Notas das aulas de Sistemas dinâmicos

Salvatore Cosentino
Departamento de Matemática - Universidade do Minho
Campus de Gualtar - 4710 Braga - PORTUGAL
gab B.4023, tel 253 604086
e-mail scosentino@math.uminho.pt
url http://w3.math.uminho.pt/~scosentino

24 de Março de 2006

Contents

1 Introdução

1.1 Sistemas dinâmicos

Uma estrutura típica de um modelo físico é a seguinte. Existe um espaço X, dito "espaço das configurações", ou "estados", do sistema (uma variedade simplética em mecânica clássica, um espaço de Hilbert em mecânica quântica, um certo espaço de funções em modelos hidrodinâmicos ...). Existe um espaço T, que chamamos "tempo", que contém um ponto chamado 0 (agora), e junto com t e s também contém $t+s$ (se é possível esperar uma hora e esperar duas horas, também deve ser possível esperar três horas). As "leis" da física definem uma dinâmica em X: uma família de transformações $\phi_t : X \to X$, definidas em cada instante $t \in T$, que verificam

$$\phi_0 = \text{identidade} \quad \text{e} \quad \phi_{t+s} = \phi_t \circ \phi_s .$$

Portanto, as leis definem uma ação $\phi : T \times X \to X$ do semigrupo "tempo" no "espaço dos estados". O ponto $\phi_t(x)$ é o estado no tempo t de um sistema que estava no estado x no tempo 0. A função $t \mapsto \phi_t(x)$ é a "trajetória" do estado inicial x. As leis podem ser "reversíveis", ou seja podem permitir decidir o que aconteceu no passado, e nesse caso o tempo é idealizado como sendo o grupo $\mathbf{R}$ ou $\mathbf{Z}$, ou irreversíveis, e neste caso o tempo é pensado como o semigrupo $\mathbf{R}_{\geq 0}$ ou $\mathbf{N}_0$. Se o tempo é contínuo, o (semi)grupo costuma ser definido por meio do seu gerador infinitesimal

$$\text{"} \lim_{t \downarrow 0} \left(\phi_t - \text{id}\right)/t \text{ "}$$

(o campo de vetores definido pela equação de Newton $F = ma$, o grupo de operadores unitários $e^{-\sqrt{-1}\hbar t H}$ gerado pelo operador hamiltoniano H, o semigrupo $e^{t\Delta}$ gerado pelo operador de Laplace-Beltrami Δ ...). Se o tempo é discreto, o (semi)grupo é gerado pela transformação

$$\phi_1 : X \to X$$

Numa experiência da física, não é necessariamente o estado do sistema que se observa. Fazer experiências quer dizer medir "observáveis", ou seja ler nos instrumentos do laboratório os valores de certas funções $\varphi : X \to \mathbf{R}$ (a distância entre dois planetas, a energia de um eletrão, a temperatura de um gás ...). A família de funções $\varphi_t = \varphi \circ \phi_t$ descreve a dinâmica do observável φ. De fato, o que se observa podem ser médias temporais do genero $\frac{1}{T}\int_0^T \varphi_t dt$, às vezes indiretamente por meio dos espetros de Fourier $\int e^{\sqrt{-1}kt}\varphi_t dt$ e de Laplace $\int e^{st}\varphi_t dt$

Exemplos físicos

Só para ter uma ideia...

Mecânica clássica. O espaço das configurações de uma partícula (pensada como um ponto material) é, de acordo com o princípio de relatividade de Galileo, $\mathbf{R}^3 \times \mathbf{R}^3$. Uma configuração é um vetor $x = (q, p)$, onde $q \in \mathbf{R}^3$ é a "posição" e $p = mq' \in \mathbf{R}^3$ o "momento", $'$ denota a derivada em ordem ao tempo e m a massa da partícula. A equação de Newton "força=massa×aceleração" se traduz no sistema de equações

$$q' = p/m \qquad p' = F$$

que definem um campo de vetores $\xi = (p/m, F)$ em $\mathbf{R}^3 \times \mathbf{R}^3$. A solução de $dx\,(t)\,/dt = \xi\,(x\,(t))$ com condição inicial $x\,(0) = x$ é a trajetória $t \mapsto \phi_t\,(x)$.

Mecânica quântica. O espaço dos estados de uma partícula é um espaço de Hilbert, por exemplo $L^2\left(\mathbf{R}^3\right)$. Um estado é uma função $q \mapsto \Psi\,(q)$, que tem a interpretação de "densidade de probabilidades de encontrar a partícula na posição q". A "energia" é um operador linear auto-ajunto $H : L^2\left(\mathbf{R}^3\right) \to L^2\left(\mathbf{R}^3\right)$, por exemplo da forma $-\left(\hbar^2/2m\right)\Delta + V\,(q)$, onde Δ é o operador de Laplace-Beltrami, $\hbar$ é a constante de Planck, m é a massa da partícula, e V é a "energia potencial". A equação de Schrodinger

$$\sqrt{-1}\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2m}\Delta\Psi + V \cdot \Psi$$

gera o grupo unitário de operadores $e^{-\sqrt{-1}tH/\hbar} : L^2\left(\mathbf{R}^3\right) \to L^2\left(\mathbf{R}^3\right)$.

Hidrodinâmica. O espaço dos estados é um espaço de funções com um certo número de derivadas parciais contínuas, por exemplo $C^k\left(\mathbf{R}^3, \mathbf{R}\right)$. Uma configuração, ou "campo", é uma função $q \mapsto u(q)$ e representa a "densidade macroscópica" de certos observáveis microscópicos (número de partículas, energia, pressão, ...). Uma equação diferencial fenomenológica descreve a evolução do campo. Por exemplo, a propagação do calor é suposta seguir a equação

$$\frac{\partial}{\partial t} u = \sigma \Delta u$$

onde Δ é o operador de Laplace-Beltrami e σ é um coeficiente que determina a velocidade de propagação. O operador diferencial Δ gera o semigrupo de operadores $e^{t\sigma\Delta} : C^k\left(\mathbf{R}^3, \mathbf{R}\right) \to C^k\left(\mathbf{R}^3, \mathbf{R}\right)$.

1.2 Problemas físicos e pequena história

O objetivo dos físicos é fazer previsões: querem saber o que acontece a um certo ponto x, ou melhor a um certo observável φ, passado um tempo t, e possivelmente dizer o que acontece quando t é grande. Eis uma lista, não exaustiva, de problemas fisicamente relevantes.

Calcular trajetórias. Resolver o "problema de Cauchy": dada uma condição inicial x, o estado do sistema no presente, determinar os estados futuros $\phi_t(x)$ com $t \geq 0$. No século XVII o Newton "inventou" o seu "methodus fluxionum" (o moderno cálculo diferencial e integral) para resolver as próprias equações e assim calcular as trajetórias dos planetas, dando uma "explicação" às leis de Kepler ...

Regularidades/periodicidades. Decidir se o sistema tem trajetórias regulares, no sentido de "previsíveis". As mais previsíveis são as trajetórias periódicas, que satisfazem $\phi_T(x) = x$ para algum tempo T dito período, e que portanto regressam a x em cada tempo múltiplo de T (a própria história do pensamento científico dos homens começou da observação das periodicidades dos astros, dando origem a cosmogonias e matemáticas em quase toda esquina do planeta). Decidir se as eventuais trajetórias regulares são observáveis, ou seja se uma pequena perturbação da condição inicial x ou da lei ϕ ainda produz uma trajetória próxima da trajetória regular, ou se estraga tudo. A procura de órbitas periódicas e a teoria das perturbações foi um dos temas favoritos dos físicos matemáticos do século XIX, particularmente interessados aos problemas da mecânica celeste. Nos anos cinquenta do século XX, Kolmogorov, e depois Arnold e Moser, provaram o resultado espetacular de que muitos sistemas hamiltonianos têm muitas órbitas "quase-periódicas".

Descrição qualitativa. Determinar o comportamento qualitativo da "maioria" das trajetórias. Acontece que, se o sistema não é extremamente simples (como um sistema kepleriano, uma partícula em um campo magnético constante, ...), é praticamente impossível "calcular" as trajectórias, embora possa ser possível provar a "existência". Os físicos devem ficar satisfeitos com uma descrição "qualitativa" das órbitas possíveis. No final do século XIX, Henri Poincaré mostrou que é possível fazer afirmações interessantes sobre o comportamento qualitativo das trajetórias utilizando informações fracas sobre a lei de evolução. O resultado mais espetacular é o seu famoso "teorema de recorrência". Outro exemplo é a classificação dos homeomorfismos do círculo, também devida a Poincaré e depois estudada por Denjoy.

Problemas numéricos. Embora seja geralmente impossível calcular trajetórias, é possivel obter trajetórias aproximadas (por exemplo, hoje em dia, utilizando um computador que "resolve" equações diferenciais, mas lembre que os astrónomos calculam "efemérides" e "calendários" desde milénios!). Um esquema muito simplificado do cálculo numérico é assim. Dada uma condição inicial x e um "passo" τ, obtemos uma aproximação $\phi'_\tau(x)$ de $\phi_\tau(x)$ com um erro que possivelmente sabemos estimar, por exemplo limitado por ε. A seguir, utilizamos o nosso valor inicial $\phi'_\tau(x)$ para estimar $\phi_{2\tau}(x)$, assim produzindo $\phi'_{2\tau}(x)$, supostamente a distância inferior a ε de $\phi_\tau(\phi'_\tau(x))$, mas geralmente a distância ainda maior de $\phi_{2\tau}(x)$... O problema é decidir se, quando n é grande, a nossa conjetura $\phi'_{n\tau}(x)$ ainda tem alguma coisa a ver com o verdadeiro $\phi_{n\tau}(x)$.

Regularidades probabilísticas. Muitos sistemas interessantes têm comportamento desordenado (por exemplo, as trajetórias podem ter dependência sensível das condições iniciais), e o estado inicial não pode ser determinado com precisão (quer por razões "a priori", quer porque todo instrumento tem a sua sensibilidade). A descrição estatística é neste caso uma necessidade e até pode simplificar a vida. Pode acontecer que o comportamento da maioria das trajetórias é tão irregular que acaba por parecer regular num sentido probabilístico. Este era o cenário imaginado por Ludwig Boltzmann, na sua teoria cinética dos gases, para justificar as lei observadas da termodinâmica. O estudo das regularidades probabilísticas dos sistemas dinâmicos é dito "teoria ergódica", em homenagem às intuições de Boltzmann, e nasceu nos anos trinta do século XX com os resultados de von Neumann, Birkhoff, Khinchin, Hopf, Kolmogorov... Em tempos mais recentes, matemáticos e físicos como Bowen, Ruelle, Sinai, descubriram ligações interessantes com a mecânica estatística de Maxwell e Gibbs...

Previsões robustas. Um sistema dinâmico pode ser pensado como uma "máquina" que pega numa condição inicial x e produz uma trajetória $t \mapsto \phi_t(x)$. O problema é decidir se uma pequena perturbação de ϕ (uma incerteza nos parâmetro da lei física), digamos ϕ', produz trajetórias "comparáveis" com as trajetórias de ϕ. Uma resposta que é particularmente apreciada pelos físicos consiste em formular resultados de "estabilidade", que digam que uma "distância" entre ϕ e ϕ' suficientemente pequena não altera a estrutura das trajetórias. Isto levanta também a questão de decidir se certos fenómenos são típicos ou não no espaço das possíveis dinâmicas. A procura de sistemas "estruturalmente estáveis" desenvolveu-se a partir das ideias de Andronov e Pontryagin, nos anos trinta do século XX. A "hiperbolicidade" enquanto chave da estabilidade estrutural foi descoberta nos anos sessenta por Anosov, Smale, Sinai ..., ao desenvolver ideias geométricas precedentes de Hadamard, Hopf , Hedlund ...

1.3 Estratégia

Para um matemático, um sistema dinâmico é uma ação $G \times X \to X$ de um (semi)grupo "grande" (tal que seja possível dar um sentido à uma expressão do género "$g \to \infty$") G sobre um espaço X. Estudar um sistema dinâmico quer dizer comprender o espaço das órbitas $G\backslash X$, ou melhor a maneira em que as diferentes órbitas Gx estão mergulhadas em X. A ênfase é no comportamento "assimptótico" das trajetórias $t \mapsto g_t x$ quando $g_t \to \infty$. Resulta que às vezes é possível fazer previsões interessantes esquecendo os "detalhes" da dinâmica, desde que X tenha alguma estrutura (uma topologia, uma métrica, uma estrutura diferenciável, simetrias, uma medida de probabilidades) que de alguma maneira precisa é respeitada pela evolução temporal, e que a lei de evolução tenha certas propriedades qualitativas. Este é o tema da teoria dos sistemas dinâmicos. A estratégia é selecionar modelos simples e tratáveis, possivelmente "descobrir" classes de sistemas com comportamento compreensível, na esperança de que sistemas "reais" tenham comportamentos comparáveis. Até esquecendo as motivações físicas, as ideias da teoria dos sistemas dinâmicos fornecem outra maneira de olhar certas estruturas matemáticas, e produzem resultados interessantes em análise, geometria, teoria de grupos, teoria de números, etc...

1.4 Composição

Isto não é um livro! Estas páginas contêm as minhas notas das aulas da disciplina de "Sistemas dinâmicos, Fractais e Caos Determinista" [1] leccionadas nos últimos quatro anos aos alunos de Ensino de Matemática. Também contêm as folhas práticas da disciplina: os "exercícios" e os "buracos" nas demonstrações. Foram escritas de maneira sintética, esquemática e informal. São baseadas essencialmente no primeiro capítulo, e em pequenos pedaços de outros, da exposição de Anatole Katok e Boris Hasselblat, *Introduction to the modern theory of dynamical systems*, Encyclopedia of mathematics and its applications, Cambridge University Press 1995. Vejam também o

[1] A palavra grega $\chi\alpha o\varsigma$, que pode ser traduzida como "abismo", contém a mesma base $\chi\alpha$- (e probabilmente deriva de) dos verbos $\chi\alpha\iota\nu\epsilon\iota\nu$ e $\chi\alpha\sigma\chi\epsilon\iota\nu$, que significam "abrir-se", "escancarar", e "abrir a boca", "bocejar" (cfr. $\chi\alpha\sigma\mu\alpha$, "abismo"). Foi utilizada em algumas cosmogonias gregas para indicar "a mistura desordenada de elementos anterior à formação do $\chi o\sigma\mu o\sigma$, o universo ordenado".

A expressão "caos determinista" é um oxímoro engraçado para indicar o aparente desordem produzido pelas leis, deterministas por definição, da física.

Quanto à palavra "fractal", foi inventada nos anos setenta pelo senhor Benoit Mandelbrot. Nada melhor de que ler o seu livro *Les object fractals: forme, hasard, et dimension*, Flammarion, Paris 1975, para tentar compreender o que ele quis dizer.

recente B. Hasselblatt and A. Katok, *A first course in dynamics: with a panorama of recent developments*, Cambridge University Press 2003. Os "problemas" são esboços de temas propostos aos alunos para estudo individual. Os "desafios" são exercícios difíceis cuja resolução implica esforço e pesquisa.

Some little ergodic theory, which assumes knowledge of integration theory, and examples or problems from physics, were not lectured (although they were the subject chosen by some of the students two years ago), and are written in english. You may want to take a look at it, if you are curious and want to have a better idea of what the whole story is worth for, or just don't print them.

Braga, 1 de Maio de 2005.
sal.

2 Sistemas dinâmicos topológicos, definições básicas

2.1 Transformações

Nestas notas, um *sistema dinâmico topológico* será uma ação de $\mathbf{N}_0$ ou de $\mathbf{Z}$ num espaço topológico (X, τ), gerada por uma transformação contínua $f : X \to X$.

As "iteradas" da transformação f são as transformações $f^n : X \to X$, definidas indutivamente por

$$f^0 = \mathrm{id} \qquad \text{e} \qquad f^n = f \circ f^{n-1} \text{ se } n \in \mathbf{N}$$

(cuidado! nesta notação $f^2(x)$ não é o quadrado de $f(x)$, mas $f(f(x))$ ...).

Em geral, se $n \in \mathbf{N}$ e A é um subconjunto de X, então $f^{-n}(A)$ denota o conjunto $\{x \in X \text{ t.q. } f^n(x) \in A\}$.

Se f é um homeomorfismo, é possível definir as iteradas $f^n : X \to X$ para todo $n \in \mathbf{Z}$.

A ação $\phi : \mathbf{N}_0 \times X \to X$, ou $\phi : \mathbf{Z} \times X \to X$ se f for um homeomorfismo, é definida por $\phi_n(x) = f^n(x)$.

Espaço dos estados. A seguir, (X, d) será um espaço métrico completo munido da topologia induzida τ, localmente compacto (todo ponto admite uma vizinhança compacta) e separável (admite um subconjunto enumerável denso, e portanto, sendo um espaço métrico, uma base enumerável da topologia). Por exemplo, domínios de $\mathbf{R}^n$, intervalos da reta, o círculo $\mathbf{R}/\mathbf{Z}$, o toro $\mathbf{R}^n/\mathbf{Z}^n$, o plano complexo $\mathbf{C}$, a esfera de Riemann $\overline{\mathbf{C}}$, conjuntos de Cantor... e produtos cartesianos de espaços finitos. Para evitar trivialidades e detalhes inúteis, assumiremos também que a cardinalidade de X seja infinita.

Propriedades genéricas. Tem interesse falar de coisas como "a maioria das trajetórias", ou "quase todas as trajetórias".

Sendo X um espaço topológico, existe a possibilidade de considerar medidas (de probabilidades ou infinitas) sobre os borelianos de X. Dada uma medida de probabilidades μ, uma propriedade é verificada em *μ-quase todo ponto* se o conjunto C dos pontos que têm a propriedade em causa tem probabilidade $\mu(C) = 1$ (se a massa de μ for infinita, a condição tem que ser substituida por $\mu(X \backslash C) = 0$).

O análogo topológico da dicotomia "probabilidade zero ou um" é possível se X é um *espaço de Baire*, ou seja um espaço topológico de Hausdorff (cada dois pontos distintos admitem vizinhanças disjuntas) onde toda interseção enumerável de abertos densos é densa. Num espaço de Baire, um subconjunto é dito *residual* (ou *gordo*) se contém uma interseção enumerável de abertos densos, e é dito *magro* se é uma reunião enumerável de subconjuntos "nowhere dense" (cuja aderência tem interior vazio), ou seja se o seu complementar é residual. Uma propriedade é dita *genérica* se o conjunto C dos pontos de X que têm a propriedade em causa é residual. O teorema de Baire diz que exemplos de espaços de Baire são os espaços métricos completos.

2.2 Trajetórias e órbitas

Estamos interessados no comportamento assimptótico da "história" de um ponto $x \in X$, a sequência de pontos

$$x \mapsto f(x) \mapsto f^2(x) \mapsto f^3(x) \mapsto ...$$

A ideia é que, se X é o espaço dos estados de um sistema físico, e se o sistema está no estado x no tempo 0, então o sistema estará no estado $f(x)$ no tempo 1, no estado $f^2(x) = f(f(x))$ no tempo 2, etc...

A *trajetória* de $x \in X$ é a sucessão $(x_n)_{n \in \mathbf{N}_0}$, a função que, dada a "condição inicial" $x_0 = x$, associa a cada tempo $n \geq 0$ o estado do sistema $x_n = f^n(x)$ no tempo n. Observe que a trajetória do ponto x é a solução da equação recursiva (ou equação às diferenças de primeira ordem)

$$x_{n+1} = f(x_n)$$

como condição inicial $x_0 = x$.

A *órbita* de $x \in X$ é a imagem da sua trajetória, ou seja o conjunto

$$\mathcal{O}_f^+(x) = \{f^n(x)\}_{n \in \mathbf{N}_0}$$

É costume abusar da linguagem e utilizar indistintamente as palavras "trajetória" e "órbita", desde que seja claro o contexto e o significado da frase.

Em geral, um ponto x pode ter mais de uma pré-imagem, e portanto o seu passado não é único. A *grande órbita* do ponto x é o conjunto

$$\mathcal{GO}_f(x) = \{x' \in X \text{ t.q. } \exists\, n, m \geq 0 \text{ t.q. } f^n(x') = f^m(x)\}$$

ou seja o conjunto dos pontos que têm eventualmente a mesma história futura de x. Observe que "estar na mesma grande órbita" é uma relação de equivalência.

Vale a pena observar desde já que o espaço quociente, dito *espaço das órbitas*, pode ser complicado, se as trajetórias forem pouco regulares. Por exemplo, se f admite uma órbita densa, então a topologia quociente no espaço das órbitas é a topologia trivial! Isto mostra que o espaço das órbitas, em quanto espaço topológico, pode não conter muita informação acerca das trajetórias.

Se f é invertível, também é útil definir a *órbita completa* (que de fato coincide com a grande órbita)

$$\mathcal{O}_f(x) = \{f^n(x)\}_{n\in\mathbf{Z}}$$

a história passada e futura de x. A relação "estar na mesma órbita completa" é uma relação de equivalência, e portanto X é uma reunião disjunta de órbitas completas.

Exercício. Estude a dinâmica, ou seja a estrutura das órbitas, de uma transformação arbitrária definida num conjunto finito. Observe que o estudo da dinâmica das transformações bijetivas consiste essencialmente no estudo dos grupos simétricos S_n.

2.3 Observáveis

Os *observáveis* são funções $\varphi : X \to \mathbf{R}$ ou $\mathbf{C}$. Se o sistema está inicialmente no estado x, e portanto é observado o valor $\varphi(x)$ de φ, passado um tempo n a observação de φ dará o valor $\varphi(f^n(x))$.

Particularmente interessantes são os observáveis que não mudam no tempo, que os físicos chamam "constantes do movimento". A função $\varphi : X \to \mathbf{R}$ é dita *invariante* se

$$\varphi \circ f = \varphi$$

ou seja se é constante em cada órbita. Observe que, se φ é invariante, $I \subset \mathbf{R}$ e $A = \varphi^{-1}(I)$, então $f^{-1}(A) = A$. A existência de uma função invariante φ contém a seguinte informação: se sabemos que $\varphi(x) = a$, então o futuro e o passado de x pertencem ao conjunto de nível $\Sigma_a = \{x \in X \text{ t.q. } \varphi(x) = a\}$, i.e. $\mathcal{GO}_f(x) \subset \Sigma_a$. As funções invariantes, portanto, reduzem o espaço disponível às trajetórias.

Também úteis são observáveis monótonos, crescentes ou decrescentes, ao variar o tempo, conhecidos em física como "funções de Lyapunov". Por exemplos, se sabemos que $\varphi \circ f \leq \varphi$, e que $\varphi(x) = a$, então o futuro de x "não sai" do conjunto de sub-nível $\Sigma_{\leq a} = \{x \in X \text{ t.q. } \varphi(x) \leq a\}$, e o passado de x "vem" de $\Sigma_{\geq a} = \{x \in X \text{ t.q. } \varphi(x) \geq a\}$.

Exercícios.
a. Mostre que, se $\varphi : X \to \mathbf{R}$ é invariante, $I \subset \mathbf{R}$ e $A = \varphi^{-1}(I)$, então $f^{-1}(A) = A$.
b. Mostre que a função característica do subconjunto $A \subset X$ é invariante sse $f^{-1}(A) = A$.

Médias temporais. A *média temporal* (ou *média de Birkhoff*) do observável φ até ao tempo $n \geq 0$ é o observável $\overline{\varphi}_n$ definido por

$$\overline{\varphi}_n(x) = \frac{1}{n+1}\sum_{k=0}^{n} \varphi\left(f^k(x)\right)$$

i.e. o valor de $\overline{\varphi}_n$ no ponto x é a média aritmética dos valores de φ na "n-órbita de x" $\left\{x, f(x), f^2(x), ..., f^n(x)\right\}$.

Se o limite

$$\overline{\varphi}(x) = \lim_{n\to\infty} \overline{\varphi}_n(x)$$

existe, tem o significado de "valor médio assimptótico" de φ ao longo da órbita de x. Observe também que $\overline{\varphi}(x) = (\overline{\varphi} \circ f)(x)$ nos pontos onde o limite existe.

Se, em particular, 1_A denota a função característica de um subconjunto $A \subset X$, então o limite

$$\overline{1_A}(x) = \lim_{n\to\infty} \frac{1}{n+1} \mathrm{card}\left\{0 \le k \le n \text{ t.q. } f^k(x) \in A\right\}$$

se existir, representa "a fração de tempo assimptótica" que a trajetória de x passa em A, ou seja "a frequência com que a trajetória de x visita o conjunto A".

2.4 Conjuntos invariantes

A função característica do subconjunto $A \subset X$ é invariante sse $f^{-1}(A) = A$. Uma definição consistente com esta observação é a seguinte.

Um subconjunto $A \subset X$ é dito *invariante* se

$$f^{-1}(A) = A$$

Esta condição implica que $f(A) \subset A$, e portanto um ponto de um conjunto invariante tem toda a sua história, futura e passada, contida no conjunto.

Observe que $\mathcal{GO}_f(x)$ é o menor conjunto invariante que contém x, e portanto um conjunto invariante é uma reunião de grandes órbitas, é composto pelas histórias possíveis passadas e futuras dos seus pontos.

Se f é invertível, $\mathcal{O}_f(x)$ é o menor conjunto invariante que contém x. Isto implica que, se $f : X \to X$ é invertível, um subconjunto $A \subset X$ é invariante sse é uma reunião de órbitas completas, i.e. se $A = \cup_{x\in A}\mathcal{O}_f(x)$.

Não há maneira de evitar o problema de distinguir entre outras noções de invariância. Um subconjunto $A \subset X$ é dito $+$*invariante* se $f(A) \subset A$ (se o futuro dos pontos de A vive em A), e é dito $-$*invariante* se $f^{-1}(A) \subset A$ (se o passado dos pontos de A está em A, ou os pontos de A "veem" de A).

Em particular, se A é +invariante é possível definir o sistema dinâmico $f|_A : A \to A$.

Exercícios.

a. Descubra as implicações entre as condições

$$f^{-1}(A) = A \quad , \quad f(A) \subset A \quad , \quad f^{-1}(A) \subset A \quad , \quad f(A) = A \quad , \quad f^{-1}(A) = A = f(A)$$

para uma transformação qualquer, uma transformação sobrejetiva e uma transformação bijetiva.

b. Considere o conjunto C igual a $\mathcal{GO}_f(x)$, $\mathcal{O}_f(x)$ ou $\mathcal{O}_f^+(x)$ para algum ponto $x \in X$, e determine as propriedades de invariância dos conjuntos C, $\overline{C}$, ∂C e C'.

c. Seja $A \subset X$. Mostre que $\cup_{n\ge 0} f^n(A)$ é um conjunto +invariante, de fato o menor conjunto +invariante que contém os pontos de A..

d. Seja $A \subset X$. Mostre que, se $f : X \to X$ é invertível, então $\cup_{n\in\mathbf{Z}} f^n(A)$ é um conjunto invariante, de fato o menor conjunto invariante que contém os pontos de A.

e. Seja $\varphi : X \to \mathbf{R}$ um observável, e seja $A \subset X$ o conjunto dos pontos $x \in X$ tais que o limite $\overline{\varphi}(x) = \lim_{n\to\infty} \overline{\varphi}_n(x)$ existe. Mostre que A é invariante, e que o observável $\overline{\varphi} : A \to \mathbf{R}$ é invariante com respeito à transformação $f|_A : A \to A$.

2.5 Conjugação topológica

Os sistemas dinâmicos topológicos $f : X \to X$ e $g : Y \to Y$ são (*topologicamente*) *conjugados* se existe um homeomorfismo $h : X \to Y$, dito *conjugação*, tal que

$$h \circ f = g \circ h$$

Observe que a condição pode também ser escrita como $f = h^{-1} \circ g \circ h$, e que é uma relação de equivalência. Por indução, vê-se que $f^n = h^{-1} \circ g^n \circ h$ para todo tempo $n \ge 0$. Em particular, uma conjugação h envia órbitas de f em órbitas de g.

A ideia é que duas transformações topologicamente conjugadas são indistinguíveis do ponto de vista topológico (estamos simplesmente a mudar o nome aos pontos do espaço dos estados), e portanto mais vale estudar a dinâmica de um representante por cada classe de equivalência.

Uma função contínua e sobrejetiva $h : X \to Y$ é dita *semiconjugação* entre os sistemas dinâmicos $f : X \to X$ e $g : Y \to Y$ se $h \circ f = g \circ h$. Neste caso, g é dito um *fator* de f. A h-imagem de toda órbita de f é uma órbita de g. Numa linguagem informal, "a dinâmica de g está contida na dinâmica de f". Esta definição é interessante sobretudo quando o conjunto dos pontos onde h não é injetiva é "pequeno".

2.6 Estabilidade estrutural

Uma "distância" natural no espaço das transformações contínuas dum espaço métrico (X, d) é a distância do sup

$$d_\infty(f, g) = \sup_{x \in X} d(f(x), g(x))$$

ou seja, f e g estão δ-próximos se $d(f(x), g(x)) < \delta$ para todo $x \in X$ (cuidado: dois pontos f e g podem estar a distância ∞ se X não é limitado!). Se X tem uma estrutura diferenciável, por exemplo se X é um domínio $V \subset \mathbf{R}^n$, ou em geral uma variedade diferenciável, podemos considerar a classe das transformações que têm derivadas parciais de ordem $\leq k$ contínuas, munida da topologia de Whitney $\mathcal{C}^k$. Sem entrar em detalhes técnicos, se X é um intervalo compacto da reta real, as topologias $\mathcal{C}^k$ são geradas pelas normas

$$\|f - g\|_{\mathcal{C}^0} = \sup_{x \in X} |f(x) - g(x)| \qquad \|f - g\|_{\mathcal{C}^1} = \sup_{x \in X} |f(x) - g(x)| + |f'(x) - g'(x)| \qquad \text{...etc.}$$

Uma transformação $f : X \to X$ é $\mathcal{C}^k$-*estruturalmente estável* se toda transformação $g : X \to X$ suficientemente próxima de f na topologia $\mathcal{C}^k$ é topologicamente conjugada a f.

Se o espaço X tem uma estrutura diferenciável, parece natural procurar conjugações diferenciáveis. O problema é que desta maneira uma divisão em classes de equivalência resulta muito fina e pouco significativa, devido a existência de "moduli"...

3 Modelos matemáticos

Para perceber melhor e apreciar os resultados da teoria, convem ter à mão alguns exemplos, nos quais testar as definições e entre os quais procurar exemplos e contra-exemplos...

3.1 Sistemas algébricos (translações em espaços homogeneos)

A maneira mais simples, de fato "tautológica", de construir ações é algébrica.

Seja G um grupo topológico, um grupo munido de uma topologia de Hausdorff tal que as operações $(g, g') \mapsto gg'$ e $g \mapsto g^{-1}$ sejam contínuas. Sejam $\Gamma \subset G$ um subgrupo, e G/Γ o espaço homogeneo munido da topologia quociente, a "maior" topologia em G/Γ tal que a projeção $\pi : G \to G/\Gamma$ seja contínua. Todo subgrupo $S \subset G$ age no espaço homogeneo, a ação $S \times G/\Gamma \to G/\Gamma$ sendo $(s, g\Gamma) \mapsto sg\Gamma$. Em particular, um subgrupo cíclico $S = \{s^n\}_{n\in\mathbf{Z}}$ gera uma ação $\phi : \mathbf{Z} \times X \to X$ definida por $\phi_n(g\Gamma) = s^n g\Gamma$. Esta ação consiste em iterar a translação $L_s : g\Gamma \mapsto sg\Gamma$.

Rotações do círculo. O círculo é o espaço quociente $S^1 = \mathbf{R}/\mathbf{Z}$ do grupo comutativo $\mathbf{R}$ pelo subgrupo $\mathbf{Z}$, munido da topologia quociente herdada da topologia euclidiana da reta. As *rotações* do círculo são as transformações $+\alpha : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ definidas por

$$x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$$

onde $\alpha \in \mathbf{R}$. Observe que o círculo $\mathbf{R}/\mathbf{Z}$ é um grupo comutativo, e que as as transformações $+\alpha$, com $\alpha \in \mathbf{R}/\mathbf{Z}$, são as translações do grupo.

A métrica euclidiana da reta "induz" uma métrica invariante d no círculo, definida por

$$\begin{aligned} d(x + \mathbf{Z}, x' + \mathbf{Z}) &= \min_{y\in\pi^{-1}\{x+\mathbf{Z}\},\ y'\in\pi^{-1}\{x'+\mathbf{Z}\}} |y - y'| \\ &= \min_{n\in\mathbf{Z}} |x - x' + n| \end{aligned}$$

onde $\pi : \mathbf{R} \to \mathbf{R}/\mathbf{Z}$ denota a projeção $x \mapsto x + \mathbf{Z}$.

As rotações do círculo são as isometrias de $(\mathbf{R}/\mathbf{Z}, d)$ que preservam a orientação.

Em notação multiplicativa, se o círculo é identificado com $\{z \in \mathbf{C}$ t.q. $|z| = 1\}$, as rotações do círculo são as transformações $z \mapsto e^{\sqrt{-1}2\pi\alpha} z$.

Exercícios.

a. Verifique que

$$d(x + \mathbf{Z}, x' + \mathbf{Z}) = \min_{n\in\mathbf{Z}} |x - x' + n|$$

é uma métrica no círculo $\mathbf{R}/\mathbf{Z}$.

b. Identifique o círculo $\mathbf{R}/\mathbf{Z}$ com o intervalo $[0, 1[$ (toda classe $x + \mathbf{Z}$ tem um e só um representante x neste intervalo), e dê uma expressão analítica para a distância $d(x + \mathbf{Z}, x' + \mathbf{Z})$ em função de x e x'.

(Observe que se $|x - x'| \le 1/2$ então $d(x + \mathbf{Z}, x' + \mathbf{Z}) = |x - x'|$)

c. Verifique que as rotações $R_\alpha : x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$, com $\alpha \in \mathbf{R}$, são isometrias, e portanto homeomorfismos, de $(\mathbf{R}/\mathbf{Z}, d)$.

Rotações do toro. Em dimensão maior, as *rotações* do toro: os homeomorfismos $+\alpha : \mathbf{R}^n/\mathbf{Z}^n \to \mathbf{R}^n/\mathbf{Z}^n$ definidos por

$$x + \mathbf{Z}^n \mapsto x + \alpha + \mathbf{Z}^n$$

onde agora $\alpha \in \mathbf{R}^n$.

3.2 Deslocamentos de Bernoulli (Bernoulli shifts)

Seja $X = \{1, 2, ..., z\}$ um "alfabeto" de $z > 1$ letras, um conjunto finito munido da topologia discreta, e seja $\Sigma^+ = X^{\mathbf{N}}$ o produto topológico de infinitas cópias de X. Os pontos de Σ^+ são denotados por $x = (x_1, x_2, ..., x_n, ...)$, com $x_n \in X$, e são "palavras" infinitas nas letras do alfabeto X.

Uma base da topologia produto τ em Σ^+ é a família $\mathcal{C}$ dos "cilindros centrados", os subconjuntos do género

$$C_\alpha = \left\{ x \in \Sigma^+ \text{ t.q. } x_1 = \alpha_1, x_2 = \alpha_2, ..., x_k = \alpha_k \right\}$$

palavras infinitas que "começam" pela palavra α, ao variar $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$ entre todas as palavras finitas nas letras de X. Observe que a família dos cilindros centrados é uma base de uma topologia porque é uma cobertura, pois $\Sigma^+ = C_1 \cup C_2 \cup ... \cup C_z$, e porque a interseção de dois cilindros centrados é um cilindro centrado ou o conjunto vazio. De fato, C_α e C_β têm interseção não vazia sse uma das duas palavras, por exemplo $\alpha = (\alpha_1, \alpha_2, ..., \alpha_k)$, é o pedaço inicial da outra palavra, no sentido em que $\beta = (\alpha_1, \alpha_2, ..., \alpha_k, \beta_{k+1} ..., \beta_{k+i})$, e neste caso $C_\alpha \cap C_\beta = C_\beta$. A ideia é que, quanto maior for o comprimento da palavra α, quanto "menor" é o cilindro C_α. Um aberto do produto topólogico Σ^+ é, por definição, uma reunião de cilindros centrados.

A topologia produto é metrizável, uma possibilidade é a métrica

$$d(x, x') = \sum_{n=1}^{\infty} \lambda^{-n} \cdot |x_n - x'_n|$$

onde $\lambda > 1$. Observe que, se λ é suficientemente grande, então os cilindros centrados são as bolas abertas, e também fechadas, de (Σ^+, d), e que d é uma ultra-métrica.

O espaço Σ^+ é um espaço métrico compacto, perfeito e totalmente desconexo, logo homeomorfo a um conjunto de Cantor.

O *deslocamento de Bernoulli* (em inglês, *Bernoulli shift)* é a transformação $\sigma : \Sigma^+ \to \Sigma^+$ que "esquece a primeira letra", definida por

$$\sigma : (x_1, x_2, x_3, ...) \mapsto (x_2, x_3, x_4, ...)$$

Observe que σ é uma transformação contínua, pois a imagem inversa de um cilindro centrado é uma reunião finita de z cilindros centrados, e portanto a imagem inversa de um aberto é uma reunião de cilindros centrados, logo um aberto. Observe também que σ não é invertível, todo ponto de Σ^+ tem z pré-imagens.

A ideia é que o alfabeto X representa os possíveis resultados de uma experiência, como lançar um dado com z faces, e um ponto de Σ^+ representa os resultados de uma sequência infinita de experiências idênticas. Iterar n vezes o deslocamento corresponde a "esquecer" os resultados das primeiras n experiências feitas.

3.3 Transformações do intervalo

São transformações $f : I \to \mathbf{R}$ definidas num intervalo I da reta real. A iteração é possível quando $f(I) \subset I$. Um estudo ingénuo (mas instrutivo!) das trajetórias pode-se fazer seguindo a história de um ponto x traçando os segmentos verticais e horizontais

$$(x, f(x)) \mapsto (f(x), f(x)) \mapsto (f(x), f^2(x)) \mapsto (f^2(x), f^2(x)) \mapsto (f^2(x), f^3(x)) \mapsto ...$$

ajudando-se com o gráfico da identidade.

É surprendente observar que, logo que f não é linear, as trajetórias podem ter comportamento mesmo complicado...

Experiência: a família quadrática. Podem começar procurando entender a dinâmica da *família quadrática*, a família das transformações

$$f_\lambda : x \mapsto \lambda x (1 - x)$$

definidas na reta real, ao variar o parâmetro real λ. Esta família contém quase tudo o que está nestas notas... é uma "palestra" para testar ideias e resultados, e o seu estudo ainda levanta hoje em dia problemas e conjeturas!

Experiência. Procure entender a dinâmica das seguintes transformações definidas em intervalos de $\mathbf{R}$ convenientes.

$$\begin{array}{lll} x \mapsto \pm x^3 & x \mapsto x^{1/3} & x \mapsto x^3 \pm x \\ x \mapsto x^2 + 1/4 & x \mapsto 3x(1-x) & x \mapsto 4x(1-x) \\ x \mapsto |1-x| & x \mapsto x^2 - 2 & x \mapsto \sin x \quad \text{e} \quad \cos x \end{array}$$

3.4 Transformações do círculo

Por alguma razão, a gente gosta de transformações definidas em espaços compactos, e a mais simples das variedades diferenciáveis compactas e sem fronteira é o círculo.

Uma transformação contínua do círculo $\mathbf{R}/\mathbf{Z}$ pode ser pensada como uma transformação contínua da reta real que "respeita" o retículo $\mathbf{Z}$. Formalmente, seja $\pi : \mathbf{R} \to \mathbf{R}/\mathbf{Z}$ a projeção $\pi(x) = x + \mathbf{Z}$. Toda função contínua $F : \mathbf{R} \to \mathbf{R}$ tal que $F(x+1) = F(x) \bmod \mathbf{Z}$ para todo $x \in \mathbf{R}$ induz uma função contínua $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ definida por

$$f(x + \mathbf{Z}) = F(x) + \mathbf{Z}$$

Por outro lado, pode-se mostrar que toda função contínua $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ admite um *levantamento*, i.e. uma função contínua $F : \mathbf{R} \to \mathbf{R}$ tal que $f \circ \pi = \pi \circ F$. O levantamento não é único, mas dois levantamentos F e G de f diferem por um inteiro, no sentido em que existe $n \in \mathbf{Z}$ tal que $F(x) = G(x) + n$ para todo $x \in \mathbf{R}$. De fato, $F - G$ é uma função contínua da reta real (que é conexa) com valores inteiros, e portanto é constante. Isto implica que o número inteiro

$$\deg(f) = F(x+1) - F(x)$$

dito *grau* de f, não depende do levantamento. O grau de f é a cardinalidade algébrica das pré-imagens $x' \in f^{-1}\{x\}$ de um ponto genérico x do círculo, onde cada x' conta ± 1 dependendo se f preserva ou inverte a orientação numa sua vizinhança.

Um homeomorfismo do círculo tem grau ± 1, dependendo se preserva ou menos a orientação. Os *recobribentos do círculo de grau* k, com $k \in \mathbf{Z} \setminus \{0\}$, são as transformações $\times k : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ definidas por

$$x + \mathbf{Z} \mapsto k \cdot x + \mathbf{Z}$$

Observe que cada ponto do círculo tem precisamente $|k|$ pré-imagens.

Em geral, uma transformação do círculo de grau k tem um levantamento que é da forma $F(x) = k \cdot x + h(x)$, onde h é uma função contínua periódica de período um, i.e. tal que $h(x+1) = h(x)$ para todo $x \in \mathbf{R}$.

4 Models from physics and other natural sciences

"... forse stima che la filosofia sia un libro e una fantasia d'un uomo, come l'*Iliade* e l'*Orlando furioso*, libri ne' quali la meno importante cosa è che quello che vi è scritto sia vero. Signor Sarsi, la cosa non istà così. La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (io dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, e conoscer i caratteri, ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto."

Galileo Galilei, *Il saggiatore*, 1623.

4.1 Structure of physical models

Flows of vector fields. The main way in which dynamical systems enter in physics is through differential equations. Let X be a differentiable manifold, and let ξ be a vector field on X. If we assume that the differential equation

$$x' = \xi$$

with any given initial condition $x(0) = x$, has solutions $t \mapsto x(t)$ which exist for any time $t \in \mathbf{R}$ (as is the case when ξ is smooth and X is compact), then the flow of ξ is the action $\phi : \mathbf{R} \times X \to X$ given by $\phi_t(x) = x(t)$. One could specialize to discrete time looking at the system at multiples integers $n\tau$ of a given time $\tau > 0$, and this amounts to iterate the invertible transformation ϕ_τ.

Most of the definitons and theorems that we are going to state and prove, possibly slightly modified, make sense for continuous flows too.

Newtonian mechanics. According to greeks, the "velocity" $q' = \frac{d}{dt}q$ of a planet, where $q \in \mathbf{R}^3$ is its position in our euclidean space and t is time, was determined by gods or whatever forced planets to move around circles. Then came Galileo, and showed that gods could at most determine the "aceleration" $q'' = \frac{d^2}{dt^2}q$, since the laws of physics should be written in the same way by an observer in any reference system at uniform rectilinear motion with respect to the fixed stars. Finally came Newton, who decided that what gods determined was to be called "force", and discovered that the trajectories of planets, fulfilling Kepler's experimental three laws, were solutions of his famous (second order differential) equation

$$mq'' = F$$

where m is the mass of the planet, and where the attractive force F between the planet and the Sun is proportional to the product of their masses and inverse proportional to the square of their distance.

Later, somebody noticed that most observed forces were "conservative", could be written as $F = -\nabla V$, for some real valued function $V(q)$ called "potential energy". There follows that Newton equations can be written as $mq'' = -\nabla V$, and that the "total energy"

$$E = \frac{1}{2}m\left|q'\right|^2 + V(q)$$

is constant along trajectories. The function $\frac{1}{2}m\left|q'\right|^2$ is called "kynetic energy" of the system.

An alternative (and indeed useful) formulation of Newtonian mechanics is the one developed by Lagrange. He defined the "Lagrangian" of the system as

$$L(q, q') = \frac{1}{2}m\left|q'\right|^2 - V(q)$$

and observed that Newton equations are equivalent to the (Euler)-Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial q'}\right) = \frac{\partial L}{\partial q}$$

The product $p = mq' = \partial L/\partial q'$ is called "momentum", and, since p/m is the gradient of the "kinetic energy" $K(p) = |p|^2/2m$, Hamilton could write Newton's second order differential equations as the system of first order differential equations

$$q' = \frac{\partial H}{\partial p} \qquad p' = -\frac{\partial H}{\partial q}$$

where $H(q,p) = K(p) + V(q)$ is the total energy as function of q and p, nowdays called "Hamiltonian". It is a simple check that the energy is a constant of the motion, since

$$\frac{d}{dt}H = \frac{\partial H}{\partial q} \cdot q' + \frac{\partial H}{\partial p} \cdot p' = \frac{\partial H}{\partial q} \cdot \frac{\partial H}{\partial p} - \frac{\partial H}{\partial p} \cdot \frac{\partial H}{\partial q} = 0$$

Hamiltonian flows. The modern abstract formulation of classical mechanics is as follows. Let (X, ω) be a symplectic manifold, i.e. a differentiable manifold X of even dimension $2n$, equipped with a smooth closed differential two-form ω such that $\omega^n \neq 0$. Darboux theorem says that locally one can choose "canonical" coordinates $(q_1, ..., q_n, p_1, .., p_n)$ such that $\omega = \sum_{k=1}^n dp_k \wedge dq_k$. Let $H : X \to \mathbf{R}$ be a smooth function, called "Hamiltonian" and thought as the "energy" of the system. Typically, it has the form "kinetic energy+potential energy", where the kinetic energy is a positive definite quadratic form in the momenta p, and the potential energy is a function V depending on the positions q and possibly on the momenta p. The Hamiltonian vector field ξ is defined by the identity $dH = i_\xi \omega$, and the *Hamiltonian flow* is the flow of ξ. In canonical coordinates, the equations of motion read

$$\frac{dq_k}{dt} = \frac{\partial H}{\partial p_k} \qquad \frac{dp_k}{dt} = -\frac{\partial H}{\partial q_k}$$

It happens that the Hamiltonian flow ϕ preserves the energy, namely $H(\phi_t(x)) = H(x)$ for any $x \in X$ and any time $t \in \mathbf{R}$, as follows form the fact that $\xi H = 0$.

Geodesic flows. The simplest mechanical system, the free motion of a particle, belongs to the class of geodesic flows. Let (M, g) be a Riemannian manifold, g beeing the Riemannian metric. Let SM be the unit tangent bundle of M. If M is geodesically complete, to every unit vector $v \in SM$ there corresponds a unique geodesic line (i.e. a local isometry) $c : \mathbf{R} \to M$ such that $c'(0) = v$. The *geodesic flow* is the action $\phi : \mathbf{R} \times SM \to SM$, defined as $\phi_t(v) = c'(t)$.

Particularly interesting are geodesic flows over homogeneous spaces. Apart from the rather trivial exemple of flat spaces, a source of interesting dynamical properties is the geodesic flow on a manifold with constant negative curvature. The proptotype is as follows. The group $G = PSL(2, \mathbf{R})$ can be seen as the orientation preserving isometry group of the Poincaré half-plane $\mathbf{H}$, equipped with the hyperbolic metric of sectional curvature -1. Its action is transitive. Since the stabilizer of a point in the half-plane is isomorphic to the group of rotations $SO(2)$, we can identify $S\mathbf{D}$ with G. Now, let Γ be a discrete cocompact subgroup of G with no torsion. The quotient space $\Sigma = \mathbf{D}/\Gamma$ is a compact Riemann surface, which comes equipped with a Riemannian metric of sectional curvature -1, and its unit tangent bundle is diffeomorphic to G/Γ. The geodesic flow on $S\Sigma$ is then the algebraic flow $\phi : \mathbf{R} \times G/\Gamma \to G/\Gamma$ defined as $\phi_t(g\Gamma) = e_t g\Gamma$, where

$$e_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$

4.2 Examples from physics and other natural sciences

Example: harmonic oscillator. The Newton equation

$$mq'' = -kq$$

on the line, where m is the mass of the particle and k is a positive constant, models small oscillations of a one-dimensional system near a stable equilibrium $q = q' = 0$ (e.g. a spring, a pendulum, ...). Let $\omega = \sqrt{k/m}$ be the resonant frequency, and define the complex variable $z = \omega q + \sqrt{-1}q'$. Newton equation then takes the form of a first order linear equation in the complex line, namely $z' = -\sqrt{-1}\omega z$, whose solution is $z(t) = e^{-\sqrt{-1}\omega t} z(0)$. In terms of the original (physical) variables, the solution of the Cauchy problem is

$$q(t) = \cos(\omega t)\, q(0) + \sin(\omega t)\, \frac{q'(0)}{\omega}$$

Hence all trajectories are closed with common period $2\pi/\omega$. Observe that orbits are ellipses in the q-q' plane, determined by the "energy"

$$E = \frac{1}{2}\left(q'^2 + \omega^2 q^2\right)$$

which is a constant of the motion.

Execícise. Discuss the Newton equation

$$mq'' = kq$$

on the line, with k positive. Does it admit periodic orbits?

Example: logistic equation. The *logistic equation* is the first order differential equation

$$x' = \lambda x\,(1 - x)$$

in the domain $x \in [0, 1]$, where λ is a real parameter. It was proposed by the mathematician Pierre-Francoise Verhulst in 1838 as a continuous time model for population growth in a limited environment. The size N of the population is supposed to follows the differenial equation

$$N' = \lambda N\,(1 - N/N_{\max})$$

where λ is the fertility (or mortality, depending on its sign) and $N_{\max}$ is the maximum allowed population. The parameter $x = N/N_{\max}$ appearing in the model is thus the relative size.

The solution of the Cauchy problem is easily seen to be

$$x\,(t) = \frac{1}{1 + \left(x\,(0)^{-1} - 1\right)e^{-\lambda t}}$$

Hence trajectories are asymptotic to $x = 1$ or $x = 0$, depending on the sign of λ.

Much more interesting is its discrete time version, the logistic map.

Problem: Lotka-Volterra predator-prey model. The Lotka-Volterra model is the system of coupled first order differential equations

$$\begin{aligned} x' &= ax - bxy \\ y' &= -cy + dxy \end{aligned}$$

defined for positive x and y, where the parameters a, b, c, d are positive constants. These equations, proposed by the physical chemist Alfred Lotka and by the mathematician Vito Volterra between 1925 and 1926, model the population dynamics of two species, x preys and y predator, in the same territory. Preys increase exponentially at rate a and are killed at rate proportional to the probability of beeing captured by a predator, while predators decrease exponentially at rate c and increase at rate proportional to the probability of capturing preys.

Discuss the possible dynamics depending on the values of the parameters.

Problem: mathematical pendulum. The Newton equation

$$I\theta'' = -mg\ell \sin\theta$$

models the motion of an idealized pendulum (meaning a point mass attached to a wire of negligible weight, under a constant gravitational force) with mass m and length ℓ, where $I = m\ell^2$ is the moment of inertia, g is the gravitational acceleration (on the Earth), and θ is the angle of the wire with the origin $\theta = 0$ located at the equilibrium point. Observe that in the limit of small oscillations we could safely replace $\sin\theta \simeq \theta$ and we are back to the harmonic oscillator. The energy

$$E = \frac{1}{2} I\,(\theta')^2 - mg\ell\cos\theta$$

is a constant of the motion. As above, we can define the resonant frequency $\omega = \sqrt{mg\ell/I}$ and write the equation as

$$\theta'' = -\omega^2 \sin\theta$$

Draw the phase portrait (i.e. the constant energy curves in the θ-θ' plane) and discuss the motion. Hint: open any book of classical mechanics.

Problem: keplerian orbits. Kepler problem deals with the motion of two point-like bodies (planets and/or stars) under mutual gravitational interaction. Let $m_1, m_2 > 0$ be their masses, and $q_1, q_2 \in \mathbf{R}^3$ their positions, respectively. Gravitational interaction is described by the conservative force $-\nabla V$ with potential energy

$$V(q_1, q_2) = G\frac{m_1 m_2}{|q_1 - q_2|}$$

where G is the gravitational constant. This force verifies the "third law of dynamics", hence the total linear and angular momentum

$$P = m_1 q_1' + m_2 q_2' \qquad \text{and} \qquad M = m_1 q_1 \wedge q_1' + m_2 q_2 \wedge q_2'$$

are conserved. This implies that the center of mass moves at uniform rectilinear speed and that the motion of the two bodies takes place in a plane orthogonal to the angular momentum M. If we choose a Galileian reference system where $P = 0$ and M is parallel to the z-axis (in particular M is supposed different from the zero vector, a case which leads to a collision ...) , the full system is described by the single vector $q_2 - q_1$ in the x-y plane, which we write in polar coordinates as $\rho e^{\sqrt{-1}2\pi\theta}$. It turns out that the two-body problem is equivalent to the motion of a single point mass $m = \frac{m_1 m_2}{m_1 + m_2}$ moving on a plane under the influence of a potential energy $V(\rho) = -G\frac{m}{\rho}$, the (conserved) energy beeing

$$E = \frac{1}{2}m\left(\rho'^2 + \rho^2\theta'^2\right) + V(\rho)$$

Observe that if one of the bodies is much bigger than the other (like the Sun and the Earth), say $m_1 \gg m_2$, then the center of mass nearly coincides with the position q_1 of the bigger body, while the reduced mass m is essentially the mass m_2 of the smaller one (hence it looks like the Earth moving around the Sun, as Galileo had suggested). Newton equations are

$$\left(\rho^2\theta'\right)' = 0$$

meaning that (twice) the "areal velocity" $A = \rho^2\theta'$ is a constant of the motion, and

$$\begin{aligned} m\rho'' &= m\rho\theta'^2 - \frac{\partial V}{\partial \rho}(\rho) \\ &= -\frac{\partial}{\partial \rho}V_A(\rho) \end{aligned}$$

where we defined the "effective potential energy" as $V_A(\rho) = m\left(A^2/2\rho - G/\rho\right)$.

Show that for $E < 0$ orbits are ellipses and the motion obeys Kepler three laws[2], while for $E \geq 0$ orbits are hyperbola's wings or parabolas, hence they are not closed. Hint: observe that the conservation of energy and angular momentum implies that

$$\rho'^2 = 2\left(\frac{E - V_A(\rho)}{m}\right)$$

draw the graph of $V_A(\rho)$ ...then open any book of classical mechanics.

[2]In *Astronomia nova*, 1609,and *Harmonices mundi*, 1619, Johannes Kepler published his three laws of planetary motions:
i) planets moves in ellipses with focus at the Sun,
ii) the radius vector describes equal areas in equal times,
iii) the squares of the periods are to each other as the cubes of the mean distance from the Sun.
It was with the purpose to derive Kepler laws from a second order differential equation $mq'' = F$ that Isaac Newton realized that the force of gravitational attraction between the Sun and a planet (hence between any two bodies!) should be proportional to m/ρ^2 (*Philosophiae naturalis principia mathematica*, 1687).

5 Órbitas regulares e perturbações

5.1 Órbitas periódicas

As órbitas mais simples são os *pontos fixos* de f, os pontos $p \in X$ tais que $f(p) = p$.

A seguir, as *órbitas periódicas.* O ponto $p \in X$ é dito *periódico* de *período* $n \geq 1$ se $f^n(p) = p$ e se n é o menor dos tempos $k \geq 1$ tais que $f^k(p) = p$. Assim, a órbita do ponto periódico p é um conjunto finito

$$\left\{p,\ f(p), f^2(p), ..., f^{n-1}(p)\right\}$$

de pontos que são permutados pela transformação f.

Um ponto x pode ter órbita finita sem ser periódico: pode acontecer que existe $k \geq 1$ tal que $f^k(x)$ é um ponto periódico. Tais pontos, que "caem" numa órbita periódica passado um tempo positivo, são ditos *pré-periódicos.*

$\mathrm{Fix}(f^n)$ denota o conjunto dos pontos fixos da transformação f^n, ou seja o conjunto dos pontos periódicos de f cujos períodos dividem n.

$$\mathrm{Per}_f = \cup_{n\geq 1}\mathrm{Fix}(f^n)$$

denota o conjunto dos pontos periódicos da transformação f. Observe que cada um dos conjuntos $\mathrm{Fix}(f^n)$ é fechado, pois f^n é contínua, mas a reunião Per_f pode não ser.

Exemplo: rotações racionais do círculo. *Uma rotação* $+\alpha : x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$ *do círculo* $\mathbf{R}/\mathbf{Z}$ *tem pontos periódicos sse* α *é racional.* Pois, se $\alpha = p/q$ com $(p, q) = 1$ e $q > 0$, então todo ponto do círculo é periódico de período q. Por outro lado, se α é irracional, não existe nenhum natural $n \geq 1$ tal que $x + \mathbf{Z} = x + n\alpha + \mathbf{Z}$, seja o que for x.

Exemplo: crescimento exponencial. Considere a equação recursiva[3]

$$x_{n+1} = \lambda x_n$$

com condição inicial $x_0 > 0$, onde λ é uma constante positiva. A solução é $x_n = \lambda^n \cdot x_0$. Logo, se $\lambda > 1$ a trajetória do todo ponto x_0 pela transformação $f : x \mapsto \lambda x$ definida nos reais positivos é divergente: a transformação f não tem pontos periódicos.

Discuta os casos em que $0 < \lambda < 1$ e $\lambda = 1$.

5.2 Teoremas de ponto fixo "topológicos"

Encontrar os pontos periódicos de uma transformação f, o seja os pontos fixos das iteradas f^n, é tudo menos que trivial. Pensem só no caso de um polinómio f de grau $k > 1$ definido na reta real. A iterada f^n é um polinómio de grau nk, e resolver a equação $f^n(x) = x$ não é fácil ...

Em dimensão um, conexos e convexos coincidem e são chamados intervalos. Este "milagre" é responsavel de dois critérios muito simples para provar a existência de pontos fixos em determinados intervalos do domínio de uma transformação contínua $f : I \to \mathbf{R}$. Numa linguagem sugestiva, dizem que se uma transformação restringe ou estica um intervalo compacto $J \subset I$, então fixa pelo menos um ponto deste intervalo. Formalmente,

Teorema de ponto fixo. *Seja* $f : I \to \mathbf{R}$ *uma transformação contínua definida num intervalo* $I \subset \mathbf{R}$.

i) Se $J \subset I$ *é um intervalo compacto tal que* $f(J) \subset J$, *então* f *tem um ponto fixo em* J.

ii) Se $J \subset I$ *é um intervalo compacto tal que* $J \subset f(J)$, *então* f *tem um ponto fixo em* J.

[3]Este é um modelo muito simples do crescimento de uma população num meio ambiente ilimitado, e λ representa a fertilidade da especie. Pode ser considerado uma aproximação e uma generalização do modelo proposto por Leonardo Pisano no *Liber abaci* em 1202, que era a equação recursiva

$$f_{n+2} = f_{n+1} + f_n$$

com condições iniciais $f_1 = f_2 = 1$, onde f_n é o número de casais de coelhos no tempo n. A solução é a sucessão 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ... Se acreditamos que $f_{n+1} \simeq \lambda f_n$ (a igualdade pode ser verdadeira apenas assimptoticamente!) somos levados à equação $f_{n+1} = \lambda f_n$ onde λ é a raiz positiva de $\lambda^2 = \lambda + 1$, um número chamado "número/razão/seção áureo/a" ...

A prova é uma aplicação elementar do teorema de Bolzano à função $f - \mathrm{id}$. É instructivo procurar demonstrações mais abstratas. Observe que, se f não tivesse pontos fixos em J, então

$$x \mapsto \frac{f(x) - x}{|f(x) - x|}$$

seria uma aplicação contínua de um intervalo (J no caso i) ou um subintervalo de J no caso ii)) sobre o espaço desconexo $\{-1, 1\}$ ...

Outros teorema de ponto fixo. Em dimensão maior, o análogo do resultado i) é o "teorema de ponto fixo de Brouwer": uma transformação contínua $f : B^n \to B^n$ da bola fechada $B^n = \{x \in \mathbf{R}^n \text{ t.q. } |x| \leq 1\} \subset \mathbf{R}^n$ tem um ponto fixo. O "absurdo" que produz a demonstração deste resultado não é tão trivial como no caso do intervalo, e utiliza ideias de "topologia algébrica": não é possível deformar de maneira contínua a bola B^n até obter a esfera $S^{n-1} = \partial B^n$. A generalização em dimensão infinita é o "teorema de ponto fixo de Shauder-Tychonov": uma transformação contínua $f : K \to K$ de um subconjunto compacto e convexo K de um espaço de Banach (ou de um espaço vetorial topológico localmente convexo) tem um ponto fixo.

5.3 Bacia de atração

Se a trajetória de x é uma sucessão convergente, o seu limite é um ponto fixo. De fato, se $f^n(x) \to p$, a continuidade de f implica que

$$f(p) = f\left(\lim_{n\to\infty} f^n(x)\right) = \lim_{n\to\infty} f^{n+1}(x) = p$$

Seja p um ponto fixo de $f : X \to X$. A *bacia de atração*, ou *conjunto estável*, de p é o conjunto dos pontos cuja trajetória é assimptótica a p, i.e.

$$W^s(p) = \left\{x \in X \text{ t.q. } \lim_{n\to\infty} f^n(x) = p\right\}$$

A unicidade do limite de uma sucessão convergente num espaço métrico implica que os conjuntos estáveis de dois pontos fixos diferentes são disjuntos.

Execícios.
a. Seja $f : \mathbf{R} \to \mathbf{R}$ a transformação linear da reta definida por $x \mapsto \lambda x$. Estude a bacia de atração do ponto fixo 0 ao variar o parâmetro λ.
b. Seja $f : \mathbf{R}^2 \to \mathbf{R}^2$ a transformação linear do plano definida por

$$x \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} x$$

Estude a bacia de atração do ponto fixo 0 ao variar os parâmetros a, b, c, d.

5.4 Dinâmica das contrações

Seja (X, d) um espaço métrico. Uma aplicação $f : X \to X$ é uma *contração* (ou λ-*contração* se é importante lembrar o valor de λ) se é Lipschitz e tem constante de Lipschitz $\lambda < 1$, ou seja se existe $0 \leq \lambda < 1$ tal que para todos $x, x' \in X$

$$d(f(x), f(x')) \leq \lambda \cdot d(x, x')$$

A dinâmica das contrações é simples, e é descrita pelo

Princípio das contrações. *Todas as trajetórias de uma contração $f : X \to X$ são sucesssões de Cauchy, e a distância entre cada duas trajetórias diminue exponencialmente no tempo. Se X é completo, então a trajetória de todo ponto converge exponencialmente para o único ponto fixo de f.*

dem. Seja $x_0 \in X$ um ponto arbitrário, e seja (x_n) a sua trajetória, i.e. $x_{n+1} = f(x_n)$. Iterando a contratividade vê-se que $d(x_{k+1}, x_k) \leq d(x_1, x_0) \cdot \lambda^k$. Usando k vezes a desigualdade do triângulo e depois a convergência da série geométrica de razão λ, vê-se que

$$\begin{aligned} d(x_{n+k}, x_n) &\leq \sum_{j=0}^{k-1} d(x_{n+j+1}, x_{n+j}) \leq d(x_1, x_0) \cdot \sum_{j=0}^{k-1} \lambda^{n+j} \\ &\leq d(x_1, x_0) \cdot \lambda^n \cdot \sum_{j=0}^{\infty} \lambda^j \leq \frac{\lambda^n}{1-\lambda} \cdot d(x_1, x_0). \end{aligned}$$

Portanto, (x_n) é uma sucessão de Cauchy, pois $\lambda^n \cdot d(x_1, x_0)/(1-\lambda)$ é menor de cada $\varepsilon > 0$ a partir de um $n(\varepsilon)$ suficientemente grande. A continuidade de f implica que o limite $p = \lim_{n\to\infty} x_n$, que existe se X é completo, é um ponto fixo de f. A unicidade do ponto fixo é evidente, pois se p e p' são fixos então $d(p, p') = d(f(p), f(p')) \leq \lambda \cdot d(p, p')$ com $\lambda < 1$, donde $d(p, p') = 0$. Por outro lado, a contratividade também diz que $d(x_n, p) \leq \lambda^n \cdot d(x_0, p)$, ou seja que a convergência $x_n \to p$ é exponencial. $\square$

Estabilidade das contrações. Uma contração $f : X \to X$ do espaço métrico completo X pode ser pensada como uma "máquina" que pega numa condição inicial x e produz o ponto fixo $p = \lim_{n\to\infty} f^n(x)$, que neste caso nem depende de x. O problema é decidir se uma pequena perturbação de f, digamos $g : X \to X$, produz um ponto fixo p' próximo de p. A resposta consiste em formular um resultado de "continuidade", que diga que um controle da "distância" entre f e g permite ter um controle da distância entre p' e p. Ora, se $d_\infty(f, g) < \delta$, a transformação g pode não ser uma contração, por quanto pequeno seja o $\delta > 0$ (para se convencer, pode traçar gráficos de contrações da reta real, e ver que numa δ-vizinhança cabem gráficos de transformações mais chatas...). Uma solução é admitir que X tenha uma estrutura diferenciável e que as transformações sejam de classe $\mathcal{C}^1$. A condição $\|f - g\|_{\mathcal{C}^1} < \delta$ implica que, se f é uma λ-contração e $\delta < 1-\lambda$, então também g é uma contração e tem constante de Lipschitz $\leq \lambda + \delta$. Mais vale procurar o teorema de estabilidade diretamente dentro do espaço das contrações. Seja $\varepsilon > 0$ arbitrário. Seja $g : X \to X$ uma $(\lambda+\delta)$-contração que satisfaz $d(f(x), g(x)) < \delta$ para todo $x \in X$, com $0 < \delta < 1 - \lambda$. Se p' é o ponto fixo de g, então em particular $g^n(p) \to p'$ quando $n \to \infty$. Utilizando a desigualdade do triângulo vê-se que

$$\begin{aligned} d(p, p') &\leq \sum_{n=0}^{\infty} d\left(g^{n+1}(p), g^n(p)\right) \leq d(g(p), p) \cdot \sum_{n=0}^{\infty} (\lambda+\delta)^n \\ &\leq \delta \cdot \sum_{n=0}^{\infty} (\lambda+\delta)^n = \frac{\delta}{1-(\lambda+\delta)} \end{aligned}$$

e esta quantidade é $< \varepsilon$ desde que δ seja suficientemente pequeno.

Classes de equivalência das contrações lineares da reta. As contrações da reta real fornecem também um exemplo simples de como pode ser utilizada a dinâmica para construir uma conjugação topológica.

Sejam $f : x \mapsto \alpha x$ e $g : x \mapsto \beta x$ duas contrações lineares de $\mathbf{R}$, com $0 < \alpha, \beta < 1$. A origem é o ponto fixo das duas contrações. O conjunto $A = [-1, -\alpha[\cup]\alpha, 1]$ é um "domínio fundamental" pela ação de f em $\mathbf{R}\backslash\{0\}$, no sentido em que, dado um $x \in \mathbf{R}\backslash\{0\}$ arbitrário, existe e é único um tempo $n(x) \in \mathbf{Z}$ tal que $f^{n(x)}(x) \in A$. Analogamente, um domínio fundamental pela ação de g em $\mathbf{R}\backslash\{0\}$ é $B = [-1, -\beta[\cup]\beta, 1]$. Seja $H : \overline{A} \to \overline{B}$ um homeomorfismo que verifique $H(-1) = -1$, $H(-\alpha) = -\beta$, $H(\alpha) = \beta$ e $H(1) = 1$ (por exemplo, podem escolher um homeomorfismo afim). É imediato verificar que a receita

$$h(x) = \begin{cases} 0 & \text{se } x = 0 \\ g^{-n(x)}(H(f^{n(x)}(x))) & \text{se } x \neq 0 \end{cases}$$

define um homeomorfismo $h : \mathbf{R} \to \mathbf{R}$. Observando que $n(x) = n(f(x)) + 1$, vê-se que

$$\begin{aligned} (h \circ f)(x) &= g^{-n(f(x))}\left(H\left(f^{n(f(x))} f(x)\right)\right) = g^{-n(x)+1}\left(H\left(f^{n(x)-1}(f(x))\right)\right) \\ &= g\left(g^{-n(x)}\left(H\left(f^{n(x)}(x)\right)\right)\right) = (g \circ h)(x) \end{aligned}$$

e portanto h é uma conjugação topológica entre f e g.

O caso em que $-1 < \alpha, \beta < 0$ pode ser tratado da mesma maneira. Não é dificil verificar que as contrações $x \mapsto \alpha x$ e $x \mapsto \beta x$ não podem ser conjugadas se $\alpha \cdot \beta < 0$, i.e. se uma é crescente e a outra é decrescente. O resultado é que todas as contrações lineares e não triviais da reta real com a mesma orientação são topologicamente conjugadas, e portanto só existem duas classes de equivalências de contrações lineares da reta.

É importante observar que uma conjugação h entre $f : x \mapsto \alpha x$ e $g : x \mapsto \beta x$ não pode ser de classe $\mathcal{C}^1$, a não ser que $\alpha = \beta$. Pois, se $f = h^{-1} \circ g \circ h$ e se h é diferenciável, então a regra da cadeia implica que $f'(0) = g'(0)$, logo que $\alpha = \beta$. Este fenómeno explica porque na definição de estabilidade estrutural é melhor pedir que a conjugação seja só contínua.

Execícios.

a. Utilize o teorema do valor médio para mostrar que uma função $f : \mathbf{R}^n \to \mathbf{R}^n$ de classe $\mathcal{C}^1$ é uma contração sse existe $\lambda < 1$ tal que $|f'(x)| \leq \lambda$ para todo $x \in \mathbf{R}^n$.

b. Prove que uma contração de um espaço métrico compacto X não pode ser invertível, desde que o espaço contenha mais de um ponto.

(Compare os diâmetros de X e de $f(X)$)

c. Dê exemplos de contrações de

$$[0,1] \qquad [0,1] \times [0,1] \qquad B_r(x) = \{y \in \mathbf{R}^n \text{ t.q. } d(x,y) < r\} \qquad S^1 = \{z \in \mathbf{C} \text{ t.q. } |z| = 1\}$$

d. Mostre que uma transformação $f : X \to X$ tal que

$$d(f(x), f(x')) < d(x, x')$$

para todos $x, x' \in X$ distintos pode não ter pontos fixos, mesmo se o espaço métrico X for completo.

(Um exemplo com X não completo é obvio: basta retirar o ponto fixo de uma contração de um espaço completo. Um exemplo "minimalista" com X completo, é um espaço formado por uma única órbita, ou seja um espaço que só contém um ponto x_0 e os pontos $x_n = f^n(x_0)$ com $n \geq 0$. Se d_n denota a distância $d(x_n, x_{n-1})$, então $d_{n+1} < d_n$. Não queremos pontos fixos, portanto a sucessão (x_n) não pode ser convergente. Uma condição suficiente é que a série $\sum d_n$ seja divergente ...)

e. Sejam $a > 0$ e $x_0 > 0$. Mostre que a sucessão (x_n) definida indutivamente por

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$$

converge para $\sqrt{a}$. Assuma que $a = 2$, escolhe um ponto inicial $x_0 > 0$, e estime a distância entre x_n e $\sqrt{2}$. Utilize o resultado para dar uma aproximação de $\sqrt{2}$ com dois dígitos decimais correctos[4].

(A sucessão é uma trajetória da transformação $f : x \mapsto (x + a/x)/2$. Observe que a restrição de f ao domínio $[\sqrt{a}, \infty[$ é uma contração, e que $f(\mathbf{R}_{>0}) \subset [\sqrt{a}, \infty[$, logo toda trajetória cai neste domínio passada uma iteração...)

f. Mostre que as contrações lineares da reta $x \mapsto \alpha x$ e $x \mapsto \beta x$ não podem ser conjugadas se $\alpha \cdot \beta < 0$, i.e. se uma é crescente e a outra é decrescente.

(Uma conjugação é um homeomorfismo da reta, em particular é monótono...)

Observação: princípio das contrações, teoremas de ponto fixo "analíticos" e métodos variacionais. O princípio das contrações é o ingrediente de muitos teoremas de existência em matemática (pense só no teorema de existência das soluções das equações diferenciais). Se queremos provar a existência e a unicidade de um ponto $p \in X$ que satisfaz uma certa propriedade P, podemos tentar escrever a condição "x satisfaz P" na forma "$f(x) = x$", onde $f : X \to X$ é uma transformação. Se existe uma métrica d tal que f seja uma contração do espaço métrico completo (X, d), então a trajetória de todo ponto $x \in X$ converge para o único ponto p que satisfaz a propriedade em causa. O método funciona porque $(f^n(x))$ é uma sucessão convergente, e a

[4]Este era o método utilizado pelos babilónios (embora posteriormente atribuído a Arquitas de Taranto, a Heron de Alexandria, ou até ao Newton!) para "calcular" o lado de um quadrado de area a. Eles chegaram a ter uma aproximação de $\sqrt{2}$ que, em notação decimal, era 1.414213, um erro de apenas $\simeq 10^{-6}$! Uma conjetura sobre a origem do método está em O. Neugebauer, *The exact sciences in antiquity*, Dover, New York 1969.

distância $d(f^n(x), p)$ converge para 0 quando $n \to \infty$. Em outras palavras, a função $x \mapsto d(x,p)$ é estritamente decrescente nas trajetórias de f, e tem um mínimo, único, no ponto p. Em falta de uma contração natural f, isto sugere uma outra estratégia. A ideia agora é tentar provar a existência de uma função φ tal que "x satisfaz P" sse "$\varphi(x) < \varphi(x')$ para todo $x' \neq x$" (ou pelo menos para todo $x' \neq x$ numa vizinhança de x, o que quer dizer que φ tem um mínimo local em x), e depois utilizar resultados de compacidade para provar que o mínimo de φ existe e, com alguma sorte, é até único. Estes métodos são ditos "princípios variacionais", e tiveram origem na física do século XIX. As próprias leis da física costumam ser enunciadas na forma de princípios variacionais. Por exemplo, em mecânica clássica, a trajetória de um ponto material com posição inicial q_0 e condição final q_T é a (derivada da) curva $q : [0,T] \to \mathbf{R}^3$, com $q(0) = q_0$ e $q(T) = q_T$, que é um ponto crítico de um certo funcional

$$\varphi(q) = \int_0^T L(q(t), q'(t))\, dt$$

dito "ação", o integral da Lagrangiana $L =$"energia cinética$-$energia potencial" ao longo da trajetória. Esta é também a estratégia utilizada para provar a existência e a unicidade das soluções das equações às derivadas parciais da física matemática: procurar uma "ação" φ que tenha um ponto crítico nas soluções.

5.5 Ordem da reta real e trajetórias

A ordem da reta real implica restrições às trajetórias de transformações monótonas.

Transformações crescentes do intervalo. Por exemplo, seja $f : I \to I$ uma transformação contínua e crescente do intervalo I. Então toda trajetória $(x_n)_{n\in\mathbf{N}_0}$ é monótona, crescente ou decrescente. A sucessão monótona $(x_n)_{n\in\mathbf{N}_0}$ só pode fazer duas coisas: ser convergente, i.e. $x_n \to p$ para algum ponto fixo p, se é limitada, ou ser divergente, no sentido em que $x_n \to \pm\infty$, se não é limitada. Em particular, se o intervalo I é compacto, a segunda possibilidade é impossível, logo toda trajetória é convergente. Isto implica que, se I é compacto, existe um compacto não vazio $F \subset I$ de pontos fixos, e que os pontos em cada componente conexa de $I \backslash F$ têm trajetórias contidas na componente conexa, e convergentes para um ponto de ∂F.

Exercícios.

a. Prove que um homeomorfismo $f : I \to I$ de um intervalo $I \subset \mathbf{R}$ não tem pontos periódicos com período superior a 2. Quando tem pontos periódicos de período 2 ?

(Se o homeomorfismo é crescente então nenhum ponto pode ter período superior a 1. De fato, as trajetórias são monótonas, portanto ou $f(x) = x$, ou $f^{n+1}(x) > f^n(x) > ... > x$ para todo $n \geq 1$, ou $f^{n+1}(x) < f^n(x) < ... < x$ para todo $n \geq 1$. Seja agora f um homeomorfismo decrescente. Não é difícil ver que f tem um, e um único, ponto fixo p, e que p divide I em dois subintervalos I_- e I_+ que são permutados pela transformação f. Observe também que, se f é decrescente, então f^2 é crescente. Seja $x \neq p$ tal que $f^2(x) \neq x$. Então as sucessões $\left(f^{2n}(x)\right)$ e $\left(f^{2n+1}(x)\right)$ são estritamente monótonas e estão em "lados" distintos de p, i.e. uma em $I_\pm$ e a outra em $I_\mp$ ...)

b. Sejam $I \subset \mathbf{R}$ um intervalo compacto e $f : I \to I$ uma função contínua e crescente. Prove que a trajetória de cada ponto de I converge para um ponto fixo de f. Discuta a dinâmica de f.

c. Estude também a dinâmica de uma transformação contínua e decrescente $f : I \to I$ definida num intervalo compacto $I \subset \mathbf{R}$.

(Observe que, se f é decrescente, então f^2 é crescente...)

Desafio. Sejam I um intervalo da reta real, e $f : I \to I$ e $g : I \to I$ dois homeomorfismos de I sem pontos periódicos. Prove que f e g são topologicamente conjugados.

Problema/curiosidade: teorema de Sharkovskii. A ordem da reta também influi na distribuição dos períodos das órbitas periódicas. Um resultado de Alexander N. Sharkovskii diz que *existe uma ordem $\prec$ nos naturais*

$$\begin{aligned} 1 &\prec 2 \prec 2^2 \prec 2^3 \prec ... \prec 2^m \prec ... \prec 2^k \cdot (2n-1) \prec ... \\ ... &\prec 2^k \cdot 3 \prec ... \prec 2 \cdot 3 \prec ... \prec 2n-1 \prec ... \prec 9 \prec 7 \prec 5 \prec 3 \end{aligned}$$

tal que, se uma função contínua $f : \mathbf{R} \to \mathbf{R}$ tem uma órbita periódica de período k e se $j \prec k$, então também tem uma órbita periódica de período j. Em particular, a existência de uma órbita de período 3 implica a existência de órbitas de todos os períodos!

5.6 Análise local, linearização

A diferenciabilidade das transformações e o princípio das contrações ajudam a comprender as trajetórias dos pontos numa vizinhança dos pontos periódicos.

Pontos fixos atrativos. Sejam $f : V \to V$ uma transformação de classe $\mathcal{C}^1$ definida num aberto $V \subset \mathbf{R}^n$, e $p \in V$ um ponto fixo de f.

Teorema. *Se $|f'(p)| < 1$, então p é "atrativo", ou seja admite uma vizinhança B tal que $f^n(x) \to p$ para todo $x \in B$.*

dem. Pela continuidade de f', existem $\lambda < 1$ e uma bola $B = B_\varepsilon(p)$ centrada em p tais que $|f'(x)| < \lambda$ para todo $x \in B$. O teorema do valor médio implica que $f\left(\overline{B}\right) \subset \overline{B}$, pois se $d(x,p) \le \varepsilon$ então

$$d(f(x), p) \le \lambda \cdot d(x,p) < \varepsilon$$

e que $d(f(x), f(x')) \le \lambda \cdot d(x,x')$ se $x, x' \in \overline{B}$. Portanto, $f|_{\overline{B}} : \overline{B} \to \overline{B}$ é uma contração, e o princípio das contrações diz que a trajetória de todo ponto de $\overline{B}$ converge para p. O resultado é que $B \subset W^s(p)$, i.e. a bacia de atração de p é uma vizinhança de p. $\Box$

Pontos fixos repulsivos. A ordem da reta real permite codificar um comportamento "oposto" à atratividade. Sejam $f : I \to I$ uma transformação de classe $\mathcal{C}^1$ definida num intervalo I da reta real, e $p \in I$ um ponto fixo de f.

Teorema. *Se $|f'(p)| > 1$, então p é "repulsivo", i.e. admite uma vizinhança B tal que a trajetória de todo ponto $x \in B$ distinto de p sai da vizinhança em tempo finito.*

dem. Pela continuidade de f', existem $\lambda > 1$ e um intervalo $B = [p - \varepsilon, p + \varepsilon]$ centrado em p tais que $|f'(x)| > \lambda$ para todo $x \in B$. Observem também que f é estritamente crescente ou decrescente em B, dependendo do sinal de $f'(p)$, e portanto envia bijetivamente intervalos em intervalos. Agora, seja $x \in B$ um ponto diferente de p, e suponhamos que $f^k(x) \in B$ para todo tempo $0 \le k \le n$. A regra da cadeia implica que as derivadas das f^k em x crescem exponencialmente, pois

$$\left|\left(f^k\right)'(x)\right| = \left|f'\left(f^{k-1}(x)\right)\right| \cdot \left|f'\left(f^{k-2}(x)\right)\right| \cdot ... \cdot |f'(x)| > \lambda^k$$

para todo $k \le n$. O teorema do valor médio implica que n não pode ser arbitrariamente grande, porque

$$d(p, f^n(x)) \ge \lambda^n \cdot d(p,x) \qquad \text{e} \qquad d(p, f^n(x)) \le \varepsilon$$

são incompatíveis quando n é grande. Portanto, existe um tempo $n \ge 1$ tal que $f^n(x) \notin B$. $\Box$

Cuidado: este resultado é "local". A condição $|f'(p)| > 1$ não contém informação acerca da bacia de atração de p.

Também, a condição $|f'(p)| > 1$ não é suficiente para decidir a repulsividade de um ponto fixo em dimensão maior, pois podem existir "direções" onde f estica as distâncias e outras onde f reduz as distâncias...

Exercícios.

a. Dê exemplos que mostram que o conjunto estável de um ponto fixo repulsivo p pode conter estritamente $\{p\}$.

b. Procure uma boa definição de *órbita periódica atrativa.*

(Observe que, se $\left\{p, f(p), ..., f^{n-1}(p)\right\}$ é uma órbita, a derivada de f^n é a mesma em todos os seus pontos pela regra da cadeia. Se $\left|(f^n)'(p)\right| < 1$, então p é um ponto fixo atrativo da transformação f^n, e portanto existe uma vizinhança B de p tal que $f^{kn}(x) \to_{k\to\infty} p$ para todo $x \in B$. Então $B \cup f^{-1}(B) \cup ... \cup f^{-(n-1)}(B)$ é uma vizinhança da órbita periódica, e as trajetórias dos seus pontos são assimptóticas à órbita de p ...)

c. Se p é um ponto fixo de $f : \mathbf{R} \to \mathbf{R}$ tal que $f'(p) = 1$, então tudo ou quase tudo pode aconteçer! O conjunto estável de p pode ser uma vizinhança de p, pode ser só $\{x\}$, ou pode conter uma "meia vizinhança" de p, um intervalo do género $[p, p \pm \varepsilon[$ ...

Estude os exemplos

$$x \mapsto x \pm x^3 \qquad \text{e} \qquad x \mapsto x \pm x^2$$

e invente outros.

Exercício/experiência: a família quadrática. A *família quadrática*, ou *logística* (do francês "logement"[5]), é a família de transformações $f_\lambda : [0,1] \to [0,1]$ definidas por

$$x \mapsto \lambda x(1-x)$$

onde o parâmetro λ tem valores no intervalo $[0,4]$.

Os pontos fixos de f_λ são 0, que é atrativo quando $0 \le \lambda < 1$, e $p_\lambda = \frac{\lambda-1}{\lambda}$, que é atrativo quando $1 < \lambda < 3$.

Se $\lambda \in [0,1]$ então toda trajetória $(f_\lambda^n(x))$ converge para 0. De fato, toda trajetória é uma sucessão decrescente e limitada, logo convergente, e o limite é o único ponto fixo 0.

Se $\lambda \in]1,3]$ então toda trajetória $(f_\lambda^n(x))$ converge para p_λ. De fato, se $1 < \lambda < 3$, existe uma vizinhança V de p_λ tal que $f_\lambda|_V$ é uma contração e tal que para todo $x \in [0,1]$ existe um tempo $n \ge 0$ tal que $f_\lambda^n(x) \in V$. O caso em que $\lambda = 3$ não é muito diferente ...

Compreender a dinâmica das transformações f_λ com $3 < \lambda \le 4$ é muito mais difícil. Se $3 < \lambda < 4$ a transformação f_λ tem órbitas periódicas de períodos 2, ou 2^2, ou 2^3, ... e quando $\lambda = 4$ a estrutura das trajetória é tão complicada que parece muito difícil fazer previsões! Uma ideia é fazer "simulações". Podem programar um computador e mandar "calcular" os primeiros 10^n pontos da trajectória de um ponto x escolhido ao acaso, esquecer os primeiros 0.9×10^n pontos da trajectória, e depois mandar desenhar os pontos do conjunto

$$\Omega_{f_\lambda}(x) = \left\{ f_\lambda^{0.9\times 10^n}(x), f_\lambda^{0.9\times 10^n+1}(x), ..., f_\lambda^{10^n}(x) \right\}$$

Dentro dos limites da precisão do computador, o conjunto $\Omega_{f_\lambda}(x)$ pode ser considerado uma aproximação do conjunto ω-limite de x, $\omega_{f_\lambda}(x)$. Façam isto para diferentes valores do parâmetro λ e produzam um gráfico de $\Omega_{f_\lambda}(x)$ versus λ ...

Problema: método de Newton, iteração de funções racionais e linearização conforme. O "método de Newton" é um método proposto por Joseph Raphson em 1690 para aproximar raízes de um polinómio $P : \mathbf{R} \to \mathbf{R}$ (o Newton só queria era resolver $x^3 - 2x - 5 = 0$). Consiste em "adivinhar" uma aproximação razoável x_0 de uma raiz, e depois seguir a sucessão (x_n) definida indutivamente por

$$x_{n+1} = x_n - \frac{P(x_n)}{P'(x_n)}$$

Em 1879 Cayley observou que o método pode ser utilizado também para aproximar raízes complexas de polinómios $P : \mathbf{C} \to \mathbf{C}$. O problema é decidir quando, ou seja para quais valores do "chute" inicial x_0, a sucessão (x_n) converge para uma raiz p. Na nossa linguagem, o método consiste em iterar a transformação racional $x \mapsto R(x) = x - P(x)/P'(x)$. Os zeros de P são pontos fixos atrativos de R, e portanto o problema agora é derminar a bacia de atração dos p.

Seja agora $R : \overline{\mathbf{C}} \to \overline{\mathbf{C}}$ uma função racional arbitrária definida na esfera de Riemann $\overline{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$. Cada ponto fixo p de R tem uma sua bacia de actração B_p. A procura de metodos rápidos para calcular iteradas, em 1871 E. Schrœder introduziu a ideia de procurar conjugações conformes da restrição $R\big|_{B_p}$ com funções racionais mais simples, do género $f : z \mapsto \lambda z$ numa vizinhança B da origem. O método consiste em resolver a equação funcional $h \circ R\big|_{B_p} = f \circ h$, onde $h : B_p \to B$ é uma função analítica. E. Schrœder, G. Kœnig e J.H. Poincaré trataram o problema com $|\lambda| \ne 1$, e depois Carl S. Siegel resolveu o caso $|\lambda| = 1$ por volta de 1940.

[5] A família quadrática nasceu como modelo de dinâmica de populações. Num meio ambiente ilimitado, uma população tem crescimento exponencial, e um modelo razoável a tempo discreto é $z_{n+1} = \lambda z_n$, onde z_n é o tamanho da população no tempo n e $\lambda > 0$ é um parâmetro que caracteriza a "fertilidade" da especie. Se o meio ambiente (a disponibilidade de espaço e comida) é limitado, parece razoável acrescentar um termo negativo $-\beta z_n^2$, onde $\beta > 0$, que toma conta da mortalidade devida à falta de recursos (a probabilidade de dois individuos estar num mesmo sitio, e portanto disputar a comida, é proporcional a z_n^2). O modelo é portanto $z_{n+1} = \lambda z_n - \beta z_n^2$. A maior população suportada pelo meio ambiente resulta ser λ/β e, se chamamos $x_n = z_n \beta/\alpha$ a população relativa, obtemos o modelo $x_{n+1} = \lambda x_n (1 - x_n)$.

5.7 Transversalidade e persistência dos pontos fixos

Sejam $f : I \to \mathbf{R}$ uma transformação de classe $\mathcal{C}^1$ definida num intervalo $I \subset \mathbf{R}$, e p um ponto fixo de f. Se $f'(p) \neq 1$, então o ponto fixo p é "isolado", ou seja admite uma vizinhança B tal que p é o único ponto fixo de f em B. De fato, um ponto fixo é uma solução da equação

$$F(x) = f(x) - x = 0$$

Se $f'(p) \neq 1$ então $F'(p) \neq 0$. O teorema da função inversa diz então que F é invertível numa vizinhança B de p, e isso implica que p é o único zero de F em B. Os pontos fixos que satisfazem a condição $f'(p) \neq 1$ são ditos *transversais*, porque a tagente ao gráfico $\operatorname{graph}(f) = \{(x,y) \text{ t.q. } y = f(x)\}$ de f em p é transversal ao gráfico da função identidade, a reta $\{(x,y) \text{ t.q. } y = x\}$.

A condição $f'(p) \neq 1$ é uma condição aberta, e isto faz suspeitar que também seja estável por pequenas perturbações de f.

Teorema. *Sejam $f : I \to \mathbf{R}$ uma transformação de classe $\mathcal{C}^1$, e p um ponto fixo transversal de f. Toda transformação $g : I \to \mathbf{R}$ suficientemente $\mathcal{C}^1$-próxima de f tem um, e um único, ponto fixo, também transversal, numa vizinhança de p.*

dem. Seja $g = f - h$ uma perturbação de f, com $\|h\|_{\mathcal{C}^1} < \delta$. Um ponto fixo de g é uma solução da equação $g(x) - x = 0$, ou seja da equação

$$F(x) = h(x)$$

onde definimos $F(x) = f(x) - x$. Sabemos que F é invertível numa vizinhança B' de p, logo um ponto fixo de g em B' é uma solução de $x = \left(F^{-1} \circ h\right)(x)$, ou seja um ponto fixo de $F^{-1} \circ h$. A estratégia é provar que $F^{-1} \circ h$ é uma contração numa vizinhança de p. Se a vizinhança $B = \overline{B_r(p)}$ é suficientemente pequena, a inversa de F tem derivada limitada, por exemplo $\left|\left(F^{-1}\right)'(x)\right| < \lambda$ em $F(B)$. Se δ é suficientemente pequeno, a derivada $\left|\left(F^{-1} \circ h\right)'(x)\right| < \lambda \cdot \delta$ é uniformemente < 1 em B, e portanto $F^{-1} \circ h$ tem boas chances de ser uma contração. O que falta verificar é que a imagem $\left(F^{-1} \circ h\right)(B)$ seja contida em B. Ora, dado $x \in B$, a desigualdade do triângulo, o teorema do valor médio e a regra da cadeia, implicam que

$$\begin{aligned} d\left(\left(F^{-1} \circ h\right)(x), p\right) &\leq d\left(F^{-1}(h(x)), F^{-1}h(p)\right) + d\left(F^{-1}(h(p)), p\right) \\ &\leq d\left(F^{-1}(h(x)), F^{-1}h(p)\right) + d\left(F^{-1}(h(p)), F^{-1}(0)\right) \\ &\leq \lambda \cdot \delta \cdot r + \lambda \cdot \delta \end{aligned}$$

(onde utilizamos o fato de que p é um ponto fixo de f) e esta quantidade é $< r$ se δ é suficientemente pequeno. O princípio das contrações enfim implica que um ponto fixo $p' \in B$ de g existe e é único. A derivada de g neste ponto está δ-próxima da derivada de f em p, e isto implica a transversalidade de p' se δ é pequeno. □

Exercícios.

a. Seja $f : \mathbf{R} \to \mathbf{R}$ uma transformação de classe $\mathcal{C}^1$, e seja p um ponto periódico de período n tal que $(f^n)'(p) \neq 1$. Toda transformação g suficientemente $\mathcal{C}^1$-próxima de f tem um ponto periódico de período n próximo de p. (Repita a demonstração anterior com f^n em vez de f)

b. Sejam $f : V \to \mathbf{R}^n$ uma transformação de classe $\mathcal{C}^1$ definida num aberto $V \subset \mathbf{R}^n$, e p um ponto fixo de f. A transversalidade de p se traduz na condição de que o operador $f'(p)$ não tenha 1 como autovalor. Prove que se o operador $f'(p)$ não tem 1 como autovalor, então o ponto fixo p é "isolado", existe uma vizinhança B de x tal que x é o único ponto fixo de f em B. Enuncie e prove um resultado de "persistência" análogo ao caso da reta.

Problema: bifurcações. Os pontos fixos não transversais não são persistentes, em presença de perturbações genéricas podem desaparecer ou mudar de natureza. Este fenómeno é chamado bifurcação. A ideia da teoria das bifurcações é tratar famílias de transformações f_λ definidas numa vizinhança de um ponto fixo, e descrever as possíveis mudanças da dinâmica ao variar o parâmetro λ.

Considere a família de transformações

$$f_\lambda(x) = x + x^2 - \lambda$$

definidas na reta real. A origem é um ponto fixo não transversal de f_0. Se $\lambda \neq 0$ é pequeno, então f_λ tem dois pontos fixos $\pm\sqrt{\lambda}$, um repulsivo e outro atrativo, quando $\lambda > 0$, ou nenhum quando $\lambda < 0$. A família

$$f_\lambda(x) = x + x^3 + \lambda x$$

mostra um comportamento diferente.

O problema é decidir quais fenomenos são "genéricos", e possívelmente "estáveis" num sentido a precisar. Admitindo a existência de um número suficiente de derivadas parciáis contínuas, uma família arbitrária de transformações tais que a origem seja um ponto fixo não transversal de f_0 é da forma

$$\begin{aligned} f_\lambda(x) &= a_\lambda + b_\lambda x + c_\lambda x^2 + ... \\ &= \left(a'\lambda + a''\lambda^2 + ...\right) + \left(1 + b'\lambda + b''\lambda^2 + ...\right)x + \left(c + c'\lambda + c''\lambda^2 + ...\right)x^2 + ... \end{aligned}$$

O caso genérico é quando $c \neq 0$ (ou seja f_0 é da forma $x + cx^2 + ...$) e uma perturbação genérica tem $a' \neq 0$ (ou seja o termo constante de f_λ é diferente de zero desde que $\lambda \neq 0$). Não é dificil convercer-se que o comportamento qualitativo desta família é o mesmo da família $x + x^2 - \lambda$: uma pequena perturbação de f_0 pode destruir o ponto fixo, numa direção, ou criar dois novos pontos fixos, na outra direção. Enuncie este resultado, e dê uma demonstração formal. Observe que procurar raizes da equação $f_\lambda(x) = x$, em função de λ, é equivalente a definir funções $\lambda \mapsto x(\lambda)$ que verifiquem $G(\lambda, x) = f_\lambda(x) - x = 0$, e a este problema responde o teorema da função implícita.

Problema: duplicação do período e cascata de Feigenbaum. Também interessante é o caso de uma família f_λ de transformações do intervalo tal que f_0 tenha um ponto fixo em 0 com $f_0'(0) = -1$. Observe que este ponto fixo é transversal, logo persistente. Por outro lado, $(-1)^2 = 1$, e portanto a derivada de f_0^2 em 0 é igual a $\left(f_0^2\right)'(0) = 1$. Isto diz que 0 não é transversal em quanto ponto fixo de f_0^2. Uma perturbação de f_0 pode produzir pontos periódicos de periodo 2, em proximidade do ponto fixo persistente 0. Para ver um exemplo, considere o caso da família

$$f_\lambda(x) = -x + x^2 + \lambda x$$

Este tipo de bifurcação é dito "duplicação do período". Ao fazer simulações num computador, Mitchell J. Feigenbaum descubriu nos anos '70 que certas famílias de transformações produzem uma "cascata" de duplicações do período, no sentido em que existe uma sucessão $\lambda_1 < \lambda_2 < ... < \lambda_n < \lambda_{n+1}...$ de valores do parâmetro λ tal que, ao passar λ_{n+1} nascem órbita de período 2^{n+1} em proximidade das órbitas de período 2^n criadas pelo valor anterior λ_n. Este fenómeno pode ser facilmente observado com a ajuda de um computador. Aliás, parece que aconteça para toda família em que podemos pensar, desde que acertamos o ponto certo onde centrar uma lupa e ve-lo. Ainda mais misterioso é o fato, também observado por Feigenbaum, de que o limite $\lambda_\infty = \lim_{n\to\infty} \lambda_n$ parece existir, é exponencial, i.e. $|\lambda_\infty - \lambda_n| \simeq \text{const} \times \delta^{-n}$ onde

$$\delta = \lim_{n\to\infty} \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n}$$

e que $\delta \simeq 4.669201609102990671853...$ independentemente da família f_λ! O misterio só foi "explicado" mais tarde por Lanford, Epstein, Dennis Sullivan...

5.8 Hiperbolicidade

Problema: transformações lineares, hiperbolicidade. As transformações lineares dos espaços euclidianos têm dinâmicas simples, e fornecem modelos para o comportamento de uma transformação diferenciável numa vizinhança de um ponto fixo. Um bom exercício é procurar entender a dinâmica de uma transformação linear do plano $T : \mathbf{R}^2 \to \mathbf{R}^2$, definida por uma matriz real 2×2. A ideia é esboçar o "retrato de fase" da transformação, ou seja descrever algumas trajetórias típicas, pelo menos uma por cada comportamento possível. Começe por estudar uma transformação diagonalizável

$$T = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$$

e reparar as diferenças no comportamento qualitativo das trajetórias ao variar os autovalores λ e μ. Ajuda observar que as curvas $x^{\log \mu} = k \cdot y^{\log \lambda}$ são invariantes. Depois, estude o caso

$$T = \lambda \cdot \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

ao variar λ e θ. Escreva a transformação em coordenadas polares (r, φ), e observe que as curvas $r \cdot \lambda^{\varphi/\theta} = k$ são invariantes. Enfim, trate o caso geral, e discuta quando duas transformações lineares são topologicamente conjugadas.

Seja agora $T : \mathbf{R}^n \to \mathbf{R}^n$ uma transformação linear arbitrária ...

Problema: linearização e hiperbolicidade. Seja $f : V \to \mathbf{R}^n$ uma transformação de classe $\mathcal{C}^1$ definida num aberto $V \subset \mathbf{R}^n$, e seja $p \in V$ um ponto fixo de f. A derivada $f'(p)$ é a transformação linear $L : \mathbf{R}^n \to \mathbf{R}^n$ que "melhor aproxima" $f(x) - p$ numa vizinhança de p. Isto sugere que tal vez seja possível que a dinâmica de f numa vizinhança B de p seja "igual" à dinâmica de L numa vizinhança A da origem. Tecnicamente, o problema é decidir se a restrição $f|_B : B \to B'$ é topologicamente conjugada à restrição $L|_A : A \to A'$, por meio de uma "conjugação local" $h : B \cup B' \to A \cup A'$. Em geral, a resposta é negativa. É possível dizer algo quando p é um ponto fixo *hiperbólico*, ou seja quando L não tem autovalores λ com módulo $|\lambda| = 1$. O resultado importante é o "teorema de Grobman-Hartman", que diz que um difeomorfismo local $f : V \to \mathbf{R}^n$ com um ponto fixo hiperbólico p é localmente topologicamente conjugado à sua parte linear $f'(p)$.

6 Statistical description of orbits

Together with the topological point of view, a source of informations about dynamical systems is their statistical description. The idea is to measure the relative size of those points whose orbits have certain definite properties. This is done looking for invariant probability measures, and the main result is the Birkhoff-Khinchin ergodic theorem. To state and prove the Birkhoff-Khinchin ergodic theorem, we need to recall many standard facts and results of integration theory. You can find most of them in the classical manuals by W. Rudin, *Real and complex analysis*, McGraw-Hill, New York 1966, or by P. Halmos, *Measure theory*, Springer-Verlag. New York 1974.

6.1 Probability measures

A *measurable space* is a pair $(X, \mathcal{E})$, a non-empty set X together with a σ-algebra of subsets $\mathcal{E}$. Recall that a (Boolean) algebra is a nonempty family $\mathcal{A}$ of subsets of X which contains X, which contains the complement of any of its elements, and which is closed under finite unions and intersections. A σ-algebra is an algebra which is also closed under countable unions and intersections. Given any family $\mathcal{C}$ of subsets of X, there exists a minimal σ-algebra $\sigma(\mathcal{C})$ which contains all the elements of $\mathcal{C}$, which is called the σ-algebra generated by $\mathcal{C}$.

If (X, τ) is a topological space, the *Borel σ-algebra* is $\sigma(\tau)$, the smallest σ-algebra which contains all open sets.

A *measure* on the measurable space $(X, \mathcal{E})$ is a σ-additive function $\mu : \mathcal{E} \to [0, \infty]$ such that $\mu(\emptyset) = 0$. Here σ-additivity means that, if (S_n) is a countable family of pairwise disjoint elements of $\mathcal{E}$, then

$$\mu(\cup_n S_n) = \sum_n \mu(S_n)$$

The triple $(\Omega, \mathcal{E}, \mu)$ is said a *measure space*, or *probability space* if it happens that $\mu(X) = 1$. Given a probability space, measurable sets $A \in \mathcal{E}$ are commonly called "events", and the number $\mu(A)$ is called "probability of the event A". Basic properties of probability measures are the following: probability measures are monotone, i.e. $\mu(S) \leq \mu(T)$ if $S \subset T$, and σ-subadditive, i.e. if (S_n) is a countable family of elements of $\mathcal{E}$ then

$$\mu(\cup_n S_n) \leq \sum_n \mu(S_n)$$

Probability measures are continuous from below and from above, in the following sense: if $S_n \uparrow S$ then $\mu(S_n) \uparrow \mu(S)$, and if $S_n \downarrow S$ then $\mu(S_n) \downarrow \mu(S)$. Both continuity properties are equivalent, and indeed a simple argument shows that they are equivalent to continuity from above at $\emptyset$: if $S_n \downarrow \emptyset$ then $\mu(S_n) \downarrow 0$. Moreover, continuity is equivalent to σ-aditivity if the set function μ is only assumed (finitely) additive.

A subset $E \subset X$ *has zero measure* if it is contained in a measurable set $S \in \mathcal{E}$ with $\mu(S) = 0$. If any set with zero measure belongs to $\mathcal{E}$, then the measure space $(X, \mathcal{E}, \mu)$ is said *complete*. Any measure space can be canonically completed, extending the measure to the σ-algebra $\overline{\mathcal{E}}$ made of $\mathcal{E}$ and of subsets of zero measure. A property (like continuity of a function, or convergence of a sequence of functions) holds *μ-a.e.* ("almost everywhere" with respect to the measure μ) if the set of points of X where it does not hold has zero measure.

Construction of probability measures. Measures are never "explicitely" given as functions on a σ-algebra. A set function $\mu : \mathcal{P}(X) \to [0, \infty]$ is an *exterior measure* if it is monotone, σ-subadditive, and if $\mu(\emptyset) = 0$. It happens that, given an exterior measure μ, the family of *μ-measurable sets*, defined as

$$\mathcal{E} = \{E \subset X \text{ such that } \mu(S) = \mu(S \cap E) + \mu(S \cap E^c) \text{ for any } S \subset X\}$$

is a σ-algebra, and that μ is a complete measure if restricted on $\mathcal{E}$ (the proof is quite long and delicate, but the only idea it uses is the following: in order to check that $E \in \mathcal{E}$ it is indeed sufficient, by virtue of monotonicity and subadditivity of μ, to check that $\mu(S) \geq \mu(S \cap E) + \mu(S \cap E^c)$ for any $S \subset X$). A strategy to construct interesting measures on uncountable spaces is: start with an exterior measure (it is very easy to produce exterior measures, for example by means of variational principles) and then check that the σ-algebra of measurable sets is sufficiently big for our purpose.

The idea of Carathéodory is the following. A *probability measure* on an algebra $\mathcal{A}$ of subsets of X is an additive function $m : \mathcal{A} \to [0,1]$ such that $m(\emptyset) = 0$, $m(X) = 1$, and such that $A_n \downarrow \emptyset$ implies $m(A_n) \downarrow 0$. Given a probability measure m on a algebra $\mathcal{A}$, the recipe

$$\mu(S) = \inf \left\{ \sum m(A_n) \text{ with } S \subset \cup_n A_n \text{ e } A_n \in \mathcal{A} \right\}$$

defines an exterior measure on $\mathcal{P}(X)$, hence the above construction produces a measure μ on the σ-algebra of μ-measurable sets, which contains $\mathcal{A}$ and so contains $\sigma(\mathcal{A})$. One then checks that $\mu(A) = m(A)$ for any $A \in \mathcal{A}$, so that μ is an "extension" of the measure m. Carathéodory's extension theorem is then stated in the following form:

Carathéodory's extension theorem. *Given a probability measure m on a algebra $\mathcal{A}$ of subsets of X, there exists a unique measure μ on $\sigma(\mathcal{A})$ which extends m.*

The following corollary of Carathéodory's theorem is also useful, for example when trying to prove that some event has a definite probability.

Approximation theorem. *Let $(X, \mathcal{E}, \mu)$ be a probability space, and let $\mathcal{A}$ be an algebra of subsets of X such that $\sigma(\mathcal{A}) = \mathcal{E}$. Then, for any $A \in \mathcal{E}$ and any $\varepsilon > 0$, we can find a $A_\varepsilon \in \mathcal{A}$ such that*

$$\mu(A \Delta A_\varepsilon) < \varepsilon.$$

Indeed, one easily sees that the family $\mathcal{C} = \{A \in \mathcal{E} \text{ s.t. } \forall \varepsilon > 0 \; \exists A_\varepsilon \in \mathcal{A} \text{ s.t. } \mu(A \Delta A_\varepsilon) < \varepsilon\}$ is a σ-algebra. Since $\mathcal{A}$ is obviously contained in $\mathcal{C}$, this implies that $\mathcal{E} = \sigma(\mathcal{A}) \subset \mathcal{C} \subset \mathcal{E}$.

Lebesgue measure. The collection $\mathcal{I}$ of intervals of the real line is a *semi-algebra*, i.e. the intersection of two elements of $\mathcal{I}$ is in $\mathcal{I}$ and the complement of an element of $\mathcal{I}$ is a union of elements of $\mathcal{I}$. The function $m : \mathcal{I} \to [0, \infty]$, defined as $m([a,b]) = |b-a|$ if a e b are finite, and ∞ if the interval is unbounded, is monotone and gives value zero to the empty set. Postulating additivity, the function m extends to a measure on the algebra $\mathcal{A}$ made of disjoint unions of elements of $\mathcal{I}$ (this is not trivial!, the proof uses the Heine-Borel theorem about compact subsets of the real line). The function $\mu : \mathcal{P}(\mathbf{R}) \to [0, \infty]$, defined as

$$\mu(E) = \inf \left\{ \sum m(C_n) \text{ with } E \subset \cup_n C_n \text{ e } C_n \in \mathcal{A} \right\}$$

is then an exterior measure on the real line. The σ-algebra $\mathcal{L}$ of μ-measurable sets, called *Lebesgue σ-algebra*, contains the Borel sets, because it contains the intervals. The restriction $\ell = \mu|_{\mathcal{L}}$, as well as $\mu|_{\mathcal{B}(\mathbf{R})}$, is called *Lebesgue measure.*

Observe that Lebesgue measure on the real line is not a probability measure, having infinite mass. Nevertheless, one can easily define probability measures on bounded intervals taking normalized restrictions of Lebesgue measure. For example, take $X = [0,1]$, and $\mathcal{E} = \mathcal{B}(X) = \{X \cap B \text{ with } B \in \mathcal{B}(\mathbf{R})\}$, the Borel subsets of the interval. The restriction of ℓ to $\mathcal{E}$ is a probability measure, called Lebesgue measure on the unit interval.

The very same construction works in $\mathbf{R}^n$, starting with the semi-algebra of "rectangles" measured by the "euclidean volume", and produces a measure ℓ on $\mathcal{B}(\mathbf{R}^n)$, also called Lebesgue measure. Lebesgue measure is the unique measure over the Borel sets of the euclidean space which is invariant under traslations, i.e. $\ell(\lambda + B) = \ell(B)$ for any $\lambda \in \mathbf{R}^n$ and any Borel set B, and which is normalized to give measure one to the unit square, i.e. $\ell([0,1]^n) = 1$.

The axiom of choice allows one to "give examples" of subsets wich are not Lebesgue-measurable (for example, the set made of one point for each orbit of an irrational rotation of the circle).

Kolmogorov extension. Let X be a finite space, equipped with the discrete topology, and let Σ^+ be the topological product $X^{\mathbf{N}} = \{x : \mathbf{N} \to X\}$, its point indentified with sequences $x = (x_1, x_2 \ldots, x_n, \ldots)$ with $x_n \in X$. Let $\mathcal{C}$ be the collection of *cylinders* of X, the subsets of the form

$$C_B = \left\{ x \in \Sigma^+ \text{ s.t. } (x_1, x_2 \ldots, x_n) \in B \right\}$$

with B an open subset of X^n. Cylinders form a basis of the product topology of Σ^+, which makes Σ^+ a compact metrizable space. In particular, the Borel σ-álgebra of Σ^+ is $\mathcal{B} = \sigma(\mathcal{C})$. Let

$\mu_1, \mu_2, \mu_3...$ be probability measures defined on the Borel sets of $X, X^2, ..., X^n, ...$ respectively. The sequence (μ_n) is said *consistent* if

$$\mu_{n+1}(B \times X) = \mu_n(B)$$

for any n and any Borel subset B of X^n. The (most elementary version of) Kolmogorov extension theorem says that

Kolmogorov extension theorem. *Given a consistent family of probability measures as above, there exists a unique probability measure μ, defined on the Borel σ-algebra of Σ^+, such that*

$$\mu(C_B) = \mu_n(B)$$

for any cilinder C_B.

The proof consists in the following two steps. First, observe that cylinders form an algebra, and use consistency of the μ_n's to verify that the formula above does define a function $\mu : \mathcal{C} \to [0,1]$ on cylinders (i.e. it does not depend on the different ways the same cylinder may be presented) which is additive and properly normalized. Then, use compactness of X to check that μ is continuous at $\emptyset$, in order to apply Carathéodory theorem. Indeed, let (A_n) be a sequence of cilinders such that $A_n \downarrow \emptyset$, and assume by contradiction that $\mu(A_n) > \delta > 0$ for any n. This implies that $A_n \neq \emptyset$ for any n, but, since the A_n are compact, then the Cantor intersection theorem says that $\cap_n A_n \neq \emptyset$, contrary to the hypothesis.

Kolmogorov theorem is the key tool in probability theory, since it allows one to construct measures which describe an infinite sequence of trials starting with some rule which gives information about the n-th trial given the knowledge of the first $n-1$. It actually works with much more general spaces and in a more general setting. Also, one can easily adapt the construction to $\prod_{n\in\mathbf{N}} X_n$, the topological product of a countable family of finite spaces. In some precise sense, this is a universal model of a dynamical system.

If $X = \{0,1\}$, then $\Sigma^+ = X^{\mathbf{N}}$ is the state space of infinite Bernoulli trials with two possible outcomes: sucess and failure. Let $\mu_1 : \mathcal{P}(X) \to [0,1]$ be a any probability measure, defined by $\mu_1(\{1\}) = p$. Kolmogorov construction can be applied postulating the independence of different trials, i.e. declaring that the family formed by the cilinders $\{x_n = 1\}$ is an independent family, and giving measure p to each $\{x_n = 1\}$. The resulting probability space $(\Sigma^+, \mathcal{B}, \mu)$ describes the infinite independent Bernoulli trials. Of course, the very same construction can be made when X is a finite space with any finite numer z of elements.

6.2 Transformations and invariant measures

Measurable transformations. A transformation $f : X \to X$ of the measurable space $(X, \mathcal{E})$ is said *measurable* if $f^{-1}(A) \in \mathcal{E}$ for any $A \in \mathcal{E}$. A measurable transformation f is said an *endomorphism* of the measurable space, or an *automorphism* if it is invertible and its inverse is measurable too.

Observe that an endomorphism f of a measurable space $(X, \mathcal{E})$ acts naturally on the space of measures on $\mathcal{E}$ by "push forward": if μ is a measure, then $f_*\mu$, defined by $(f_*\mu)(A) = \mu\left(f^{-1}(A)\right)$ for any $A \in \mathcal{E}$, is also a measure.

Let f be an endomorphism of the measurable space $(X, \mathcal{E})$. A probability measure μ on $\mathcal{E}$ is *invariant* (w.r.t. the transformation f) if $f_*\mu = \mu$, namely if

$$\mu\left(f^{-1}(A)\right) = \mu(A)$$

for any $A \in \mathcal{E}$. If this happens, we also say that f is an *endomrphism* (resp. an *automorphism*) of the probability space $(X, \mathcal{E}, \mu)$. The meaning of this definition is that "mean values" of integrable observables $\varphi : X \to \mathbf{R}$ with respect to invariant probability measures do not change with time, in the sense that $\int_X \varphi d\mu = \int_X (\varphi \circ f)\, d\mu$.

Given an endomorphism f of the probability space $(X, \mathcal{E}, \mu)$, one says that an event $A \in \mathcal{E}$ is *invariant mod* 0 if $\mu\left(A \bigtriangleup f^{-1}(A)\right) = 0$. The set of invariant mod 0 events form a sub-σ-algebra of $\mathcal{E}$, denoted by $\mathcal{E}_f$.

How to prove that a measure is invariant. The very definition of invariance does not help too much if we want to prove that a certain measure μ on the σ-algebra $\mathcal{E}$ is invariant w.r.t. the measurable transformation $f : X \to X$. The trick is the following. Suppose that we can prove that $\mu\left(f^{-1}(C)\right) = \mu(C)$ for any $C \in \mathcal{C}$, where $\mathcal{C}$ is some subset of $\mathcal{E}$. Caratheodory theorem implies that $f_*\mu$ and μ are the same measure on the σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$. On the other side, the family of those $A \in \mathcal{E}$ such that $\mu\left(f^{-1}(A)\right) = \mu(A)$ is easily seen to be a σ-algebra. Hence, if it happens that $\sigma(\mathcal{C}) = \mathcal{E}$, then μ is actually invariant. In other words, in order to prove that μ is invariant it is sufficient to check that $\mu\left(f^{-1}(C)\right) = \mu(C)$ for any C belonging to a family of subsets of X which generate the σ-algebra $\mathcal{E}$.

Observables as random variables. When dealing with a endomorphism $f : X \to X$ of the probability space $(X, \mathcal{E}, \mu)$, one should consider *measurable* observables $\varphi : X \to \mathbf{R}$ (or $\mathbf{C}$), those functions such that $\varphi^{-1}(A) \in \mathcal{E}$ for any Borel set $A \subset \mathbf{R}$. In the context of probability theory they are called "random variables", and the sequence of observables $\varphi \circ f^n$ may be interpreted as a "random process". If φ is integrable, the Lebesgue integral $\int_X \varphi d\mu$ is interpreted as the "mean value" of φ. Of course, invariance of a measurable observable must be intended modulo sets of zero measure. Then, one can consider the Banach spaces $L^p(\mu)$ of (equivalence classes of) observables equipped with the L^p-norm

$$\|\varphi\|_p = \left(\int |\varphi|^p \, d\mu\right)^{1/p}$$

and use the full power of integration theory to get informations about the dynamical system. In particular, $L^2(\mu)$ is a Hilbert space if equipped with the inner product

$$\langle \varphi, \psi \rangle = \int_X \varphi \overline{\psi} d\mu$$

Conditional mean. Recall that, given a measurable space $(X, \mathcal{E})$, a measure ν is said *absolutely continuous* w.r.t. the measure μ if $\nu(A) = 0$ whenever $\mu(A) = 0$. The following technical result (which may be proved using Hilbert space techniques) is particularly useful:

Radon-Nikodym theorem. *Let $(X, \mathcal{E}, \mu)$ be a probability space, and let ν be a finite measure over $\mathcal{E}$ which is absolutely continuous with respect to μ. Then there exists a nonnegative integrable random variable ρ (called the Radon Nikodym derivative of ν w.r.t. μ and denoted by $d\nu/d\mu$) such that*

$$\nu(A) = \int_A \rho d\mu$$

for any $A \in \mathcal{E}$.

A particularly important tool, taken from the theory of probability, is the conditional mean. Let $(X, \mathcal{E}, \mu)$ be a probability space, and let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{E}$. Given an integrable random variable φ, there exists a unique random variable $\varphi_{\mathcal{F}}$, called the *conditional mean* of φ w.r.t. $\mathcal{F}$, which is $\mathcal{F}$-measurable (i.e. the inverse image of any Borel set belongs to $\mathcal{F}$) and such that

$$\int_A \varphi_{\mathcal{F}} d\mu = \int_A \varphi d\mu$$

for any $A \in \mathcal{F}$. Indeed, if $\varphi \geq 0$, then one can define $\varphi_{\mathcal{F}}$ as equal to the Radon-Nikodym derivative of the measure $A \mapsto \int_A \varphi d\mu$, defined on $\mathcal{F}$, with respect to the restriction $\mu|_{\mathcal{F}}$. The general case is treated by linearity, writing φ as a difference of two non-negative random variables. Uniqueness is intended μ-a.e., i.e. modulo sets of zero probability. The conditional mean is monotone, namely if $\varphi \geq 0$ then $\varphi_{\mathcal{F}} \geq 0$, and preserves the mean value, since $\int_X \varphi_{\mathcal{F}} d\mu = \int_X \varphi d\mu$. It can be considered as a "projection" of φ onto the space of $\mathcal{F}$-measurable random variable, preserving the mean value. In particular, if $\mathcal{N}$ is the trivial σ-algebra made of events of measure 0 or 1, then $\varphi_{\mathcal{N}}$ is constant a.e. and equal to $\int_X \varphi d\mu$.

Topological dynamical systems and Borel measures. If we are interested in the dynamics of a continuous transformation $f : X \to X$ of a topological space X, it is natural to consider the Borel σ-algebra $\mathcal{B}$, the smallest σ-algebra of subsets of X which contain all open sets. The map f

is then an endomorphism of $(X, \mathcal{B})$. Probability measures on $\mathcal{B}$ are said *Borel probability measures.* If, moreover, X is a compact metric space, one can consider the space $\mathcal{C}^0(X, \mathbf{R})$ of bounded continuous real valued functions of X (observe that, since X is compact, any continuous function is automatically bounded), equipped with the sup norm

$$\|\varphi - \psi\|_\infty = \sup_{x \in X} |\varphi(x) - \psi(x)|$$

These observables are clearly integrable w.r.t. to any Borel probability measure μ, and the mean value map

$$\varphi \mapsto \int_X \varphi d\mu$$

is a bounded, positive definite (in the sense that $\int_X \varphi d\mu \geq 0$ for any $\varphi \geq 0$) linear functional on $\mathcal{C}^0(X, \mathbf{R})$. The basic fact about Borel measures is the converse of that, namely

Riesz-Markov representation theorem. *Let X be a compact metric space. Given any bounded and positive definite linear functional L on $\mathcal{C}^0(X, \mathbf{R})$ such that $L(1) = 1$, there exists a unique Borel probability measures μ such that*

$$L(\varphi) = \int_X \varphi d\mu$$

for any $\varphi \in \mathcal{C}^0(X, \mathbf{R})$.

The space of invariant probability measures. The space Prob of probability measures on a measurable space $(X, \mathcal{E})$ has a natural convex structure: convex combinations of probability measures are also probability measures. An arbitrary measurable transformation $f : X \to X$ of a measurable space may not admit any invariant probability measure. On the other side, if μ_0 and μ_1 are invariant probability measures, so are their convex combinations $\mu_t = (1-t)\mu_0 + t\mu_1$ for any $t \in [0, 1]$. This means that the set Prob_f of invariant probabilty measures on $\mathcal{E}$ is a convex set: if it contains two points, it contains the whole segment between them.

Now, let (X, d) be a compact metric space and let $\mathcal{B}$ its Borel σ-algebra. The space Prob of probability measures on $\mathcal{B}$ can be equipped with a natural topology, called the *weak* topology*, which says essentially that two measures are near if they give nearby mean values to some well behaved observables. Formally, one says that a sequence of measures (μ_n) converge weakly* to a measure μ, which we denote simply as $\mu_n \to \mu$, if

$$\int_X \varphi d\mu_n \to \int_X \varphi d\mu$$

for any (bounded) continuous function $\varphi : X \to \mathbf{R}$. The space $\mathcal{C}^0(X, \mathbf{R})$ of bounded continuous real valued functions on X, equipped with the sup norm, is a separable Banach space. In particular, it admits a countable set of points $\{\varphi_n\}_{n \in \mathbf{N}}$ which is dense in its unit sphere. Given that, one defines, for any couple of Borel probability measures μ and ν, a distance

$$d(\mu, \nu) = \sum_{n=1}^{\infty} 2^{-n} \cdot \left| \int_X \varphi_n d\mu - \int_X \varphi_n d\nu \right|$$

It turns out that d is indeed a metric, and that it induces the weak* topology on Prob. The important fact (somewhere called "Helly's theorem"), which follows from the Ascoli-Arzela theorem together with the above Riesz-Markov representation theorem, is that Prob, equipped with the weak* topology, is a compact space: any sequence (μ_n) of Borel probability measures admits a weakly* convergent subsequence $\mu_{n_i} \to \mu$.

Now, we are in position to prove the existence of invariant probability measures for certain well behaved dynamical systems.

Krylov-Bogolyubov theorem. *A continuous transformation $f : X \to X$ of a metrizable compact space X admits at least one Borel invariant probability measure.*

proof. Take any Borel probability measure μ_0 on X, and inductively define a family of probability measures μ_n by $\mu_{n+1} = f_* \mu_n$. Consider the family of Cesaro means

$$\overline{\mu}_n = \frac{1}{n+1} \sum_{k=0}^{n} \mu_k$$

Since the space of Borel probability measures on a compact metrizable space is compact w.r.t. weak* convergence, there exist a weakly* convergent subsequence $\overline{\mu}_{n_i} \to \mu$. One then easily sees that

$$\begin{aligned} \int_X (\varphi \circ f)\, d\mu &= \lim_{i\to\infty} \frac{1}{n_i+1} \sum_{k=0}^{n_i} \int_X (\varphi \circ f)\, d\mu_k \\ &= \lim_{i\to\infty} \frac{1}{n_i+1} \sum_{k=0}^{n_i} \int_X \varphi d\mu_{k+1} \\ &= \lim_{i\to\infty} \frac{1}{n_i+1} \sum_{k=0}^{n_i} \int_X \varphi d\mu_k + \frac{1}{n_i+1} \left(\int_X \varphi d\mu_{n_i+1} - \int_X \varphi d\mu_0 \right) \\ &= \int_X \varphi d\mu \end{aligned}$$

for any bounded continuous observable φ, hence that μ is an invariant measure. □

6.3 Invariant measures and time averages

The relevance of invariant measures when studying the dynamics of continuous transformations is due to the following crucial observations.

Assume that, for a given point $x \in X$, the time averages

$$\overline{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \varphi\left(f^k(x)\right)$$

do exist for any bounded continuos observable φ. One easily shows that the funcional $\mathcal{C}^0_{\mathrm{b}}(X, \mathbf{R}) \to \mathbf{R}$ defined by $\varphi \mapsto \overline{\varphi}(x)$ is linear, bounded and positive definite. There follows from the Riesz-Markov representation theorem that there exists a unique Borel probability measure μ_x on X such that

$$\overline{\varphi}(x) = \int_X \varphi d\mu_x$$

for any $\varphi \in \mathcal{C}^0_{\mathrm{b}}(X, \mathbf{R})$. The invariance property $\overline{\varphi}(x) = (\overline{\varphi} \circ f)(x)$ then implies that $\int_X (\varphi \circ f)\, d\mu_x = \int_X \varphi d\mu_x$ for any φ, hence that μ_x is an invariant probability measure. In the language of physicists, this says that "time averages" along the orbit of x are equal to "space averages" with respect to the measure μ_x.

One is thus lead to consider the following questions. Do there exist points x for which time averages exists? Given an invariant measure μ, do there exist, and how many, points x such that $\mu = \mu_x$?

Example: periodic orbits. Let p be a periodic point with period n. The time average $\overline{\varphi}(p)$ of any observable φ exists, and is equal to the arithmetic mean of φ along the orbit, namely

$$\overline{\varphi}(p) = \frac{1}{n} \sum_{k=0}^{n-1} \varphi\left(f^n(p)\right)$$

If μ_p denotes the normalized sum $\frac{1}{n}\sum_{k=0}^{n-1} \delta_{f^n(p)}$ of Dirac masses placed on the orbit of p, this amount to say that $\overline{\varphi}(p) = \int_X \varphi d\mu_p$.

Let p be a fixed point, and $\varphi : X \to \mathbf{R}$ be an observable which is continuous at p. If $x \in W^s(p)$, then the time average $\overline{\varphi}(x)$ exists and is equal to $\varphi(p)$, i.e. time averages of points in the basin of attraction of p are described by the Dirac measure $\mu_p = \delta_p$.

The Birkhoff-Khinchin ergodic theorem. Ergodic theorems are the milestones of ergodic theory, and deal with various type of convergence of the time means $\overline{\varphi}_n$ for certain classes of observables φ. In particular, the Birkhoff-Khinchin "individual" ergodic theorem must be thougth as the generalization of the Kolmogorov strong law of large numbers, as it says that time means of certain well-behaved observables exist almost everywhere. The Birkhoff-Khinchin ergodic theorem was actually preceeded by the von Neumann's "statistic" ergodic theorem, which says that

von Neumann "statistic" ergodic theorem. *Let U be a unitary operator on a Hilbert space H, let $H_U = \{v \in H \text{ s.t. } Uv = v\}$ denote the closed subspace of those vectors which are fixed by U, and $P_U : H \to H_U$ denote the orthogonal projection onto H_U. Then, for any vector $v \in H$ we have*

$$\lim_{n\to\infty} \left\| \frac{1}{n+1} \sum_{k=0}^{n} U^k v - P_U v \right\|_H = 0$$

If $f : X \to X$ is an endomorphism of the probability space $(X, \mathcal{E}, \mu)$, one can consider the "shift" operator $U : L^2(\mu) \to L^2(\mu)$ given by $(U\varphi)(x) = \varphi(f(x))$. It is clearly unitary, its fixed point set is the space of invariant L^2-observable. The von Neumann theorem then asserts convergence of time means $\overline{\varphi}_n \to \overline{\varphi}$ in $L^2(\mu)$. Here, we prove the

Birkhoff-Khinchin "individual" ergodic theorem. *Let $f : X \to X$ be an endomorphism of the probability space $(X, \mathcal{E}, \mu)$, and let $\varphi \in L^1(\mu)$ be an integrable observable. Then the limit*

$$\overline{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \varphi\left(f^k(x)\right)$$

exists for μ-almost any $x \in X$. Moreover, the observable $\overline{\varphi}$ is in $L^1(\mu)$, is invariant, and satisfies

$$\int \overline{\varphi} d\mu = \int \varphi d\mu$$

proof. (by A. Garsia) Let $\mathcal{E}_f$ be the invariant σ-algebra. For any $\psi \in L^1$, set $\psi_n = \max_{k\le n} \sum_{k=0}^{n-1} \varphi \circ f^k$ and observe that $E_\psi = \{x \in X \text{ s.t. } \psi_n(x) \to \infty\} \in \mathcal{E}_f$. One easily sees that the sequence $\psi_{n+1} - \psi_n \circ f$ is decreasing, and converges to ψ at the points of E_ψ. The monotone convergence theorem and the invariance of μ imply that

$$0 \le \int_{E_\psi} (\psi_{n+1} - \psi_n)\, d\mu = \int_{E_\psi} (\psi_{n+1} - \psi_n \circ f)\, d\mu \to \int_{E_\psi} \psi d\mu = \int_{E_\psi} \psi_{\mathcal{E}_f} d\mu\,|_{\mathcal{E}_f}$$

In particular, if $\psi_{\mathcal{E}_f} < -\varepsilon < 0$ then $\mu(E_\psi) = 0$. On the other side,

$$\limsup \frac{1}{n} \sum_{k=0}^{n-1} \psi \circ f^k(x) \le \limsup \frac{1}{n} \psi_n \le 0$$

on $X \backslash E_\psi$. Applying twice these observations to the observables $\varphi - \varphi_{\mathcal{E}_f} - \varepsilon$ and $-\varphi + \varphi_{\mathcal{E}_f} - \varepsilon$, with $\varepsilon > 0$, we find

$$\limsup \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k(x) - \varphi_{\mathcal{E}_f} - \varepsilon \le 0 \qquad \liminf \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^k(x) - \varphi_{\mathcal{E}_f} + \varepsilon \ge 0$$

μ-almost everywhere. Since ε was arbitrary, the limit $\overline{\varphi}(x)$ exists and is equal to $\varphi_{\mathcal{E}_f}(x)$ for μ-almost every x. The rest of the theorem then follows easily from the properties of the conditional mean. □

6.4 Examples

Haar measures. Any locally compact topological group G admits a *Haar measure*, a measure μ on its Borel sets which is left-invariant, i.e. satisfies $L_g\mu = \mu$ for any $g \in G$. Moreover, the Haar measure is unique up to a constant factor. It is an exercise that μ is a finite measure, hence can be renormalized to give a probability measure, iff G is compact. There follows that translations on compact topological groups admits invariant probability measures.

On the other side, for some groups G, called *unimodular*, the Haar measure μ is both left and right invariant. If $\Gamma \subset G$ is a lattice, i.e. a subgroup such that $\mu(G/\Gamma) < \infty$, then the normalized Haar measure on the homogeneous space G/Γ is an invariant probability measure for any left translation $g\Gamma \mapsto sg\Gamma$.

Rotations of the circle. Lebesgue probability measure ℓ on the circle is invariant for the rotations $+\alpha : x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$, with $\alpha \in \mathbf{R}$. Indeed, rotations of the circle are isometries, and the Lebesgue measure $\ell(I)$ of an interval is its "lenght".

Coverings of the circle. Lebesgue probability measure ℓ on the circle is invariant for the maps $\times\lambda : x + \mathbf{Z} \mapsto \lambda \cdot x + \mathbf{Z}$, with $\lambda \in \mathbf{Z}\backslash\{0\}$. This comes from the fact that the inverse image of a sufficiently small interval I with lenght $\ell(I)$ is the disjoint union of $|\lambda|$ intervals with lenght $\ell(I)/|\lambda|$.

Bernoulli shifts. Consider the Bernoulli shift $\sigma : \Sigma^+ \to \Sigma^+$ over the alphabet $X = \{1, 2, ..., z\}$. Let p be a "probability on X", i.e. a finite set of nonegative numbers $p_1, p_2, ..., p_z$ such that $p_1 + p_2 + ... + p_z = 1$. Given a centered cilinder C_α, we define $\mu(C_\alpha)$ as equal to the product $p_{\alpha_1} p_{\alpha_2} ... p_{\alpha_n}$. This function μ extends in a unique way as a finitely additive function on the algebra $\mathcal{A}$ generated by the centered cilinders, the algebra which contains all finite unions of centered cilinders as well as the empty set and Σ^+. One then show that μ is σ-additive on $\mathcal{A}$ (for example, showing that if a decreasing sequence $A_1 \supset A_2 \supset ...$ has empty intersection then $\mu(A_n) \to 0$). Since centered cilinders generates the topology of Σ^+, Carathéodory theorem implies that there exists a unique extension, which we still call μ, of this measure on the Borel σ-algebra of Σ^+. This measure is called the *Bernoulli measure* defined by p.

As for the "physical" meaning of this measure, you may imagine that X represents the possible outcomes when tossing a coin with z sides, and p_k is the probability of obtaining the k-th side. Then points in Σ^+ represent the outcomes of an infinite sequence of tossings, and the very definition of μ says that each trial is described by the probability p, and each trial is "independent" from any finite collection of different trials.

It is not surprising that μ is indeed an invariant probability measure. This comes from the fact that the inverse image $\sigma^{-1}(A)$ of any $A \in \mathcal{A}$ is the disjoint union of z elements $B_1, B_2, ..., B_z$ of the algebra (obtained from A chosing the first letter in z different ways) with measures $\mu(B_k) = p_k \cdot \mu(A)$, so that

$$\mu\left(\sigma^{-1}(A)\right) = \sum_{k=1}^{z} p_k \cdot \mu(A) = \mu(A)$$

Absolutely continuous invariant measures for maps and flows. Let U be a domain in some euclidean $\mathbf{R}^n$, and let vol denote the Lebesgue measure on U, given locally as $d\mathrm{vol} = dx = dx_1 dx_2 ... dx_n$. A local diffeomorphism $f : U \to U$ of class $\mathcal{C}^1$ preserves the measure vol iff

$$\sum_{x \in f^{-1}\{x'\}} \frac{1}{|\det f'(x)|} = 1$$

for any point $x' \in U$, as one can check using the change of coordinates formula. Also interesting is to see wheather f preserves an absolutely continuous measure $\mu = \rho\mathrm{vol}$, and this happens iff the "density" ρ satisfies the equation

$$\sum_{x \in f^{-1}\{x'\}} \frac{\rho(x)}{|\det f'(x)|} = \rho(x')$$

for any point $x' \in U$.

Now, let ϕ be the flow of a vector field $\xi = \sum_{k=1}^{n} \xi_k \frac{\partial}{\partial x_k}$ on U. The above obviously applies, considering the Jacobian of the diffeomorphisms ϕ_t. Since

$$\det \phi_t' = \int_0^t \mathrm{div}\xi \circ \phi_s ds$$

we get the result that Lebesgue measure vol is invariant under the flow of ξ iff

$$\operatorname{div}\xi = \sum_{k=1}^{n} \frac{\partial \xi_k}{\partial x_k} = 0$$

In general, the absolutely continuous measure $\mu = \rho\text{vol}$ is invariant under the flow of ξ iff its density satisfies $\operatorname{div}(\rho\xi) = 0$.

Hamiltonian flows. Consider a symplectic manifold (X, ω). Liouville measure $d\text{vol} = \omega^n$ is invariant under the Hamiltonian flow of any Hamiltonian function H. If X has finite volume, it can be normalized to give an invariant probability measure.

Geodesic flows. Consider a geodesic flow on the unit tangent bundle $\pi : SM \to M$ of the Riemannian manifold (M, g). Let $d\text{vol} = \sqrt{g}dx$ denote the Riemannian volume form on M, and let $d\sigma_m$ denotes the Lebesgue probability measure on the sphere $S_m M = \pi^{-1}\{m\}$. The Liouville measure μ, defined locally as $d\text{vol}(m) \times d\sigma_m$, is invariant under the geodesic flow.

Gauss map. Any irrational real number $x \in]0, 1]$ has a "continued fraction representation" of the form

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_n + \cfrac{1}{\dots}}}}}$$

where the a_n are nonnegative integers. The equality sign and the "infinite fraction" above mean that the

$$r_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_n}}}}$$

converge to x as $n \to \infty$. The sequence of rationals r_n (any such r_n provide the best rational approximation for x with denominator less or equal than that of r_n, as you may have been teached in a course on number theory) is inductively constructed as follows. First, observe that if $a_1 = [1/x]$ and $x_1 = 1/x - a_1$ we may write

$$x = \frac{1}{a_1 + x_1}$$

with $x_1 \in [0, 1]$. Then, since $x_1 \neq 0$, for otherwise x would be rational, we may define $a_2 = [1/x_1]$ and $x_2 = 1/x_1 - a_2$ to get

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + x_2}}$$

Inductively, we see that

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_n + x_n}}}}$$

where $x_n = 1/x_{n-1} - a_n$ and $a_n = [1/x_{n-1}]$. This amounts to say that the sequence (x_n) is the trajectory of x under the *Gauss map* $g :]0, 1] \to]0, 1]$, defined as

$$x \mapsto 1/x - [1/x]$$

Observe that g is not defined at the origin, hence to iterate g we need to avoid all the preimages of 0, which are the rationals. This is not a problem if we want to study the statistical properties of g with respect to Lebesgue measure, since rationals form a subset of zero measure. The Gauss map admits an absolutely continuous invariant measure $\mu = \rho dx$, defined as

$$\mu(A) = \frac{1}{\log 2} \cdot \int_A \frac{1}{1+x} dx$$

for any Borel subset $A \subset]0, 1]$. The denominator $\log 2$ is there to normalize the measure, so we just have to check the invariance criterium for the density $\rho(x) = 1/(1+x)$. Since any $x' \in]0, 1]$

has one preimage $x_k = 1/(x' + k)$ in each interval $]1/(k+1), 1/k]$, we compute

$$\begin{aligned}
\sum_{x \in g^{-1}\{x'\}} \frac{\rho(x)}{|\det g'(x)|} &= \sum_{k \geq 1} \frac{x_k^2}{1 + x_k} \\
&= \sum_{k \geq 1} \left(\frac{1}{x' + k} - \frac{1}{x' + k + 1} \right) \\
&= \frac{1}{1 + x'} = \rho(x')
\end{aligned}$$

and we are done.

7 Recorrências

O objectivo da dinâmica topológica é a caracterização (topológica!) das possíveis histórias dos pontos. Isso começa a ser interessante quando a órbita de x não é finita.

7.1 Comportamento assimptótico das órbitas infinitas: conjuntos ω e α limite

A coisa mais simples que pode acontecer, é que a trajetória de x seja uma sucessão convergente, e neste caso o seu limite é um ponto fixo.

As trajetórias podem não ser convergentes, mas pelo menos ter subsucessões convergentes. O conjunto *ω-limite* de x é o conjunto dos pontos limites da trajetória de x, ou seja

$$\omega_f(x) = \cap_{n=0}^{\infty} \overline{\cup_{k\geq n}\{f^k(x)\}}$$

o conjunto dos pontos $x' \in X$ tais que existe uma subsucessão $n_i \to \infty$ de tempos tal que $f^{n_i}(x) \to x'$ quando $i \to \infty$. Observe que, se a órbita de x não é finita, então $\omega_f(x) = \mathcal{O}_f^+(x)'$. O conjunto $\omega_f(x)$ é fechado e +invariante.

$\mathrm{Lim}_f = \cup_{x\in X}\omega_f(x)$ denota o conjunto dos pontos ω-limites. Observe que, se x é periódico, então $\omega_f(x)$ é a sua órbita, e portanto

$$\mathrm{Per}_f \subset \mathrm{Lim}_f$$

Se f é invertível, o conjunto *α-limite* de x é definido por $\alpha_f(x) = \omega_{f^{-1}}(x)$, ou seja é o conjunto dos pontos $x' \in X$ tais que existe uma subsucessão $n_i \to \infty$ de tempos tal que $f^{-n_i}(x) \to x'$ quando $i \to \infty$. Neste caso, os conjuntos $\omega_f(x)$ e $\alpha_f(x)$ são fechados e invariantes. $\mathrm{Lim}_{f^{-1}} = \cup_{x\in X}\alpha_f(x)$ denota o conjunto dos pontos α-limites.

Estes conjuntos podem ser vazios. Por exemplo, $\omega_f(x) = \emptyset$ quer dizer que a trajetória de x anda pelo espaço X sem voltar muitas vezes numas vizinhanzas dos pontos que já visitou, e isto é possível desde que X não seja compacto. Se $\mathrm{Lim}_{f^{\pm 1}}$ não forem vazios, têm a interpretação dos conjuntos onde as trajetórias "morrem" e "nascem", respetivamente (donde a notação ω e α).

(X compacto $\Rightarrow \omega_f(x) \neq \emptyset$) Se X é compacto, então a trajetória $(f^n(x))_{n\in\mathbf{N}_0}$ de todo ponto $x \in X$ admite subsucessões convergentes, e portanto $\omega_f(x) \neq \emptyset$. Analogamente, se f é um homeomorfismo, $\alpha_f(x) \neq \emptyset$ para todo ponto $x \in X$. Em particular, os conjuntos $\mathrm{Lim}_{f^{\pm 1}}$ não são vazios.

Exercícios.

a. Prove que $\omega_f(x)$ é fechado e +invariante. Prove que, se f é um homeomorfismo, então $\omega_f(x)$ e $\alpha_f(x)$ são fechados e invariantes.

b. Dê exemplos que mostram que $\omega_f(x)$ e $\alpha_f(x)$ podem ser vazios.

c. Mostre que $\mathrm{Per}_f \subset \mathrm{Lim}_f$.

7.2 Pontos recorrentes

O ponto x é *recorrente* se $x \in \omega_f(x)$. Observe que x é recorrente se, dada uma vizinhança arbitrária B de x, existe um tempo $n \geq 1$ tal que $f^n(x) \in B$, ou seja se a trajetória de x volta a visitar toda vizinhança de x. Isto implica que a trajetória de x passa infinitas vezes numa vizinhança arbitrária de x. Rec_f denota o conjunto dos pontos recurrentes de f.

Se f é um homeomorfismo, também tem interesse o conjunto $\mathrm{Rec}_{f^{-1}}$, o conjunto dos pontos x tais que $x \in \alpha_f(x)$.

Um ponto periódico é recorrente, logo

$$\mathrm{Per}_f \subset \mathrm{Rec}_f$$

Exercícios.

a. Defina uma ordem parcial em X da seguinte maneira: $x \prec x'$ se para todas vizinhanças U de x e V de x' existe um tempo $n \geq 1$ tal que $f^n(U) \cap V \neq \emptyset$. Prove que x é recorrente sse $x \prec x$.

b. Mostre que $\mathrm{Per}_f \subset \mathrm{Rec}_f$.
c. Dê exemplos que mostram que Rec_f e $\mathrm{Rec}_{f^{-1}}$ podem ser vazios.

Pseudo-trajetórias e "chain-recurrent points". Dado $\varepsilon > 0$, a sequência de pontos $\{x_k\}_{k=0,1,...,n}$ com $x_0 = x$ é dita *ε-pseudo-trajetória* de x se

$$d\left(x_{k+1}, f\left(x_k\right)\right) < \varepsilon$$

para todo $k = 0, 1, 2, ..., n-1$. O ponto $x \in X$ é dito *chain-recurrent* se, para todo $\varepsilon > 0$, existe uma ε-trajetória $\{x_k\}_{k=0,1,...,n}$ com $x_0 = x_n = x$. $\mathrm{Rec}_f^{\varepsilon}$ denota o conjunto dos pontos chain-recurrent de f.

7.3 Invariant measures and recurrent points: Poincaré theorem

If f satisfies a condition (natural in physics) like "preserving a probability measure", then there are a lot of recurrent points, actually almost any point is recurrent. If, moreover, the probability measure is diffuse, i.e. any non-empty open set has positive measure, then the set of recurrent points is also dense. These results, discovered by Jules Henri Poincaré around 1890, motivated the modern theory of dynamical systems. They show how weak informations on the transformation f may yeld significative qualitative information about "almost all" orbits of the system. Here follow the precise statements, together with all the necessary technical details. If you don't know the meaning of some words, like "measurable" or "Borel set", don't worry, just try to understand what's going on. Poincaré himself didn't know, yet!

Poincaré recurrence theorem (probabilistic version). *Let $f : X \to X$ be an endomorphism of a probability space $(X, \mathcal{E}, \mu)$, and let $A \in \mathcal{E}$. Then the set*

$$A^{\mathrm{rec}} = \{x \in A \text{ t.q. } f^n(x) \in_{\text{i.o.}} A\}$$

of those points of A whose orbit passes through A infinitely often has total probability, namely $\mu\left(A^{\mathrm{rec}}\right) = \mu\left(A\right)$.

proof. For $k \geq 1$, let $B_k = \{x \in A \text{ s.t. } f^n(x) \notin A \ \forall \ n \geq k\}$ be the set of those points of A which never return in A after $n \geq k$ iterates. Observe that $B_k = A \cap \left(\cap_{n \geq k} f^{-n}\left(X \backslash A\right)\right)$, and that $A^{\mathrm{rec}} = A \backslash \left(\cup_{k \geq 1} B_k\right)$, and this shows in particular that A^{rec} is measurable. One sees that $f^{-nk}(B_k) \cap B_k = \emptyset$ for any $n \geq 1$ (for a point x in the intersection would be a point of B_k with $f^{kn}\left(x\right) \in A$ for some $kn \geq k$), and this implies that $f^{-nk}(B_k) \cap f^{-mk}(B_k) = \emptyset$ for any $n > m \geq 0$. The sets $f^{-nk}(B_k)$ all have the same measure $\mu\left(B_k\right)$, because μ is invariant, and are pairwise disjoint. This implies that $\mu\left(B_k\right) = 0$, because

$$\sum_{n \geq 1} \mu\left(B_k\right)) = \sum_{n \geq 1} \mu\left(f^{-nk}(B_k)\right) = \mu\left(\cup_{n \geq 1} f^{-nk}(B_k)\right) \leq \mu\left(X\right) = 1$$

and so $\mu\left(A^{\mathrm{rec}}\right) = \mu\left(A\right)$. $\square$

Now, let $f : X \to X$ be a continuous transformation of a metrizable topological space X, and let μ be an invariant Borel probability measure. If X admits a countable basis $(U_i)_{i \in \mathbf{N}}$, we can apply the above theorem to every open set U_i. This easily implies that the set of recurrent points has full measure, i.e.

$$\mu\left(\mathrm{Rec}_f\right) = 1$$

In particular, since any set of full measure is dense in the support of a Borel measure, we get the following general result.

Poincaré recurrence theorem (topologic version). *Let $f : X \to X$ be a continuous transformation of a separable metrizable topological space X. The support of any invariant Borel probability measure μ is contained in the closure of the set of recurrent points, namely*

$$\mathrm{supp}\left(\mu\right) \subset \overline{\mathrm{Rec}_f}.$$

If, in particular, f admits an invariant measure μ which is diffuse (i.e. gives positive measure to any nonempty open set) then the set of recurrent points is dense in X, namely

$$\overline{\mathrm{Rec}_f} = X.$$

Observe that if f is a homeomorphism, then the same applies to $\mathrm{Rec}_{f^{-1}}$, and the support of any invariant Borel probability measure is contained in the closure of $\mathrm{Rec}_f \cap \mathrm{Rec}_{f^{-1}}$.

If you don't like the above proof, here is another, perhaps more "visual", of the last statement. Assume that the continuous map $f : X \to X$ preserves a diffuse Borel probability measure μ. For each $n \geq 1$, let

$$R_n = \left\{ x \in X \text{ s.t. } \exists\, k \geq 1 \text{ s.t. } d(f^k(x), x) < 1/n \right\}$$

be the set of "$1/n$-recurrent" points. Of course, $\mathrm{Rec}_f = \cap_{n=1}^{\infty} R_n$. The sets R_n are clearly open. To show that Rec_f is dense we must show that each R_n is, since then the Baire theorem implies that also their countable intersection is dense. So, take any nonempty ball B with diameter $< 1/n$. Its inverse images $f^{-1}(B)$, $f^{-2}(B)$, $f^{-3}(B)$,... have all the same measure $\mu(B) > 0$. Since $\mu(X) = 1$, they cannot be disjoint. There follows that there exist $k > 0$ and $n \geq 0$ such that $f^{-(n+k)}(B) \cap f^{-n}(B) \neq 0$, and this implies that B contains a $1/n$-recurrent point (for a point x in the intersection has both images $f^n(x)$ and $f^{n+k}(x) = f^k(f^n(x))$ in B, hence at distance $< 1/n$). Since B was arbitrary, this proves that each R_n is dense, and Baire theorem implies that Rec_f is dense too.

7.4 Conjunto não-errante

Por volta dos anos trinta, George D. Birkhoff teve a ideia de dividir o espaço dos estados de um sistema dinâmico em duas classes de pontos com dinâmicas qualitativamente distintas.

O ponto x é *errante* se admite uma vizinhança disjunta de todas as suas iteradas, i.e. se existe um aberto U que contém x tal que $U \cap f^n(U) = \emptyset$ para todo tempo $n \geq 1$. O ponto x não é errante se para toda vizinhança U de x existe um tempo $n \geq 1$ tal que $f^n(U) \cap U \neq \emptyset$.

O *conjunto não-errante* NW_f (do inglês "non-wandering set") é o conjunto dos pontos x que não são errantes. A ideia informal é que conjunto não-errante é onde acontece a dinâmica interessante, enquanto o conjunto errante é o conjunto dos pontos que a dinâmica esquece.

O conjunto não-errante NW_f é fechado (o cojnunto dos pontos errantes é aberto quase por definição, pois, se x é errante, todo ponto duma sua vizinhança é errante) e +invariante. Contém os ω-limites de todos os pontos de X, assim como os pontos recorrentes. As inclusões são

$$\mathrm{Per}_f \subset \mathrm{Lim}_f \subset \mathrm{NW}_f \qquad \text{e} \qquad \mathrm{Per}_f \subset \mathrm{Rec}_f \subset \mathrm{NW}_f$$

Se f é um homeomorfismo, NW_f, que é igual a $\mathrm{NW}_{f^{-1}}$, é também invariante, e contém os ω- e α-limites de todos os pontos de X.

(X compacto $\Rightarrow \mathrm{NW}_f \neq \emptyset$) Se X é compacto, então $\mathrm{NW}_f \neq \emptyset$, porque todo ponto $x \in X$ tem $\omega_f(x) \neq \emptyset$ e porque $\mathrm{Lim}_f \subset \mathrm{NW}_f$.

Exercícios.

a. Prove que o conjunto não-errante de um homeomorfismo é fechado, invariante e contém os ω- e α-limites de todos os pontos.

b. Mostre que, se f é um homeomorfismo, então $\mathrm{NW}_f = \mathrm{NW}_{f^{-1}}$.

c. Dê exemplos que mostram que NW_f pode ser vazio.

d. Mostre que $\mathrm{Per}_f \subset \mathrm{Rec}_f \subset \mathrm{NW}_f \subset \mathrm{Rec}_f^{\varepsilon}$. Dê exemplos que mostram que as inclusões podem ser estrictas.

d. Mostre que $\mathrm{Per}_f \subset \mathrm{Rec}_f \subset \mathrm{NW}_f$ e portanto $\overline{\mathrm{Per}_f} \subset \overline{\mathrm{Rec}_f} \subset \mathrm{NW}_f$. Mais dificil é arranjar exemplos que mostram que as inclusões podem ser estrictas.

e. Determine os conjuntos não errantes das transformações lineares do plano.

8 Transitividade e órbitas densas

Na sua genial tentativa de justificar a termodinâmica a partir da hipótese molecular, Ludwig Boltzmann fez a sua famosa "hipótese ergódica". Ele conjeturou que "a superficie de energia constante de um sistema de muitas partículas interagentes (um gas) é composta de uma única órbita" e esta órbita "passa em cada região da superficie um tempo assinptoticamente proporcional ao volume da região". A primeira parte da hipótese equivale a dizer que a ação do tempo no espaço dos estados é "transitiva": para cada dois estados x e x' existe um tempo t tal que $\phi_t(x) = x'$. Ora, se o tempo é $\mathbf{Z}$, isto não pode acontecer, a não ser que X seja enumerável. Em realidade o Boltzmann pensava no caso contínuo, onde o tempo é $\mathbf{R}$, mas mesmo assim esta hipótese é falsa se X tem dimensão maior de um e se a ação é suficientemente suave para evitar fenómenos patológicos como curvas de Peano (e, aliás, as órbitas dos sistemas físicos são soluções de equações diferenciais, logo curvas diferenciáveis). O que sim pode acontecer, é que o sistema admita órbitas densas, e que estas sejam muitas. Este fenómeno, formalizado nas definições à seguir, é o correspetivo topológico da ergodicidade, e o seu signficado é que o espaço dos estados é essencialmente um único pedaço. A segunda parte da hipótese é de natureza probabílistica, e é formalizada na definição de medida invariante ergódica.

8.1 Transitividade

Transformações transitivas. Seja X um espaço métrico completo e separável. Uma transformação contínua $f : X \to X$ é *(topologicamente) +transitiva* se verifica uma das condições equivalentes:
i) para cada dois abertos não vazios $U, V \subset X$ *existe um tempo* $n \geq 0$ *tal que* $f^n(U) \cap V \neq \emptyset$*,*
ii) existe um ponto $x \in X$ *tal que* $\omega_f(x) = X$*,*
iii) existe um conjunto residual de pontos $x \in X$ *tais que* $\omega_f(x) = X$*.*

As implicações iii) $\Rightarrow$ ii) $\Rightarrow$ i) são obvias, pois se $\omega_f(x) = X$, então a trajetória de x passa uma infinidade de vezes por todos os abertos não vazios de X. Para provar que i) $\Rightarrow$ iii), a primeira observação é que a condição i) é equivalente a dizer que, para todo aberto não vazio V, a sua órbita $\cup_{n\geq 0} f^{-n}(V)$ é densa, e ainda mais, as suas órbitas $\cup_{n\geq k} f^{-n}(V) = \cup_{n\geq 0} f^{-n}\left(f^{-k}(V)\right)$ são densas para todo $k \geq 0$. Agora, seja $(U_i)_{i\in\mathbf{N}}$ uma base enumerável da topologia de X . A família dos $\cup_{n\geq k} f^{-n}(U_i)$, com $k \geq 0$ e $i \geq 1$, é uma família de abertos densos em X. A sua interseção enumerável $R = \cap_{i\in\mathbf{N}} \cap_{k\geq 0} \cup_{n\geq k} f^{-n}(U_i)$ é um conjunto residual, e um ponto $x \in R$ tem uma trajetória que passa infinitas vezes por cada um dos abertos U_i, i.e. $\omega_f(x) = X$.

Também, é facil de ver que i) implica que X não tem pontos isolados (desde que não tenha cardinalidade finita, caso trivial em que X é composto por una única órbita). A ausência de pontos isolados implica que, de fato, $\mathcal{O}_f^+(x)' = X$ se $x \in R$.

(+transitivo $\Rightarrow \mathrm{NW}_f = X$*)* Se $f : X \to X$ é +transitiva, então o seu conjunto não-errante é X, porque NW_f contém os conjuntos ω-limite dos pontos de X.

(+transitivo $\Rightarrow \mathrm{Rec}_f$ *residual)* Observe também que uma transformação +transitiva tem muitos pontos recorrentes, de fato um conjunto residual, porque se $\omega_f(x) = X$ então $x \in \omega_f(x)$.

Homeomorfismos transitivos. Existe uma noção mais fraca de transitividade, que só é significativa para as transformações invertíveis. Um homeomorfismo $f : X \to X$ é um *homeomorfismo (topologicamente) transitivo* se verifica uma das condições equivalentes:

i) para cada dois abertos não vazios $U, V \subset X$ existe um tempo $n \in \mathbf{Z}$ tal que $f^n(U) \cap V \neq \emptyset$,

ii) existe um ponto $x \in X$ tal que $\overline{\mathcal{O}_f(x)} = X$,

iii) existe um conjunto residual de pontos $x \in X$ tais que $\overline{\mathcal{O}_f(x)} = X$.

As implicações iii) $\Rightarrow$ ii) $\Rightarrow$ i) são obvias, pois, se a órbita completa de x é densa, passa pelo menos uma vez por todos os abertos não vazios de X. Para provar que i) $\Rightarrow$ iii), a primeira observação é que a condição i) é equivalente a dizer que a órbita $\cup_{n\in\mathbf{Z}} f^n(V)$ de todo aberto não vazio V é densa em X. Agora, seja $(U_i)_{i\in\mathbf{N}}$ uma base enumerável da topologia de X . A família dos $\cup_{n\in\mathbf{Z}} f^n(U_i)$ é uma família de abertos densos. A sua interseção enumerável $R = \cap_{i\in\mathbf{N}} \cup_{n\in\mathbf{Z}} f^n(U_i)$ é um conjunto residual, e um ponto $x \in R$ tem uma trajetória completa que passa pelo menos uma vez por cada um dos abertos U_i, i.e. $\overline{\mathcal{O}_f(x)} = X$.

Observe que $f : X \to X$ é um homeomorfismo transitivo sse f^{-1} é um homeomorfismo transitivo. Um homeomorfismo transitivo pode não ser +transitivo, e, aliás, pode até não ter pontos recorrentes e ter conjunto não-errante vazio, desde que X não seja compacto!

(transitivo $\Leftrightarrow$ "dinamicamente conexo") Mais interessante é observar que um homeomorfismo $f : X \to X$ é um homeomorfismo transitivo sse X não contém uma reunião disjunta de dois subconjuntos abertos invariantes e não vazios. A implicação $\Rightarrow$ é trivial. Para provar a implicação $\Leftarrow$, observe que, se $U, V \subset X$ são dois abertos não vazios, então $\cup_{n\in\mathbf{Z}} f^n(U)$ e $\cup_{n\in\mathbf{Z}} f^n(V)$ são abertos, invariantes e não vazios. Se não forem disjuntos, existem $n, m \in \mathbf{Z}$ tais que $f^n(U) \cap f^m(V) \neq \emptyset$, o que implica $f^{n-m}(U) \cap V \neq \emptyset$.

(transitivo $\Rightarrow$ as funções contínuas invariantes são triviais) Se $f : X \to X$ é um homeomorfismo transitivo, então toda função contínua $\varphi : X \to \mathbf{R}$ invariante é constante. De fato, se φ não é constante, então assume pelo menos dois valores, $a < b$. Logo existe $c = (a + b)/2$ tal que $\{\varphi < c\}$ e $\{\varphi > c\}$ são invariantes, abertos, disjuntos, e não vazios, mas isto contradiz o resultado anterior.

Exercícios.

a. Prove as implicações iii) $\Rightarrow$ ii) $\Rightarrow$ i) na definição de "transformação +transitiva".

b. Prove que, se $f : X \to X$ é +transitiva, então $\mathrm{NW}_f = X$.

c. Prove que se $f : X \to X$ é +transitiva, então o conjunto Rec_f dos pontos recorrentes é residual.

d. Prove as implicações iii) $\Rightarrow$ ii) $\Rightarrow$ i) na definição de "homeomorfismo transitivo".

e. Prove que um homeomorfismo $f : X \to X$ é transitivo sse X não contém uma reunião disjunta de dois subconjuntos abertos invariantes e não vazios.

f. Prove que, se $f : X \to X$ é um homeomorfismo transitivo, então toda função contínua $\varphi : X \to \mathbf{R}$ invariante é constante.

g. Dê exemplos de homeomorfismos $f : X \to X$ que sejam transitivos mas que não sejam +transitivos.

Desafios.

a. *(transitivo e* $\mathrm{NW}_f = X \Leftrightarrow$ *+transitivo)* Mostre que um homeomorfismo $f : X \to X$ é +transitivo sse é um homeomorfismo transitivo e o seu conjunto não-errante é X. A implicação $\Leftarrow$ é obvia ...

b. Pode acontecer que uma transformação $f : X \to X$ seja topologicamente +transitiva mas tenha uma iterada f^n, com $n > 1$, que não é topologicamente +transitiva. Um exemplo trivial é uma permutação de um espaço finito, pois alguma iterada é a identidade. Em geral, se X é compacto, o que acontece é o seguinte: existe uma cobertura $X = X_1 \cup X_2 \cup ... \cup X_k$, onde k é um inteiro que divide n e os X_i são subconjuntos compactos com interseções $X_i \cap X_j$ nowhere dense se $i \neq j$, tal que $f(X_i) = X_{i+1 \bmod k}$ e as restrições $f^n|_{X_i}$ são topologicamente +transitivas. A ideia é escolher um ponto $x \in X$ tal que $\omega_f(x) = X$, e definir $X_i = \omega_{f^n}\left(f^i(x)\right)$ ...

8.2 Minimalidade

A transitividade implica que muitos pontos têm órbitas densas. Isto não impede que exitam pontos x cujas órbitas $\mathcal{O}_f^+(x)$ tenham aderências estritamente contidas em X.

Conjuntos minimais. Seja $f : X \to X$ uma transformação contínua. Um subconjunto fechado e não vazio $K \subset X$ é dito *minimal* se é +invariante e se não contém subconjuntos próprio fechados e +invariantes.

A órbita de um ponto periódico é um exemplo de um conjunto minimal.

Se K é minimal, então a órbita de todo $x \in K$ é densa em K, pois caso contrário a aderência $\overline{\mathcal{O}_f^+(x)}$ seria um subconjunto próprio de K, fechado e +invariantes. Isto implica que em particular $x \in \omega_f(x)$, e portanto todo ponto de um conjunto minimal é recorrente.

Se Min_f denota a reunião dos subconjuntos minimais de X, as inclusões são

$$\mathrm{Per}_f \subset \mathrm{Min}_f \subset \mathrm{Rec}_f$$

Uma transformação arbitrária $f : X \to X$ pode não admitir conjuntos minimais (pense numa translação da reta real).

(X compacto $\Rightarrow \mathrm{Min}_f \neq \emptyset$) Se o espaço X é compacto, podemos considerar a família $\mathcal{C}$ dos subconjunto $C \subset X$ que são fechados, não vazios e +invariantes, munida da ordem parcial "$\subset$". A família contém pelo menos um elemento, o próprio X. Pelo lema de Zorn [6] (observe que toda cadeia $... \subset C_{i+1} \subset C_i \subset ...$ de elementos de $\mathcal{C}$ tem um limite inferior, porque uma interseção de compactos encaixados é um compacto não vazio, e a invariância é preservada), $\mathcal{C}$ contém um elemento minimal K, e este elemento minimal é um conjunto minimal. Em geral, se uma transformação $f : X \to X$ admite um compacto $C \subset X$ tal que $f(C) \subset C$, então admite pelo menos um conjunto minimal $K \subset C$.

(X compacto $\Rightarrow \mathrm{Rec}_f \neq \emptyset$) Corolario é que uma transformação $f : X \to X$ definida num compacto admite pelo menos um ponto recorrente (que pode ser único!), pois $\mathrm{Min}_f \subset \mathrm{Rec}_f$.

[6] O lema de Zorn é um teorema de existência equivalente ao axioma da escolha.

Seja Ω um conjunto não vazio. Uma *ordem parcial* em Ω é uma relação $\preceq$ reflexiva ($x \preceq x\ \forall x \in \Omega$), anti-simétrica ($x \preceq y$ e $y \preceq x \Rightarrow x = y$) e transitiva ($x \preceq y$ e $y \preceq z \Rightarrow x \preceq z$). Um *conjunto parcialmente ordenado* é um par $(\Omega, \preceq)$, um conjunto não vazio Ω munido de uma ordem parcial $\preceq$.

Uma *ordem* em Ω é uma ordem parcial $\preceq$ tal que $\forall x, x' \in \Omega$ temos $x \preceq x'$ ou $x' \preceq x$. Um *conjunto (totalmente) ordenado* é um par $(\Omega, \preceq)$, um conjunto não vazio Ω munido de uma ordem $\preceq$.

Seja $(\Omega, \preceq)$ um conjunto parcialmente ordenado. Uma *cadeia* em Ω é um subconjunto não vazio $C \subset \Omega$ tal que a restrição de $\preceq$ define uma ordem em C, i.e. tal que $\forall c, c' \in C$ temos $c \preceq c'$ ou $c' \preceq c$.

Sejam $(\Omega, \preceq)$ um conjunto parcialmente ordenado e seja $A \subset \Omega$ um subconjunto não vazio. Um elemento $s \in \Omega$ é dito um *limite superior* (*inferior*) de A se $s \preceq a\ \forall a \in A$ (se $a \preceq s\ \forall a \in A$). Um elemento $m \in A$ é dito *elemento maximal* (minimal) de A se nenhum outro elemento de A é maior (menor) que m, ou seja se $\forall a \in A\ m \preceq a$ ($a \preceq m$) $\Rightarrow a = m$.

Lema de Zorn. *Seja Ω um conjunto não vazio e parcialmente ordenado por $\preceq$, tal que toda cadeia $C \subset \Omega$ tem limite superior (inferior). Então Ω contém um elemento maximal (minimal).*

Transformações minimais. Uma transformação contínua $f : X \to X$ é *minimal* se verifica uma das condições equivalentes:
i) toda órbita $\mathcal{O}_f^+(x)$ é densa em X,
ii) X não contém um subconjunto próprio fechado e +invariante, e portanto é um conjunto minimal.

A equivalência i) $\Leftrightarrow$ ii) acima é obvia. Se X é um espaço discreto, a minimalidade implica que X é composto por uma unica órbita, que pode ser finita. Caso contrário, uma transformação minimal não tem pontos periódicos.

(minimal $\Rightarrow$ +transitiva) Obviamente, uma transformação minimal é +transitiva.

Homeomorfismos minimais. O homeomorfismo $f : X \to X$ é um *homeomorfismo minimal* se toda órbita completa $\mathcal{O}_f(x)$ é densa em X, ou seja se X não contém um subconjunto próprio fechado e invariante.

Os homeomorfismos minimais são transitivos. A discussão acima pode ser repetida tirando os "+"... Em particular, um homeomorfismo $f : X \to X$ definido num espaço compacto admite pelo menos um subconjunto minimal K, ou seja neste caso um subconjunto fechado, não vazio e invariante que não contém subconjuntos próprios fechados e invariantes.

Exercícios.
a. Dê exemplos de transformações $f : X \to X$ tais que $\mathrm{Min}_f = \emptyset$.
b. Prove as implicações i) $\Leftrightarrow$ ii) na definição de "transformação minimal".
c. Prove as implicações i) $\Leftrightarrow$ ii) na definição de "homeomorfismo minimal".

8.3 Rotações irracionais do círculo

Teorema. *Uma rotação irracional do círculo é um homeomorfismo minimal.*

dem. Seja $+\alpha : x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$ uma rotação do círculo $\mathbf{R}/\mathbf{Z}$, com $\alpha \notin \mathbf{Q}$. Se $+\alpha$ não é minimal, então existe uma órbita $\mathcal{O}_{+\alpha}(x)$ cuja aderência F não é igual a $\mathbf{R}/\mathbf{Z}$. O complementar $(\mathbf{R}/\mathbf{Z}) \setminus F$ é aberto, invariante e não vazio. Pela classificação dos abertos da reta, sabemos que $(\mathbf{R}/\mathbf{Z}) \setminus F$ é uma reunião enumerável de intervalos abertos (um "intervalo" do círculo é, por definição, um subconjunto conexo) não vazios e dois a dois disjuntos. Seja I o (ou um dos) que tem comprimento maior (porque existe?). Como as rotações preservam o comprimento, as imagens $(+\alpha)^n(I)$ com $n \in \mathbf{Z}$ não podem coincidir (porque $+\alpha$ não tem pontos periódicos) nem intersetar-se (porque a reunião seria um intervalo de comprimento maior). Logo elas são disjuntas e têm todas o mesmo comprimento $|I| > 0$, mas isto é impossível porque o círculo tem comprimento finito. $\square$

Observe que a demonstração acima utiliza a σ-aditividade da medida de Lebesgue.

A informação "aritmética" deste teorema é que, dado $\alpha \notin \mathbf{Q}$, toda órbita

$$\{x + n\alpha + \mathbf{Z}, \text{ com } n \in \mathbf{Z}\}$$

é densa no círculo $\mathbf{R}/\mathbf{Z}$. De fato, é possível provar mais: toda órbita é "equidistribuida" no círculo, no sentido em que, dada uma função integrável φ no intervalo $[0, 1]$, as médias aritmética

$$\frac{1}{n+1}\sum_{j=0}^{n} \varphi\,(x + n\alpha \bmod 1)$$

covergem para o integral $\int_0^1 \varphi\,(x)\,dx$ (veja o teorema de Kronecker e Weyl no capitulo sobre a ergodicidade).

Existe também a possibilidade de dar uma leitura "algébrica" deste resultado. Basta observar que a órbita de 0, a identidade do grupo, é o subgrupo cíclico de $\mathbf{R}/\mathbf{Z}$ gerado por $\alpha + \mathbf{Z}$. Portanto o teorema diz que *os subgrupos próprios e fechados de $\mathbf{R}/\mathbf{Z}$ são os subgrupos finitos.*

Exercício. Embora o resultado seja óbvio, é instructivo provar que as rotações racionais não são topologicamente transitivas usando o critério das funções invariantes, porque a demonstração pode extender-se às translações do toro. (Basta observar que, se $\alpha = p/q$ com p e q inteiros, a função $x \mapsto \sin(2\pi q x)$ está bem definida no círculo, é contínua, não é constante, e é invariante pela rotação $+\alpha$)

Desafio: rotações do toro. Seja $+\alpha : \mathbf{R}^n/\mathbf{Z}^n \to \mathbf{R}^n/\mathbf{Z}^n$ a translação do toro definida por

$$(x_1, x_2, ..., x_n) + \mathbf{Z}^n \mapsto (x_1 + \alpha_1, x_2 + \alpha_2, ..., x_n + \alpha_n) + \mathbf{Z}^n$$

Prove que, se os α_i não são racionalmente independentes, i.e. se existe $k \in \mathbf{Z}^n \setminus \{0\}$ tal que $\sum_{i=1}^n k_i \alpha_i = 0 \mod \mathbf{Z}$, então $+\alpha$ não é topologicamente transitiva. Mais difícil é provar que a translação $+\alpha$ é minimal sse os α_i são racionalmente independentes, i.e. se não existe $k \in \mathbf{Z}^n \setminus \{0\}$ tal que $\sum_{i=1}^n k_i \alpha_i = 0 \mod \mathbf{Z}$.

9 Homeomorfismos do círculo

O estudo dos campos de vetores no toro $\mathbf{R}^2/\mathbf{Z}^2$ levou Poincaré a considerar a necessidade de classificar as possíveis dinâmicas dos homeomorfismos do círculo. Um modelo é constituido pelas rotações $+\alpha$, cujo comportamento é determinado pela racionalidade ou menos do "número de rotação" α. Se α é racional, todo ponto é periódico, logo as órbitas são finitas. Se α é irracional, o sistema é minimal, e portanto toda órbita é densa. A chave para compreender homeomorfismos genéricos é a definição de um "invariante" que jogue o papel de α ...

9.1 Número de rotação

Sejam $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ um homeomorfismo do círculo, e $F : \mathbf{R} \to \mathbf{R}$ um levantamento de f. A seguir, assumimos que f "preserva a orientação", ou seja que $\deg(f) = 1$, e portanto F é uma função contínua e estritamente crescente que verifica $F(x+1) = F(x) + 1$ para todo $x \in \mathbf{R}$.

O *número de rotação* de f é

$$\rho(f) = \tau(F) \mod \mathbf{Z}$$

onde $\tau(F)$ é o *número de translação* de F, definido por

$$\tau(F) = \lim_{n\to\infty} \frac{F^n(x) - x}{n}$$

onde x é um ponto arbitrário da reta. A prova de que a definição acima faz sentido, e que o número de rotação é um invariante topológico, está contida nas seguintes observações.

O limite $\tau(F)$ existe. A transformação F e as suas iteradas F^n são homeomorfismos crescentes da reta que satisfazem $F^n(x+1) = F^n(x)+1$ para todo $x \in \mathbf{R}$. Em particular, $F^n - \mathrm{id}$ são funções periódicas de período um. Isto implica que

$$\max_{x,x'} |(F^n(x) - x) - (F^n(x') - x')| \le 1$$

pois, pela periodicidade basta calcular o máximo no intervalo $[0,1]$, e sabemos que F^n é crescente e que a imagem $F^n([0,1])$ é um intervalo de comprimento um. Seja agora $a_n = F^n(x) - x$. A desigualdade acima implica que a sucessão (a_n) é "quase-subaditiva", i.e. satisfaz

$$a_{n+m} \le a_n + a_m + c$$

para todos $n, m \ge 0$, onde c é uma constante. De fato,

$$\begin{aligned} F^{n+m}(x) - x &= F^n(F^m(x)) - F^m(x) + F^m(x) - x \\ &= F^n(x) - x - F^n(x) + x + F^n(F^m(x)) - F^m(x) + F^m(x) - x \\ &\le F^n(x) - x + F^m(x) - x + 1 \end{aligned}$$

e portanto basta escolher $c = 1$. A existência do limite $\lim_{n\to\infty} a_n/n$ é equivalente à existência do limite $\lim_{n\to\infty} b_n/n$, onde $b_n = a_n + c$. A sucessão (b_n) é subaditiva, ou seja satisfaz $b_{n+m} \le b_n + b_m$. A sucessão (b_n) é crescente, e pela subaditividade satisfaz $b_n \le nb_1$. Portanto, a sucessão (b_n/n) é limitada, logo existe $\lambda = \underline{\lim}_{n\to\infty} b_n/n$. Dado $\varepsilon > 0$, existe um natural N tal que $b_N/N < \lambda + \varepsilon$. Seja agora $n = kN + r$, com k inteiro não negativo e $0 \le r < N$, e seja $B = \max_{1\le i<N} b_i$. Utilizando a subaditividade temos

$$\begin{aligned} b_n/n &\le (b_{kN} + b_r)/n \le (kb_N + b_r)/n \\ &\le b_N/N + b_r/n \le \lambda + \varepsilon + B/n \end{aligned}$$

Pela arbitrariedade de ε, a desigualdade acima implica que $\overline{\lim}_{n\to\infty} b_n/n \le \lambda$, e portanto que o limite $\lim b_n/n$ existe e é igual a λ.

O limite $\tau(F)$ não depende do ponto x. Já vimos que $|(F^n(x) - x) - (F^n(x') - x')| \le 1$, portanto

$$\left| \frac{F^n(x) - x}{n} - \frac{F^n(x') + x}{n} \right| \le 1/n$$

para todos x, x' e n. Isto implica que $\tau(F)$ é independente do ponto x escolhido na sua definição.

$\rho(f)$ **não depende do levantamento** F**.** Observe que dois levantamentos F e G de f diferem por um inteiro, ou seja $G(x) = F(x) + k$ para algum $k \in \mathbf{Z}$. Isto implica que $\tau(F) = \tau(G) + k$, pois $G^n(x) - x = F^n(x) - x + nk$. Portanto o número de rotação $\rho(f)$ está bem definido, não depende do levantamento escolhido.

$\rho(f)$ **é invariante para conjugações topológicas.** Seja $h : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ uma conjugação entre os homeomorfismos f e g. Se H é um levantamento de h e F é um levantamento de f, é imediato verificar que $H \circ F \circ H^{-1}$ é um levantamento de g. Não é difícil mostrar que a diferença $\left(H \circ F \circ H^{-1}\right)^n(x) - F^n(x)$ é limitada, independentemente de x e de n. Basta observar que $\left(H \circ F \circ H^{-1}\right)^n = H \circ F^n \circ H^{-1}$, que $|H(x) - x|$ e $\left|H^{-1}(x) - x\right|$ são limitados por uma constante independente de x, e utilizar a desigualdade do triângulo. Isto implica que $\tau(F) = \tau\left(H \circ F \circ H^{-1}\right)$, e portanto que $\rho(f) = \rho(g)$.

Exercícios.
a. Determine o número de rotação de uma rotação do círculo.
b. Seja f um homeomorfismo do círculo. Mostre que $\rho(f^q) = q \cdot \rho(f) \bmod \mathbf{Z}$. (Observe que, se F é um levantamento de f, então F^q é um levantamento de f^q ...)

9.2 Teorema de classificação de Poincaré

O número de rotação contém a seguinte informação acerca da dinâmica de f.

Teorema. *O número de rotação $\rho(f)$ é racional sse f tem pontos periódicos.*

dem. $\Leftarrow$ Se $F^q(x) = x + p$ com q e p inteiros, então $F^{nq}(x) - x = np$ para todo n, e portanto $\tau(F) = p/q$.
$\Rightarrow$ Observando que $\rho(f^q) = q \cdot \rho(f) \bmod \mathbf{Z}$, basta provar que $\rho(f) = 0$ implica que f tem um ponto fixo. Se f não tem pontos fixos, então a função $F - \mathrm{id}$ tem valores em $\mathbf{R} \backslash \mathbf{Z}$. Em particuar, existe um levantamento tal que $F - \mathrm{id}$ tem valores no intervalo $]0, 1[$ (porque $F - \mathrm{id}$ é contínua e o seu domínio é conexo). Observando que $F - \mathrm{id}$ é periódica de período um, deduzimos que o seu máximo e o seu mínimo são diferentes de 1 e 0 respetivamente, i.e. existe $\varepsilon > 0$ tal que $\varepsilon < F(x) < 1 - \varepsilon$ para todo $x \in [0, 1]$. Iterando as desigualdades, isto implica que $n\varepsilon < F^n(0) < n(1 - \varepsilon)$ e portanto que $\tau(F)$ não é inteiro. $\Box$

É possível provar com pouco esforço que, se $\rho(f)$ é racional, então todos os pontos periódicos de f têm o mesmo período. Portanto, para compreender a estrutura das órbitas de um homeomorfismo com número de rotação racional é suficiente compreender as órbitas de um homeomorfismo f com pontos fixos. Se $F = \mathrm{Fix}(f)$, então f induzes homeomorfismos em cada componente conexa I de $(\mathbf{R}/\mathbf{Z}) \backslash F$. As imagens $f^n(x)$ dos pontos $x \in I \subset (\mathbf{R}/\mathbf{Z}) \backslash F$ convergem para pontos de $\partial I \subset F$ quando $n \to \pm\infty$. A estrutura de F é arbitrária: todo subconjunto compacto do círculo é o conjunto dos pontos fixos de um homeomorfismo.

A dinâmica dos homeomorfismos com número de rotação irracional é descrita pelo seguinte resultado.

Teorema de Poincaré. *Seja $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ um homeomorfismo do círculo (que preserva a orientação) com número de rotação irracional. Então*
i) ou f é minimal, i.e. a órbita de todo ponto é densa no círculo,
ii) ou existe um conjunto invariante $K \subset \mathbf{R}/\mathbf{Z}$, compacto, perfeito e com interior vazio (i.e. um conjunto de Cantor), tal que o conjunto ω-limite de todo ponto do círculo é igual a K.

dem. Pelo lema de Zorn, a família dos subconjuntos não vazios do círculo que são compactos e invariantes, parcialmente ordenada pela inclusão, admite um elemento minimal K. Pela minimalidade, a órbita de todo ponto de K é densa em K. A fronteira ∂K e o conjunto derivado K' são compactos, invariantes e contidos em K, logo têm que ser vazios ou iguais a K. O homeomorfismo não tem pontos periódicos, logo K não pode ser finito. Pelo teorema de Bolzano-Weierstrass $K' \neq \emptyset$, logo $K' = K$, i.e. K é perfeito. Se $\partial K = \emptyset$, então $K = \mathbf{R}/\mathbf{Z}$ e portanto f é minimal. Se, por outro lado, $\partial K = K$, então K tem interior vazio. Seja $x \in (\mathbf{R}/\mathbf{Z}) \backslash K$ e seja I a componente conexa de $(\mathbf{R}/\mathbf{Z}) \backslash K$ que contém x. As imagens $f^n(I)$ são dois a dois disjuntas (sempre porque f não tem pontos periódicos), e portanto diam $(f^n(I)) \to 0$ quando $n \to \infty$. Se $x' \in \partial I \subset K$, então $\omega_f(x') = K$, e a observação anterior implica que também $\omega_f(x) = K$, pois $d(f^n(x), f^n(x')) \leq \operatorname{diam}(f^n(I)) \to 0$ quando $n \to \infty$. Em particular, isto mostra que o conjunto minimal K é único. $\square$

Mais interessante ainda é o seguinte resultado.

Problema: teorema de classificação de Poincaré. *Seja $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ um homeomorfismo do círculo (que preserva a orientação) com número de rotação irracional.*
i) Se f é minimal, então f é topologicamente conjugado à rotação $+\rho(f)$.
ii) Se f não é minimal, então a rotação $+\rho(f)$ é um fator de f.

Se f é minimal, podemos construir uma conjugação H entre uma órbita de f e uma órbita de $+\rho(f)$, e depois definir uma conjugação $h : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ por continuidade, utilizando o fato das órbitas serem densas. Isto é possível porque as órbitas de f têm "a mesma ordem" das órbitas de $+\rho(f)$. Se f não é minimal, é possível construir uma semiconjugação $h : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ tal que o próprio $\mathbf{R}/\mathbf{Z}$ seja a imagem $h(K)$ do conjunto minimal de f. De alguma maneira, a semiconjugação "esquece" $(\mathbf{R}/\mathbf{Z}) \setminus K$, o conjunto errante de f.

9.3 Difeomorfismos do círculo e teorema de Denjoy

Um homeomorfismo $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ com número de rotação irracional pode não ser minimal, logo ter um conjunto não errante $(\mathbf{R}/\mathbf{Z}) \setminus K$, composto por intervalos abertos I com imagens $f^n(I)$ disjuntas. Se f é de classe $\mathcal{C}^1$, um controle sobre a derivada f' e o teorema do valor médio ajudam a estimar os comprimentos dos $f^n(I)$. O resultado, obtido por Arnaud Denjoy nos anos trinta, é o seguinte.

Problema: teorema de Denjoy. *Um homeomorfismo $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ com número de rotação irracional, de classe $\mathcal{C}^1$ e com derivada de variação limitada, é minimal, e portanto topologicamente conjugado à rotação $+\rho(f)$.*

A ideia é provar que, se f' tem variação limitada e I é um "intervalo errante", os comprimentos dos $f^n(I)$ são uniformemente $> \varepsilon$ para algum $\varepsilon > 0$. Sendo disjuntos, isto leva a um absurdo. O próprio Denjoy mostrou como construir homeomorfismos de classe $\mathcal{C}^1$, com derivada f' de classe α-Holder e $\alpha < 1$ arbitrário, que têm número de rotação irracional sem seres minimais.

Problema. Discuta a dinâmica da família de tranformações do círculo $f_{\alpha,\varepsilon} : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ definidas por

$$x + \mathbf{Z} \mapsto x + \alpha + \frac{\varepsilon}{2\pi}\sin(2\pi x) + \mathbf{Z}$$

ao variar os parâmetros α e ε.

10 Perda de memória, independência assimptótica e "mixing"

10.1 Órbitas desordenadas

Dicotomia: pontos regulares ou não regulares. Seja $f : X \to X$ uma transformação do espaço métrico (X, d). As iterações de f dividem de maneira natural o espaço X em duas classes de pontos, dependendo se as órbitas são "estáveis" ou "instáveis" por pequenas perturbações da condição inicial.

O ponto $x \in X$ é *regular* se, para todo $\varepsilon > 0$, admite uma vizinhança B tal que para todo $x' \in B$ e todo tempo $n \geq 0$

$$d\left(f^n\left(x\right), f^n\left(x'\right)\right) < \varepsilon$$

i.e. se a família de transformações $\{f^n\}_{n \in \mathbf{N}_0}$ é equicontínua em x. Do ponto de vista físico isto quer dizer que, se os instrumentos têm sensibilidade ε, as trajetórias de um ponto em cada ε-bola são suficientes para descrever todas as trajetórias dos pontos regulares. Em particular, se X é compacto e todo ponto é regular, um número finito de trajetórias descreve o comportamento de todas as trajetórias a menos de um erro ε, por tempos arbitrariamente grandes.

O ponto $x \in X$ *não é regular* se existe $\delta > 0$ tal que para quaisquer vizinhança B de x existem $x' \in B$ e um tempo $n \geq 0$ tais que

$$d\left(f^n\left(x\right), f^n\left(x'\right)\right) > \delta$$

O significado desta condição é que f "tem dependência sensível das condições iniciais" nas vizinhanças de x. Num certo sentido, as trajetórias de pontos numa vizinhança arbitrária de x "perdem memória" de x.

Dependência sensível das condições iniciais. Se o conjunto dos pontos não regulares for compacto, o δ acima pode ser escolhido independente do ponto. Isto sugere a seguinte definição.

A transformação $f : X \to X$ *tem dependência sensível das condições iniciais* se todos os pontos de X são "uniformemente" não regulares, i.e. se existe $\delta > 0$ tal que para todo $x \in X$ e toda vizinhança B de x, existem $x' \in B$ e um tempo $n \geq 0$ tais que

$$d\left(f^n\left(x'\right), f^n\left(x\right)\right) > \delta$$

O significado físico deste fenómeno é: não inporta quanto pequena seja a sensibilidade ε dos nossos instrumentos, as trajetórias de dois pontos que nos consideramos "indistinguíveis", ou seja a distância $< \varepsilon$, distam mais de um certo δ, independente de ε, passado um certo tempo n.

Problema: conjuntos de Julia e de Fatou. A dicotomia acima é particularmente significativa para os endomorfismos da esfera de Riemann $\overline{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$, as transformações racionais $R : \overline{\mathbf{C}} \to \overline{\mathbf{C}}$, definidas por

$$z \mapsto P\left(z\right) / Q\left(z\right)$$

onde P e Q são polinómios. Um ponto $z \in \overline{\mathbf{C}}$ é dito "regular" se admite uma vizinhança U tal que a família $\{R^n|_U\}_{n \geq 1}$ é uma família normal (i.e. toda sucessão de elementos da família admite uma subsucessão localmente uniformemente convergente). O conjunto F dos pontos regulares, que é um subconjunto aberto da esfera de Riemann, é dito *conjunto de Fatou.* O conjunto complementar, o fechado $J = \overline{\mathbf{C}} \backslash F$, é dito *conjunto de Julia.* O conjunto de Julia é onde acontece a dinâmica "desordenada" de R. Se a transformação é da forma $R(z) = z^n$ com $n > 1$ (ou conformemente conjugada a um polinómio deste tipo), então J é um círculo. Se R é mais complicado, acontece que J é tipicamente um conjunto muito irregular: um conjunto de Cantor, uma curva não retificável de dimensão de Hausdorff > 1, ou um conjunto ainda mais esquisito. O estudo destes fenómenos começou por volta de 1918-19, com os trabalhos de Gaston Julia e Pierre Fatou. A compreensão da dinâmica dos endomorfismos da esfera de Riemann (devida essencialmente às técnicas disponíveis de análise complexa) é um dos maiores sucessos da moderna teoria dos sistemas dinâmicos. Uma introdução excelente está nas notas de John Milnor, *Dynamics in one complex variable*, IMS-SUNY Stony Brook 1990.

10.2 Mixing topológico

A transformação $f : X \to X$ é *topologicamente mixing* (ou seja, “misturadora”) se para cada dois abertos não vazios $U, V \subset X$ existe um tempo $n \geq 0$ tal que para todo tempo $k \geq n$

$$f^k(U) \cap V \neq \emptyset$$

(ou seja, basta esperar um certo tempo finito n para ver pontos de U cujas órbitas visitam V).

O mixing topológico captura a ideia de que o futuro $f^k(U)$, com $k > 1$, de cada aberto U é assimptoticamente ”independente” do seu presente, pois interseta estavelmente cada outro aberto não vazio V.

(mixing ⇒ +transitivo) Uma transformação topologicamente mixing é topologicamente +transitiva. Em particular, $\mathrm{NW}_f = X$, e $\omega_f(x) = X$ é uma propriedade genérica.

(mixing ⇒ dependência sensível) O mixing topológico é uma propriedade ainda mais forte de que a dependência sensível das condições iniciais. Pois, seja $f : X \to X$ uma transformação topologicamente mixing do espaço métrico (X, d) com pelo menos dois pontos, e sejam U e V dois abertos disjuntos a distância $> 2\delta$, com $\delta > 0$. Dado $x \in X$, a órbita de toda vizinhança B de x interseta os dois abertos a partir de um certo tempo $n \geq 0$, e portanto existe um ponto $x' \in B$ tal que $d(f^n(x'), f^n(x)) > \delta$.

Em particular, uma isometria não pode ser topologicamente mixing.

Desafios.

a. Existe um homeomorfismo minimal (portanto topologicamente transitivo) que não é topologicamente mixing?

b. Existe uma transformação topologicamente transitiva que não é nem minimal nem topologicamente mixing?

c. Uma transformação $f : X \to X$ é dita *weakly mixing* se a ”transformação produto” $f \times f : X \times X \to X \times X$, definida por

$$(x, x') \mapsto (f(x), f(x'))$$

é topologicamente mixing. Mostre que uma transformação weak mixing (de um espaço X que contém mais do que um ponto) tem dependência sensível das condições iniciais. Mostre que todas as iteradas f^n de uma transformação weak mixing de um espaço compacto são +transitivas. Prove que

$$\text{mixing} \;\Rightarrow\; \text{weak mixing} \;\Rightarrow\; +\text{transitivo}$$

e dê exemplos que mostram que as implicações contrárias são falsas.

Exemplo/exercício: a transformação tenda. A *transformação tenda* [7] é a transformação $T : [0,1] \to [0,1]$ definida por

$$T(x) = \begin{cases} 2x & \text{se } x < 1/2 \\ 2 - 2x & \text{se } x \geq 1/2 \end{cases}$$

A iteração de T é simples, pois a composição de duas transformações afins é uma transformação afin.

Não é difícil provar por indução que em cada um dos intervalos $I_{k,n} = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ com $k = 0, 1, 2, ..., 2^n - 1$ a transformação T^n tem a forma

$$x \mapsto T^n(x) = \begin{cases} 2^n x + k & \text{se } k \text{ é par} \\ -2^n x + k + 1 & \text{se } k \text{ é impar} \end{cases}$$

Em particular, T^n é uma bijeção estritamente crescente ou decrescente de $I_{k,n}$ sobre o intervalo $[0,1]$. O teorema de ponto fixo implica que T^n tem um e um único ponto fixo em cada um dos intervalos $I_{k,n}$ (que é repulsivo, pois a derivada de T^n é $2^n > 1$, e só coincide com um dos extremos quando $k = 0$), e portanto que $|\mathrm{Fix}(T^n)| = 2^n$. Além disso, sendo que todo aberto não vazio

[7] Um padeiro estica, dobra e amassa repetidamente a sua massa com o objetivo de ”misturar”, ou seja de chegar a ter uma mistura de farinha e água e outros ingredientes que seja o quanto mais possível homogénea... Isto é mais ou menos o que faz a transformação tenda. Por alguma razão, o nome de *transformação do padeiro* é reservado ao seu análogo bidimensional e invertível.

$U \subset [0,1]$ contém um dos intervalos $I_{k,n}$ se n é suficientemente grande, os pontos periódicos de T são densos no intervalo $[0,1]$.

A transformação T é topologicamente mixing. De fato, todo aberto não vazio $U \subset [0,1]$ contém um dos intervalos $I_{k,n}$ com n suficientemente grande, logo $T^n(U) = [0,1]$ e portanto $T^k(U) = [0,1]$ para todo tempo $k \geq n$ porque T é sobrejetiva. Isto implica que $T^k(U) \cap V \neq \emptyset$ para todo tempo $k \geq n$ e para todo aberto não vazio $V \subset [0,1]$. Portanto existe um conjunto residual (i.e. grande!) de pontos x tais que $\omega_f(x) = [0,1]$, ou seja cuja trajetória é essencialmente imprevisível!

Exercícios.

a. Verifique que $h : x \mapsto \sin^2(\pi x/2)$ realiza uma conjugação topológica entre a transformação tenda T e a transformação $f_4 : [0,1] \to [0,1]$ da família quadrática, definida por $f_4(x) = 4x(1-x)$. Deduza ....

b. Discuta a dinâmica da transformação $S : [0,1] \to [0,1]$ definida por

$$S(x) = \begin{cases} 2x & \text{se } x < 1/2 \\ 2x-1 & \text{se } x \geq 1/2 \end{cases}$$

Cuidado, S não é contínua!, mas não é muito diferente da transformação tenda...

10.3 Dinâmica dos deslocamentos de Bernoulli

O shift de Bernoulli $\sigma : \Sigma^+ \to \Sigma^+$ sobre um alfabeto finito $X = \{1, 2, ..., z\}$ é o protótipo de uma transformação topologicamente mixing, e tem todas as propriedades típicas desta classe de transformações.

Uma base da topologia produto em $\Sigma^+ = X^{\mathbf{N}}$ é a família dos "cilindros centrados", a família dos subconjuntos $C_\alpha = \{x = (\alpha, *)\} =$"palavras infinitas que começam pela palavra α", ao variar α entre todas as palavras finitas nas letras do alfabeto X. Ora, se $U \subset \Sigma^+$ é um aberto não vazio, existe um cilindro centrado $C_\alpha \subset U$, e, se $|\alpha|$ denota o comprimento da palavra α, é imediato ver que $\sigma^n(C_\alpha) = \Sigma^+$ para todo $n \geq |\alpha|$, logo a fortiori $\sigma^n(U)$ interseta todo aberto não vazio a partir do tempo $|\alpha|$. Isto prova que σ é topologicamente mixing.

Sendo σ topologicamente mixing, logo +transitiva, um ponto genérico de Σ^+ tem órbita densa.

Uma curiosidade é que neste exemplo é de fato imediato "construir" um ponto com órbita densa. Basta enumerar o conjunto das palavras finitas nas letras de X, por exemplo $\alpha_1, \alpha_2, \alpha_3, ...$, e depois observar que a trajetória do ponto $x = (\alpha_1, \alpha_2, \alpha_3, ...)$ passa por todos os elementos da base da topologia.

Menos obvio é construir um ponto x tal que $\omega_\sigma(x) = \Sigma^+$, que também sabemos ser uma propriedade genérica! Um exemplo é o ponto

$$x = (\alpha_1, \alpha_1, \alpha_2, \alpha_1, \alpha_2, \alpha_3, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, ...)$$

cuja trajetória passa infinitas vezes por todos os elementos da base da topologia.

O shift de Bernoulli tem também muitos pontos periódicos. Per_f é um conjunto enumerável e denso em Σ^+. De fato, dada uma palavra finita α arbitrária, o ponto $x = (\alpha, \alpha, \alpha, ...)$ é periódico, e o seu período é um inteiro positivo que divide $|\alpha|$. Isto prova que a cardinalide de $\mathrm{Fix}(\sigma^n)$ é igual a cardinalidade das palavras de comprimento n nas letras do alfabeto X. Além disso, todo cilindro centrado contém um ponto periódico, pois C_α contém $(\alpha, \alpha, \alpha, ...)$, logo os pontos periódicos são densos em Σ^+.

Além de pontos periódicos e de pontos cuja órbita é densa, o shift de Bernoulli admite pontos cuja órbita é densa em subconjuntos próprios de Σ^+. Por exemplo, a restrição de σ ao subconjunto $(X \backslash \{1\})^{\mathbf{N}} \subset \Sigma^+$ formado pelas palavras infinitas que não contêm a letra "1" é uma transformação topologicamente mixing (basta repetir a demonstração anterior), logo um ponto genérico $x \in (X \backslash \{1\})^{\mathbf{N}}$ tem órbita densa em $(X \backslash \{1\})^{\mathbf{N}}$ ...

Exercícios.

a. Discuta em detalhes a dinâmica do shift de Bernoulli:

Considere o produto cartesiano $\Sigma^+ = X^{\mathbf{N}}$, onde $X = \{1, 2, ..., z\}$ é um alfabeto finito. Verifique que a familia dos cilindros centrados é uma base de uma topologia em Σ^+, dita topologia produto.

Verifique que o shift de Bernoulli $\sigma : \Sigma^+ \to \Sigma^+$ sobre um alfabeto finito é uma transformação contínua (observe que a imagem inversa de um cilindro centrado é uma reunião finita de z cilindros centrados, e deduza que a imagem inversa de um aberto é um aberto).

Prove que $\sigma : \Sigma^+ \to \Sigma^+$ é uma transformação expansora se a métrica em Σ^+ é definida por $d(x, x') = \sum_{n=1}^{\infty} \lambda^{-n} \cdot |x_n - x'_n|$, onde $\lambda > 1$.

Mostre que $|\mathrm{Fix}(\sigma^n)| = |X|^n$.

Mostre que todo cilindro centrado de Σ^+ contém um ponto periódico de σ, e que portanto o conjunto dos pontos periódicos Per_σ é denso em Σ^+.

Prove que $\sigma : \Sigma^+ \to \Sigma^+$ é topologicamente mixing.

Dê exemplos de pontos $x \in \Sigma^+$ tais que $\omega_\sigma(x) = X$.

De exemplos de pontos não pré-periódicos $x \in \Sigma^+$ tais que tais que a aderência da órbita $\mathcal{O}_\sigma^+(x)$ seja um subconjunto próprio de Σ^+.

b. Seja $\Sigma = X^{\mathbf{Z}}$, o espaço das palavras infinitas $x = (..., x_{-2}, x_{-1}, x_0 x_1, x_2, ..., x_n, ...)$ nas letras do alfabeto finito $X = \{1, 2, ..., z\}$, munido da topologia produto. Verifique que o shift $\sigma : \Sigma \to \Sigma$, definido por $(\sigma(x))_k = x_{k+1}$, é um homeomorfismo.

Determine a cardinalidade de $\mathrm{Fix}(\sigma^n)$, e prove que os pontos periódicos são densos em Σ.

Prove que $\sigma : \Sigma \to \Sigma$ é topologicamente mixing.

Problema: cadeias de Markov topológicas e codificação. A restrição do shift σ a um subconjunto invariante de Σ^+ (ou Σ) é dita um *sistema dinâmico simbólico*. A maneira mais simples de produzir subconjuntos invariantes é por meio de matrizes de transição (a ideia vem da teoria das cadeias de Markov). Seja $A = (a_{ij})$ uma "matiz de transição", ou seja uma matriz $z \times z$ com entradas 0 ou 1. Seja

$$\Sigma_A^+ = \left\{ x \in \Sigma^+ \text{ t.q. } a_{x_n x_{n+1}} = 1 \ \forall \ n \geq 0 \right\}$$

E facil de ver que $\sigma\left(\Sigma_A^+\right) \subset \Sigma_A^+$. A restrição $\sigma_A = \sigma\Big|_{\Sigma_A^+} : \Sigma_A^+ \to \Sigma_A^+$ é dita *cadeias de Markov topológicas* (ou *subshift of finite type*). A ideia é que o alfabeto representa os possíveis "estados" do sistema, e a "transição" entre o estado i e o estado j é possível se $a_{ij} = 1$.

Uma palavra finita $x_1, x_2, ..., x_n$ é *admissivel* se $a_{x_k x_{k+1}} = 1$ para todos $k = 1, 2, ..., n-1$, i.e. se é um pedaço de uma história possível do estado x_1. O número de palavras admissiveis de comprimento $n+1$ que começam pela letra i e terminam pela letra j é igual a $(A^n)_{ij}$. Isto mostra que $|\mathrm{Fix}(\sigma_A^n)| = \mathrm{tr} A^n$. Este número pode ser estimado utilizando o teorema de Perron-Frobenius.

A cadeia de Markov topológica σ_A é dita *transitiva* se existe um tempo $k \geq 1$ tal que todas as entradas de A^k (e portanto as entradas de A^n se $n \geq k$) são positivas. O resultado relevante é que uma cadeia de Markov topológica transitiva é topologicamente mixing, e tem órbitas periódicas densas.

Os sistemas dinâmicos simbólicos são modelos abstratos de sistemas dinâmicos. Uma das ideias centrais da teoria é procurar "codificar" um sistema "concreto" $f : X \to X$ com um sistema simbólico. Uma estratégia possível é dividir X em subconjuntos fechados $B_1, B_2, ..., B_z$ tais que a historia de todo $x \in X$, definida por $(x_0, x_1, x_2, ..., x_n, ...)$ onde $x_n = i$ sse $f^n(x) \in B_i$, determina univocamente o ponto x. Então, se $A = (a_{ij})$ é a matriz de 0 e 1's definida por $a_{ij} = 1$ se $B_j \subset f(B_i)$, e se f é suficientemente "desordenada", a cada história $(x_0, x_1, x_2, ..., x_n, ...) \in \Sigma_A^+$ corresponde um único ponto $x = \cap_{n=0}^{\infty} f^{-n}(B_{x_n})$. A esperança é que f seja conjugado a σ_A. Se X é conexo, os B_i têm que ter interseções não vazias, e portanto a correspondência entre Σ_A^+ e X não pode ser biunívoca. Mesmo assim, fora das histórias ambíguas, a dinâmica de f é descrita pela dinâmica de σ.

Problema: a transformação de padeiro. A *transformação do padeiro* é $f : [0,1] \times [0,1] \to [0,1] \times [0,1]$ definida por

$$(x, y) \mapsto \begin{cases} (2x, y/2) & \text{se } 0 \leq x \leq 1/2 \\ (2x - 1, (y+1)/2) & \text{se } 1/2 < x \leq 1 \end{cases}$$

Discuta a sua dinâmica. Considere o shift de Bernoulli $\sigma : \Sigma \to \Sigma$ no produto cartesiano $\Sigma = \{0,1\}^{\mathbf{Z}}$. Mostre que a aplicação $h : \Sigma \to [0,1] \times [0,1]$ definida por

$$x = (..., x_{-2}, x_{-1}, x_0 x_1, x_2, ..., x_n, ...) \mapsto \left(\sum_{n=0}^{\infty} \frac{x_{-n}}{2^n}, \sum_{n=1}^{\infty} \frac{x_n}{2^n} \right)$$

é uma semiconjugação entre σ e f. Deduza ...

10.4 Digressão: conjuntos de Cantor

As aderências das órbitas de transformações suficientemente "desordenadas" podem ter uma estrutura complicada, e, se forem desconexas, são tipicamente conjuntos de Cantor.

Um *conjunto de Cantor* é um espaço métrico compacto, perfeito e totalmente desconexo. O arquétipo é o *conjunto de Cantor standard* (o "middle-third Cantor set")

$$K = \left\{ \sum_{n=1}^{\infty} \frac{x_n}{3^n} \text{ com } x_n \in \{0,2\} \right\} \subset [0,1]$$

o conjunto dos números entre 0 e 1 cuja representação em base 3 utiliza só as letras 0 e 2.

Outra definição é $K = [0,1] \backslash \cup_{k=1}^{\infty} I_k$, onde os intervalos abertos I_k são definidos iterativamente da seguinte maneira: I_1 é o terço central $]1/3, 2/3[$ de $[0,1]$, I_2 e I_3 são os terços centrais dos intervalos de $[0,1] \setminus I_1$, a saber $]1/9, 2/9[$ e $]7/9, 8/9[$, ...etc.

Mais uma definição é $K = \cap_{k \geq 0} K_n$, onde

$$K_n = \left\{ \sum_{n=1}^{\infty} \frac{x_n}{3^n} \text{ com } x_1, x_2, ..., x_k \in \{0,2\} \text{ e } x_i \in \{0,1,2\} \text{ se } i > k \right\}$$

Observe que $... \subset K_{n+1} \subset K_n \subset ... \subset K_0 = [0,1]$, e que cada K_n é uma reunião disjunta de 2^n intervalos fechados de comprimento 3^{-n}. Em particular, K é compacto, pois é uma interseção enumerável de compactos encaixados.

As estranhas propriedades do conjunto de Cantor tem demonstrações muito simples observando que K é homeomorfo ao produto topológico $\{0,2\}^{\mathbf{N}}$, i.e. ao espaço do deslocamentos de Bernoulli num alfabeto de duas letras. O homeomorfismo é simplesmente

$$\{0,2\}^{\mathbf{N}} \ni x = (x_1, x_2, ..., x_n, ...) \mapsto \sum_{n=1}^{\infty} x_n/3^n \in K$$

K não tem pontos isolados, e portanto $K' = K$, i.e. é perfeito. De fato, um ponto $x \in K$ pertençe a uma interseção $\cap_{n\geq 1} J_n$, onde $J_n = [a_n, b_n]$ são certas componentes conexas dos compactos K_n. Logo, pelo menos as duas sucessões distintas (a_n) e (b_n) de pontos de K convergem para x, pois $|b_n - a_n| = 3^{-n} \to 0$ quando $n \to \infty$.

A componente conexa de todo ponto $x \in K$ é $\{x\}$, i.e. K é totalmente desconexo. De fato, sejam x e x' dois pontos distintos de K. Se n é suficientemente grande, i.e. se $3^{-n} < d(x, x')$, os pontos x e x' estão em duas componentes conexas distintas de K_n.

A função $\{0,2\}^{\mathbf{N}} \to \{0,2\}^{\mathbf{N}} \times \{0,2\}^{\mathbf{N}}$ definida por

$$(x_1, x_2, ..., x_n, ...) \mapsto ((x_1, x_3, ..., x_{2n-1}, ...), (x_2, x_4, ..., x_{2n}, ...))$$

induz um homeomorfismo de K sobre $K \times K$. Por indução, vê-se que K é homeomorfo a K^n para todo $n \in \mathbf{N}$. De fato, não é difícil provar que K é homeomorfo a $K^{\mathbf{N}}$.

Observe que $\{0,2\}^{\mathbf{N}}$ é homeomorfo a $\{0,1\}^{\mathbf{N}}$, e que a representação binária dos reais entre 0 e 1 é uma aplicação contínua de $\{0,1\}^{\mathbf{N}}$ sobre o intervalo $[0,1]$. Portanto, existe uma função contínua de K sobre o intervalo $[0,1]$, e em particular, pelo teorema de Schröder-Bernstein, K tem a cardinalidade do intervalo.

Outra propriedade muito apreciada do conjunto de Cantor é a "auto-similaridade". Assim como um intervalo compacto da reta é homeomorfo a todo seu subintervalo (não trivial) compacto, o conjunto de Cantor K contém muitos subconjuntos próprios homeomorfos a K. Por exemplo, a aplicação $x \mapsto 3x$ define um homeomorfismo de $K \cap [0,1/3]$ sobre K (isto não é casual, mas tem muito a ver com a dinâmica da transformação $\times 3$ no círculo). De fato, todo aberto não vazio do conjunto de Cantor contém uma "cópia" do próprio conjunto. Formalmente, todo intervalo aberto da reta I tal que $I \cap K \neq \emptyset$ contém um subconjunto $J \subset I \cap K$ homeomorfo a K. Pois, se um intervalo aberto contém um ponto de K, então contém pelo menos uma das componentes conexas de K_n, digamos $J_n = [a_n, b_n]$, desde que n seja suficientemente grande. Não é difícil depois arranjar um homeomorfismo (por exemplo afim, da forma $x \mapsto 3^n (x - a_n)$) de J_n sobre K.

Uma curiosidade: se o comprimento de K podesse ser calculado como o limite dos comprimentos dos K_n (o que é de fato possível, basta definir cuidadosamente o que é o "comprimento", mas esta é uma história comprida... e chama-se teoria da integração), seria $|K| = \lim_{n\to\infty} |K_n| = \lim_{n\to\infty} 2^n \cdot 3^{-n} = 0$. O conjunto de Cantor é "muito pequeno", mesmo tendo "o mesmo número de pontos" do intervalo!

Exercício. Prove em detalhes todas as propriedades do conjunto de Cantor enunciadas acima (exceto a última!).

Problema: Cantor invariante pela família quadrática. Seja $f_\lambda : \mathbf{R} \to \mathbf{R}$ a transformação definida por $x \mapsto \lambda x (1 - x)$, onde $\lambda > 0$. A trajetória dos pontos $x \notin I = [0, 1]$ diverge, de fato $|f_\lambda^n (x)| \to \infty$ quando $n \to \infty$. O conjunto dos pontos da reta que têm órbitas limitadas é

$$\Lambda = \cap_{n\geq 0} f_\lambda^{-n} (I)$$

Se $\lambda > 4$, então $f_\lambda^{-1} ([0, 1])$ é a reunião disjunta de dois intervalos fechados I_0 e I_1 contidos em I. Se λ é suficientemente grande, o módulo da derivada f_λ' é uniformemente > 1 nos pontos de I_0 e I_1. Mostre que, se λ é suficientemente grande, $f_\lambda^{-(n+1)} (I)$ é uma reunião disjunta de 2^{n+1} intervalos compactos estritamente contidos, em pares, nos 2^n intervalos de $f_\lambda^{-n} (I)$. Deduza que Λ é um conjunto de Cantor, e que a transformação $f_\lambda|_\Lambda : \Lambda \to \Lambda$ é topologicamente conjugada ao shift de Bernoulli $\sigma : \Sigma^+ \to \Sigma^+$ no alfabeto $\{0, 1\}$.

10.5 Transformações expansoras

A dependência sensível das condições iniciais pode ser induzida por propriedades mais fortes. Uma maneira obvia é "obrigar" f a esticar as distâncias, e isto pode ser feito de muita maneiras.

Uma transformação contínua $f : X \to X$ é *expansiva* se existe $\delta > 0$ tal que para todos $x, x' \in X$ distintos existe um tempo $n \geq 0$ tal que

$$d(f^n(x), f(x')) > \delta$$

Uma transformação contínua $f : X \to X$ é *expansora* se existem $\mu > 1$ e $\varepsilon > 0$ tais que para todos $x, x' \in X$ distintos com $d(x, x') < \varepsilon$

$$d(f(x), f(x')) > \mu \cdot d(x, x')$$

Esta é uma condição local, porque se ε fosse infinito nenhum espaço compacto admitiria transformações expansoras. Por outro lado, é precisamente nos espaços compactos que a expansividade causa recorrências interessantes das trajetórias: os pontos querem fugir uns dos outros, mas não têm muito espaço por onde ir, e acabam se reencontrando de vez em quando...

A existência de transformações expansoras implica fortes restrições topológicas sobre o espaço X. Se X é uma variedade, o recobrimento universal de X tem que ser $\mathbf{R}^n$ e o grupo fundamental de X não pode ser arbitrário. Por exemplo, entre as superfícies fechadas e orientáveis, só o toro admite transformações expansoras!

Exercício. Dê exemplos de transformações expansoras de $\mathbf{R}$, de $\mathbf{R}/\mathbf{Z}$ e de $\mathbf{R}^2/\mathbf{Z}^2$.

Desafios.

a. Prove que não existe nenhuma transformação expansiva $f : I \to I$ definida num intervalo compacto $I \subset \mathbf{R}$. Observe que uma tal transformação seria localmente injetiva, logo estritamente crescente ou decrescente...

b. Uma transformação expansora de um espaço compacto pode ser um homeomorfismo? A resposta é sim, mas só se o espaço compacto tiver cardinalidade finita! Arranjar um exemplo não é difícil. Por outro lado, mostrar que um espaço compacto e infinito não admite homeomorfismos expansores não é trivial...

Exemplo/exercício: expansão decimal. Seja F a função $x \mapsto 10 \cdot x$ definida na reta real. Mostre que ela "induz" uma transformação contínua $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ do círculo, por meio de

$$f (x + \mathbf{Z}) = F (x) + \mathbf{Z}$$

(ou seja, mostre que esta expressão é indepedente do representante x escolhido para o ponto $x + \mathbf{Z}$ do círculo, e mostre que f é contínua).

Mostre que f é expansora, se o círculo é munido da métrica standard herdada da métrica euclidiana da reta. (Observe que, se $d(x, x') < 1/20$, então $d(f(x), f(x')) = 10 \cdot d(x, x')$ ...)

Seja $x = 0, x_1 x_2 x_3 ...$, com $x_n \in \{0, 1, 2, ..., 9\}$, uma representação decimal de $x \in [0, 1[$, pensado como um ponto de $\mathbf{R}/\mathbf{Z}$. Mostre que $f(0, x_1 x_2 x_3 ... + \mathbf{Z}) = 0, x_2 x_3 x_4 ... + \mathbf{Z}$.

Procure os pontos fixos, periódicos, e pré-periódicos de f.

Calcule a cardinalidade de $\mathrm{Fix}(f^n)$. Mostre que os pontos periódicos de f são densos no círculo.

Um número x é dito "periódico" se a sua representação decimal é da forma

$$x = b_n b_{n-1} ... b_0 . x_1 x_2 ... x_n \, (a_1 a_2 ... a_k)$$

Existem número não periódicos? Quantos? Sabe fazer exemplos?

Mostre que, para todo $\varepsilon > 0$ e todo $x \in \mathbf{R}/\mathbf{Z}$, existem $x' \in \mathbf{R}/\mathbf{Z}$ e um tempo $n \geq 0$ tais que

$$d(x, x') < \varepsilon \qquad \text{e} \qquad d(f^n(x), f^n(x')) > 1/4$$

ou seja, que a transformação f tem a propriedade de dependência sensível das condições iniciais. (Observe que, se $d(x, x') < 1/2 \cdot 10^{-n}$, então $d(f^n(x), f^n(x')) = 10^n \cdot d(x, x')$ ...)

Mostre que, para todo intervalo não vazio $I \subset \mathbf{R}/\mathbf{Z}$, existe um tempo $n \geq 0$ tal que $f^k(I) = \mathbf{R}/\mathbf{Z}$ para todo tempo $k \geq n$. Deduza que f é topologicamente mixing.

Seja $b = (b_1 b_2 ... b_n)$ uma palavra finita no alfabeto $\{0, 1, 2, ..., 9\}$. Prove que existe um conjunto residual de pontos $x \in [0, 1[$ tais que a representação em base 10 de x contém a palavra b uma infinidade de vezes (no sentido em que, se x é da forma $0.x_1 x_2 x_3 ... x_k ...$, existem uma infinidade de $k \geq 0$ tais que $(x_{k+1} x_{k+2} ... x_{k+n}) = (b_1 b_2 ... b_n)$). Prove que existe um conjunto residual de pontos $x \in [0, 1[$ tais que a representação em base 10 de x contém todas as palavras finitas no alfabeto $\{0, 1, 2, ..., 9\}$ uma infinidade de vezes. Dê exemplos.

(Emile Borel provou um teorema muito mais forte: o conjunto dos números $x \in [0, 1[$ tais que a representação em base 10 de x contém cada palavra finita b com frequência assimptótica igual a 1 sobre o comprimento de b tem probabilidade (medida de Lebesgue) igual a 1. Veja a observação sobre números normais no capítulo sobre a ergodicidade.)

10.6 Transformações expansoras do círculo

A multiplicação por 10 é um caso particular da seguinte classe de transformações.

Seja λ um inteiro tal que $|\lambda| > 1$. A *transformação expansora standard* de grau λ é a transformação $\times\lambda : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$, definida por

$$x + \mathbf{Z} \mapsto \lambda \cdot x + \mathbf{Z}$$

A transformação $\times\lambda$ é topologicamente mixing. De fato, todo aberto não vazio $U \subset \mathbf{R}/\mathbf{Z}$ contêm um intervalo I de comprimento $|I| > |\lambda|^{-n}$, para algum n suficientemente grande, e $(\times\lambda)^k (I) = \mathbf{R}/\mathbf{Z}$ para todo tempo $k \geq n$.

A transformação $\times\lambda$ tem um conjunto enumerável e denso de pontos periódicos.

A transformação $\times\lambda$ é um fator do shift de Bernoulli sobre um alfabeto de $|\lambda|$ letras (e o conjunto onde a semicojugação falha de ser injetiva é pequeno!).

As transformações expansoras, além de pontos periódicos densos e de órbitas densas, admitem trajetórias que se acumulam em conjuntos bem mais complicados. Por exemplo, a transformação $\times 3$ preserva o conjunto de Cantor standard K (pensado como um subconjunto do círculo), i.e. $\times 3(K) \subset K$. Agora, a restrição $\times 3 \,|_K : K \to K$ é topologicamente conjugada ao shift de Bernoulli $\sigma : \Sigma^+ \to \Sigma^+$ sobre o alfabeto $\{0, 2\}$ (quase!, de fato é um fator, pois a semiconjugação obvia não é injetiva no ponto $(2, 2, 2, 2, ...)$, que é igual a $(0, 0, 0, 0, ...)$ mod 1), que é topologicamente mixing, logo existem (e muitos!) pontos de $K \subset \mathbf{R}/\mathbf{Z}$ cuja $\times 3$-órbita é densa em K.

Seja agora $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ uma transformação expansora de classe $\mathcal{C}^1$, ou seja tal que um seu levantamento $F : \mathbf{R} \to \mathbf{R}$ é de classe $\mathcal{C}^1$. Sendo F' uma função periódica de período um, a expansividade de f implica que existe $\mu > 1$ tal que $|F'(x)| > \mu$ em todo $x \in \mathbf{R}$, e que F' não muda de sinal. Em particular, o grau de f tem módulo > 1, porque

$$|\deg(f)| = |F(1) - F(0)| = \left| \int_0^1 F'(t)dt \right| = \int_0^1 |F'(t)|\, dt > 1.$$

Teorema. *Toda transformação expansora $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ de classe $\mathcal{C}^1$ e grau λ é topologicamente conjugada à transformação expansora standard $\times\lambda$.*

dem. Por simplicidade, assumimos que λ seja um inteiro > 1. A ideia é construir uma conjugação entre os conjuntos das pré-imagens de um ponto fixo pelas iteradas de f e $\times\lambda$, e aproveitar do fato deles ser densos para extender a conjugação em todo o círculo. Sejam $x_k^i = i/\lambda^k$ com $i = 0, 1, ..., \lambda^k - 1$. Então $\times\lambda(x_k^i) = x_{k-1}^{i'}$, onde i' é o único inteiro entre 0 e $\lambda^{k-1} - 1$ tal que $i = i' \bmod \lambda^{k-1}$. Sejam F um levantamento de f, e p o ponto fixo de F. Como F é estrictamente crescente e $F(p+1) = p + \lambda$, existem $p = y_1^0 < y_1^1 < ... < y_1^{\lambda-1} < p + 1$ tais que $F(y_1^i) = p + i$. Indutivamente (em k) definimos os pontos y_k^i com $i = 0, 1, ..., \lambda^k - 1$ tais que

$$y_{k-1}^i = y_k^{\lambda i} < y_k^{\lambda i+1} < ...y_k^{\lambda i+\lambda-1} < y_k^{\lambda i+\lambda} = y_{k-1}^{i+1}$$

e $F(y_k^i) = y_{k-1}^{i'}$, onde i' é o único inteiro entre 0 e $\lambda^{k-1} - 1$ tal que $i = i' \bmod \lambda^{k-1}$. Para cada intervalo $I_k^i = \pi\left(\left[y_k^i, y_k^{i+1}\right]\right)$ temos que $f^k(I_k^i) = \mathbf{R}/\mathbf{Z}$. Como f é expansora, i.e. existe $\mu > 1$ tal que $|F'(x)| > \mu$ em todo ponto x, cada um desses intervalos tem comprimento $\left|I_k^i\right| < \mu^{-k}$, e portanto a família de pontos $\left\{y_k^i\right\}_{k\in\mathbf{N},\ i=0,1,...,m^k-1}$ é densa em $[p, p+1]$. A função

$$H : \left\{y_k^i\right\}_{k\in\mathbf{N},\ i=0,1,...,\lambda^k-1} \to \left\{x_k^i\right\}_{k\in\mathbf{N},\ i=0,1,...,\lambda^k-1}$$

definida por $H(y_k^i) = x_k^i$ é estritamente monótona. A densidade dos pontos $\left\{y_k^i\right\}$ e $\left\{x_k^i\right\}$ permite extender H como um homeomorfismo $H : [p, p+1] \to [0, 1]$, logo como um homeomorfismo $h : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$. Vê-se facilmente que $\times\lambda \circ h = h \circ f$. $\square$

Em particular, dada uma transformação expansora do círculo $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}/\mathbf{Z}$ de classe $\mathcal{C}^1$, toda transformação g suficientemente próxima de f na topologia $\mathcal{C}^1$ é topologicamente conjugada a f, porque a expansividade é uma condição aberta, e o grau é localmente constante (pois é uma função contínua com valores inteiros). Acabamos de provar o seguinte

Teorema. *As transformações expansoras do círculo de classe $\mathcal{C}^1$ são $\mathcal{C}^1$-estruturalmente estáveis.*

10.7 Problema: automorfismos hiperbólicos do toro

A expansividade não é necessária para induzir o mixing topológico. Foi o Dmitri Victorovich Anosov que, a partir dos exemplos geométricos dos fluxos geodésicos em superficies de curvatura negativa estudados por Hadamard, Hopf, ..., descobriu nos anos sessenta uma classe muito grande de transformações "desordenadas" e estruturalmente estáveis. O protótipo é a familia dos automorfismos hiperbólicos do toro.

O toro de dimensão dois é o espaço quociente $\mathbf{R}^2/\mathbf{Z}^2$. Uma aplicação linear $A : \mathbf{R}^2 \to \mathbf{R}^2$, definida por uma matriz 2×2 com coeficientes inteiros, induz uma aplicação contínua $f : \mathbf{R}^2/\mathbf{Z}^2 \to \mathbf{R}^2/\mathbf{Z}^2$, definida por

$$f\left(x + \mathbf{Z}^2\right) = A\left(x\right) + \mathbf{Z}^2$$

Se a matriz A tem determinante ±1, então também a sua inversa tem coeficientes inteiros, logo respeita o retículo $\mathbf{Z}^2$, e portanto f é invertível, é um "automorfismo" do toro. A existência de tais homeomorfismos é devida a razões aritméticas: as linhas e as colunas destas matrizes são pares de intéiros relativamente primos. Os automorfismos do toro que preservam a orientação, i.e. induzidos por matrizes com determinante $+1$, formam o grupo $SL(2, \mathbf{Z})$. Um exemplo é o automorfismo f induzido pela matriz

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

É imediato verificar que, se $x \in \mathbf{R}^2/\mathbf{Z}^2$ é um ponto periódico, então as suas coordenadas são racionais. Por outro lado, o conjunto dos pontos do toro cujas coordenadas são múltiplos inteiros de $1/n$ é um conjunto finito e +invariante. Isto implica que todos os pontos com coordenadas racionais são periódicos, e portanto que os pontos periódicos são densos. Também, é possível mostrar que $|\mathrm{Per}_n(f)| = |\det(A^n - \mathrm{id})|$. Os autovalores de A são

$$\lambda_+ = \frac{3+\sqrt{5}}{2} \quad \text{e} \quad \lambda_- = \frac{3-\sqrt{5}}{2}$$

Como A é simétrica, os vetores próprios são ortogonais. Resulta que $\mathbf{R}^2$ é a soma direta $E_+ \oplus E_-$ dos espaços próprios de A. A transformação A estica os vetores de E_+ pelo fator λ_+ e contrai os vetores de E_- pelo fator λ_-. Observe que f preserva a área, pois $\lambda_+ \cdot \lambda_- = 1$, mas não preserva as "formas". As projeções da linhas $x + E_\pm \subset \mathbf{R}^2$ em $\mathbf{R}^2/\mathbf{Z}^2$ contêm as órbitas de uma translação minimal do toro (porque $\lambda_\pm$ são irracionais) e portanto são densas. Seja R um pequeno quadrado com lados de comprimento ε paralelos às linhas $E_\pm$. A imagem $f^n(R)$ é um "retângulo" com lados $\varepsilon \cdot \lambda_+^n$ e $\varepsilon \cdot \lambda_-^n$, paralelos às linhas $E_\pm$, respetivamente. Quando n cresce, o complementar $\left(\mathbf{R}^2/\mathbf{Z}^2\right) \backslash f^n(R)$ não contém bolas de raio maior de ε, onde $\varepsilon \to 0$ quando $n \to \infty$, logo $f^n(R)$ interseta estavelmente cada aberto não vazio do toro. Isto mostra que f é topologicamente mixing. O resultado realmente interessante é o teorema de Anosov, que diz que f é $\mathcal{C}^1$-estruturalmente estável.

10.8 Problema: entropia topológica

O fenómeno da dependência sensível das condições iniciais por ser "quantificado", e isto produz um importante invariante topológico chamado "entropia"... Seja $f : X \to X$ uma transformação contínua do espaço topológico compacto X. Se d é uma métrica que induz a topologia de X, podemos definir uma famíla de métricas d_n, dependentes do tempo $n \geq 0$, por meio de

$$d_n(x, x') = \max_{0 \leq k \leq n} d\left(f^k(x), f^k(x')\right)$$

Ou seja, $d_n(x, x')$ é "a máxima distância entre as n-órbitas de x e x'". Dado $\varepsilon > 0$, sejam $S(n, \varepsilon)$ a mínima cardinalidade de um subconjunto $A \subset X$ tal que as d_n-bolas de raio ε centradas nos pontos de A cobrem X, e seja $T(n, \varepsilon)$ a máxima cardinalidade de um subconjunto $B \subset X$ tal que as d_n-bolas de raio ε centradas nos pontos de B são disjuntas. Se pensamos em ε como a precisão dos nossos instrumentos, $S(n, \varepsilon)$ representa "o número mínimo de n-órbitas necessárias para descrever as órbitas de todos os pontos de X com erro $\leq \varepsilon$", e $T(n, \varepsilon)$ representa "o número máximo de n-órbitas que os nossos instrumentos com sensibilidade ε conseguem distinguir". Se X é compacto, estes números são finitos, e crescem quando n cresce e quando ε decresce. A *entropia topológica* de f é

$$h_{\mathrm{top}}(f) = \lim_{\varepsilon \downarrow 0} \overline{\lim}_{n \to \infty} \frac{\log S(n, \varepsilon)}{n} = \lim_{\varepsilon \downarrow 0} \underline{\lim}_{n \to \infty} \frac{\log T(n, \varepsilon)}{n}$$

É possível provar que, de fato, este número não depende da métrica utilizada na sua definição, mas só da topologia de X e da transformação f. Isto implica que $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(f)$ se f e g são topologicamente conjugadas. Além disso, não é dificil provar que, como esperado, se g é um fator de f então $h_{\mathrm{top}}(g) \leq h_{\mathrm{top}}(f)$. As isometrias têm entropia topológica igual a zero, pois nesse caso as métricas d_n não dependem de n. Um bom exercício é calcular a entropia topológica das transformações expansoras do círculo, e descubrir que $h_{\mathrm{top}}(f) = \log(\deg(f))$, assim como dos shift de Bernoulli $\sigma : \Sigma^+ \to \Sigma^+$ num alfabeto de z letras, e descubrir que $h_{\mathrm{top}}(\sigma) = \log z$.

11 Ergodicity and convergence of time means

11.1 Ergodicity

Let $f : X \to X$ be an endomorphism of the measurable space $(X, \mathcal{E})$. The invariant probability measure μ is said *ergodic* if any of the following equivalent conditions is satisfied:

i) for any observable $\varphi \in L^1(\mu)$, the time average

$$\overline{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \varphi\left(f^k(x)\right)$$

exists and is equal to the mean value $\int_X \varphi d\mu$ for μ-almost any $x \in X$,

ii) any invariant event $A \in \mathcal{E}$ has probability $\mu(A) = 0$ or 1, namely the invariant σ-algebra $\mathcal{E}_f$ is equal to the trivial σ-algebra $\mathcal{N}$ generated by events of zero measure,

iii) any invariant (measurable) observable φ is constant μ-a.e.

If this happens, one also says that f is an *ergodic endomorphism* of the probability space $(X, \mathcal{E}, \mu)$.

Condition i) is the physical meaning of ergodicity, as it says that "time averages are almost everywhere constant and equal to space averages". In particular, taking φ equal to the characteristic function of any event A, almost any trajectory spend in A a fraction of time asymptotically proportional to $\mu(A)$, as dreamed by Boltzmann in his ergodic hypothesis.

Condition ii) is what one usually check in order to prove ergodicity of a probability measure. To see that i) $\Rightarrow$ ii), let A be an invariant event, and φ its characteristic function. Invariance of A implies that φ is invariant, hence that $\overline{\varphi} = \varphi$. There follows fom i) that $\mu(A) = \int_X \varphi d\mu = \varphi(x)$ for some $x \in X$, hence that $\mu(A) = 0$ or 1, the only values of characteristic functions.

Conditions ii) and iii) are clearly equivalent, since any invariant event defines an invariant function (its characteristic function), and conversly level sets of invariant functions are invariant events.

Finally, in order to show that iii) $\Rightarrow$ i), let $\varphi \in L^1(\mu)$ be an integrable observable. According to the Birkhoff-Khinchin ergodic theorem, the time average $\overline{\varphi}(x)$ exists for μ-almost any $x \in X$ and $\int_X \overline{\varphi} d\mu = \int_X \varphi d\mu$. Since $\overline{\varphi}$ is invariant mod 0, by iii) it is constant with probability one. This implies that $\overline{\varphi}(x) = \int_X \varphi d\mu$ for μ-almost any $x \in X$.

Warning. Ergodic dynamical systems exist, and some are listed below. On the other side, to show that a "physically interesting" system is ergodic turns out to be extremely difficult, and very few examples are known. The most famous are some "billards", systems of hard spheres inside a billard table interacting via elastic collisions, studied by Yakov Sinai in the sixties...

11.2 Ergodic measures as extremal measures, ergodic decomposition

We already saw that the space Prob_f of invariant probability measure is a convex and closed subset of the compact space Prob. Here, we observe that ergodic measures are the "indecomposable" elements of this set.

Proposition. *Ergodic invariant measures are the extremals of* Prob_f. *Namely, an invariant measure μ is ergodic iff it cannot be written as*

$$\mu = t\mu_1 + (1-t)\mu_0$$

where $t \in]0,1[$ and μ_0 and μ_1 are distinct invariant measures.

proof. First, observe that if ν is an invariant measure which is absolutely continuous w.r.t. the ergodic measure μ, then $\nu = \mu$. Indeed, one easily verifies that the Radon-Nykodim derivative $\rho = d\nu/d\mu$ is an invariant function, and ergodicity of μ implies that it is constant and equal to one μ-a.e. Now, let μ be an ergodic measure, and assume that $\mu = t\mu_1 + (1-t)\mu_0$ for some $t \in]0,1[$. Since both μ_0 and μ_1 are absolutely continuous w.r.t. μ, they coincide with μ, hence, are not different. To prove the converse, assume that the invariant measure μ is not ergodic, hence there exists an invariant event C such that $0 < \mu(C) < 1$. Let μ_0 and μ_1 be the "conditional probability measures" defined as $\mu_1(A) = \mu(A \cap C)/\mu(C)$ and $\mu_0(A) = \mu(A \cap C^c)/\mu(C^c)$. Clearly they are different, both are invariant, and $\mu = \mu(C)\mu_1 + (1-\mu(C))\mu_0$. □

In the first lines of the above proof, we actually showed that any two ergodic invariant measure μ and ν are either equal or "mutually singular", namely, if $\mu \neq \nu$ then there exists a measurable set A such that $\mu(A) = \nu(A^c) = 1$ and $\mu(A^c) = \nu(A) = 0$. This suggests that maybe any invariant measure could be "disintegrated" along a partition whose atoms are the support of all the different ergodic measure, in other word that μ is a "convex combination", namely an integral, of the ergodic measures. This is true, sometimes, but both its statement and proof are quite technical: we just quote the result.

Ergodic decomposition theorem. *Let $f : X \to X$ be a continuous transformation of the compact metrizable space X. There exists a partition $\mathcal{P} = \{P_e\}_{e\in E}$ of X (modulo sets of zero measure) into invariant measurable sets indexed by a Lebesgue space E, and a measurable map $E \ni e \mapsto \mu_e \in \mathrm{Prob}_f$ with values in the space of ergodic Borel probability measures and with the property that $\mu_e(P_e) = 1$ for any $P_e \in \mathcal{P}$, such that any invariant Borel probability measure μ can be written as an integral*

$$\mu = \int_E \mu_e d\overline{\mu}(e)$$

where $\overline{\mu}$ is some probability measure on E.

Observe that the above theorem contains the statement that any continuous transformation of a compact space admits at least one ergodic Borel probability measure.

11.3 Examples

Example: Bernoulli shift. *Let $\sigma : \Sigma^+ \to \Sigma^+$ be the Bernoulli shift over the alphabet $X = \{1, 2, ..., z\}$, and p any probability on X. The Bernoulli invariant measure μ defined by p is ergodic.* First observe that, given two centered cilinders C_α and C_β, the definition of μ implies that there exists a time $n \geq 1$ such

$$\mu\left(C_\alpha \cap \sigma^{-k}(C_\beta)\right) = \mu(C_\alpha)\cdot\mu\left(\sigma^{-k}(C_\beta)\right) = \mu(C_\alpha)\cdot\mu(C_\beta)$$

whenever $k \geq n$. Indeed, one can take $n = |\alpha| + 1$, and the above reflect the "independence" of the different trials encoded in the construction of the Bernoulli measure. By aditiviity, the same holds true for any couple of elements of $\mathcal{A}$, the algebra made of finite unions of centered cilinders. Now, assume that $A \in \mathcal{B}$ is invariant. Since any Borel set $A \in \mathcal{B}$ can be aproximated in measure by an elements of $\mathcal{A}$, given any $\varepsilon > 0$ one can find an $A_\varepsilon \in \mathcal{A}$ such that $\mu(A\Delta A_\varepsilon) < \varepsilon$. Using the above result, we can find an $n \geq 1$ such that

$$\mu\left(A_\varepsilon \cap \sigma^{-n}(A_\varepsilon)\right) = \mu(A_\varepsilon)\cdot\mu\left(\sigma^{-n}(A_\varepsilon)\right) = \mu(A_\varepsilon)^2$$

where the last equality comes from invariance of μ. Then, observe that the symmetric difference between $A \cap \sigma^{-n}(A)$ and $A_\varepsilon \cap \sigma^{-n}(A_\varepsilon)$ is contained in $(A\Delta A_\varepsilon) \cup \sigma^{-n}(A\Delta A_\varepsilon)$. This gives

$$\begin{aligned} \left|\mu\left(A \cap \sigma^{-n}(A)\right) - \mu\left(A_\varepsilon \cap \sigma^{-n}(A_\varepsilon)\right)\right| &\leq \mu(A\Delta A_\varepsilon) + \mu\left(\sigma^{-n}(A\Delta A_\varepsilon)\right) \\ &\leq 2\cdot\mu(A\Delta A_\varepsilon) < 2\varepsilon \end{aligned}$$

which, together with

$$\left|\mu(A)^2 - \mu(A_\varepsilon)^2\right| \leq 2\cdot\mu(A\Delta A_\varepsilon) < 2\varepsilon$$

gives

$$\left|\mu(A) - \mu(A)^2\right| < 4\varepsilon$$

Since $\varepsilon > 0$ was arbitrary, we just showed that the measure of any invariant Borel set A satisfies $\mu(A) = \mu(A)^2$, hence it is either 0 or 1. Observe that this proof is very similar to the argument in the Kolmogorov zero-one law for tail events in the theory of stochastic processes.

Now, let φ_k be the the characteristic function of $\{x \in \Sigma^+ \text{ s.t. } x_1 = k\}$. The observables $\varphi_k \circ \sigma^n$ form a sequence of independent and identically distributed random variables with mean p_k. One can interprete the event $\{\varphi_k \circ \sigma^n = 1\} = \{x \in \Sigma^+ \text{ s.t. } x_n = k\}$ as "sucess in the n-th trial", where the probability of sucess in each trial is p_k. The Birkhoff-Khinchin ergodic theorem, together with the ergodicity of μ, gives the result that

$$\mu \left\{ x \in \Sigma^+ \text{ s.t. } \frac{1}{n+1} \left(\varphi_k + \varphi_k \circ \sigma^1 + \varphi_k \circ \sigma^2 + ... + \varphi_k \circ \sigma^n \right)(x) \to p_k \right\} = 1$$

which is the Kolmogorov strong law of large numbers.

Example: expanding endomorphisms of the circle. *Let* $\times\lambda : x + \mathbf{Z} \mapsto \lambda \cdot x + \mathbf{Z}$*, with* $\lambda \in \mathbf{Z}$ *and* $|\lambda| > 1$*, be an expanding endomorphism of the circle. Lebesgue probability measure* ℓ *is an ergodic measure for* $\times\lambda$. To prove ergodicity, let A be an invariant Borel set, and assume that $\ell(A) < 1$. We must show that the complement $B = (\mathbf{R}/\mathbf{Z}) \backslash A$, that has positive measure, has indeed probability one. The argument goes as follows: if $\ell(\mathbf{B}) > 0$, then, according to Lebesgue density theorem, B contains nearly all the mass of some nonempty interval. Namely, given any $\varepsilon > 0$, we can find an open interval I_n with lenght $\ell(I_n) = |\lambda|^{-n}$ and centered at a density point of B such that

$$\ell(B \cap I_n) > (1 - \varepsilon) \cdot \ell(I_n)$$

Now observe that the restriction $(\times\lambda)^n |_{I_n}$ is an injective map sending I_n onto the circle minus one point, in particular, $\ell((\times\lambda)^n (I_n)) = 1$. Since $\times\lambda$ uniformly dilatates lenghts by a factor $|\lambda|$, there follows that

$$\frac{\ell((\times\lambda)^n (B \cap I_n))}{\ell((\times\lambda)^n (I_n))} = \frac{\ell(B \cap I_n)}{\ell(I_n)}$$

Since, moreover, A is invariant, its complement B is +invariant, and this implies that the left-hand side above is equal to $\ell(B)$. There follows that

$$\ell(B) = \frac{\ell(B \cap I_n)}{\ell(I_n)} > (1 - \varepsilon)$$

and, since ε was arbitrary, that $\ell(B) = 1$.

Observation/curiosity: normal numbers. In particular, Lebesgue measure ℓ is ergodic w.r.t. multiplication by 10 in the unit circle. Identify the circle with the interval $[0, 1[$, and let $x = 0, x_1 x_2 x_3 ...$ be the base 10 expression of a point of the circle, which is unique outside a subset of Lebesgue measure zero. For $k = 0, 1, 2, ..., 9$, let φ_k be the characteristic function of the interval $[k/10, (k+1)/10[$, i.e. the observable which is equal to $\varphi_k(x) = 1$ if $x_1 = k$ and $\varphi_k(x) = 0$ otherwise. The time mean of φ_k is

$$\frac{1}{n+1} \sum_{j=0}^{n} \varphi_k \left((\times 10)^j (x) \right) = \frac{1}{n+1} \cdot \text{card} \{ 1 \leq j \leq n+1 \text{ s.t. } x_j = k \}$$

that is the number of k's within the first $n+1$ digits of the decimal expansion of x. The limit as $n \to \infty$, if it exists, is the "asymptotic frequency" of k's contained in the expansion of x. Ergodicity of μ implies that there exists a set $A_k \subset [0, 1[$ of Lebesgue measure one where the limit $\overline{\varphi_k}(x)$ exists and is equal to $\int \varphi_k d\ell = 1/10$. Since the intersection $A_0 \cap A_1 \cap ... \cap A_9$ has still probability one, the result is that Lebesgue almost any number $x \in [0, 1[$ contains in its decimal expansion any of the letters $0, 1, 2, ..., 9$ with asymptotic frequency $1/10$.

Actually, one could repeat the same argument considering any finite word $b = b_1 b_2 ... b_n$ in the alphabeth $\{0, 1, 2, ..., 9\}$, and show that there is a set $A_b \subset [0, 1[$ of probability one such that the base 10 expansion of any $x \in A_b$ contains the word b with asymptotic frequency 10^{-n}. A real number x whose base 10 expansion contains any finite word with the right asymptotic frequency is called 10-*normal* (meaning "normal in base 10"). Since finite words in the alphabeth $\{0, 1, 2, ..., 9\}$ are countable, and a countable union of zero measure sets still has zero measure, we just showed that *Lebesgue almost any real number* x *is* 10*-normal.* This observation, and indeed the stronger statement that *Lebesgue almost any real number is normal in every base* $m \geq 2$, is due to Emile Borel (1909).

It is not difficult to give examples of normal numbers, actually of series whose sum is a normal number. Much more difficult is to show that a "given" real number, such as π or e ..., is normal. Here we quote Mark Kac: "as is often the case, it is much easier to prove that an overhelming majority of objects possess a certain property that to exhibit even one such object..."

11.4 Unique ergodicity

Unique ergodicity. A homeomorphism $f : X \to X$ of a compact metric space (X, d) is *uniquely ergodic* if it admits one, and only one, invariant Borel probability measure μ. The above discussion implies that this unique invariant measure is ergodic.

This notion is the probabilistic counterpart of minimality, and indeed both minimality and unique ergodicity are often observed simultaneously (this means that, although equivalence of the two is false, it is not easy to think at a couterexample!). Observe that we defined unique ergodicity in the context of continuous transformations. The reason is that this notion is interesting due to the following

Oxtoby's theorem. *Let $f : X \to X$ be a homeomorphism of a compact metric space X. The following statements are equivalent:*

i) f is a uniquely ergodic,

ii) there exists an invariant Borel probability measure μ such that, for any continuous observable φ, the time averages $\overline{\varphi}(x)$ exist and are equal to $\int_X \varphi d\mu$ for any initial condition $x \in X$,

iii) there exists an invariant Borel probability measure μ such that, for any continuous observable φ, the convergence

$$\frac{1}{n+1}\sum_{k=0}^{n}\varphi\left(f^k(x)\right) \to \int_X \varphi d\mu$$

as $n \to \infty$ holds and is uniform in $x \in X$.

Example: Kronecker-Weyl equidistribution theorem. *An irrational rotation of the circle is uniquely ergodic.* Indeed, let $+\alpha : x + \mathbf{Z} \mapsto x + \alpha + \mathbf{Z}$ be an irrational rotation. We must check that time means of continuous observables φ converge uniformly to the average $\int \varphi d\ell$, where ℓ is Lebesgue probability measure on the circle. Since, according to Weierstrass theorem, trigonometric polinomials are dense in the space of continuous functions of the circle, it suffices to check that the above holds for any of the functions $x \mapsto \varphi_k(x) = e^{\sqrt{-1}2\pi k x}$ with $k \in \mathbf{Z}$. A computation gives, for $k \neq 0$,

$$\left|\frac{1}{n+1}\sum_{j=0}^{n}\varphi_j\left((+\alpha)^j(x)\right)\right| = \left|\frac{1}{n+1}\sum_{j=0}^{n}e^{\sqrt{-1}2\pi k j\alpha}\right| \leq \frac{2}{n+1}\cdot\frac{1}{\left|1-e^{\sqrt{-1}2\pi k\alpha}\right|} \to 0$$

uniformly in x, while the time averages of φ_0 are constant and equal to 1. Hence, the time means of each φ_k converge uniformly to their space means as times goes to infinity, and we are done. The theorem owes its name to the fact that

$$\frac{1}{n+1}\sum_{j=0}^{n}\varphi(x+j\alpha) \to \int \varphi d\ell$$

uniformly for any continuous function φ on the circle, and this is interpreted as saying that the sequence of points $\{x, x+\alpha, x+2\alpha, x+3\alpha, ...\}$ is "equidistributed" w.r.t. Lebesgue measure.

Now, consider the torus $X = \mathbf{R}^n/\mathbf{Z}^n$ of dimension $n \geq 2$, and the linear flow $\phi_t : x + \mathbf{Z}^n \mapsto x + t\alpha + \mathbf{Z}^n$ defined by the differential equation

$$x' = \alpha$$

where $\alpha \in \mathbf{R}^n$. The "frequency vector" $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$ is said *non resonant* if the scalar product $\langle \alpha, k\rangle = \sum_{i=1}^{n}\alpha_i \cdot k_i \neq 0$ for any $k \in \mathbf{Z}^n \setminus \{0\}$. As above, one can approximate any continuous function on the torus with trigonometric functions. One then checks that

$$\frac{1}{T}\int_0^T e^{\sqrt{-1}2\pi\langle k, x+t\alpha\rangle}dt = \frac{e^{\sqrt{-1}2\pi\langle k,x\rangle}}{T\sqrt{-1}\langle k,x\rangle}\left(e^{\sqrt{-1}2\pi T\langle k,x\rangle}-1\right) \to 0$$

as $T \to \infty$, for any $k \in \mathbf{Z}^n \setminus \{0\}$, while the time mean of the observable 1 is constant and equal to one. There follows that *a non resonant linear flow on the torus is uniquely ergodic w.r.t. to Lebesgue measure.*

11.5 Continued fractions and Gauss map

References

[HK03] B. Hasselblatt and A. Katok, *A first course in dynamics: with a panorama of recent developments*, Cambridge University Press 2003.

[KH95] A. Katok and B. Hasselblat, *Introduction to the modern theory of dynamical systems*, Encyclopedia of mathematics and its applications, Cambridge University Press 1995.