galilean group we may assume I is centered at 0.

Now rescale the parabola by $(\xi_1, \xi_2) \mapsto (\xi_1/\delta_0, \xi_2/\delta_0^2)$. This has a nontrivial Jacobian but it's not a big deal. Then the claim is

$$\operatorname{Dec}_6(R(\omega,\delta):\omega\in\phi(\Sigma))\lesssim\operatorname{Dec}_6(R(\omega,1):\omega\in\phi(\Sigma/\delta_0)).$$

Thus we at least have a local induction in scales. In fact it implies

$$D(\delta_1 \delta_2) \lesssim D(\delta_1) D(\delta_2)$$

by using the triangle inequality to split Σ into small subsets where we can induct on scales. The trouble is that the implied constant in this estimate could be huge, and so every time we apply the induction hypothesis we could get a horrible buildup. So we need to replace this estimate with a slightly better estimate; annoyingly, nobody seems to know how to do this in general, other than trial and error.

Definition 20.7. Let $0 \leq \delta \leq \rho_1, \rho_2 \leq \nu \leq 1$. Let $M_{2,4}(\delta, \nu, \rho_1, \rho_2)$ be the best constant such that for any δ -separated sets Σ_1, Σ_2 contained in intervals I_1, I_2 of lengths ρ_1, ρ_2 , such that $d(I_1, I_2) \geq \nu$, and $f_{\omega_1}, g_{\omega_2}$ collections of Schwartz functions whose Fourier support is contained in $R(\omega_1, \delta), R(\omega_2, \delta)$,

$$\int_{\mathbb{R}^2} |\sum_{\omega_1 \in \phi(\Sigma_1) f_{\omega_1}}|^2 |\sum_{\omega_2 \in \phi(\Sigma_2)}|^4 \le M_{2,4}(\delta,\nu,\rho_1,\rho_2)^6 \sum_{\omega_1 \in \phi(\Sigma_1)} ||f_{\omega_1}||_{L^6}^2 \sum_{\omega_2 \in \phi(\Sigma_2)} ||g_{\omega_2}||_{L^6}^2$$

We might as well think of $\nu = 1$ but it is convenient to be able to perturb it. Confusingly, we want to hold δ fix and induct on ρ . The reason for this madness is because we want to use Hilbert space theory, and this is possible because of the sum of squares in f_{ω_1} . Otherwise we would use 3 + 3 instead of 2 + 4. In fact the L^4 term will look like a constant and then this reduces to basic facts about L^2 .

By Holder, we have

$$\int_{\mathbb{R}^2} |\sum_{\omega_1 \in \phi(\Sigma_1) f_{\omega_1}}|^2 |\sum_{\omega_2 \in \phi(\Sigma_2)}|^4 \le ||\sum_{\omega_1} f_{\omega_1}||_{L^6}^2 ||\sum_{\omega_1} ||_{L^6}^4.$$

Therefore

 $M_{2,4}(\delta,\nu,\rho_1,\rho_2) \le D(\delta).$

We will mainly be interested in when $\rho_1 = \rho_2 = \nu \gtrsim 1$.

Lemma 20.8. If $0 \le \delta \le \nu \le 1$ then

$$D(\delta) \lesssim \nu^{-O(1)} M_{2,4}(\delta,\nu,\nu,\nu) + D(\delta/\nu).$$

In this scenario, the $\nu^{-O(1)} \approx 1$ and $D(\delta/\nu)$ terms are basically irrelevant and so, heuristically, this lemma says that

$$D(\delta) \lesssim M_{2,4}(\delta,\nu,\nu,\nu)$$

which is what we want, because since the L^4 term is basically a constant, this reduces the L^6 decoupling problem to an L^2 problem. This gives a hint as to why 6 is the optimal exponent.

Proof. Let Σ be a δ -separated subset of [0, 1] and for each $\omega \in \phi(\Sigma)$, let $f_{\omega} \in L^{6}$ have Fourier support in $R(\omega, \delta)$. Normalizing, we assume $\sum_{\omega} ||f_{\omega}||_{L^{6}}^{2} = 1$. We then must show

$$\int_{\mathbb{R}^2} |\sum_{\omega} f_{\omega}|^6 \lesssim \nu^{-O(1)} M_{2,4}(\delta,\nu,\nu,\nu)^6 + D(\delta/\nu)^6.$$

Partition Σ into components Σ_I where each Σ_I is contained in some subinterval I of [-1, 1] of length ν . Let \mathcal{I} index the intervals, so card $\mathcal{I} = O(1/\nu)$. Then

$$|\sum_{\omega \in \phi(\Sigma)} f_{\omega}(x)| \le \sum_{I \in \mathcal{I}} |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}(x)|.$$

Then for each x, the pigeonhole principle furnishes an I_x such that

$$|\sum_{\omega \in \phi(\Sigma)} f_{\omega}(x)| \lesssim |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}(x)|,$$

the "narrow case", or there are I,J such that $d(I,J) \geq \nu$ and

$$|\sum_{\omega \in \phi(\Sigma)} f_{\omega}(x)| \lesssim \nu^{-O(1)} |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}(x)| |\sum_{\omega \in \phi(\Sigma_J)} f_{\omega}(x)|.$$

In particular,

$$|\sum_{\omega \in \phi(\Sigma)} f_{\omega}(x)| \lesssim \nu^{-O(1)} \sum_{d(I,J) \ge \nu} |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}|^2 |\sum_{\omega \in \phi(\Sigma_J)} f_{\omega}|^4 + \sum_I |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}|^6.$$

By definition,

$$\int_{\mathbb{R}^2} \sum_{\omega \in \phi(\Sigma_I)} f_{\omega}|^2 |\sum_{\omega \in \phi(\Sigma_J)} f_{\omega}|^4 \lesssim M_{2,4}(\delta, \nu, \nu, \nu).$$

Meanwhile, by the parabolic rescaling estimate,

$$\int_{\mathbb{R}^2} |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}|^6 \lesssim D(\delta/\nu)^6 (\sum_{\omega \in \phi(\Sigma_I)} ||f_{\omega}||_{L^6}^2)^3$$

but by our normalization, after summing over I, this implies that

$$\sum_{I} \int_{\mathbb{R}^2} |\sum_{\omega \in \phi(\Sigma_I)} f_{\omega}|^6 \lesssim D(\delta/\nu)^6.$$

Thus if we can get a good estimate in the endpoint case $\rho_1 = \rho_2 = \nu$ we are in good shape, as

$$D(\delta) \approx M_{2,4}(\delta, \nu, \nu, \nu)$$

when ν is much larger than δ . On the other hand it's only easy to estimate $M_{2,4}(\delta, \rho_1, \rho_2, \nu)$ when ρ_1, ρ_2 are very small.

Our next lemma allows to assume that *one* of the ρ 's is very small:

Lemma 20.9. Suppose $0 \le \delta \le \rho_2^2 \le \rho_1' \le \rho_1 \le \nu \le 1$ and $\delta \le \rho_2 \le \nu$ then

$$M_{2,4}(\delta,\nu,\rho_1,\rho_2) \lesssim \nu^{-O(1)} M_{2,4}(\delta,\nu,\rho_1',\rho_2).$$

Proof. It suffices to show that

$$\int_{\mathbb{R}^2} |\sum_{\omega_1 \in \phi(\Sigma_1)} f_{\omega_1}|^2 |\sum_{\omega_2 \in \phi(\Sigma_2)} g_{\omega_2}|^4 \lesssim \nu^{-O(1)} M_{2,4}(\delta,\nu,\rho_1,\rho_2)^6$$

whenever Σ_1, Σ_2 are of the appropriate lengths ρ_1, ρ_2 and separation δ, ν and the $f_{\omega_i}, g_{\omega_i}$ have Fourier support in $R(\omega_i, \delta)$. Here we normalize

$$\sum_{\omega_1} ||f_{\omega_1}||_{L^6}^2 = ||\sum_{\omega_2} g_{\omega_2}||_{L^6}^4 = 1.$$

We are viewing the integral we are trying to bound as a weighted sum of L^2 -norms where the L^4 -norms are weights.

Partition Σ_1 into δ -separated subsets $\Sigma_{1,I'}$ of intervals I' of length ρ'_1 of bounded overlap. The left-hand side of the claim is given by $||\sum_{I'} F_{I'}G^2||^2_{L^2}$ where $F_{I'} = \sum_{\omega_1 \in \phi(\Sigma_{1,I'})} f_{\omega_1}$ and $G = \sum_{\omega_2} g_{\omega_2}$. But

$$||F_{I'}G^2||_{L^2}^2 \le M_{2,4}(\delta,\nu,\rho_1',\rho_2)\sum_{\omega_1}||f_{\omega_1}||_{L^6}^2.$$

So we must prove

$$||\sum_{I'} F_{I'} G^2||_{L^2}^2 \lesssim \nu^{-O(1)} \sum_{I'} ||F_{I'} G^2||_{L^2}^2.$$

After a Galilean transform we may assume J is centered at 0 and thus I is at least $\geq \nu$ from the origin. In this setting G is Fourier supported on a small rectangle of height $O(\rho_2^2 + \delta)$, so it is almost constant in the vertical direction. A commutation I missed means this implies the lemma.

Lemma 20.10. For any $0 \le \delta \le \rho_1 \le \rho_2 \le \nu \le 1$,

$$M_{2,4}(\delta,\nu,\rho_1,\rho_2) \lesssim \sqrt{M_{2,4}(\delta,\nu,\rho_2,\rho_1)D(\delta/\rho_2)}.$$

This follows by Hoelder's inequality. In fact, if f is the sum of the f_{ω} , $\omega \in \phi(I_1)$ where $|I_1| = \rho_1$ and similarly for g and ρ_2 , this lemma is equivalent to the inequality

$$\int_{\mathbb{R}^2} |f|^2 |g|^4 \le \int_{\mathbb{R}^2} |f|^4 |g|^2 \int_{\mathbb{R}^2} |g|^6.$$

This is true by the L^6 version of Hoelder, if we normalize $||f||_{L^6} = ||g||_{L^6} = 1$ and then rescale appropriately. This allows us to induct on the scales ρ_j by ping-ponging back and forth between ρ_j .

Let α be the least exponent such that for every $\delta > 0$, $D(\delta) \leq \delta^{-\alpha - o(1)}$. Then

$$\alpha = \limsup_{\delta \to 0} -\frac{\log D(\delta)}{\log \delta}.$$

Lemma 20.11. One has $\alpha = 0$.

Proof. Suppose not; we will find an α' such that $D(\delta) \leq \delta^{-\alpha'-o(1)}$, a contradiction. Let $\nu = \delta^{\varepsilon}$ for some $\varepsilon > 0$ which is independent of δ and to be chosen later. Then

$$D(\delta) \lesssim \delta^{-O(\varepsilon)} M_{2,4}(\delta,\nu,\nu,\nu) + \delta^{(-1-\varepsilon)\alpha - o(1)}.$$

Since $\alpha > 0$, the second term is harmless. We then need to control the first term.

By an above lemma,

$$M_{2,4}(\delta,\nu,\nu,\nu) \lesssim \delta^{-O(\varepsilon)} M_{2,4}(\delta,\nu,\nu^2,\nu).$$

If $a \leq O(\varepsilon^{-1})$, we have

$$M_{2,4}(\delta,\nu,\nu^{2a},\nu^{a}) \le M_{2,4}(\delta,\nu,\nu^{a},\nu^{2a})^{1/2} D(\delta/\nu^{a})^{1/2}$$

and

$$D(\delta/\nu^a)^{1/2} \lesssim (\delta/\nu^a)^{-(\delta+o(1))/2} = \delta^{-\alpha/2 + a\alpha\varepsilon/2 - o(1)}$$

Thus

$$M_{2,4}(\delta,\nu,\nu^{2a},\nu^{a}) \lesssim M_{2,4}(\delta,\nu,\nu^{2a},\nu^{a})^{1/2} \delta^{-\alpha/2-o(1)} \delta^{a\alpha\varepsilon/2}$$

and by an above lemma,

$$M_{2,4}(\delta,\nu,\nu^{2a},\nu^{a}) \lesssim \nu^{-O(1)} M_{2,4}(\delta,\nu,\nu^{4a},\nu^{2a})^{1/2} \delta^{-\alpha/2-o(1)} \delta^{a\alpha\varepsilon/2},$$

or

$$(\delta^{\alpha} M_{2,4}(\delta,\nu,\nu^{2a},\nu^{a}))^{1/a} \lesssim (\delta^{\alpha} M_{2,4}(\delta,\nu,\nu^{4a},\nu^{2a}))^{1/2a} \delta^{\alpha\varepsilon/2 - O(\varepsilon/a) - o(1)}.$$

Here if a is large then we gain a small power of δ^{ε} . If $k \leq -\log_2(4\varepsilon)$ then

$$\delta^{\alpha} M_{2,4}(\delta,\nu,\nu^2,\nu) \lesssim_k (\delta^{\alpha} M_{2,4}(\delta,\nu,\nu^{2^{k+1}},\nu^{2^k}))^{2^{-k}} \delta^{k\alpha\varepsilon} - O(\varepsilon) - o_k(1).$$

Here $o_k(1)$ is a constant that $\to 0$ as $\delta \to 0$ as k, ε are held constant. Thus

$$\delta^{\alpha} M_{2,4}(\delta,\nu,\nu^2,\nu) \lesssim \delta^{k\alpha\varepsilon/2 - O(\varepsilon) - o(1)}$$

or

$$D(\delta) \lesssim \delta^{-\alpha + k\varepsilon\alpha/2 - O(\varepsilon) + o(1)} + \delta^{(\varepsilon - 1)\alpha + o(1)}.$$

This gives the desired contradiction.

This completes the proof of the main theorem.

Chapter 21

Pseudodifferential calculus

Let S denote the Schwartz space of \mathbb{R} . Consider the position and momentum operators $X, D: S \to S$. We normalize them by Xf(x) = xf(x) and $Df(x) = f'(x)/(2\pi i)$. Previously we have taken physical space and frequency space as separate but isomorphic, with the isomorphism $\mathbb{R} \to \mathbb{R}$ sending X to D. But now we will be interested in considering the "phase space", where we work with position and momentum together. This should be viewed as analoguous to how in music theory we indicate notes by their place in time, and their frequencies.

However, there is no perfectly rigorous phase space, by the uncertainty principle: we cannot describe a point in phase space arbitrarily well because then it would be a point in physical space, hence is delocalized to frequency space.

21.1 The Kohn-Nirenberg quantization

We are interested in developing a functional calculus for the unbounded operators X, D. This is easy for polynomials; for example, the quantum harmonic oscillator $D^2 + X^2$ is a polynomial in X, D. Intuitively this should be associated to the polynomial

$$a(x,\xi) = \xi^2 + x^2,$$

and then we write a(X, D) for the quantum harmonic oscillator.

In general, given some function a of phase space, we let

$$a(X,D)f(x) = \int_{-\infty}^{\infty} a(x,\xi)\hat{f}(\xi)e^{2\pi i\xi} d\xi$$

be the induced Fourier multiplier. The map $a \mapsto a(X, D)$ is called the *Kohn-Nirenberg* quantization of the symbol a. This is at first defined for $f \in S$ and

$$\partial_x^j \partial_{\xi}^{\ell} a(x,\xi) \lesssim_{a,j} \langle x \rangle^{O_{a,j,\ell}(1)} \langle \xi \rangle^{O_{a,j,\ell}(1)}.$$

More generally, a functional calculus for functions defined on phase space is known as a *quantization* of phase space. Most interesting observables from quantum mechanics are in fact given by the Kohn-Nirenberg quantization.

We denote the seminorms of \mathcal{S} by

$$||u||_{\alpha,\beta} = \sup_{x} |x^{\alpha}\partial^{\beta}u(x)|$$

Lemma 21.1. If a satisfies the above estimates then a(X, D) is a bounded operator on \mathcal{S} .

Proof. Let $f \in S$ and α, β be given; we must show that $||a(X, D)f||_{\alpha,\beta}$ is bounded by finitely many Schwartz seminorms of f. By the Leibniz rule,

$$\begin{aligned} ||a(X,D)f||_{\alpha,\beta} &= \sup_{x} \left| x^{\alpha} \partial^{\beta} \int_{-\infty}^{\infty} a(x,\xi) \hat{f}(\xi) e^{2\pi i x\xi} d\xi \right| \\ &\lesssim_{\beta} \sup_{x} \sum_{\gamma \leq \beta} \left| x^{\alpha} \int_{-\infty}^{\infty} (\partial_{x}^{\gamma} a)(x,\xi) \hat{f}(\xi) \xi^{\beta-\gamma} e^{2\pi i x\xi} d\xi \right|. \end{aligned}$$

We treat each of these integrals separately. Since everything is Schwartz we can integrate by parts to get, for any μ ,

$$\int_{-\infty}^{\infty} (\partial_x^{\gamma} a)(x,\xi) \hat{f}(\xi) \xi^{\beta-\gamma} e^{2\pi i x\xi} d\xi = \frac{(-1)^{\mu}}{(2\pi i x)^{\mu}} \int_{-\infty}^{\infty} \partial_{\xi}^{\mu} ((\partial_x^{\gamma} a)(x,\xi) \hat{f}(\xi) \xi^{\beta-\gamma}) e^{2\pi i x\xi} d\xi.$$

Since $\partial_{\xi}^{\nu}\partial_{x}^{\gamma}a$ is of polynomial growth in ξ for any $\nu \leq \mu$ and \hat{f} is Schwartz (in fact $f \mapsto \hat{f}$ is bounded on \mathcal{S}) we have

$$\left| \int_{-\infty}^{\infty} \partial_{\xi}^{\mu} ((\partial_{x}^{\gamma} a)(x,\xi) \hat{f}(\xi) \xi^{\beta-\gamma}) e^{2\pi i x \xi} d\xi \right| \lesssim_{a,\beta,\mu} \sum_{\delta,\delta' \le C_{\beta,\mu}} ||f||_{\delta,\delta'}$$

Therefore

$$\left| \int_{-\infty}^{\infty} (\partial_x^{\gamma} a)(x,\xi) \hat{f}(\xi) \xi^{\beta-\gamma} e^{2\pi i x\xi} d\xi \right| \lesssim_{a,\beta,\mu} \langle x \rangle^{-\mu} \sum_{\delta,\delta' \le C_{\beta,\mu}} ||f||_{\delta,\delta'}$$

Taking $\mu = \alpha$,

$$|a(X,D)f||_{\alpha,\beta} \lesssim_{a,\alpha,\beta} \sum_{\delta,\delta' \le C_{\beta,\alpha}} ||f||_{\delta,\delta'}.$$

Since $C_{\beta,\alpha} < \infty$ this is as desired.

Lemma 21.2. If a, b are symbols such that a(X, D) = b(X, D) then a = b.

Proof. If a(X, D) = b(X, D) then for every $f \in S$,

$$\int_{-\infty}^{\infty} (a(x,\xi) - b(x,\xi)) \hat{f}(\xi) e^{2\pi i x\xi} d\xi = 0.$$

In particular this is true when

$$f(\xi) = e^{2\pi i\xi y} \chi(\xi) g(y,\xi)$$

where χ is a smooth approximation to the indicator function of [-N, N], $y \in \mathbb{R}$ is to be determined, and $g(y, \xi)$ is a smooth approximation to the function which is 1 when
$a(x,\xi)/b(x,\xi) > 0$ and -1 otherwise. Plugging in y = x and taking the limits as the approximations become rough, we have

$$\lim_{N \to \infty} \int_{-N}^{N} |a(x,\xi) - b(x,\xi)| \, d\xi = 0$$

But this implies that a = b.

A fundamental defect with any quantization is that it is not an algebra morphism:

$$ab(X, D) \neq a(X, D)b(X, D).$$

This is because $[x, \xi] = 0$ but $2\pi i[X, D] = -1$. This motivates semiclassical analysis, because in the limit where frequency is very large, $[X, D] \to 0$, and this is formalized by introducing a semiclassical parameter h and taking $h \to 0$. As a useful consequence of $2\pi i[X, D] = -1$ we prove the Heisenberg uncertainty principle:

Theorem 21.3 (Heisenberg uncertainty principle). If $x, \xi \in \mathbb{R}$ and $f \in S$, then

$$||(X-x)f||_{L^2}||(D-\xi)f||_{L^2} \ge \frac{||f||_{L^2}^2}{4\pi}.$$

Proof. By the Cauchy-Schwarz inequality we have

$$|\langle [X-x]f, [D-\xi]f \rangle| \le ||(X-x)f||_{L^2} ||(D-\xi)f||_{L^2}.$$

Since x, ξ commute with everything, $[X - x, D - \xi] = [X, D] = -1/2\pi i$. So

$$|\langle [X - x, D - \xi] f, f \rangle| = \frac{||f||_{L^2}^2}{2\pi}.$$

But also

$$2|\langle [X-x]f, [D-\xi]f \rangle| = |\langle [X-x, D-\xi]f, f \rangle| = \frac{||f||_{L^2}^2}{2\pi}$$

since X is self-adjoint while D is anti-self-adjoint.

Physically, $||(X - x)f||_{L^2}/||f||_{L^2}$ should be viewed as the variance Δx of the observable X when applied to the wavefunction f. Similarly for $\Delta \xi$. So the Heisenberg uncertainty principle says

$$(\Delta x)(\Delta \xi) \ge \frac{1}{4\pi}.$$

The intuitive reason for the uncertainty principle is that f is a function of one variable, yet we want to view it as a function of phase space (two variables), yet there is no smooth bijection $\mathbb{R} \to \mathbb{R}^2$.

Fortunately, $ab(X, D) \approx a(X, D)b(X, D)$ where \approx means "equal up to lower order terms" here. For example [X, D] is a zeroth-order operator. We need certain estimates on the growth of the symbol to make "lower order terms" a rigorous definition, and then prove the above approximate commutativity relation.

There are other interesting quantizations, but they essentially behave equivalently to the Kohn-Nirenberg quantization, so we restrict to this quantization in what follows.

Definition 21.4. Let $\alpha \in \mathbb{R}$. A symbol *a* has *order* α if it is smooth and

$$\partial_x^j \partial_\xi^\ell a(x,\xi) \lesssim_{a,j,\ell} \langle \xi \rangle^{\alpha-\ell}.$$

We let S^{α} denote the space of all symbols of order α . In this case we say a(X, D) is known as an *pseudodifferential operator* of order α .

Thus if $a \in S^{\alpha}$, then $a(x,\xi)$ is bounded in x and grows like ξ^{α} , even after it is differentiated. If one tiles phase space by rectangles which have x-width 1 and ξ -width $\sim \xi$, then a restricted to any rectangle is approximately constant. Thus at high frequencies $\xi \to \infty$, the effect of the uncertainty principle is unimportant, since there a is so close to constant that we can localize it quite well.

We let

$$||a||_{\alpha,j,\ell} = \sup_{x,\xi} \frac{\partial_x^j \partial_\xi^\ell a(x,\xi)}{\langle \xi \rangle^{\alpha-\ell}}$$

denote the (j, ℓ) th Hormander seminorm for symbols of order α . For example, fractional differential operators are pseudodifferential operators.

A smooth linear differential operator is a pseudodifferential operator provided that its coefficients are bounded.

Example 21.5. The quantum harmonic oscillator $D^2 + X^2$ is not a pseudodifferential operator but it is locally a pseudodifferential operator of order 0.

21.2 The Calderon-Vallaincourt theorem

We start by proving a generalization of the Hormander-Mikhlin multiplier theorem which shows that pseudodifferential operators of null order are bounded on L^p . The assumption on order is best-possible here; clearly X, D are not bounded on L^p for any p, and this remains for fractional operators such as D^{ε} .

Theorem 21.6 (Calderon-Vallaincourt). Every pseudodifferential operator of order 0 extends to a bounded operator on L^2 .

Actually, we can use Calderon-Zygmund theory to extend to this to L^p : once a Calderon-Zygmund operator is bounded on L^q for some q, it is bounded on L^p for every p. But we omit the proof.

We will give two proofs of the Calderon-Vallaincourt theorem, both of which use sequences of operators which are "almost orthogonal." One way to quantify the extent to which operators are almost orthogonal is if they meet the hypotheses of the Cotlar-Stein lemma:

Lemma 21.7 (Cotlar-Stein). Suppose that $T_1, \ldots, T_n : H \to H'$ are linear operators such that the maps $T_i^*T_j : H \to H$ satisfy the estimate

$$\sum_j ||T_i^*T_j||^{1/2} \leq A$$

uniformly in i and similarly the $T_iT_j^*:H'\to H'$ satisfy the estimate

$$\sum_{j} ||T_i T_j^*||^{1/2} \le B$$

uniformly in j. Then

$$||\sum_{j} T_{j}|| \le \sqrt{AB}.$$

Proof. Since B(H) is a C^* -algebra we have

$$|T_i|| = ||T_i T_i^*||^{1/2} \le A.$$

Similarly $||T_i|| \leq B$ and hence $||T_i|| \leq \sqrt{AB}$, the geometric mean. We want to iterate this estimate.

Let $T = \sum_{j} T_{j}$. Iterating and using the fact that B(H) is a C^{*}-algebra again, we see that

$$||T|| \le ||(T^*T)^m||^{1/(2m)}$$

This is in particular true when m is a power of 2. We have the bound

$$||T|| \leq \sqrt[2m]{\sum_{j} ||T_{i_1}^* T_{i_2} \cdots T_{i_{2m-1}}^* T_{i_{2m}}||}.$$

Here the i_j are a substring of some permutation of n. We split this up as

$$||T|| \leq \sqrt[2m]{\sum_{j} ||T_{i_1}^* T_{i_2}|| \cdots ||T_{i_{2m-1}}^* T_{i_{2m}}||}.$$

This gives a bound $||T|| \leq n^{1/2m} A$. Similarly we can bound $||T|| \leq n^{1/2m} B$. Taking geometric means we have

$$||T|| \le n^{1/2m} \sqrt{AB},$$

and iterating we see $||T|| \leq \sqrt{AB}$.

If the T_i have pairwise orthogonal images then $T_i^*T_j = 0$ whenever $i \neq j$; similarly for the adjoints. So if A, B are minimal possible (say $A = \sum_j ||T_i||^2$) then the operators T_i are actually orthogonal, and the size of A measures "how orthogonal" they are.

The Cotlar-Stein lemma also has a useful corollary, which we will not need for the Calderon-Vallaincourt theorem.

Corollary 21.8. Suppose that we have a sequence of operators $T_i: H \to H'$ satisfying the bounds

$$\sum_{j \le J} ||T_i^*T_j||^{1/2} \le A$$
$$\sum_{j \le J} ||T_iT_j^*||^{1/2} \le B$$

uniformly in i, J and similarly

$$\sum_{i \le I} ||T_i T_j^*||^{1/2} \le B$$

uniformly in I, j. Then the infinite series $T = \sum_{j} T_{j}$ converges in the weak operator topology, and

$$||T|| \le \sqrt{AB}.$$

Proof. The estimate on the infinite series follows from the fact that the estimate is uniformly true on partial sums by the Cotlar-Stein lemma. Now we must show that if $f \in H, g \in H'$, then

$$\sum_{i} \langle g, T_i f \rangle$$

is a convergent series. Suppose not, so there are f, g such that for every C > 0 there is an $I \in \mathbb{N}$ such that

$$\left|\sum_{i\leq I} \langle g, T_i f \rangle\right| > C.$$

This is in particular true when $C = ||f||||g||\sqrt{AB}$, which is a contradiction by the Cotlar-Stein lemma.

We can actually prove convergence in the strong operator topology, but omit the proof.

Now recall that a symbol of order 0 is a smooth function $a : \mathbb{R}^2 \to \mathbb{R}$ such that $\partial_x a(x,\xi)$ is bounded and $\partial_{\xi}^{\ell} a(x,\xi) = O(\xi^{-\ell})$. Let a be a symbol of order 0.

We want to run a Littlewood-Paley decomposition on a, because these hypotheses imply that a is "approximately constant" on each of the dyadic pieces of \mathbb{R} . Let ϕ be a bump function on [-1, 1] which is 1 on [-1/2, 1/2] and consider the Littlewood-Paley decomposition

$$a = \sum_{k} a_{k}$$

where $a_0 = a(x)\phi$ and

$$a_k(x,\xi) = (\phi(\xi/2^k) - \phi(\xi/2^{k+1}))a(x,\xi).$$

To avoid technicalities, we don't want to have to worry about the infinite sum; by Fatou's lemma, we must then show

$$||\sum_{k\leq K} a_k(X,D)||_{L^2} \lesssim_a ||f||_{L^2}$$

uniformly in K. Each a_k lives on a strip of (x,ξ) where $\xi \sim 2^k$. Then

$$\partial_x^j \partial_\xi^\ell a_k(x,\xi) \lesssim_{a,j,\ell} 2^{-k\ell}.$$

Therefore

$$K_k(x,y) = \int_{-\infty}^{\infty} a_k(x,\xi) e^{2\pi i (x-y)\xi} d\xi$$

has compact support and hence is the integral kernel of $a_k(X, D)$ by Fubini's theorem.

We bound K_k . First, a_k is bounded and supported on the interval $[2^k, 2^{k+1}]$ of length 2^k so by the triangle inequality,

$$|K_k(x,y)| \le ||a_k(x)||_{L^1} \le_a 2^k$$

Integrating by parts in ξ ,

$$K_k(x,y) = \frac{-1}{2\pi i(x-y)} \int_{-\infty}^{\infty} \partial_{\xi} a_k(x,\xi) e^{2\pi i(x-y)\xi} d\xi \lesssim_a \frac{2^k}{2^k |x-y|}.$$

Iterating and using the previous bound to prevent blowup,

$$|K_k(x,y)| \lesssim_{a,\ell} \frac{2^k}{\langle 2^{-\ell k} | x - y |^\ell}$$

We now bound $\partial_x K_k(x, y)$. If the derivatives fall on a_k nothing happens. If they fall on the phase, we gain a term $2\pi i\xi$, but $\xi \sim 2^k$, so we lose a factor of 2^k . Running the above argument again,

$$\left|\partial_x K_k(x,y)\right| \lesssim_{a,\ell} \frac{2^{2k}}{\langle 2^{-\ell k} | x - y |^{\ell}}$$

The same thing happens for ∂_y . It follows from Young's inequality that

 $||a_k(X,D)||_{L^2 \to L^2} \lesssim_a 1.$

It remains to remove the dependence on the number of terms in the Fatou sum K. To prove this uniform bound we will show that the $a_k(X, D)$ have "almost orthogonal" ranges and coranges (i.e. row spaces), in two different ways.

First proof. We claim

$$||a_k(X,D)a_j(X,D)^*f||_{L^2} \lesssim_a 2^{-|j-k|} ||f||_{L^2}$$

and similarly for $a_k(X, D)^* a_j(X, D)^* f$. The Cotlar-Stein lemma then gives the theorem. Assume $j \leq k$ and work with $a_k(X, D)a_j(X, D)^*$; the other three cases are similar.

By Fubini's theorem,

$$K_{kj^*}(x,z) = \int_{-\infty}^{\infty} \overline{K_k}(y,x) K_j(y,z) \, dy$$

is the integral kernel of $a_k(X, D)a_j(X, D)^*$. We want to bound it using integration by parts. First,

$$K_{j}(y,z) = \partial_{y} \int_{-\infty}^{\infty} \frac{a_{j}(y,\xi)}{2\pi i\xi} e^{2\pi i (y-z)\xi} d\xi - \int_{-\infty}^{\infty} \frac{\partial_{y} a_{j}(y,\xi)}{2\pi i\xi} e^{2\pi i (y-z)\xi} d\xi.$$

The second term here is harmless since a is a symbol; we have

$$\int_{-\infty}^{\infty} \frac{\partial_y a_j(y,\xi)}{2\pi i\xi} e^{2\pi i(y-z)\xi} d\xi \sim_a 2^{-j} \int_{-\infty}^{\infty} a_j(y,\xi) e^{2\pi i(y-z)\xi} d\xi$$

since we have a ξ in the denominator and are integrating over $[2^j, 2^{j+1}]$. Thus

$$\int_{-\infty}^{\infty} \overline{K_k}(y,x) \int_{-\infty}^{\infty} \frac{\partial_y a_j(y,\xi)}{2\pi i\xi} e^{2\pi i (y-z)\xi} d\xi \sim 2^{-j+k}.$$

This was desired. Similarly,

$$\int_{-\infty}^{\infty} \overline{K_k}(y,x) \partial_y \int_{-\infty}^{\infty} \frac{a_j(y,\xi)}{2\pi i\xi} e^{2\pi i (y-z)\xi} d\xi dy = -\int_{-\infty}^{\infty} \partial_y \overline{K_k}(y,x) \int_{-\infty}^{\infty} \frac{a_j(y,z)}{2\pi i\xi} e^{2\pi i (y-z)\xi} d\xi dy \sim 2^{-j+k}$$

Here we used the fact that we took a complex conjugate to get ξ in the numerator when we differentiated k.

Second proof. Consider the integral kernel

$$L_j(z,x) = \int_{2^j}^{2^{j+1}} e^{2\pi i(z-x)\xi} d\xi$$

of the Littlewood-Paley projector P_j to the scale $[2^j, 2^{j+1}]$. Let F_{jk} be the integral kernel of the composite

$$F_{jk}(z,y) = \int_{-\infty}^{\infty} L_j(z,x) K_k(x,y) \, dx$$
$$= \int_{-\infty}^{\infty} \int_{2^k}^{2^{k+1}} \int_{2^j}^{2^{j+1}} a_k(x,\eta) e^{2\pi i (x-y)\eta} e^{2\pi i (z-x)\xi} \, d\xi \, d\eta \, dx.$$

We used the Fourier support property of a_k .

Lemma 21.9. One has

$$||F_{jk}||_{L^2} \lesssim_a 2^{-|j-k|}$$

Proof. Note that

$$||F_{jk}||_{L^2}^2 = \iint_{\mathbb{R}^2} \left| \int_{-\infty}^{\infty} \int_{2^k}^{2^{k+1}} \int_{2^j}^{2^{j+1}} a_k(x,\eta) e^{2\pi i (x-y)\eta} e^{2\pi i (z-x)\xi} d\xi d\eta dx \right|^2 dy dz.$$

We now integrate by parts in dx to see

$$\int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} a_{k}(x,\eta) e^{2\pi i (x-y)\eta} e^{2\pi i (x-y)\xi} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\xi}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\xi}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\xi}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\xi}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{j+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{j}}^{2^{k+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{k}}^{2^{k+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{k}}^{2^{k+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{\infty} \int_{2^{k}}^{2^{k+1}} \int_{2^{k}}^{2^{k+1}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{2^{k}} \int_{2^{k}}^{2^{k}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{2^{k}} \int_{2^{k}}^{2^{k}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{2^{k}} \int_{2^{k}}^{2^{k}} \partial_{x} (a_{k}(x,\eta) e^{2\pi i (x-y)\eta}) \frac{e^{2\pi i (x-y)\eta}}{2\pi i \eta} \, d\xi \, d\eta \, dx = \int_{-\infty}^{2^{k}} \int_{2^{k}}^{2^{k}} \partial_{x$$

We have $\partial_x a_k(x,\eta) \lesssim_a 1$ and $a_k(x,\eta) \lesssim_a 1$. Similarly we have $\eta \sim 2^k$ and $\xi \sim 2^j$, so the triple integral is

$$\lesssim_a \frac{1+2^j}{2^k} \int_{-\infty}^{\infty} \int_{2^k}^{2^{k+1}} \int_{2^j}^{2^{j+1}} e^{2\pi i (z-x)\xi} e^{2\pi i (x-y)\eta} d\xi d\eta dx.$$

This new triple integral is

$$\lesssim_a \int_{-\infty}^{\infty} \frac{e^{2^{j+k}\pi ix}}{\langle (x-y)(x-z)\rangle} \, dx$$

whose L^2 norm is $\leq_a 1$. Therefore if $k \geq j$ the L^2 norm is $\leq_a 2^{-|k-j|}$. The proof is similar when $j \geq k$, but we want to differentiate $e^{2\pi i (x-y)\eta}$ instead.

We conclude that

$$||P_j a_k(X, D)|| \lesssim_a 2^{-|k-j|}.$$

Similarly for $a_k(X, D)P_j$, where the roles of η and ξ in the above proof are swapped. TODO

21.3 Composition estimates

Lemma 21.10. Let a(X, D) and b(X, D) be pseudodifferential operators of order α, β respectively. Then the composite a(X, D)b(X, D) is a pseudodifferential operator of order $\alpha + \beta$.

Definition 21.11. The symbol of the composite a(X, D)b(X, D) is denote a * b.

Proof of lemma. We let

$$a_R(x,\xi) = a(x,\xi)\phi(x/R)\phi(\xi/R)$$

where ϕ is a standard bump function; then the a_R obeys the same symbol estimates as a uniformly in R, and $a_R \to a$ locally uniformly as $R \to \infty$. If f is Schwartz, then b(X, D)f is as well, so $a_R(X, D)b(X, D)f$ converges pointwise to a(X, D)b(X, D)f.

Suppose that $a_R(X, D)b(X, D) = c_R(X, D)$ is a pseudodifferential operator with order uniform in R. Then the Ascoli theorem guarantees that there is a limit c(X, D) of a subsequence $c_R(X, D)$ as $R \to \infty$, and then c(X, D) is a limit of not just a subsequence but the full sequence. Here we are taking limits in the weak operator topology, which is all we need because we just require that limits be unique, and the weak operator topology is Hausdorff. By a similar argument, we can replace b(X, D) with $b_R(X, D)$. So we may assume that a, bare compactly supported, and hence justify any use of Fubini's theorem.

We have

$$\widehat{b(X,D)}f(\eta) = \iint_{\mathbb{R}^2} b(y,\xi)\widehat{f}(\xi)e^{2\pi i(y\xi-y\eta)} d\xi dy$$

hence

$$a(X,D)b(X,D)f(x) = \iiint_{\mathbb{R}^3} a(x,\eta)b(y,\xi)\hat{f}(\xi)e^{2\pi i(y\xi-y\eta+x\eta)} d\xi dy d\eta$$

By Fubini's theorem,

$$a * b(x,\xi) = \iint_{\mathbb{R}^2} a(x,\eta) b(y,\xi) e^{2\pi i (y-x)(\xi-\eta)} dy d\eta$$

We must now show the symbol estimate

$$\partial_x^j \partial_{\xi}^{\ell}(a * b)(x, \xi) \lesssim_{a, b, \alpha, \beta, j, \ell} \langle \xi \rangle^{\alpha + \beta - \ell}$$

We readily check

$$\partial_x(a*b) = \partial_x * b + a\partial_x b,$$

and similarly in ξ . Moreover ∂_x sends S^{α} to itself and ∂_{ξ} sends S^{β} to $S^{\beta-1}$. So by induction in j and another in ℓ , we see that we just need to check the case $j = \ell = 0$, i.e.

$$(a * b)(x, \xi) \lesssim_{a,b,\alpha,\beta,j,\ell} \langle \xi \rangle^{\alpha+\beta}.$$

We bound a * b using the method of stationary phase.

Applying a smooth partition of unity to a in the frequency variable, we split a into two functions, one supported on the set of (x, η) such that $\langle \xi \rangle \sim \langle \eta \rangle$, and its complement. This cutoff preserves the symbol estimates for a.

We now treat the case $\langle \xi \rangle \sim \langle \eta \rangle$. Let

$$K_x(z) = \int_{-\infty}^{\infty} a(x,\eta) e^{2\pi i x \eta} \, d\eta,$$

and then we have

$$(a * b)(x, \xi) = \int_{-\infty}^{\infty} K_x(x - y)b(y, \xi)e^{2\pi i(y - x)\xi} dy.$$

Integrating by parts and using our case assumption,

$$K_x(z) \lesssim_{a,\alpha} \langle \xi \rangle^{\alpha+1} \langle z\xi \rangle^{-2}.$$

Since we have the estimate $b(y,\xi) \leq_b \langle \xi \rangle^{\beta}$, the claim follows.

In the other case,

$$(a * b)(x,\xi) = \iint_{\mathbb{R}^2} \frac{a(x,\eta)}{(2\pi i(\eta-\xi))^m} \partial_y^m b(y,\xi) e^{2\pi i(y-x)(\xi-\eta)} \, dy \, d\eta.$$

This follows by iterated integration by parts. Let

$$K_{x,m}(z) = \int_{-\infty}^{\infty} \frac{a(x,\eta)}{(2\pi i(\eta-\xi))^m} e^{2\pi i z\eta} d\eta,$$

 \mathbf{SO}

$$(a*b)(x,\xi) = \int_{-\infty}^{\infty} K_{x,m}(x-y)\partial_y^m b(y,\xi)e^{2\pi i(y-x)\xi} dy,$$

and if m is large enough we have

$$K_{x,m}(z) \lesssim_{a,\alpha} \langle \langle \xi \rangle^{\alpha+1} \langle z\xi \rangle^{-2}$$

as before.

Example 21.12. If $a(x,\xi) = a(x)$ and similarly for b,

$$(a * b)(x, \xi) = a(x) \iint_{\mathbb{R}^2} b(y) e^{2\pi i (y-x)(\xi-\eta)} \, dy \, d\eta = a(x)b(x)$$

by the Fourier inversion formula, as we would hope.

If $a(x,\xi) = a(\xi)$ and similarly for b, we similarly have a * b = ab.

Theorem 21.13. Let $a \in S^{\alpha}, b \in S^{\beta}$. Then $a * b - ab \in S^{\alpha+\beta-1}$. Moreover, $a * b - ab - \partial_{\xi} a \partial_x b \in S^{\alpha+\beta-2}$.

TODO Prove me.

Higher-order versions of this theorem are also true.

Corollary 21.14 (quantum correspondence). Let $\{\cdot, \cdot\}$ denote the Poisson bracket and $a, b \in S^{\alpha}, S^{\beta}$ respectively. Then

$$[a(X, D), b(X, D)] - \{a, b\}(X, D) \in S^{\alpha + \beta - 1}.$$

	-	-	
			L
			L
			н

TODO Prove me.

This suggests that if we view a symbol a as a classical observable, then a(X, D) is its quantization. In semiclassical analysis we are interested in the extremely high frequency regime, where lower-order symbols are negligible. The correspondence principle says that in the extremely high frequency regime, classical and quantum mechanics behave identically.

Corollary 21.15. Let $a \in S^{\alpha}$. If a is compactly supported then the function \tilde{a} ,

$$\tilde{a}(x,\xi) = \iint_{\mathbb{R}^2} \overline{a}(y,\eta) e^{2\pi i (x-y)(\eta-\xi)} \, dy \, d\eta,$$

is a symbol of order α and $\tilde{a}(X, D)$ is the adjoint of a(X, D).

In particular, even if a is not compactly supported, there is a unique pseudodifferential operator $a(X, D)^*$ of order α which is the adjoint of a(X, D). Moreover, $a(X, D)^* - \overline{a}(X, D)$ is a pseudodifferential operator of order $\alpha - 1$.

21.4 The Gabor transform

Part VI General relativity

Chapter 22

Lorentzian geometry

Lorentzian geometry is a generalization of Riemannian geometry that was motivated by physical axioms.

22.1 Axioms of special relativity

Definition 22.1. A reference frame is a coordinate system for $\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3$. A reference frame is said to be *inertial* if the motion of a body without external influence forms a straight line in \mathbb{R}^4 . Otherwise, the reference frame is said to be *accelerated*.

Axiom 22.2. All laws of physics are invariant under change of inertial reference frame.

Axiom 22.3. The speed of light in a vacuum is invariant under change of inertial reference frame.

We denote the speed of light in a vacuum by c.

22.2 Lorentz transformations

Definition 22.4. A *Lorentz transformation* is a smooth transformation which fixes the origin and is homotopic to the identity, which carries an inertial reference frame to an inertial reference frame.

Definition 22.5. Let $p = (t, x), q = (t', x') \in \mathbb{R}^4$. The spacetime interval $\Delta s = [p, q]$ between p, q is the distance

$$\Delta s^2 = \Delta x^2 - c^2 \Delta t^2$$

where $\Delta t = t' - t$ and $\Delta x = x' - x$.

It is not hard to check that Lorentz transformations are linear (since they preserve straight-line trajectories). Moreover, spacetime intervals are also preserved by Lorentz transformations. By the second axiom of relativity, the quadratic polynomials associated to Δs and its Lorentz transform, say $\Delta s'$, have the same roots. So there is an $\alpha \neq 0$ such that

$$(\Delta s)^2 = \alpha (\Delta s')^2.$$

Moreover, this constant appears for any choice of s, s', so by "reciprocity", $\alpha^2 = 1$. Since Lorentz transformations are homotopic to the identity, which clearly has $\alpha > 0$, we have $\alpha = 1$. Therefore the claim holds.

Example 22.6 (twin paradox). Let A, B be two twins born in space. They are separated at birth (spacetime P), and B moves away from A but then suddenly turns around (at spacetime Q) and meets A again (at spacetime R). Then it appears that both A is older than B (from the point of view of A) and B is older than A (from the point of view of B). However, one can check that in fact A is older than B, since B had an accelerated reference frame (when he turned around at Q) and so has an incorrect perception of the universe. This can be checked using the spacetime interval invariance.

Let us consider the Lorentzian metric

$$ds^2 = dx^2 - c^2 dt^2,$$

where as usual we write $s = (t, x) \in \mathbb{R} \times \mathbb{R}^3 = \mathbb{R}^4$ for a point in spacetime. This is a linear combination of the Riemannian metrics dx^2 and dt^2 . We will also write m for the indefinite quadratic form induced by ds^2 on the tangent bundle. On the other hand we will write δ for the positive-definite quadratic form induced by the Riemannian metric dx^2 . Therefore (\mathbb{R}^3, δ) is just Euclidean space.

Definition 22.7. A Lorentzian manifold is a smooth manifold equipped with a smoothly varying quadratic form on each tangent space. The Lorentzian manifold (\mathbb{R}^4, m) is known as Minkowski spacetime.

Now let γ be a curve in \mathbb{R}^4 , which we think of as parametrized by [0, 1]. We denote the tangent vector by $\dot{\gamma}$.

Definition 22.8. The proper time of the curve γ is

$$\int_0^1 \frac{\sqrt{-m(\dot{\gamma}(\sigma), \dot{\gamma}(\sigma))}}{c} \, d\sigma.$$

Definition 22.9. Let v be a tangent vector over \mathbb{R}^4 . If m(v, v) < 0, we say that v is *timelike*. If m(v, v) = 0, then v is *lightlike* or *null*. Otherwise, v is *spacelike*. If v is not spacelike, we say that v is *causal*. If every tangent vector to a curve is timelike (lightlike, etc.), we say that the curve itself is timelike (lightlike, etc.)

Notice that a vector v has speed $\leq c$ iff v is causal. So causal curves are those trajectories of objects which are allowed by the laws of physics.

22.3 Riemannian geometry

We dive deeper into Lorentzian geometry. Throughout this chapter we fix a Lorentzian spacetime (M, g). In other words, (M, g) is locally isomorphic as a Lorentzian manifold to the Minkowski spacetime (\mathbb{R}^4, m) . We always assume that M is orientable.

Throughout these notes we use Einstein's conventions that a repeated index is summed over:

$$\omega_{\mu}v^{\mu} = \langle \omega, v \rangle.$$

We let $\{\partial_0, \ldots, \partial_3\}$ be the standard basis of the tangent bundle TM, and $\{dx_0, \ldots, dx_3\}$ be the standard basis of the cotangent bundle T^*M . Thus we have a pairing

$$dx^{\mu}\partial_{\nu} = \delta^{\mu}_{\nu}.$$

Definition 22.10. A (p,q)-tensor at $x \in M$ is an element of $(T_x^*M)^{\otimes p} \otimes (T_xM)^{\otimes q}$. A (p,q)-tensor field is a section of the tensor bundle

$$(T^*M)^{\otimes p} \otimes (TM)^{\otimes q} \to M.$$

In local coordinates, we write

$$T^{\beta_1,\ldots,\beta_q}_{\alpha_1,\ldots,\alpha_p}(x)$$

for the $(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q)$ th coordinate of a (p, q)-tensor field evaluated at x.

We now need the notion of a linear connection. A linear connection, morally, is a "way to differentiate a vector field against another vector field." Let $\mathcal{T}(M)$ denote the space of vector fields $M \to TM$.

Definition 22.11. A linear connection is a $(C^{\infty}(M), \mathbb{R})$ -bilinear map $\nabla : \mathcal{T}(M)^2 \to \mathcal{T}(M)$, written $(X, Y) \mapsto \nabla_X Y$ (though we write $\nabla_{\alpha} = \nabla_{\partial_{\alpha}}$) satisfying the Leibniz rule

$$\nabla_X(fY) = (\nabla_X f)Y + f\nabla_X Y = df(X)Y + f\nabla_X Y.$$

Let's consider the easy example of a Euclidean connection.

Definition 22.12. Assume (M, g) is Euclidean space. The *Euclidean connection* on M is the linear connection

$$\nabla_X Y^j \partial_j = X Y^j \partial_j.$$

Since X is a first-order derivation at each point, XY^j is the partial derivative of Y in the direction of X. Thus Euclidean connections are a very natural thing to study, and in case $X = \partial_{\alpha}$, ∇_X is just the map that sends Y to its derivative in some direction. The Euclidean connection has the useful property that $\nabla g = 0$.

Definition 22.13. The Levi-Civita connection ∇ is the unique connection on M such that $\nabla g = 0$ and which satisfies $\nabla_X Y - \nabla_Y X = [X, Y]$.

Theorem 22.14. The Levi-Civita connection is well-defined.

Definition 22.15. The *Riemann curvature tensor* is the (3, 1)-tensor field

$$R^{\delta}_{\alpha\beta\gamma}\partial_{\delta} = \nabla_{\alpha}\nabla_{\beta}\partial_{\gamma} - \nabla_{\beta}\nabla_{\alpha}\partial_{\gamma}$$

It is pretty clear that

$$R^{\delta}_{\alpha\beta\gamma} = -R^{\delta}_{\beta\alpha\gamma},$$

and

$$R_{\alpha\beta\gamma\delta} = R_{\alpha\beta\delta\gamma}.$$

Theorem 22.16 (Bianchi). One has

$$R_{\alpha\beta\gamma\delta} + R_{\beta\gamma\alpha\delta} + R_{\gamma\beta\alpha\delta} = 0$$

and

$$\nabla_{\alpha} R_{\beta\gamma\delta\epsilon} + \nabla_{\beta} R_{\gamma\alpha\delta\epsilon} + \nabla_{\gamma} R_{\alpha\beta\delta\epsilon} = 0.$$

Definition 22.17. The *Christoffel symbol* Γ is defined by

$$\nabla_{\alpha}\partial_{\beta} = \Gamma^{\gamma}_{\alpha\beta}\partial_{\gamma}.$$

We have the Kozsul formula

$$\Gamma^{\gamma}_{\alpha\beta} = \frac{g^{\gamma\delta}}{2} (\partial_{\alpha}g_{\beta\delta} + \partial_{\beta}g_{\delta\alpha} - \partial_{\delta}g_{\alpha\beta}).$$

22.4 Causality

Let (M, g) be a Lorentzian spacetime as above. Recall that we have a causal structure on the tangent bundle of M, which gives rise to a pair of light cones in each tangent space. Taking the exponential map, we get a causal structure on curves in M.

Definition 22.18. The spacetime (M, g) is *time-orientable* if there is a continuous choice of light cone for each tangent space.

Let us fix a time-orientation. The vectors in the chosen lightcone point to the "future."

Definition 22.19. Let $S \subseteq M$. The chronological future $I^+(S)$ is the set of points in M that can be reached by a curve through the exponential map of the open future light cones of S. The chronological past is defined similarly, but with the open past light cone. The causal future and causal past are defined similarly, but for the closed light cones.

Chapter 23

The Cauchy problem in general relativity

We pose the Einstein equation as an initial-value problem on the Lorenztian manifold (M, g).

23.1 The Einstein equation

If \mathcal{L} is a Lagrangian density, then \mathcal{L} does not have to be integrable, so long as we only take only compactly supported perturbations when we carry out the calculus of variations. That is why we emphasize that \mathcal{L} is only a "density" rather than a summable quantity.

Throughout, we let $\delta g_{\alpha\beta}$ be a compatibly supported perturbation of the metric tensor $g_{\alpha\beta}$.

Definition 23.1. Let \mathcal{L} be a Lagrangian density. The *energy-momentum tensor* associated to \mathcal{L} is the tensor $T_{\alpha\beta}$ given by

$$\int \frac{d\mathcal{L}}{ds}(\cdot, g + s\delta g) \, dV(g + s\delta g) + \int T^{\alpha\beta} \delta g_{\alpha\beta} dV = 0.$$

Theorem 23.2 (Noether). If $T_{\alpha\beta}$ is an energy-momentum tensor, then the divergence

$$\nabla^{\alpha}T_{\alpha\beta} = 0.$$

Noether's theorem can be interpreted as a generalization of the conservation laws of energy, mass (which is just a form of energy by Einstein's special theory of relativity), and momentum. It is a special case of Noether's theorem that for any Lie action of \mathbb{R} on a Lagrangian density, there is an associated conserved quantity in the respective Euler-Lagrange equations.

A key point in general relativity is that "curvature is energy-momentum", yet the energymomentum tensor $T_{\alpha\beta}$ is divergence-free. So the curvature tensor appearing in general relativity should be a divergence-free covariant 2-tensor. Thus we must define a 2-tensor which measures curvature.

Definition 23.3. Let $R^{\alpha}_{\beta\gamma\delta}$ be the Riemann curvature tensor of (M, g). The *Ricci curvature tensor* of (M, g) is given by

$$\operatorname{Ric}_{\alpha\beta} = R^{\mu}_{\alpha\mu\beta}$$

The scalar curvature is given by $R = \operatorname{Ric}_{\alpha}^{\alpha}$.

Definition 23.4. The *Einstein tensor* of (M, g) is

$$G_{\alpha\beta} = \operatorname{Ric}_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R.$$

The *Einstein equation* is the equation

$$G_{\alpha\beta} = \frac{8\pi G}{c^4} T_{\alpha\beta}.$$

We will normalize c = 1, and then take $G = 1/(4\pi)$, so the Einstein equation will read as $G_{\alpha\beta} = 2T_{\alpha\beta}$. Notice the similarity to the Gauss-Poisson equation for gravity

$$4\pi G \nabla^{\alpha} g_{\alpha} = \rho$$

where ρ is the mass density of the universe and g_{α} is the gravitational field. We have

$$\nabla^{\alpha}G_{\alpha\beta} = 0$$

by the second Bianchi identity.

We interpret $T_{\alpha\beta} = 0$ as meaning that the universe is a vacuum. In this case we have $\operatorname{Ric}_{\alpha\beta} = 0$. Now $\operatorname{Ric}_{\alpha\beta} = 0$ is a geometric PDE, but we want to frame it as an initial-value problem where the initial data consists of a 3-manifold, and the resulting 4-manifold comes from gluing together the 3-manifolds together in time.

This interpretation gives another derivation of the Einstein equation. Assume $T_{\alpha\beta} = 0$; then the universe should have no curvature.

Definition 23.5. The Einstein-Hilbert Lagrangian density is $\mathcal{L}_{EH} = R \, dV$.

The Einstein equation should be the Euler-Lagrange equation minimizing the Einstein-Hilbert action.

Theorem 23.6. The Euler-Lagrange equation corresponding to the Einstein-Hilbert Lagrangian density is the Einstein equation.

Proof. Let δg be a compactly supported perturbation of hte metric tensor as above. Then

$$\frac{\delta}{\delta s}(g+s\delta g)^{\mu\nu} = -\delta g^{\mu\nu}.$$

Similarly

$$\frac{\delta}{\delta g} dV(g) = \delta g^{\alpha\beta} \ dV(g).$$

In coordinates, we have

$$\operatorname{Ric}_{\beta\nu} = \partial_{\alpha}\Gamma^{\alpha}_{\beta\nu} - \partial_{\beta}\Gamma^{\alpha}_{\alpha\nu} + \Gamma^{\mu}_{\alpha\gamma}\Gamma^{\gamma}_{\mu\nu} + \Gamma^{\alpha}_{\beta\gamma}\Gamma^{\gamma}_{\alpha\nu}$$

After a tedious computation in normal coordinates (where g = m and $\Gamma = 0$ at the origin) we have

$$\frac{\delta}{\delta s}\operatorname{Ric}_{\alpha\beta}(g+s\delta g)| = \frac{1}{2}g^{\mu\nu}(\partial_{\alpha}\delta g_{\beta\nu} + \partial_{\beta}\delta g_{\nu\alpha} - \partial_{\nu}\delta g_{\alpha\beta}).$$

Applying the Lebiniz rule we have

$$\int_{M} \frac{\delta}{\delta s} \mathcal{L}_{EH}(g + s\delta g) = \frac{1}{2} \int_{M} g^{\alpha\beta} R \delta g_{\alpha\beta} - \operatorname{Ric}^{\alpha\beta} \delta g_{\alpha\beta} \, dV$$

so the claim follows by lowering indices.

23.2 Initial-data sets

Henceforth we fix a time-orientation of (M, g).

Definition 23.7. A *time function* is a smooth function t on M such that for every futurepointing timelike vector field X^{α} , $g_{\alpha\beta}\nabla^{\alpha}tX^{\beta} > 0$.

We fix a time function as well.

Definition 23.8. The *initial-time slice* Σ_0 is the level set of the equation t = 0.

Because the metric signature is (-, +, +, +), the initial-time slice will be a 3-manifold. We denote its induced Riemannian metric by \overline{g} and induced Levi-Civita connection by $\overline{\nabla}$.

Definition 23.9. Let Σ be a Riemannian submanifold of codimension 1. Let *n* denote the future-pointing unit normal vector to Σ . Then the *second fundamental form* $\overline{k}_{\alpha\beta}$ is given by

$$\overline{k}_{\alpha\beta}u^{\alpha}v^{\beta} = -g_{\alpha\beta}u^{\alpha}\cdot\nabla_{v}n^{\beta}.$$

The second fundamental form measures how fast the unit normal vectors change as we move along unit tangent vectors; it is a measure of the extrinsic curvature of Σ in M.

Theorem 23.10 (Gauss-Codazzi). One has $R(\overline{g})_{ijk\ell} + \overline{k}_{ij}\overline{k}_{j\ell} - \overline{k}_{i\ell}\overline{k}_{jk} = R(g)_{ijk\ell}$ and $\overline{\nabla}_i\overline{k}_{j\ell} - \overline{\nabla}_j\overline{k}_{i\ell} = CR(g)_{ik\ell 0}$.

Thus, if $(\Sigma, \overline{g}, \overline{k})$ is to be an initial-data set, it had better satisfy the Gauss-Codazzi equations for the 4-manifold we want to embed it into.

Definition 23.11. Let $T_{\alpha\beta}$ be a symmetric, divergence-free 2-tensor. An *initial-data set* $(\Sigma, \overline{g}, \overline{k})$ corresponding to the energy-momentum tensor $T_{\alpha\beta}$ is the data of:

- 1. A 3-manifold Σ ,
- 2. a Riemannian metric \overline{q} on Σ with Riemann curvature tensor \overline{R} ,
- 3. and a symmetric 2-tensor \overline{k} on Σ ,

satisfying the Gauss-Codazzi constraints

$$\overline{R} + (\operatorname{tr} \overline{k})^2 + \overline{k}^{ij} \overline{k}_{ij} = 4\varphi^2 T_{tt}$$
$$\nabla^i \overline{k}_{ij} - \nabla_j \operatorname{tr} \overline{k} = 2\varphi T_{jt}$$

where

$$\varphi = \frac{\overline{k}_{ij}}{2\partial_t \overline{g}_{ij}}$$

Definition 23.12. Let $(\Sigma, \overline{g}, \overline{k})$ be an initial-data set corresponding to the energy-momentum tensor $T_{\alpha\beta}$. A *development* of $(\Sigma, \overline{g}, \overline{k})$ is an isometric embedding $\iota : \Sigma \to M$, where M = (M, g) is a Lorentzian (1 + 3)-manifold solving the Einstein equation

$$\operatorname{Ric}_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R = T_{\alpha\beta}$$

and \overline{k} is the second fundamental form of ι .

We think of the initial-data set as being the initial conditions of the Einstein equation and (M, g) as being the solution.

Definition 23.13. Let $S \subseteq M$ be a spacelike hypersurface. Then S is a Cauchy hypersurface if every maximal causal curve in M intersects S at exactly one point. Moreover, the *domain* of dependence is the maximal submanifold $D \subseteq M$ such that S is a Cauchy hypersurface of D.

For example, the initial-time slice Σ_0 of Minkowski spacetime is a Cauchy hypersurface. In fact, if *B* is a ball in Σ_0 , then the future-pointing causal cone based at *B* is the domain of dependence of *B*.

Definition 23.14. Let $(\Sigma, \overline{g}, \overline{k})$ be a respective initial-data set. Let (M, g) be a Lorentzian (1+3)-manifold. A globally hyperbolic development $\iota : \Sigma \to M$ is a development of $(\Sigma, \overline{g}, \overline{k})$ such that $\iota(\Sigma)$ is a Cauchy hypersurface of (M, g).

23.3 Well-posedness for the vacuum equation

Throughout this section we work with the Einstein vacuum equation $\operatorname{Ric}_{\alpha\beta} = 0$.

First, we recall the theorem that quasilinear wave equations are well-posed.

Theorem 23.15. Fix a Lorentzian metric g and consider the PDE

$$g^{\mu\nu}(x,\varphi(x))\partial_{\mu}\partial_{\nu}\varphi(x) = N(x,\varphi(x),\partial\varphi(x))$$
$$(\varphi,\partial_{t}\varphi)(x) = (\varphi_{0},\varphi_{1})(x)$$

where φ is an unknown. Let s > d/2 + 1. If $(\varphi_0, \varphi_1) \in H^s \times H^{s-1}(\Sigma_0)$, then there is an maximal eclipse time T > 0 and a unique solution $\varphi \in H^s([0, T] \times \Sigma_0)$.

For $T_{\alpha\beta}$ a tensor, we write

$$\hat{T}_{\alpha\beta} = T_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}g^{\alpha\beta}T_{\alpha\beta}$$

Theorem 23.16 (Choquet-Bruhat-Geroch). Let $(\Sigma, \overline{g}, \overline{k})$ be a smooth initial-data set with $\operatorname{Ric}_{\alpha\beta} = 0$. Then there is a unique maximal globally hyperbolic development (M, g, ι) of $(\Sigma, \overline{g}, \overline{k})$; i.e. a globally hyperbolic development such that for any globally hyperbolic development $(\hat{M}, \hat{g}, \hat{\iota})$, there is an isometric embedding $\Phi : \hat{M} \to M$ such that the diagram



commutes.

By lower-order terms we mean those of first or zeroth order (those which may serve as quasilinear perturbations of the d'Alembertian, which is a second-order linear operator). The idea of the proof is to write $\text{Ric}_{\alpha\beta}$ as a quasilinear wave equation and use local well-posedness to construct local solutions, then glue all the local solutions together using Zorn's lemma.

Proof. We have

$$\operatorname{Ric}_{\alpha\beta} = \partial_{\mu}\Gamma^{\mu}_{\alpha\beta} - \partial_{\alpha}\Gamma^{\mu}_{\mu\beta}$$

= $\frac{1}{2}g^{\mu\nu}\partial_{\mu}\partial_{\nu}g_{\alpha\beta} - \frac{1}{2}g^{\mu\nu}\partial_{\alpha}\partial_{\beta}g_{\mu\nu} + \frac{1}{2}g^{\mu\nu}\partial_{\alpha}\partial_{\nu}g_{\beta\nu} + \frac{1}{2}g^{\mu\nu}\partial_{\beta}\partial_{\mu}g_{\alpha\nu}$
= $\frac{1}{2}\partial^{\nu}\partial_{\nu}g_{\alpha\beta} + \partial_{\alpha}\Gamma_{\beta} + \partial_{\beta}\Gamma_{\alpha}$

where

$$\Gamma_{\beta} = \frac{1}{2}g^{\mu\nu}\partial_{\mu}g_{\beta\nu} - \frac{1}{2}\partial_{\beta}g_{\mu\nu} = \frac{1}{2}g^{\mu\nu}\Gamma^{\alpha}_{\mu\nu}g_{\alpha\beta}.$$

Let

$$S_{\alpha\beta} = \operatorname{Ric}_{\alpha\beta} - \partial_{\alpha}\Gamma_{\beta} - \partial_{\beta}\Gamma_{\alpha}.$$

Then the equation $S_{\alpha\beta} = 0$ is a quasilinear wave equation, so is locally well-posed, and has a solution on a submanifold M of $\mathbb{R} \times \Sigma$.

Recall that $\operatorname{Ric}_{\alpha\beta}$ is the Einstein tensor and hence

$$\nabla^{\alpha} \widehat{\mathrm{Ric}}_{\alpha\beta} = 0.$$

Taking the hat and divergence of both sides of the definition of $S_{\alpha\beta}$, we have

$$\nabla^{\alpha} (\widehat{\nabla_{\alpha} \Gamma_{\beta} + \nabla_{\beta} \Gamma_{\alpha}}) = 0.$$

But

$$\nabla_{\alpha} \widehat{\Gamma_{\beta}} + \overline{\nabla}_{\beta} \Gamma_{\alpha} = \nabla_{\alpha} \Gamma_{\beta} + \nabla_{\beta} \Gamma_{\alpha} - g_{\alpha\beta} g^{\mu\nu} \nabla_{\mu} \Gamma_{\nu}$$

 \mathbf{SO}

$$0 = \nabla^{\alpha} \nabla_{\alpha} - \Gamma_{\beta} + \nabla^{\alpha} \nabla_{\beta} \Gamma_{\alpha} - g^{\mu\nu} \nabla_{\beta} \nabla_{\mu} \Gamma_{\nu} = \nabla^{\alpha} \nabla_{\alpha} \Gamma_{\beta}.$$

Therefore Γ solves the wave equation. Since the wave equation is well-posed, it suffices to show therefore that $\Gamma|_{\Sigma} = 0$ and $\partial_t \Gamma|_{\Sigma} = 0$. For i, j spatial coordinates, we set $g_{ij}|_{\Sigma} = 0$ and $g_{tt}|_{\Sigma} = -1$, $g_{ti}|_{\Sigma} = 0$, $\partial_t g_{ij}|_{\Sigma}|_{\Sigma} = 2\bar{k}_{ij}$, and $\partial_t g_{t\alpha}|_{\Sigma} = 0$. Then $\Gamma|_{\Sigma} = \partial_t \Gamma|_{\Sigma} = 0$ by the Gauss-Codazzi equations. Thus with these initial conditions, $S_{\alpha\beta} = \operatorname{Ric}_{\alpha\beta}$ so the Einstein equation reduces to the quasilinear wave equation $S_{\alpha\beta} = 0$, and the solution manifold Msolves the vacuum Einstein equation, which is therefore locally well-posed.

Now let \mathcal{M} be the class of globally hyperbolic developments of $(\Sigma, \overline{g}, k)$, ordered by isometric embeddings which commute with the inclusions ι . This class is proper, but taking a quotient by isometry, we arrive at a poset. Taking injective limits, we show that every chain has an upper bound, so \mathcal{M} has a maximal element $\iota : \Sigma \to M$, the *set-theoretic* maximal globally hyperbolic development. It remains to show that ι is maximal in the sense of the definition of maximal globally hyperbolic development (and hence unique).

Let $\hat{\iota}: \Sigma \to M$ be a set-theoretically maximal globally hyperbolic development. We must construct a isometric embedding $\Phi: \hat{M} \to M$ making the diagram commute. By a *partial isometric embedding* of \hat{M} into M we mean a isometric embedding $\hat{U} \to M$ for some open set $\hat{U} \subseteq \hat{M}$. By local well-posedness, every point is contained in a neighborhood which admits a partial isometric embedding that makes the diagram commute, and by local uniqueness, they satisfy the cohomological conditions in the definition of a sheaf. Therefore there is a global partial isometric embedding, which is of course Φ .

We define the *development-theoretic union* $M \cup \hat{M} = M \coprod \hat{M} / \Phi$, where the coproduct \coprod is the sense of disjoint union. All conditions in the definition of a globally hyperbolic development are easily checked for $M \cup \hat{M}$ except that $M \cup \hat{M}$ is Hausdorff.

Assume that $M \cup M$ is not Hausdorff at a point $x \in M \cup M$. Then $x \in \partial(M \cup M)$, and by a difficult computation in Lorentzian geometry, there is a spacelike hypersurface S which touches ∂M exactly at x. Away from x, S and $\Psi(S)$ determine the same initial-data set. But by continuity, S and $\Psi(S)$ determine the same initial-data set at x as well.

But x lifts to a regular point in $M \coprod \hat{M}$ (and let us assume without loss of generality that x then lifts to a regular point in M), so there is a globally hyperbolic development extending from a S-neighborhood of x by local well-posedness. Since M is set-theoretically maximal, x does not lift to a point of ∂M . Therefore $x \notin \partial(M \cup \hat{M})$, a contradiction.

It follows that $M \cup \hat{M}$ is Hausdorff, and hence a globally hyperbolic development which contains M. So $M \cup \hat{M} = M$, and it follows that $\hat{M} = M$. So M is a globally hyperbolic development.

Chapter 24

General relativity in spherical symmetry

We now make a simplifying assumption to get rid of annoying obstructions in Lorentzian geometry: that of spherical symmetry.

Definition 24.1. A spacetime (M, g) is *spherically symmetric* if there is a SO(3)-action on M by g-isometries such that every SO(3)-orbit is a manifold of dimension at most 2.

Then the only possible orbits are fixed points and spheres of positive radius. For $p \in M$ we let S_p denote the orbit of p, and let r(p) denote the radius of S_p , which can be intrinsically defined by

$$r(p) = \sqrt{\frac{\mu(S_p)}{4\pi}},$$

 μ denoting area. Then zeroes of r are fixed points of SO(3).

When r(p) > 0, the induced metric on S_p is given by

$$g|_{S_p} = r^2 \underline{g}$$

where \underline{g} is the Riemannian metric of the unit 2-sphere S^2 .

We let $\mathcal{Q} = M/SO(3)$, so \mathcal{Q} is a Lorentzian (1+1)-manifold with boundary $\Gamma = \partial \mathcal{Q}$. Then

$$g = g_{\mathcal{Q}} + r^2 \underline{g}.$$

24.1 Double-null pairs

Definition 24.2. A *double-null pair* on M is a pair of SO(3)-invariant smooth functions $u, v : M \to \mathbb{R}$, increasing in time, such that

$$g^{\alpha\beta}du_{\alpha}du_{\beta} = g^{\alpha\beta}dv_{\alpha}dv_{\beta} = 0$$

and such that du, dv are linearly independent on every cotangent space. If we view u, v as coordinates on M and let θ, φ be the usual polar coordinates on S^2 , the tuple (u, v, θ, φ) is known as a system of *double-null pair coordinates*.

Assuming that M has double-null pair coordinates,

$$\underline{g} = d\theta^2 + \sin^2\theta \ d\varphi^2$$

and

$$g_{\mathcal{Q}} = -\Omega^2 du \cdot dv$$

for some function Ω .

Definition 24.3. The function Ω determined by double-null pair coordinates is called the *null lapse* of the double-pull pair.

Let us assume that every spacetime admits a double-null pair coordinate system. We have $g_{uu} = g_{vv} = 0$ and $g_{uv} = -\frac{1}{2}\Omega^2$.

In particular we have

$$g = \begin{bmatrix} 2^{-1}\Omega^2 & & \\ 2^{-1}\Omega^2 & & \\ & & r^2 & \\ & & & r^2 \end{bmatrix}$$

so $2\sqrt{-\det g}$ = We will always write the angular coordinates with capital letters.

By reparametrizing u, v to have bounded range, we can embed Q into a compact subset of the Minkowski spacetime \mathbb{R}^{1+1} .

Theorem 24.4. Let (u, v, θ, φ) be a double-null coordinate system with null lapse Ω and let ϕ be a spherically symmetric scalar field. Then

$$\Box_g \phi = -4\Omega^{-2} (\partial_u \partial_v \phi + r^{-1} \partial_u r \partial_v \phi + r^{-1} \partial_v r \partial_u \phi).$$

Besides this, we can write the Einstein tensor $G_{\alpha\beta}$ as

$$G_{uu} = \operatorname{Ric}_{uu} = -2r^{-1}\Omega^{2}\partial_{u}(\Omega^{-2}\partial_{u}r)$$

$$G_{vv} = \operatorname{Ric}_{vv} = -2r^{-1}\Omega^{2}\partial_{v}(\Omega^{-2}\partial_{v}r)$$

$$G_{uv} = \operatorname{Ric}_{uv} - \frac{1}{2}g_{uv}R = 2r^{-1}\partial_{u}\partial_{v}r + 2r^{-2}\partial_{u}r\partial_{v}r + \Omega^{2}r^{-2}$$

$$G_{AB} = \operatorname{Ric}_{AB} - \frac{1}{2}g_{AB} = -4r^{2}Omega^{2}\underline{g}_{AB}(\Omega^{-1}\partial_{u}\partial_{v}\Omega - \Omega^{-2}\partial_{u}\Omega\partial_{v}\Omega + r^{-1}\partial_{u}\partial_{v}r)$$

and all other entries determined by symmetry or vanishing.

From this, it is easy to see that the Einstein vacuum equation in spherical symmetry can be expressed as

$$\partial_u (\Omega^{-2} \partial_u r) = 0$$
$$\partial_v (\Omega^{-2} \partial_v r) = 0$$
$$\partial_u \partial_v r + 2r^{-1} \partial_u r \partial_v r + (2r)^1 \Omega = 0$$
$$\partial_u \partial_v \Omega - \Omega r^{-2} (2 \partial_u r \partial_b r + 2^{-1} \Omega^2) = 0.$$

The wave operator in spherical symmetry has principal part $\partial_u \partial_v$. So we view the first two Einstein equations as constraint equations (called *Raychaudhuri equations*) and the last two Einstein equations as quasilinear wave equations.

24.2 Local rigidity

We show that the Raychaudhuri equations form a strong constraint on the sort of solutions we are allowed to study.

Lemma 24.5. Consider the quasilinear wave equation

$$\partial_u \partial_v \Phi = N(\phi, \partial \Phi)$$

where N is C^1 . If Φ is a C^1 solution to the equation with Φ prescribed on the future-pointing lightcone centered at a point $p \in M$, then Φ is unique.

Since this is a wave equation, we expect to need $\partial \Phi$ as initial data as well in order for Φ to be unique. But the lightcone consists exactly of characteristic curves of Φ , one of which determines $\partial_u \Phi$ and the other determines $\partial_v \Phi$ – and they must be compatible since p touches both curves.

Proof. Notice that

$$\partial_u \partial_v \Psi (\partial_u + \partial_v) \Psi = .5 \partial_v (\partial_u \Psi)^2 + .5 \partial_u (\partial_v \Psi)^2.$$

Assume that Φ, Φ' are solutions and let $\Psi = \Phi - \Phi'$. Then

$$\partial_u \partial_v \Psi = \Psi \partial_\Phi N(\Phi, \partial \Phi) + O(\Psi, \partial \Psi).$$

Integrate along a future-pointing "diamond" whose first vertex is p = (0, 0) and whose sides are given by (or are perpendicular to – we call these C_u and C_v) the characteristic curves C_0, \underline{C}_0). This gives

$$\frac{1}{2}\left(\int_{C_u} (\partial_v \Psi)^2 + \int_{C_v} (\partial_u \Psi)^2 - \int_{C_0} (\partial_v \Psi)^2 - \int_{\underline{C}_0} (\partial_u \Psi)^2\right) \le C \int_D |\Psi| |\partial \Psi| + C \int_D |\partial \Psi|^2.$$

The integrals along C_0 and \overline{C}_0 are 0 because $\Phi = \Phi'$ there by assumption. So we have

$$\int_{C_u} (\partial_v \Psi)^2 + \int_{C_v} (\partial_u \Psi)^2 \le C \int_D |\Psi| |\partial \Psi| + \int_D |\partial \Psi|^2.$$

We use Gronwall's inequality to control the right-hand side. Now

$$\Psi(u,v) = \int_0^v \partial_v \Psi(u,\cdot)$$

which then vanishes. So this energy estimate gives $\Psi = 0$.

Theorem 24.6 (Birkhoff). Up to gauge symmetry, the solution to the Einstein vacuum equation in spherical symmetry near a point $p \in M$ is determined by r(p), the signs $\sigma_{\nu}, \sigma_{\lambda}$ of $\partial_u r(p)$ and $\partial_v r(p)$, and $g^{\mu\nu}\partial_{\mu}\partial_{\nu}r(p)$.

Proof. If u, v are a future-pointing double-null pair, and we transform them to (\tilde{u}, \tilde{v}) where $\tilde{u} > 0$ only depends on u and similarly for \tilde{v} , then (\tilde{u}, \tilde{v}) are a future-pointing double-null pair. This transformation results in the transformation of Ω by

$$-\Omega^2 \, du \, dv = -\tilde{\Omega}^2 \, d\tilde{u} \, d\tilde{v} = -\tilde{\Omega}^2 \tilde{u}' \tilde{v}' \, du \, dv$$

so $\tilde{\Omega}^2 \tilde{u}' \tilde{v}' = \Omega^2$. Let \underline{c}_p , c_p be the curves of the future-pointing lightcone along u, v from p. Then we can choose \tilde{u}, \tilde{v} so that $\tilde{\Omega} = 1$ on \underline{c}_p, c_p . Let us henceforth work in the coordinates \tilde{u}, \tilde{v} (so $\Omega = \tilde{\Omega}$).

By the lemma, we only need to determine Ω and r along \underline{c}_p, c_p , and this will uniquely determine the solution inside any diamond with two sides that lie along \underline{c}_p, c_p ; then make the diamond as big as we need.

Assume $\sigma_{\nu} = \sigma_{\lambda} = 0$. Applying the Raychaudhuri equations along \underline{c}_p, c_p , we have $\partial_u \partial_u r = 0$ on \underline{c}_p , so by the initial conditions we have $\partial_u r = 0$ on \underline{c}_p . Similarly $\partial_v r = 0$ on c_p . Thus r = r(p).

If $\sigma_{\nu} \neq 0$, $\sigma_{\lambda} \neq 0$, then we again have $\partial_u \partial_u r = 0$ along \underline{c}_p . The only remaining degree of freedom is our freedom to choose $g^{\mu\nu}\partial_{\mu}\partial_{\nu}r(p)$, which ends up determining $\partial_u r$. Therefore we know the value of r along \underline{c}_p, c_p .

Finally assume $\sigma_{\nu} \neq 0$ but $\sigma_{\lambda} = 0$. The proof is similar to the previous cases.

We have proven uniqueness in the future-pointing lightcone of p. By time-reversal symmetry we obtain uniqueness in the past-pointing lightcone. By spherical symmetry, we can switch u and v with -u and -v and run the same argument for the "left-pointing lightcone" and the "right-pointing lightcone" which is all four cones that are around p.

The point is that $r, \sigma_{\nu}, \sigma_{\lambda}$, and $\mu_0 = g^{\mu\nu} \partial_{\mu} r \partial_{\nu} r(p)$ are geometric data, and don't depend on the choice of coordinates; but everything that isn't determined by these terms is determined by our choice of coordinates. Actually, μ_0 is only needed when σ_{ν} and σ_{λ} are nonzero.

Definition 24.7. The *Hawking mass* is a function m on spherically symmetric spacetime defined by

$$g^{\mu\nu}\partial_{\mu}r\partial_{\nu}r = 1 - 2mr^{-1}$$

Note that m does not depend on a choice of coordinates since neither does r.

Lemma 24.8. The Hawking mass is constant on connected components.

Proof. We have

$$-4\partial_u r \partial_v r \Omega^{-2} = g^{\mu\nu} \partial_\mu r \partial_\nu r$$

so by implicit differentation we have

$$-2r^{-1}\partial_u m + 2\partial_u r r^{-2} m = -4\partial_u r \Omega^{-2} \partial_u \partial_v r.$$

Doing a bunch of algebra we see $\partial_u m = \partial_v m = 0$.

When m = 0 we will end up with Minkowski spacetime. If m > 0 one can show that the null lapse is given by

$$\Omega^2 = -\sigma_\nu \sigma_\lambda (1 - 2mr^{-1}).$$

This gives a certain metric that we call the Schwarzschild metric.

Definition 24.9. The Schwarzschild metric is the metric

$$g = -\sigma_{\nu}\sigma_{\lambda}(1 - 2mr^{-1}) \ dudv + r^2g.$$

One can construct a maximal Schwarzschild spacetime. In fact if we define μ by $1 - \mu = g^{\alpha\beta}\partial_{\alpha}\partial_{\beta}r$, then $r\mu$ is constant, and in fact we take $r\mu = 2m$. Doing some algebra and using the Raychaudhuri equations, we have $\Omega^2 = |1 - 2mr^{-1}|$. If $r \to r_0$ as $v \to \infty$, then $\partial_v r \to 0$, i.e. $r_0 = 2m$.

We now rephrase Birkhoff's theorem.

Corollary 24.10. If (M, g) is a spherically symmetric solution to the Einstein vacuum equation then (M, g) is locally isometric to an open subset of a Schwarzschild spacetime.

We think of Schwarzschild spacetime as an easy example of a black hole spacetime, for m > 0. Choosing our signs correctly, we have

$$g = -(1 - 2mr^{-1}) \, dudv + r^2g = -dudv + r^2g + o(1)$$

so that g approximates Minkowski spacetime for r large enough (or m small enough; if m = 0 it is Minkowski spacetime, with the singularity r = 0 artifically added by the choice of coordinates). That is, g is asymptotically flat, so models a gravitational system (where the mass is concentrated in a compact set – say, all the mass is inside some star.) In particular, if the observer is not massless, then the observer is at $r = \infty$. Drawing the Penrose diagram, our causal past, looking in from $r = \infty$, is r > 2m. Thus no matter how far in the future we are, we lie in the causal complement of the region r < 2m.

Definition 24.11. The boundary r = 2m of a black hole is called the *event horizon*. The region r < 2m is called a *black hole*.

In fact we have $R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta} \geq Cm^2r^{-6}$, so the spacetime has infinite curvature at 0.

But assume m < 0. Then r = 0 is a "naked singularity", which lies in our causal past. A major conjecture, the *weak cosmic censorship conjecture*, is that for any physically meaningful spacetime, naked singularities do not exist. Note that the Schwarzschild spacetime with m < 0 is not a counterexample, because such a spacetime somehow has negative mass, which is absurd.

24.3 Einstein-Maxwell equations

Let $F_{\mu\nu}$ be a real-valued 2-form on M, the *electromagnetic field*. If (M, g) is Minkowski spacetime, we can take $E_i = F_{0i}$ and $B_i = \epsilon_{ijk} F^{jk}/2$, the Hodge dual of E, to recover the electric and magnetic fields.

Definition 24.12. The Maxwell equations are the system $\nabla^{\mu}F_{\nu\mu} = 0, dF = 0.$

Let us assume that F is spherically symmetric; i.e. if $R \in SO(3)$ then $R^*F_{\mu\mu} = F_{\mu\nu}$, where we think of SO(3) as the symmetry group of (M, g). We will write

$$F = F_{uv} \, du \wedge dv + F_{\theta\varphi} \, d\theta \wedge d\varphi.$$

One can use algebraic topology to prove that that F_{uv} is completely determined by u, v and $F_{\theta\varphi}$ is completely determined by a function of u, v as well as $\sin \theta$. Since dF = 0, $\partial_u F_{\theta\varphi} = 0$, and $\partial_v F_{\theta\varphi} = 0$. So actually $F_{\theta\varphi} = m \sin \theta \ d\theta \wedge d\varphi$ for some constant m. Also,

$$0 = \nabla^{\mu} F_{u\mu} = -2\Omega^{-2} \partial_u (r^2 \Omega^{-2} F_{uv})$$

and similarly for v. Thus $\partial_u(r^2\Omega^{-2}F_{uv}) = 0$ and similarly for v. Thus $F_{uv} = e\Omega^2 r^{-2}$ for some constant e.

Theorem 24.13 (Weyl?). Every spherically symmetric solution F of the Maxwell equation is given by

$$F = e\Omega^2 r^{-2} \, du \wedge dv + b\sin\theta \, d\theta \wedge d\varphi.$$

Thus an electromagnetic field is completely determined by the pair (e, m). In Minkowski spacetime, $e\Omega^2 r^{-2} du \wedge dv = er^{-2} dt \wedge dr$ which is the electric field given by a point charge at the origin, while $b \sin \theta d\theta \wedge d\varphi$ is the magnetic flux through a sphere.

We now derive the Einstein-Maxwell system from the principle of least action. The Lagrangian density of the Einstein vacuum equation was $R \, dV$ while the Lagrangian density of the Maxwell equation is $-F_{\alpha\beta}F^{\alpha\beta}/2$. Thus we have

$$\operatorname{Ric}_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = 2T_{\alpha\beta}$$
$$T_{\alpha\beta} = F_{\alpha\mu} F^{\mu}_{\beta} - \frac{1}{4} g_{\alpha\beta} F_{\mu\nu} F^{\mu\nu}$$
$$\nabla^{\alpha} F_{\alpha\beta} = dF = 0.$$

As usual, $T_{\alpha\beta}$ is the energy-momentum tensor of the Maxwell equation. We now compute $T_{\alpha\beta}$ by

$$T_{uu} = T_{vv} = 0$$

$$T_{uv} = 4^{-1} \Omega^2 r^{-4} (b^2 - e^2)$$

since

$$F_{\mu\nu}F^{\mu\nu} = 2(g^{uv})^2(F_{uv})^2 + g^{AA'}g^{BB'}F_{AB}F_{A'B'} = 2r^{-4}(e^2 + b^2).$$

Moreover, T_{AB} is proportional to T_{uv} . Thus, the Einstein-Maxwell equations in spherical symmetry, obtained by plugging into the Raychaudhuri and Einstein spherically symmetric equations, is

$$0 = -2r^{-1}\Omega^{2}\partial_{u}(\Omega^{-2}\partial_{u}r)$$

$$0 = -2r^{-1}\Omega^{2}\partial_{v}(\Omega^{-2}\partial_{v}r)$$

$$0 = 2r^{-1}\partial_{u}\partial_{v}r + 2r^{-2}\partial_{u}r\partial_{v}r + \Omega^{2}2^{-1}r^{-2} - 2^{-1}\Omega^{2}r^{-4}(e^{2} + b^{2})$$

Theorem 24.14. Let (M, g, F) be a spherically symmetric solution to the Einstein-Maxwell system. Then for any $p \in M$, the solution is determined in an open neighborhood O of p by $r, \mu, \sigma_{\nu}, \sigma_{\lambda}, e = 2r^2 \Omega^{-2} F_{uv}$, and $b = \csc \theta F_{\theta\varphi}$.

Proof. The electromagnetic field is rigid since we are in spherical symmetry. Now run the proof of Einstein vacuum equation rigidity (using the Raychaudhuri equations, which are the same as before) but with $T_{\alpha\beta}$ given by (e, b).

Lemma 24.15. $d(1-\mu) \wedge dr = 0$.

Proof. Same as in the vacuum case, because $\operatorname{Ric}_{uu} = \operatorname{Ric}_{vv} = 0$.

As a result, there are constants C such that $1 - \mu = 1 - 2Cr^{-1} + 2C(e^2 + b^2)r^{-2}$. In fact, $1 - \mu = g^{\alpha\beta}\partial_u r \partial_v r = -4\partial_u r \partial_v r \Omega^{-2}$ so we have

$$d(1-\mu) = -4d(\partial_u r \partial_v r \Omega^{-2}) = -4\partial_u \partial_v r \Omega^{-2}(\partial_u r \ du + \partial_v r \ dv).$$

Therefore

$$d(1-\mu) = -4\Omega^{-2}\partial_u\partial_v r \ dr.$$

Using the Einstein-Maxwell equations we see that if $f(r) = -r^{-1}(1-\mu) + r^{-1} - (e^2 + b^2)r^{-3}$ then $d(1-\mu) = f(r) dr$. In addition, if $h = 1 - \mu$ then h' = f so $d(rh)/dr = 1 - (e^2 + b^2)r^{-2}$ whence $rh = C + r + (e^2 + b^2)r^{-1}$. This proves the above claim.

We now search for a global solution to the Einstein-Maxwell equation. We do this in Eddington-Finkelstein coordinates, which just means that $\partial_u r \Omega^{-2}$ is constant in one dimension and $\partial_v r \Omega^{-2}$ is constant in the other dimension. This is possible because of the Raychaudhuri equations. In these coordinates, $\Omega^2 = |1 - 2Cr^{-1} + (e^2 + b^2)r^{-2}|$ and $g = -\Omega^2 du \, dv + r^2 g$.

We look at the sign of the discriminant $C^2 - Q^2$ where $Q^2 = e^2 + b^2$. If 0 < |Q| < C then there are two solutions to the equation $1 - 2CR^{-1} + Q^2r^{-2} = 0$, namely $r_{\pm} = C \pm \sqrt{C^2 - Q^2}$. This is called the *subextremal case*.

By the Raychaudhuri equations the signs of $\partial_u r$ and $\partial_v r$ cannot change along the u and v directions respectively. So we can fix a sign for each and see what happens.

First take the case $\partial_u r < 0$, $\partial_v r > 0$. Assume that $r \to r_+$ as $u \to \infty$, $v \to -\infty$; then $r \to \infty$ as $u \to -\infty$, $v \to \infty$. Thus along every null curve, r tends to r_+ in one direction and ∞ in the other direction. On the other hand, if we take "initial data" $r = r_-$, then we hit r = 0 for some finite u, which is a singularity.

If $\partial_u r < 0$, $\partial_v r < 0$, then as $u \to -\infty$, $v \to -\infty$, $r \to r_+$. Similarly as $u \to \infty$, $v \to \infty$, $r \to r_-$.

Gluing together the above Penrose diagrams we construct all possible solutions to the Einstein-Maxwell equations in spherical symmetry. We have to make sure that Ω is continuous along the gluings, which can be guaranteed by a clever change of coordinates. The maximal such simply connected solution is called the *maximal Reissner-Nordstrom space-time*. It is not compact.

In the superextremal case $Q^2 > C^2$ we recover the negative-mass Schwarzschild solution. Finally we consider the extremal case $Q^2 = C^2$. The resulting maximal solution is the Bertotti-Robinson spacetime. One can show that

$$\partial_u \partial_v \log \Omega = 2r^{-2} \partial_u \partial_v r + ((2r^2)^{-1} - e^2 r^{-4})\Omega^2$$

using the equation for the angular Einstein tensor $\operatorname{Ric}_{AB} - g_{AB}R/2$ and the angular energymomentum $T_{AB} = Q^2$. One then shows that $\partial_u \partial_v \log \Omega = K\Omega^2$ for some $K = (2r_0^2)^{-1} - e^2 r_0^{-4}$. This is a constant-curvature spacetime.

Theorem 24.16 (Birkhoff for Einstein-Maxwell). If (M, g, F) is a spherically symmetric solution to the Einstein-Maxwell equation, then each point of M is contained in an open set which is isometric to an open set of either the maximal Reissner-Nordstrom spacetime, the Bertotti-Robinson spacetime, a Schwarzschild spacetime, or the Minkowski spacetime.

Chapter 25

Cosmic censorship

In GR, we are interested in two regimes: isolated gravitational systems (asymptotically flat spacetimes; there is a singularity at one point and everything else is a vacuum, so we are studying the dynamical structure) and cosmological systems (where we are modeling the entire universe, and we want to study the topological structure). For now, we will study the isolated case, and view it as a Cauchy problem.

Notice that the Cauchy problem behaves quite strange in the negative Schwarzschild spacetime (M, g). Suppose we have an initial-data set Σ for M; then, a geodesic in Σ along which $r \to 0$ cannot be extended to the future. Drawing the Penrose diagram we see that the negative Schwarzschild spacetime is "not deterministic," i.e. Σ does not uniquely determine the future because we cannot extend it into the future-pointing lightcone of the black hole.

At least in a positive Schwarzschild spacetime, these "incomplete geodesics" are inside the black hole region. Therefore the observer at infinity cannot see the singularity, where we cannot extend an initial data set to the future. But in the negative Schwarzschild spacetime, the observer sees the singularity. But negative Schwarzschild spacetimes have negative mass by definition, which makes no sense physically.

We thus state the weak cosmic censorship conjecture: an observer at infinity cannot see a singularity in a "typical" physically meaningful spacetime.

We call the future boundary of a Penrose diagram (limiting points of radial null geodesics along which $r \to \infty$) the *null infinity* of the spacetime. A spacetime has *complete null infinity* if the lengths of geodesics parallel to null infinity tend to ∞ as $r \to \infty$. In the negative Schwarzschild spacetime, the null infinity was incomplete because the null infinity was the limit of the causal future of the initial-data set, which was compact.

We will be deliberately vague about what we mean by a reasonable Einstein-matter system, but it will be the Einstein equation coupled to physically meaningful Lagrangian densities (i.e. the Maxwell density, the vacuum density, etc.) Similarly for physically-meaningful initial-data set but in particular the initial-data set should be a geodesically complete manifold. (This means that you can "follow a geodesic forever"; or in other words the domain of the exponential map $T\Sigma \rightarrow \Sigma$ is defined on all of $T\Sigma$.) This rules out the punctured line and manifolds with boundary, because those have singularities we can run into in finite distance, which does not seem physically reasonable.

By generic we mean in the sense of the Baire category theorem. In fact, Christodoulou has proven that a naked singularity is unstable, and under a slight perturbation of g necessarily

collapses into a black hole, and so is hidden from the observer at infinity, as in the positive Schwarzschild spacetime.

Conjecture 25.1 (weak cosmic censorship). Given a generic physically-meaningful initialdata set to a reasonable Einstein-matter system in an asymptoically flat universe, the future maximal globally hyperbolic development has complete null infinity.

Recall that by definition, the maximal globally hyperbolic development ends at the spacelike hypersurface wherein the development fails to be unique. This region is called a *Cauchy horizon*. In a Reissner-Nordstrom black hole, there is a Cauchy horizon, so that a test particle falling into a black hole is NOT unique.

Conjecture 25.2 (strong cosmic censorship). Given a generic physically-meaningful initialdata set to a reasonable Einstein-matter system in an asymptotically flat universe, the future maximal globally hyperbolic development is inextendible as a smooth Lorentzian manifold.

25.1 Einstein-Maxwell-charged scalar field equations

The most complicated model of the cosmic censorship conjectures is the *Einstein-Maxwell-charged scalar field* equation. A scalar field ϕ is a section of a complex line bundle E whose structure group is the orthogonal group O(1). This gives rise to a connection D on E and

$$F_{\alpha\beta} = [D_{\alpha}, D_{\beta}].$$

Locally, we have $D_{\alpha} = \partial_{\alpha} + iA_{\alpha}$. The action is given by

$$\rho(\phi, D, g) = \int R \, dV(g) - \int F^{\alpha\beta} F_{\alpha\beta} \, dV_g - 2 \int \langle D^{\alpha}\phi, D_{\alpha}\phi \rangle \, dV_g$$

where $\langle \phi, \psi \rangle = \operatorname{Re}(\phi \overline{\psi})$ is the natural real-valued inner product on a complex line bundle.

In case $\phi = 0$, the Einstein-Maxwell-charged scalar field reduces to the Einstein-Maxwell system, but it is dynamical because it solves the wave equation with connection D, namely

$$D_{\alpha}D^{\alpha}\phi = 0.$$

However, the Einstein-Maxwell-charged scalar field is too hard to study directly, so we restrict to subsystems thereof.

Example 25.3. The Einstein-scalar field equation or Christodoulou model is the Einstein-Maxwell-charged scalar field with trivial Maxwell tensor, $F_{\alpha\beta} = 0$. Then $D = \partial$, so we do not need to worry about the curvature of the line bundle. That is, we can think of ϕ as a mapping $\phi : M \to \mathbb{R}$. It is the model that we will study when we consider the weak cosmic censorship conjecture.

Example 25.4. The *Einstein-Maxwell-uncharged scalar field* equation or *Daferemos model* is the system obtained by decoupling ϕ from F. In other words, $\phi : M \to \mathbb{R}$ (so the curvature of the line bundle is trivial). It is the model where Reichner-Nordstrom spacetimes make sense, so we study the strong cosmic censorship here.

We now study the (relativistic) kinetic theory of the Einstein equation. Let M be a spacetime, so T^*M , the cotangent bundle, has a natural symplectic form, namely

$$\omega = dx^{\alpha} \wedge dp_{\alpha}.$$

Here x^{α} is a coordinate system on $U \subseteq M$ and we view a covector as $p_{\alpha} dx^{\alpha}$. Given $H \in C^{\infty}(T^*M)$ we define the Hamiltonian vector field by

$$(X^H)^{\alpha} = \omega^{\alpha\beta} dH_{\beta}.$$

In case $H = 2^{-1}p^{\alpha}p_{\alpha}$ then X^{H} is the vector field on the cotangent bundle whose flow restricts to the Hamiltonian flow on M. We apply the Legrende transform $(x^{\alpha}, p_{\alpha}) \mapsto (x^{\alpha}, p^{\alpha})$ we get a flow on TM for which $\dot{x}^{\alpha} = p^{\alpha}, \dot{p}^{\alpha} = -\Gamma^{\alpha}_{\beta\gamma}p^{\beta}p^{\gamma}$. We let Φ^{H} denote the induced flow of X_{H} .

Definition 25.5. A Vlasov field is a positive measure μ on T^*M which is invariant under the pullback by Φ_t^H for every t; that is,

$$\mu = (\Phi_t^H)^* \mu.$$

In Newtonian mechanics, one assumes that the Vlasov field is absolutely continuous with respect to the natural volume form ϵ induced by the symplectic form μ . Using the Radon-Nikodym theorem, we find an f so that $\mu = f\epsilon$. Since Φ^H preserves ϵ we just need to check that $X^H f = 0$, the Vlasov equation.

Now T^*M is foliated by level hypersurfaces of H, and $X^H H = 0$, so Φ^H preserves the foliation of T^*M . Now a null geodesic is one arising from the flow restricted to H = 0, and timelike geodesics are those for which H = -1. To restrict to future-pointing geodesics we assume $p^0 < 0$. Thus we define P_0^+ to be the (x, p) with H(x, p) = 0 and $p^0 < 0$. Similarly for P_1^+ where we have H(x, p) = -1. These level hypersurfaces are 7-manifolds and we search for a top form on them. Now

$$\epsilon_{P_{\sigma}^+} = c \ dH \wedge \omega \wedge \omega \wedge \omega$$

for some function c allowed to depend on $\sigma\{0, -1\}$. For $\mu = f \epsilon_{P_{\sigma}^+}, X^H f = 0$ iff

$$p^{\alpha}\partial_{\alpha}f = \partial_{\alpha}g^{\beta\gamma}p_{\beta}p_{\gamma}\partial_{\alpha}f = 0.$$

Of course if $\sigma = 0$ then we are thinking of our particle as a photo (no mass) so we say that this is the "massless" case and $\sigma = -1$ is the "massive" case.

Definition 25.6. Let μ be a Vlasov field which is absolutely continuous with respect to $\epsilon_{P_{\sigma}^+}$. The associated energy-momentum tensor $T_{\alpha\beta}$ of μ is given weakly by (with φ a test function)

$$\int_M T_{\alpha\beta} x\varphi(x) \ dV(g) = \int_{T^*M} p_\alpha p_\beta \varphi(x) \ d\mu.$$

The *number density* is

$$\int_{M} N_{\alpha}(x)\varphi(x) \ dV(g) = \int_{T^*M} p_{\alpha}\varphi(x) \ d\mu$$

Since μ is absolutely continuous, it is supported on the 7-manifold which is a level hypersurface of H. If f is the Radon-Nikodym derivative, then

$$T_{\alpha\beta}(x) = \int_{T^*M} p_{\alpha} p_{\beta} f(-\det g)^{-1/2} \epsilon_{P_{\sigma}^+} |_{T^*_x M}$$

and

$$N_{\alpha}(x) = \int_{T^*M} p_{\alpha} f(-\det g)^{-1/2} \epsilon_{P_{\sigma}^+} |_{T^*_x M}.$$

To couple the Vlasov field to the Maxwell equation we take $N_{\beta} = \nabla^{\alpha} F_{\alpha\beta}$ and

$$2H = g^{\alpha\beta}(p_{\alpha} + A_{\alpha})(p_{\beta} + A_{\beta}).$$

Example 25.7. A subsystem of the Einstein-Vlasov system is the *Einstein-null dust system*. It is too simple to be realistic but is useful to demonstrate computations. The interpretation is that everything travels along radial null geodesics. So there is no mass, and the physical system consists solely of radiation moving radially.

An null dust field which is outgoing is characterized by having energy-momentum tensor $T_{\alpha\beta}$ such that

$$T_{uu}^{out} \ge 0$$

with other components zero. Similarly $T_{vv}^{in} \leq 0$ for incoming null dust fields. Now $\nabla^{\alpha}T_{\alpha\beta} = 0$ so $\partial_{v}T_{uu}^{out} = 0$. (Similarly $\partial_{u}T_{vv}^{in} = 0$.)

We will assume that there are two noninteracting null dusts, one incoming and one outgoing. That is, the Einstein-null dust equation is given by

$$\operatorname{Ric}_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R = 2(T_{\alpha\beta}^{in} + T_{\alpha\beta}^{out}).$$

25.2 The structure of toy models

Two basic papers about the a priori characterizations of solutions to spherically symmetric toy models are Daferemos "Spherically symmetric spacetimes with a trapped surface" and Komnemi "The global structure of a spherically symmetric charged scalar field spacetime". Let us give a shallow introduction to this theory.

We will let (M, g) be the 1+3-dimensional maximal globally hyperbolic development with a spherically symmetric initial data set Σ_0 . Let (Q, g_Q) be the quotient of (M, g) by SO(3). We will assume that Σ_0 is diffeomorphic to \mathbb{R}^3 or $\mathbb{R} \times S^2$. (The latter is the initial-data set of the spacetimes for which we will study the strong cosmic censorship conjecture.)

If $\Sigma_0 = \mathbb{R}^3$, then by algebraic topology, there is a fixed point of SO(3). In other words, the set $\Gamma = \{r = 0\}$ has

$$\Gamma \cap \Sigma_0 = \{p\}.$$

On the other hand, if $\Sigma_0 = \mathbb{R} \times S^2$, then SO(3) cannot have any fixed points.

By global hyperbolicity, there is a future-pointing double null pair (u, v) on Q. The existence of a double null pair implies that there is an embedding $Q \to \mathbb{R}^2$, i.e. a Penrose diagram. We will write \overline{Q} for the closure of Q inside \mathbb{R}^2 ; i.e. if Q is not a closed manifold then we will take it to be a manifold with boundary.

Definition 25.8. $T_{\alpha\beta}$ obeys the *dominant energy condition* if for every causal, futurepointing vectors x, y,

$$T_{\alpha\beta}x^{\alpha}y^{\beta} \ge 0.$$

In fact, $T_{\alpha\beta}\dot{\gamma}^{\alpha} = J_{\beta}$ should be interpreted as the "energy-momentum" along γ . In fact, the coordinate of J_{β} along γ is the energy along γ . So the dominant energy condition says that there is positive energy.

In spherical symmetry, the dominant energy condition is equivalent to $T_{uu} \ge 0$, $T_{vv} \ge 0$, $T_{uv} \ge 0$. It follows that $\operatorname{Ric}_{uu} \ge 0$, $\operatorname{Ric}_{vv} \ge 0$. Since

$$\operatorname{Ric}_{uu} = -2r^{-1}\Omega^2 \partial_u (\Omega^{-2} \partial_u r)$$

it must be that the sign of $\partial_u r$ is preserved, and similarly for $\partial_v r$. Thus r is monotone in u and v separately.

Henceforth we assume the dominant energy condition. It therefore makes sense to also assume the antitrapping condition:

Definition 25.9. Σ_0 obeys the antitrapping condition if: if $\Sigma_0 = \mathbb{R}^3$ then $\partial_u r < 0$ on Σ_0 ; if $\Sigma_0 = \mathbb{R} \times S^2$ then $\partial_u r < 0$ on some Σ'_0 a connected subset of Σ which meets the ideal endpoint of Σ_0 on the right.

Let

$$Q' = \{ (u, v) \in Q : \exists u_0 \ (u_0(v), v) \in \Sigma_0 \}.$$

Then the antitrapping condition implies that $\partial_u r < 0$.

Definition 25.10. Assume the antitrapping condition. $(u, v) \in Q'$ is trapped if $\partial_v r < 0$. $(u, v) \in Q'$ is regular if $\partial_u r > 0$. (u, v) is marginally trapped if $\partial_v r = 0$.

Because of these conventions, we say that u is incoming and v is outgoing.

In a black hole, every point is trapped. The event horizon is marginally trapped. Formally, if T is the set of trapped points, and (u, v') lies in the future of (u, v), then $(u, v) \in T$ implies $(u, v') \in T$.

Theorem 25.11 (Penrose singularity theorem). Suppose that T is nonempty. Then there is an incomplete outgoing null geodesic.

Recall that a geodesic γ is complete if for every t such that $\gamma_{\dot{\gamma}}\dot{\gamma}(t) = 0$ (t is an affine parameter), $\gamma(t)$ exists. This is not the case if γ runs into a boundary. That is, there is an incomplete geodesic the exponential map $TM \to M$ fails to be defined far away from the origin of each tangent space. Incomplete null geodesics can be interpreted physically as meaning that a light wave fails to exist after traveling a finite distance.

Proof. Let $(u_0, v_0) \in T$ be trapped, and let (u_1, v_1) be the endpoint of the outgoing null geodesic from (u_0, v_0) . This is finite because we embedded Q in \mathbb{R}^2 . Now we compute

$$\int_{v_0}^{v_1} \Omega^2(u,v) \ dv$$

and use the Raychaudhuri equations and the trapping conditions to conclude that the integral is the integral of a bounded function over a compact set. So it's finite, hence an affine parameter. $\hfill \Box$

Actually the Penrose singularity theorem holds in much greater generality. The existence of trapped surfaces is an open condition on the moduli space of all initial-data sets, which implies that there is a *stable* singularity, which necessarily follows from the existence of black holes.

25.3 Penrose inequalities

We generalize the result that says that the mass of the universe is positive if there are no black holes, to bound the mass of the universe in terms of the radius of the black hole. We follow Daferemos's paper "Spherically symmetric spacetimes with a trapped surface".

Let (M, g) be a spherically symmetric solution to the Einstein equation, which is the future maximally globally hyperbolic development of a spherically symmetric initial-data set Σ_0 , where Σ_0 is either homeomorphic to \mathbb{R}^3 to $\mathbb{R} \times S^2$. Let Q = M/SO(3) be the Penrose diagram of (M, g). We will assume the dominant energy condition on the energy-momentum tensor T (i.e. $T_{uu} \ge 0$, $T_{vv} \ge 0$, $T_{uv} \ge 0$). We also assume that there are no antitrapped spheres (which for $\Sigma_0 = \mathbb{R}^3$ means that $\partial_u r < 0$ on Σ_0 .)

Let Q' be the set of points in the Penrose diagram which are in the image of an incoming null curve from Σ_0 . By the Raychaudhuri equations and the assumption on antitrapped spheres, $\partial_u r < 0$ on Q'.

Let A be the apparent horizon, i.e. those (u, v) for which $\partial_v r(u, v) = 0$. Every event horizon is contained in the apparent horizon.

Let U be the set of u such that $\sup_{v} r(u, v) = \infty$. Thinking of Q as a bounded subset of \mathbb{R}^2 we let ζ^+ be the set of $(u, v) \in \partial U$ such that $u \in U$.

Definition 25.12. ζ^+ is the *future null infinity* of Q.

Lemma 25.13. If ζ^+ is nonempty, then it is a connected incoming null curve emanating from the interior.

Proof. If $(u, v) \in \zeta^+$ then we can find a point on Σ_0 whose lightcone includes (u, v).

Lemma 25.14. $J^{-}(\zeta^{+}) \subseteq R$, the set of regular points.

So a particle cannot end up in the future null infinity if it is trapped or lies on an event horizon.

Recall that the Hawking mass m at (u, v) is defined by

$$1 - \frac{2m}{r} = -4\frac{\partial_u r \partial_v r}{\Omega^2}.$$

Lemma 25.15. One has $\partial_u m = 2r^2 \Omega^{-2} (T_{uv} \partial_u r - T_{uu} \partial_v r)$ and similarly for v.

Proof. Use the Einstein equations in spherical symmetry and the dominant energy and noantitrapping conditions. \Box

Corollary 25.16. Inside $R \cup A$, $\partial_u m \leq 0$ and $\partial_v m \geq 0$.

Lemma 25.17. Inside Q', the sign of $\partial_v r$ is the sign of $1 - 2mr^{-1}$.

So in particular, a point is trapped provided that $1 - 2mr^{-1}$.

Definition 25.18. Fix $(u, v) \in J^{-1}(\zeta^+)$. Define the *Bondi mass*

$$M(u) = \lim_{v \to v_{\zeta^+}} m(u, v).$$

The ADM mass is

$$M_{ADM} = \lim_{u \to \Sigma_0} M(u).$$

So the Bondi mass is the mass observed by someone standing at the future end of a curve for which u is constant. The ADM mass is the mass observed by an observer at $\partial \Sigma_0$. Here "feeling mass" means experiencing a gravitational field.

Theorem 25.19 (positive mass theorem). If $\Sigma_0 = \mathbb{R}^3$ then $M_{ADM} \ge 0$.

Proof. Either $\Sigma_0 \subseteq R$ or not. If not, then there is a point on Σ_0 which does not end up at ζ^+ , and in particular there is a point (u_0, v_0) on $\partial R \cap \Sigma_0$ which lies in the apparent horizon. So at that point, $1 = 2mr^{-1}$. Since r > 0, m > 0, and the monotonicity properties above guarantee that the regular points also have positive mass. The observer at infinity can only feel things in his causal past, in particular $\Sigma_0 \cap R$, so we're done.

If $\Sigma_0 \subseteq R$, note that since g is smooth, r is Lipschitz. So

$$1 - 2mr^{-1} = g(\partial r, \partial r)$$

is bounded, whence $m \to 0^+$ as $r \to 0$ on Σ_0 . By monotonicity, $m \ge 0$ on Σ_0 , and so $M_{ADM} \ge 0$.

Note that in the case $\Sigma_0 \subseteq R$, we used the fact that $\Sigma_0 = \mathbb{R}^3$ so show that g is smooth and that Σ_0 is connected. We used the dominant energy condition and the antitrapping to guarantee monotonicity.

Corollary 25.20 (Riemannian Penrose inequality). Let S_R be a minimal sphere in Σ_0 of radius R > 0, and assume that the second fundamental form is 0. Then

$$M_{ADM} \ge \frac{R}{2}.$$

In \mathbb{R}^3 there are no minimal spheres so we take $R \to 0$. The positive mass theorem is sharp, because the ADM mass of Minkowski spacetime is 0.

Definition 25.21. The generalized extension principle is the assumption that for every $p \in \overline{Q}, q \in I^{-}(p), q \neq p$, if

$$D = J^+(q) \cap J^-(p) \setminus p,$$

then D has finite volume and

$$0 < \inf_D r < \sup_D r < \infty.$$

The generalized extension principle holds for any reasonable spacetime.

Example 25.22. The generalized extension principle is not true for the Einstein null dust spacetime.

Definition 25.23. The event horizon H^+ is the future boundary of $J^-(\zeta^+)$.

Then one has

$$\lim_{v \to \zeta^+} r = \sup_{H^+} r.$$

Definition 25.24. The final Bondi mass is

$$M_f = \lim_{u \to u_{\Box}} M(u) = \inf_u M(u),$$

the limit taken as u goes to the future.

The fact that this is an infimum follows from the monoticity assumptions.

Theorem 25.25 (Penrose event horizon inequality). Assume the generalized extension principle. Then

$$\sup_{H^+} r \le 2M_f.$$

We think of the sup as the radius of the black hole. Unravelling the definitions, we obtain a lower bound on all Bondi masses that follows from the size of the black hole.

The idea of the proof is that if we have control of $1-2mr^{-1}$, then we use the Raychaudhuri equation

$$-4\partial_u r \Omega^{-2} = \frac{1 - 2mr^{-1}}{\partial_v r}$$

to control $\partial_v r$ in terms of $\partial_u r$. We then use the definition of the Hawking mass to control the integral of the energy-momentum tensor. So if the conclusion of the Penrose event horizon inequality fails, we can find a (u, v) on the event horizon such that $r > 2M_f$. Since M and r obey similar monoticity conditions, we obtain an absurd bound on the mass.

25.4 Recent progress on strong censorship

Conjecture 25.26. For a generic asymptotically flat initial data set for a "reasonable" Einstein-matter Lagrangian, the future maximal globally hyperbolic development is inextendible as a "suitably regular" Lorentzian manifold.

Example 25.27. The Reichner-Nordstrom spacetime is a (highly nongeneric) counterexample.

To prove the strong cosmic censorship conjecture one would need to show that the Cauchy horizon of any counterexample must be unstable in the sense that a small perturbation of the initial-data set would necessarily destroy the Cauchy horizon. To do this, we first characterize its stability properties.
Example 25.28. Let us study the stability properties of Cauchy horizons in the Einstein-Maxwell null dust Lagrangian. Recall that null dust is defined by its energy momentum tensor. That is, T_{uu}^{out} is nonzero and can depend on u and v but all other $T_{\alpha\beta}^{out} = 0$. This is the outgoing null dust. By Noether's theorem, $\nabla^{\mu}T_{\mu\nu}^{out} = 0$ which implies that T_{uu} does not depend on v. We also have a term T_{vv}^{out} which can be nonzero and only depends on v.

The Einstein-Maxwell-null dust equation is $G_{\alpha\beta} = 2T_{\alpha\beta}$ like usual, where where

$$T_{\alpha\beta} = T_{\alpha\beta}^{out} + T_{\alpha\beta}^{in} + T_{\alpha\beta}^{Max}$$

where T^{Max} is the Einstein-Maxwell energy-momentum tensor. But $T_{uu}^{Max} = T_{vv}^{Max} = 0$, so the null dust coordinates do not interact with the Einstein-Maxwell coordinates.

First we treat the case $T^{out}|_{\Sigma_0} = 0$. Assume that $T^{in}|_{\Sigma_0}$ is small and rapidly decaying. We claim that this will deform into a large perturbation of the Cauchy horizon. In these assumptions, the metric tensor is given by the *Vaidya metric*. In Eddington-Finkelstein coordinates for v (i.e. v is normalized so $\Omega^2(\partial_u r)^{-1} = -2$, which is possible by the Raychaudhuri equations),

$$g = -\Omega^2 \, du dv + r^2 \underline{g} = -\Omega^2 ((\partial_u r)^{-1} \partial_u r + (\partial_v r)^{-1} \partial_v r) \, du dv + r^2 \underline{g}$$

= 2 drdv - 2\overline{v}_r dv^2 + r^2 g.

We introduce the *modified Hawking mass* ϖ defined by

$$g^{\alpha\beta}\partial_{\alpha}r\partial_{\beta}r = 1 - 2\varpi r^{-1} + \varpi^2 r^{-2}.$$

Then $\partial_u \varpi = -2r^2 \partial_v r \Omega^{-2} T_{uu}$ and similarly for v. In Eddington-Finkelstein coordinates,

$$\partial_v \varpi = r^2 T_{vv}.$$

Thus we arrive at the Vaidya metric

$$g = 2 dr dv - (1 - 2\varpi r^{-1} + Q^2 r^{-2} dv^2 + r^2 \underline{g})$$

$$\partial_v \varpi = r^2 T_{vv}.$$

We now normalize so $T_{vv}|_{\Sigma_0}(v) = \varepsilon v^{-p}$ for p, ε parameters. Then $\partial_v \varpi = \varepsilon v^{-p}$. We assume p > 1; then $\partial_v \varpi$ is L^1 and ϖ stays finite up to the Cauchy horizon. Now we do not have $r \to 0$ at the Cauchy horizon, so the Vaidya metric cannot blow up. But $v \to \infty$ at the Cauchy horizon, and Ric blows up at the Cauchy horizon. In double-null coordinates,

$$\Omega^2 = O(e^{-2Kv})$$

for some K, when u is held fixed. So if $L = \Omega^2 \partial_v$, L "should be" a well-behaved vector field, yet L experiences exponential growth as we approach the Cauchy horizon. Now $\operatorname{Ric}(L, L) = 2r^{-2}v^{-p}O(e^{4Kv})$ which blows up at the Cauchy horizon.

So Ric is highly unstable near the Cauchy horizon, even though the Vaidya metric itself is stable. But a theorem of Poisson and Israel in the early 90's shows that the blowup of ϖ near the Cauchy horizon is generic, so the Einstein-Maxwell null dust system with no outgoing radiation is highly unrealistic. In fact if T^{out} is nonzero on Σ_0 then $\varpi = \infty$ on the Cauchy horizon. This phenomenon is known as the mass inflation scenario. **Example 25.29.** We now treat the more realistic Einstein-Maxwell uncharged scalar field system. Here we have introduced a scalar field ϕ on M which is governed by the wave equation $\Box_g \phi = 0$. When we say it is uncharged we mean that we do not use the Maxwell tensor to introduce curvature on the line bundle that ϕ maps M into (so we can take that line bundle to just be \mathbb{R}).

Theorem 25.30 (Kommemi). Assume that Σ_0 is homeomorphic to $\mathbb{R} \times S^2$, and is asymptotically flat on two ends. Then there are at most two Cauchy horizons, and there must be a complete null infinity. If the spacetime is C^2 -extendible, then it must be extendible through the Cauchy horizon in the sense that there must be a timeline geodesic γ in the extension which meets the Cauchy horizon on the interior of γ .

Let us consider the initial-value problem inside the black hole region. We normalize v so that $-2\partial_v r = \Omega^{-2}$ on the event horizon, and we assume that there is a p > 1/2 such that $\phi = O(v^{-p})$ on the event horizon. If ϕ is smooth, then Daferemos proved that there is a Cauchy horizon, and that g is continuous up to the Cauchy horizon. Moreover, if ϕ is compactly supported on Σ_0 then $\partial_v \phi = O(v^{-4})$. This estimate on ϕ is known as the *Price law rate*.