Real Analysis Notes

Sebastian Sanchez

3/2: **Generalizing the Caratheodory Construction**.

Let X be a set. Let $\mathcal{A}_{0_0}$ be an ''algebra of sets'' on X. (I.e., $\varnothing \in \mathcal{A}_{0_0}$, closed under complements, ***finite*** unions).

Also recall, pre-measure $\rightsquigarrow$ outer measure $\rightsquigarrow$ measure.

Definition 0.0.1. $\mu_0 : \mathcal{A}_{0_0} \longrightarrow \mathbb{R}$ is a pre-measure if

- $\mu_0(\varnothing) = 0$
- $\mu_0(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^n \mu_0(A_i)$ if $\{A_i\}$ are pairwise disjoint
- $\mu_0(A) \geq 0, \forall A \in \mathcal{A}_{0_0}$.

Remark. We didn't have this precisely when building m on $\mathbb{R}$. As in for $X = \mathbb{R}$, we had (something like... my notes seem off here...) $\mathcal{A}_{0_0}$ = (finite unions of) open intervals/rays [not closed under complements].

Example 0.0.2. Let $(X, \mathcal{A}_0, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. Our goal is to build a measure space, $(X \times Y, \mathcal{A}_0 \otimes \mathcal{B}, \mu \otimes \nu)$. To this end we define $(\mathcal{A}_0 \otimes \mathcal{B})_0 := \bigcup_{i=1}^n A_i \times B_i$, for $A_i \in \mathcal{A}_0, B_i \in \mathcal{B}$. In other words, we define this particular σ-algebra to be finite unions of disjoint rectangles.

Claim: $(\mathcal{A}_0 \otimes \mathcal{B})_0$ is an algebra of sets. Indeed, we can clearly see that we have the empty set. For complements, see that

$$(X \times Y) \setminus (A \times B) = A \times (Y \setminus B) \cup (X \setminus A) B \cup (X \setminus B) \times (Y \setminus B)$$

But we want to show this for some arbitrary union of rectangles. Now,

$$(X \times Y) \setminus \left(\bigcup_{i=1}^n A_i \times B_i \right) = \bigcap_{i=1}^{n''} (X \times Y) \setminus (A_i \times B_i) = \bigcap_{i=1}^{n} \bigcup_{j=1}^{3} C_{ij} \times D_{ij}.$$

We'll prove intersections and unions separately.

Proof. Intersection of Rectangles: $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$. Note that for each term of the product belongs to its respective σ-algebra , that is, on the RHS of the equality. Moreover,

$$\left(\bigcup_i A_i \times B_i \right) \cap \left(\bigcup_j A_j \times B_j \right) = \left(\left(\bigcup A_i \right) \cap \left(\bigcup C_j \right) \right) \times \left(\left(\bigcup B_i \right) \cap \left(\bigcup D_j \right) \right).$$

To make this clearer, observe that $(i_1, j_1) \neq (i_2, j_2) \implies$ that $i_1 \neq i_2$ or $j_1 \neq j_2$. If $i_1 \neq i_2 \implies$ use disjointedness of $A_i \times B_i$, and if $j_1 \neq j_2 \implies$ use disjointedness of $C_j \times D_j$. $\square$

Next, for closure under unions, recall that $\bigcup_i A_i \times B_i = (X\times Y)\setminus(\bigcap_i (X\times Y)\setminus(A_i\times B_i))$. Define $(\mu\otimes\nu)_0\left(\bigcup_i A_i\times B_i\right) := \sum_i \mu(A_i)\nu(B_i)$. We call this the ''measure formula for rectangles'' and note that when the measure in question is the *probability* measure, this is like saying we are ''forcing events to be independent.''

Claim: ν_0 is a a pre-measure. Briefly we note a conventions that $0\cdot\infty := 0$. We'll show that $A\times B = \bigsqcup_{i=1}^n C_i\times D_i, \mu(A)\nu(B) = \sum_{i=1}^n \mu(C_i)\nu(D_i)$. We'll use that $\chi_{A\times B} = \sum_{i=1}^n \chi_{C_i}\chi_{D_i}$.

Proof. Fix $x\in X$. Define $f_x : Y \longrightarrow \mathbb{R}$ by $f_x(y) = \chi_{A\times B}(x,y) = 1 \iff x \in A, y\in B$. Another way to see this, $f_x(y) = \sum_i \chi_{C_i}(x)\chi_{D_i}(y) = \begin{cases} \sum \chi_{D_i}(y) & x\in C_i \\ 0 & x\notin C_i \end{cases} \rightsquigarrow$ integrate $\displaystyle\int_Y f_x(y)\,\mathrm{d}\mu = \begin{cases} \sum \mu\left(D_i\right)_j & x\in C_i \\ 0 & x\notin C_i\end{cases} = \chi_{C_i}\cdot\sum \nu(D_i)$.

Define $h : X\longrightarrow\mathbb{R}$ by $h(x) = \displaystyle\int_Y f_x(y)\,\mathrm{d}\nu = \chi_{C_i}\cdot\sum\nu(D_j) \implies \int_X h\,\mathrm{d}\mu = \sum_{i,j}\mu(C_i)\nu(D_j)$.

But $\displaystyle\int_X h(x)\,\mathrm{d}\mu = \int_X\left(\int_Y \chi_{A\times B}(x,y)\,\mathrm{d}\nu\right)\mathrm{d}\mu = \int_X\left(\int_Y \chi_A(x)\chi_B(y)\,\mathrm{d}\nu(y)\right)\mathrm{d}\mu(x) = \int_X \chi_A(x)\left(\int_Y \chi_B(y)\,\mathrm{d}\nu\right)\mathrm{d}\mu = \int_X \chi_A(A)\nu(B)\,\mathrm{d}\mu = \mu(A)\nu(B)$. □

3/3 - **Change of Schedule**.

- Standard lecture now on Wednesday.
- Monday is the new problem set day.

Last time: talked of *pre-measures*, with the example of the product pre-measure.

Today: Tracing the progression from pre-measure $\rightsquigarrow$ outer measure $\rightsquigarrow$ measure.

Definition 0.0.3. Let X be a set, $\mathcal{A}_{0_0}$ is an algebra of sets, $\mu_o : \mathcal{A}_{0_0} \longrightarrow \mathbb{R}$ is a pre-measure, then the outer measure induced by μ_o is the function $\mu^* : 2^X \longrightarrow \mathbb{R}$ defined to be $\mu^*(A) := \inf \left\{ \sum_{i=1}^{\infty} \mu_o(A_i) : A \text{ is a cover of } A \right\}$.

Notice that if S is said to be the top-half of a semi-circle, and suppose we try to cover S with rectangles ''length-wise,'' then we can observe $\mu^*(S) \approx \sum \text{area(rect)}$.

Definition 0.0.4. $A \subset X$ is called *μ-measurable* if $\forall B \subset X, \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A)$.

Note outer measure still satisfies monotonicity: if $A \subset B$ then $m^*(A) \leq m^*(B)$, and subadditivity: $\mu^*(A \cup B) \leq \mu^*(A) + \mu^*(B)$.

Theorem 0.0.5. *If $\mathcal{A}_0$ is a collection of μ-measurable sets, $\mathcal{A}_0$ is a σ-algebra.*

Proof. *Same as Lebesgue proof.* □

Theorem 0.0.6. *$\mathcal{A}_{0_0} \subset A_0, m^*(A) = \mu_o(A)$, where $\mathcal{A}_{0_0}$ denote the μ-measurable sets of m^*, $\forall A \in \mathcal{A}_{0_0}$.*

Proof. We'll show $m^* = m_o(A), \forall A \in \mathcal{A}_{0_0}$ first. The easy direction goes: $m^*(A) \leq m_o(A)$, since $\{A\}$ covers A.

For the hard direction, we'll show that if the collection A_i covers A, $\sum_{i=1}^{\infty} \mu_o(A_i) \leq \mu_o(A)$. Assume for contradiction that $\sum_{i=1}^{\infty} \mu_o(A_i) < m_o(A)$. Define $A_i' = A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j \right) \in \mathcal{A}_{0_0}$ (sidenote: this is in particular the same proof doesn't work with the inf definition)

By monotonicity: $\mu_o(A_i') \leq \mu_o(A_i)$ and $\sum \mu_o(A_i') \leq \sum \mu_o(A_i) < \mu_o(A)$, with the A_i''s being pairwise disjoint. Define $A_i'' := A_i' \cap A \implies \sum_{i=1}^{\infty} \mu_o(A_i'') \leq \mu_o(A)$. Now ''take a break'' and define $B_i = \bigcup_{j=1}^{i} A_j''$, so $\mu_o(B_i) = \sum_{j=1}^{i} \mu_o(A_j'')$ but we can see then that taking limits term-wise we have that this must indicate that $\mu_o(A) = \sum_{j=1}^{\infty} \mu_o(A''_j) < \mu_o(A)$ but bounding a strict inequality and and equals sign here gives as a contradiction. □

Observe that in the last proof we also made use of the lemma we proved from the homework that if $B_1 \subset B_2 \subset B_3 \cdots \subset$ is a nested sequence of sets then this implies that $\mu^*(B_i) \xrightarrow{i\to\infty} \mu^*\left(\bigcup_{i=1}^{\infty}\right)$.

We want to now show that: if $A \in \mathcal{A}_{0_0}, B \subset X \implies m^*(B) \geq m^*(B \cap A) + m^*(B \setminus A)$.

Proof. Let $\{A_i\}$ be a cover of B by elements of A_{0_0}, then $\{A_i \cap A\}$ covers $B \cap A$ and $\{A \cap (X \setminus A)\}$ covers $B \setminus A$. Consequently, $\sum \mu_o(A_i) = \sum \mu_o(A_i \cap A) + \sum m_o\big(A_i \cap (X \setminus A)\big)$ but then finally by the definition of outer measure we can notice that $m^*(B) \geq m^*(B \cap A) + m^*(B \setminus A)$. I.e., $\sum \mu_o(A_i \cap A) \geq m^*(B \cap A)$ and $\sum m_o\big(A_i \cap (X \setminus A)\big) \geq m^*(B \setminus A)$. □

Next time: if μ and ν are Borel measures on $X \times Y$, so is $\mu \otimes \nu$. Also, the Bernoulli measure. Eventually we will look at the the infinite product measure space of $(\{0,1\}, 2^{\{0,1\}}, \mu_{p,1-p})$.

Remark. You can build ∞-product measures.

3/4 - **More on product measure spaces, also intro the Bernoulli measure**

Recall let X be a metric space.

Definition 0.0.7. X is called *separable* if there exists a countable subset $A \subset X$ such that $\overline{A} = X$ (i.e., X has a countable dense subset).

Example 0.0.8. $\mathbb{R}$ is the base space, and $\mathbb{Q}$ is its countable dense subset.

$C^0([a, b])$...

Definition 0.0.9. If X and Y are metric spaces, define $d_{X\times Y}\big((x_1, y_1), (x_2, y_2)\big) := d_X(x_1, x_2) + d_Y(y_1, y_2)$. (Alternatively, $\sqrt{d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2}$)

Exercise (perhaps to self): show that the class of open sets and notions of convergence are the same for both.

In fact, there exists some C so that $\frac{1}{C}d(\cdot, \cdot) \leq d'(\cdot, \cdot) \leq Cd(\cdot, \cdot)$.

Lemma 0.0.10. *If X and Y are separable so is $X \times Y$.*

Proof. Let $A \subset X$ be dense and countable, and also let $B \subset Y$ be dense and countable. Then we claim that $A \times B := \{(a, b) : a \in A, b \in B\}$ is countable and dense in $X \times Y$.

To see that it is countable, recall that the product of 2 countable sets is countable.

To see that it is indeed dense, fix $(x, y) \in X \times Y$. Since $\overline{A} = X$, there exists a sequence (a_k) such that $a_k \to x \in X$ and likewise, because $\overline{B} = Y$, there exists (b_k), a sequence such that, $b_k \to y \in Y$. Now, we want to show that $(a_k, b_k) \to (x, y)$.

Indeed, it would suffice to show that $d\big((a_k, b_k), (x, y)\big) \to 0$. But by definition, $d\big((a_k, b_k), (x, y)\big) = d_X(a_k, x) + d_Y(b_k, y) \to 0$ and this implies that $\overline{A \times B} = X \times Y$. □

One could say that ‘‘separateness” passes to finite products (if this proof could be extended to such via induction).

Lemma 0.0.11. *If $U \subset X$ is open in X and $V \subset Y$ is open in Y, then $U \times V$ is open in $X \times Y$.*

Proof. We'll show that $\forall (x, y) \in U \times V$ there exists some $\varepsilon > 0$ such that $B_\varepsilon\big((x, y)\big) \subset U \times V$. Since U is open, there exists some $\varepsilon_1 > 0$ such that $d_X(x', x) < \varepsilon_1 \implies x' \in U$

and similarly we can say that there exists some $\varepsilon_2 > 0$ so that $d_Y(y, y') < \varepsilon_2 \implies y' \in V$.

Then if $d\big((x', y'), (x, y)\big) < \min\{\varepsilon_1, \varepsilon_2\} \implies (x', y') \in U \times V$. □

Theorem 0.0.12. *Let $(X, \mathcal{A}_{0_x}, \mu)$ and $(Y, \mathcal{A}_{0_y}, \nu)$ be a measure space, and X and Y also be separable metric spaces, and A_{0_x}, A_{0_Y} both contain the Borel σ-algebras of X and Y respectively. Then $\underbrace{A_{0_x} \otimes A_{0_y}}_{\sigma\text{-alg. for measurable sets w.r.t.} \mu\otimes\nu}$ contains the Borel σ-algebra on $X \times Y$.*

Proof. (Attempt). Let W be open in $X \times Y$. Let $\mathcal{A}_0 \subset X \times Y$ be countable and dense, and $A = \mathcal{A}_0 \cap W$.

For each $(x, y) \in W$, pick $U_x \times V_y$ such that $(x, y) \in U_x \times V_y \subset W$. For each (x, y) pick $(x_i, y_i) \in A \cap (U_x \cap V_y)$ ..... □

BERNOULLI MEASURE

Recall $\sum_d := \{(x_n)_{n\in\mathbb{Z}} : x_n \in \{1, \ldots, d\}, \forall n\} = \{1, \ldots, d\}^{\mathbb{Z}} =$ ‘‘functions from x into $\{1, \ldots, d\}$.”

Definition 0.0.13. *Cylinder sets.* If $w = w_{-n}, w_{-n+(n-1)}, \ldots w_0, w_1, \ldots w_{n-1}, w_n$, then we have the cylinder $C_w := \{(x_k) : x_k = w_k, \forall |k| \leq n\}$ are open.

Fix *probability vector* $p = (p_1, \ldots p_d)$ (i.e, $\sum p_i = 1, p_i \geq 0$). Let $\mathcal{A}_{0_0}$ be the algebra of finite unions of disjoint cylinders. Then we declare $\mu_{p,0}(C_w) = \prod_{k=-n}^{n} p_{w_k}$. (One example is heads in the 0,1 position, then a cylinder is $C_{011} \cup C_{111}$.)

Claim: this is a well defined pre-measure. $\mathcal{A}_{0_0}$ is an algebra. Indeed, see that $C_w \implies \Sigma_a \setminus C_w$ is a finite union of cylinders. Moreover, $C_w \cup C_v$ are finite unions of disjoint cylinders.

This makes us ask a question, how many cylinders are there specifying coordinates $-n$ through n? The answer is $2^{2n+1} < \infty$. Finally, see that $\Sigma_d \setminus C_w = \bigsqcup_{v\subset w} C_v$. (Set specifier for the big square cup was too hard for me to see).

For unions, consider if $v = (v_{-n}, \ldots, v_n)$ and $w = (w_{-m}, \ldots, w_m)$. Without loss of generalization, assume that $n \leq m$. Let $A(v) = \{\text{words}, (u_{-n}, \ldots, u_n\} : u_k = v_k, \forall k \leq n\}$.

Example 0.0.14. $v = 010 \implies A_2(v) = \{00100, 00101, 10100, 10101\}$. Notice that it's the letters ''starting after the first" where v is embedded in each word.

$\implies C_v = \bigsqcup_{u \in A_d(v)} C_u$. And we note that $A_m(v)$ has $d^{2(m-n)}$. This is because we have d choices of $m - n$ new coordinates. In fact, this shows any $A \subset \mathcal{A}_{0_0}$ is the disjoint union of cylinders of the same length. At length n, the cylinders each have measure $\prod_{i=-n}^{n} p_{w_i}$. Finally we want to note how this is related to the binomial distribution and such. See that $\left(\sum_{i=1}^{d} p_i\right)^{2n+1} = \sum_{(w_{-n},\dots w_n)} \prod p_{w_i}$, another reference point is that, we can view the following from the perspective of combining words and symbols: $(x + y)(x + y) = xx + xy + yx + yy$.

3/6 - **More on product and Bernoulli measure concerns**

Theorem 0.0.15. *If $\mathcal{A}_{0_X}, \mathcal{A}_{0_Y}$ both contain their Borel σ-algebras, so does $A_{0_X} \otimes A_{0_Y}$ (where we assume that X, Y are separable).*

Note that unless we are building metric spaces to suit some universal property it is safe to usually speak of them in general terms as being separable (?).

Recall X is separable $\iff$ X has a countable dense subset. $(A_{0_X} \otimes A_{0_Y})_0$ is said to be the pre-product algebra, or equivalently, the finite unions of product sets of forms $U \times V, U \in A_{0_X}, V \in A_{0_Y}$. Note also: $A_{0_X} \otimes A_{0_Y}$ contains countable unions of product sets.

Proof. It would suffice to show that every open $W \subset X \times Y$ is the union of countably many product boxes $W = \bigcup_i U_i \times V_i$ where U_i, V_i are open.

Let $A \subset X$ be dense and countable. Then we can write $A = \{x_1, x_2, \ldots\}$ and likewise letting $B \subset Y$ dense and countable we can write $B = \{y_1, y_2, \ldots\}$. For $(i, j, k, l) \in \mathbb{N}^4$ we let $U_{i,j,k,l} = B_{\frac{1}{k}}(x_i) \times B_{\frac{1}{l}}(y_i)$. Notice this is ‘‘still a box.” Now, for each open $W \subset X \times Y$, we let $S_w = \{(i, j, k, l) : U_{i,j,k,l} \subset W\}$. We claim that if we take a union, $\bigcup_{S_w} U_{i,j,k,l} = W$. The ‘‘forward” containment is easy to show, it follows by definition/construction. For the other direction, we fix $(x, y) \in W$, and note by openness that $\exists \varepsilon > 0$ such that $B_\varepsilon\big((x, y)\big) \subset W$. Pick $n \in \mathbb{N}$ so that $\frac{1}{n} < \frac{\varepsilon}{4}$. Since A is dense in X, $\exists i_0$ such that $d(x_{i_0}, x) < \frac{1}{n}$, and similarly, we can find a j_0 such that $d(y_{j_0}, y) < \frac{1}{n}$.

We claim now that $(i_0, j_0, n, n) \in S_w$ and additionally claim that $(x, y) \in U_{i_0,j_0,n,n}$. First we show that this first claim holds. To that end, we suppose that $(x', y') \in U_{i_0,j_0,n,n} \implies d\big((x, y), (x', y')\big) \leq d\big((x, y), (x_{i_0}, y_{j_0})\big) + d\big((x_{i_0}, y_{j_0}), (x', y')\big)$. Now we can apply definitions again, and remark that this is indeed an application of the triangle inequality to see that: $d\big((x, y), (x_{i_0}, y_{j_0})\big) + d\big((x_{i_0}, y_{j_0}), (x', y')\big) \leq d(x, x_{i_0}) + d(y, y_{i_0}) + d(x_{i_0}, x') + d(y_{i_0}, y')$. But notice that each of the four terms (With the latter to by the definition of $U_{i_0,j_0,n,n}$, the former two by construction) are all bounded by $\frac{1}{n}$. Indeed we then have that $d(x, x_{i_0}) + d(y, y_{i_0}) + d(x_{i_0}, x') + d(y_{i_0}, y') = \frac{4}{n} < 4 \cdot \frac{\varepsilon}{4} = \varepsilon \implies (x', y') \in B_\varepsilon\big((x, y)\big) \subset W$.

Now it remains to show that $(x, y) \in U_{i_0,j_0,n,n}$. But this essentially follows by definition since these two coordinates are $\frac{1}{n}$-close to x_{i_0} and y_{j_0}, respectively. So by definition/construction of our product set, this sort of ordered pair must also belong (?). ◻

Back to Bernoulli

Recall Cylinder sets: If given a finite word $w = w_{-n}, w_{-(n_1)}, \ldots w_0, w_1, \ldots w_{n-1}, w_n$, then we have a cylinder, $C_w = \{x \in \Sigma_d : x_k = w_k : \forall |k| \leq n\}$.

Example 0.0.16. That of Σ_d when $d = 2$.

$p_0^3 + p_0^2 p_1 + p_0^2 p_1 + p_0 p_1^2 + p_1 p_0^2 + p_0 p_1^2 + p_0 p_1^2 + p_1^3 = p_0^3 + 3p_o^2 p_1 + 3p_0 p_1^2 + p_1^3$.

Morals:

- The longer the word, the smaller (cardinality-wise) the cylinder.
- Cylinders can always be written as disjoint unions of longer cylinders.

Finally, to demonstrate a couple things, including the importance of being well-defined, we establish our new goal, to show that if $A = \bigsqcup_{i=n}^{n} C_{w_i}$ is a finite disjoint union of cylinders, $\mu(A) := \sum_{i=1}^{n} \prod_{j=}^{\ell(w)} p_{w_j}$. How might this fail well defined-ness? $A = \bigsqcup_{i=1}^{n} C_{w_i}$ and $A = \bigsqcup_{j=1}^{m} C_{v_j} \underbrace{\implies}_{?} \mu(A) = \mu(A)$.

Example 0.0.17. $\mu(C_0) = \mu(C_{000}) + \mu(C_{001}) + \mu(C_{100}) + \mu(C_{101})$

(Recall that by our convention the ‘‘middle” coordinate in the notation denotes the ‘‘parent-node” so-to-speak).

3/17 - **Bernoulli Measure Space**.

$\Sigma_d := \{(x_n)_{n\in\mathbb{Z}} : x_n \in \{0,\ldots,d-1\}\forall n\}$. For example, $\cdots 1010\underline{0}11101\cdots \in \Sigma_2$.

Algebra: Cylinder sets

If w is a *finite word*, that is, of the form, $w_{-n},\ldots,w_{n-1},w_n$, then $C_w = \{(x_n)_{n\in\mathbb{Z}} : x_i = w_i \quad \forall |i| \le n\} \subset \Sigma_d$. Let $\mathcal{A} = \{\text{finite disjoint unions of cylinders}\}$.

For a *premeasure*, we first fix $p = (p_0,\ldots,p_{d-1})$ such that $\sum_{i=0}^{d-1} p_i = 1$ and $p_i \ge 0$. Now we define $\mu_0(C_w) = \prod_{i=-n}^{n} P_{w_i}$. As usual, we define an outer measure by $\mu^*(C_w) = \inf\left\{\sum_{n=1}^{\infty}\mu_0(E_n) : E_n \in \mathcal{A} \text{ and } C_w \subset \bigcup_{i=1}^{\infty} E_n\right\}$. Notwithstanding that μ^* also is such that $\mu^*(\varnothing) = 0, A \subset B \implies \mu^*(A) \le \mu^*(B)$, and , in terse terms, countable subadditivity: $\mu^*(\bigcup A_i) = \sum \mu^*(A_i)$.

As usual we then extend using the Caratheodory criterion and obtain a measure μ, which we call the *Bernoulli measure*.

Example 0.0.18. Let $d = 2$, $p = (1,0)$. Now this implies that $\mu = \delta_{\bar{0}} = \delta_{000\ldots}$ (?), or in other words, that $\mu(C_w) = \begin{cases} 0 & \text{if } w \text{ contains } 1 \\ 1 & \text{if } w \text{ is all zeros} \end{cases}$

Example 0.0.19. Let $d = 2$ and $p = (1/2, 1/2)$, $A = \{$sequences such that every 0 is followed by a 1

Then $A = \bigcap\left(\bigcup_{\substack{\text{w admissible}\\ |w|=n}} C_w\right)$. Meaning that, for every n, we have that $x_w,\ldots,x_{-w}$ is admissible (?).

w is admissible if every 0 in w is followed by a 1.

Let $A_n = \bigcup_{\substack{\text{w admissible}\\ |w|=n}} C_w$. Notice that the right hand side is actually a disjoint union.

Now, $\mu(C_w) = \dfrac{1}{2^{2n+1}}$ and $\mu(A_n) = \dfrac{\#\text{ of admissible } w}{2^{2n+1}}$.

Let a_m = # of sequences $w_1,\ldots,w_m$ admissible with $w_m = 0$, and similarly, let b_m = # of sequences $w_1,\ldots,w_m$ admissible with $w_m = 1$. See that $a_1 = b_1 = 1$, and more generally there is a relation $a_{m+1} = b_m \iff a_m = b_{m-1}$ and $b_{m+1} = a_m + b_m$.

We may note then a familiar form, $b_{m+1} = b_{m-1} + b_m$ gives the *Fibonacci sequence*. So if, as we may know, $\phi = 1.68\ldots$, we have that $b_m \sim \phi^m, a_m \sim \phi^m$, and that $a_m + b_m \sim \phi^m$. Consequently, $\mu(A) \leq \mu(A_n) \approx \dfrac{\phi^{2n+1}}{2^{2n+1}} = \left(\dfrac{\phi}{2}\right)^{2n+1} \to 0$.

Example 0.0.20. Let $d = 2$ and $p = (p_1, p_2)$, $p_i > 0, \quad i = 1, 2$. Let $P = \{\text{sequences } (x_n) \text{ such that } \exists \ell \text{ where } x_n {+} \ell = x_n, \forall n\}$. Take for example, $\cdots 0101010101 \cdots$ or $\cdots 01000100010001 \cdots$.

Let $P = \displaystyle\bigcup_{\ell \in \mathbb{N}} \{(x_n) : x_n = w_{n \pmod{\ell}}, w_1, \ldots, w_\ell \text{ finite}\}$. But now observe, P is countable. And since $\mu\big(\{(x_n)\}\big) = \mu(Cx_{-n}, \ldots, x_n) = \displaystyle\prod_{i=-n}^{n} P_{x_i} \to 0 \leq (\max P_i)^{2n+1} \to 0$.

3/18 - **Integration with regard to Bernoulli**

Example 0.0.21. If μ is the $(\frac{1}{2}, \frac{1}{2})$-Bernoulli measure on Σ_2 and $f : \Sigma_2 \longrightarrow \mathbb{R}$ is defined by $f\big((x_n)\big) = \sum_{n=-\infty}^{\infty} a_n x_n$ for some non-negative sequence $(a_n) \in \mathbb{R}$ with $\sum a_n < \infty$, find $\int_{\Sigma_2} f \,\mathrm{d}\mu$. (E.g., $a_n = \frac{1}{2^n}, a_n = \frac{1}{n^2}$, etc...).

We fix $\ell \in \mathbb{N}$ (our intent here is that ℓ will represent the length of cylinders to partition the space). If w is a finite word on C_w with symmetric length ℓ, or in symbols, $w = w_{-(\ell-1)} \cdots w_0 \cdots w_\ell$ then see that $\sup\limits_{(x_n)\in C_w} f\big((x_n)\big) = \sum\limits_{n=-\ell}^{\ell} a_n w_n + \sum\limits_{|n|>\ell} a_n = f(\cdots |||w_{-\ell} \cdots w_0 \cdots w_\ell |||| \cdots)$. While, $\inf\limits_{(x_n)\in C_w} f\big((x_n)\big) = \sum\limits_{n=-\ell}^{\ell} a_n w_n$. It follows then that taking such a sum over the words of length ℓ, in the value of the infimum multiplied by the characteristic function of the cylinder set is less than or equal to f, which is itself less than or equal a sum over the words of length ℓ, in the value of the *supremum*; or in symbols: $\sum\limits_{|w|=\ell} \left(\sum\limits_{n=-\ell}^{\ell} a_n w_n\right) \chi_{C_n} \leq f \leq \sum\limits_{|w|=\ell} \left(\sum\limits_{n=-\ell}^{\ell} a_n w_n + \sum\limits_{|n|>\ell} a_n\right)$.

Now we must integrate, recall that for the Bernoulli measure we consider $2^{-(2\ell+1)}$. Consequently, we have that $\sum_{|w|=\ell} \left(\sum_{n=-\ell}^{\ell} a_n w_n\right) \cdot 2^{-(2\ell+1)} \leq \int f \,\mathrm{d}\mu \leq \sum_{|w|=\ell} \left(\sum_{n=-\ell}^{\ell} a_n w_n + \sum_{|n|>\ell} a\right.$ $2^{-(2\ell+1)}$. Where then $2^{-(2\ell+1)} \sum_{n=-\ell}^{\ell} \underbrace{\sum\limits_{|w|=\ell} a_n w_n}_{=2^{2\ell} a_n} \leq \int f \,\mathrm{d}\mu \leq \cdots \implies \frac{1}{2} \sum_{-\ell}^{\ell} a_n \leq \int f \,\mathrm{d}\mu \leq \left(\frac{1}{2} \sum_{-\ell}^{\ell} a_n\right) + \sum_{|n|>\ell} a_n$. And this implies that $\int f \,\mathrm{d}\mu = \frac{1}{2} \sum_{-\infty}^{\infty} a_n = \frac{1}{2} \sum_{-\ell}^{\ell} a_n$.

In general, when $(X, \mu, \mathcal{A}_0)$ is a measure space and $\mu(X) = 1$, then the space is a called a *probability space*.

A function $f : X \longrightarrow \mathbb{R}$ is called an *observable* or *random* variable.

- $\int f \,\mathrm{d}\mu$ is the *expected value* of f.
- $f_*\mu$, the push-forward of the measure μ, is the *distribution* of the random variable.
- $A \subset X$ is called an *event*.

- $\mu(A)$ is the *probability* of A.

Spaces of Measurable Functions

We first naively attempt to show that a certain specific space is a normed vector space. Let $L^1(X,\mu)$ denote the *set* of integrable functions in a measure space $(X,\mu,\mathscr{A}_0)$. Next, we define $\|f\|_{L^1} := \int_X |f|\,\mathrm{d}\mu$. What we want to show is that $\|\cdot\|_{L^1}$ is a norm on $L^1(X,\mu)$.

For the triangle inequality, see that $\|f+g\|_{L^1} = \int_X |f+g|\,\mathrm{d}\mu \le \int |f|+|g|\,\mathrm{d}\mu = \int_X |f|\,\mathrm{d}\mu + \int_X |g|\,\mathrm{d}\mu = \|f\|_{L^1} + \|g\|_{L^1}$. Next then for showing Homogeneity see that $\lambda \in \mathbb{R} \implies \|\lambda f\|_{L^1} = \int |\lambda f|\,\mathrm{d}\mu = |\lambda| \int |f|\,\mathrm{d}\mu = |\lambda| \cdot \|f\|_{L^1}$. While non-negativity passes by definition. But notice this is not quite a norm but instead a *semi-norm*. Indeed, when we consider the ‘‘zero-property” we see trouble in one direction, $\|f\|_{L^1} = 0 \not\Longrightarrow f = 0$. For instance, consider $f = \chi_{\{0\}}$ on $(\mathbb{R}, \text{Leb.})$. So let us try again.

Let $L^1(X,\mu)$ denote the set of *equivalence classes* of integrable functions in a measure space $(X,\mu,\mathscr{A}_0)$. We say that $f \sim g$ if $\exists A \subset X$ with $\mu(X \setminus A) = 0$ and $f_{|A} = g_{|A}$.

So our claim is that $\sim$ as defined above is an equivalence relation, that $L^1(X,\mu)$ has a well defined vector space structure (i.e., elements are equivalence classes closed under vector addition and scalar multiplication), that if f, g continuous and $f \sim g$, then $f = g$, and that $\|\cdot\|_{L^1}$ is well-defined on $L^1(X,\mu)$.

Example 0.0.22. $\sim$ is a transitive relation. Suppose that $f \sim g$ and $g \sim h$. Then $\exists A_1$ of full measure such that $f_{|A_1} = g_{|A_1}$ and $\exists A_2$ of full measure such that $g_{|A_2} = h_{|A_2}$. It follows that on $A_1 \cap A_2$ we have $f(x) = g(x) = h(x)(?) \implies f_{|A_1\cap A_2} = h_{|A_1\cap A_2}$. Furthermore, $\mu(X \setminus (A_1 \cap A_2) = \mu\big((X\setminus A_1) \cup (X \setminus A_2)\big) \le \mu(X\setminus A_1) + \mu(X \setminus A_2) = 0$.

We claim now that if $\|f\|_{L^1} = 0 \implies f \sim 0$. If $\int_X |f|\,\mathrm{d}\mu = 0$, then Chebyshev corollary, $|f = 0|$ a.e. $\implies f = 0$ a.e. $\implies f \sim 0$.

Example 0.0.23. Consider $L^1(\mathbb{R}, \delta_0)$. Where δ_0 is the Dirac δ-function with $x = 0$. See then that $f \sim g \iff f(0) = g(0) \implies \dim\big(L^1, \delta_0)\big) = 1$, a very interesting fact.

<u>Things to come</u>

- Define L^p-spaces.

- Prove L^p-spaces are complete [actually thus *Banach spaces*].
- Duality in L^p-spaces
- Fourier analysis.

3/20 - **L^p spaces & the triangle-inequality for L^p-norm(s) (using convexity)**.

Recall: $L^1(X,\mu)$ = space of *equivalence classes* of integrable functions $f : X \longrightarrow \mathbb{R}$.

$\|f\|_{L^1} = \int_X |f| \, d\mu$ is a well-defined *norm*.

The equivalence relation: $f \sim g \iff f = g$ a.e. $\iff \exists A \subset X$ such that $\mu(X \setminus A) = 0$ and $f_{|A} = g_{|A}$.

Lemma 0.0.24. *If $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ are continuous, μ is a Borel measure such that* $\operatorname{supp}(\mu) = \mathbb{R}$ *and $f \sim g$, then $f = g$.*

Proof. If $f \sim g$, then $\exists A \subset \mathbb{R}$ such that $\mu(\mathbb{R} \setminus A) = 0$ and $f_{|A} = g_{|A}$. We claim then that A is dense. Assume for contradiction then that A is not dense, i.e., $\overline{A} \neq \mathbb{R}$. Then this implies that $\exists x_0 \in \mathbb{R}$ such that $x_0 \notin \overline{A}$. Thus $\exists \varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \cap A = \emptyset$. By definition, since $x_0 \in \operatorname{supp}(\mu)$, $\mu\big((x_0 - \varepsilon, x_0 + \varepsilon)\big) > 0$. Since $(x_0 - \varepsilon, x_0 + \varepsilon) \subset \mathbb{R} \setminus A$, $\mu(\mathbb{R} \setminus A) \geq \mu\big((x_0 - \varepsilon, x_0 + \varepsilon)\big) > 0$. A contradiction. □

Lemma 0.0.25. *This actually a sublemma. If* $\operatorname{supp}(\mu) = X$ *and* $\mu(X \setminus A) = 0$, *then A is dense.*

We go on to now claim that if $f, g : X \longrightarrow \mathbb{R}$ are continuous, $A \subset X$ is dense, and $f_{|A} = g_{|A} \implies f = g$.

Proof. Let $x \in X$, Then there exists a sequence x_k in A such that $x_k \to x$ and so by continuity we have that $f(x_k) \to f(x)$, and likewise, $g(x_k) \to g(x)$. But because sequences in this context are unique, we note that $f(x_k) = g(x_k) \implies f(x) = g(x)$. □

L^p-norms.

Fix $p \geq 1$. Let $L^p(X,\mu)$ = space of equivalence classes of functions such that $\int_X |f|^p \, d\mu < \infty$. We claim then that $\|f\|_{L^p} = \left(\int_X |f|^p \, d\mu\right)^{1/p}$ is a norm. Indeed, it is easy to see that the function satisfies the homogeneity, non-negativity, and ‘‘zero bi-conditional” properties of a norm. What remains to show is whether the triangle-inequality holds here, and for this we need to introduce convexity and make use of convexity arguments.

Definition 0.0.26. $f : \mathbb{R} \longrightarrow \mathbb{R}$ is *convex* if $\forall a, b \in \mathbb{R}$ and $t \in [0,1]$, we have that $f\big(ta + (1-t)b\big) \leq tf(a) + (1-t)f(b)$.

One way to interpret this definition is to see it as meaning that $f\big(\text{weighted average}\big) \leq$ weighted average $\cdot f(x)$. Yet another is to say that the *graph* of f is below its ‘‘secant lines” everywhere.

Lemma 0.0.27. *If $f \in C^2$ and $f''(x) \geq 0$, then f is convex.*

Proof. Fix $a, b \in \mathbb{R}$. Set $g(t) = f\big(ta + (1-t)b\big) - tf(a) - (1-t)f(b)$. See then that this implies that $g'(x) = f'\big(ta + (1-t)b\big)(a-b) - f(a) + f(b) \implies g''(x) = f''\big(ta + (1-t)b\big)(a-b)^2 \geq 0$. Furthermore, $g(0) = f(b) - f(b) = 0$ and $g(1) = f(a) - f(a) = 0$. Since $g'' \geq 0$ we note that g' is increasing. By the mean value theorem then, $\exists x_0 \in (a,b)$ such that $g'(x_0) = 0$. This implies that g is *decreasing* on $[a, x_0]$ and that g is increasing on $[x_0, b]$. So $g \leq 0$. □

Corollary 0.0.28. *$h(x) = x^p$ is continuous on $[0, \infty)$ when $p \geq 1$.*

Proof. $h''(x) = p(p-1)x^{p-2}$. □

Finally then we show the triangle-inequality for the L^p norm (thus on L^p spaces).

Proof. Fix $f, g \in L^p(X, \mu)$. Then, $|f(x) + g(x)|^p = \left| t\dfrac{f(x)}{t} + (1-t)\dfrac{g(x)}{(1-t)} \right|^p \leq t\left|\dfrac{f(x)}{t}\right|^p + (1-t)\left|\dfrac{g(x)}{(1-t)}\right|^p = t^{1-p}|f(x)|^p + (1-t)^{1-p}|g(x)|^p$. Notice then if we integrate this last expression we have $\|f+g\|^p_{L^p} \leq t^{1-p}\|f\|^p_{L^p} + (1-t)^{1-p}\|g\|^p_{L^p} = (**)$. Now set $t := \dfrac{||f||_{L^p}}{||f||_{L^p} + ||g||_{L^p}} \leq 1$ and note that this implies $1 - t = \dfrac{||g||_{L^p}}{||f||_{L^p} + ||g||_{L^p}}$. So then, $(**) = \dfrac{||f||^{1-p}_{L^p}}{(||f||_{L^p} + ||g||_{L^p})^{1-p}}\|f\|^p_{L^p} + \dfrac{||g||^{1-p}_{L^p}}{(||f||_{L^p} + ||g||_{L^p})^{1-p}}\|g\|^p_{L^p} = \dfrac{(\|f\|_{L^p} + \|g\|_{L^p})^1}{(\|f\|_{L^p} + \|g\|_{L^p})^{1-p}} = (\|f\|_{L^p} + \|g\|_{L^p})^p$. So we conclude that $\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$. □

Theorem 0.0.29. *(**Jensen’s inequality**). Let $(X, \mu, \mathcal{A}_0)$ be a probability space, $g : X \longrightarrow \mathbb{R}$ be measurable, and $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be C^2 and convex. It follows that $\varphi\left(\int_x g\, d\mu\right) \leq \int_X \varphi\left(g(x)\right)\, d\mu$.*

3/24 - **$\mathbf{L^p}$-spaces are complete, Typewriter sequence**.

Theorem 0.0.30. *If $(X, \mu, \mathcal{A}_0)$ is a measure space, then $L^p(X, \mu)$ is complete.*

Proof. We will show that if (f_n) is Cauchy in L^p, then $f_n \to f$ for some $f \in L^p$ (i.e., $\|f_n - f\|_{L^p} \to 0$ Throughout we note that $\|\cdot\|$ denotes $\|\cdot\|_{L^p}$. For step 1, note without loss of generalization we may assume $\|f_n - f_{n+1}\| < \frac{1}{4^n}$. Indeed, for each n, choose $m_n \geq m_{n-1}$ such that if $k, \ell \geq m_n$, then $\|f_k - f_\ell\| < \frac{1}{4^n}$. Then, if $g_n = f_{m_n}, g_m$ satisfies $\|g_{n+1} - g_n\| = \|f_{m_{n+1} - f_{m_n}}\| < \frac{1}{4^n}$ for some f.

Now we claim that $f_n \to f$ in L^p as well. Indeed, let $\varepsilon > 0$. Choose n_1 such that $n \geq n_1 \implies \|g_n - f\| < \frac{\varepsilon}{2}$. Choose n_2 such that $\frac{1}{4^{n_2}} < \frac{\varepsilon}{2}$. Then if $N = \max\{n_1, n_2\}$ and if $k \geq m_N$, $\|f_k - f\| \leq \|f_k - g_{m_n}\| + \|g_{m_N} - f\| < \frac{1}{4^{m_n}} + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. '

For step 2 we set up the use of Borel-Cantelli. We assume the ''rapid-convergence Cauchy property" holds. That is, $\|f_n - f_{n+1}\| < \frac{1}{4^n}$, so $\int |f_n - f_{n+1}|^p \, \mathrm{d}\mu < \dfrac{1}{4^{np}}$. By Chebyshev's inequality, $\mu\left(\left\{x : |f_n(x) - f_{n+1}(x)|^p \geq \frac{1}{2^{np}}\right\}\right) \leq 2^{np} \int |f_n - f_{n+1}|^p \, \mathrm{d}\mu \implies \mu\left(\left\{x : |f_n(x) - f_{n+1}(x)|^p \geq \dfrac{1}{2^n}\right\}\right) < 2^{np} \dfrac{1}{4^{np}} = \dfrac{1}{2^{np}}$. Since $\sum \frac{1}{2^{np}} < \infty, \exists A \subset X$ of full measure such that $\forall x \in A, |f_k(x) - f_{k+1}(x)| < \frac{1}{2^k}$, for all but finitely many k's $\iff \forall x \in A\ \exists k(x)$ such that if $k \geq k(x), |f_k(x) - f_{k+1}(x)| < \frac{1}{2^k}$.

Next, for step 3 we will construct a candidate for f. Fix $x \in A$. Then if (WLOG) $m > n \geq k(x)$ we take a look at what will be a telescoping sum, $|f_n(x) - f_m(x)| = \left|\sum_{j=m}^{n-1} f_{j+1}(x) - f_j(x)\right| \leq \sum_{j=m}^{n} |f_{j+1}(x) - f_j(x)| < \sum_{j=m}^{n} 2^{-j} < \sum_{j=m}^{\infty} 2^{-j} \xrightarrow{m \to \infty} 0 \implies \big(f(x)\big)$ is Cauchy, and therefore $f_n(x) \to f(x)$ for some $f(x)$. We will now apply Fatou's lemma for step 4. Because $|f_n - f_m|^p \xrightarrow{n \to \infty} |f - f_m|^p$ a.e., by Fatou, $(**) = \lim_{n \to \infty} \int |f_n - f_m|^p \, \mathrm{d}\mu \geq \int |f - f_m|^p \, \mathrm{d}\mu$. In fact, $(**) = \lim_{n \to \infty} \|f_n - f_m\|^p \leq \left(\lim_{n \to \infty} \sum_{j=m}^{n-1} \|f_{j+1} - f_j\|\right)^p < \left(\sum_{j=1}^{\infty} 2^{-j}\right)^p \implies \|f - f_m\| < \sum_{j=m}^{\infty} 2^{-j} \to 0.$ $\square$

Corollary 0.0.31. *If (f_n) is Cauchy in L^p, it has a subsequence which converges a.e.*

Note that L^p convergence $\not\Longrightarrow$ a.e. convergence automatically. That is, we require the rate of convergence to be "fast enough" (summable).

Example 0.0.32. Typewriter Sequence

Suppose we have a sequence of functions that ‘‘goes back and forth like a typewriter” i.e., on the unit interval halves its upper bound ‘‘each iteration” and then bounces back and forth the ‘‘current” 0

two bounded halves: $f_1 = \chi_{[0,1/2]}, f_2 = \chi_{[1/2,1]}, f_3 = \chi_{[0,1/4]}, f_4 = \chi_{[1/4,1/2]}, \ldots$. Notice that $\left(f_n(x)\right)$ never converges for a sequence of real numbers! But $\displaystyle\int f_n = \frac{1}{2^{\log_2 n}} \sim \frac{1}{n} \implies \|f_n\|_{L^p} = \frac{1}{n^{1/p}} \to 0$. That is, we have found a failure of Borel-Cantelli. (Noting the issue is that we need to pass to a subsequence and verify summability).

In the coming material we will consider a candidate for a *bounded functional* on L^p spaces. If $g \in L^q(X, \mu)$ define $\phi_g : L^p \longrightarrow \mathbb{R} \quad \phi_g(f) = \displaystyle\int_x f g \, \mathrm{d}\mu$.

Theorem 0.0.33. *ϕ_g is well-defined for $g \in L^q$ for specific q, we call it the ‘‘polar conjugate of q” and every bounded linear functional on L^p is of this form, $\left(L^p(X,\mu)\right)^{\phi_g} \cong L^q(X,\mu) \quad \left(L^p(X,\mu)\right)^{\phi_g} \cong L^p(X,\mu)$.*

(?) (?) Check the above, sloppily handwritten from me and not heard well from me.

3/25 - **Conjugates, Inequalities, and Hilbert Spaces: end of "essentials."**

Recall: $\|f\|_{L^p} = \left(\int_x |f|^p\right)^{1/p}$ is a norm, $L^p(X,\mu) := \{f \text{ measurable}, \|f\|_{L^p}\} < \infty\}$, and the theorem that $L^p(X,\mu)$ is complete.

Definition 0.0.34. $p, q > 1$ are conjugate if $\frac{1}{p} + \frac{1}{q} = 1$ ($\iff q = \frac{p}{p-1} \iff q(p-1) = p \iff p(q-1) = q$).

Note: p is said to be self-conjugate $\iff \frac{1}{p} + \frac{1}{p} = 1 \iff \frac{2}{p} = 1 \iff p = 2$.

Young's Inequality let p, q be conjugate and $a, b > 0$ then, $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. Where equality holds $\iff a^p = b^q$.

Proof. Let $g(x) - \frac{1}{p}x^p + \frac{1}{q} - x \implies g'(x) = x^{p-1}$, so $g'(x) > 0$ when $x > 1$ and $g'(x) < 0$ when $0 < x < 1$. Furthermore, $g(1) = \frac{1}{p} + \frac{1}{q} - 1 = 0$. In fact, $g \geq 0$ for all positive inputs.

Plug in $x = \frac{a}{b^{q-1}}$. See then that this implies $\frac{1}{p}\left(\frac{a}{b^{q-1}}\right)^p + \frac{1}{q} - \frac{a}{b^{q-1}} \geq 0$. Equivalently, $\frac{1}{p}\frac{a^p}{b^{q+p-1}} + \frac{1}{q} \geq \frac{a}{b^{q-1}}$, and since $p, q > 1$ we can deduce that $\frac{1}{p}\frac{a^p}{b^q} + \frac{1}{q} \geq \frac{a}{b^{q-1}} \implies \frac{1}{p}a^p + \frac{1}{q}b^q \geq ab$.

For the case then of the equality $\frac{a}{b^{q-1}} = 1 \iff a = b^{q-1} \iff a^p = b^{p(q-1)} = b^q$. □

We make brief note that for the above to hold there is a case of $p = 1 \implies q = \infty$.

Hölder's Inequality let p, q be conjugate. If $f \in L^p(X,\mu)$, then $\int_X |fg|\, d\mu \leq \|f\|_{L^p}\|g\|_{L^q}$.

Note: if $p = q = 2$ then this is what is also referred to as the *Cauchy-Schwartz inequality*, $|\langle x, y\rangle| \leq \|x\|\|y\|$, which holds in any *inner product space*.

Proof. Divide both sides by $\|f\|_{L^p}\|g\|_{L^q}$, we get $\int_X \left|\frac{f}{\|f\|_{L^p}}\frac{g}{\|f\|_{L^q}}\right\| d\mu \leq 1$. Then, we may assume without loss of generalization (because $\int \frac{F}{\|F\|_{L^p}}\frac{G}{\|G\|_{L^q}} \leq 1 \implies \int |FG| \leq \|F\|_{L^p}\|G\|_{L^q}$) that $\|f\|_{L^p} = \|g\|_{L^q} = 1$. It follows that $\int_X |f|\,|g|\, d\mu \leq \int \frac{|f|^p}{p} + \frac{|g|^q}{q}\, d\mu = \frac{1}{p}\int |f|^p\, d\mu + \frac{1}{q}\int |g|^q\, d\mu = \frac{1}{p} + \frac{1}{q} = 1$. □

Definition 0.0.35. If $g \in L^q(X,\mu)$, define $\phi_g : L^p(X,\mu) \longrightarrow \mathbb{R}$ by $\phi_g(f) = \int_X fg \, \mathrm{d}\mu$.

We want to explore the prospect that this means then that ϕ_g is linear and bounded. Indeed, since $\frac{|\phi_g(f)|}{\|f\|_p} \leq \frac{\int_X |fg| \, \mathrm{d}\mu}{\|f\|_p}$

Theorem 0.0.36. *(Riesz Representation Theorem). If ϕ is a bounded linear functional on $L^p(X,\mu)$, $\exists! g \in L^q(X,\mu)$ such that $\phi = \phi_g$.*

....

If $p = q = 2$, then $L^2(X,\mu)$ is an example of a *Hilbert space*. Additionally, every Hilbert space is an example of a Banach space.

Definition 0.0.37. $\mathcal{H}$ is a Hilbert space if it is a vector space with an inner product $\langle \cdot, \cdot \rangle$ such that:

(1) $\langle \cdot, \cdot \rangle$ is *bilinear* and symmetric

$$\langle f_1 + f_2, g \rangle = \langle f_1, g \rangle + \langle f_2, g \rangle$$
$$\langle f, g_1 + g_2 \rangle = \langle f, g_1 \rangle + \langle f, g_2 \rangle$$
$$\langle \lambda f, g \rangle = \lambda \langle f, g \rangle \text{ and } \langle f, \lambda g \rangle = \bar{\lambda} \langle f, g \rangle$$
$$\langle f, g \rangle = \overline{\langle g, f \rangle}$$

(2) If $\|f\| = \sqrt{\langle f, f \rangle}$, then $\mathcal{H}$ is complete with regard to this norm.

For $L^2(X,\mu)$ we set $\langle f, g \rangle = \int_X f\bar{g} \, \mathrm{d}\mu \implies \|f\| = \sqrt{\langle f, f \rangle} = \left(\int_X f\bar{f} \, \mathrm{d}\mu \right)^{1/2} = \left(\int_X |f|^2 \, \mathrm{d}\mu \right)^{1/2} = \|f\|_{L^2}$.

Finally, we note a quick example: $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is linear, $\phi(v_1, \ldots, v_n) = a_1 v_1 + \ldots + a_n v_n = a \cdot v$. With that, we mark this the end of the ‘‘essential material.”

3/27 - **Intro to (complete) orthonormal sets in Hilbert spaces.**

Plan:

- Hilbert Spaces and Fourier Analysis
- Ergodic theory and dynamics on circles.
- and if there is time, information theory and entropy.

Recall: $\mathcal{H}$ is a Hilbert space if it is a vector space over the field of complex numbers with a complete inner product, $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C}$, i.e.,

- $f \mapsto \langle f, g \rangle$ and $f \mapsto \langle g, f \rangle$ are linear.
- $\langle g, f \rangle = \overline{\langle f, g \rangle}$
- $\langle f, f \rangle \geq 0$ and $\langle f, f \rangle = 0 \iff f = 0$ (what we call *positive definite*.

Relations in finite dimensions If $\mathcal{H} \cong \mathbb{C}^n$, every bilinear form is given by $\langle v, w \rangle = v^T A \bar{w}$ for A a non-matrix object.

- Symmetry of $\langle \cdot, \cdot \rangle \iff$ symmetry of A. That is, $A = \bar{A}^T$.
- Positive definite of $\langle \cdot, \cdot \rangle \iff$ positive definite of A.

A subset $\{U_n\} \subset \mathcal{H}$ is *orthonormal* if $\langle \varphi_m, \varphi_n \rangle = \delta_{mn} = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases}$. We say that $\{U_n\}$ is a *complete orthonormal* set $\iff \{U_n\}$ is orthonormal and $\langle \psi, \varphi_n \rangle = 0 \iff \psi = 0$. In other words we can use the orthonormal set to check if an element of the Hilbert space is 0.

Definition 0.0.38. Let $\ell^2 := L^2(\mathbb{Z}, \gamma)$ where γ is the counting measure, i.e., $\gamma(A) =$ #of elements in A. Let $x, y \in \ell^2 \implies x, y : \mathbb{Z} \longrightarrow \mathbb{C} \implies x = (x_n), y = (y_n)$ with values in $\mathbb{C}$. That is, $x \in \ell^2 \iff \sum_{-\infty}^{\infty} |x_n|^2 < \infty$ and we note that we can also write that this sum is $\int_{\mathbb{Z}} \left|x^2\right| d\gamma$ (?). Now we have that $\langle x, y \rangle = \sum_{-\infty}^{\infty} x_n \bar{y_n}$.

A complete orthonormal set $\delta^n : \mathbb{Z} \longrightarrow \mathbb{C}$, given by $\delta^n(k) = \begin{cases} 0 & \text{if } k \neq n \\ 1 & \text{if } k = n \end{cases}$, $\langle \delta^n, \delta^m \rangle =$ $\sum_{k=-\infty}^{\infty} \delta(k)^n \bar{\delta^m}(k) = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases}$. We claim then that this set $\{\delta^n\}_{n \in \mathbb{Z}}$ is complete, that is, $\langle (x_k), \delta^n \rangle = \sum_{k=-\infty}^{\infty} x_k \bar{\delta^n}(k) = x_n$. We check that if $\langle x_k, \delta^n \rangle = x_n = 0, \forall n \implies (x_k) = 0$.

(Briefly we say as a digressive aside here that if $p = 2$ then by Hölders inequality in this $p = 2$ space we have orthogonality.)

Definition 0.0.39. Let X be a circle with circumference 1 and let m be the arc-length measure. Think of this as $[0, 1]$ with end-points glued together. Now, we say that functions on $[0, 1] \leftrightarrow$ 1-periodic functions on $\mathbb{R}$.

This motivates $L^2\big([0, 1], m\big)$ being considered as 1-periodic on $\mathbb{R}$, or in other words, a set of functions on a circle.

Recall, $e^{ix} = \cos(x) + i\sin(x), \quad \varphi_n(x) := e^{2\pi inx}$ form a complete orthonormal set for $L^2\big([0, 1], m\big)$

[An idea: group action from a fundamental group?]

Orthonormality, see that $\langle \varphi_n, \varphi_m \rangle - \displaystyle\int_{[0,1]} e^{2\pi ix} e^{-2\pi ix}\, \mathrm{d}x = \int_0^1 e^{2\pi i(n-m)x}\, \mathrm{d}x,\ n \in \mathbb{Z}.$

Now we have two cases. First, if $n = m$ then this integral is equal to $\int_0^1 1dx = 1$, otherwise if $n \neq m$ we have that this equals $\frac{1}{2\pi i(n-m)} e^{2\pi i(n-m)x}\big|_0^1 = \frac{1}{2\pi i(n-m)}(1-1) = 0$.

[Here we made not that showing completeness is harder and will be done later. Quadratic decay was also mentioned.]

Theorem 0.0.40. ***(Perseval's Theorem)*** *If $\{\varphi_n\}$ is an orthonormal set, then $\forall \psi \in \mathcal{H}, \quad \psi = \sum_n \langle \psi, \varphi_n \rangle \varphi_n$.*

We note that the bounds of this sum can be $(-\infty, \infty)$. Compare then in $\mathbb{C}^n$, there, $v = \sum_{i=1}^n (v \cdot e_i) e_i = \sum_{i=1}^n v_i e_i = (v_1, v_2, \ldots, v_n)$.

3/31 - **Bessel's inequality and Perseval's theorem**.

Recall, $\mathcal{H}$ is a symbols stands for a *Hilbert space*, which is a vector space with a complete inner-product. $\{U_n\}$ is orthonormal if $\langle \varphi_n, \varphi_m \rangle = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases} = (*)$. We say that $\{U_n\}$ is *complete orthonormal* if $(*)$ holds and $\langle \psi, \varphi_n \rangle = 0 \forall n \iff \psi = 0$.

Theorem 0.0.41. ***(Bessel's inequality)****. If* $\{\varphi_n\}$ *is orthonormal, then* $\|\psi\| \geq \left(\sum |\langle \psi, \varphi_n \rangle|^2\right)^{1/2}$.

Recall also that $\|\psi\| = (\langle \psi, \psi \rangle)^{1/2}$

Proof. We'll show that $\langle \psi, \psi \rangle \geq \sum_n |\langle \psi, \varphi_n \rangle|^2$. Consider the case where $\{\varphi_n\}$ has finitely many members, $0 \leq \|\psi - \sum_n \langle \psi, \varphi_n \rangle \varphi_n\|^2 - \langle \psi - \sum_n \langle \psi, \varphi_n \rangle \varphi_n, \psi - \sum_m \langle \psi, \varphi_m \rangle \varphi_m \rangle$
$= \langle \psi, \psi \rangle - \sum_m \langle \psi, \varphi_n \rangle \langle \varphi_n, \psi \rangle - \sum_m \overline{\langle \psi, \varphi_m \rangle} \langle \psi, \varphi_m \rangle + \sum \langle \psi, \varphi_n \rangle \overline{\langle \varphi_m, \varphi_n \rangle} \langle \varphi_m, \varphi_n \rangle$
$= \|\psi\|^2 - \sum_m |\langle \psi, \varphi_m \rangle|^2 - \sum_n |\langle \psi, \varphi_n \rangle|^2 + \sum_n |\langle \psi, \varphi_n \rangle|^2$.
So, the upshot is then that $\implies \|\psi\|^2 - \sum_n |\langle \psi, \varphi_m \rangle|^2 \geq 0 \implies \|\psi\|^2 \geq \sum_n |\langle \psi, \varphi_m \rangle|^2 \implies \|\psi\| \geq \sqrt{\sum |\langle \psi, \varphi_m \rangle|^2}$, what was needed to be shown.

[Prof. mentioned a complete and orthonormal set of polynomials with applications. I do not want to be a smart-alek so I did not say it out loud but I believe he was referring to the *Legendre* polynomials as one might recall from a numerical analysis class].

We now claim that the infinite case follows from the finite case. That is, for the infinite case we note that since we have already established that $\|\psi\|^2 \geq$ all partial sums, then by the monotone convergence theorem, we can infer that $\sum_n |\langle \psi, \varphi_n \rangle|^2 \leq \|\psi\|^2 < \infty$. □

Briefly we recall the monotone convergence theorem for sequences. From Wikipedia: If $(a_n)_{n \in \mathbb{N}}$ is a monotone sequence of real numbers (meaning that, either $a_n \geq a_{n+1}$ or $a_n \leq a_{n+1}, \forall n$), then the following are equivalent: (a_n) has a limit in $\mathbb{R}$ such that $\lim_{n \to \infty} a_n < \infty$ and there exists some constant $B > 0$ such that $|a_n| \leq B, \forall n$. In other words, (a_n) is bounded. Moreover, if (a_n) is nondecreasing then $\lim_{n \to \infty} a_n = \sup_n a_n$ and if (a_n) is nonincreasing then $\lim_{n \to \infty} a_n = \inf_n a_n$.

Corollary 0.0.42. *If* $\{\varphi_n\}$ *is orthonormal,* $\sum_n |\langle \psi, \varphi_n \rangle|^2 < \infty, \ \forall \psi \underbrace{\implies}_{H.W.} \sum_n \langle \psi, \varphi_n \rangle \varphi_n \in \mathcal{H}$.

Theorem 0.0.43. ***(Parseval)*** *If $\{\varphi_n\}$ is complete orthonormal and $\psi \in \mathcal{H}$, then $\psi = \sum_n \langle \psi, \varphi_n \rangle \varphi_n$.*

Proof. Recall that $v = 0 \iff \langle v, \varphi_m \rangle = 0,\ \forall n$ (by completeness). Thus, it would suffice to show that $\langle \psi - \sum_n \langle \psi, \varphi_n \rangle \varphi_n, \varphi_n \rangle = 0,\ \forall n$. But, $\langle \psi - \sum \langle \psi, \varphi_n \rangle \varphi_n, \varphi_n \rangle = \langle \psi, \varphi_n \rangle - \sum_n \langle \psi, \varphi_n \rangle \langle \varphi_n, \varphi_n \rangle$
$= \langle \psi, \varphi_n \rangle - \langle \psi, \varphi_n \rangle = 0.$ □

It may be instructive to remark that completeness again here does not mean the same thing as Cauchy implying convergence, so-to-speak.

Corollary 0.0.44. $\|\psi\| = \left(\sum_n |\langle \psi, \varphi_n \rangle|^2\right)^{1/2}$ *if $\{\varphi_n\}$ is a complete orthonormal.*

Next, we will sketch that $\varphi_n(x) = e^{2\pi i n x},\ n \in \mathbb{Z}$ is a complete orthonormal (basis?), $x \in L^2\big([0,1], m\big)$.

> Step 1. This is sometimes called the *Stone-Weierstrass theorem*. What we would want to do in step is show finite linear combinations of φ_n are dense in $C^0\big([0,1]\big)$.
>
> Step 2. We show that any C^0-dense subset of $C^0\big([0,1]\big)$ is also dense in $L^2\big([0,1], m\big)$.
>
> Step 3. $\{\varphi_n\}$ is complete $\iff$ {finite linear combinations of φ_n} is dense in $\mathcal{H}$ (and we recall that here we are fixated on $L^2\big([0,1], m\big)$).

We will finish up today with proving one direction of Step 3.

Proof. We will show that the reverse implication holds. Suppose that for all φ and that $\forall \varepsilon > 0$, there exists constants, $(c_n)_{n=1}^{m_n}$, so that $\left\|\sum_{n=1}^{m_n} c_n \varphi_n - \psi\right\| < \varepsilon$. We will show that $\langle \psi, \varphi_n \rangle = 0,\ \forall n \implies \psi = 0 \implies \left\|\psi - \sum_{n=1}^{m_n} c_n \varphi_n\right\| = \|\psi\|^2 + \sum_{n=1}^{m_n} |c_n|^2 < \varepsilon^2$, and this holds for all ε. But then this implies that $\|\psi\|^2 < \varepsilon^2 \implies \|\psi\| = 0 \implies \psi = 0$. □

We end today by noting that ‘‘this is a good place to look when we need a subspace of a Hilbert space which is closed.”

4/1- $\{\sum \mathbf{a_k e^{2\pi i k x}}\}$ is complete orthonormal, Ergodic Theory

[The final is Wednesday April 29, 10:30-12:30]

Step 1 Show that $\left\{\sum a_k e^{2\pi i k x}\right\}$ is dense in $C^0\big([0,1]\big)$, with regard to $\|\cdot\|_{C^0}$.

Step 2 Show that if $\mathbb{A} \subset C^0\big([0,1]\big)$ is dense, then it is dense in $L^2\big([0,1], m\big)$.

Step 3 Show that $\left\{\sum a_k e^{2\pi i k x}\right\}$ is complete orthonormal $\iff$ $(\left\{\sum a_k e^{2\pi i k x}\right\}$ is dense in L^2,

$$\langle \varphi_n, \varphi_m \rangle = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases}$$

Let X be a compact metric space. An *algebra of functions* in $C^0(X)$ is a subset $\mathbb{A} \subset C^0(X)$ such that

- $\mathbb{A}$ is closed under additions, scalar multiplication [$\mathbb{A}$ is a vector space]
- $\mathbb{A}$ is closed under multiplication.

Example 0.0.45. $\mathbb{A} = \{\text{polynomials on } [0,1]\}$

$\mathbb{A} = \left\{\sum_{k=-n}^{n} a_k e^{2\pi i k x}\right\}$

Indeed see for this last example that, $e^{2\pi i k x} \cdot e^{2\pi i \ell x} = e^{2\pi i (k+\ell) x}$, $(e^{2\pi x})^L$

$\mathbb{A} = C^0(X)$

$\mathbb{A} = \left\{\sum_{i=1}^{n} a_i \chi_{C_{w_i}} : C_{w_i} \subset \Sigma_d \text{ are cylinders}\right\}$

An algebra $\mathbb{A}$ is said to *separate points* if $\forall x \neq y \in X, \exists f \in \mathbb{A}$ such that $f(x) \neq g(x)$. In other words we can somewhat identify distinct points using the functions from $\mathbb{A}$.

Theorem 0.0.46. *(Stone-Weierstrass) Let $\mathbb{A}$ contain constant functions. Then* $\operatorname{cl}_{C^0} \mathbb{A} = C^0(X) \iff$ *A separates points.*

[Here we briefly note that the the summation bounds for $\left\{\sum_{k=-n}^{n} a_k e^{2\pi i k x}\right\}$ have $-n$ on the bottom index due to considerations having to do with the real and imaginary parts of complex numbers. That is, there is a technical ‘‘reason” as to why this is the case.]

For now, this is the end of Step 1. So next we check with Step 2.

If $S \subset C^0(X)$ is such that $\operatorname{cl}_{C^0} S = C^0(X) \implies \operatorname{cl}_{L^2} S = L^2\big([0,1], m\big)$. In fact, this is true generally for such an L^2 space though we specialize here to the $[0,1]$ case.

The next big idea is to take $f \in L^2(X, m)$ and find a simple function such that it approximates f well. If so then we can apply the simple approximation theorem and then have that $\exists \sum_{i=1}^{n} a_i \chi_{A_i}$ such that $\left\| f - \sum_{i=1}^{n} a_i \chi_{A_i} \right\|_{L^2} < \varepsilon$. Notice we use the fact that we can get arbitrarily close to an integral assuming the measurability of functions (?).

Each A_i is measurable, so there exists compact K_i such that $K_i \subset A_i$ and $m(A_i \setminus K_i) < \varepsilon$, and there also exists (compact ?) open U_i such that $A_i \subset U_i$ and $m(U_i \setminus A_i) < \varepsilon$.

Lemma 0.0.47. *If U is open and $K \subset U$ is compact, there exists a function $\varphi : X \longrightarrow \mathbb{R}$ such that*

- $\varphi(x) = 1 \quad \forall x \in K$
- $\varphi(x) = 0 \quad \forall x \notin U$
- $0 \leq \varphi(x) \leq 1 \quad \forall x \in X$

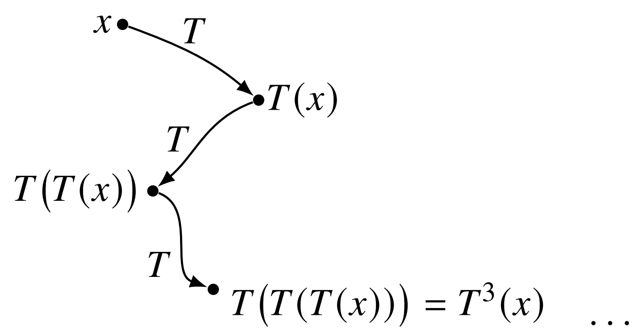
Example 0.0.48. (**Bump functions**) Consider where $K = [a, b], U = (a - \varepsilon, a + \varepsilon$. We can draw a number line encoding the end points of these sets. We note a roughly ‘‘trapezoidal” outline, except where it is like we have chopped the bottom of the trapezoid away and instead extended it’s resulting ‘‘legs” infinitely in the direction they were ‘‘heading anyway,” just flat-line wise.

We note the remarkable fact that we can always ‘‘make” bump functions smooth, that is they always exist in C^∞ but they are *not analytical*. Indeed, consider a ‘‘transition point” such as the left end-point of the interval U, that is, $a - \varepsilon$. If the bump function defined using U was analytical, then the Taylor series about $a - \varepsilon$ (which for brevity we will let ξ denote) *‘‘reconstructs” the function in a neighborhood*, that is a local point (within what would be the radius of convergence) could be written $\varphi(x) = \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(\xi)}{n!}(x - \xi)^n$. But at our ξ, we see that while φ is identically 0 on the left hand side, that is not the case on the right hand side. Given that the bump function is smooth however, it has derivatives of all orders on ξ which are constantly 0. One can observe that however, every neighborhood of ξ includes points that are not 0, so φ cannot be analytic at this point, since its Taylor polynomial being constantly 0 on all orders cannot coincide with non-zero values of φ.

Thus a bump function serves as an example that it is possible for a function to coincide on all orders of its Taylor polynomial, centered at a specific point, and yet fail to agree on any other point with that particular Taylor polynomial.

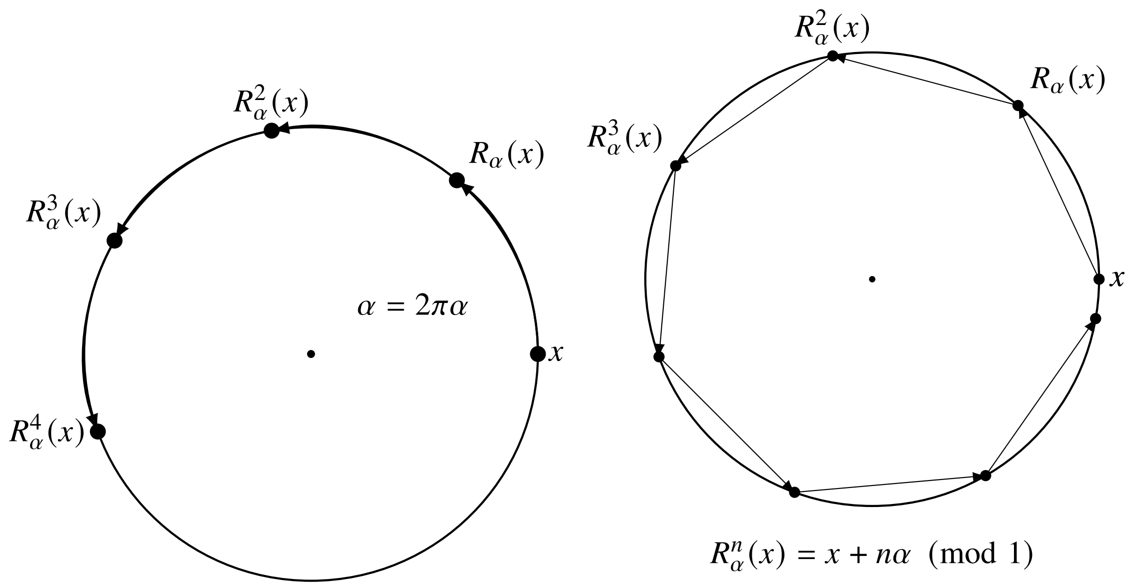
Ergodic Theory

Let $(X, \mathcal{A}_0, \mu)$ be a probability space. We think of the map $T : X \longrightarrow X$ as a *time-evolution*.



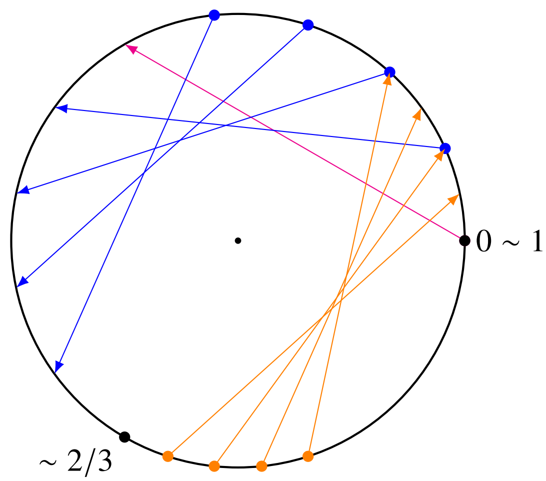
T is a called *measure preserving* if for all measurable $A \subset X$, we have both that $\mu(\{x \in X : T(x) \in A\}) = \mu(A)$ and $\mu(T^{-1}(A)) = \mu(A)$ are satisfied. We also note here that ''time'' can be take to be the *orbit of* x (?), that is, $\{x, T(x), T^2(x), T^3(x), \dots\}$. Finally we note that we are not assuming that each iteration is independent but that they do have the same probability.

Example 0.0.49. (*rotations*)

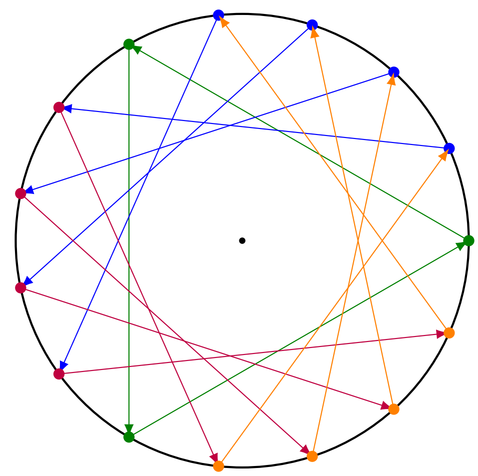


In [0, 1] coordinates this transformed becomes $R_\alpha(x) = \begin{cases} x+\alpha & x \in [0, 1-\alpha) \\ x+\alpha-1 & x \in [1-\alpha, 1] \end{cases}$.

We can illustrate rotations by $2/3$ of the angle 2π.



A different perspective of the class example

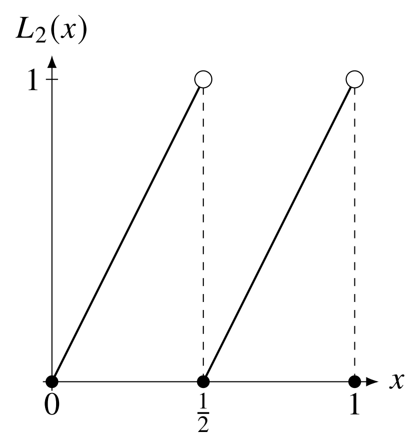


The same transformation, different sample

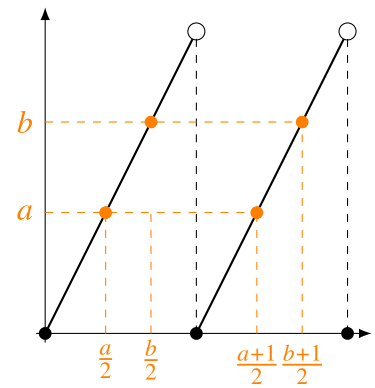
[I spent some time doing this and because it's a lot to write out, I will say it is worth considering what it ‘‘takes” to translate this perspective to the one shown in class, visually. With your mind.]

[Since algebra naturally comes to mind in these constructions and considerations, we briefly note that such rotations are all *endomorphisms* of the measure space.]

Example 0.0.50. (*Expanding maps*). Fix $d \in \mathbb{N}$. Let L_d be a linear expanding map, defined by $L_d(x) = d_x - \lfloor d_x \rfloor$. The following is an illustration for the map in the case $d = 2$.



Next, we claim that L_d is measure preserving, specifically for the Lebesgue measure m. In fact, $m\left(L_d^{-1}\left([a,b]\right)\right) = m\left(\bigsqcup_{k=0}^{d-1}\left[\frac{a+k}{d}, \frac{b+k}{d}\right]\right)$.



[I do believe that the ‘‘optional” problem from Homework 8 could be resolved using knowledge of such expanding maps.]

4/3 - **Koopman, unitary, and adjoint operators**.

Recall Let $(X, \mathcal{A}_0, \mu)$ be a probability space.

- $T : X \longrightarrow X$ is a *measurable* if $T^{-1}(A) \in \mathcal{A}_0, \forall A \in \mathcal{A}_0$.
- $T : X \longrightarrow X$ is a *measure preserving* if $\mu\big(T^{-1}(A)\big) = \mu(A), \forall A \in \mathcal{A}_0$.

Example 0.0.51. Let Σ_d be a sequence space, $\{\ldots, x_{-n}, x_{-(n-1)}, \ldots, x_n, \ldots, \ldots : x_i \in \{0, \ldots, d-1\}\}$. Suppose also that μ denotes the the *Bernoulli* measure. Now the map $T : \Sigma_d \longrightarrow \Sigma_d \quad T\big((x_n)\big)_k = x_{k+1}$, is commonly known as a *shift map* is such that, for instance $(\ldots 010\underline{1}01010\ldots) = \ldots 010101\underline{0}10\ldots, T^2(\ldots 010\underline{1}01010\ldots) = T(\ldots 01010\underline{1}010\ldots) = \ldots 0101010\underline{1}0\ldots \implies T^2(x) = x$.

Observe then that T is measure preserving if the measure μ is a Bernoulli measure. Recall that a Bernoulli measure of a cylinder an be expressed by $\mu\big(C_w\big) = \prod_i p_{w_i}$, for a cylinder C_w centered anywhere.

Lemma 0.0.52. *T is measure preserving $\iff T_*\mu = \mu$. That is, the measure by the push-forward of T is equal to the measure. (?).*

Proof. Recall that $\big(T_*\mu\big)(A) = \mu\big(T^{-1}(A)\big)$. □

Lemma 0.0.53. *Let $T : X \longrightarrow X$ be measurable, then T is measure preserving $\iff$ for all integrable $f : X \longrightarrow \mathbb{R}$ we have that $\int_X f\, d\mu = \int_X f \circ T\, d\mu$.*

Notice that another way to see this is "T acting as a change of coordinates." Moreover, all together $f \circ T : X \longrightarrow \mathbb{R}$ since $X \xrightarrow[T]{} X \xrightarrow[f]{} \mathbb{R}$.

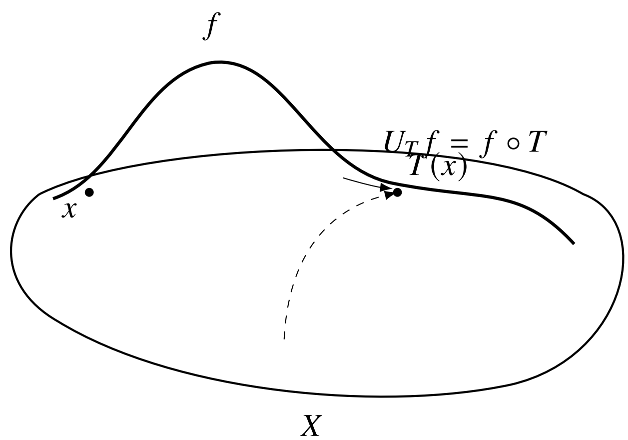
Proof. We begin with the reverse direction. Take $f = \chi_A$, then $f \circ T = \chi_A\big(T(x)\big) = \begin{cases} 1 & \text{if } T(x) \in A \\ 0 & \text{if } T(x) \notin A \end{cases} = \begin{cases} 1 & \text{if } x \in T^{-1}(A) \\ 0 & \text{if } x \notin T^{-1}(A) \end{cases} = \chi_{T^{-1}(A)}(x)$. But now by definition the measure of A can by expressed as an integral, and so our assumption would then imply that $\mu(A) = \int \chi_A\, \mathrm{d}\mu = \int_X \chi_A \circ T\, \mathrm{d}\mu = \int_A \chi_{T^{-1}(A)}(x)\, \mathrm{d}\mu = \mu\big(T^{-1}(A)\big)$.

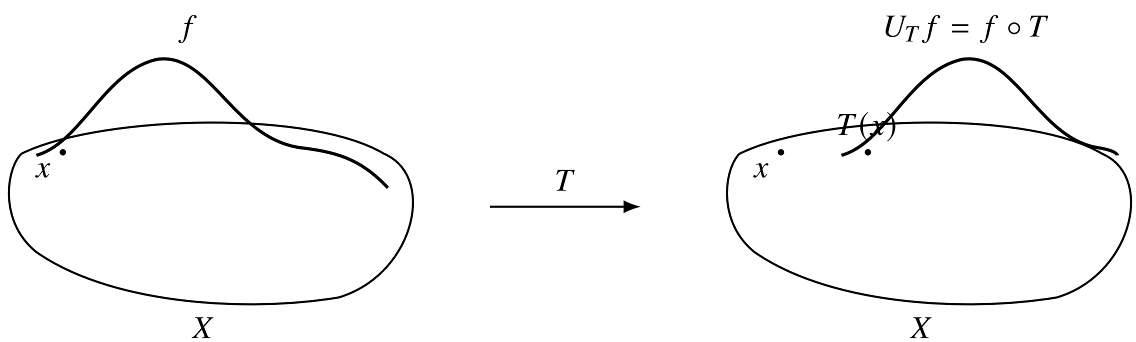
For the forward implication, recall that if f is integrable then the integral is arbitrarily well approximated by integrals of simple functions. By the same observations as above, if $\mu(A) = \mu\big(T^{-1}(A)\big), \forall A$, then for all simple φ it is the case that $\int \varphi = \int \varphi \circ T$.

Indeed, if $\varphi = \sum a_i \chi_{A_i}$, then $\varphi \circ T(x) = \sum a_i \chi_{A_i}\big(T(x)\big) \implies \int \varphi = \int \varphi \circ T = \sum a_i \mu(A_i)$. Note then that after taking limits it follows that $\int f = \int f \circ T$. □

Definition 0.0.54. If T is measure preserving, let $U_T : L^2(X,\mu) \longrightarrow L^2(X,\mu)$ be $U_T(f) = f \circ T$. [Picture here of spinning disk pile of sand, we mentally visualize rotating the disk, i.e., like the ground moving under the function, carrying the function with us].

U_T is called the *Koopman operator* associated with T.





Note, if $f \in L^2$, then $h(x) = |f(x)|^2$ is integrable. But, $\int h \circ T = \int h \implies \int \big|f\big(T(x)\big)\big|^2 = \int |f|^2 \implies \int |U_T f(x)|^2 = \int |f|^2 \implies \|U_T(f)\|_{L^2} = \|f\|_{L^2}$, i.e., U_T is *norm preserving*. Furthermore, $\langle U_T f, U_T g\rangle = \displaystyle\int \big(U_T f\big)(x)\overline{\big(U_T g\big)(x)}\,\mathrm{d}\mu = \displaystyle\int f\big(T(x)\big)\overline{g\big(T(x)\big)}\,\mathrm{d}\mu = \displaystyle\int f\bar{g}\,\mathrm{d}\mu = \langle f, g\rangle$. This is enough to motivate the following definition.

Definition 0.0.55. The operator $\phi : \mathcal{H} \longrightarrow \mathcal{H}$ such that $\langle \phi f, \phi g \rangle = \langle f, g \rangle$ is called unitary.

Moreover a unitary operator ϕ is also linear. This is what we mean by ‘‘operator.”

Lemma 0.0.56. *If ϕ is a bounded operator on $L^2(X, \mu)$, it has what is called an adjoint, that is an operator $\phi^* : L^2(X, \mu) \longrightarrow L^2(X, \mu)$ uniquely satisfying $\langle \phi f, g \rangle = \langle f, \phi^* g \rangle$. Furthermore, U is unitary $\iff U \circ U^* := UU^* = U^* U = \mathrm{id}$.*

Example 0.0.57. Let $X = \{1, \ldots, n\}$ denote a measure space. Let μ be the counting measure, then $L^2(X, \mu) = \mathbb{C}^n$ and the inner product is $\langle v, w \rangle = \sum v_i \overline{w_i}$, that is, the usual *Hermitian* inner product on an n-dimensional complex space. [All linear maps on complex metric spaces are bounded (?)]. Suppose $A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ is linear, then $A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \ldots & \ldots & a_{nn} \end{pmatrix} \implies a_{ij} = \langle Ae_j, e_i \rangle$. Now, A^* satisfies $\langle Ae_j, e_i \rangle = \langle e_j, A^* e_i \rangle \implies \langle A^* e_i, e_j \rangle = \overline{\langle Ae_j, e_i \rangle} = \overline{a_{ij}}$. Thus in finite dimensions, this adjoint operator A^* is the *conjugate transpose*.

Notice then that our last example gives new way to see something familiar, with $A^* = \overline{A^T}$. Moreover we then note that once we have a fixed orthonormal basis and inner product, we only then do we have enough structure to serve as the basis for a ‘‘transpose” operation.

Next time we will continue the discussion on Koopman maps.

4/6 - **Invertible Unitary operators and their adjoint(s), Spectra**.

Recall operators on a Hilbert space, where $V :=$ a normed vector space. We say an operator is a linear self-map, $\phi : V \longrightarrow V$. ϕ is bounded if $\|\phi\| = \sup_{v \neq 0} \frac{\|\phi(v)\|}{\|v\|} < \infty$.

Let $B(v)$ be the set of bounded operators on V.

Theorem 0.0.58. *Let $\mathcal{H}$ be a Hilbert space and $\phi \in B(\mathcal{H})$. Then there exists a unique bounded operator ϕ^* such that $\langle \phi(v), w \rangle = \langle v, \phi^*(w) \rangle, \forall v, w \in \mathcal{H}$.*

ϕ^* is called the adjoint of ϕ. For the following proof, we will need the following generalization of the Riesz Representation theorem.

Theorem 0.0.59. *If $\mathcal{H}$ is a Hilbert space and $f : \mathcal{H} \longrightarrow \mathbb{C}$ is a bounded linear functional (i.e., $|f(v)|/\|v\| < B$) $\implies \exists! w \in \mathbb{H}$ such that $f(v) = \langle v, w \rangle$.*

Next we will show the existence of the adjoint.

Proof. Fix $w \in \mathcal{H}$. Observe that $f_w(v) = \langle \phi(v), w \rangle$ is bounded and linear.

Linear $f_w(v_1+v_3) = \langle \phi(v_1+v_2), w \rangle = \langle \phi(v_1)+\phi(v_2), w \rangle = \langle \phi(v_1), w \rangle + \langle \phi(v_2), w \rangle = f_w(v_1) + f_2(v_2)$

Bounded $\frac{|f_w(v)|}{\|v\|} = \frac{\langle \phi(v), w \rangle}{\|v\|} \leq \frac{\|\phi(v)\| \, \|w\|}{\|v\|} \leq \|\varphi\| \, \|w\|$ with the penultimate inequality given by Cauchy-Schwartz, the last since ϕ, w is fixed and so f_w is bounded. (?).

Thus by Riesz, $\exists! \phi^*(w)$ such that $f_w(v) = \langle v, \phi^*(w) \rangle$, i.e., $\langle \phi(v), w \rangle = \langle v, \phi^*(w) \rangle$.

Next we will check that ϕ^* is linear and bounded. Now, $\phi^*(w_1 + w_2)$ is the unique vector such that $\underbrace{\langle \phi(v), w_1 + w_2 \rangle}_{= f_{w_1+w_2(v)}} = \langle v, \phi^*(w_1+w_2) \rangle, \forall v \in \mathcal{H}$. But $\langle \phi(v), w_1+w_2 \rangle = \langle \phi(v), w_1 \rangle + \langle \phi(v), w_2 \rangle = \langle v, \phi^*(w_1) \rangle + \langle v, \phi^*(w_2) \rangle = \langle v, \phi^*(w_1) + \phi^*(w_2) \rangle$. By uniqueness of Riesz, $\phi^*(w_1 + w_2) = \phi^*(w_1) + \phi^*(w_2)$. A similar argument shows that $\phi^*(\lambda w) = \lambda \phi^*(w)$, we must simply be a little careful pulling out λ with bars.

Now we need to estimate $\|\phi^*(w)\|$. Note $\langle \phi^*(w), \phi^*(w) \rangle = \left\langle \phi\big(\phi^*(w)\big), w \right\rangle \implies |\langle \phi^*(w), \phi^*(w) \rangle| = \left| \left\langle \phi\big(\phi^*(w)\big), w \right\rangle \right|$. Consequently, $\|\phi^*(w)\|^2 \leq \left\| \phi\big(\phi^*(w)\big) \right\| \, \|w\| \leq \|\phi\| \, \|\phi^*(w)\| \, \|w\| \implies \|\phi^*(w)\| \leq \|\phi\| \, \|w\| \implies \frac{\|\phi^*(w)\|}{\|w\|} \leq \|\phi\|$. ∎

Additionally, $\|\phi^*\| = \|\phi\|$, a fact we will not justify right this moment.

Theorem 0.0.60. *An operator on a Hilbert space, that is,* $\phi : \mathcal{H} \longrightarrow \mathcal{H}$*, is unitary* $\iff \phi^*\phi = \text{id}$.

Notice that if ϕ is invertible, then $\phi^* = \phi^{-1}$.

However an important note to make is that in infinite dimensional vector spaces we can have a self-map which is injective but *not surjective*. Indeed, this is a case where we may instead of a notion of a *one sided inverse*. (?).

Proof. We begin by showing the direct implication and assume that ϕ is unitary. Then $\forall v, w$ we have that $\langle \phi(v), \phi(w) \rangle = \langle v, w \rangle$, and so $\langle v, \phi^*\phi(w) \rangle = \langle v, w \rangle$. By the uniqueness give in Riesz, $\phi^*\phi(w) = w$.

For the reverse direction, see that $\langle v, w \rangle = \langle v, \phi^*\phi(w) \rangle = \langle \phi(v), \phi(w) \rangle$. So ϕ, an operator, which preserves the inner product, $\langle \cdot , \cdot \rangle$ is unitary. □

Thus, the adjoint of an (invertible) unitary operator is its inverse. A wonderful property. (?).

Spectra

The idea here is to describe a notion that generalizes eigenvalues.

Recall that λ is an *eigenvalue* $\iff \det(A - \lambda\,\text{id}) = 0 \iff A - \lambda\,\text{id}$ is *non-invertible*.

Definition 0.0.61. If $\phi : \mathcal{H} \longrightarrow \mathcal{H}$ is an operator, let $\text{Spec}(\phi) := \{\lambda : \phi - \lambda\,\text{id} \text{ has no bounded inverse}\}$. λ is an *eigenvalue* if $\phi - \lambda\,\text{id}$ is not injective.

Briefly, recall that ‘‘non-injective” in this context means that $\phi(v) = \lambda v = 0 \iff \phi(v) = \lambda v$.

Example 0.0.62. $\phi = U_{R_\alpha} : L^2\big([0,1], m\big) \longrightarrow L^2\big([0,1], m\big) \quad U_{R_\alpha}(f) = f \circ R_\alpha$.

Now see that $\text{spec}\big(U_{R_\alpha}\big) = S^1 \subset \mathbb{C}$, and its set of eigenvalues is $\{e^{2\pi i n\alpha} : n \in \mathbb{Z}\}$.

An example of unitary maps which are not invertible are the expanding maps (?). Moreover, their spectra is *not* the unit circle.

4/7 - **Finishing up on *Spectra***

Let $\phi : \mathcal{H} \longrightarrow \mathcal{H}$ be a bounded operator.

Definition 0.0.63. $x \in \operatorname{Spec}(\phi) \iff \phi - \lambda \operatorname{id}$ does <u>not</u> have a bounded inverse.

Example 0.0.64. $R_\alpha : [0,1) \longrightarrow [0,1) \quad R_\alpha(x) = x + \alpha \pmod 1 \quad \phi = U_{R_\alpha}$ (i.e., $U_{R_\alpha}(f) = f \circ R_\alpha$.

We claim that $\varphi_n \left(e^{2\pi i n x}\right)$ is an eigenfunction of the eigenvalue $e^{2\pi i n x}$. So we have that $\operatorname{Spec}(\phi) = \begin{cases} \left\{e^{2\pi i n x} : n \in \mathbb{Z} & \text{if } x \in \mathbb{Q}\right\} \\ S^1 & \text{if } x \notin \mathbb{Q} \end{cases}$

Note that $U_{R_\alpha}(\varphi_n) = \varphi_n(R_\alpha(x)) = \varphi_n(x+\alpha) = e^{2\pi i n(x+\alpha)} = e^{2\pi i n x + 2\pi i n \alpha} = e^{2\pi i n x} e^{2\pi i n \alpha} = e^{2\pi i n \alpha}\varphi_n(x)$.

Next then we have the example of $\alpha \sim \frac{1}{3} \implies e^{2\pi i 3\alpha} \approx 1 \implies U_{R_\alpha}(\varphi_3) \approx \varphi_3$.

To really get at the spectrum here we analyze the $\left(U_{R_\alpha} - e^{2\pi i n \alpha} \operatorname{id}\right)\varphi_n = 0$.

<u>Case</u> $\alpha \notin \mathbb{Q}$ We pick $\beta \in [0,1)$. Let us then invert $\left(\operatorname{id} - e^{2\pi i n \beta} \operatorname{id}\right)$. See then that $\left(U_{R_\alpha} - e^{2\pi i n \beta} \operatorname{id}\right)\varphi_n = \left(e^{2\pi i n \alpha}\varphi_n - e^{2\pi i n \beta}\varphi_n\right) = \left(e^{2\pi i n \alpha} - e^{2\pi i n \beta}\right)\varphi_n \implies \left(U_{R_\alpha} - e^{2\pi i n \beta} \operatorname{id}\right)^{-1}\varphi_n = \frac{1}{e^{2\pi i n \alpha} - e^{2\pi i n \beta}}\varphi_n$.

So we claim that $\left\|\left(U_{R_\alpha} - e^{2\pi i n \beta} \operatorname{id}\right)^{-1}\right\| \geq \sup_n \frac{\left\|\left(U_{R_\alpha} - e^{2\pi i n \beta} \operatorname{id}\right)^{-1}\right\|}{\|\varphi_n\|} \implies \sup_n \left|\frac{1}{e^{2\pi i n \alpha} - e^{2\pi i n \beta}}\right| = \infty$. Importantly, this is due to a claim we will not prove right this moment but basically what it says is that if $\alpha \notin \mathbb{Q}$, then $\forall \beta$ and $\forall \varepsilon > 0, \exists n, q$ such that $|n\alpha - \beta - q| < \varepsilon$.

Proposition 0.0.65. *If ϕ is unitary then* $\operatorname{Spec}(\phi) \subset D^2 = \{z : |z| \leq 1\}$. *If ϕ is unitary and invertible, then* $\operatorname{Spec}(\phi) \subset S^1 = \{z : |z| = 1\}$.

Proof. We'll show that if $|\lambda| > 1$, then $\phi - \lambda \operatorname{id}$ has a bounded inverse. In fact, to demonstrate the trick almost universally used for these sorts of arguments, we briefly digress to a case of *power series*. Consider that $\frac{1}{x - \lambda} = \frac{-1}{\lambda\left(1 - \frac{x}{\lambda}\right)} = -\frac{1}{\lambda}\sum_{n=0}^{\infty}\left(\frac{x}{\lambda}\right)^k$.

Now we define $\psi = -\sum_{n=0}^{\infty} \lambda^{-(k+1)} \phi^k$. Briefly we emphasize as sort of a reminder that $\phi^k = \underbrace{\phi \circ \phi \circ \phi \circ \ldots \circ \phi}_{k \text{ times}}$. In other words then, $\psi(v) = -\sum_{n=0}^{\infty} \lambda^{-(k+1)} \phi^k(v)$. It is easy to see nearly by inspection that ψ is linear (?).

Note that $\left\|\phi^k(v)\right\| = \|v\| \implies \|\psi(v)\| \leq \sum_{n=0}^{\infty} |\lambda|^{-(k+1)} \|v\| = \frac{\lambda}{(1-\lambda)\lambda} \|v\| \implies \|\psi\| \leq \frac{\lambda}{(1-|\lambda|)|\lambda|}$. For the rest of the argument we note that by completeness we get convergence of partial sums so long as these norms decay, such as we have just shown. Therefore we consider then that $\psi \circ (\phi - \lambda \operatorname{id}) = \psi\left(\phi(v) - \lambda v\right) = -\sum_{n=0}^{\infty} \lambda^{-(k+1)} \phi^k \left(\phi(v) - \lambda v\right) = -\sum_{n=0}^{\infty} \lambda^{-(k+1)} \phi^{k+1}(v) - \lambda^{-k} \phi^k(v)$. Notice that this last expression denotes a *telescoping sum*, so sending $k \to \infty$ shows only the first term to ‘‘survive.” Thus, $\operatorname{id}(v) = v$. In general, the spectrum of v in such spaces is contained in $B_{\|v\|}(v)$.

We also check that $(\phi - \lambda \operatorname{id}) \circ \psi = \operatorname{id}$. In fact, the same trick applied *mutatis mutandis* shows that $(\phi - \lambda \operatorname{id})^{-1} = \sum_{n=0}^{\infty} \lambda^k \phi^{-k}$ if $|\lambda| < 1$. [Fact here about telescoping sum sum, invertability?] □

Example 0.0.66. Let $d \in \mathbb{Z}$. Find Spec $\left(U_{L_d}\right)$, where $L_d : [0, 1) \longrightarrow [0, 1)$ is defined by $x \mapsto dx \pmod 1$, the expanding map of degree $2 = d$. [Here is the ‘‘rubber band” picture].

$U_{L_d}(\varphi_n)(x) = \varphi_n\left(U_{L_d}\right) = \varphi(dx) = e^{2\pi i n d x} = \varphi_{nd}$. I.e., $\left(U_{L_d}\right)^k(\varphi_n) = \varphi_{d^k m}$ [here is incredible picture of many $\varphi_i, i = -6, -5, \ldots 0, 1, 2, \ldots 6$ demonstrating why is not invertible. We see that φ_0 is an eigenfunction with eigenvalue 1.] The image is Span $\left(\frac{\text{Fourier coefficient}}{d}\right)$, as is it the case for all Koopman operators since constant functions are invariant (?).

Hence, U_{L_d} is <u>not</u> invertible. [picture of ‘‘copying and pasting sequentially and then normalizing the double copy of a curve” f under the operator U_{L_d}]. Spec $\left(U_{L_d}\right) = D^2$.

Transfer operators, $\left(U_{L_d}\right)^*$ are a whole big industry. Such concerns lead into broader discussions about spectral analysis and tricks in general. It may be remarked in fact, what we have just encountered as the notion of a *spectrum*, may be the most *accessible*.

The general notion of a spectrum is as the set of prime ideals for a commutative ring R. In fact, given a linear operator T on V, a finite-dimensional vector space over a field K, there is a particularly neat fact to relate the general notion to ours. We can view V as a *module* over a *polynomial ring* in one variable.

Given the canonical homomorphism, $\mathrm{ev}_T : K[x] \longrightarrow \mathrm{End}(V), \quad p(x) \mapsto p(T)$ we induce obtain module where $xv := T(v)$, $p(x)v = p(T)v$. Now let $K[T] := \mathrm{im}\big(\mathrm{ev}_T\big) = \big\{p(T) : p \in K[x]\big\} \subseteq \mathrm{End}(V)$. The *kernel* of $K[T]$ is the ideal (m_T) where m_T is the *minimal polynomial*. By the usual isomorphism theorems we have ${}^{K[x]}\big/{}_{(m_T)} \cong K[T]$. The *annihilator* of V as such a $K[x]$ module is $\mathrm{Ann}(V) := \{p(x) \in K[x] : p(T) = 0\}$. By definition, the annihilator of the module is actually (m_T). Now, the *prime ideals* of $K[x]$ are: $\begin{cases} (0) \\ (x-\lambda) & \text{for } \lambda \in K \end{cases}$. Finally, see that the prime ideals containing (m_T) are *precisely* those $(x-\lambda)$ such that $(x-\lambda)\big| m_T$, and these are *eigenvalues*. Briefly then, $\lambda \in \mathrm{Spec}(T) \iff (x-\lambda) \supseteq (m_T) \iff x-\lambda$ fails to be an invertible action on V.

A usual nice (beginner) example for polynomial rings is the case of when we have the field extension $\mathbb{Q}(\sqrt{2}) = \Big\{a + b\sqrt{2} : a, b \in \mathbb{Q}\Big\}$. Define $T = \alpha v$ where $\alpha = \sqrt{2}$. Taking $\mathbb{Q}(\sqrt{2})$ to be a vector space (it has 2 dimensions, a basis of $\{1, \sqrt{2}\}$), and observe the minimal polynomial of this case, the annihilator of the $\mathbb{Q}[x]$ module, is $m_T = x^2 - 2$. *However,* $x^2 - 2$ doesn't factor in $\mathbb{Q}$. Thus, $\{\lambda \in \mathbb{Q} : (x-\lambda) \supseteq (m_T)\} = \varnothing \implies \mathrm{Spec}(T) = \varnothing$. However, consider then instead the field of scalars to be $\mathbb{C}$ which reveals the roots $\pm\sqrt{2}$. Indeed, now we can write $x^2 - 2 = (x-\sqrt{2})(x+\sqrt{2}) \implies (x-\sqrt{2}) \supset (m_T), (x+\sqrt{2}) \supset (m_T) \implies \mathrm{Spec}(T) = \{-\sqrt{2}, \sqrt{2}\}$.

Briefly, since it is simple to define, for a commutative ring R there is a notion of topology we can put on $\mathrm{Spec}(R)$. Let I be an ideal in R and let C_I be the set of prime ideals such that $C_I \supset I$. The set $\{C_I : I \text{ is an ideal of } R\}$ treated as a collection of *closed sets* for a topology is called the *Zariski topology*. These topological spaces are always compact but not always Hausdorff.

Returning to functional analysis, we note some common tactics may be to change the norm so that we may obtain *spectral gaps* (?).

More generally, $\phi_\mu : L^2(\mathbb{C}, \mu) \longrightarrow L^2(\mathbb{C}, \mu) \quad \phi_\mu(f)(z) = zf(z)$. *Every* unitary operator can be written this way, and it tells us how the operator, *sees* the support, that is, the spectrum. (?)

4/10 - **Ergodic T, Von Neumann ergodicity criterion & theorem**.

Setting Let $(X, \mathcal{A}_0, \mu)$ be a probability space. Let $T : X \longrightarrow X$ be measure preserving. Some examples of this situation include R_α and L_d apropos to $m[0,1)$, and σ the shift-map in a sequence space Σ_d. $L_d(x) = dx \pmod 1$.

Definition 0.0.67. T is called *ergodic* if every T-invariant subset A has $\mu(A) = 0$ or $\mu(A) = 1$.

Recall, A is T-*invertible* $\iff T^{-1}(A) = A$. Informally we can then identify the idea Ergodicity with *irreducibility*. [non-ergodic $\implies T_{|_A} : A \longrightarrow A$ (?)].

Example 0.0.68. $\{0\}$ is invariant under L_d but $m(\{0\}) = 0$ (doesn't contradict ergodic (?)).

Example 0.0.69. $\{T^n(x) : n \in \mathbb{Z}\}$ is T-invariant and here we also note, doesn't contradict ergodicity.

Lemma 0.0.70. *Let $(X, \mathcal{A}_0, \mu)$ be a probability space, and $T : X \longrightarrow X$ be measure μ-preserving. Then T is ergodic $\iff$ the only T-invariant functions in $L^2(X, \mu)$ are almost everywhere constant.*

Recall that f is called T-invariant $\iff f \circ T = f$.

Proof. The converse direction is the easier one. We assume every T-invariant f is almost everywhere constant. Let A be T-invariant. [A moral here is that sets can be linked to functions by indicator functions]. Then χ_A is T-invariant since $\chi_A \circ T = \chi_{T^{-1}(A)} = \chi_A$, since A is T-invariant. Consequently, χ_A is almost everywhere constant. Thus, $A = \left(\chi_{A^{-1}}\right)(\{1\})$ has either measure 0 or 1.

For the other direction, we briefly outline the idea of identifying points in X, say a, with pull-backs of the form $f^{-1}\big((-\infty, a)\big)$. Indeed we can picture gradually ''dropping'' the ''line-sets'' of full measure, by way of a (?). Let $f : X \longrightarrow \mathbb{C}$ be T-invariant. It suffices to consider $f : X \longrightarrow \mathbb{R}$ by taking $\Re$ and $\Im$ parts.

For $a \in \mathbb{R}$, let $y_a = f^{-1}\big((-\infty, a)\big) = \{x \in X : f(x) < a\}$. Now y_a s T-invariant. Since if $T(x) \in y_a, f\big(T(A)\big) = f(x) < a \implies \mu(y_a) = 0$ or 1. Define $B \subset \mathbb{R}$ to be $B = \{a : \mu(y_a) = 1\}$. Set $a_0 = \inf B$. Note, $\forall n, \mu\left(y_{a_0} + y_n\right) = 1$. Set $A_0 = \bigcap_n y_{a_0} + 1/n \implies \mu(A_0) = 1$. Furthermore, if $x \in A_0 \implies \forall n, f(x) < a_0 + 1/n \implies f(x) \geq 0(?)$. Thus, ''bringing it down.''

Furthermore, $\forall n$ we have that $m\left(y_{a_0} - 1/n\right) = 0$. Let $A_1 = \bigcap_n \left(X \setminus (y_{a_0} - 1/n)\right)$. Then if $x \in A, f(x) \geq a_0 - 1/n \implies f(x)f(x) \geq a_0$ (?). Thus $A = A_0 \cap A_1$ has full measure and $a_0 \leq f(x) \leq a_0, \forall x \in A \implies f(x) = a_0$ on A. □

Theorem 0.0.71. *(**Von Neumann ergodicity criterion**). T is ergodic $\iff U_T$ has 1 as an eigenvalue with a multiplicity of 1.*

In other words, there exists a *unique* eigenfunction of eigenvalue, $\lambda = 1$, up to a scalar.

Proof.

$$\begin{gathered} f \text{ is an eigenfunction with } \lambda = 1 \\ \Updownarrow \\ U_T(f) = 1 \cdot f - f(?) \\ \Updownarrow \\ f \circ T = f \end{gathered}$$

Now apply the next lemma

Theorem 0.0.72. *(**Von Neumann ergodicity theorem**). Let $\mathcal{I}(T)$ denote the vector space of T-invertible functions. Then, $\forall f \in L^2(X, \mu)$ we have that $\frac{1}{N}\sum_{n=0}^{N-1} {U_T}^n f \xrightarrow{L^2} \pi_{\mathcal{I}(T)}(f)$, where $\pi_{\mathcal{I}(T)}(f)$ is the orthogonal projection onto $\mathcal{I}(T)$. When T is ergodic, the constant function is converges to is $\displaystyle\int_X f \, d\mu$.*

□

Some krazy functions one may consider to putting to Desmos sometime are: $f(x) = \cos(2\pi x)$, $\frac{1}{n}\sum_{n=0}^{N-1} f\left(2^N x - \lfloor 2^N x \rfloor\right)$, and $\frac{1}{n}\sum_{n=0}^{N-1} f\left(x + n\sqrt{2} - \lfloor x + n\sqrt{2} \rfloor\right)$

4/13 -

Recall We let $(X, \mathcal{A}, \mu)$ be a probability space and $T : X \longrightarrow X$ be measure preserving. Also, $U_T : L^2(X, \mu) \longrightarrow L^2(X, \mu)$ with $U_T f = f \circ T$.

Theorem 0.0.73. ***(Von Neumann)*** *For all $f \in L^2(X, \mu)$, we have that $\frac{1}{N}\sum_{n=0}^{N-1} {U_T}^n f \xrightarrow{L^2} \pi_{\mathcal{I}(f)}$, where $\mathcal{I} \subset L^2(X, \mu)$ are the T-invariant functions.*

Theorem 0.0.74. ***(Birkhoff)*** *Assume T is ergodic. Then, $\forall f \in L^1(X, \mu)\, \exists A_f \subset X$ such that $\mu(A_f) = 1$.*

Also, $\forall x \in A_f$, we have that $\frac{1}{N}\sum_{n=0}^{N-1} f\big(T^n(x)\big) \longrightarrow \int f\, d\mu$. That is, "$L^2$ convergence is almost everywhere."

Theorem 0.0.75. ***(Kronecker/Von Neumann)*** *Assume T is ergodic. Then $L^2(X, \mu)$ has a basis if eigenvectors $\iff$ T is isomorphic to a translation on a compact abelian group.*

$$\begin{array}{ccc} A \subset X & \xrightarrow{\quad T \quad} & A \subset X \\ \Big\downarrow H & & \Big\downarrow H \\ B \subset G & \xrightarrow{\quad +g \quad} & B \subset G \end{array}$$

Mixing

Consider that $\frac{\mu(T^{-n}(A) \cap B)}{\mu(B)} \to \frac{\mu(A)}{1}$. Accordingly we obtain then that $\mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B)$.

Definition 0.0.76. T is called *mixing* if $\forall A, B \subset X$, we have that $\mu(T^{-n}(A) \cap B) \to \mu(A)\mu(B)$ for all measurable A, B.

Lemma 0.0.77. *If T is mixing $\implies$ T is ergodic.*

Proof. We'll show that if $A \subset X$ is T-invariant, then $\mu(A) = 0$ or $\mu(A) = 1$.

We take $B = A$, as in the definition of mixing. It follows that if A is T-invariant, $A = T^{-n}(A)$. Consequently we have that $\mu(A) = \mu\big(A \cap T^{-n}(A)\big) \to \mu(A)\mu(T^{-n}A) = \mu(A)^2$. In other words, now have that $\mu(A) = \mu(A)^2 \implies \mu(A) = 0$ or $\mu(A) = 1$. $\square$

Spectral Characterization: Matrix coefficients/correlations

Definition 0.0.78. If $U : \mathcal{H} \longrightarrow \mathcal{H}$ is unitary and $f, g \in \mathcal{H}$, then $U_{fg} := \langle f, U_g \rangle$ is called the *matrix coefficient* or *correlation* for f, g.

Example 0.0.79. Take $\mathcal{H} = \mathbb{C}^n$ and U to be an $n \times n$ unitary matrix. Additionally, $U_{e_i e_j} = (i, j)$-th entry of U.

Note that $|U_{fg}| = |\langle f, U_g \rangle| \le \|f\| \|U_g\| = \|f\| \|g\| \le 1$, with equality if f, g are unit vectors.

Definition 0.0.80. $L^2_0(X, \mu) = \{f \in L^2(X, \mu) : \int f \, \mathrm{d}\mu = 0\} = \{\text{constants}\}^\perp$.

Theorem 0.0.81. *The following are equivalent:*

(1) T is a mixing.

(2) $\forall f, g \in L^2(X, \mu)$*, we have* $\displaystyle\int f \cdot (g \circ T^n) \, d\mu \to \left(\int f \, d\mu\right)\left(\int g \, d\mu\right)$

(3) $\forall f, g \in L^2_0(X, \mu)$ *it is the case that* $\langle f, {U_T}^n g \rangle \to 0$.
In other words, the matrix coefficients of ${U_T}^n$ *acting on* $L^2_0(X, \mu)$ *all converge to 0. The phrases then "decay of matrix coefficients" and "decay of correlations" come to mind.*

Proof. $\big((2) \implies (1)\big)$ Fix measurable $A, B \subset X$ so that χ_A, χ_B are measurable. It follows that $\mu\big(A \cap T^{-n}(B)\big) = \displaystyle\int \chi_{A \cap T^{-n}(B)} \, \mathrm{d}m = \int \chi_A \chi_{T^{-n}B} \, \mathrm{d}\mu = \int \chi_A \chi_B \circ T^n \, \mathrm{d}\mu \to \left(\int \chi_A \, \mathrm{d}\mu\right)\left(\int \chi_B \, \mathrm{d}\mu\right) = \mu(A)\mu(B)$.

$\big((1) \implies (2)\big)$ If mixing, the same argument shows (2) for characteristic functions. Both sides are bilinear in f, g. $\square$

4/14 - **Rates of Convergence for correlations and more norms**.

Recall T is *mixing* $\iff \mu(A \cap T^{-n}(B)) \implies (?)(A), \forall A, B \subset X \iff \langle f, U_T{}^n g\rangle \to 0, \forall f, g \in L_0^2(X, \mu)$. Independence may be gleaned here $\mu(C \cap D) = \mu(C)\mu(D)$.

Theorem 0.0.82. *Let $L_d : [0,1) \longrightarrow [0,1)$ be the linear expanding map $L_d(x) = dx \pmod 1, |d| \geq L_d(?) \implies f, g \in C^1([0,1)), \left|\langle f, U_{L_d}{}^n g\rangle\right| \leq \lambda^n \|f\|_{C^1} + \|g\|_{C^1}$, if $fg = 0(?)$, for some $\lambda < 1$.*

It is generally true for expanding maps that $|T(x)| \geq \alpha > 1$ for some fixed $\alpha > 1$.

We now reflect upon the fact that all of the interesting Banach spaces we've build contain C^∞ functions as a dense subspace. L^p, C^r. In particular, we can identify these spaces as ‘‘completions” of C^∞ functions with regard to some norm.

Example 0.0.83. Recall $\varphi_n = 2^{2\pi i n x}$ is a complete orthonormal set on $L^2([0,1), m) \implies \forall f \in L^2([0,1)), f = \sum_{n=-\infty}^{\infty} \hat{f}_n \varphi_n$.

If $f \in C_1^1, f' = \sum_{n=-\infty}^{\infty} \hat{f}_n \cdot 2\pi i n \varphi_n$. Then, $\forall x \in [0,1), |f'(x)| \leq \sum_{n=-\infty}^{\infty} \left|\hat{f}_n 2\pi i n \varphi_n\right| = 2\pi \sum_{n=-\infty}^{\infty} |n| \left|\hat{f}_n\right|$. Here is a new definition for a norm, $\|f\| = \sum_{n=-\infty}^{\infty} |n| \left|\hat{f}_n\right|$. Let $W = \left\{f \in L^2 : \sum_{n=-\infty}^{\infty} |n| \left|\hat{f}_n\right| < \infty\right\}$. We may call some certain spaces then, *Sobolev spaces*.

Lemma 0.0.84. $C^1([0,1)) \subset W$.

Proof. $\|\cdot\|_{L^1} \leq \|\cdot\|_W$ (?). If $f \in W$ then $f_N = \sum_{-N}^{N} \hat{f}_N \to f$, with regard to $\|\cdot\|_W$, □

I.e., $\|f_N - f\|_W \to 0 \implies \|f_N - f\|_{C^1} \leq \|f_N - f\|_W \to 0$ [Brief aside here about compact embeddings].

Other useful norms

Hölder norm $\|f\|_\theta = \|f\|_{C^0} + \sup_{x \neq y} \dfrac{|f(x) - f(y)|}{|x-y|^\theta}$, usually for $0 < \theta < 1$. In fact, in the case where $\theta > 1$, we just get constant functions. Now, $\|\cdot\|_{C^0} \leq \|\cdot\|_\theta \leq C\|\cdot\|_{C^1}$. We may also see that $\|f\|_\theta = \|f\|_{C^0} + \sup_{x \neq y} \dfrac{|f(x) - f(y)|}{|x-y|}|x-y|^{1-\theta}$.

Bounded variation Let $\|f\|_{BV} = \sup_{a=0\leq x_1<x_2<\ldots<x_n=b} \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)|$. This norm in a sense the total ‘‘length” travaled say, by some graph in the plane. Every time the line goes down the ‘‘value” flips up. [For each function which has bounded variation, there exists a function which doesn’t have bounded variation (?)].

Revisiting measures

- Absolute continuity & Radon-Nikodym derivatives.
- Fubini’s theorem, which is hard to prove because the criterion of measurability is hard to understand.

4/17 - **Radon-Nikodym derivatives & absolute continuity**

Definition 0.0.85. Let $X, \mathcal{A}_0$ be a set with a σ-algebra. If μ, ν are measures defined on $\mathcal{A}_0$, we say that ν is *absolutely continuous* with regard to μ if $\mu(A) = 0 \implies \nu(A) = 0, \forall A \in \mathcal{A}_0$. We write $\nu \ll \mu$.

If $\nu \ll \mu$ and $\mu \ll \nu$, we say that the measures are *equivalent*. One way to think of this is that we symmetrize a partial order and this implies that we get equivalence classes (?).

Example 0.0.86. See that $\delta_{\{0\}}$, the Dirac measure at 0 and m the Lebesgue measure are such that $m \not\ll \delta_{\{0\}}$ and $\delta_{\{0\}} \not\ll m$. Therefore, there is not a total order but a partial order. I.e., $\delta_{\{0\}}\big((0,1)\big) = 0$, but $m\big((0,1)\big) = 1$. Also, $m\big(\{0\}\big) = 0$ but $\delta_{\{0\}}\big(\{0\}\big) = 1$ is of full measure.

Example 0.0.87. We will describe a construction that always gives an absolutely continuous measure. Let $(X, \mathcal{A}_0, \mu)$ be a measure space, and $f : X \longrightarrow \mathbb{R}_{\geq 0}$ be measurable. Define $\mu_f(A) = \displaystyle\int_A f \, \mathrm{d}\mu = \int_X f \cdot \chi_A \, \mathrm{d}\mu$. If f is identically 1, then $\mu_f = \mu$. In fact, if μ_f is a probability measure $\iff \int f \, \mathrm{d}\mu = 1$.

<u>Observe</u>: $\mu_f \ll \mu, \forall f$. (?)So, $\mu(A) = 9 \implies \displaystyle\int_A f \, \mathrm{d}\mu = 0 \implies \mu_f(A) = 0$.

Theorem 0.0.88. ***(Radon-Nikodym)****. If $\nu \ll \mu$ then $\exists! f$ which is non-negative and measurable, and such that $\nu = \mu_f$. f is called the Radon-Nikodym derivative and is denoted $\dfrac{d\nu}{d\mu}$.*

Writing in a sort of ‘‘Leibnizian notation” aids us in understanding the following equation, itslef really a change of variables formula (?). Indeed, given that $\displaystyle\int_X g \, \mathrm{d}\mu = \int_X \left(g \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right) \mathrm{d}\mu$, then we have that $\displaystyle\int_X \chi_A \, \mathrm{d}\nu = \nu(A) = \int_X \chi_A \, \mathrm{d}\mu = \int_X \left(\chi_A \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right) \mathrm{d}\mu$.
If $\mu \ll$ Leb, and μ is a probability measure, we call it an absolutely continuous probability distribution on $\mathbb{R}$. And here we collide with familiar probability, see that $\dfrac{\mathrm{d}\mu}{\mathrm{d}\nu}$ is a *probability density function*, a pdf (!!!). In general, the cumulative distribution functions, the *cdf* of μ, is $g(x) = \mu\big((-\infty, x)\big)$

We now switch gears to our next topic.

Theorem 0.0.89. ***(Fubini)***. *Let $f : X \times Y \longrightarrow \mathbb{R}$ be integrable with regard to $\mu \otimes \nu$, then for almost every $x \in X$ we have that $f_x(y) := f(x, y)$ is ν-integrable. Furthermore,* $\displaystyle\int_{X\times Y} f \, d(\mu \otimes \nu) = \int_X \left[\int_Y f_x(y) \, d\nu(y) \right] d\mu(x)$.
There is a ''sneaky assumption'' here. We assume that the space $(Y, \mathcal{A}_{0Y}, \mu)$ is complete. (?).

That is, if $\nu(A) = 0$ and $B \subset A \implies B \in \mathcal{A}_{0Y}$. [Recall that Caratheodory gives a complete space and we need f to be measurable with regard to the complete space (?)].

Breifly we note that $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$, the standard generative vectors of the tensor product between A and B (?), $(\mu \otimes \nu) \left(\bigsqcup_i A_i \times B_i \right) = \sum \mu(A_i)\nu(B_i)$. We breifly acknowledge *Tonelli's theorem* here but will at the moment not delve further.

fin..

4/20 - **fin**

- The final exam is on Wednesday, April 29, from 10:30 am to 12:30 pm.
- There will be 8 problems.
  - Problems 1-3 will be of the multi-step sort, asking to supply definitions, examples, statements, and theorems of the topics, in order:
    - 1. Caratheodory's construction of measures.
    - 2. L^p-spaces.
    - 3. Shift-spaces and Bernoulli measure.
  - Problems 4-6 are problems we have ‘‘seen before.” In other words, homework problems or past midterm problems. From the following sources:
    - 4. Homework 12.
    - 5. Midterm.
    - 6 Homework 9.
  - Problems 7-8 will be ‘‘new problems,” on:
    - 7. Bounded operators as a normed vector space.
    - 8. Multiplication operators.
      - *(on a Hilbert space,) key idea in spectral theory besides Koopman operator*.
      - Proofs will be *fundamental* using fundamental ideas.

The following is a listing of the topics that will appear, though this time not strictly in order of appearance.

- Hilbert and L^2 spaces.
  - Definition of a Hilbert space (complete inner product space).
  - $L^2(X, \mu)$ is an example. $\langle f, g\rangle := \int_X f\bar{g}\,\mathrm{d}\mu$. Indeed, recall that we motivate some of what we do here by comparing the finite case, here we really take a sum of multiple coordinates (?). When X is a finite measure space, then theorem (?).
  - Recall something from the homework about closed subspaces.
  - The Riesz Representation theorem, only the statement not the whole proof. Recall it is used for adjoints. Orthogonal projections. Indeed, as an aside consider that in the Hilbert spaces we have one case, in another case we

recall that $1/p + 1/q = 1$, think about dual spaces and how this is analogous in this case to the inner product definition (?).

- Orthonormal sets.
  - Definition of a (complete) orthonormal set.
  - Recall that ‘‘we have a case of expanding an arbitrary element” (?) using Bessel’s inequality and Perseval’s theorem/identity.
  - As an aside, it is quite remarkable, one may even say shocking that we can look at coefficients and sum them up to still the infinite case (from a finite one?) (?). That we can’t detect where is a space is discrete or infinite space, that is, in the Hilbert case (?).
  - Examples:
    - Recall that for the family $\varphi_n(x) = e^{2\pi i n x}$ on $[0, 1)$ that completeness is hard to prove but showing it is normal is not so.
    - $\bar{\delta}(m) = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases}$ on $\ell^2(X)$.
  - As a brief aside then, it is quite something, analysis on a Hilbert space, because we can do computations (?).
- L^p-spaces.
  - Definition of L^p norm, L^p space.
  - Young, Hölder (and so Cauchy-Schwartz), and Jensen inequalities.
  - Completeness, that is, recall of the main theorem we proved: that for all measure spaces, the set of L^p functions is a complete normed vector space (that is *Banach*). In fact, one such space in particular comes up in the topic, *Hilbert and L^2-spaces*.
  - Recall that a counter-example to almost everywhere convergence is is the Typewriter sequence. Indeed, this sequences shows that L^p convergence almost everywhere *does not* imply almost everywhere convergence in general. Only for subsequences that is...
  - When is $L^p(X, \mu) \subset L^q(X, \mu)$?
  - There is an aside here about bounded functions implying convergence (?).
  - If the space in question is a *probability space*(?)
  - Basically, know the conclusion of some relevant homework problem(s).
- Caratheodory’s construction.

- Premeasures.
- Algebra's of sets.
- Outer measure induced by a premeasure.
- Measurability criterion
  - A is *measurable* if $\forall B, \quad m^*(B) = m^*(B \setminus A) + m^*(A \cap B)$. In other words, A is measurable if it ''cuts up any set well.''
  - Recall also that the class of measurable sets is then a σ-algebra and... (?)
  - An interesting aside is there *always* exists some non-measurable set.
- Examples:
  - Product measure, Bernoulli measure, finite measures.
  - An interesting aside is that measures on finite spaces are such that... a premeasure because of... every algebra is a σ-algebra (?).

- Ergodic and spectral theory.
  - Unitary operators. The Koopman operator was our main example in particular.
  - Spectrum and eigenvalues/eigen-functions for bounded operators.
  - Ergodicity and mixing for measure/*probability*-preserving transformations. Here some bridges we saw were:
    - Von Neumann ergodicity criterion and theorem.
    - We characterized mixing as decay of coefficients/matrix of coefficients.

    Some of this involved looking at situations such as (And here the order may be switched but it doesn't matter because of ''bars''), $\langle {U_T}^n f, g\rangle \to 0, \forall f, g \in L^2(X, \mu)$, i.e., ''with zero average on a metric space'' (?).

Remark. We conclude with a philosophical remark. We reflect that we started this course by trying to do something simple, that is, we attempted to evaluate the size of a set. We find that this was actually not that consistent to do. In light of this, we investigated and motivated many methods and precise steps to make it so. The process of undertaking such investigations, that is, *analysis*, has potential to import on other areas of life. The author of these notes interjects also that *beauty* itself needs no qualification and justification, and anything done well in terms of motivation and execution is imbued with something like that. *It justifies itself.*