

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

**SULL' ERGODICITA' DI UNA CLASSE
DI SISTEMI CAOTICI**

Relatore:
Prof. Marco Lenci

Presentata da:
Giovanni Canestrari

Anno Accademico 2017/2018

Sommario

Viene discussa l'importanza dell'Ipotesi Ergodica per sistemi termodinamici nell'ambito della Meccanica Statistica. Viene poi introdotta a livello matematico la Teoria Ergodica e vengono presentati alcuni esempi di sistemi dinamici e contestualmente alcuni metodi per dimostrarne l'ergodicità. In questo ambito viene introdotto il metodo di Hopf per la dimostrazione dell'ergodicità della mappa del 'Gatto di Arnold'. Viene in seguito presentata una generalizzazione discontinua di tale mappa ed illustrati i problemi nella realizzazione del metodo di Hopf che si presentano nel caso discontinuo. Viene quindi esposto il Teorema di Sinai e discusso il metodo di Sinai per l'ergodicità. Successivamente viene presentato il Teorema di Ergodicità Locale come conseguenza del metodo di Sinai. Infine, grazie a quest'ultimo teorema, viene studiato il caso più generale di un sistema dinamico formato da una generica mappa lineare iperbolica discontinua definita su un insieme di poligoni in \mathbb{R}^2 .

Indice

1	Introduzione	3
1.1	Misure termodinamiche	3
1.2	Il teorema di Liouville e l'ipotesi ergodica	6
1.3	Termodinamica e statistica	10
2	Introduzione alla teoria ergodica	11
2.1	Trasformazioni che conservano la misura	11
2.2	Ergodicità	13
2.3	Media temporale e media in fase	13
2.4	Esempi di sistemi dinamici ergodici	18
2.4.1	Rotazioni sul cerchio	19
2.4.2	Mappa del gatto di Arnold	21
3	L'esempio di Liverani e Wojtkowski	27
3.1	Una generalizzazione della mappa del gatto di Arnold	27
3.2	Varietà stabili e instabili	31
3.3	L'argomento di Hopf per le mappe discontinue	34
3.4	Il metodo di Sinai	38
4	Ergodicità per mappe iperboliche lineari a tratti	47
4.1	Definizione della mappa	48
4.2	Il teorema di ergodicità locale	50
4.3	Una estensione del teorema di ergodicità locale	54
A	Lemma di Borel-Cantelli	59
B	Dimostrazione del teorema di Sinai	60
	Bibliografia	63

Capitolo 1

Introduzione

In Termodinamica dimostrare che un sistema fisico è ergodico è di fondamentale importanza per poter implementare un modello teorico che descriva le quantità termodinamiche empiriche a partire dalla conoscenza dinamica dello stato del sistema. Laddove non sia possibile dimostrare che un sistema gode della proprietà dell'ergodicità (e questo è il caso della maggioranza dei sistemi termodinamici di interesse) è necessaria assumerla come ipotesi di lavoro. In questo caso si parla di Ipotesi Ergodica. Tale Ipotesi Ergodica può però essere confermata a posteriori dalla bontà della corrispondenza tra le previsioni ottenute dal modello teorico ed i dati sperimentali. Vediamo adesso nel dettaglio come l'ipotesi che un sistema sia ergodico permetta di connettere la conoscenza dinamica-microscopica del sistema a quella dello stato termodinamico.

1.1 Misure termodinamiche

Consideriamo un sistema classico il cui stato è descritto da una N -pla di parametri Lagrangiani (o coordinate generalizzate) $\bar{q} = (q_1, q_2, \dots, q_N)$ e dalla N -pla dei relativi momenti coniugati $\bar{p} = (p_1, p_2, \dots, p_N)$. Diciamo che N è il numero di gradi di libertà del sistema. Per tale sistema lo Spazio delle Fasi (SdF), ovvero l'insieme delle possibili configurazioni in cui può trovarsi il sistema, è dato dall'insieme delle $2N$ -ple $\bar{z} = (\bar{p}, \bar{q})$. In generale le coordinate del sistema potrebbero essere vincolate ad assumere solo certi valori, in modo da essere confinate nello spazio Ξ e dunque gli stati che può assumere il sistema sono:

$$SdF = \{ \bar{z} | \bar{q} \in \Xi, \bar{p} \in \mathbb{R}^N \}.$$

L'evoluzione del sistema all'interno dello SdF è descritta dalle equazioni di Newton o, più elegantemente, dalle equazioni di Hamilton, comprensive dei vincoli cui il sistema è

soggetto:

$$\begin{aligned}\dot{p}_j &= -\frac{\partial H(\bar{p}, \bar{q})}{\partial \dot{q}_j} \\ \dot{q}_j &= \frac{\partial H(\bar{p}, \bar{q})}{\partial p_j}\end{aligned}\tag{1.1.1}$$

Chiaramente non tutte le configurazioni dello spazio delle fasi saranno accessibili a un qualsiasi sistema: se, ad esempio, l'energia è conservata, il moto avviene su una superficie equienergetica $\Sigma(E)$ del sistema:

$$\Sigma(E) = \{\bar{z} : H(\bar{z}) = E\},$$

dove E è il valore iniziale dell'energia, determinato dalle condizioni iniziali. Se questo è vero per un sistema completamente isolato, è anche vero che il completo isolamento è una condizione più matematica che fisica, se non altro perché il semplice atto di misura di una quantità termodinamica, implica una qualche forma di interazione (e dunque scambio di energia), tra il sistema e l'ambiente. E' allora preferibile sostituire la iper-superficie del moto $\Sigma(E)$ con il volume nello spazio delle fasi $\Omega(E)$ così definito:

$$\Omega(E) = \left\{ (\bar{q}, \bar{p}) \in SdF \mid E - \frac{1}{2}\Delta \leq H(\bar{q}, \bar{p}) \leq E + \frac{1}{2}\Delta \right\},\tag{1.1.2}$$

dove Δ rappresenta la minima incertezza sull'energia non fisicamente eliminabile in principio. L'evoluzione temporale del sistema si ottiene risolvendo le $2N$ equazioni differenziali di Hamilton ed è dunque determinato univocamente dalle condizioni iniziali.

$$\bar{z} = \bar{z}(\bar{z}_0, t).$$

Definiamo il flusso di fase Φ^t come la trasformazione che associa ad ogni condizione iniziale il suo evoluto al tempo t :

$$\Phi^t(z_0) = \bar{z}(\bar{z}_0, t).$$

La conoscenza dinamica del sistema richiede dunque la conoscenza di $2N$ condizioni iniziali. Tuttavia l'esperienza empirica mostra che per i sistemi termodinamici più semplici (ad esempio per i gas perfetti) lo stato termodinamico è descritto da una sola *coppia* di parametri. In qualche modo c'è dunque una grande perdita di informazione nel passaggio da una descrizione meccanica-microscopica a una termodinamica-macroscopica. La descrizione formale di questa perdita di informazione si da assumendo che ogni funzione termodinamica f relativa ad una certa funzione dinamica $f(\bar{z})$ si ottenga da una media temporale:

$$\bar{f}_{sperimentale} = \bar{f}(T, V) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' f(\bar{z}(t', \bar{z}_0)),\tag{1.1.3}$$

dove sono stati scelti la temperatura e il volume come parametri che descrivono lo stato termodinamico del sistema. Tale interpretazione è coerente con l'idea secondo la quale una misura sperimentale termodinamica non registri i valori istantanei delle quantità fisiche di interesse, ma piuttosto una media temporale per tempi molto lunghi (nel limite, infiniti) [3].¹ Se prendiamo un sistema preparato in una certa condizione iniziale e lo mettiamo in un contenitore isolato e racchiuso in un volume finito, è un fatto sperimentale che quantità termodinamiche quali temperatura, pressione, etc... variano dapprima in maniera complicata ma, dopo aver aspettato a sufficienza, esse divengono pressoché costanti. Consideriamo il classico esempio di un gas in un cilindro munito di pistone. Supponiamo che il gas sia al tempo $t = 0$ tutto confinato nella metà inferiore del cilindro. Se misuriamo con un dinamometro la forza cui è soggetto il pistone in funzione del tempo, in $t = 0^+$ non misureremo alcuna forza, in $t > 0$, man mano che il gas raggiunge il pistone, misureremo una forza molto variabile dipendente dai moti convettivi del gas all'interno del cilindro. Dopo un certo tempo, a causa degli urti tra le molecole, i moti macroscopici si smorzano e la loro energia diventa energia termica. Le molecole non si muovono più *in gruppo*, ma tendono ad andare ognuna per conto proprio. A questo punto il gas ha raggiunto l'equilibrio termodinamico e la sua pressione è quella della legge di stato del gas. Tuttavia dobbiamo ricordare che l'equilibrio *termodinamico* non implica in nessun modo quello *dinamico*. La forza registrata dal dinamometro collegato al pistone non sarà costante: le molecole vicino al pistone, agitate dal moto termico, sbattono in modo discontinuo sulla parete del pistone causando fluttuazioni della misura. Tuttavia se il numero di molecole è grande, tali fluttuazioni saranno molto piccole rispetto al valore medio della forza misurato. Dal momento che però sperimentalmente le misure termodinamiche si stabilizzano deve per forza essere che quello che in realtà si ottiene dagli strumenti di misura è una media della grandezza misurata su di un certo intervallo di tempo. Più questo intervallo di tempo è grande più le fluttuazioni all'equilibrio diventano insignificanti rispetto al membro sinistro della (1.1.3). E' dunque ragionevole identificare il valore di una grandezza termodinamica all'equilibrio come la media su di un periodo di tempo molto grande dell'equivalente grandezza dinamica.

Tenendo ben presente l'evidenza empirica che è possibile effettuare misure termodinamiche, un fatto davvero singolare è che se decidiamo di identificare tali misure termodinamiche con le medie temporali della (1.1.3), abbiamo che esiste una precisa controparte matematica dell'esistenza di tale limite e dunque della possibilità di attuare misure termodinamiche. Come vedremo più in dettaglio nel prossimo capitolo, il teorema di Birkhoff afferma che, sotto blande ipotesi sulla funzione f , per quasi tutte le condizioni iniziali (cioè, per quasi tutti i punti \bar{z}_0 nello Spazio delle Fasi da cui si origina l'evoluzione

¹All'atto pratico il limite $t \rightarrow \infty$ si traduce in tempi anche molto brevi rispetto alla scala umana. Un termometro di quelli usati per misurare la febbre ha bisogno di circa dieci minuti per avere una buona misura della temperatura corporea. La chiave di interpretazione sta nel considerare il rapporto $\frac{t}{\tau}$, dove τ è il tempo di rilassamento che definisce la scala dei tempi per raggiungere l'equilibrio termico, con legge esponenziale $e^{-\frac{t}{\tau}}$. Dunque con $t \rightarrow \infty$ si intende semplicemente $t \gg \tau$.

del sistema) la media temporale nella (1.1.3) esiste ed è invariante sotto l'azione del flusso di fase generato dalle equazioni canoniche (1.1.1) [4]. Il teorema di Birkhoff richiede di introdurre il *framework* di Sistema Dinamico, che è sostanzialmente una generalizzazione di sistema Hamiltoniano (in cui il flusso di fase è dato dalle (1.1.1)). Per quanto riguarda il nostro studio, la caratteristica decisiva dei sistemi Hamiltoniani è quella presentata dal teorema di Liouville.

1.2 Il teorema di Liouville e l'ipotesi ergodica

Sia dato un ensemble di sistemi, ovvero un sistema insieme a un grande numero di copie mentali dello stesso. Ad ogni tempo t i vari membri dell'ensemble si disporranno in tutti i possibili microstati compatibili con un dato macrostato del sistema. Nello Spazio delle Fasi l'immagine corrispondente è quella di una moltitudine di punti, ognuno rappresentativo di una copia dell'ensemble, che occupano la *regione permessa* dello SdF (ad esempio se il sistema è conservativo, tutti questi punti devono appartenere a $\Sigma(E)$). Col passare del tempo, ciascun membro dell'ensemble è soggetto a un continuo cambiamento di microstato e, corrispondentemente, ogni punto dello SdF si muove lungo una traiettoria che definisce il flusso di fase. Sia d la dimensionalità di un certo sistema fisico. Per un sistema di $\frac{N}{d}$ particelle, possiamo illustrare tale movimento tramite la funzione ρ *densità degli stati*, definita in modo che $\rho(\bar{q}, \bar{p}; t)d^N \bar{q}d^N \bar{p}$ sia il numero di punti rappresentativi di un certo microstato all'interno di un volume $d^N \bar{q}d^N \bar{p}$ dello SdF centrato nel punto (\bar{q}, \bar{p}) . Chiaramente la funzione di densità $\rho(\bar{q}, \bar{p}; t)$ simboleggia il modo con cui i membri dell'ensemble sono distribuiti tra tutti i possibili microstati compatibili con un macrostato fissato al passare del tempo. Possiamo dare una interpretazione probabilistica alla funzione ρ se imponiamo la condizione di normalizzazione:

$$\int_{SdF} \rho(\bar{q}, \bar{p})d^N \bar{q}d^N \bar{p} = 1. \quad (1.2.1)$$

Facendo questo, grazie alla funzione di distribuzione ρ , è possibile definire una misura di probabilità μ sullo SdF. Sia $\mathcal{U} \subset SdF$, viene spontaneo definire

$$\mu(\mathcal{U}) = \int_{\mathcal{U}} \rho(\bar{q}, \bar{p})d^N \bar{q}d^N \bar{p}. \quad (1.2.2)$$

Si può dimostrare che μ è una misura grazie alle proprietà dell'integrale e alla non negatività della funzione ρ . L'interpretazione probabilistica è allora la seguente: data una certa funzione di distribuzione ρ possiamo affermare che la probabilità di avere il nostro sistema termodinamico \mathcal{S} in un determinato volume dello SdF è la seguente:

$$Prob\{\mathcal{S} \in \mathcal{U}\} = \int_{\mathcal{U}} \rho(\bar{q}, \bar{p})d^N \bar{q}d^N \bar{p} = \mu(\mathcal{U}). \quad (1.2.3)$$

In quest'ambito è possibile definire la *media in fase* $\langle f \rangle$ di una certa quantità fisica $f(\bar{q}, \bar{p})$ come:

$$\langle f \rangle = \int f(\bar{q}, \bar{p}) \rho(\bar{q}, \bar{p}; t) d^N \bar{q} d^N \bar{p}, \quad (1.2.4)$$

dove l'integrale è esteso a tutto lo SdF.

Il teorema di Liouville dimostra che i punti nello SdF si muovono lungo le superfici ad energia costante come farebbero le molecole di un fluido incompressibile.

Teorema 1.2.1. (di Liouville)[2] *La funzione $t \mapsto \rho(\bar{q}, \bar{p}; t)$ è costante:*

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0. \quad (1.2.5)$$

L'equazione (1.2.5) può anche essere scritta come

$$\mu(\Phi^t \mathcal{U}) = \mu(\mathcal{U}), \quad (1.2.6)$$

per ogni $\mathcal{U} \subset SdF$.

In generale la media in fase di f può essere essa stessa una funzione del tempo. Per escludere questa possibilità definiamo un ensemble stazionario come un ensemble in cui ρ non dipende esplicitamente dal tempo:

$$\frac{\partial \rho}{\partial t} = 0. \quad (1.2.7)$$

In questo caso infatti si vede subito dalla (1.2.4) che $\langle f \rangle$ non può dipendere da t e ρ definisce uno stato in equilibrio termodinamico. Se in particolare siamo interessati ad una misura effettuata su un corpo in equilibrio termodinamico, abbiamo che la (1.2.7) assieme alla (1.2.5) implicano:

$$\{\rho, H\} = 0. \quad (1.2.8)$$

Sia Ω il sottoinsieme dello spazio delle fasi, $\Omega \subset SdF$, interessato dall'evoluzione dinamica dei microstati del sistema, ovvero la regione nello spazio delle fasi in cui

$$E - \frac{1}{2}\Delta \leq H(\bar{q}, \bar{p}) \leq E + \frac{1}{2}\Delta. \quad (1.2.9)$$

Un modo in cui la (1.2.8) è banalmente soddisfatta è quello in cui

$$\rho(\bar{q}, \bar{p}) = \begin{cases} \text{costante se } E - \frac{1}{2}\Delta \leq H(\bar{q}, \bar{p}) \leq E + \frac{1}{2}\Delta \\ 0 \text{ altrimenti.} \end{cases} \quad (1.2.10)$$

Tale scelta per la ρ è solitamente indicata come il postulato di *probabilità uguali a priori* per i vari possibili microstati (o per i vari elementi di volume nello SdF) e la funzione di

distribuzione ρ così definita è detta *funzione di distribuzione microcanonica* ed il corrispondente ensemble *ensemble microcanonico*. In pratica, fintanto che un certo microstato è compatibile con le condizioni macroscopiche fissate del sistema, allora avrà la stessa probabilità di realizzarsi di ogni altro. Vedremo tra poco che esiste una stretta relazione tra questa scelta per la funzione di distribuzione ρ (che per ora abbiamo chiamato postulato) e l'ergodicità (o ipotesi di ergodicità) di un certo sistema termodinamico. Per adesso diamo la seguente definizione non rigorosa² di trasformazione ergodica di un sistema termodinamico.

Definizione 1.2.2. (di ergodicità) Un sistema termodinamico \mathcal{S} è detto ergodico se per ogni osservabile termodinamico f^3 vale [3]:

$$\bar{f}(z_0) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau f(\Phi^t(z_0)) = \text{costante}, \quad (1.2.11)$$

per quasi ogni $z_0 \in SdF$.

Dalla (1.2.2) e (1.2.4) possiamo ottenere la μ definita dalla ρ microcanonica. Tenendo presente che in questo caso $\rho = 0$ al di fuori della *iper-shell* del moto abbiamo che se \mathcal{U} è un sottoinsieme dello SdF, allora :

$$\mu(\mathcal{U}) = \int_{\Omega \cap \mathcal{U}} d^N \bar{q} d^N \bar{p}, \quad (1.2.12)$$

avendo posto la condizione di normalizzazione (1.2.1), che in questo caso diventa:

$$\mu(SdF) = \mu(\Omega) = 1.$$

La media in fase per l'*ensemble microcanonico* di una certa grandezza fisica f sarà invece data dalla (1.2.4) ed è data dalla media pesata sulla probabilità μ costante sui valori che

²Riprenderemo più rigorosamente questi argomenti nel prossimo capitolo

³Tale definizione non è rigorosa perché non si è detto esattamente cos'è un osservabile termodinamico. Dal punto di vista matematico bisognerebbe richiedere che \bar{f} sia costante quasi dovunque per ogni $f \in L^1(\Omega, \mu)$. Ad ogni modo, vista la definizione (1.2.9) di Ω , la stessa Hamiltoniana (che oltretutto è uno dei più importanti osservabili termodinamici) è una funzione non costante in Ω . A tale obiezione si potrebbe rispondere affermando che l'energia (insieme agli altri osservabili termodinamici) sono funzioni già di per loro *quasi* costanti, ovvero assumono per continuità valori molto prossimi tra loro in quasi tutte le *zone* dello SdF accessibili al sistema. Dunque, anche avendo a disposizione un sistema non completamente ergodico, il fatto di interessarsi tra tutte le funzioni appartenenti ad $L^1(\Omega, \mu)$ ad un numero ristrettissimo di esse che sono già *quasi* costanti sullo SdF, permette comunque di ottenere la costanza delle medie temporali e l'uguaglianza di queste con la media in fase. In effetti l'ipotesi ergodica è comunque più di quello che serve alla termodinamica: si potrebbe infatti pensare ad un sistema decomponibile (quindi non ergodico) su cui però le medie temporali delle funzioni di interesse termodinamico (energia, pressione, etc...) sono costanti con tale costante uguale nelle due componenti ergodiche. In questo caso è però possibile immaginare l'esistenza di funzioni non 'fisiche' le cui medie temporali assumono valori diversi nelle due diverse componenti ergodiche [1].

assume tale funzione sui microstati compatibili con un macrostato dato (cioè sulla regione Ω dello SdF):

$$\langle f \rangle = \int_{\Omega} f(\bar{q}, \bar{p}) d^N \bar{q} d^N \bar{p}. \quad (1.2.13)$$

Vediamo che la (1.2.12) assegna una probabilità costante ad ogni microstato relativo ad uno stesso macrostato termodinamico: negli ensemble microcanonici infatti non ci sono microstati privilegiati, tutti quelli compatibili con un certo macrostato danno lo stesso contributo alla media in fase.

Il calcolo teorico delle medie temporali delle quantità termodinamiche (1.1.3) è praticamente impossibile per un sistema di molte particelle perché implicherebbe una soluzione esplicita del flusso di fase Φ^t . In questo senso l'ipotesi ergodica diventa cruciale per poter confrontare le previsioni teoriche con i risultati sperimentali. Infatti, dal momento che l'ensemble che stiamo considerando (quello da cui vogliamo estrarre una misura termodinamica) è stazionario, per ogni quantità fisica f , $\langle f \rangle$ è indipendente dal tempo per la (1.2.7) ed abbiamo allora, per lo meno formalmente ⁴:

$$\begin{aligned} \langle f \rangle &= \frac{1}{\mu(\Omega(E))} \int_{\Omega(E)} f(\bar{z}) d^N \bar{q} d^N \bar{p} = \lim_{t \rightarrow \infty} \frac{1}{\mu(\Omega(E))} \frac{1}{t} \int_0^t d\tau \int_{\Omega(E)} d^N \bar{q} d^N \bar{p} f(\bar{z}) \\ &= \lim_{t \rightarrow \infty} \frac{1}{\mu(\Omega(E))} \frac{1}{t} \int_0^t d\tau \int_{\Omega(E)} f(\Phi^t(\bar{z})) d^N \bar{q} d^N \bar{p} \\ &= \frac{1}{\mu(\Omega(E))} \int_{\Omega(E)} d^N \bar{q} d^N \bar{p} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau f(\Phi^t(\bar{z})) = \bar{f}, \end{aligned} \quad (1.2.14)$$

dove $\bar{z} = (\bar{q}, \bar{p})$. La terza uguaglianza è il teorema di Liouville mentre la penultima uguaglianza è possibile qualora si richieda che per quasi ogni $\bar{z} \in \Omega$, $\bar{f} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau f(\Phi^t(\bar{z}))$ sia costante, cioè proprio qualora si supponga che un certo sistema termodinamico sia ergodico [3]. L'ipotesi ergodica ci fornisce dunque una ricetta per dedurre teoricamente i risultati delle misure termodinamiche⁵. Questo perché:

- seleziona la distribuzione *microcanonica* di probabilità (ed il corrispondente *ensemble microcanonico*) per un sistema termodinamico isolato tra tutte le distribuzioni compatibili con il teorema di Liouville
- garantisce l'uguaglianza tra la media temporale e quella in fase per le quantità termodinamiche.

⁴Vedremo nel prossimo Capitolo che questa proprietà è una delle conseguenze del fondamentale Teorema 2.3.4 Capitolo 2 Ergodico di Birkhoff.

⁵In realtà è possibile formulare diverse obiezioni alla idea secondo cui 'l'ipotesi ergodica sia la giustificazione teorica all'uso dell'ensemble microcanonico per la descrizione degli stati di equilibrio'. Per una discussione approfondita vedi [1] e la nota 3 nella pagina precedente.

1.3 Termodinamica e statistica

Da quanto visto può risultare interessante fare, con le dovute cautele, le seguenti identificazioni:

$$\begin{aligned}
 &\text{esistenza delle misure termodinamiche} \iff \text{Teorema di Birkhoff}^6 \\
 &\quad \text{ensemble statistico} \iff \text{misura di probabilità} \\
 &\quad \text{ensemble microcanonico} \iff \text{misura di probabilità costante} \\
 &\text{teorema di Liouville} \iff \text{trasformazioni che conservano la misura} \\
 &\quad \bar{f}_{\text{sperimentale}} = \langle f \rangle_{\text{teorica}} \iff \text{ipotesi ergodica.}
 \end{aligned}$$

Nei capitoli successivi cercheremo di inquadrare meglio e più rigorosamente il concetto di ergodicità, senza dimenticare la fondamentale importanza che tale ipotesi riveste per poter costruire un modello termodinamico dei sistemi fisici riconducibile alle leggi della Meccanica (classica o quantistica). Questa prospettiva, intuita dal genio di Boltzmann e ancora prima dai pionieri⁷ della *teoria cinetica dei gas*, vale per eccesso gli sforzi fatti in passato e quelli che molti ricercatori stanno tutt'ora facendo nel campo della teoria ergodica. Lo scopo del presente lavoro è quello di ricapitolare la dimostrazione dell'ergodicità di un particolare sistema dinamico caratterizzato da linee di discontinuità della sua mappa tramite il metodo di Sinai e di ampliare la applicazione di questo metodo allo studio di un dominio più generale di quello considerato in un primo momento. Per completezza si riporta che, anche se non approfondite in questo lavoro, ci sono varie critiche all'idea della ipotesi ergodica come fondamento della Meccanica Statistica (vedi ad esempio nota 3 a pagina 8) e sono conosciuti approcci alternativi a tale disciplina.

⁶Abbiamo già menzionato che il Teorema di Birkhoff garantisce l'esistenza del limite nel membro destro della (1.1.3), enunceremo più precisamente questo teorema nel prossimo Capitolo.

⁷Tra i quali figurano ad esempio Bernoulli, Herapath, Joule, Krönig che derivò la legge per la pressione di un gas perfetto in una scatola $p = \frac{1}{3}nmc^2$, con c velocità media delle molecole assunta costante, Clausius e Maxwell che ricavò teoricamente la distribuzione della velocità delle molecole di un gas perfetto. Per una introduzione storica alla Meccanica Statistica si veda l'introduzione di [2].

Capitolo 2

Introduzione alla teoria ergodica

In questo capitolo spiegheremo più precisamente cosa si intende per trasformazione ergodica. Enunceremo inoltre il teorema di Birkhoff che descrive le trasformazioni ergodiche come quelle per cui la media temporale è uguale alla media in fase. Come abbiamo visto nell'introduzione è questa la caratterizzazione delle trasformazioni ergodiche cruciale per la Termodinamica. Seguirà poi la discussione di due modelli molto semplici di trasformazioni ergodiche: la rotazione sul cerchio unitario di un angolo irrazionale e la famosa mappa del gatto di Arnold. Nell'ambito di quest'ultimo esempio presenteremo l'argomento di Hopf per mappe continue, evidenziandone i limiti nell'estensione per mappe discontinue. E' infine presentata l'idea generale alla base della *catena di Hopf*, che verrà approfondita nei capitoli successivi.

2.1 Trasformazioni che conservano la misura

Per definire una trasformazione ergodica, bisogna in primo luogo che questa sia misurabile.

Definizione 2.1.1. (di trasformazione misurabile) Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}))$ uno spazio misurabile. $\mathfrak{B}(\mathcal{M})$ è la σ -algebra dei sottoinsiemi misurabili di \mathcal{M} . Allora:

- (a) Una trasformazione $T: \mathcal{M} \rightarrow \mathcal{M}$ è misurabile se $T^{-1}(\mathfrak{B}(\mathcal{M})) \subseteq \mathfrak{B}(\mathcal{M})$ (cioè se $B \in \mathfrak{B}(\mathcal{M}) \Rightarrow T^{-1}B \in \mathfrak{B}(\mathcal{M})$)
- (b) Sia $T: \mathcal{M} \rightarrow \mathcal{M}$. Diciamo che T è un automorfismo se T è biettiva e sia T che T^{-1} sono misurabili.

Le iterazioni positive $\{T^n\}$, $n \geq 0$, di una trasformazione misurabile T formano un semigrupp; tutte le iterazioni T^n , $n \in \mathbb{Z}$, di un automorfismo T formano un gruppo. Per ogni $x \in \mathcal{M}$ la sequenza $\{T^n(x)\}$ è chiamata traiettoria o orbita di x . Indichiamo con $\mathfrak{M}(\mathcal{M})$ l'insieme di tutte le misure di probabilità su $(\mathcal{M}, \mathfrak{B}(\mathcal{M}))$.

Definizione 2.1.2. (di trasformazione che conserva la misura) Sia data una misura $\mu \in \mathfrak{M}(\mathcal{M})$. Diciamo che una trasformazione $T : \mathcal{M} \rightarrow \mathcal{M}$ conserva la misura se $\mu(T^{-1}B) = \mu(B) \forall B \in \mathfrak{B}(\mathcal{M})$ (diciamo che μ è T -invariante e scriviamo $T\mu = \mu$). Se T è un automorfismo, $T\mu = \mu$ è equivalente a $T^{-1}\mu = \mu$, di modo che se T conserva una certa misura $\mu \in \mathfrak{M}(\mathcal{M})$ allora anche T^{-1} conserva μ .

L'ambiente generale su cui lavoreremo è quello dei sistemi dinamici.

Definizione 2.1.3. (di sistema dinamico che preserva una misura) Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}))$ un sistema misurabile e μ una misura su \mathcal{M} . Sia inoltre T una trasformazione da \mathcal{M} in se stesso, $T : \mathcal{M} \rightarrow \mathcal{M}$. Se T conserva la misura μ , allora $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ è detto sistema dinamico che preserva una misura.

Una proprietà che è condivisa da tutte le trasformazioni che conservano una misura finita è la ricorrenza:

Teorema 2.1.4. (di Poincaré) [7] Sia $T : \mathcal{M} \rightarrow \mathcal{M}$ una trasformazione che conserva la misura sullo spazio di probabilità $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$. Sia $E \in \mathfrak{B}$ con $\mu(E) > 0$. Allora quasi tutti i punti di E ritornano inifinitamente spesso in E sotto iterazioni positive di T (cioè, esiste $F \subset E$ con $\mu(F) = \mu(E)$ tale che per ogni $x \in F$ esiste una sequenza $n_1 < n_2 < \dots$ di numeri naturali con $T^{n_i}(x) \in F$ per ogni i).

La nozione di ergodicità di una data mappa si fonda su quella di insiemi invarianti (mod 0) per quella mappa. Siano A e B due insiemi definiamo la *differenza simmetrica* tra A e B , che indichiamo con $A\Delta B$, come:

$$A\Delta B = (A \cup B) \setminus (A \cap B).$$

Osserviamo che $\mu(A\Delta B) = 0$ è equivalente a dire che A e B coincidono a meno di una misura nulla di punti. Diamo adesso la seguente:

Definizione 2.1.5. (di insieme invariante) Diciamo che un insieme misurabile $E \subset \mathcal{M}$ è T -invariante se $T^{-1}E = E$. Si noti che questo accade se $T(E) \subset E$ e $T(E^c) \subset E^c$, dove $E^c = \mathcal{M} \setminus E$. Supponiamo che T preservi la misura μ . Allora un insieme misurabile E è detto T -invariante (mod 0) se $\mu(T^{-1}E\Delta E) = 0$. Se E è T -invariante (mod 0), allora esiste un insieme \tilde{E} , T -invariante, tale che $\tilde{E} = E$ (mod 0).

Da ora in poi, se non diversamente specificato, quando parleremo di insieme invariante sottointenderemo la invarianza mod 0.

Definizione 2.1.6. (di sistema dinamico decomponibile) Sia $T : \mathcal{M} \rightarrow \mathcal{M}$. Diciamo che il sistema dinamico $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ è decomponibile se è possibile scrivere \mathcal{M} come unione di due insiemi disgiunti di misura positiva, $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, $\mathcal{M}_1 \cap \mathcal{M}_2 = \emptyset$, tali che

$$T^{-1}(\mathcal{M}_1) = \mathcal{M}_1 \text{ mod } 0 \quad T^{-1}(\mathcal{M}_2) = \mathcal{M}_2 \text{ mod } 0.$$

Una funzione che assume lo stesso valore su tutte le orbite di una data mappa è detta invariante rispetto a quella mappa.

Definizione 2.1.7. (di funzione invariante) Una funzione (o osservabile) $f : \mathcal{M} \rightarrow \mathbb{R}$ è T -invariante se $f = f \circ T$. Se T preserva la misura μ , allora diciamo che una funzione $f : \mathcal{M} \rightarrow \mathbb{R}$ è T -invariante (mod 0) se $f(x) = f(T(x))$ quasi dovunque secondo la misura μ .

A volte capita che lo studio di una mappa definita sull'insieme \mathcal{M} possa essere condotto, senza perdere proprietà importanti della mappa, su insiemi più piccoli di \mathcal{M} . Questa idea di decomposizione e semplificazione del dominio di una certa mappa è fondante per l'ergodicità.

2.2 Ergodicità

Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ uno spazio di probabilità e $T : \mathcal{M} \rightarrow \mathcal{M}$ una trasformazione che conserva la misura. Sia $E \in \mathfrak{B}(\mathcal{M})$ ed $E^c = \mathcal{M} \setminus E$. Supponiamo che $T^{-1}E = E$, allora anche $T^{-1}(E^c) = E^c$ e possiamo studiare T concentrandoci sulle due trasformazioni più semplici $T|_E$ e $T|_{E^c}$. Se $0 < \mu(E) < 1$ questo ha semplificato lo studio di T . Se invece $\mu(E) = 0$ o $\mu(E^c) = 0$ possiamo ignorare E o E^c e non abbiamo significativamente semplificato lo studio di T , dal momento che ignorare un insieme di misura nulla è consentito in teoria della misura. Questo porta all'idea di studiare quelle trasformazioni che non possono essere decomposte e di provare ad esprimere ogni trasformazione che conserva la misura in termini di altre indecomponibili. Tali trasformazioni indecomponibili sono chiamate ergodiche.

Definizione 2.2.1. (di trasformazione ergodica) Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ uno spazio di probabilità. Una trasformazione T che conserva la misura su $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ è chiamata ergodica se gli unici membri E di $\mathfrak{B}(\mathcal{M})$ tali che $T^{-1}(E) = E$ soddisfano $\mu(E) = 0$ o $\mu(E) = 1$.

In altre parole un sistema è ergodico se non è decomponibile.

2.3 Media temporale e media in fase

Vedremo che nel caso di mappe iperboliche non sarà semplice dimostrare l'ergodicità così come la abbiamo definita. Illustreremo però un metodo, detto *catena di Hopf* o *delle linee a zig zag* o *us-path* (*stable-unstable path*), parecchio ingegnoso, che permetterà di dimostrare l'appartenenza di un insieme ad un'unica componente ergodica¹, dimostrando

¹La nozione di componente ergodica di un sistema dinamico che preserva la misura è ben definita matematicamente (si veda ad esempio [7]), tuttavia esula dagli scopi di questa tesi richiamarla. Qui

che la media temporale nel passato e nel futuro di una qualsiasi funzione continua è costante su quasi tutto l'insieme.

Definizione 2.3.1. (di media temporale) Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ un sistema dinamico. Sia $f : \mathcal{M} \rightarrow \mathbb{R}$. La media temporale nel futuro f^+ di f è definita $\forall x \in \mathcal{M}$ da:

$$f^+(x) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)),$$

qualora il limite esista. Viceversa la media temporale nel passato f^- di f è definita $\forall x \in \mathcal{M}$ da:

$$f^-(x) = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^{-n}(x)),$$

qualora il limite esista.

Osservazione 2.3.2. E' possibile definire una media temporale nel futuro e nel passato anche nel caso di sistemi dinamici continui. Sia $T_t : \mathcal{M} \rightarrow \mathcal{M}$ un gruppo a un parametro di diffeomorfismi che preservano la misura. In questo caso, se esistono:

$$f^+(x) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(T_t(x)) dt,$$

per $x \in \mathcal{M}$, $t \in \mathbb{R}$.

$$f^-(x) = \lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T f(T_{-t}(x)) dt,$$

per $x \in \mathcal{M}$, $t \in \mathbb{R}$.

Definizione 2.3.3. (di media in fase) Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ uno spazio di probabilità ($\mu(\mathcal{M}) = 1$). Definiamo la media in fase di una funzione $f : \mathcal{M} \rightarrow \mathbb{R}$, come:

$$\bar{f} = \int_{\mathcal{M}} f(x) d\mu.$$

In generale la media temporale e la media in fase sono diverse, ma se la trasformazione T è ergodica si dimostra che coincidono. A tal proposito riportiamo il seguente teorema, uno dei primi e più importanti risultati della teoria ergodica.

Teorema 2.3.4. (*Teorema Ergodico di Birkhoff*)[7] Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ un sistema dinamico che conserva la misura. Supponiamo che T sia una trasformazione misurabile. Sia inoltre $f : \mathcal{M} \rightarrow \mathbb{R}$ ed $f \in L^1(\mathcal{M}, \mu)$, una funzione a valori reali (o complessi) μ -sommabile su \mathcal{M} . Allora:

la citiamo solo per la sua valenza intuitiva. In particolare in questo lavoro non ci occuperemo mai di componenti ergodiche di misura nulla e quindi l'espressione *componente ergodica* sarà sinonimo di 'insieme invariante mod 0 non decomponibile in insiemi invarianti di misura positiva'.

- (i) per quasi ogni punto $x \in \mathcal{M}$, esisitono i limiti f^+ della Definizione 2.3.1;
- (ii) f^+ è T -invariante, nel senso che, $\forall x \in \mathcal{M}$ in cui f^+ esiste si ha che $f^+(T^n(x)) = f^+(x)$ per ogni $n \in \mathbb{Z}^+$
- (iii)

$$\int_{\mathcal{M}} f^+(x) d\mu = \int_{\mathcal{M}} f(x) d\mu.$$

Se T è una trasformazione *invertibile*, è chiaramente possibile enunciare il Teorema di Birkhoff anche per le medie temporali nel passato.

Corollario 2.3.5. [7] *Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ un sistema dinamico che conserva la misura. Supponiamo che T sia un automorfismo, ovvero che T sia invertibile e T^{-1} sia misurabile. Sia inoltre $f : \mathcal{M} \rightarrow \mathbb{R}$ ed $f \in L^1(\mathcal{M}, \mu)$, una funzione a valori reali (o complessi) μ -sommabile su \mathcal{M} . Allora:*

- (i) per quasi ogni punto $x \in \mathcal{M}$, esisitono i limiti f^+ ed f^- della Definizione 2.3.1;
- (ii) f^+ ed f^- sono T -invarianti, nel senso che, $\forall x \in \mathcal{M}$ in cui f^+ esiste si ha che $f^+(T^n(x)) = f^+(x)$ per ogni $n \in \mathbb{Z}^+$ e $\forall x \in \mathcal{M}$ in cui f^- esiste si ha che $f^-(T^{-n}(x)) = f^-(x)$ per ogni $n \in \mathbb{Z}^+$
- (iii)

$$\int_{\mathcal{M}} f^+(x) d\mu = \int_{\mathcal{M}} f(x) d\mu, \quad \int_{\mathcal{M}} f^-(x) d\mu = \int_{\mathcal{M}} f(x) d\mu.$$

Corollario 2.3.6. *Supponiamo che valgano tutte le ipotesi del Teorema di Birkhoff, incluso il fatto che T sia un automorfismo in modo che valga anche il Corollario (2.3.5). Allora*

$$f^+(x) = f^-(x),$$

per quasi ogni $x \in \mathcal{M}$.

Dimostrazione. Sia

$$\mathfrak{A}_+ = \{x \in \mathcal{M} | f^+(x) > f^-(x)\}.$$

Per definizione \mathfrak{A}_+ è un insieme invariante (mod 0), in quanto f^+ ed f^- sono quasi ovunque invarianti per il teorema di Birkhoff. Abbiamo allora che

$$\int_{\mathfrak{A}_+} [f^+(x) - f^-(x)] d\mu = \int_{\mathfrak{A}_+} f(x) d\mu - \int_{\mathfrak{A}_+} f(x) d\mu = 0, \quad (2.3.1)$$

dove la prima uguaglianza è giustificata sempre dal teorema di Birkhoff. Infatti se \mathfrak{A}_+ è un insieme invariante e T è un automorfismo, allora sono soddisfatte tutte le ipotesi del teorema di Birkhoff sul sistema dinamico (\mathfrak{A}_+, μ, T) ed in particolare il punto (iii)

del teorema è la prima uguaglianza dell'espressione sopra. Siccome l'integrando è per definizione positivo, la (2.3.1) implica che $\mu(\mathfrak{A}_+) = 0$, cioè $f^+ \leq f^-$ quasi dovunque secondo la misura μ . Lo stesso identico argomento può essere ripetuto per l'insieme:

$$\mathfrak{A}_- = \{x \in \mathcal{M} \mid f^-(x) \geq f^+(x)\}.$$

Il risultato è che anche $\mu(\mathfrak{A}_-) = 0$, e che quindi $f^+ \geq f^-$ quasi dovunque secondo la misura μ . Combinando quanto ottenuto si vede che deve necessariamente essere che $f^+ = f^-$ quasi dovunque secondo la misura μ . \square

Corollario 2.3.7. *Un sistema dinamico $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ è ergodico se e solo se per ogni funzione a valori reali (o complessi) μ -sommabile $f \in L^1(\mathcal{M}, \mu)$ la media temporale è uguale alla media in fase quasi ovunque:*

$$f^+(x) = \bar{f}(x) \quad \text{q.o.} \quad (2.3.2)$$

Dimostrazione.

(\Rightarrow) Vediamo che se il sistema non è ergodico allora $f^+(x) \neq \bar{f}(x)$ su un insieme di misura positiva.

Se (\mathcal{M}, μ, T) non è ergodico allora è decomponibile, ovvero \mathcal{M} può essere scritto come unione di due insiemi disgiunti $\mathcal{M}_1, \mathcal{M}_2$ tali che

$$T(\mathcal{M}_1) = \mathcal{M}_1 \text{ mod } 0 \quad T(\mathcal{M}_2) = \mathcal{M}_2 \text{ mod } 0.$$

Sia $f : \mathcal{M} \rightarrow \mathbb{R}$ una funzione così definita:

$$f(x) = \begin{cases} 1 & x \in \mathcal{M}_1 \\ 0 & x \in \mathcal{M}_2. \end{cases}$$

Chiaramente $f \in L^1(\mathcal{M}, \mu)$ dal momento che è la funzione caratteristica di $\mathcal{M}_1 \subset \mathcal{M}$ e $\mu(\mathcal{M}) = 1$. Abbiamo però che $f^+(x) = f(x)$ per quasi ogni $x \in \mathcal{M}$ ed $f(x)$ assume due valori diversi su due sottoinsiemi di \mathcal{M} di misura positiva. Allora $f^+(x)$ non potrà mai essere costante quasi dovunque, come richiesto dalla (2.3.2).

(\Leftarrow) Vediamo che se $f^+(x) \neq \bar{f}(x)$ su un insieme di misura positiva, allora il sistema non è ergodico.

Se esiste una funzione $f \in L^1(\mathcal{M}, \mu)$ (che assumiamo senza perdita di generalità a valori reali) la cui media temporale (che ricordiamo essere quasi-ovunque definita) non

è quasi-ovunque costante ² possiamo definire:

$$\mathcal{M}_1 = \{x \in \mathcal{M} | f^+(x) < a, a \in \mathbb{R}\}, \quad \mathcal{M}_2 = \{x \in \mathcal{M} | f^+(x) \geq a, a \in \mathbb{R}\}.$$

Deve chiaramente esistere un a tale che

$$\mu(\mathcal{M}_1) > 0, \quad \mu(\mathcal{M}_2) > 0.$$

Per il teorema di Birkhoff f^+ è (mod 0) invariante in \mathcal{M} ; ne consegue che, sempre mod 0,

$$T^{-1}\mathcal{M}_1 = \mathcal{M}_1, \quad T^{-1}\mathcal{M}_2 = \mathcal{M}_2,$$

ed il sistema è decomponibile. \square

E' interessante notare che la definizione storicamente esatta (e anche visivamente più immediata, nonché fisicamente più significativa) è proprio quella presentata nel Corollario 2.3.7, che è riassumibile nel celebre detto '*ergodicità significa media temporale = media in fase*'.

Osservazione 2.3.8. Si può dire anche di più di quanto già detto. Infatti riguardando la dimostrazione del Corollario 2.3.7, si vede che continua a funzionare se al posto di considerare f^+ con $f \in L^1(\mathcal{M}, \mu)$ consideriamo una qualsiasi funzione T -invariante $g \in L^1(\mathcal{M}, \mu)$. Abbiamo allora che:

$$(\mathcal{M}, \mu, T) \text{ è ergodico} \Leftrightarrow \forall g \in L^1(\mathcal{M}, \mu) \text{ } T\text{-invariante, } g = \text{costante mod } 0.$$

Corollario 2.3.9. Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ un sistema dinamico, con \mathcal{M} spazio topologico tale che lo spazio delle funzioni continue sia denso in $L^1(\mathcal{M}, \mu)$. Allora se per ogni funzione continua $f : \mathcal{M} \rightarrow \mathbb{R}$, f^+ è quasi ovunque costante su \mathcal{M} secondo la misura μ , allora $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu, T)$ è ergodico.

Dimostrazione. Consideriamo un sottoinsieme misurabile T -invariante A .

$$T^{-1}A = A \quad \text{mod } 0.$$

Sia g la funzione caratteristica di A

$$g(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

ed

$$f_n \rightarrow g \text{ in } L^1(\mathcal{M}, \mu)$$

²Se la misura è normalizzata (e dunque siamo in uno spazio di probabilità) e se la media temporale nel futuro è uguale a una costante, la stessa costante è anche uguale alla media in fase della funzione.

una sequenza di approssimazioni della funzione caratteristica continue. Possiamo usare il fatto che la media temporale è continua rispetto alla norma di L^1 , per dimostrare che la media temporale di g deve essere costante su \mathcal{M} . In effetti se indichiamo con $\|\cdot\|_1$ la norma di L^1 , allora, per uno degli asserti del Teorema di Birkhoff 2.3.4:

$$\|f_n^+ - g^+\|_1 = \|f_n - g\|_1.$$

Ma $g^+ = g$, poiché A è un insieme invariante, quindi

$$\|f_n^+ - g\|_1 = \|f_n - g\|_1, \quad (2.3.3)$$

che può essere reso più piccolo di qualsiasi $\epsilon > 0$, per n sufficientemente grande. Ma le medie temporali $f_n^+(x)$ di $f_n(x)$ sono costanti per quasi ogni $x \in \mathcal{M}$ e inoltre per n sufficientemente grande, $0 < f_n^+ < 1$, poiché

$$f_n^+ = \int_{\mathcal{M}} f_n d\mu \xrightarrow{n \rightarrow \infty} \int_{\mathcal{M}} g d\mu.$$

Si osserva allora che il membro sinistro della (2.3.3) vale $\mu(A)(1 - f_n^+) + (1 - \mu(A))f_n^+ \geq \min\{\mu(A), 1 - \mu(A)\}$. Dunque

$$\min\{\mu(A), 1 - \mu(A)\} \leq \|f_n^+ - g\|_1 = \|f_n - g\|_1 \xrightarrow{n \rightarrow \infty} 0.$$

Abbiamo quindi dimostrato che $\min\{\mu(A), 1 - \mu(A)\} < \epsilon, \forall \epsilon > 0$. Ne consegue che o $\mu(A) = 0$ o $\mu(A) = 1$. Quindi ogni insieme invariante ha misura 0 o 1. \square

Per approfondire la teoria ergodica e argomenti collegati si guardi [6] e [7].

2.4 Esempi di sistemi dinamici ergodici

In questa sezione presentiamo due semplici sistemi dinamici ergodici. Tale (apparente) semplicità non impedisce a questi sistemi di poter rappresentare ottimi modelli generali e anzi mette in luce parecchie caratteristiche comuni a tutti i sistemi ergodici.

Per poter procedere abbiamo bisogno di alcune nozioni ausiliarie, in particolare possiamo iniziare dando la seguente:

Definizione 2.4.1. (di relazione di equivalenza) Dati due insiemi A e B , il loro prodotto cartesiano, $A \times B$, è l'insieme delle coppie ordinate $\{(a, b) | a \in A, b \in B\}$. Si definisce relazione binaria R su un insieme A un sottoinsieme di $A \times A$. Due elementi a^* e b^* sono messi in relazione da R se $(a^*, b^*) \in R$. In questo caso si dice che a^* è in relazione con b^* e si scrive $a^* R b^*$. Una relazione di equivalenza, che si indica generalmente con \sim , è una particolare relazione tra gli elementi di un insieme A , che gode delle seguenti proprietà:

- (a) $a^* \sim a^*, \forall a^* \in A$,
- (b) se $a^* \sim b^*$, allora $b^* \sim a^*, \forall a^*, b^* \in A$,
- (c) se $a^* \sim b^*$ e $b^* \sim c^*$, allora $a^* \sim c^*, \forall a^*, b^*, c^* \in A$.

Se un insieme A è munito di una relazione di equivalenza è possibile definire particolari sottoinsiemi di A chiamati *classi di equivalenza*.

Definizione 2.4.2. Un sottoinsieme di A che contiene tutti e soli quegli elementi di A che sono equivalenti a un certo elemento $a^* \in A$ è detto classe di equivalenza. Indichiamo tale classe di equivalenza con $[a^*]_{\sim}$. Si presti attenzione al fatto che nella scrittura precedente a^* è un qualsiasi elemento della classe di equivalenza $[a^*]_{\sim}$.

E' evidente che l'insieme dei membri di tutte le classi di equivalenza su un certo insieme deve essere l'insieme stesso. Infatti nel peggiore dei casi ogni elemento appartiene comunque alla classe di equivalenza di cui è l'unico esponente. Da un punto di vista pratico, se è stata definita una relazione di equivalenza su un insieme, vuol dire che non ci interessa distinguere tutti gli elementi dell'insieme ma solo quelli appartenenti a classi di equivalenza distinte. Questo porta a considerare un insieme i cui elementi sono le classi di equivalenza.

Definizione 2.4.3. (di insieme quoziente) Sia A un insieme non vuoto ed \sim una relazione di equivalenza su A . Si dice insieme quoziente di A rispetto alla relazione \sim , l'insieme denotato con A/\sim i cui elementi sono le classi di equivalenza di A indotte dalla relazione \sim :

$$A/\sim = \{[a]_{\sim} | a \in A\}.$$

2.4.1 Rotazioni sul cerchio

Sia $\mathbb{S}^1 \subset \mathbb{R}^2$, $\mathbb{S}^1 = \{x \in \mathbb{R}^2 | \|x\| = \frac{1}{2\pi}\}$, dove $\|\cdot\|$ è la norma indotta dalla metrica euclidea. \mathbb{S}^1 è il cerchio unitario. Un altro modo per pensare ad \mathbb{S}^1 è quello di considerare la retta reale \mathbb{R} e di definire la relazione di equivalenza \sim , ponendo $x \sim y$ se $x - y \in \mathbb{Z}$. Abbiamo allora che $\mathbb{S}^1 = \mathbb{R}/\sim$. Al posto di $\mathbb{S}^1 = \mathbb{R}/\sim$ useremo la notazione $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, come da tradizione.

Per avere un sistema dinamico dobbiamo in primo luogo definire una misura su \mathbb{S}^1 . Costruiamo una corrispondenza tra sottoinsiemi di \mathbb{S}^1 e sottoinsiemi di $[0, 1)$: $\forall A \subset \mathbb{S}^1$, definiamo $A^* = \{a \in [0, 1) | [a]_{\sim} \in A\}$, $A^* \subset [0, 1)$. Sia $B \subset \mathbb{S}^1$, definiamo $\mu(B) = \text{Leb}_1(B^*)$. Diciamo che B è μ -misurabile se B^* è Leb_1 -misurabile. Indichiamo con $\mathfrak{B}(\mathbb{S}^1)$ la σ -algebra dei sottoinsiemi di \mathbb{S}^1 μ -misurabili. Si ha per definizione che $\mathfrak{B}(\mathbb{S}^1) = \{A \in \mathbb{S}^1 | A^* \text{ è } \text{Leb}_1 \text{ misurabile}\}$. Adesso che abbiamo una σ -algebra e una misura su \mathbb{S}^1 resta solo da definire una mappa per avere un sistema dinamico.

Sia $T : \mathbb{S}^1 \rightarrow \mathbb{S}^1$, così definita:

$$T([x]_{\sim}) = [x + a]_{\sim}. \quad (2.4.1)$$

La (2.4.1) può anche essere scritta come $T(x) = x + a \pmod{1}$, intendendo che sia il dominio che il codominio sono le classi di equivalenza $\pmod{1}$. D'ora in poi scriveremo x invece che $[x]_{\sim}$, intendendo l'intera classe di equivalenza cui x appartiene. Si vede bene che μ è T -invariante, in quanto T rappresenta le rotazioni sul cerchio \mathbb{S}^1 .

Teorema 2.4.4. *La rotazione $T(x) = x + a \pmod{1}$ del cerchio \mathbb{S}^1 è ergodica se e solo se a è irrazionale.*

Dimostrazione. Distinguiamo due casi.

(1) a razionale. Poniamo $a = \frac{p}{q}$, $p, q \in \mathbb{Z}$, $q > 0$, p e q primi tra loro. $f(x) = e^{2\pi i q x}$ è una funzione non costante, misurabile e invariante. Per l'Osservazione 2.3.8 paragrafo 2.3, abbiamo che il sistema non è ergodico.

(2) a irrazionale. Sia A un insieme invariante di misura positiva. Dimostriamo che $\mu(A) = 1$. Dal momento che $\mu(A) > 0$, per il teorema della densità di Lebesgue, deve necessariamente esistere un punto $x_0 \in A$ ed un intorno di x_0 , $I =]x_0 - \delta, x_0 + \delta[$, tale che

$$\lim_{\delta \rightarrow 0} \frac{\mu(A \cap I)}{\mu(I)} = 1.$$

In altre parole, esiste un punto di A per cui posso scegliere un intorno sufficientemente piccolo tale che la misura dei punti appartenenti a quell'intorno e non appartenenti ad A è piccola a piacere. Abbiamo allora per la definizione di limite che $\forall \epsilon, 0 < \epsilon < 1$ è possibile scegliere $\delta(\epsilon)$ abbastanza piccolo che valga:

$$\mu(A \cap I) \geq (1 - \epsilon)\mu(I).$$

$\forall \delta < \delta(\epsilon)$. Usando il fatto che $TA = A$ (poiché A è invariante e T invertibile) e l'invarianza di μ , otteniamo:

$$\mu(A \cap T^n I) \geq (1 - \epsilon)\mu(T^n I).$$

Se n_1, n_2, \dots, n_k , $k \in \mathbb{N}$, sono interi per i quali $T^{n_1} I, T^{n_2} I, \dots, T^{n_k} I$ sono disgiunti, otteniamo:

$$\mu(A) \geq \sum_{i=1}^k \mu(A \cap T^{n_i} I) \geq (1 - \epsilon)\mu\left(\bigcup_{i=1}^k T^{n_i} I\right).$$

D'altra parte, per il Teorema di Jacobi l'orbita di ogni punto di \mathbb{S}^1 è densa in \mathbb{S}^1 . In particolare l'orbita di uno qualsiasi dei due estremi di I è densa in \mathbb{S}^1 . Dal momento che $\mu(I) = 2\delta$, devono esistere degli interi n_1, n_2, \dots, n_k , tali che gli insiemi $T^{n_1} I, T^{n_2} I, \dots, T^{n_k} I$ sono disgiunti e coprono tutto \mathbb{S}^1 a meno di un insieme

di misura $2 \times 2\delta$ ³. Di conseguenza si ha che

$$\mu\left(\bigcup_{i=1}^k T^{n_i} I\right) \geq 1 - 4\delta,$$

e

$$\mu(A) \geq (1 - 4\delta)(1 - \epsilon).$$

Siccome però ϵ può essere scelto piccolo a piacere e δ è arbitrario purché $\delta < \delta(\epsilon)$, allora ne consegue che

$$\mu(A) = 1.$$

□

2.4.2 Mappa del gatto di Arnold

Siano $x, y \in \mathbb{R}^2$. Definiamo una relazione di equivalenza \sim tra i punti appartenenti ad \mathbb{R}^2 , dicendo che $x \sim y$ se $x - y \in \mathbb{Z}^2$. Chiamiamo toro unitario, \mathbb{T}^2 , l'insieme delle classi di equivalenza $\{[x]_\sim | x \in \mathbb{R}^2\}$, cosicché $\mathbb{T}^2 = \mathbb{R}^2 / \sim$ o, come si indica tradizionalmente, $\mathbb{T}^2 = \mathbb{R}^2 / \mathbb{Z}^2$.

Analogamente al caso di \mathbb{S}^1 , $\forall A \subset \mathbb{T}^2$, definiamo $A^* = \{a \in [0, 1) \times [0, 1) | [a]_\sim \in A\}$ e, ove possibile, (cioè quando B^* è Leb_2 -misurabile), $\forall B \subset \mathbb{T}^2$, definiamo $\mu(B) = Leb_2(B^*)$. Anche in questo caso dunque $\mathfrak{B}(\mathbb{T}^2) = \{A \in \mathbb{T}^2 | A^* \text{ è } Leb_2 \text{ misurabile}\}$.

Sia $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ così definita:

$$T\left[\begin{pmatrix} x_0 \\ x_1 \end{pmatrix}\right]_\sim = \left[\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}\right]_\sim,$$

dove $\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \in \mathbb{R}^2$.

³Dal momento che gli estremi di I sono punti la cui orbita è densa in \mathbb{S}^1 , è possibile scegliere gli interi n_i in modo che tra $T^{n_i} I$ e $T^{n_{i+1}} I$ ci sia una misura piccola a piacere. Per comodità poniamo tale misura uguale a $4 \times \delta^2$. Il numero di *zone scoperte* di \mathbb{S}^1 è all'incirca uguale alla lunghezza di \mathbb{S}^1 diviso la lunghezza di I : $N_{intervalliscoperti} \sim \frac{1}{2\delta}$. Si ricordi infatti che T conserva la misura. Non è possibile però controllare il risultato dell'ultima iterazione: non è possibile cioè sapere se dopo aver ricoperto tutto il cerchio a meno di spazi vuoti anche molto piccoli, sia possibile piazzare l'ultima iterazione $T^{n_k} I$ senza che si sovrapponga con la prima. Nel peggiore dei casi dobbiamo aggiungere alla *lunghezza scoperta* la lunghezza di I , cioè $2 \times \delta$. Abbiamo dunque che la massima lunghezza scoperta può comunque sempre essere resa minore di $N_{intervalliscoperti} \times 4 \times \delta^2 + 2\delta = \frac{1}{2\delta} \times 4 \times \delta^2 + 2 \times \delta = 2 \times 2\delta$.

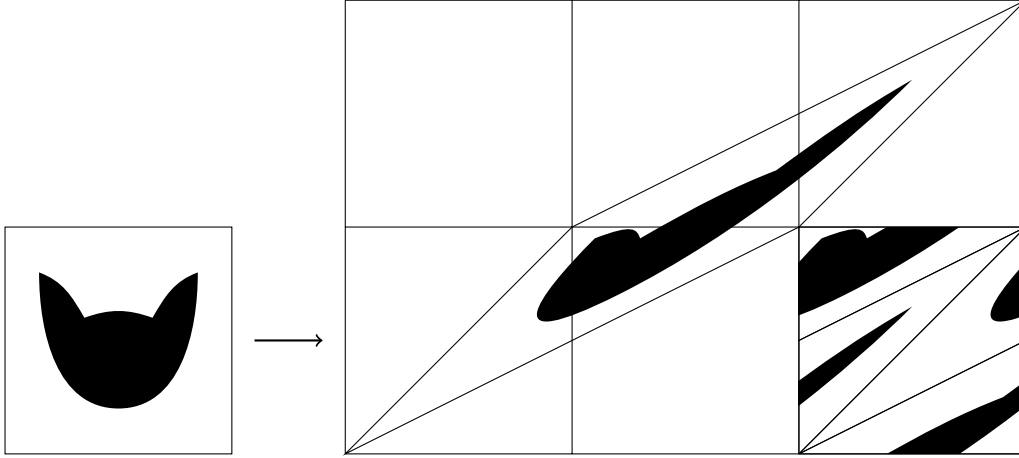


Fig.1: una rappresentazione della mappa del *Gatto di Arnold*. Il fatto che il gatto 'stirato' possa essere 'ripiegato' alla perfezione nel quadrato unitario di \mathbb{R}^2 dipende dal fatto che tale mappa è ben definita sul toro.

Vediamo che l'azione di T è ben definita su \mathbb{T}^2 , in quanto non dipende dalla scelta del rappresentante di una classe di equivalenza. Infatti se $k, p \in \mathbb{Z}$, allora

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_0 + k \\ x_1 + p \end{pmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} k \\ p \end{pmatrix} \sim \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}. \quad (2.4.2)$$

Dunque se $\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \sim \begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$, allora anche $T\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \sim T\begin{pmatrix} y_0 \\ y_1 \end{pmatrix}$ e T è una mappa ben definita su \mathbb{T}^2 . Consideriamo adesso una applicazione $\hat{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ così definita:

$$\hat{T} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix},$$

dove $\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \in \mathbb{R}^2$. Dal momento che la matrice $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ ha determinante unitario, abbiamo che \hat{T} conserva la misura di Lebesgue in \mathbb{R}^2 . Sia allora A un qualsiasi sottoinsieme di \mathbb{T}^2 , ricordando come abbiamo definito la misura μ su \mathbb{T}^2 , abbiamo che

$$\mu(T(A)) = \text{Leb}_2((T(A))^*) = \text{Leb}_2(\hat{T}(A^*)) = \text{Leb}_2(A^*) = \mu(A).$$

Siccome T è invertibile, abbiamo che T conserva la misura μ .

Passiamo adesso a considerare il differenziale della mappa. T è una applicazione lineare e dunque il suo differenziale è in ogni punto di \mathbb{T}^2 costante ed è definito dalla stessa matrice che definisce T :

$$dT|_{\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}} = \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix},$$

con $\begin{pmatrix} x \\ y \end{pmatrix}$ appartenente allo spazio tangente del punto $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$. Possiamo decomporre tale spazio tangente in due sottospazi dati dallo *span* dei due autovettori di $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$:

$$\bar{u}_1 = \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}, \quad \bar{u}_2 = \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}.$$

Dunque gli autovettori della mappa costituiscono in ogni punto una base dello spazio tangente. Le proprietà interessanti sono le seguenti:

- (i) in ogni punto $x \in \mathbb{T}^2$, lo spazio tangente si decompone nella somma di due sottospazi ortogonali: il sottospazio contraente E^s ed il sottospazio dilatante E^u ;
- (ii) i sottospazi E^s ed E^u sono invarianti sotto l'azione del differenziale della mappa dT ; ⁴
- (iii) vale la seguente proprietà:

$$\begin{aligned} dT\bar{\zeta} &= \lambda_1\bar{\zeta}, \forall \bar{\zeta} \in E^s, \\ dT\bar{\zeta} &= \lambda_2\bar{\zeta}, \forall \bar{\zeta} \in E^u, \end{aligned}$$

dove $\lambda_1 < 1$ è l'autovalore relativo ad \bar{u}_1 . Abbiamo che E^s è generato dall'autovettore \bar{u}_1 , mentre E^u è generato dall'autovettore \bar{u}_2 . Gli autovalori relativi a \bar{u}_1 ed \bar{u}_2 sono

$$\lambda_1 = \frac{3 - \sqrt{5}}{2}; \quad \lambda_2 = \frac{3 + \sqrt{5}}{2}.$$

Una qualsiasi applicazione lineare che abbia determinate unitario e due autovalori, uno maggiore e l'altro minore di uno è detta iperbolica. Nel caso della mappa T (del gatto di Arnold) e delle altre applicazioni lineari *continue* esiste un metodo piuttosto elaborato per dimostrarne l'ergodicità: il metodo di Hopf. Presentiamo nella prossima sezione il metodo di Hopf per la mappa del gatto di Arnold, considerando immediata la sua traduzione per qualsiasi mappa lineare iperbolica *continua*. Per approfondire altre proprietà della *mappa del Gatto di Arnold* si faccia riferimento al Capitolo 7 di [8].

Il metodo di Hopf per la mappa del gatto di Arnold

Sia $x \in \mathbb{R}^2$, $x = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix}$, indichiamo con $W^u(x)$ e $W^s(x)$ la linea in \mathbb{T}^2 passante per x e avente la direzione rispettivamente del vettore instabile e del vettore stabile.

$$W^u(x) = \left\{ \left[x_0 + t, x_1 + \frac{\sqrt{5}-1}{2}t \right]_{\sim} \in \mathbb{T}^2 \mid t \in \mathbb{R} \right\}$$

⁴Questa proprietà non è altro che la definizione di autospazio. Si ricordi infatti che i sottospazi E^s ed E^u sono gli autospazi generati dai vettori \bar{u}_1 e \bar{u}_2 .

$$W^s(x) = \left\{ [x_0 + t, x_1 - \frac{\sqrt{5} + 1}{2}t]_{\sim} \in \mathbb{T}^2 \mid t \in \mathbb{R} \right\}.$$

Chiamiamo $W^s(x)$ e $W^u(x)$ *varietà stabile* e *instabile*, rispettivamente. Chiamiamo invece *foliazione stabile* o *instabile* l'insieme di tutte le varietà (stabili o instabili) passanti per ogni $x \in \mathbb{T}^2$.

$$W^s = \{W^s(x)\}_{x \in \mathbb{T}^2},$$

$$W^u = \{W^u(x)\}_{x \in \mathbb{T}^2}.$$

Siccome può risultare abbastanza faticoso lavorare sempre con le classi di equivalenza, possiamo semplificarci la vita se al posto di ogni classe scriviamo il suo rappresentante appartenente al quadrato $[0, 1[\times [0, 1[$. Usando questa convezione abbiamo:

$$W^u(x) = \left\{ (x_0 + t, x_1 + \frac{\sqrt{5} - 1}{2}t) - [x_0 + t, x_1 + \frac{\sqrt{5} - 1}{2}t] \in \mathbb{T}^2 \mid t \in \mathbb{R} \right\}$$

$$W^s(x) = \left\{ (x_0 + t, x_1 - \frac{\sqrt{5} + 1}{2}t) - [x_0 + t, x_1 - \frac{\sqrt{5} + 1}{2}t] \in \mathbb{T}^2 \mid t \in \mathbb{R} \right\},$$

dove $[\cdot, \cdot]$ indica la coppia $([\cdot], [\cdot])$ e $[\cdot]$ denota il più grande tra gli interi minori del numero tra parentesi. Le varietà di x hanno la seguente proprietà. Se $y \in W^s(x)$ allora la distanza⁵ delle immagini ripetute di x e y sotto la mappa T tende a zero:

$$d(T^n x, T^n y) = \frac{1}{\lambda_2^n} d(x, y) \rightarrow 0 \quad \text{per } n \rightarrow \infty. \quad (2.4.3)$$

Viceversa se $y \in W^u(x)$:

$$d(T^{-n} x, T^{-n} y) = \frac{1}{\lambda_2^n} d(x, y) \rightarrow 0 \quad \text{per } n \rightarrow \infty. \quad (2.4.4)$$

Consideriamo allora una funzione continua⁶ $f : \mathbb{T}^2 \rightarrow \mathbb{R}$. Se $y, z \in W^u(x)$ per un certo $x \in \mathbb{T}^2$ allora per la continuità di f e per la (2.4.4) si ha:

$$|f(T^{-n}(y)) - f(T^{-n}(z))| \rightarrow 0 \quad \text{per } n \rightarrow \infty. \quad (2.4.5)$$

Dalle Definizione 2.3.1 si vede come la (2.4.5) implichi che, se ben definita, $f^-(y) = f^-(z)$. Analogamente abbiamo che se $y, z \in W^s(x)$ per un certo $x \in \mathbb{T}^2$ allora $f^+(y) = f^+(z)$, sempre se queste quantità esistono.

⁵Possiamo considerare la distanza tra due punti del toro come la distanza euclidea tra i due rappresentanti di quei punti appartenenti al quadrato $[0, 1[\times [0, 1[$.

⁶E dunque anche uniformemente continua siccome \mathbb{T}^2 è un compatto.

Ricordiamo che in questo caso $f^{+(-)} : \mathbb{T}^2 \rightarrow \mathbb{R}$. Possiamo però considerare il sollevamento, $\hat{f}^{+(-)}$, di $f^{+(-)}$ ad \mathbb{R}^2 , ovvero $\hat{f}^{+(-)} : \mathbb{R}^2 \rightarrow \mathbb{R}$, così definita:

$$\hat{f} = f \circ \pi, \quad (2.4.6)$$

dove $\pi : \mathbb{R}^2 \rightarrow \mathbb{T}^2$ è la funzione che associa ad ogni $x \in \mathbb{R}^2$ la sua classe di equivalenza:

$$\pi(x) = [x]_{\sim}.$$

Procedendo su questa strada abbiamo che

$$W^u(x) = \pi(\widehat{W^u(x)}) \quad W^s(x) = \pi(\widehat{W^s(x)}),$$

dove

$$\widehat{W^u(x)} = \left\{ (x_0 + t, x_1 + \frac{\sqrt{5}-1}{2}t) \in \mathbb{R}^2 \mid t \in \mathbb{R} \right\}$$

$$\widehat{W^s(x)} = \left\{ (x_0 + t, x_1 - \frac{\sqrt{5}+1}{2}t) \in \mathbb{R}^2 \mid t \in \mathbb{R} \right\}.$$

In pratica $W^{u(s)}(x)$ è l'avvolgimento sul toro delle rette di \mathbb{R}^2 $\widehat{W^{u(s)}(x)}$. Per costruzione se $y \in \widehat{W^u(x)} \subset \mathbb{R}^2$:

$$\hat{f}^-(y) = (f^- \circ \pi)(y) = f^-(\pi(y)),$$

con $\pi(y) \in W^u(x) \subset \mathbb{T}^2$. Abbiamo dunque che se $f^{+(-)}$ è costante su $W^{s(u)}(x)$, allora il suo sollevamento $\hat{f}^{+(-)}$ è costante su $\widehat{W^{s(u)}(x)}$. Possiamo allora affermare che grazie alle proprietà (2.4.3) e (2.4.4) della mappa T che $\forall x \in \mathbb{R}^2$, \hat{f}^- è costante su tutto $\widehat{W^u(x)}$ mentre \hat{f}^+ è costante su tutto $\widehat{W^s(x)}$. Ma $\widehat{W^u(x)}$ e $\widehat{W^s(x)}$ individuano una coppia di coordinate di \mathbb{R}^2 e dunque \hat{f}^+ è funzione di una sola coordinata mentre \hat{f}^- è funzione solo dell'altra. Inoltre per il Corollario 2.3.6 abbiamo che $f^+ = f^-$ quasi dovunque su \mathbb{T}^2 e dunque, per la definizione di sollevamento di una funzione (2.4.6), si ha che anche $\hat{f}^+ = \hat{f}^-$ quasi ovunque in \mathbb{R}^2 secondo la misura Leb_2 . L'unica possibilità per cui \hat{f}^+ possa essere funzione di una sola coordinata, \hat{f}^- funzione dell'altra coordinata ed $\hat{f}^+ = \hat{f}^-$ quasi dovunque è quello in cui $\hat{f}^+ = \hat{f}^- = \text{costante}$ quasi dovunque. Sempre per la Definizione 2.4.6, questo implica che $f^+ = f^- = \text{costante}$ quasi dovunque in \mathbb{T}^2 secondo la misura μ e quindi, per il Corollario 2.3.9, abbiamo che (\mathbb{T}, μ, T) è ergodico.

Nei prossimi capitoli considereremo mappe non continue. Vedremo che in quel caso le varietà $W^u(x)$ e $W^s(x)$ (con x punto del dominio \mathcal{M} che considereremo) per cui valgono le (2.4.3) e (2.4.4) hanno tutte lunghezza finita⁷ e dunque non potranno mai essere viste come coordinate di \mathbb{R}^2 e l'intero metodo di Hopf sembra fallire. Chiudiamo però il Capitolo con la seguente:

⁷Vedremo infatti che *nel migliore dei casi* saranno segmenti con estremi sul bordo del dominio compatto \mathcal{M} .

Osservazione 2.4.5. (della catena di Hopf) Sia $W^u(x_1)$ una varietà instabile, caratterizzata dalla proprietà che f^- è costante su di essa. Anche se $W^u(x_1)$ è piuttosto piccola è possibile (vedremo anzi che in un certo senso è anche estremamente probabile), che su $W^u(x_1)$ ci sia un punto x_2 per il quale $f^+(x_2)$ ed $f^-(x_2)$ esistono, $f^+(x_2) = f^-(x_2)$ ⁸ e per il quale passi una varietà stabile $W^s(x_2)$. Nel prossimo Capitolo definiremo come *f1-tipico* un punto con proprietà analoghe a x_2 . Siccome f^+ è costante su $W^s(x_2)$ ed $f^+(x_2) = f^-(x_2)$, ed inoltre vedremo che sarà possibile scegliere anche $W^u(x_1)$ e $W^u(x_2)$ in modo che siano ‘tipiche’ (cioè per quasi ogni $x \in W^u(x_1)$ e $x \in W^u(x_2)$, nel senso della misura di Lebesgue unidimensionale, $f^+(x) = f^-(x)$), allora $f^+|_{W^s(x_2)} = f^-|_{W^u(x_1)} = \text{costante}$. Potremmo continuare cercando varietà instabili che intersechino $W^s(x_2)$ in punti per i quali f^+ ed f^- esistono ed $f^+ = f^-$. Si viene così a costruire una sorta di catena su cui $f^+ = f^- = \text{costante}$ su domini alternati: la *catena di Hopf* [9]. Siamo ancora ben lontani da poter dire che $f^+ = f^- = \text{costante}$ sul quasi tutti i punti del dominio della mappa, ma l’idea si rivelerà essere quella vincente.

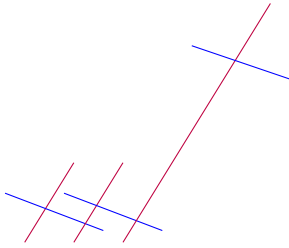


Fig.2: in figura sono rappresentate in rosso le varietà instabili e in blu le varietà stabili. Nel caso di una mappa discontinua vedremo che tutte le varietà (stabili e instabili) hanno lunghezza finita, anche molto piccola. Inoltre non per tutti i punti in \mathcal{M} passerà una varietà e le varietà hanno lunghezze differenti le une dalle altre. Ad ogni modo se le intersezioni avvengono in punti *f-tipici*, $f^+ = \text{costante}$ sulle linee stabili ed $f^- = \text{costante}$ sulle linee instabili, con le due costanti che coincidono.

⁸Infatti, per il teorema di Birkhoff, quasi tutti i punti godono di queste due proprietà.

Capitolo 3

L'esempio di Liverani e Wojtkowski

In questo Capitolo presenteremo l'argomento di Hopf per mappe discontinue. In particolare studieremo l'effetto di una mappa iperbolica T simile a quella del *Gatto di Arnold* che si differenzia da quest'ultima per il fatto che la matrice A che contribuisce alla definizione di T è in generale composta da numeri non interi, nella base canonica di \mathbb{R}^2 . L'appartenenza a \mathbb{Q} piuttosto che a \mathbb{Z} degli elementi della matrice A impedisce di definire il toro come dominio di T , dato che in generale la (2.4.2) non sarà più valida. Se tuttavia pensiamo alla mappa del Gatto di Arnold come una mappa da $[0, 1) \times [0, 1)$ a $[0, 1) \times [0, 1)$, vediamo che essa agisce come l'effetto di una matrice iperbolica A seguita da uno *spezzettamento* di $A([0, 1) \times [0, 1))$ e traslazioni rigide dei pezzi (vedi figura 1). Quando si tenta di generalizzare questo tipo di mappa per matrici iperboliche a coefficienti non interi, la differenza sostanziale è dovuta all'insorgere di linee di discontinuità della mappa.

Nella sezione 3.1 introdurremo un particolare dominio \mathcal{M} su cui è possibile definire una mappa T data dall'effetto di una matrice iperbolica seguita da traslazioni. Nelle sezioni 3.2 e 3.3 affronteremo il problema di definire varietà stabili e instabili su \mathcal{M} (che, come già menzionato, hanno lunghezza finita per problemi di continuità delle orbite) e vedremo per *quanti* punti di \mathcal{M} passano tali varietà. Nella sezione 3.4 presenteremo il *Metodo di Sinai* per dimostrare l'appartenenza di \mathcal{M} ad una sola componente ergodica ¹.

3.1 Una generalizzazione della mappa del gatto di Arnold

Sia \mathcal{M} un sottoinsieme compatto di \mathbb{R}^2 definito da

$$\mathcal{M} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 0 \leq x_1 + ax_2 \leq 1, 0 \leq x_2 \leq 1\} \quad (3.1.1)$$

¹Rimandiamo alla nota 1 del Capitolo 2 per la spiegazione di cosa si intenda per 'componente ergodica' in questo lavoro.

$a \in \mathbb{R}$, su cui è definita una mappa affine $T : \mathcal{M} \rightarrow \mathcal{M}$. Per definire l'azione di T su \mathcal{M} definiremo più avanti la partizione di \mathcal{M} :

$$\mathcal{P}^+ = \{\mathcal{M}_i^+ | i = 1, 2, 3, 4\}, \quad (3.1.2)$$

$$\mathcal{M} = \bigcup_{i=1}^4 \mathcal{M}_i^+, \quad (3.1.3)$$

tali che su ogni sotto-dominio \mathcal{M}_i^+ la mappa T agisca nella seguente maniera:

$$T|_{\mathcal{M}_i^+}(x) = Ax + b_i, \quad (3.1.4)$$

dove $b_i \in \mathbb{R}^2$ ed $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è la seguente matrice

$$A = \begin{bmatrix} 1 & a \\ a & 1 + a^2 \end{bmatrix}. \quad (3.1.5)$$

Vediamo subito che A ha $\det A = 1$ ed è diagonalizzabile con autovalori $\lambda_1(a)$, $\lambda_2(a)$ tali che $\lambda_1 < 1$ e $\lambda_2 = \frac{1}{\lambda_1} > 1$. Quindi A è una matrice iperbolica. Definiamo, $\forall i = 1, 2, 3, 4$,

$$\mathcal{N}_i = A(\mathcal{M}_i^+). \quad (3.1.6)$$

Quindi $A(\mathcal{P}^+) = \{\mathcal{N}_i | i = 1, 2, 3, 4\}$ è una partizione di $A(\mathcal{M})$. Supponiamo che esista una seconda partizione di \mathcal{M} , \mathcal{P}^- :

$$\mathcal{P}^- = \{\mathcal{M}_i^- | i = 1, 2, 3, 4\}, \quad (3.1.7)$$

tale che $\forall i = 2, \dots, N$:

$$\mathcal{M}_i^- = \mathcal{N}_i + b_i, \quad (3.1.8)$$

dove nel nostro caso scegliamo di definire:

$$\begin{aligned} \mathcal{N}_1 &= \{(x_1, x_2) \in \mathbb{R}^2 | x_2 \leq 1, x_1 \leq 1, x_1 + ax_2 \geq 1, x_2 + ax_1 \geq 0\} \\ \mathcal{N}_2 &= \{(x_1, x_2) \in \mathbb{R}^2 | x_2 \geq 1, ax_1 - x_2 \geq -1, ax_2 + x_1 \leq 2\} \\ \mathcal{N}_3 &= \{(x_1, x_2) \in \mathbb{R}^2 | x_1 \leq 1, ax_2 + x_1 \geq 2, ax_1 - x_2 \geq -1\} \\ \mathcal{N}_4 &= \{(x_1, x_2) \in \mathbb{R}^2 | x_2 \leq 1, x_1 \geq 1, x_1 + ax_2 \leq 1, x_2 + ax_1 \geq 0\}. \end{aligned} \quad (3.1.9)$$

Questa scelta definisce implicitamente anche \mathcal{P}^+ , poiché $\mathcal{M}_i^+ = A^{-1}(\mathcal{N}_i)$. Dalla figura 3 si vede come, definendo

$$\begin{aligned} b_1 &= (-1, 0) \\ b_2 &= (0, -1) \\ b_3 &= (-1, -1) \\ b_4 &= (0, 0), \end{aligned} \quad (3.1.10)$$

gli insiemi dati da (3.1.8) formano effettivamente una partizione (\mathcal{P}^-). Possiamo allora descrivere così l'azione della mappa T su \mathcal{M} : ogni \mathcal{M}_i^+ viene mandato dalla matrice nel corrispondente \mathcal{N}_i (nel fare ciò \mathcal{M}_i^+ cambia forma in quanto l'effetto di A è quello di 'stirare' gli insiemi su cui viene applicata lungo la direzione instabile), il quale viene traslato di un vettore dipendente dall'indice i , ottenendo \mathcal{M}_i^- (vedi figura 3). Abbiamo infatti:

$$T(\mathcal{M}_i^+) = A(A^{-1}(\mathcal{N}_i)) + b_i = \mathcal{N}_i + b_i = \mathcal{M}_i^-. \quad (3.1.11)$$

E, rispetto alle partizioni di \mathcal{M} , \mathcal{P}^+ e \mathcal{P}^- , possiamo scrivere:

$$T(\mathcal{P}^+) = \mathcal{P}^-, \quad (3.1.12)$$

intendendo con questa notazione che $T(\mathcal{P}^+) = \{T\mathcal{M}_i^+ | i = 1, 2, 3, 4\}$. E' importante notare che per definire T basta specificare una tra le partizioni \mathcal{P}^+ , \mathcal{P}^- e $A(\mathcal{P}^+)$ (le altre due venendo fissate dalla 3.1.8) e la lista di vettori $\{b_i | i = 1, 2, 3, 4\}$. Tuttavia, fissata ad esempio \mathcal{P}^+ non è detto che esista una lista di vettori per cui \mathcal{P}^- è effettivamente una partizione di \mathcal{M} . Risulta in effetti comodo (come abbiamo precedentemente fatto) definire in primo luogo la partizione $A(\mathcal{P})$ tale per cui esistano dei vettori b_i tramite i quali la (3.1.8) definisca una vera partizione \mathcal{P}^- di \mathcal{M} , e, solo in un secondo momento, definire la partizione \mathcal{P}^+ , grazie alla (3.1.6). Dal momento che $\det A \neq 0$, esiste l'inversa di T , $T^{-1} : \mathcal{M} \rightarrow \mathcal{M}$, così definita:

$$T^{-1} |_{\mathcal{M}_i^-} (x) = A^{-1}x - A^{-1}b_i.$$

Dalla (3.1.11) si vede che

$$T^{-1}(\mathcal{M}_i^-) = \mathcal{M}_i^+.$$

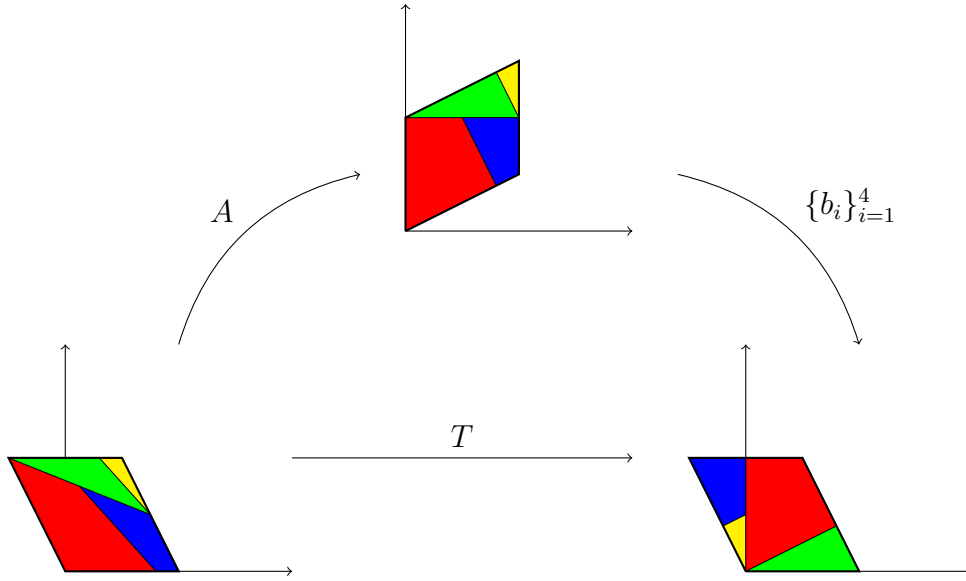


Fig 3: Per $a = \frac{1}{2}$, la figura mostra l'effetto della mappa T sul dominio $\mathcal{M} = \{(x_1, x_2) \in \mathbb{R}^2 | 0 \leq x_1 + \frac{1}{2}x_2 \leq 1, 0 \leq x_2 \leq 1\}$, data dalla matrice $A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{4} \end{bmatrix}$ seguita dalle traslazioni $\{b_i\}$ della (3.1.10). L'insieme dei poligoni in basso a sinistra rappresenta la partizione \mathcal{P}^+ , quelli in basso a destra la partizione \mathcal{P}^- , in alto i poligoni \mathcal{N}_i . Ognuno dei poligoni \mathcal{M}_i^+ di \mathcal{P}^+ viene mappato in un poligono corrispondente (dello stesso colore) identificando così la partizione $A(\mathcal{P}^+)$ di $A(\mathcal{M})$. Le traslazioni rigide permettono di riottenere l'insieme \mathcal{M} di partenza cosicché l'effetto finale è quello di una mappa $T : \mathcal{M} \rightarrow \mathcal{M}$. Si noti che la partizione \mathcal{P}^- dell'insieme $\mathcal{M} = T(\mathcal{M})$, visto cioè come immagine della mappa T , è diversa da quella di partenza \mathcal{P}^+ .

Sia $\mathfrak{B}(\mathcal{M})$ la σ -algebra degli insiemi misurabili secondo Lebesgue contenuti in \mathcal{M} . Indichiamo allora con $Leb_2(\cdot)$ la misura di Lebesgue (bidimensionale) definita su $\mathfrak{B}(\mathcal{M})$. Osservando che $Leb_2(\mathcal{M}) = 1$ abbiamo che $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), Leb_2(\cdot))$ è uno spazio di probabilità. Inoltre, dal momento che A è una applicazione con determinante unitario e T è data dalla composizione di A con traslazioni rigide, abbiamo che T è misurabile e conserva la misura $Leb_2(\cdot)$. Ha senso allora chiedersi se T è una trasformazione ergodica sullo spazio $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), Leb_2(\cdot))$. Supponiamo che il parametro a , che fissa la mappa T , non sia intero. Introduciamo ora due importanti notazioni che useremo spesso da qui in poi. Indichiamo con \mathcal{S}^+ l'insieme dei punti appartenenti ad \mathcal{M} su cui la mappa T è discontinua. E' facile verificare che

$$\mathcal{S}^+ = \bigcup_{i=1}^4 \partial \mathcal{M}_i^+.$$

Viceversa introduciamo l'insieme di punti \mathcal{S}^- per cui T^{-1} è discontinua. Anche in questo caso si vede che

$$\mathcal{S}^- = \bigcup_{i=1}^4 \partial \mathcal{M}_i^-.$$

3.2 Varietà stabili e instabili

Vediamo adesso cosa può essere salvato dell'argomento di Hopf nel caso di mappe discontinue per dimostrare l'ergodicità di T . In questo caso abbiamo ancora una direzione stabile ed una instabile ma se una linea parallela, ad esempio, alla direzione instabile viene tagliata da \mathcal{S}^- in più parti e se $x, y \in \mathcal{M}$ appartengono a parti diverse, non è per nulla detto che la distanza delle immagini di x e y sotto ripetizioni della mappa T^{-1} tenda a zero. Come visto nel paragrafo precedente infatti, in generale:

$$d(T^{-n}y, T^{-n}x) \not\rightarrow 0, \text{ per } n \rightarrow +\infty. \quad (3.2.1)$$

Ad ogni modo possiamo costruire comunque delle varietà stabili e instabili W^s, W^u , sulle quali le medie temporali, rispettivamente nel futuro e nel passato, di una funzione continua sono costanti. Vediamo come si possa costruire la varietà, ad esempio, instabile procedendo induttivamente.

Per $x \in \mathcal{M} \setminus (\mathcal{S}^+ \cup \mathcal{S}^-)$, definiamo $W_1^u(x)$ come il segmento aperto appartenente alla retta passante per x della stessa direzione dell'autovettore instabile e che ha entrambi gli estremi su \mathcal{S}^- . La preimmagine $T^{-1}W_1^u(x)$ è di un fattore λ più corta di $W_1^u(x)$ e può essere tagliata in due o tre pezzi da \mathcal{S}^- . Prendiamo il pezzo che contiene $T^{-1}(x)$. Notiamo che questo segmento non è altro che $T^{-1}W_1^u(x) \cap W_1^u(T^{-1}(x))$. Se calcoliamo la sua immagine sotto la mappa T quello che otteniamo è la seconda approssimazione della varietà instabile, cioè $W_2^u(x)$:

$$W_2^u(x) = T(T^{-1}W_1^u(x) \cap W_1^u(T^{-1}(x))). \quad (3.2.2)$$

A meno che $T^{-1}(x) \in \mathcal{S}^-$ la seconda approssimazione di varietà stabile è ancora un segmento aperto contenente x con estremi su $\mathcal{S}^- \cup T\mathcal{S}^-$ e naturalmente $W_2^u(x) \subseteq W_1^u(x)$. Data $W_n^u(x)$ con $n = 1, 2, \dots$ definiamo la $(n+1)$ -esima-varietà instabile $W_{n+1}^u(x)$ come

$$W_{n+1}^u(x) = T^n(T^{-n}W_n^u(x) \cap W_n^u(T^{-n}(x))). \quad (3.2.3)$$

Se $x \notin \bigcup_{i=0}^{\infty} T^i \mathcal{S}^-$ allora questa procedura induttiva porterà ad una sequenza nidificata di segmenti aperti contenenti x

$$W_1^u \supseteq W_2^u \supseteq \dots \supseteq W_n^u \supseteq \dots$$

con estremi su $\bigcup_{i=0}^{\infty} T^i \mathcal{S}^-$, oppure sulla sua chiusura. Possiamo anche immaginare la costruzione delle varietà instabili nel seguente modo. Prima di tutto consideriamo il

segmento $W_1^u(x)$ con estremi in \mathcal{S}^- . Poi guardiamo se $T\mathcal{S}^-$ interseca $W_1^u(x)$, nel qual caso $W_2^u(x)$ sarà il segmento contenente x ed eliminiamo gli altri segmenti. Ripetiamo la procedura per $T^2\mathcal{S}^-$, vedendo se taglia $W_2^u(x)$ e chiamando la parte contenente x (che potrebbe anche essere l'intera $W_2^u(x)$) $W_3^u(x)$. Si procede così in modo che la varietà passante per x possa in teoria essere accorciata infinite volte. Il fatto che $x \notin \bigcup_{i=0}^{\infty} T^i\mathcal{S}^-$ assicura che x stia sempre strettamente all'interno di tutti i segmenti $W_n^u(x)$ con $n = 1, 2, \dots$. E' piuttosto impressionante che, per quasi ogni x , questo processo induttivo accorci il segmento solo un numero finito di volte. Più precisamente vale la seguente:

Proposizione 3.2.1. *Per quasi ogni $x \in \mathcal{M} \setminus \bigcup_{i=0}^{\infty} T^i\mathcal{S}^-$ la sequenza di varietà instabili approssimate di x si stabilizza, cioè esiste $M = M(x) \in \mathbb{N}$ tale che*

$$\bigcap_{i=1}^{\infty} W_i^u(x) = \bigcap_{i=1}^{M(x)} W_i^u(x) \quad (3.2.4)$$

Dimostrazione. Sia, per $t > 0$,

$$X_t = \{x \in \mathcal{M}^- \mid d(x, \mathcal{S}^-) \leq t\} \quad (3.2.5)$$

dove in questo caso $d(\cdot, \cdot)$ è la distanza tra un punto e un insieme definita da:

$$d(b, A) = \inf_{a \in A} \tilde{d}(b, a)$$

con $\tilde{d}(\cdot, \cdot)$ ordinaria distanza euclidea. Dal momento che \mathcal{S}^- è un'unione finita di segmenti abbiamo

$$Leb_2(X_t) \leq \text{costante} \cdot t. \quad (3.2.6)$$

Se scegliamo $t_n = \frac{1}{n^2}$ abbiamo

$$\sum_{n=1}^{\infty} Leb_2(X_{t_n}) < \infty. \quad (3.2.7)$$

E, siccome T conserva la misura ($\det T = 1$), allora anche

$$\sum_{n=1}^{\infty} Leb_2(T^n X_{t_n}) < \infty. \quad (3.2.8)$$

Segue dal lemma di Borel-Cantelli (vedi Appendice A) che quasi ogni x appartiene a un numero finito degli insiemi:

$$T^1 X_{t_1}, T^2 X_{t_2}, \dots, T^n X_{t_n}, \dots$$

Questo significa che esiste un insieme $\mathcal{M}_0 \subset \mathcal{M}$, con $Leb_2(\mathcal{M}_0) = 1$, tale che $\forall x \in \mathcal{M}_0$, l' 'evento':

$$d(T^{-n}(x), \mathcal{S}^-) > \frac{1}{n^2}. \quad (3.2.9)$$

si realizza per tutti gli $n \in \mathbb{N}$, tranne al più un numero finito di volte. Scegliendo $c(x) > 0$ sufficientemente piccolo allora possiamo fare in modo che la (2.0.10) valga per tutti i valori di n cosicchè

$$d(T^{-n}(x), \mathcal{S}^-) > \frac{c(x)}{n^2} \quad (3.2.10)$$

per ogni $n = 1, 2, \dots$. Ricordiamo che λ_2 è l'autovalore relativo all'autovettore che definisce la direzione instabile e $\lambda_2 > 1$. Se vogliamo che ad esempio $W_2^u(x)$ sia più piccola di $W_1^u(x)$ allora dobbiamo avere che

$$d(T^{-1}(x), \mathcal{S}^-) < \frac{\text{lunghezza}(W_1^u(x))}{\lambda_2}. \quad (3.2.11)$$

Infatti quando applichiamo T^{-1} a $W_n^u(x)$ per vedere se il risultante segmento interseca \mathcal{S}^- , la lunghezza diminuisce di un fattore λ_2 :

$$\text{lunghezza}(T^{-1}W_n^u(x)) = \frac{\text{lunghezza}(W_n^u(x))}{\lambda_2}, \quad (3.2.12)$$

ed allora, in generale, ogni volta che avviene un taglio, cioè ogni volta che $W_{n+1}^u(x)$ è più corta di $W_n^u(x)$ si deve avere

$$d(T^n(x), \mathcal{S}^-) < \frac{\text{lunghezza}(W_n^u(x))}{\lambda_2^n}. \quad (3.2.13)$$

Ma allora

$$\frac{c(x)}{n^2} < \frac{\text{lunghezza}(W_n^u(x))}{\lambda_2^n} \leq \frac{\text{lunghezza}(W_1^u(x))}{\lambda_2^n} = \frac{A(x)}{\lambda_2^n} \quad (3.2.14)$$

con $A(x)$ costante che dipende da x . Siccome $\lambda_2 > 1$ la (3.2.14) può essere valida solo per un numero finito di valori di n . \square

Sia allora \mathcal{U} l'insieme dei punti per i quali le successive approssimazioni delle varietà instabili si arrestano per un certo M . Abbiamo che \mathcal{U} ha misura piena

$$Leb_2(\mathcal{M} \setminus \mathcal{U}) = 0. \quad (3.2.15)$$

Solo per gli $x \in \mathcal{U}$ definiamo le varietà instabili prendendo quello che rimane dopo un numero M finito di tagli

$$W^u(x) = \bigcap_{i=1}^{\infty} W_i^u(x) = \bigcap_{i=1}^{M(x)} W_i^u(x). \quad (3.2.16)$$

La proprietà fondamentale per le quali abbiamo costruito queste varietà è che adesso, sebbene la varietà possa essere estremamente piccola, per tutti gli $y \in W^u(x)$ vale:

$$d(T^{-n}(y), T^{-n}(x)) = \frac{d(x, y)}{\lambda_2^n} \rightarrow 0 \text{ per } n \rightarrow \infty \quad (3.2.17)$$

Tutto quello che abbiamo fatto per le varietà instabili può essere ripetuto per quelle stabili. In particolare se \mathcal{V} è l'insieme degli x per le quali le successive approssimazioni delle varietà stabili si arrestano per un certo numero naturale M , allora \mathcal{V} ha misura piena, cioè:

$$\text{Leb}_2(\mathcal{M} \setminus \mathcal{V}) = 0, \quad (3.2.18)$$

e definiamo le varietà stabili solo per gli $x \in \mathcal{V}$ come:

$$W^s(x) = \bigcap_{i=1}^{\infty} W_i^s(x) = \bigcap_{i=1}^{M'(x)} W_i^s(x), \quad (3.2.19)$$

dove stavolta:

$$W_{n+1}^s(x) = T^{-n}(T^n W_n^s(x) \cap W_n^s(T^n(x))), \quad (3.2.20)$$

e $M'(x)$ è l'analogo per la varietà stabile di $M(x)$. In analogia con la (3.2.17), si ha:

$$d(T^n(y), T^n(x)) = \frac{d(x, y)}{\lambda_2^n} \rightarrow 0 \text{ per } n \rightarrow \infty. \quad (3.2.21)$$

Siano $x, y \in \mathcal{M} \setminus (\mathcal{S}^+ \cup \mathcal{S}^-)$ punti vicini a piacere appartenenti alla stessa varietà instabile $W^u(z)$, $z \in \mathcal{M}$. Analizziamo le orbite di x e y per le prime 2 iterazioni della mappa T^{-1} . Se applico T^{-1} ad \mathcal{M} è possibile che $T^{-1}W^u$ venga 'tagliata' in più pezzi da \mathcal{S}^- . Notiamo che in questo caso è possibile che $T^{-1}(x)$ e $T^{-1}(y)$ appartengano a sottodomini diversi della mappa T^{-1} (cioè a due diversi \mathcal{M}_i^-). Se questo è il caso all'iterazione successiva di T^{-1} , non possiamo dire più nulla sulla distanza tra $T^{-2}(x)$ e $T^{-2}(y)$. Dopo sole 2 iterazioni della mappa, due punti appartenenti alla stessa varietà instabile $W^u(z)$ possono trovarsi anche molto lontani tra loro. In queste condizioni l'argomento di Hopf, così come è stato presentato nel precedente capitolo, non può funzionare.

3.3 L'argomento di Hopf per le mappe discontinue

Sia $f : \mathcal{M} \rightarrow \mathbb{R}$ una funzione continua (e dunque anche uniformemente continua, siccome \mathcal{M} è un compatto). Per ogni funzione continua $f : \mathcal{M} \rightarrow \mathbb{R}$ la media temporale nel passato f^- è costante sulle varietà instabili per la (2.4.5) e, analogamente, la media temporale nel futuro f^+ è costante sulle varietà stabili.

Definizione 3.3.1. Sia $x \in \mathcal{M}$, diciamo che x è un punto *f1-tipico* se esiste $f^+(x)$, $f^-(x)$, $f^+(x) = f^-(x)$ ed esiste la varietà stabile e instabile passante per x nel senso della Proposizione 3.2.1.

Come vedremo a breve, quasi tutti i punti in \mathcal{M} sono *f1-tipici*.

Definizione 3.3.2. Sia $x \in \mathcal{M}$, diciamo che x è un punto *f2-tipico* se è *f1-tipico* ed esiste una varietà stabile ed una instabile passante per x tale che tutti i punti in esse sono *f1-tipici* a parte che per un insieme di punti di misura di Lebesgue unidimensionale nulla.

Definizione 3.3.3. Una varietà $W^u(x)$ o $W^s(x)$ si dice *f-tipica* se la misura di Lebesgue unidimensionale dei punti non *f1-tipici* appartenenti ad essa è nulla.

$$Leb_1(\{y|y \in W(x), y \text{ non è } f1\text{-tipico}\}) = 0 \quad (3.3.1)$$

Risulta chiaro che un punto è *f2-tipico* se e solo se è *f1-tipico* ed entrambe le varietà (stabile e instabile) cui appartiene sono *f-tipiche*. Come conseguenza diretta del Teorema di Birkhoff e della Proposizione 3.2.1, possiamo fare questa prima affermazione sulla probabilità che un punto di \mathcal{M} scelto a caso sia *f1-tipico*.

Teorema 3.3.4. *Quasi tutti i punti appartenenti ad \mathcal{M} sono f1-tipici.*

Dimostrazione. Possiamo definire i seguenti insiemi:

$$\mathcal{A} = \{x \in \mathcal{M} | \text{esiste } f^+(x)\}$$

$$\mathcal{B} = \{x \in \mathcal{M} | \text{esiste } f^-(x)\}$$

$$\mathcal{C} = \{x \in \mathcal{M} | f^+(x) = f^-(x)\}$$

$$\mathcal{D} = \{x \in \mathcal{M} | \text{esiste } W^u(x) \text{ nel senso della Proposizione 3.2.1}\}$$

$$\mathcal{E} = \{x \in \mathcal{M} | \text{esiste } W^s(x) \text{ nel senso della Proposizione 3.2.1}\}.$$

Un punto $x \in \mathcal{M}$ è *f1-tipico* se $x \in \mathcal{A} \cap \mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \cap \mathcal{E}$ e dunque l'insieme \mathcal{F} dei punti *f1-tipici* è uguale a $\mathcal{A} \cap \mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \cap \mathcal{E}$ e $Leb_2(\mathcal{F}) = Leb_2(\mathcal{A} \cap \mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \cap \mathcal{E})$. Abbiamo però che per il teorema di Birkhoff gli insiemi \mathcal{A} , \mathcal{B} e \mathcal{C} hanno misura piena e per la proposizione 3.2.1 anche \mathcal{D} e \mathcal{E} hanno misura piena. Segue che \mathcal{F} ha misura piena e

$$Leb_2(\mathcal{M} \setminus \mathcal{F}) = 0.$$

□

Tuttavia per formare una catena di Hopf non basta avere a disposizioni punti *singoli* per cui f^+ ed f^- sono ben definite e uguali. I punti più utili sono quelli che permettono di costruire in modo agevole delle *intersezioni* in punti *f1-tipici* tra *due varietà distinte*. A tal proposito iniziamo con l'enunciare il seguente:

Teorema 3.3.5. *Quasi tutti i punti appartenenti ad \mathcal{M} sono $f2$ -tipici.*

Dimostrazione. Il Teorema 3.3.4 ci assicura che l'insieme dei punti non $f1$ -tipici ha misura di Lebesgue bidimensionale nulla. Dunque se \mathcal{Z} è l'insieme dei punti non $f1$ -tipici allora:

$$Leb_2(\mathcal{Z}) = 0. \quad (3.3.2)$$

Supponiamo per assurdo che l'enunciato del teorema sia falso e mostriamo che questo implicherebbe che la (3.3.2) è falsa. Se non è vero che quasi tutti i punti di \mathcal{M} sono $f2$ -tipici allora esiste un sottoinsieme $\mathcal{G} \subseteq \mathcal{M}$ che ha misura piena di punti non $f2$ -tipici tale che

$$Leb_2(\mathcal{G}) > 0, \quad (3.3.3)$$

cioè di misura di Lebesgue bidimensionale non nulla. D'altra parte, per quanto già dimostrato, tutti i punti, eccetto che per un insieme di misura nulla, di \mathcal{G} sono $f1$ -tipici. Abbiamo allora che quasi tutti i punti dell'insieme \mathcal{G} sono $f1$ -tipici per il Teorema 3.3.4 e quasi tutti i punti di \mathcal{G} sono non $f2$ -tipici. Siccome l'intersezione di due insiemi di misura piena è ancora di misura piena abbiamo che quasi tutti i punti di \mathcal{G} hanno la proprietà di essere $f1$ -tipici ma non $f2$ -tipici. Dunque se

$$\mathcal{B} = \{x \in \mathcal{M} \mid x \text{ è } f1\text{-tipico ma non } f2\text{-tipico}\}, \quad (3.3.4)$$

allora

$$Leb_2\{\mathcal{G} \setminus \mathcal{B}\} = 0. \quad (3.3.5)$$

Chiaramente $\mathcal{B} \subset \mathcal{G}$. Vediamo adesso che, come conseguenza del teorema di Fubini, esiste sicuramente un segmento l diretto come le varietà stabili di misura di Lebesgue unidimensionale positiva, tale che vi è una misura unidimensionale positiva di punti di l appartenenti a \mathcal{B} .

Consideriamo un rettangolo \mathcal{V} contenuto in \mathcal{G} di lati l^+ ed l^- , con l^+ diretto come la direzione stabile (verticale) ed l^- diretto come quella instabile (orizzontale), l^+ ed l^- sufficientemente piccoli per cui $\mathcal{V} \subset \mathcal{M}$. Abbiamo che la misura di tale rettangolo è positiva, se i lati non sono degeneri (cioè non si riducono a punti). Dal momento che $\mathcal{V} \subseteq \mathcal{G}$ abbiamo necessariamente che

$$Leb_2(\mathcal{V} \setminus \mathcal{B}) = 0.$$

Vediamo che esiste un segmento $l \subset \mathcal{G}$, di misura unidimensionale positiva, diretto come la foliazione stabile, tale che $l \cap \mathcal{B}$ è di misura positiva. Supponiamo per assurdo che tale segmento non esista. Avremmo allora:

$$\begin{aligned} Leb_2\{\mathcal{V}\} &= \int_{l^-} dLeb_1(y) \int_{l^+(y)} dLeb_1(x) = \int_{l^-} dLeb_1(y) \int_{l^+(y) \setminus \mathcal{B}} dLeb_1(x) \\ &= Leb_2(\mathcal{V} \setminus \mathcal{B}) = 0, \end{aligned} \quad (3.3.6)$$

dove $l^+(y)$ è il segmento parallelo ad l^+ (dunque diretto come la foliazione stabile) che interseca l^- nel punto y . L'equazione (3.3.6) è assurda in quanto avevamo scelto i lati l^+ ed l^- non degeneri. Chiaramente la seconda uguaglianza è valida solo se si assume l'ipotesi che non esista un segmento $l \subset \mathcal{G}$ diretto come le varietà stabili la cui misura unidimensionale di punti appartenenti a \mathcal{B} è positiva. Possiamo quindi concludere che tale segmento l esiste e per esso sia vero che:

$$Leb_1(l \cap \mathcal{B}) = Leb_1(l) > 0. \quad (3.3.7)$$

Costruiamo adesso un insieme $\mathcal{C} \subset \mathcal{M}$ tale che

$$Leb_2(\mathcal{C}) > 0,$$

e abbia misura piena di punti non *f1-tipici*. Sappiamo che per quasi ogni punto $x \in l$ passa una varietà instabile non *f-tipica*. Infatti per la (3.3.7) quasi tutti i punti di l sono *f1-tipici* e non *f2-tipici*. Sia \mathcal{A} l'insieme dei punti *f1-tipici* e sia:

$$\mathcal{C} = \bigcup_{x \in (l \cap \mathcal{B})} (W^u(x) \setminus \mathcal{A}) \quad (3.3.8)$$

Abbiamo che

$$Leb_2(\mathcal{C}) = \int_{l \cap \mathcal{B}} Leb_1(W^u(x) \setminus \mathcal{A}) dLeb_1(x).$$

Per costruzione tutti i punti di \mathcal{C} sono non *f1-tipici*. Dal momento che tutti gli $x \in l^+ \cap \mathcal{B}$ sono non *f2-tipici*, $Leb_1(W^u(x) \setminus \mathcal{A}) > 0 \forall x \in l \cap \mathcal{B}$. Inoltre, per la (3.3.2), $Leb_1(l \cap \mathcal{B}) > 0$. Abbiamo allora che per il Lemma che riportiamo qua di seguito $Leb_2(\mathcal{C}) > 0$ e si ottiene l'assurdo desiderato. \square

Lemma 3.3.6. *Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ uno spazio di misura e sia $f > 0$ quasi dovunque secondo la misura μ . Sia inoltre $A \in \mathfrak{B}(\mathcal{M})$, $\mu(A) > 0$. Allora:*

$$\int_A f d\mu > 0.$$

Dimostrazione. Poniamo:

$$A_n = \left\{ x \in A \mid f(x) \geq \frac{1}{n} \right\}.$$

Chiaramente $A = \bigcup_{n \in \mathbb{N}} A_n$ e $A_n \subseteq A_{n+1}$. Inoltre possiamo pensare a $\mu(A)$ come il limite (crescente) della successione $\mu(A_n)$: abbiamo allora che esiste sicuramente $\bar{n} \in \mathbb{N}$ tale che $\mu(A_{\bar{n}}) > \epsilon > 0$. In questo caso possiamo scrivere:

$$\int_A f d\mu \geq \int_{A_{\bar{n}}} f d\mu \geq \frac{1}{\bar{n}} \mu(A_{\bar{n}}) > 0.$$

\square

Sappiamo allora che quasi tutti i punti di \mathcal{M} sono non solo *f1-tipici* ma anche *f2-tipici*. Questo significa che le varietà (stabili e instabili) in \mathcal{M} hanno tutte misura piena di punti *f1-tipici* eccetto che per un insieme di varietà i cui punti costituiscono un insieme di misura bidimensionale di Lebesgue nulla. Vediamo allora quale potrebbe essere una prima idea (non immediatamente attuabile nei termini in cui verrà esposta ora) per dimostrare l'ergodicità di T . Se $W^s(x)$ non è una di quelle varietà particolari, cioè x è *f2-tipico*, allora il set

$$C_1 = \bigcup_{y \in W^s(x), y \text{ è } f1\text{-tipico}} W^u(y) \quad (3.3.9)$$

ha misura positiva e $f^- = f^+ = \text{costante}$ su C_1 . Possiamo procedere aggiungendo tutte le varietà stabili passanti attraverso i punti *f1-tipici* in C_1 per ottenere C_2 e via dicendo. Ma non c'è a priori ragione di aspettarsi che saremo in grado di ricoprire tutto \mathcal{M} in questo modo. Ci si può infatti immaginare che ci sia una linea che divida due insiemi invarianti del nostro sistema e che tutte le varietà, stabili e instabili, non oltrepassino in alcun modo questa linea. Questa è la ragione per cui il metodo di Hopf può non funzionare: si può dimostrare che ogni componente ergodica² ha misura positiva ma non è chiaro che ve ne sia una sola [5].

3.4 Il metodo di Sinai

Abbiamo visto nella sezione precedente che l'argomento di Hopf non è sufficiente per provare l'ergodicità di una mappa discontinua T dal momento che le varietà stabili e instabili potrebbero essere *troppo corte*. Il metodo di Sinai raggiunge l'obiettivo di dimostrare che la maggior parte delle varietà stabili e instabili sono in un certo senso lunghe abbastanza. Il primo passo è di formulare precisamente cosa si intende con *sufficientemente lunga*. Come fatto in precedenza trattiamo solo il caso di varietà instabili, considerando la traduzione per quelle stabili automatica. Sia $\mathcal{U} \subseteq \mathcal{M}$ un piccolo quadrato con i lati rispettivamente paralleli alle direzioni stabili e instabili (per semplificare la geometria si può pensare a queste direzioni come la direzione verticale e orizzontale, rispettivamente). Per ogni $0 < c < 1$ costruiamo una sequenza $\mathcal{G}_n(c), n = 1, 2, \dots$ di ricoprimenti di \mathcal{U} nel seguente modo. Senza perdita di generalità sia

$$\mathcal{U} = \{(u, v) \in \mathbb{R}^2 \mid -b < u < b, -b < v < b\} \quad (3.4.1)$$

$b \in \mathbb{R}$ abbastanza piccolo da fare in modo che $\mathcal{U} \subset \mathcal{M}$. Consideriamo i punti del reticolo $\mathcal{N}(n, c)$ definito da

$$\mathcal{N}(n, c) = \left\{ \frac{c}{n}(m, k) \in \mathcal{U} \mid m, k \in \mathbb{Z} \right\}. \quad (3.4.2)$$

²Rimandiamo alla nota 1 del Capitolo 2 per la nozione di 'componente ergodica' in questo lavoro.

Possiamo ora definire i ricoprimenti $\mathcal{G}_n(c)$ come una collezione di quadrati aventi centro nei punti del reticolo $\mathcal{N}(n, c)$ e lati di lunghezza $\frac{1}{n}$ paralleli ai lati di \mathcal{U} . Il numero c è un indice di quanto sono vicini i centri dei quadrati di $\mathcal{G}_n(c)$, in particolare più c è piccolo più due quadratini adiacenti si sovrappongono. Inoltre, se $c < \frac{1}{2}$, $\mathcal{G}_n(c)$ è un ricoprimento di \mathcal{U} , altrimenti $\mathcal{G}_n(c)$ ricopre un quadrato più piccolo contenuto in \mathcal{U} . Il parametro c sarà in futuro scelto molto piccolo cosicché più quadrati di $\mathcal{G}_n(c)$ si sovrappongono per una gran parte. Ad ogni modo, una volta che c è fissato, un punto di \mathcal{U} può appartenere al massimo a $k(c)$ quadrati in $\mathcal{G}_n(c)$, dove $k(c)$ è un numero fisso indipendente da $n = 0, 1, 2, \dots$; si può facilmente stimare che $k(c) \leq (\frac{1}{c} + 1)^2$ ma non ci servirà nessuna stima esplicita.

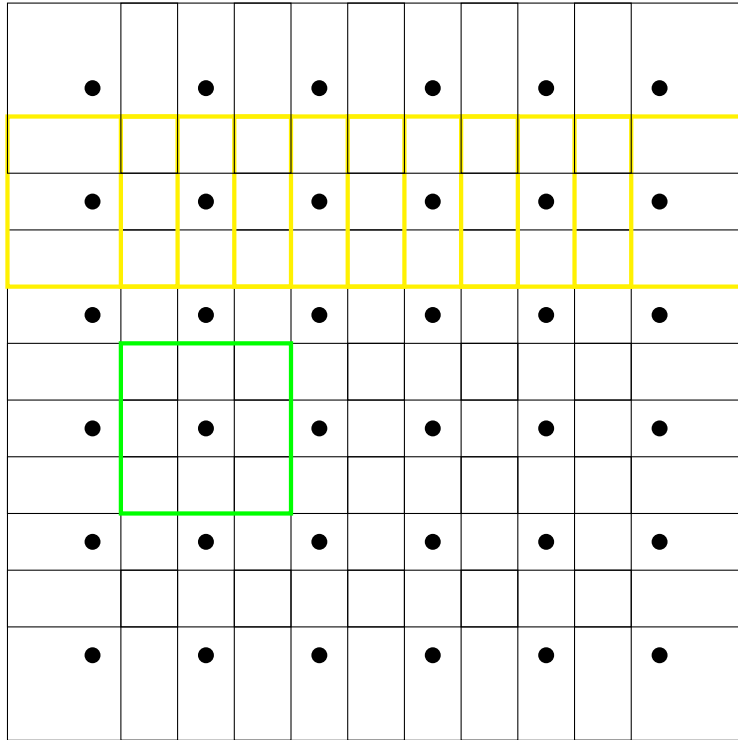


Fig. 4: in figura i punti neri rappresentano la griglia $\mathcal{N}(\frac{2}{3}, \frac{2}{3})$, centrati nei punti abbiamo invece la collezione $\mathcal{G}_{\frac{2}{3}}(\frac{2}{3})$. In verde è rappresentato un quadrato $R \in \mathcal{G}_{\frac{2}{3}}(\frac{2}{3})$, mentre in giallo troviamo una *riga* di quadrati $\{R_i\}_{i=1}^6$. I quadrati $R \in \mathcal{G}_{\frac{2}{3}}(\frac{2}{3})$ si sovrappongono per $1 - c = \frac{1}{3}$ della loro area. Inoltre, come si vede dal disegno, ogni punto può appartenere al massimo a quattro quadrati contemporaneamente.

Definiamo due quadrati in $\mathcal{G}_n(c)$ adiacenti se la distanza tra i loro centri è $\frac{c}{n}$, cioè se sono ottenuti da due punti del reticolo $\mathcal{N}(n, c)$ orizzontalmente o verticalmente consecutivi. E' importante ricordare che due quadrati adiacenti si sovrappongono per una percentuale $1 - c$ della loro area. Si possono anche definire particolari sottoinsiemi di $\mathcal{G}_n(c)$ come una colonna o una riga di quadrati. Ad esempio, una sequenza $\{R_i\}_{i=1}^l$ di

$\mathcal{G}_n(c)$ è chiamata colonna di quadrati se, per ogni $i = 1, \dots, l-1$, R_i ed R_{i+1} sono adiacenti, R_{i+1} è sopra R_i e non c'è nessun quadrato di $\mathcal{G}_n(c)$ sotto R_1 o sopra R_l (vedi figura 4).

Per ogni quadrato $R \in \mathcal{G}_n(c)$ introduciamo i bordi stabili ed instabili, $\partial_s R$ e $\partial_u R$ rispettivamente: $\partial_s R$ è l'unione dei due segmenti di bordo di R che sono posti lungo la direzione stabile (verticale) e $\partial_u R$ è l'unione dei segmenti di bordo di R posti nella direzione instabile (orizzontale). Dato un punto $x \in R$, la varietà instabile passante per x , $W^u(x)$, potrebbe intersecare entrambi, uno solo o nessuno dei segmenti di bordo in $\partial_s R$. Nel primo caso ($W^u(x)$ interseca entrambi i segmenti di bordo di $\partial_s R$) diciamo che $W^u(x)$ è lunga (o *connettente*) in R , altrimenti (secondo e terzo caso) diciamo che $W^u(x)$ è corta in R .

Definizione 3.4.1. Dato α , $0 < \alpha < 1$, chiamiamo un quadrato α -connettente se la misura dell'insieme di punti $x \in R$ le cui varietà instabili $W^u(x)$ sono lunghe in R è almeno α parte dell'area di R , ovvero:

$$Leb_2(\{x \in R | W^u(x) \text{ è lunga in } R\}) \geq \alpha \cdot Leb_2(R). \quad (3.4.3)$$

Sinai formula la proprietà che la maggior parte delle varietà instabili è sufficientemente lunga nella seguente maniera.

Teorema 3.4.2. (di Sinai)[5] *Esiste $\alpha_0 < 1$ tale che per ogni α , $0 < \alpha \leq \alpha_0$ ed ogni c , $0 < c < 1$,*

$$\lim_{n \rightarrow +\infty} n \cdot Leb_2(\bigcup \{R \in \mathcal{G}_n(c) | R \text{ non è } \alpha\text{-connettente}\}) = 0 \quad (3.4.4)$$

In altre parole se α è sufficientemente piccolo, allora l'unione di quadrati di $\mathcal{G}_n(c)$ che è non α -connettente ha misura $o(\frac{1}{n})$ per $n \rightarrow \infty$.

Dimostrazione. La dimostrazione di questo teorema è un caso particolare della dimostrazione riportata nella Appendice B, in cui si pone $K = 0$. \square

Osservazione 3.4.3. Questa formulazione del teorema di Sinai (che vale per un qualsiasi quadrato $\mathcal{U} \subset \mathcal{M}$, anche molto grande) fa affidamento sul fatto che le direzioni dei segmenti in $\mathcal{S}^+ \setminus \partial\mathcal{M}$ sono contenute nel settore $\mathfrak{C}_+ \subset \mathbb{R}^2$,

$$\mathfrak{C}_+ = \{(\xi, \eta) \in \mathbb{R}^2 | |\xi| \leq \chi(a)|\eta|\}$$

dove $\chi(a) \in \mathbb{R}$ dipende dal parametro a che definisce la mappa T , ma $\forall a$ $0 < \chi(a) < 1$ è un numero fissato. Inoltre ξ è la coordinata della direzione instabile mentre η quella direzione stabile. Viceversa, perché valga la formulazione presentata sopra del Teorema di Sinai, le direzioni dei segmenti in $\mathcal{S}^- \setminus \partial\mathcal{M}$ devono essere contenute nel settore $\mathfrak{C}_- \subset \mathbb{R}^2$,

$$\mathfrak{C}_- = \{(\xi, \eta) \in \mathbb{R}^2 | |\eta| \leq \chi(a)|\xi|\}$$

dove $0 < \chi(a) < 1$. Nel prossimo capitolo saremo interessati in una generalizzazione in cui questa condizione viene a mancare.

Vediamo adesso come sia possibile utilizzare questo teorema per ottenere informazioni sulle componenti ergodiche di \mathcal{M} . Si noti anche come la Definizione 3.4.1 e il Teorema 3.4.2 possano essere entrambi riscritti per le varietà stabili.

Teorema 3.4.4. *Il quadrato $\mathcal{U} \subset \mathcal{M}$ (per il quale vale il teorema di Sinai per entrambe le varietà stabili e instabili) appartiene ad una sola componente ergodica di T .*

Dimostrazione. Per chiarezza dividiamo la dimostrazione in passi.

(i) *Definiamo $W^{u(s)}(R)$.*

Fissiamo un α sufficientemente piccolo da far valere il teorema di Sinai per quadrati α -connettente in entrambe le versioni di foliazioni stabile e instabile. Fissiamo poi $c < \alpha$. Come conseguenza l'intersezione di due quadrati adiacenti α -connettente in $\mathcal{G}_n(c)$ contiene un insieme di varietà lunghe (in uno dei due quadrati) di misura positiva. Ricordiamo infatti che la misura dell'intersezione di due quadrati adiacenti è pari a $1 - c$ della misura di uno dei quadrati e un quadrato α -connettente ha una misura di varietà connettenti pari ad α parti la misura del quadrato. Dal momento che la parte di un quadrato ricoperta da varietà connettenti non può essere completamente contenuta nella frazione c dello stesso quadrato che non fa parte dell'intersezione, abbiamo che, se $c < \alpha$, allora ci deve essere una misura di varietà connettenti nell'intersezione almeno pari ad $\alpha - c$. Per semplicità chiamiamo un quadrato α -connettente (sia per le varietà stabili che per quelle instabili) semplicemente quadrato connettente.

Consideriamo una qualsiasi funzione f continua su \mathcal{M} . Per ogni quadrato connettente R definiamo:

$$W^{u(s)}(R) = \{x \in R \mid W^{u(s)}(x) \text{ è } f\text{-tipica e lunga in } R\}.$$

Per il Teorema 3.3.5 quasi tutti i punti di \mathcal{M} sono f 2-tipici dunque l'insieme delle varietà non f -tipiche è un insieme di misura bidimensionale nulla. Una conseguenza di questo fatto è che se un quadrato in $\mathcal{G}_n(c)$ è α -connettente (e cioè la misura delle varietà lunghe nel quadrato è almeno α -parte della misura del quadrato) allora è anche α -connettente di varietà f -tipiche, ovvero la misura delle varietà lunghe ed f -tipiche nel quadrato è almeno α -parte della misura del quadrato. Questo fatto sarà utile nel prosieguo della dimostrazione.

(ii) *Dimostriamo che $f^+ = \text{costante}$ su tutto $W^s(R)$ e $f^- = \text{costante}$ su tutto $W^u(R)$, con le due costanti che coincidono³.*

³Ricordiamo che la funzione f è uniformemente continua (in quanto continua su un compatto), e dunque f^+ assume lo stesso valore su *tutti* i punti di una varietà locale stabile ed f^- su *tutti* i punti di una varietà locale instabile. Se f non fosse uniformemente continua, non potremmo facilmente affermare ad esempio che $f^+ = \text{costante}$ per tutti i punti di un insieme di misura positiva.

Sebbene non possiamo applicare l'argomento di Hopf all'intero dominio \mathcal{M} , possiamo usarlo nel quadrato *connettente* R per dimostrare che f^+ è costante su *tutto* $W^s(R)$ ed f^- è costante su *tutto* $W^u(R)$ e che tali due costanti coincidono. Si noti che in questo caso si ha che $f^{+(-)}$ è costante su tutti i punti (e non semplicemente quasi tutti i punti) di $W^{s(u)}(R)$. Vediamo come sia possibile affermare ciò. Innanzitutto f^+ è costante su ciascuna delle varietà stabili di $W^s(R)$. Fissiamo adesso una varietà instabile $W^u(z_1)$ di $W^u(R)$. Chiamiamo O_1 l'insieme dei punti appartenenti a varietà di $W^s(R)$ che intersecano $W^u(z_1)$ in punti non *f1-tipici* (per la definizione di *f1-tipicità*, *f2-tipicità* e varietà *f-tipiche* si torni alla sezione 3.3). Dimostriamo che

$$Leb_2(O_1) = 0$$

Se indichiamo con $W^{u*}(z_1)$ la varietà $W^u(z_1)$ a cui sono stati tolti i punti non *f1-tipici*:

$$W^{u*}(z_1) = W^u(z_1) \setminus \{x \in W^u(z_1) | x \text{ non è } f1\text{-tipico}\}$$

allora per il teorema di Fubini:

$$\begin{aligned} Leb_2(W^s(R)) &= \int_{W^u(z_1)} dLeb_1(y) \int_{W^s(y)} dLeb_1(x) = \int_{W^{u*}(z_1)} dLeb_1(y) \int_{W^s(y)} dLeb_1(x) \\ &= Leb_2(W^s(R) \setminus O_1) \end{aligned}$$

Dove la seconda uguaglianza è lecita siccome $W^u(z_1)$ è per ipotesi *f-tipica*. Escludendo dunque queste varietà particolari (ovvero escludendo i punti appartenenti a O_1) il valore di f^+ sulle varietà stabili deve necessariamente coincidere con il valore costante di f^- sulla particolare varietà instabile $W^u(z_1)$ che abbiamo scelto (infatti nei punti *f1-tipici* $f^+ = f^-$). Concludiamo che f^+ assume un valore costante quasi dovunque su $W^s(R)$ e che tale valore costante è uguale a quello che assume f^- sulla varietà instabile $W^u(z_1)$. Consideriamo adesso una qualsiasi altra varietà $W^u(z_2) \in W^u(R)$ e dimostriamo che il valore (costante) che f^- assume su questa varietà deve necessariamente essere uguale al valore che f^- assume su $W^u(z_1)$.

Infatti abbiamo che se

$$f^- |_{W^u(z_1)} = A$$

allora quasi dovunque su $W^s(R)$ deve essere:

$$f^+ |_{W^s(R)} = A \quad \text{q.d.} \quad (3.4.5)$$

In particolare le varietà di $W^s(R)$ per le quali non possiamo ancora essere sicuri che $f^+ = A$ sono un insieme di misura di Lebesgue bidimensionale nulla. Definiamo l'insieme $W^s(R)^* \subseteq W^s(R)$

$$W^s(R)^* = W^s(R) \setminus O_1$$

dei punti per i quali vale dappertutto la (3.4.5).

$$f^+ |_{W^s(R)^*} = A.$$

Sia $O_2 = W^s(R)^* \cap W^u(z_2)$. Dimostriamo che O_2 ha misura unidimensionale positiva. Dal momento che $W^s(R)^*$ ha misura bidimensionale positiva, per il teorema di Fubini

$$0 < \text{Leb}_2(W^s(R)^*) = \int_{W^u(z_2) \cap W^s(R)^*} d\text{Leb}_1(y) \int_{W^s(y)} d\text{Leb}_1(x)$$

da cui si ottiene:

$$\text{Leb}_1(O_2) > 0.$$

Se eliminiamo da O_2 i punti non *f1-tipici* otteniamo un nuovo insieme O'_2 la cui misura di Lebesgue unidimensionale è ancora strettamente positiva ($W^u(z_2)$ è *f1-tipica*).

$$\text{Leb}_1(O'_2) > 0.$$

O'_2 è l'insieme dei punti *f1-tipici* di intersezione tra $W^u(z_2)$ e tutte quelle varietà di $W^s(R)$ per le quali siamo già certi che $f^+ = A$. Siccome tale insieme non è vuoto (e anzi è di misura unidimensionale positiva) siamo certi che f^- assuma su $W^u(z_2)$ lo stesso valore costante A che assumeva su $W^u(z_1)$ e su quasi tutte le varietà di $W^s(R)$. Siccome due qualsiasi varietà ($W^u(z_1)$ e $W^u(z_2)$) appartenenti a $W^u(R)$ assumono lo stesso valore costante, dobbiamo dedurre che f^- è costante con lo stesso valore su *tutto* $W^u(R)$. Prendendo ora due qualsiasi varietà stabili di $W^s(R)$, $W^s(z'_1)$ e $W^s(z'_2)$, applicando un ragionamento in tutto e per tutto analogo a questo abbiamo che f^+ assume lo stesso valore costante su queste due varietà. Si ha allora che necessariamente

$$f^+ |_{W^s(R)} = f^- |_{W^u(R)} = \text{costante} = A,$$

con tale costante A unica.

Riassumendo abbiamo che se consideriamo le varietà appartenenti a un quadrato $R \in \mathcal{G}_n(c)$ vale il seguente fatto:

*se due qualsiasi varietà, una stabile e l'altra instabile, $W^s(z)$ e $W^u(z')$ si intersecano in un punto *f1-tipico*, grazie all'uniforme continuità di f , $f^+ |_{W^s(z)} = f^- |_{W^u(z')} = \text{costante}$. Inoltre su tutte le varietà *f-tipiche* stabili (instabili), la $f^{+(-)}$ ha lo stesso valore costante.*

Si presti attenzione al fatto che non possiamo asserire con certezza che ad esempio $f^+ = \text{costante}$ su *tutto* $W^u(R)$, con R quadrato α -connettente. Infatti nei punti non *f1-tipici* di $W^u(R)$ (che sono comunque un insieme di misura nulla) non sappiamo nemmeno se f^+ esiste. Discorso ben diverso vale per la f^- su $W^u(R)$: infatti in una varietà instabile basta che la f^- sia ben definita in un punto e per la (3.2.17) essa esiste in *ogni*

punto di quella varietà e su questa è costante.

(iii) Dimostriamo che se R_1 ed R_2 sono due quadrati adiacenti e connettenti allora $f^+|_{W^u(R_1)} = f^-|_{W^u(R_1)} = f^+|_{W^u(R_2)} = f^-|_{W^u(R_2)} = \text{costante}$ ⁴.

Per due quadrati *connettenti* adiacenti R_1 e R_2 si ha che f^+ è costante su $W^s(R_1) \cup W^s(R_2)$ ed f^- è costante su $W^u(R_1) \cup W^u(R_2)$. Infatti almeno una delle intersezioni $W^u(R_1) \cap W^s(R_2)$ oppure $W^s(R_1) \cap W^u(R_2)$ ha misura bidimensionale positiva e quindi, per il Teorema 3.3.4, vi è una misura positiva anche di punti *f1-tipici* nell'intersezione $W^u(R_1) \cap W^s(R_2)$ (o $W^s(R_1) \cap W^u(R_2)$), obbligando così il valore costante di f^+ ed f^- ad essere uguale nei due quadrati R_1 ed R_2 (nei rispettivi domini chiaramente) e non solo all'interno dei due quadrati. Ricordiamo infatti che un quadrato *connettente* R è caratterizzato da almeno α -parte della sua misura occupata da varietà stabili lunghe in R e da almeno α -parte della sua misura occupata da varietà instabili lunghe in R : dal momento che l'intersezione tra due quadrati in $\mathcal{G}_n(c)$ occupa una misura di $1 - c$ per entrambi i quadrati ed abbiamo posto $c < \alpha$, deve necessariamente succedere che una misura bidimensionale positiva di varietà stabili (instabili) *f-tipiche* di R_1 si interseca con una misura bidimensionale positiva di varietà instabili (stabili) *f-tipiche* di R_2 ⁵. Dato che questa intersezione ha misura positiva⁶ ed avviene tra varietà *f-tipiche*, i punti appartenenti ad essa sono quasi tutti *f1-tipici* e dunque a seconda che R_1 ed R_2 siano uno di fianco all'altro o uno sopra l'altro:

$$f_{|W^s(u)(R_1)}^{+(-)} = f_{|W^u(s)(R_2)}^{-(+)} = \text{costante}.$$

Ricapitolando abbiamo ottenuto due risultati utili:

- dato un quadrato α -connettente R , $f^+|_{W^s(R)} = f^-|_{W^u(R)} = \text{costante}$
- dati due quadrati α -connettenti R_1 e R_2 adiacenti in $\mathcal{G}_n(c)$ e scelto un $c < \alpha$, $f_{|W^s(u)(R_1)}^{+(-)} = f_{|W^u(s)(R_2)}^{-(+)} = \text{costante}$.

(iv) Dimostriamo che la media temporale di f è quasi ovunque costante in \mathcal{U} .

Dopo questa osservazione possiamo procedere dimostrando che la media temporale di f è

⁴In realtà per ottenere questo risultato è sufficiente che anche uno solo tra R_1 ed R_2 sia *connettente*. Ma possiamo assumere senza perdita di generalità che lo siano entrambi.

⁵Avevamo menzionato alla fine del punto (i) il fatto che, in una varietà *f-tipica*, un quadrato α -connettente è anche α -connettente di varietà *f-tipiche*.

⁶E' chiaro che se R_1 è α_1 -connettente ed R_2 è α_2 -connettente allora la misura dell'intersezioni delle varietà lunghe ed *f-tipiche* dei due quadrati vale per lo meno $(\alpha_1 - c) \times \alpha_2$.

quasi dovunque costante nel quadrato \mathcal{U} . Ricordiamo che, per semplificare la geometria, abbiamo ridefinito come *verticale* la direzione *stabile* e *orizzontale* la direzione *instabile*. Quando si parlerà di *colonna* di quadrati si intenderà una collezione di quadrati uno sopra l'altro nella direzione *stabile* (verticale) viceversa una riga di quadrati è una collezione di quadrati che si estende nel verso della direzione *instabile* (orizzontale). Siano allora $y, z \in \mathcal{U}$ due punti *f2-tipici* con varietà *f-tipiche* $W^u(y)$ e $W^u(z)$ rispettivamente. Dal momento che per il Teorema 3.3.5 sappiamo che quasi tutti i punti di $\mathcal{U} \subseteq \mathcal{M}$ sono *f2-tipici*, ci basta dimostrare che $f^-(y) = f^+(z)$. Diciamo che $W^u(y)$ ($W^s(z)$) interseca completamente una colonna (riga) di quadrati di $\mathcal{G}_n(c)$ se è *connettente* in *anche uno solo* dei quadrati della colonna (riga). Il Teorema 3.4.2 di Sinai ci permette di affermare che, per n sufficientemente grande, $W^u(y)$ interseca completamente almeno una colonna di quadrati *connettenti* in $\mathcal{G}_n(c)$ e $W^s(z)$ interseca completamente almeno una riga di quadrati *connettenti* in $\mathcal{G}_n(c)$. Infatti supponiamo per assurdo che qualsiasi colonna di quadrati di $\mathcal{G}_n(c)$ intersecata completamente da $W^u(y)$ contenga almeno un quadrato non *connettente*. Dal momento che il numero di colonne intersecate da $W^u(y)$ (orizzontale) cresce linearmente con n (infatti la lunghezza di $W^u(y)$ è costante per $n \rightarrow +\infty$, mentre la ‘densità lineare’ di quadrati è proporzionale ad n , siccome ogni quadrato ha lato di lunghezza $\frac{1}{n}$) e la misura di un qualsiasi quadrato in $\mathcal{G}_n(c)$ è $\frac{1}{n^2}$, si otterrebbe che la misura dell’unione dei quadrati non *connettenti* è un $O(\frac{1}{n})$ per $n \rightarrow +\infty$. Si noti che i quadrati di $\mathcal{G}_n(c)$ possono sovrapporsi su parte della propria misura ma ogni punto $x \in \mathcal{G}_n(c)$ può appartenere al massimo a $k(c)$ quadrati indipendentemente da n . Se $\mathfrak{F}_n \subset \mathcal{G}_n(c)$ è l’insieme dei quadrati non *connettenti* per un certo valore di n , nell’ipotesi dell’assurdo si avrebbe:

$$Leb_2\left(\bigcup_{Q \in \mathfrak{F}_n} Q\right) \geq \frac{1}{k(c)} \cdot n \cdot Leb_2(Q) = \text{costante} \cdot \frac{1}{n} \quad \text{per } n \rightarrow +\infty.$$

In contraddizione con il teorema di Sinai secondo il quale

$$Leb_2\left(\bigcup_{Q \in \mathfrak{F}_n} Q\right) = o\left(\frac{1}{n}\right) \quad \text{per } n \rightarrow +\infty.$$

Fissiamo allora una colonna ed una riga di quadrati *connettenti* che sono intersecati completamente da $W^u(y)$ e $W^s(z)$ rispettivamente. Sia R l’unico quadrato che appartiene a entrambe la colonna e la riga. Sia inoltre R_1 il quadrato in cui $W^u(y)$ è lunga (appartenete alla colonna fissata) e R_2 il quadrato in cui $W^s(z)$ è lunga (appartenete alla riga fissata). Siccome la riga e la colonna che stiamo considerando sono formate da quadrati *connettenti*, per costruzione $y \in W^u(R_1)$ ed f^- è costante sull’unione $W^u(R_1) \cup W^u(R)$. Allo stesso modo $z \in W^s(R_2)$ ed f^+ è costante sull’unione $W^s(R_2) \cup W^s(R)$. poiché però per quanto detto sopra, R è *connettente*, allora $f^-|_{W^u(R)} = f^+|_{W^s(R)} = \text{costante}$, e dunque si ha che $f^-(y) = f^+(z)$. Considerata l’arbitrarietà della scelta dei punti *f2-tipici*, y e z , otteniamo che la media temporale di f deve essere quasi dovunque costante in \mathcal{U} .

(v) *Dimostriamo che \mathcal{U} non è decomponibile.*

Abbiamo visto che f^+ è quasi ovunque costante in \mathcal{U} , con f funzione continua arbitraria. Ma $\mathcal{U} \subset \mathcal{M}$. Ricordiamo che il sistema dinamico su cui abbiamo lavorato è $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), Leb_2, T)$ con $\mathfrak{B}(\mathcal{M})$ σ -algebra dei sottoinsiemi di \mathcal{M} misurabili secondo Lebesgue. La topologia definita su \mathcal{M} è quella generata dagli aperti canonici di \mathbb{R}^2 sottoinsiemi di \mathcal{M} . Siccome \mathcal{U} è un sottoinsieme aperto di \mathbb{R}^2 , è risaputo che l'insieme $C_0^\infty(\mathcal{U})$ delle funzioni continue, infinitamente derivabili, con tutte le derivate continue a supporto compatto, viste come elementi di $L^p(\mathcal{U}, Leb_2)$ è denso in $L^p(\mathcal{U}, Leb_2)$, $p \geq 1$. Siccome l'insieme $C(\mathcal{U})$ delle funzioni continue su \mathcal{U} è più grande di $C_0^\infty(\mathcal{U})$, $C_0^\infty(\mathcal{U}) \subset C(\mathcal{U})$, allora anche $C(\mathcal{U})$ è denso in $L^1(\mathcal{U}, Leb_2)$. Sono dunque verificate le ipotesi del Corollario 2.3.9 e abbiamo che \mathcal{U} appartiene ad un' unica componente ergodica. \square

Per il metodo di Sinai nell'ambito dei biliardi si veda [10].

Capitolo 4

Ergodicità per mappe iperboliche lineari a tratti

Gli argomenti presentati nel Capitolo precedente mostrano che la mappa definita nella Sezione (3.1) è ergodica. Ci si potrebbe aspettare allora che se il dominio non ha linee di discontinuità nella direzione stabile o instabile (affinché valga il Teorema di Sinai 3.4.2), allora una generica mappa formata da una matrice iperbolica più traslazioni rigide dei pezzi del codominio debba necessariamente essere ergodica. Consideriamo la mappa $T : \mathcal{M} \rightarrow \mathcal{M}$ rappresentata in figura 5. E' possibile pensare all'azione della mappa T sul suo dominio \mathcal{M} come l'azione separata di T sui due sottodomini \mathcal{M}_1 e \mathcal{M}_2 ,

$$\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2,$$

con \mathcal{M}_1 dato dall'unione dei poligoni rosso, blu verde e giallo ed \mathcal{M}_2 dato dall'unione dei poligoni arancione, marrone, viola e rosa. Se per un attimo non consideriamo \mathcal{M}_1 ed \mathcal{M}_2 come facenti parte di \mathcal{M} , si vede come le mappe $T|_{\mathcal{M}_1}$ e $T|_{\mathcal{M}_2}$ agiscono esattamente nello stesso modo sui rispettivi domini. Non si è fatto altro che accostare nello stesso spazio (\mathbb{R}^2) due sistemi dinamici identici, che in un primo momento potevano essere pensati come appartenenti a spazi diversi, e *'sincronizzarne'* l'evoluzione temporale, di modo che l'effetto globale sia quello di una unica mappa T .

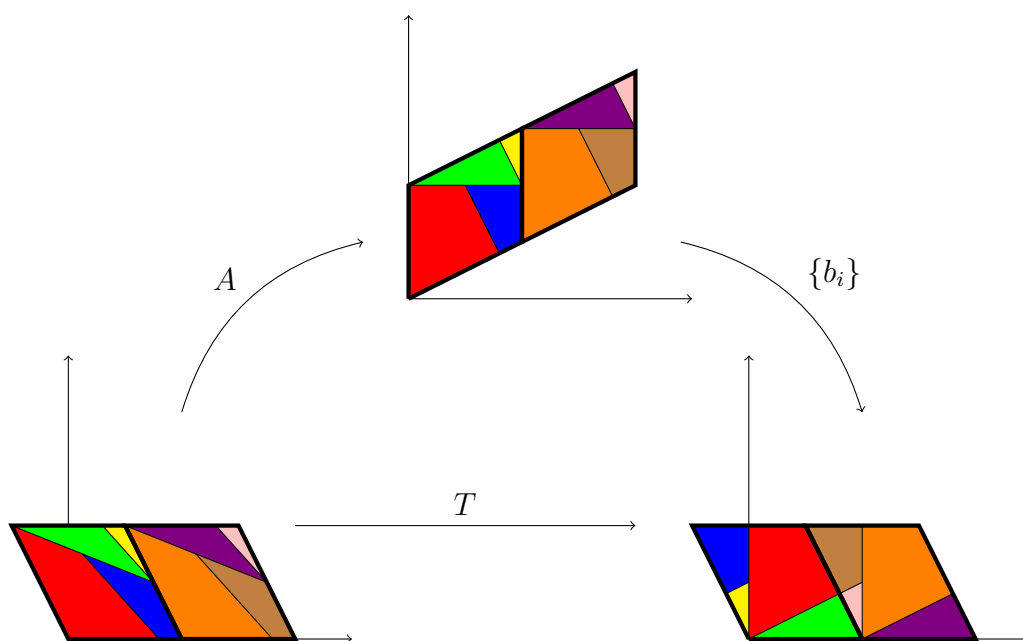


Fig 5: una estensione della mappa del Capitolo precedente, ottenuta affiancando due insiemi identici.

Chiaramente la mappa T non è ergodica in quanto è per costruzione decomponibile. Infatti, osservando i colori, si vede dalla figura che

$$\begin{aligned} T(\mathcal{M}_1) &= \mathcal{M}_1, \\ T(\mathcal{M}_2) &= \mathcal{M}_2, \\ \text{Leb}_2(\mathcal{M}_1) &= \text{Leb}_2(\mathcal{M}_2) = \frac{1}{2} > 0. \end{aligned}$$

Abbiamo dunque costruito un semplice esempio di mappa *simile* a quella presentata nella Sezione 3.1, tuttavia non ergodica. Nel presente Capitolo ci proponiamo di studiare le condizioni in cui un sistema dinamico di questo genere è ergodico. Il primo passo sarà allora quello di definire più precisamente la classe di mappe con cui lavoreremo.

4.1 Definizione della mappa

Sia \mathcal{M} un poligono in \mathbb{R}^2 su cui è definita una mappa affine $T : \mathcal{M} \rightarrow \mathcal{M}$. Per definire l'azione di T su \mathcal{M} definiamo due partizioni di \mathcal{M} ,

$$\mathcal{P}^+ = \{\mathcal{M}_i^+ | i = 1, \dots, N\} \tag{4.1.1}$$

$$\mathcal{P}^- = \{\mathcal{M}_i^- | i = 1, \dots, N\}, \quad (4.1.2)$$

i cui elementi sono poligoni e tali che su ogni sotto-dominio \mathcal{M}_i^+ la mappa T agisca nella seguente maniera:

$$T(\mathcal{M}_i^+) = \mathcal{M}_i^- \quad (4.1.3)$$

$$T|_{\mathcal{M}_i^+}(x) = Ax + b_i \quad (4.1.4)$$

dove $b_i \in \mathbb{R}^2$ ed $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ è una matrice iperbolica, cioè ha determinate unitario e due autovalori λ_1, λ_2 tali che $\lambda_1 < 1$ e $\lambda_2 = \frac{1}{\lambda_1} > 1$. Grazie alla (4.1.3) possiamo scrivere:

$$T(\mathcal{P}^+) = \mathcal{P}^-,$$

intendendo con questa notazione che $T(\mathcal{P}^+) = \{T\mathcal{M}_i^+ | i = 1, 2, \dots, N\}$. Abbiamo dunque che per definire la mappa T è sufficiente definire una qualsiasi tra le partizioni \mathcal{P}^+ o \mathcal{P}^- , dal momento che l'altra è determinata dalla relazione (4.1.3). Si ricordi infatti che T è invertibile. Tuttavia, assegnata una matrice iperbolica A , risulta abbastanza complicato determinare *direttamente* una partizione di \mathcal{P}^+ (\mathcal{P}^-) e un insieme di vettori $\{b_i\}$ tali che la partizione $T(\mathcal{P}^+) = \mathcal{P}^-$ ($T(\mathcal{P}^-) = \mathcal{P}^+$) continui *effettivamente* ad essere una partizione di \mathcal{M} . In analogia con quanto fatto nel Capitolo 3, è più semplice definire una partizione di $A(\mathcal{M})$ (in cui A è la matrice iperbolica nella definizione della mappa T) che indicheremo con $A(\mathcal{P}^+)$,

$$A(\mathcal{P}^+) = \{\mathcal{N}_{i=1} | i = 1, 2, \dots, N\} \quad (4.1.5)$$

$$A(\mathcal{M}) = \bigcup_{i=1}^N \mathcal{N}_i, \quad (4.1.6)$$

e un insieme di vettori $\{b_i\}$ tali che

$$\mathcal{M}_i^- = \mathcal{N}_i + b_i \quad \mathcal{M}_i^+ = A^{-1}(\mathcal{N}_i). \quad (4.1.7)$$

Una volta fissata $A(\mathcal{P}^+)$ è poi possibile riottenere le partizioni \mathcal{P}^+ e \mathcal{P}^- grazie alle equazioni (4.1.7). Si noti che se i poligoni \mathcal{N}_i formano una partizione di $A(\mathcal{M})$, allora grazie alla (4.1.7) gli insiemi \mathcal{M}_i^+ formano trivialmente una partizione di \mathcal{M} . E' proprio questo che rende preferibile partire da una partizione di $A(\mathcal{M})$. Verifichiamo inoltre che la definizione (4.1.7) di $A(\mathcal{P}^+)$ sia coerente con quella della mappa T , cioè valga la (4.1.3):

$$T(\mathcal{M}_i^+) = A(A^{-1}(\mathcal{N}_i)) + b_i = \mathcal{N}_i + b_i = \mathcal{M}_i^-. \quad (4.1.8)$$

Ricapitolando abbiamo che per identificare completamente la mappa T definita dalle (4.1.1), (4.1.2), (4.1.3) e (4.1.4) è sufficiente fornire una tra le partizioni \mathcal{P}^+ , \mathcal{P}^- e $A(\mathcal{P}^+)$, essendo le altre due definite dalle relazioni (4.1.3) e (4.1.7).

Se dunque immaginiamo \mathcal{M} spezzettato negli N sottodomini \mathcal{M}_i^+ abbiamo che T agisce su ognuno di questi come una matrice seguita da una traslazione b_i che dipende in generale da quale sottodominio sto considerando, in modo che l'effetto finale sia quello di ottenere nuovamente l'insieme di partenza. Pensando ad \mathcal{M} come formato da un 'puzzle' abbiamo che ognuno dei pezzi del puzzle viene 'stirato' dalla mappa iperbolica A e disposto tra gli altri pezzi stiracchiati in modo tale da ricreare l'immagine di partenza. In analogia con quanto fatto nel capitolo precedente, definiamo anche l'inversa $T^{-1} : \mathcal{M} \rightarrow \mathcal{M}$,

$$T^{-1}|_{\mathcal{M}_i^-}(x) = A^{-1}x - A^{-1}b_i$$

tale che

$$T^{-1}(\mathcal{M}_i^-) = \mathcal{M}_i^+.$$

4.2 Il teorema di ergodicità locale

$\forall i = 1, \dots, N$ chiamiamo $\partial\mathcal{M}_i^+$ il bordo di \mathcal{M}_i^+ . Generalizzando quanto scritto nel capitolo precedente, definiamo \mathcal{S}^+ come l'unione di tutti i bordi dei pezzi di cui è formato \mathcal{M} :

$$\mathcal{S}^+ = \bigcup_{i=1}^N \partial\mathcal{M}_i^+. \quad (4.2.1)$$

Sia poi \mathcal{S}_K^+ l'unione di \mathcal{S}^+ con le sue prime $K - 1$ retroimmagini:

$$\mathcal{S}_K^+ = \bigcup_{i=0}^{K-1} T^{-i}\mathcal{S}^+, \quad (4.2.2)$$

ed allo stesso modo per \mathcal{S}^- e \mathcal{S}_K^- si ha:

$$\mathcal{S}^- = \bigcup_{i=1}^N \partial\mathcal{M}_i^-, \quad (4.2.3)$$

$$\mathcal{S}_K^- = \bigcup_{i=0}^{K-1} T^i\mathcal{S}^-. \quad (4.2.4)$$

Se $K = 0$ poniamo per definizione $\mathcal{S}_K^+ = \mathcal{S}_K^- = \emptyset$. Abbiamo allora che $\mathcal{S}_K^+ \cup \mathcal{S}_K^-$ costituisce una sorta di reticolo che spezzetta l'insieme \mathcal{M} . Indichiamo con Z_i i più piccoli sottoinsiemi di \mathcal{M} il cui bordo appartiene a $\mathcal{S}_K^+ \cup \mathcal{S}_K^-$, abbiamo:

$$Z_\alpha = \mathcal{M}_{j_0}^+ \cap T^{-1}\mathcal{M}_{j_1}^+ \cap \dots \cap T^{-K}\mathcal{M}_{j_{K-1}}^+ \cap \mathcal{M}_{i_0}^- \cap T^1\mathcal{M}_{i_1}^- \cap \dots \cap T^K\mathcal{M}_{i_{K-1}}^- \quad (4.2.5)$$

con $j_0, j_1, \dots, j_{K-1} \in \{1, 2, \dots, N\}$ e $i_0, i_1, \dots, i_{K-1} \in \{1, 2, \dots, N\}$. Tali insiemi Z_α definiscono una partizione \mathcal{P}_K di \mathcal{M} :

$$\mathcal{P}_K = \{Z_\alpha\}_{\alpha \in \{1, 2, \dots, N\}^{2K}}, \quad (4.2.6)$$

dove α denota un (multi)indice del tipo $(j_0, j_1, \dots, j_{K-1}, i_0, i_1, \dots, i_{K-1}) \in \{1, 2, \dots, N\}^{2K}$. Possiamo poi figurarci il reticolo $\mathcal{S}_K^+ \cup \mathcal{S}_K^-$ come l'unione dei bordi dei vari Z_α :

$$\mathcal{S}_K^+ \cup \mathcal{S}_K^- = \bigcup_{\alpha} \partial Z_\alpha. \quad (4.2.7)$$

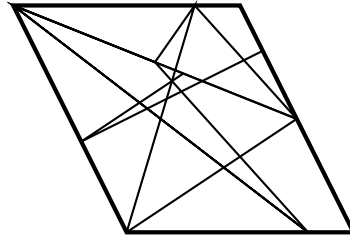


Fig. 6: possibile rappresentazione della partizione \mathcal{P}_K : i poligoni in cui è suddiviso \mathcal{M} sono gli insiemi Z_α . L'insieme dei tagli compreso il bordo di \mathcal{M} costituisce $\mathcal{S}_K^+ \cup \mathcal{S}_K^- = \bigcup_{\alpha} \partial Z_\alpha$.

Il numero degli elementi Z_α non vuoti di \mathcal{P}_K è una qualche funzione anche molto complicata f di K , $i \in \{1, 2, \dots, f(K)\}$ e dipende in generale anche dalla forma di \mathcal{M} e dalla sua partizione \mathcal{P}^+ .

Nel Capitolo 3 abbiamo visto come dimostrare l'ergodicità della mappa nell'esempio di Liverani e Wojtkowski. Questo è stato possibile grazie al fondamentale contributo del Teorema di Sinai 3.4.2. Abbiamo già menzionato nell'Osservazione 3.4.3 che tale teorema, così scritto, vale solo per il particolare esempio di Liverani e Wojtkowski¹ e, siccome adesso studiamo il caso di una *generica mappa lineare iperbolica*, è necessaria una riformulazione del Teorema di Sinai 3.4.2. Sia $\mathcal{U} \subset \mathcal{M}$, un quadrato sufficientemente piccolo da essere compreso in \mathcal{M} . Costruiamo l'insieme di ricoprimenti $\mathcal{G}_n(c)$ di \mathcal{U} esattamente come si è fatto per la mappa di Liverani e Wojtkowski (per il lettore che non avesse letto i capitoli precedenti rimandiamo alla Sezione 3.4 del Capitolo 3 per la costruzione di $\mathcal{G}_n(c)$). Siano dati i seguenti settori dello spazio tangente di ogni punto $x \in \mathcal{M}$ (\mathbb{R}^2):

$$\mathfrak{C}_+ = \{(\xi, \eta) \in \mathbb{R}^2 \mid |\xi| \leq \chi|\eta|\}$$

$$\mathfrak{C}_- = \{(\xi, \eta) \in \mathbb{R}^2 \mid |\eta| \leq \chi|\xi|\},$$

dove abbiamo scelto le coordinate dello spazio tangente dei punti di \mathcal{M} in modo che la retta $\eta = 0$ individui la direzione instabile mentre la retta $\xi = 0$ individui quella stabile

¹La ragione di questo fatto sarà spiegata nella Appendice B.

e $\chi \in \mathbb{R}$ è un numero fissato, con però $0 < \chi < 1$. Si noti che così facendo il cono \mathfrak{C}_+ è simmetrico rispetto alla direzione stabile (che pensiamo come verticale) ed il cono \mathfrak{C}_- rispetto a quella instabile (che pensiamo come orizzontale). Supponiamo che, per un certo $K = K^* \geq 0$, le direzioni delle linee di singolarità contenute in $\mathcal{S}_\infty^+ \setminus \mathcal{S}_{K^*}^+$ siano tutte contenute nel settore \mathfrak{C}_+ e, viceversa, supponiamo che per lo stesso K^* di sopra le linee di singolarità contenute nell'insieme $\mathcal{S}_\infty^- \setminus \mathcal{S}_{K^*}^-$ abbiano tutte direzione contenuta nel settore \mathfrak{C}_- ². Vale allora il seguente:

Teorema 4.2.1. *(di Sinai) Sia $K = K^*$. Se \mathcal{U} non interseca nessuna linea di singolarità di \mathcal{S}^+ e delle sue prime $K - 1$ iterazioni nel passato né interseca nessuna linea di \mathcal{S}^- e delle sue prime $K - 1$ iterazioni nel futuro*

$$\mathcal{U} \cap (\mathcal{S}_K^+ \cup \mathcal{S}_K^-) = \emptyset, \quad (4.2.8)$$

allora esiste $\alpha_0 < 1$ tale che per ogni α , $0 < \alpha \leq \alpha_0$ ed ogni c , $0 < c < 1$,

$$\lim_{n \rightarrow +\infty} n \cdot \text{Leb}_2(\bigcup \{R \in \mathcal{G}_n(c) \mid R \text{ non è } \alpha\text{-connettente}\}) = 0$$

In altre parole se α è sufficientemente piccolo, allora l'unione di quadrati di $\mathcal{G}_n(c)$ che è non α -connettente ha misura $o(\frac{1}{n})$ per $n \rightarrow \infty$.

Dimostrazione. La dimostrazione di questo teorema è riportata in Appendice B. \square

Se dunque consideriamo un quadrato \mathcal{U} che non ha intersezioni con $\mathcal{S}_K^+ \cup \mathcal{S}_K^-$ vale il Teorema di Sinai. E' chiaro allora che possiamo ripetere *esattamente alla lettera* le argomentazioni esposte nella dimostrazione del teorema 3.4.4 ed affermare:

Teorema 4.2.2. *Il quadrato \mathcal{U} per cui valga l'equazione 4.2.8 appartiene a una sola componente ergodica.*

Dimostrazione. Come sopra accennato, si può ripetere alla lettera la dimostrazione del teorema 3.4.4 nel Capitolo 3. \square

Per l'arbitrarietà del quadrato \mathcal{U} abbiamo che un qualsiasi insieme individuato dal reticolo $\mathcal{S}_K^+ \cup \mathcal{S}_K^-$ appartiene alla stessa componente ergodica. Quello che abbiamo ottenuto può sostanzialmente essere riassunto nel seguente teorema di *ergodicità locale*.

²Per essere veramente precisi si sarebbe dovuto scrivere $\mathcal{S}_\infty^+ \setminus \mathcal{S}_{K^*}^+ \setminus \partial M$ e $\mathcal{S}_\infty^- \setminus \mathcal{S}_{K^*}^- \setminus \partial M$ invece che $\mathcal{S}_\infty^+ \setminus \mathcal{S}_{K^*}^+$ e $\mathcal{S}_\infty^- \setminus \mathcal{S}_{K^*}^-$. Infatti nel caso (come quello nell'esempio di Liverani e Wojtkowski) in cui $K^* = 0$, tutte le linee di discontinuità sono contenute nei settori \mathfrak{C}_+ e \mathfrak{C}_- , eccetto quelle di ∂M . Ma è chiaro che le linee di bordo del dominio complessivo di una mappa sono in un certo senso *naturalmente* linee di divisione tra componenti ergodiche e quindi non devono essere considerate.

Teorema 4.2.3. *(di Ergodicità Locale) Siano \mathcal{M} e T definite come sopra (Sezione 4.1). Allora, se per ogni $i \in \{1, 2, \dots, N\}$ $\partial\mathcal{M}_i^+$ non ha lati diretti lungo la direzione stabile né quella instabile, esiste un numero naturale $K \in \mathbb{N}$ sufficientemente grande tale per cui la suddivisione in componenti ergodiche di \mathcal{M} non può essere più fine di quella data dalla partizione \mathcal{P}_K .*

Dimostrazione. Grazie al teorema 4.2.2 rimane solo da dimostrare che per la classe di mappe definite nella Sezione 4.1, esista effettivamente un $K \geq 0$ per cui le linee di singolarità di $\mathcal{S}_\infty^+ \setminus \mathcal{S}_K^+$ appartengano a \mathfrak{C}_+ e quelle di $\mathcal{S}_\infty^- \setminus \mathcal{S}_K^-$ appartengano a \mathfrak{C}_- . Tuttavia si può facilmente verificare che l'effetto rilevante della mappa T su un segmento è quello di avvicinarlo alla direzione instabile, tranne che per una linea diretta lungo la direzione stabile che non cambia inclinazione. Viceversa l'effetto della mappa T^{-1} su un segmento è quello di avvicinarlo alla direzione stabile, tranne che per una linea diretta lungo la direzione instabile che non cambia inclinazione³. Sia \mathcal{M} suddiviso nella partizione \mathcal{P}^+ , $T : \mathcal{M} \rightarrow \mathcal{M}$ definita dalla (4.1.4), \mathcal{S}^+ ed \mathcal{S}^- definiti dalle (4.2.1) e (4.2.3). \mathcal{S}^+ ed \mathcal{S}^- sono composti da un numero finito di segmenti *dei quali per ipotesi nessuno è disposto lungo la direzione stabile né instabile*. Se nessuno tra i segmenti che compongono \mathcal{S}^+ è parallelo alla direzione instabile, deve esistere, per quanto detto sopra, un numero naturale K_1 tale che tutti i segmenti in $\bigcup_{i=K_1}^{+\infty} T^{-i}\mathcal{S}^+$ hanno direzione contenuta nel settore \mathfrak{C}_+ . D'altra parte se nessuno tra i segmenti che compongono \mathcal{S}^- è parallelo alla direzione stabile, deve esistere un numero naturale K_2 tale che tutti i segmenti in $\bigcup_{i=K_2}^{+\infty} T^i\mathcal{S}^-$ hanno direzione contenuta in un settore \mathfrak{C}_- . Scegliendo $K = \max\{K_1, K_2\}$ abbiamo che tutte le linee di

$$\mathcal{S}_\infty^+ \setminus \mathcal{S}_K^+ = \bigcup_{i=K}^{+\infty} T^{-i}\mathcal{S}^+$$

$$\mathcal{S}_\infty^- \setminus \mathcal{S}_K^- = \bigcup_{i=K}^{+\infty} T^i\mathcal{S}^-$$

hanno direzione contenuta in \mathfrak{C}_+ e \mathfrak{C}_- , rispettivamente. Ma per il teorema 4.2.2 abbiamo che un quadrato $\mathcal{U} \subset \mathcal{M}$ appartiene ad un'unica componente ergodica se non interseca

³Infatti, siccome le traslazioni $\{b_i\}$ non cambiano la direzione dei segmenti, l'effetto di 'rotazione' di questi ultimi è dovuto alla sola matrice A . Per capire come cambia la direzione di una linea a seguito di applicazioni di A , si può studiare l'effetto di A su un vettore di \mathbb{R}^2 . Sia $\bar{v} \in \mathbb{R}^2$, abbiamo che \bar{v} può essere scritto nella base ortonormale individuata dalla direzione stabile e instabile di A , $\bar{v} = (v_{in}, v_s)$. In questa base, A è per definizione diagonale ed è facile vedere come cambiano le componenti di \bar{v} per applicazioni ripetute di A . La componente lungo la direzione instabile aumenta di un fattore $\lambda_2 > 1$ e quella lungo la direzione stabile diminuisce di un fattore $\lambda_1 < 1$ per ogni applicazione di A , $A\bar{v} = (\lambda_2 v_{in}, \lambda_1 v_s)$. Se applico A ad un vettore con sola componente sulla direzione stabile, questo sarà ridotto di un fattore $\lambda_1 < 1$ per ogni applicazione di A ma non cambierà direzione in quanto la componente sulla direzione instabile rimane $0 = \lambda_2 \cdot 0$ a seguito di ogni applicazione, $A\bar{v} = (0, \lambda_1 v_s)$. Un discorso analogo ma invertito vale per A^{-1} .

linee appartenenti a \mathcal{S}_∞^+ e \mathcal{S}_∞^- che non appartengono a \mathfrak{C}_- e \mathfrak{C}_+ , rispettivamente. Fissato una volta per tutte il parametro χ che definisce \mathfrak{C}_+ e \mathfrak{C}_- in modo che valga il teorema 4.2.1, abbiamo però dimostrato che esiste $K \geq 0$ tale che tutte le linee di $\mathcal{S}_\infty^+ \setminus \mathcal{S}_K^+$ sono comprese in \mathfrak{C}_+ e tutte le linee di $\mathcal{S}_\infty^- \setminus \mathcal{S}_K^-$ sono comprese in \mathfrak{C}_- . Dunque la suddivisione in componenti ergodiche è al più fine come quella data dagli aperti connessi di $\mathcal{M} \setminus (\mathcal{S}_K^+ \cup \mathcal{S}_K^-)$, con \mathcal{S}_K^+ ed \mathcal{S}_K^- definiti dalle (4.2.2) e (4.2.4), che costituiscono gli insiemi che formano la partizione \mathcal{P}_K . \square

In altre parole, se esiste un sottoinsieme $E \subset \mathcal{M}$, invariante rispetto a T allora esso deve essere unione degli insiemi appartenenti alla partizione \mathcal{P}_K (con K individuato dal Teorema di Ergodicità Locale) e

$$\partial E \subset \mathcal{S}_K^+ \cup \mathcal{S}_K^-. \quad (4.2.9)$$

Inoltre i segmenti appartenenti a $\bigcup_{i=K}^{+\infty} T^{-i}\mathcal{S}^+ \cup \bigcup_{i=K}^{+\infty} T^i\mathcal{S}^-$ possono avere intersezioni con ∂E solo in punti isolati:

$$\partial E \cap \left(\bigcup_{i=K}^{+\infty} T^{-i}\mathcal{S}^+ \cup \bigcup_{i=K}^{+\infty} T^i\mathcal{S}^- \right) = \emptyset \text{ eccetto che per punti isolati.} \quad (4.2.10)$$

4.3 Una estensione del teorema di ergodicità locale

Sappiamo dunque che la suddivisione in componenti ergodiche di \mathcal{M} non può essere più fine di quella data dalla partizione in aperti connessi di $\mathcal{M} \setminus (\mathcal{S}_K^- \cup \mathcal{S}_K^+)$. Possiamo però spingerci oltre ed affermare quanto segue.

Teorema 4.3.1. *Se per ogni $i \in \{1, 2, \dots, N\}$ $\partial\mathcal{M}_i^+$ non ha lati diretti lungo la direzione stabile né quella instabile, se non esiste un sottoinsieme stretto di indici $\mathfrak{J} \subset \{1, 2, \dots, N\}$ tale che*

$$\bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+ = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^- \quad (4.3.1)$$

la mappa $T : \mathcal{M} \rightarrow \mathcal{M}$ è ergodica. Viceversa l'ergodicità di T implica la non esistenza di un sottoinsieme stretto di indici $\mathfrak{J} \subset \{1, 2, \dots, N\}$ tale che

$$\bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+ = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^-. \quad (4.3.2)$$

In un certo senso quanto ottenuto significa che una mappa T sull'insieme \mathcal{M} definite dalle (4.1.1), (4.1.2), (4.1.3) e (4.1.4) o è 'palesamente' non ergodica oppure è ergodica: se esiste un insieme invariante sotto l'effetto della mappa è molto facile trovarlo, perché per forza dato dall'unione di un certo numero di poligoni tra gli N che formano l'insieme \mathcal{M} secondo la (4.3.1).

Dimostrazione. (\Rightarrow)

Supponiamo per assurdo che esista un insieme invariante $E \subset \mathcal{M}$ strettamente incluso in \mathcal{M} . Siccome T è invertibile, possiamo scrivere:

$$T(E) = E, \quad (4.3.3)$$

$$T^{-1}(E) = E. \quad (4.3.4)$$

Per il Teorema di Ergodicità Locale 4.2.3 sappiamo che è possibile trovare un K arbitrariamente grande per cui la suddivisione in componenti ergodiche di \mathcal{M} non può essere più fine di quella della partizione \mathcal{P}_K . Indichiamo con Z_i gli insiemi non nulli appartenenti alla partizione \mathcal{P}_K , $i = 1, 2, \dots, f(K)$. Allora,

$$E = \bigcup_{i \in \mathfrak{H} \subset \{1, 2, \dots, f(K)\}} Z_i \quad (4.3.5)$$

per un certo sottoinsieme di indici \mathfrak{H} . Si ricordi che E è un sottoinsieme stretto di \mathcal{M} , $E \subset \mathcal{M}$, dunque la (4.3.5) non può estendersi su *tutti* gli Z_i ma solo su un sottoinsieme proprio di indici.

Dalla (4.2.6) sappiamo che gli Z_i sono poligoni in quanto l'immagine e la retroimmagine di un poligono sotto una mappa lineare è ancora un poligono e l'intersezione tra due poligoni o è un insieme di misura nulla oppure è un poligono. Inoltre gli insiemi \mathcal{M}_i^+ sono poligoni per ipotesi. Abbiamo allora dalla (4.3.5) che E è una unione finita di poligoni e per costruzione ogni lato di E giace su $T^{-i}\mathcal{S}^+$ o $T^i\mathcal{S}^-$ per un qualche $i \in \{0, 1, \dots, K-1\}$.

Indichiamo allora i segmenti che formano il bordo di E (∂E) con $\{\gamma_1, \gamma_2, \dots, \gamma_p\}$, $p \in \mathbb{N}$ non meglio precisato, e supponiamo che per un certo indice $s \in \{1, 2, \dots, p\}$, γ_s abbia una intersezione nulla con \mathcal{S}^- eccetto che per punti isolati. In questo caso T^{-1} è continua su quasi tutti i punti di γ_s e quasi tutti i punti di $T^{-1}\gamma_s$ appartengono a ∂E . In effetti se $x \in \partial E$ e x è un punto di continuità di T^{-1} , in questo caso è ovvio che $T^{-1}x \in \partial E$. Infatti se ∂E è localmente un segmento in ogni intorno di x c'è un più piccolo intorno contenuto in E e un altro disgiunto da E . Dal momento che E è un insieme invariante le immagini di questi due intorni sotto la mappa T^{-1} avranno rispettivamente una misura di punti piena dentro e fuori E . Ma per la continuità di T^{-1} in x , per ogni intorno B di $T^{-1}(x)$ è possibile trovare un intorno A di x tale che $T^{-1}A \subset B$. Segue che in ogni intorno di $T^{-1}(x)$ ci sono punti fuori e dentro E e dunque $T^{-1}(x)$ è un punto di frontiera di E . Un ragionamento analogo si può fare per i segmenti in ∂E che hanno intersezioni con \mathcal{S}^+ solo in punti isolati, ottenendo cioè:

$$T^{-1} : \partial E \setminus \mathcal{S}^- \rightarrow \partial E, \quad (4.3.6)$$

$$T : \partial E \setminus \mathcal{S}^+ \rightarrow \partial E. \quad (4.3.7)$$

Supponiamo che per un certo $s \in \{1, \dots, p\}$, $\gamma_s \subset \partial E$, $\gamma_s \subset T^{-K+1}\mathcal{S}^+$, con K sufficientemente grande perché valga la (4.2.9) e la (4.2.10). Supponiamo che γ_s intersechi \mathcal{S}^- al più in punti isolati, allora per la (4.3.6) quasi tutti i punti di $T^{-1}\gamma_s$ devono appartenere a ∂E . Avremmo però che, per definizione, $T^{-1}\gamma_s \subset T^{-K}\mathcal{S}^+$ e dunque si avrebbe che una misura unidimensionale positiva di punti di ∂E appartiene a $T^{-K}\mathcal{S}^+$ in contraddizione con la (4.2.10). Abbiamo allora che se esiste un indice s per cui γ_s appartiene a $T^{-K+1}\mathcal{S}^+$ allora γ_s deve appartenere anche a \mathcal{S}^- . Viceversa se γ_s interseca \mathcal{S}^+ al più in punti isolati, allora per la (4.3.7) quasi tutti i punti di $T\gamma_s$ devono appartenere a ∂E . Avremmo però che, per definizione, $T\gamma_s \subset T^K\mathcal{S}^-$ e dunque si avrebbe che una misura unidimensionale positiva di punti di ∂E appartiene a $T^K\mathcal{S}^-$ in contraddizione con la (4.2.10). Abbiamo allora che se esiste un indice s per cui γ_s appartiene a $T^{K-1}\mathcal{S}^-$ allora γ_s deve appartenere anche a \mathcal{S}^+ .

Si è trovato allora che tutti i punti del bordo dell'insieme invariante $E \subset \mathcal{M}$ che appartengono a $T^{-K+1}\mathcal{S}^+ \cup T^{K-1}\mathcal{S}^-$ appartengono anche a $\mathcal{S}_+ \cup \mathcal{S}_-$ e dunque

$$\partial E \subset \mathcal{S}_{K-1}^+ \cup \mathcal{S}_{K-1}^- \quad (4.3.8)$$

Possiamo applicare lo stesso ragionamento di sopra per dimostrare che $\partial E \subset \mathcal{S}_{K-2}^+ \cup \mathcal{S}_{K-2}^-$, ottenendo così dopo $K - 1$ passaggi che

$$\partial E \subset \mathcal{S}^+ \cup \mathcal{S}^- \quad (4.3.9)$$

Proseguendo con la dimostrazione, vediamo che se $\gamma_s \subset \mathcal{S}^+$ e γ_s ha intersezioni con \mathcal{S}^- solo in punti isolati allora T^{-1} è continua su quasi ogni punto di γ_s e dunque per la (4.3.6) $T^{-1}\gamma_s \subset \partial E$ con $T^{-1}\gamma_s \subset T^{-1}\mathcal{S}^+$, in disaccordo con la (4.3.9). Viceversa se $\gamma_s \subset \mathcal{S}^-$ e γ_s interseca \mathcal{S}^+ solo in punti isolati allora T è continua quasi ovunque su γ_s e dunque, per la (4.3.6), $T\gamma_s \subset \partial E$ con $T\gamma_s \subset T\mathcal{S}^-$, impossibile per la (4.3.9). Dunque si ha che

$$(a) \quad \gamma_s \subset \mathcal{S}^+ \longrightarrow \gamma_s \subset \mathcal{S}^-$$

$$(b) \quad \gamma_s \subset \mathcal{S}^- \longrightarrow \gamma_s \subset \mathcal{S}^+$$

per ogni segmento $\gamma_s \subset \partial E$. Si ottiene infine grazie alla (4.3.9):

$$\partial E \subset \mathcal{S}^+ \cap \mathcal{S}^- \quad (4.3.10)$$

Se però $\partial E \subset \mathcal{S}^+$ allora dalla (4.2.1) abbiamo che

$$E = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+, \quad (4.3.11)$$

con $\mathfrak{J} \subset \{1, 2, \dots, N\}$, dove l'inclusione stretta è importante perché E deve essere un sottoinsieme stretto di \mathcal{M} , $E \subset \mathcal{M}$. Se invece $\partial E \subset \mathcal{S}^-$ allora dalla (4.2.3) abbiamo che

$$E = \bigcup_{i \in \mathfrak{K}} \mathcal{M}_i^-, \quad (4.3.12)$$

con $\mathfrak{K} \subset \{1, 2, \dots, N\}$ in generale diverso da \mathfrak{J} .

Dalla (4.1.8), (4.3.3), (4.3.11) e (4.3.12) si trova che

$$TE = T\left(\bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+\right) = \bigcup_{i \in \mathfrak{J}} T(\mathcal{M}_i^+) = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^- = E = \bigcup_{i \in \mathfrak{K}} \mathcal{M}_i^-, \quad (4.3.13)$$

e dunque $\mathfrak{K} = \mathfrak{J}$. Abbiamo dimostrato la validità della (4.3.1).

(\Leftarrow)

Se esiste un sottoinsieme di indici $J \subset \{1, 2, \dots, N\}$ tale che vale la (4.3.2), allora sia

$$E = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+. \quad (4.3.14)$$

Abbiamo che

$$TE = T\left(\bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^+\right) = \bigcup_{i \in \mathfrak{J}} T(\mathcal{M}_i^+) = \bigcup_{i \in \mathfrak{J}} \mathcal{M}_i^- = E, \quad (4.3.15)$$

per la (4.3.2). Questo conclude la dimostrazione. \square

Osservazione 4.3.2. La (4.3.10) è molto importante perché essa costituisce la condizione necessaria e sufficiente affinché un insieme di misura positiva E sia invariante sotto l'azione della mappa T . Questo ci suggerisce una procedura generale per vedere se la mappa T è ergodica sul suo dominio \mathcal{M} :

- costruisco \mathcal{S}^+ evidenziando i bordi dei poligoni che formano \mathcal{M}
- costruisco \mathcal{S}^- facendo agire la mappa T su tali poligoni ed evidenziando i nuovi bordi
- controllo che nessuna parte di $\mathcal{S}^+ \cap \mathcal{S}^-$ possa essere il bordo di un insieme $E \subset \mathcal{M}$ di misura positiva. Per farlo basta vedere se su tale intersezione si possa costruire un qualsiasi percorso *chiuso*, diverso da quello che forma il bordo di \mathcal{M} . Se vi fossero uno o più percorsi chiusi allora il sistema è decomponibile in insiemi invarianti che hanno bordi appartenenti a $\mathcal{S}^+ \cap \mathcal{S}^-$
- posso riapplicare i punti precedenti per ogni componente così trovata fino a quando non vi sono sottoinsiemi invarianti ed ho scomposto \mathcal{M} nelle sue componenti ergodiche.

Se riguardiamo la figura 5 a inizio Capitolo, possiamo in effetti notare che $\mathcal{S}^+ \cap \mathcal{S}^-$ non è vuota ma anzi costituisce il bordo degli insiemi invarianti sotto la mappa T , \mathcal{M}_1 ed \mathcal{M}_2 . L'insieme $\mathcal{S}^+ \cap \mathcal{S}^-$ è evidenziato in nero in figura 5. Siccome tali insiemi non contengono a loro volta linee appartenenti ad $\mathcal{S}^+ \cap \mathcal{S}^-$ (eccetto che per il loro stesso bordo) possiamo affermare che costituiscono due componenti ergodiche distinte e non ulteriormente decomponibili⁴ dell'insieme \mathcal{M} . Per costruire $\mathcal{S}^+ \cap \mathcal{S}^-$ basta considerare le due partizioni del dominio in basso a destra e in basso a sinistra, sovrapponendo queste due immagini si vede come le uniche linee di bordo dei poligoni che formano \mathcal{M} comuni ad entrambe \mathcal{P}^+ e \mathcal{P}^- sono quelle evidenziate in grassetto. Sempre considerando la figura 5, abbiamo che \mathcal{M}_1 è dato dai poligoni rosso, verde, giallo e blu e \mathcal{M}_2 dai poligoni arancione, marrone, viola e rosa. Il fatto che $\partial\mathcal{M}_1$ e $\partial\mathcal{M}_2$ siano un sottoinsieme di $\mathcal{S}^+ \cap \mathcal{S}^-$ significa che sono verificate le condizioni del Teorema 4.3.1 per la non ergodicità. Infatti i due gruppi di poligoni di \mathcal{M} che formano \mathcal{M}_1 ed \mathcal{M}_2 vengono mandati in se stessi per effetto della mappa T . In particolare se etichettiamo con i colori i poligoni che formano \mathcal{M} si vede subito che:

$$\begin{aligned} \mathcal{M} &= \bigcup_{i=\text{rosso,blu,verde,giallo,arancione,marrone,viola,rosa}} \mathcal{M}_i^{+(-)}, \\ \mathcal{M}_1 &= \bigcup_{i=\text{rosso,blu,verde,giallo}} \mathcal{M}_i^+ = \bigcup_{i=\text{rosso,blu,verde,giallo}} \mathcal{M}_i^- = T\mathcal{M}_1, \\ \mathcal{M}_2 &= \bigcup_{i=\text{arancione,marrone,viola,rosa}} \mathcal{M}_i^+ = \bigcup_{i=\text{arancione,marrone,viola,rosa}} \mathcal{M}_i^- = T\mathcal{M}_2. \end{aligned}$$

Le condizioni per l'ergodicità del Teorema 4.3.1 sono sicuramente rispettate se nessuna linea di bordo dei poligoni che formano \mathcal{M} viene mappata in un'altra linea di bordo che non appartenga a $\partial\mathcal{M}$.

Corollario 4.3.3. *Se*

$$\mathcal{S}^+ \cap \mathcal{S}^- = \partial\mathcal{M}, \quad (4.3.16)$$

allora T è ergodica.

Dimostrazione. L'unico circuito chiuso in $\mathcal{S}^+ \cap \mathcal{S}^-$ è proprio $\partial\mathcal{M}$ dunque \mathcal{M} non è decomponibile per l'Osservazione 4.3.2 sopra. \square

La condizione (4.3.16) significa che nessuna linea di \mathcal{S}^+ viene mappata in $\mathcal{S}^+ \setminus \partial\mathcal{M}$. Essa è solo sufficiente ma non necessaria per l'ergodicità. Tuttavia può essere più facile da verificare rispetto alle ipotesi del Teorema 4.3.1.

⁴La non ulteriore decomponibilità di \mathcal{M}_1 e \mathcal{M}_2 era comunque già chiara già dai risultati ottenuti nel Capitolo 3.

Appendice A

Lemma di Borel-Cantelli

Sia $(\mathcal{M}, \mathfrak{B}(\mathcal{M}), \mu)$ uno spazio di misura e $\{X_n\}_{n \in \mathbb{N}}$ una successione di sottoinsiemi misurabili di \mathcal{M} . Definiamo l'insieme di eventi che si verificano infinitamente spesso $\mathfrak{S}(\{X_n\}_{n \in \mathbb{N}})$ come:

$$\mathfrak{S}(\{X_n\}_{n \in \mathbb{N}}) = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} X_n.$$

Allora:

$$\sum_{n=0}^{\infty} \mu(X_n) < \infty \implies \mu(\mathfrak{S}(\{X_n\}_{n \in \mathbb{N}})) = 0.$$

Dimostrazione. Si ha:

$$\mu(\mathfrak{S}(\{X_n\}_{n \in \mathbb{N}})) = \mu\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} X_n\right) = \lim_{N \rightarrow \infty} \mu\left(\bigcup_{n=N}^{\infty} X_n\right),$$

dove l'ultimo passaggio è permesso perché gli insiemi $A_N = \bigcup_{n=N}^{\infty} X_n$ sono decrescenti in N . Ma per la *subadditività* della misura μ :

$$\lim_{N \rightarrow \infty} \mu\left(\bigcup_{n=N}^{\infty} X_n\right) \leq \lim_{N \rightarrow \infty} \sum_{n=N}^{\infty} \mu(X_n).$$

Siccome la successione $S_M = \sum_{n=0}^M \mu(X_n)$ è convergente allora è anche di Cauchy e dunque $\forall \epsilon > 0$, esiste $M^* = M^*(\epsilon)$ tale che $|S_P - S_Q| < \epsilon$, $\forall P, Q \geq M^*(\epsilon)$. Segue immediatamente che il resto di una serie convergente può essere reso piccolo a piacere e dunque

$$\mu(\mathfrak{S}(\{X_n\}_{n \in \mathbb{N}})) = 0.$$

□

Appendice B

Dimostrazione del teorema di Sinai

La dimostrazione del *Teorema di Sinai* non richiede una costruzione geometrica rigida per i ricoprimenti $\mathcal{G}_n(c)$; essa continua a valere per una qualsiasi sequenza di ricoprimenti di quadrati di lato che scala come $\frac{1}{n}$, *fintanto che c'è un limite uniforme al numero di quadrati che ricoprono un qualsiasi punto*. Tuttavia, la struttura a griglia dei centri dei quadrati di $\mathcal{G}_n(c)$ ci permette di lavorare con *righe e colonne* di quadrati, come si è visto nelle applicazioni del Teorema di Sinai. Senza perdita di generalità, dimostriamo il teorema nella versione delle varietà instabili, dal momento che l'argomento per la controparte stabile non cambia. Per avere il teorema valido in entrambe le versioni basterà scegliere come α_0 il minimo tra i due valori di α_0 individuati nelle due versioni:

$$\alpha_0 = \min \{ \alpha_{0stabile}, \alpha_{0instabile} \}.$$

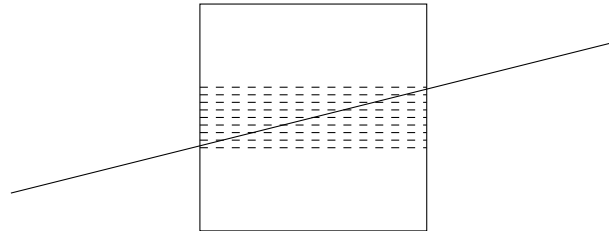
Il primo passo nella dimostrazione è proprio la scelta di α_0 . A questo proposito ricordiamo che per ipotesi tutte le linee di discontinuità per T^{-1} che possono intersecare \mathcal{U} devono appartenere a $\mathcal{S}_\infty^- \setminus \mathcal{S}_K^-$, con K scelto sufficientemente grande, in modo che tali linee abbiano direzione contenuta nel settore \mathfrak{C}_- . Poniamo allora:

$$\alpha_0 = \frac{1}{2}(1 - \chi) > 0^1.$$

La ragione di questa richiesta è che, per ogni quadrato con lati orizzontali e verticale attraversato da una linea con direzione contenuta in \mathfrak{C}_- , l'area ombreggiata in figura 7 (vedi pagina seguente) non supera le $1 - 2\alpha$ parti dell'area del quadrato, se $\alpha \leq \alpha_0$. Osserviamo che tutti i segmenti in $\bigcup_{i=K}^\infty T^i \mathcal{S}^-$ hanno direzione contenute nel settore \mathfrak{C}_- . Infatti una mappa lineare iperbolica spinge le linee verso la direzione instabile, eccetto che per una linea che sia disposta lungo la direzione stabile, la quale non cambia direzione (per una spiegazione di questo fatto, si guardi la nota 3 nel Capitolo 4). Osserviamo

¹Si ricordi che $\chi < 1$. Per definizione di \mathfrak{C}_- si veda la Sezione 4.2 Capitolo 4.

inoltre che per la stessa ragione tale numero $K \geq 0$ esiste². Dalla costruzione delle varietà stabili e instabili (Proposizione 3.2.1) si vede che una varietà instabile ha estremi su \mathcal{S}^- o sulle sue successive immagini secondo la mappa T . Dunque, se una varietà instabile è corta in $R \in \mathcal{G}_n(c)$ (ricordiamo che questo vuol dire che non tale varietà non interseca entrambi i lati appartenenti al bordo stabile di R , $\partial_s R$), allora R deve necessariamente essere intersecato da

$$\mathcal{S}_\infty^- \setminus \mathcal{S}_K^- = \bigcup_{i=K}^{\infty} T^i \mathcal{S}^-.$$


The diagram shows a square with a diagonal line running from the bottom-left to the top-right. A series of horizontal hatched lines are drawn across the square, representing the singularities $T^i \mathcal{S}^-$. The hatching is concentrated in the lower-left portion of the square, illustrating the intersection of the square with the set $\mathcal{S}_\infty^- \setminus \mathcal{S}_K^-$.

Fig. 7: in figura un elemento di $\mathcal{G}_c(n)$ tagliato da una *sola* linea di \mathcal{S}_M^- . Tale quadrato deve necessariamente essere α -connettente con la scelta di $\alpha \leq \alpha_0 = \frac{1}{2}(1 - \chi(a))$. Abbiamo infatti che esiste sicuramente una parte di quadrato non ombreggiata essendo $\chi(a)$ strettamente minore di 1 per ogni a e per ogni linea di \mathcal{S}_M^- .

Sebbene questa non sembri una restrizione troppo severa essa ha conseguenze assai importanti. La ragione è che, euristicamente, le linee di singolarità $T^i \mathcal{S}^-$ *diventano sempre più orizzontali* per $i \rightarrow \infty$ e non possono *tagliare efficacemente* le varietà instabili, che sono a loro volta orizzontali. A questo punto possiamo dividere la dimostrazione in due passi: innanzitutto mostriamo che un numero finito di linee di singolarità può produrre solo un numero sufficientemente piccolo (vedremo a breve in che senso) di quadrati non α -connettenti, e in un secondo momento ci occuperemo della misura di quadrati non α -connettenti dalla *coda infinita* di \mathcal{S}_∞^- . Più precisamente:

I) Affermiamo che, per qualsiasi $M \geq K$ fissato, le linee di singolarità

$$\mathcal{S}_M^- = \bigcup_{i=K}^M T^i \mathcal{S}^-$$

possono, da sole, produrre solo (pochi) quadrati di $\mathcal{G}_n(c)$ non α -connettenti tali che la loro misura totale è un $O(\frac{1}{n^2})$ per $n \rightarrow \infty$. Per rendere tale argomento preciso introduciamo la nozione ausiliaria di quadrato M -bad nel ricoprimento $\mathcal{G}_n(c)$.

Definizione B.0.1. Diciamo che un quadrato $R \in \mathcal{G}_n(c)$ è M -bad se la misura dell'insieme di punti $y \in R$ tali che la varietà instabile $W^u(y)$ ha un estremo in $R \cap \mathcal{S}_M^-$ (cioè tale che $W^u(y)$ è corta in R) è più grande di $1 - 2\alpha$ parti del quadrato.

²In realtà, per includere nella dimostrazione anche il caso stabile, il K 'corretto' è $K = \max\{K_{stabile}, K_{instabile}\}$.

In poche parole, un quadrato R è M -bad se non è *connettente* a causa delle linee di singolarità di \mathcal{S}_M^- . Se un quadrato R interseca solo un segmento in \mathcal{S}_M^- , allora la misura di punti in R le cui varietà instabili hanno estremi nell'intersezione di tale segmento con R (cioè la misura di punti la cui varietà instabile è stata *tagliata* da una linea in \mathcal{S}_M^-) non supera $1 - 2\alpha_0 = \chi < 1$ parti del quadrato, semplicemente perché l'(unica) linea che interseca R appartiene al settore \mathfrak{C}_- (vedi figura 7). Dunque possiamo affermare che un quadrato M -bad deve intersecare *almeno due* segmenti in \mathcal{S}_M^- . Ma l'insieme di linee di singolarità \mathcal{S}_M^- è una collezione finita di segmenti chiusi, con solo un numero finito di punti di intersezione (cioè punti appartenenti ad almeno due segmenti di \mathcal{S}_M^-). Sufficientemente lontano dalla loro intersezione, due segmenti sono così *distanti* da non poter intersecare entrambi uno stesso quadrato, se quest'ultimo è sufficientemente piccolo. Dunque, per n *sufficientemente grande*, un quadrato M -bad in $\mathcal{G}_c(n)$, non può essere più distante di $\frac{A}{n}$ da uno dei punti di intersezione, dove A è una costante che dipende da a (forma del dominio iniziale: inclinazione iniziale delle linee di singolarità), c (coefficiente che determina il *grado di sovrapposizione* dei quadrati in $\mathcal{G}_c(n)$), α (quanta percentuale di area deve avere un quadrato formata da punti con varietà stabili che non hanno un estremo in \mathcal{S}_M^- , per essere considerato *connettente*) ed M (il numero di immagini di \mathcal{S}^- nel futuro che considero)³. Dal momento che in \mathcal{S}_M^- ci sono solo un numero Q finito di punti di intersezione possiamo stimare per eccesso la misura di quadrati M -bad con la somma delle aree dei cerchi di raggio $\frac{A}{n}$ centrati nei punti di intersezione. Segue allora che la misura totale di quadrati M -bad non supera $\frac{B}{n^2}$ con $B = Q\pi A^2$. Dunque, per un qualsiasi M finito *fissato*, la misura dei quadrati M -bad è effettivamente un $O(\frac{1}{n^2})$. In questo modo ci siamo occupati del numero *finito* di linee di singolarità in \mathcal{S}_M^- .

- II) Vediamo adesso gli effetti della 'coda' $\bigcup_{i=M+1}^{\infty} T^i \mathcal{S}^-$. Supponiamo che un quadrato $R \in \mathcal{G}_c(n)$ non sia né α -connettente né M -bad. Allora, almeno α -parti della sua misura è coperta da varietà *corte* in R con estremi in

$$R \cap \bigcup_{i=M+1}^{\infty} T^i \mathcal{S}^-.$$

Sia $W^u(y)$ una varietà di questo tipo, cioè corta in R con un estremo su $T^i \mathcal{S}^-$, $i \geq M + 1$. Allora:

$$T^{-i}(W^u(y) \cap R) \subset X_{t_i},$$

dove $t_i = \frac{1}{n\lambda_2^i}$ e, come prima, $X_t = \{x \in \mathcal{M} | d(x, \mathcal{S}^-) \leq t\}$. Questo perché sotto l'azione di T^{-1} una varietà instabile si contrae di un fattore $\lambda_2 > 1$ e la lunghezza

³In ogni caso la dipendenza della costante A rispetto ai parametri in gioco non è cruciale. L'importante è che A sia costante rispetto ad n .

di $W^u(y)$ in R non supera $\frac{1}{n}$. In virtù di questa osservazione, possiamo affermare che ciascun quadrato che non è né α -connettente né M -Bad ha almeno α -parti della sua misura coperta da:

$$\bigcup_{i=M+1}^{\infty} T^i X_{t_i}.$$

Dunque, se \mathfrak{T} è l'insieme di quadrati di $\mathcal{G}_n(c)$ né α -connettenti né M -bad, si ha:

$$\frac{\alpha}{k(c)} \cdot \text{Leb}_2\left(\bigcup_{R \in \mathfrak{T}} R\right) \leq \text{Leb}_2\left(\bigcup_{i=M+1}^{\infty} T^i X_{t_i}\right) \leq \bigcup_{i=M+1}^{\infty} \text{Leb}_2(T^i X_{t_i}).$$

dove $k(c)$ è il numero massimo di quadrati di $\mathcal{G}_n(c)$ che può ricoprire un punto in \mathcal{U} . Chiaramente, essendo \mathcal{S}^- , una unione finita di linee:

$$\text{Leb}_2(T^i X_{t_i}) = \text{Leb}_2(X_{t_i}) \leq \text{costante} \cdot t_i.$$

Ma allora la misura dell'unione di quadrati di $\mathcal{G}_c(n)$ che non sono né α -connettenti né M -bad non supera:

$$k(c) \cdot \frac{1}{\alpha} \sum_{i=M+1}^{\infty} \frac{\text{costante}}{n\lambda_2^i} = \frac{1}{n} \left(\frac{k(c)}{\alpha} \sum_{i=M+1}^{\infty} \frac{\text{costante}}{\lambda_2^i} \right).^4$$

Abbiamo quindi stimato la misura di quadrati in $\mathcal{G}_c(n)$ che non sono né α -connettenti né M -bad con la lunghezza dei lati di uno di questi quadrati moltiplicato per il resto di una serie convergente (si ricordi che $\lambda_2 > 1$). Scegliamo allora un $\epsilon > 0$ arbitrario. Possiamo allora scegliere e fissare un numero naturale positivo $M^* = M(\epsilon)$ sufficientemente grande da rendere la serie sopra menzionata minore di $\frac{\epsilon}{2n}$, cioè:

$$\frac{k(c)}{\alpha} \sum_{i=M+1}^{\infty} \frac{\text{costante}}{\lambda_2^i} < \frac{\epsilon}{2}.$$

Vediamo come questo concluda la dimostrazione. Per stimare la misura dell'unione di quadrati non α -connettenti di $\mathcal{G}_c(n)$ li dividiamo nelle due classi I) e II), ovvero dei quadrati che sono M -bad e di quelli che non lo sono. Per la classe II) possiamo scegliere un $M = M^*$ sufficientemente grande tale che la misura dei quadrati non α -connettenti sia minore di $\frac{\epsilon}{2n}$, poiché tale misura è sicuramente minore di $\frac{1}{n}$ moltiplicato per la coda di una serie convergente; per la classe I) invece si ha, per quanto visto sopra, che la misura dei quadrati M -bad, con $M = M^*$ fissato sopra, è un $O(\frac{1}{n^2})$. Possiamo quindi scegliere $n_0 = n_0(\epsilon, M^*)$ tanto grande che, per ogni $n > n_0$, la misura dell'unione di quadrati M -bad in $\mathcal{G}_c(n)$ è minore di $\frac{\epsilon}{2n}$. Abbiamo allora che nel complesso la misura dei quadrati non α -connettenti è minore di $\frac{\epsilon}{n}$. Questo conclude la dimostrazione. Per una dimostrazione più rigorosa e in ambito più generale del Teorema di Sinai si guardi [5].

⁴Qui la costante è uguale alla lunghezza totale di \mathcal{S}^- .

Bibliografia

- [1] G. Gallavotti (1999) “*Statistical Mechanics - Short Treatise*”, Cap. 1 - Cap.2.
- [2] R. K. Pathria e Paul D. Beale (2011), “*Statistical Mechanics - Third edition*”, Academic Press, Cap. 1 - Cap 2.
- [3] Loris Ferrari (2015), “*Elementi di Termodinamica Statistica*”, Cap. 1 - Cap. 2 - Cap.3.
- [4] R. Soldati (2006), “*Note per il Corso di Meccanica Statistica*”, Cap. 1 - Cap 2.
- [5] Carlangelo Liverani e Maciej P. Wojtkowski (1995) “*Ergodicity in Hamiltonian Systems*”, Dynamics Reported - Exposition in Dynamical Systems, New Series vol. 4, pag. 130-145.
- [6] V.I. Arnold A. Avez (1968), “*Ergodic Problems of Classical Mechanics*”, Advanced Book Classics, Cap. 1- Cap.2.
- [7] P. Walters (1981), “*An Introduction to Ergodic Theory*”, Springer, Cap. 0 - Cap. 1.
- [8] N. Chernov e R. Markarian (2006), “*Chaotic Billiards*”, American Mathematical Society, Cap. 6 - Appendice C.
- [9] B. Hasselblatt e A. Katok (2003), “*A First Course in Dynamics*”, Cambridge, Cap. 4, Cap. 5, Cap. 6, Cap. 7.
- [10] Ya. G. Sinai (1970), “*Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*”, Uspekhi Mat. Nauk, 25:2(152) (1970), 141–192; Russian Math. Surveys, 25:2 (1970), 137–189.

Ringraziamenti

Desidero innanzitutto ringraziare il prof. Marco Lenci per tutto il tempo e la voglia che ha dedicato a questo lavoro e la bella fisica e matematica che mi ha fatto conoscere. Voglio poi ringraziare tutti coloro che mi hanno aiutato: Ema per la compagnia nelle notti di studio al *Sant'Orsola*, *SUNNY* per avermi introdotto alla grafica di Latex, Tabby per avermi chiarito alcuni dubbi sulle σ -algebre, Edo per l'aiuto informatico, mio fratello Nico per aver riletto varie volte le bozze e Lumi per avermi risolto così tanti problemi che non saprei nemmeno elencarli. Un ringraziamento anche alla mia famiglia, a tutti i *Ragazzi di Via Mascarella*, agli amici di *Che Cosa Mangio* ed a Marco, Andrea ed Isa.