

I. Introduzione alla teoria ergodica

(2001/2002)

1. Prologo: il problema ergodico in Boltzmann e Gibbs

1.1. Motivazioni

La *teoria ergodica* è un ramo della matematica, oggi molto sviluppato e a sua volta ben ramificato, il cui inizio si fa comunemente risalire ai lavori di Von Neuman e Birkhoff, verso la fine degli anni '20. Le motivazioni e alcune idee di fondo provengono tuttavia da Boltzmann e Gibbs, fondatori assieme a Maxwell della meccanica statistica, che in diverso modo introdussero la nozione fondamentale di *insieme statistico* (*ensemble*; una probabilità in un opportuno spazio delle fasi) per descrivere lo stato macroscopico di un sistema a molti gradi di libertà. Come è noto, il problema fisico da cui nasce la meccanica statistica è quello di comprendere il comportamento macroscopico di un sistema termodinamico a partire dalla sua struttura microscopica, usando come strumento essenziale il calcolo delle probabilità. L'interpretazione stretta di questo programma¹⁾ consiste dunque nel tentativo di “dedurre” in qualche modo la termodinamica dalla meccanica, o come spesso si dice, di *ridurre* l'una all'altra.

A prima vista l'impresa può apparire disperata, e anzi alcuni ben noti paradossi sembrano indicare che i due mondi della meccanica e della termodinamica siano *a priori* inconciliabili. Ricordiamo in particolare

- il *paradosso di Loschmidt*, o della reversibilità: i processi meccanici microscopici sono reversibili, ovvero per ogni processo meccanico esiste un processo inverso; i processi termodinamici appaiono invece irreversibili. Da dove nasce l'irreversibilità, e come è possibile, in queste condizioni, la riduzione?
- Il *paradosso di Zermelo*, o della ricorrenza: nei sistemi meccanici conservativi (quelli che conservano il volume nello spazio delle fasi; in particolare i sistemi hamiltoniani), con spazio delle fasi limitato, per tutti i dati iniziali a meno di un insieme di misura di Lebesgue nulla, il sistema ritorna infinite volte arbitrariamente vicino alle condizioni iniziali (Poincaré, *teorema della ricorrenza*; si veda il paragrafo 3.1). In che modo scompare la ricorrenza a livello macroscopico?

¹⁾ Fu soprattutto Boltzmann a sentire questo problema in modo acuto. Alcune idee di Boltzmann, profondissime, furono a lungo incomprese, e forse solo oggi, dopo l'ampio sviluppo che ha avuto la teoria dei sistemi dinamici, è possibile apprezzarle appieno.

La questione è in effetti sottile, e in essa gioca un ruolo importante la distinzione, mascherata nei paradossi sopra ricordati,²⁾ tra *stato meccanico* o microscopico (un punto in un opportuno spazio delle fasi) e *stato termodinamico* o macroscopico, tutto da definire, nel quale deve in qualche modo entrare la probabilità. Sono problemi delicati, e non è facile anche solo formularli in modo matematicamente chiaro; la teoria ergodica, almeno nel suo nascere, si può considerare come un tentativo in questo senso.

Vediamo allora rapidamente alcune idee di Boltzmann e Gibbs, che hanno lasciato traccia profonda nelle nozioni di base della teoria ergodica.³⁾

1.2. Il punto di vista di Boltzmann

Consideriamo un sistema termodinamico costituito da un numero N molto elevato di sottosistemi identici, ciascuno a l gradi di libertà; il sistema completo ha allora $n = lN$ gradi di libertà. E' tradizione indicare con μ lo spazio delle fasi $2l$ -dimensionale del singolo sottosistema, e con

$$\Gamma = \mu^N, \quad \dim \Gamma = 2n,$$

lo spazio delle fasi dell'intero sistema. Se indichiamo con

$$x^{(i)} = (p^{(i)}, q^{(i)}) = (p_1^{(i)}, \dots, p_l^{(i)}, q_1^{(i)}, \dots, q_l^{(i)}) \in \mu$$

le coordinate canoniche dell' i -esimo sottosistema, allora lo *stato microscopico* dell'intero sistema è rappresentato da una N -pla (ordinata) di punti $x^{(i)}$ in μ , o equivalentemente da un unico punto x in Γ ,

$$x = (p, q) = (p_1, \dots, p_n, q_1, \dots, q_n) \in \Gamma;$$

l'evoluzione microscopica appare allora indifferentemente come un movimento in Γ o una N -pla di movimenti in μ . Il movimento in Γ è evidentemente soluzione delle equazioni differenziali microscopiche del moto; queste si suppongono essere hamiltoniane, con hamiltoniana del tipo

$$H(p, q) = \sum_{i=1}^N h(p^{(i)}, q^{(i)}) + V(q), \quad (p, q) \in \Gamma,$$

²⁾ Il valore dei paradossi è sempre quello di mostrare che un problema non è stato posto in modo sufficientemente chiaro: le contraddizioni nascono regolarmente da confusioni nascoste.

³⁾ Questa breve esposizione è tratta dai seguenti lavori, di cui si suggerisce la consultazione:

- G.E. Uhlenbeck e G.W. Ford, *Lectures in Statistical Mechanics*, Lezione I (American Math. Soc., Providence, Rhode Island, 1963);
- J.L. Lebowitz e O. Penrose, *Modern ergodic theory*, Physics Today, feb. 1973, pag. 23;
- G. Benettin, L. Galgani e A. Giorgilli, *Appunti di Meccanica Razionale*, inizio del Capitolo 5 (dispense non pubblicate);
- G. Gallavotti, *Meccanica Statistica, e Insiemi Statistici*, voci per il *Dizionario delle Scienze Fisiche*, Istituto dell'Enciclopedia Italiana "G. Treccani" (Roma, 1993). Questi e altri scritti di Meccanica Statistica dello stesso autore (alcuni dei quali sulla teoria ergodica) sono raccolti in *Meccanica Statistica*, Quaderno del CNR n. 50 (1995).

ove h è l'hamiltoniana del singolo sottosistema (la stessa per tutti), mentre V è un opportuno potenziale di interazione tra i sottosistemi (si pensi alle collisioni in un gas). La presenza di V è essenziale perché i sottosistemi interagiscano e il sistema complessivo evolva in modo significativo: tuttavia si ammette che V sia sempre piccolo in valore rispetto al primo termine, e con buona approssimazione ininfluenza ai fini dei calcoli di energia.

Una prima importante idea di Boltzmann, che conduce alla sua nozione di stato macroscopico, è che “molti” stati microscopici vadano considerati macroscopicamente equivalenti e tra loro identificati, e che l'unica cosa rilevante a livello macroscopico sia la distribuzione dei punti in μ . Più precisamente:

- stati microscopici che differiscono per spostamenti sufficientemente piccoli dei punti in μ devono corrispondere al medesimo stato macroscopico. La maniera più semplice di tenerne conto è quella di suddividere μ in *cellette* $\omega_1, \omega_2, \dots$ sufficientemente piccole, tutte del medesimo volume ω , e considerare macroscopicamente equivalenti gli stati microscopici che differiscono per spostamento dei punti di μ dentro una celletta.
- Stati microscopici corrispondenti alla permuta dei punti in μ vanno anch'essi considerati macroscopicamente equivalenti. Poiché delle permutate entro la medesima celletta già si tien conto nel momento in cui si spostano i punti entro le cellette in tutti i modi possibili, ciò che qui si aggiunge è la permuta dei punti da una celletta all'altra.

Dopo queste identificazioni, e in questa visione discretizzata, lo stato macroscopico resta completamente caratterizzato dai *numeri di occupazione* N_1, N_2, \dots delle cellette, o se si vuole dalle densità

$$f_j = \frac{N_j}{N\omega}, \quad \sum_j f_j \omega = 1.$$

L'idea, purtroppo non facile da formalizzare in una operazione di limite, è che si possa raffinare la suddivisione in celle, mandando a zero il volume ω e aumentando contemporaneamente N , in modo da arrivare a una distribuzione continua di probabilità in μ , caratterizzata da una opportuna densità di probabilità $f : \mu \rightarrow \mathbb{R}$. Tale densità in μ rappresenta idealmente, per Boltzmann, uno stato macroscopico del sistema: in realtà però Boltzmann non si preoccupa molto del limite,⁴⁾ e ragiona volentieri sui numeri di occupazione N_j .

Si osservi che tutti gli stati microscopici tra loro equivalenti hanno quasi la stessa energia: infatti, spostamenti dei punti entro una celletta alterano di poco l'energia totale, e scambi dei punti tra cellette diverse non la alterano affatto; se denotiamo con ε_j l'energia della j -esima celletta,⁵⁾ ad esempio l'energia corrispondente al suo baricentro, allora l'energia totale è prossima al valore

$$E = \sum_j N_j \varepsilon_j. \tag{1.1}$$

⁴⁾ Boltzmann sembra a volte avere una visione finitista, sorprendentemente priva della preoccupazione di recuperare in qualche modo il continuo. Anche la dinamica microscopica in Γ è sostituita all'occorrenza da Boltzmann con una dinamica discreta (un numero altissimo ma finito di configurazioni, e passi temporali finiti, proprio come si fa oggi nelle simulazioni al computer). L'idea di fondo è che realizzazioni anche molto diverse della dinamica microscopica devono dar luogo alla stessa fenomenologia macroscopica, ovvero alla stessa termodinamica, quando siano preservate poche caratteristiche essenziali, come la conservazione dell'energia e del volume nello spazio delle fasi.

⁵⁾ si osservi che ε_j è ben definita solo nell'approssimazione in cui si trascura V .

A ogni scelta dei numeri di occupazione N_j delle singole celle corrisponde evidentemente un insieme $W(N_1, N_2, \dots)$ in Γ , disposto entro un sottile strato $\Gamma_{E \pm \Delta E}$ attorno alla superficie di energia costante Σ_E , con E data dalla (1.1) e ΔE determinato in qualche modo dalla dimensione delle celle in μ . Per effetto della dinamica i numeri di occupazione variano nel tempo, ma per la conservazione dell'energia il moto è confinato allo strato $\Gamma_{E \pm \Delta E}$, che dunque rappresenta, per così dire, il vero spazio delle fasi del sistema.

L'idea centrale di Boltzmann è ora che, a meno che il sistema non entri in regioni specialissime, di volume estremamente piccolo e per questo ritenute *a priori* poco probabili (su questo punto, che è il più cruciale per il legame con la teoria ergodica, torneremo tra un momento), la nuvola dei punti in μ evolve mantenendo praticamente costanti i numeri di occupazione, ovvero le densità f_j , e con essi lo stato macroscopico. Il ragionamento si svolge in questo modo: innanzitutto, indicando ancora con il simbolo W il volume del generico insieme W , si trova immediatamente

$$W(N_1, N_2, \dots) = \frac{N!}{N_1! N_2! \dots} \omega^N$$

(infatti, il fattore ω^N tiene conto degli spostamenti dei sottosistemi entro una stessa cella; il coefficiente combinatorio tiene conto degli scambi tra celle diverse). Si “dimostra” poi che

- i) il massimo W^* di $W(N_1, N_2, \dots)$, a E e N fissati, si trova per $N_j = N_j^*$, con

$$N_j^* = C N \omega e^{-\beta \varepsilon_j}, \quad C^{-1} = \sum_j e^{-\beta \varepsilon_j} \omega, \quad (1.2)$$

ovvero per $f_j = f_j^*$, con

$$f_j^* = C e^{-\beta \varepsilon_j}, \quad (1.3)$$

ove β è un moltiplicatore di Lagrange determinato dall'energia specifica E/N ;

- ii) per piccole variazioni dei numeri di occupazione, ovvero per

$$N_j = N_j^* + \delta N_j, \quad \delta N_j \ll N_j,$$

si ha

$$W \simeq W^* \prod_j e^{-\frac{1}{2} \frac{(\delta N_j)^2}{N_j}}. \quad (1.4)$$

Queste proprietà si deducono senza difficoltà assumendo che i numeri N_j siano molto grandi, in modo che sia buona l'approssimazione di Stirling $\log N_j! \simeq N_j(\log N_j - 1)$, e inoltre trattando tali numeri come se fossero variabili continue.⁶⁾

⁶⁾ Per dedurre le (1.2) e (1.4) cerchiamo il massimo di $\mathcal{S} := \log W$ sotto il doppio vincolo

$$\sum_j N_j = N, \quad \sum_j \varepsilon_j N_j = E,$$

con il metodo dei moltiplicatori di Lagrange. Nell'approssimazione di Stirling si ha

$$\mathcal{S} = C_{N,\omega} - \sum_j N_j (\log N_j - 1),$$

ove $C_{N,\omega}$ è una costante che dipende solo da N e ω ; segue subito che il massimo è dato da

$$\log N_j + \alpha + \beta \varepsilon_j = 0,$$

La (1.4) mostra che W ha distribuzione gaussiana in ciascuno dei numeri di occupazione N_j , e la larghezza della gaussiana è $\sigma_j = \sqrt{N_j} \ll N_j$. Si vede così che *variazioni relative anche molto piccole dei numeri di occupazione determinano grandi variazioni di W* . Ad esempio, se per ogni j si ha $\delta N_j \simeq \sigma_j$, ovvero

$$\frac{\delta N_j}{N_j} \simeq \frac{1}{\sqrt{N_j}} \ll 1,$$

il volume si riduce di un fattore enorme, pari all'esponenziale del numero di celle occupate. In definitiva, la quasi totalità dello spazio delle fasi (o meglio dello strato dinamicamente accessibile) corrisponde a numeri di occupazione *molto* ben definiti, dati dalla (1.2), o se si vuole a densità f_j in μ altrettanto ben definite, con fluttuazioni

$$\frac{\delta f_j}{f_j} = \frac{\delta N_j}{N_j} \simeq \frac{1}{\sqrt{N_j}}$$

del tutto irrisorie. Lo stato macroscopico corrispondente alla scelta privilegiata (1.2) dei numeri di occupazione è chiamato *stato di Maxwell–Boltzmann*; il fattore esponenziale nelle (1.2), (1.3) è il ben noto *fattore di Boltzmann*. La situazione è simbolicamente illustrata in figura 1.1, ove compaiono il grande insieme W^* corrispondente allo stato di Maxwell–Boltzmann, e molti insiemi piccoli corrispondenti agli stati macroscopicamente distinguibili dallo stato di Maxwell–Boltzmann.

A questo punto Boltzmann introduce un'ipotesi fondamentale di carattere dinamico,⁷⁾ nota come *ipotesi ergodica*: la dinamica microscopica è tale che il punto x in Γ , rappresentativo dello stato microscopico del sistema, vaga su tutto lo strato $\Gamma_{E \pm \Delta E}$, e spende in ogni volume W un tempo mediamente proporzionale a W stesso. Pertanto, se il sistema è osservato a un istante scelto a caso nell'arco di un tempo molto lungo (per dar tempo alla traiettoria di esplorare l'intera superficie di energia costante), la *probabilità* di trovare il

con α e β determinati dalle condizioni di vincolo, e la (1.2) è immediata; β si verifica essere positivo. Si tratta di un massimo perché la matrice delle derivate seconde di \mathcal{S} è

$$\frac{\partial^2 \mathcal{S}}{\partial N_j \partial N_k} = -\frac{\delta_{jk}}{N_j},$$

diagonale con tutti gli elementi negativi. Infine, con uno sviluppo di Taylor arrestato al secondo ordine si ha

$$\mathcal{S}(N_1^* + \delta N_1, N_2^* + \delta N_2, \dots) \simeq \mathcal{S}(N_1^*, N_2^*, \dots) + \sum_j \frac{\partial \mathcal{S}}{\partial N_j} \delta N_j + \frac{1}{2} \sum_{jk} \frac{\partial^2 \mathcal{S}}{\partial N_j \partial N_k} \delta N_j \delta N_k;$$

il termine del primo ordine si annulla per variazioni compatibili con i vincoli (altrimenti non avremmo un massimo!), e la (1.4) segue immediatamente.

⁷⁾ Su questo punto in realtà il pensiero di Boltzmann è piuttosto complicato, e non facile da ricostruire. Sulla nozione di ergodicità in Boltzmann sembrano essere avvenute vistose incomprensioni anche da parte di autori molto qualificati come Ehrenfest, che pure era allievo di Boltzmann. In particolare, l'idea che una singola traiettoria (una curva continua) debba passare per *tutti* i punti di una superficie di energia costante — idea matematicamente bizzarra, e soprattutto fisicamente inutile — a differenza di quanto riportato in molti libri di testo non è attribuibile a Boltzmann (se non all'interno della sua dinamica discretizzata, ove la cosa ha perfettamente senso). Per una discussione critica di questo punto si veda il lavoro sopra citato di G. Gallavotti, o anche, dello stesso autore, *L'hypothèse ergodique et Boltzmann*, in *Dictionnaire Philosophique des Presses Univ. Françaises*, p. 1081–1085 (Paris, 1988); *Ergodicity, Ensembles, Irreversibility and beyond*, Journ. Stat. Phys. **78**, 1571–1589 (1995).

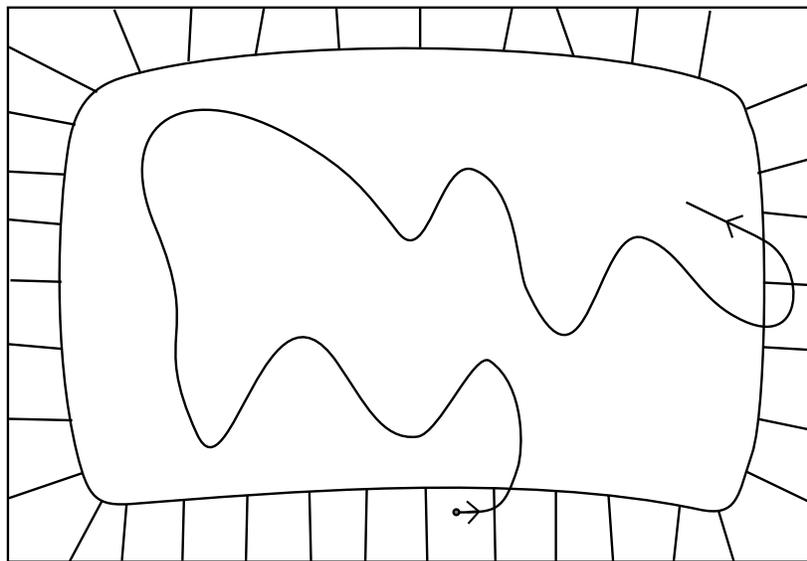


Figura 1.1. Lo stato di Maxwell–Boltzmann domina in $\Gamma_{E \pm \Delta E}$.

sistema in un generico insieme coincide con il suo volume W , a meno della normalizzazione. Questa interpretazione del volume nello spazio delle fasi come probabilità *a priori* di un insieme di stati microscopici è comunemente chiamato principio di *equiprobabilità degli stati microscopici*. A questo principio, come è noto, è appesa l'intera meccanica statistica, in ogni sua formulazione;⁸⁾ l'ipotesi ergodica di Boltzmann ne rappresenta, si può dire, una possibile giustificazione dinamica.

In pratica — è questa la conclusione di Boltzmann — comunque il sistema sia preparato, anche in condizioni molto lontane dall'equilibrio termodinamico (ad esempio, con riferimento alla figura 1.1, in uno dei minuscoli insiemi che contornano W^*), la dinamica microscopica spingerà presto il sistema in W^* , e in questo insieme esso spenderà poi la stragrande quantità del tempo, a meno di piccole fluttuazioni estremamente rare. Il transiente, in pratica il tempo necessario ad allontanarsi dalle condizioni iniziali per arrivare in W^* , rappresenta nel pensiero di Boltzmann l'avvicinamento all'equilibrio (misure fatte in questo intervallo di tempo darebbero risultati fortemente influenzati dalle condizioni iniziali del sistema); la situazione di equilibrio è invece quella in cui, in accordo con il principio di equiprobabilità degli stati microscopici, ogni stato macroscopico ha probabilità di essere realizzato proporzionale al suo volume in Γ . Dal punto di vista pratico, l'equilibrio è del tutto indistinguibile dal singolo stato di Maxwell–Boltzmann: in linea di principio però l'equilibrio è una collezione di stati macroscopici, e benchè uno di essi sia di gran lunga dominante, le fluttuazioni non sono *a priori* escluse.

⁸⁾ Boltzmann stesso, come è noto, fu in grado di dedurre da questo principio le leggi della termodinamica. Per questo introdusse una interpretazione microscopica delle fondamentali grandezze termodinamiche U , P , V , T e S (costruendo così un *modello* di termodinamica), e mostrò in particolare che l'equiprobabilità degli stati microscopici implica il secondo principio, $dU + PdV = TdS$. Vale la pena di ricordare che l'entropia S resta definita, microscopicamente, da $S = k_B \mathcal{S} = k_B \log W$, con $k_B > 0$ (costante di Boltzmann), e dunque, a meno di una inessenziale costante additiva, da

$$S = -k_B N \omega \sum_j f_j \log f_j .$$

L'ipotesi ergodica è bel lontana dall'essere dimostrabile in pratica, in sistemi fisicamente realistici: ma ha il *grandissimo* valore di mostrare che in linea di principio ha senso discutere di probabilità in un sistema deterministico, e che il comportamento macroscopico non è incompatibile con la dinamica microscopica. Si osservi in particolare che, in questa visione di Boltzmann, i due paradossi sopra riportati sono sciolti: per il secondo cade del tutto la contraddizione (le fluttuazioni corrispondenti al ritorno del sistema vicino alle condizioni iniziali sono magari rare al limite dell'assurdo,⁹⁾ ma in linea di principio sono ammesse anche a livello macroscopico); per quanto riguarda il primo, la risposta è simile: i processi inversi al raggiungimento dell'equilibrio esistono anche a livello macroscopico, e coincidono precisamente con le fluttuazioni ammesse da Boltzmann, anche se sono così rari da non avere alcuna rilevanza pratica.

1.3. Il punto di vista di Gibbs

La nozione di Gibbs di stato macroscopico è abbastanza diversa da quella di Boltzmann; in essa la probabilità gioca un ruolo più essenziale, per così dire più primitivo. Mentre Boltzmann pensa fundamentalmente allo spazio μ , e associa lo stato macroscopico a una distribuzione f dei sottosistemi nello spazio μ dove ciascuno di essi è definito, Gibbs invece lavora direttamente in Γ , e identifica lo stato macroscopico con una distribuzione di probabilità ρ in tale spazio; l'interpretazione di ρ è che per un generico $W \subset \Gamma$ la probabilità *a priori* che sia realizzato fisicamente uno degli stati microscopici $x \in W$ è

$$P(W) = \int_W \rho dV ,$$

ove $dV = dp_1, \dots, dp_n dq_1, \dots, dq_n$ è il volume¹⁰⁾ in Γ . Ogni stato macroscopico è così una misura in Γ , con densità ρ .

La differenza rispetto a Boltzmann è notevole: mentre Boltzmann pensa al comportamento di un singolo sistema che evolve su Γ — una singola traiettoria, con un dato iniziale non conosciuto ma in linea di principio determinato, che con le sue evoluzioni in Γ assegna la probabilità ai diversi sottoinsiemi — Gibbs invece considera a ogni istante di tempo una famiglia (*ensemble*) di sistemi in evoluzione, repliche mentali indipendenti dello stesso sistema fisico in stati microscopici diversi, distribuiti in Γ con una opportuna densità di probabilità ρ . L'idea è che in ogni esperimento la preparazione del sistema a $t = 0$ determina non un dato iniziale, ma una distribuzione iniziale¹¹⁾ ρ_0 in Γ (lo stato macroscopico iniziale); per effetto della dinamica ciascuna condizione iniziale evolve poi indipendentemente — non c'è evidentemente interazione tra queste repliche immaginarie dello stesso sistema — come farebbe un fluido di particelle non interagenti, e resta così determinata a ogni istante una opportuna distribuzione ρ_t .

Dalla conservazione del volume nello spazio delle fasi (teorema di Liouville; si veda il paragrafo 2.1, esempio 2.4) segue facilmente per ρ_t l'equazione di evoluzione

$$\rho_t(x) = \rho_0(\Phi^{-t}(x)) , \quad x \in \Gamma , \quad (1.5)$$

⁹⁾ Si veda il paragrafo 3.1 per ulteriori commenti su questo punto.

¹⁰⁾ Ci riferiamo qui al volume euclideo in coordinate canoniche qualsiasi (tale volume non dipende dalla particolare scelta delle coordinate canoniche).

¹¹⁾ La probabilità in questo approccio non nasce dunque, come in Boltzmann, dalla scelta casuale dell'istante al quale si osserva il sistema, ma piuttosto dall'ignoranza dello stato microscopico iniziale.

avendo indicato con $x \mapsto \Phi^t(x)$ l'evoluzione microscopica (la soluzione delle equazioni di Hamilton all'istante t , con dato iniziale x). Un modo di dedurre la (1.5) è il seguente: è chiaro che lo stato microscopico del sistema si trova in W all'istante t , se e solo se esso si trovava in $W_0 = \Phi^{-t}(W)$ a $t = 0$; per ogni $W \subset \Gamma$ si ha allora

$$\int_W \rho_t(x) dV = \int_{\Phi^{-t}(W)} \rho_0(x) dV$$

(con ciò si è semplicemente affermato che in presenza di una dinamica deterministica la probabilità si conserva). Si esegue ora a destra il cambiamento di variabile di integrazione $x = \Phi^{-t}(x')$, $dV = (\det J^{-t})dV'$, ove J^t indica per ogni t la matrice jacobiana $\frac{\partial \Phi^t}{\partial x}$; tale cambiamento riporta il dominio di integrazione a W . Per il teorema di Liouville sulla conservazione del volume dello spazio delle fasi si ha $\det J^t = 1$ per ogni t ; abolendo gli apici, e tenendo conto dell'arbitrarietà di W , si trova immediatamente la (1.5).¹²⁾

E' ora naturale la ricerca degli stati (o distribuzioni, o *ensembles*) di equilibrio, ovvero quelli in cui $\rho_t(x)$ in ogni punto x non dipende da t . Un esempio di distribuzione di equilibrio si ottiene prendendo un qualunque guscio di energia costante $\Gamma_{E \pm \Delta E}$, e ponendo

$$\rho^*(x) = \begin{cases} \text{cost} & \text{in } \Gamma_{E \pm \Delta E} \\ 0 & \text{altrove} \end{cases} ;$$

è infatti evidente dall'equazione di evoluzione (1.5) che ponendo $\rho_0 = \rho^*$ si ha $\rho_t = \rho^*$ per ogni t . La situazione di equiprobabilità degli stati microscopici (nel senso già visto in Boltzmann) è dunque, per Gibbs, uno stato di equilibrio. E' evidente che tale stato non è unico: presa infatti

$$\rho^*(x) = F(H(x)) ,$$

ove $F : \mathbb{R} \rightarrow \mathbb{R}^+$ è arbitraria (a meno della normalizzazione) e H è l'hamiltoniana, si verifica immediatamente che ρ^* è di equilibrio. In effetti, la dinamica conserva l'energia, dunque non ha senso lavorare in Γ , e ci si deve restringere, come in Boltzmann, a un guscio "sottile al limite della indistinguibilità fisica", o meglio a una singola superficie di energia costante Σ_E . Come sarà illustrato tra breve (paragrafo 2.1, esempio 2.4) la conservazione del volume nello spazio delle fasi Γ induce una misura conservata μ su ciascuna superficie di energia costante; questa non coincide con l'area euclidea $d\Sigma$ di Σ_E in coordinate canoniche, ma differisce da essa per una densità proporzionale a $1/\|\nabla H\|$:

$$d\mu = C \frac{d\Sigma}{\|\nabla H\|} ,$$

ove $\|\cdot\|$ denota la norma euclidea.¹³⁾

¹²⁾ In effetti, questa equazione altro non è che l'equazione di continuità per la probabilità in forma lagrangiana, che per un generico sistema di equazioni differenziali $\dot{x} = f(x)$ si scrive

$$d\rho/dt + \rho \nabla \cdot f = 0 ;$$

per i sistemi hamiltoniani si ha $\nabla \cdot f = 0$, dunque ρ_t è costante lungo le traiettorie e si ha la (1.5).

¹³⁾ Benché l'area $d\Sigma$ e la norma $\|\nabla H\|$ dipendano singolarmente dalla scelta delle coordinate canoniche (e singolarmente appaiano anche dimensionalmente non corrette), $d\mu$ ne è indipendente (e l'espressione risulta dimensionalmente corretta).

A questo punto si può introdurre, al posto della densità ρ in Γ , una densità superficiale in Σ_E , che continuiamo a denotare ρ , e attribuire a ogni dominio $A \subset \Sigma_E$ la probabilità

$$P(A) = \int_A \rho(x) d\mu ;$$

come prima, ρ è soggetta all'equazione di evoluzione (1.5). Ora è evidente che esiste la distribuzione di equilibrio ρ^* costante su Σ_E , precisamente (se μ è normalizzata) $\rho^*(x) = 1$ per ogni $x \in \Sigma_E$, ed è naturale porsi due domande:

- i. se tale distribuzione di equilibrio sia unica;
- ii. se in più si possa garantire il raggiungimento dell'equilibrio a partire da stati macroscopici qualsiasi, ovvero se, in qualche senso da precisare, risulti¹⁴⁾

$$\lim_{t \rightarrow \infty} \rho_t \rightarrow \rho^* .$$

Questa nozione, vedremo, si può formalizzare in un contesto matematicamente preciso, e corrisponde a una delle idee rilevanti della teoria ergodica, quella di sistema mescolante.

Chiudiamo qui questa breve introduzione, il cui scopo, come si è detto, era di illustrare il contesto fisico dal quale è poi emersa la teoria ergodica, e passiamo senz'altro a una esposizione formale della teoria; testi consigliati per l'approfondimento di quanto si dirà, e dai quali cui sono tratte queste note, sono i seguenti:

- V.I. Arnol'd e A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York 1968). Asciutto ma molto bello, in particolare sono interessanti le numerose appendici. E' il testo che seguiamo di più.
- P.R. Halmos, *Ergodic Theory* (Chelsea, New York 1956).
- I.P.Cornfeld, S.V. Fomin e Ya.G. Sinai, *Ergodic Theory* (Springer, 1982). Un manuale molto completo.
- Ya.G. Sinai (editore) e altri, *Dynamical Systems II – Ergodic Theory* (Springer, 1992). Ancor più completo, ma terribilmente asciutto e telegrafico.
- G. Gallavotti, *Aspetti della teoria ergodica, qualitativa e statistica del moto* (Pitagora, Bologna 1981).

¹⁴⁾ Per la (1.5), ρ_t mantiene su ogni traiettoria il valore iniziale, dunque non può esservi un limite puntuale. Ma può avvenire che per una classe rilevante di funzioni risulti

$$\int_{\Sigma_E} f(x) \rho_t(x) d\mu \longrightarrow \int_{\Sigma_E} f(x) d\mu ,$$

ovvero che il limite possa essere definito in senso debole.

2. Sistemi dinamici classici e astratti

2.1. Nozione ed esempi

Sia M una varietà differenziabile¹⁵⁾ compatta sulla quale è definita una misura regolare¹⁶⁾ normalizzata μ , e sia $\Phi := \{\Phi^t\}$, $t \in \mathbb{R}$ oppure $t \in \mathbb{Z}$, un gruppo a un parametro di diffeomorfismi su M , che preservano la misura: precisamente, tali che

$$\begin{aligned}\Phi^t \circ \Phi^s &= \Phi^{t+s}, & \Phi^0 &= \text{Id}, \\ \mu(\Phi^{-t}(A)) &= \mu(A)\end{aligned}\tag{2.1}$$

per ogni t, s in \mathbb{R} o \mathbb{Z} , e ogni $A \subset M$ misurabile. Con la notazione $\Phi^{-t}(A)$ si intende l'insieme

$$\Phi^{-t}(A) := \{x \in M : \Phi^t(x) \in A\}.\tag{2.2}$$

Definizione. La collezione (M, μ, Φ) , con le proprietà sopra indicate, è chiamata *sistema dinamico classico* (continuo o discreto, a seconda che sia $t \in \mathbb{R}$ o $t \in \mathbb{Z}$).

Una generalizzazione di un certo interesse è quella in cui lo spazio M in cui agisce Φ non è dotato della struttura di varietà, ma è uno spazio astratto provvisto di una misura (normalizzata) μ ; corrispondentemente, si chiede che le applicazioni Φ^t del gruppo siano soltanto automorfismi di M (cioè applicazioni $M \rightarrow M$, definite e invertibili quasi ovunque, che preservano la misura).

Definizione. La collezione¹⁷⁾ (M, μ, Φ) , con le proprietà sopra indicate, è chiamata *sistema dinamico astratto*, o semplicemente *sistema dinamico* (continuo o discreto).

Il caso in cui Φ^t sia regolare a tratti (cioè definito e regolare al di fuori di un numero finito di sottovarietà di M) è ancora considerato un sistema dinamico classico. Una ulteriore naturale generalizzazione (per la quale non vale la pena di introdurre nomi nuovi) è quella in cui viene meno l'invertibilità delle applicazioni, e $\Phi = \{\Phi^t\}$, $t \geq 0$, è un semigruppato di endomorfismi (ovvero applicazioni di uno spazio in sé che preservano la misura), regolari nel caso di sistema dinamico classico. Prenderemo in considerazione solo occasionalmente sistemi non invertibili.

Nel caso discreto, la proprietà gruppale implica che sia $\Phi^t = (\Phi^1)^t$; le trasformazioni del gruppo (o semigruppato) corrispondono dunque all'iterazione di una mappa Φ^1 , e per semplicità denoteremo $\Phi = \Phi^1$. Tra i sistemi continui il caso tipico è quello in cui $\Phi^t(x)$ è la soluzione al tempo t e con dato iniziale x di un'equazione differenziale (autonoma) $\dot{x} = f(x)$ su M ; la proprietà gruppale (2.1) è allora automaticamente soddisfatta.

¹⁵⁾ La classe di differenziabilità non gioca un ruolo importante in quello che andremo a dire. La classe C^2 , anche a tratti, per M , μ e Φ sarà comunque sempre sufficiente.

¹⁶⁾ Se $dV = dx_1, \dots, dx_n$ denota il volume euclideo in una carta qualsiasi della varietà, si chiede che sia $d\mu = \rho dV$, con densità ρ regolare.

¹⁷⁾ Nel caso astratto si trova di frequente anche la notazione $(M, \mathcal{A}, \mu, \Phi)$, utile a mettere in evidenza la σ -algebra degli insiemi misurabili sulla quale è definita μ .

Vediamo allora innanzitutto alcuni esempi semplici e significativi.

Esempio 2.1 (Moto quasi periodico su \mathbb{T}^2). Prendiamo:¹⁸⁾

$$\begin{aligned} M &:= \mathbb{T}^2 := \mathbb{R}^2 / \mathbb{Z}^2 \\ \mu &= \text{misura di Lebesgue} \\ \Phi^t(x, y) &= (x + v_x t, y + v_y t) \text{ mod } 1 . \end{aligned} \tag{2.3}$$

Il sistema è classico, continuo, invertibile; si tratta evidentemente del moto uniforme sul toro, corrispondente all'equazione differenziale $\dot{x} = v_x, \dot{y} = v_y$ (si veda la figura 2.1).

Se il rapporto $\alpha = v_x/v_y$ è razionale, $\alpha = p/q$ con $p, q \in \mathbb{Z}$, allora tutte le orbite sono periodiche, precisamente di periodo $T = p/v_x = q/v_y$; in caso contrario sono aperte e, come vedremo, dense sul toro. Lo studio è strettamente collegato al seguente altro esempio:

Esempio 2.2 (Traslazione su \mathbb{T}^1). Prendiamo:

$$\begin{aligned} M &:= \mathbb{T}^1 := \mathbb{R}/\mathbb{Z} \\ \mu &= \text{misura di Lebesgue} \\ \Phi(x) &= x + \alpha \text{ mod } 1 . \end{aligned} \tag{2.4}$$

Il sistema è classico, discreto, invertibile. La connessione con l'esempio precedente è immediata (figura 2.2): presa una qualunque sezione N di \mathbb{T}^2 del tipo $y = \text{cost.}$ (*sezione di Poincaré*), ad esempio $y = 0$, alle orbite di \mathbb{T}^2 corrispondono successioni di punti in N , tali che a $x \in N$ segue $\Phi(x)$ dato dalla (2.4); infatti, per $T_y = 1/v_y$, si ha

$$\Phi^{T_y}(x, y) = (x + \alpha \text{ mod } 1, y) , \quad \alpha = v_x/v_y .$$

E' evidente che per α razionale le orbite sono periodiche; mostriamo che

Proposizione 2.1. *Se α è irrazionale, allora le orbite sono dense in \mathbb{T}^1 .*

Dimostrazione. Mostriamo innanzitutto che per ogni $\varepsilon > 0$ esiste s tale che¹⁹⁾

$$\text{dist}(x, \Phi^s(x)) < \varepsilon . \tag{2.5}$$

Infatti, fissato x , gli infiniti punti dell'orbita (che non si chiude) hanno certamente una accumulazione, perciò esistono $k, l \in \mathbb{Z}$ tali che $\text{dist}(\Phi^k(x), \Phi^l(x)) < \varepsilon$. Ma Φ preserva la distanza, e dunque (applicando Φ^{-k}) $\text{dist}(x, \Phi^{l-k}(x)) < \varepsilon$, ovvero è soddisfatta la (2.5) con $s = l - k$. Se ne deduce che $\Phi^s(x), \Phi^{2s}(x), \dots$, procedono "a passi più corti di ε ",

¹⁸⁾ Con il termine "misura di Lebesgue" sul toro si intende qui la misura corrispondente alla misura di Lebesgue $dx dy$ sul piano (detta anche *misura di Haar*). La notazione $(x, y) \text{ mod } 1$ indica il punto del toro corrispondente al punto (x, y) del piano; in pratica $x \text{ mod } 1$ è la parte frazionaria di x , ovvero $x - [x]$, ove $[x]$ (parte intera di x) denota il più grande intero non superiore a x .

¹⁹⁾ La distanza su \mathbb{T}^1 è definita, come è naturale, da $\text{dist}(x, y) = \min(|x - y|, 1 - |x - y|)$ (la più piccola delle distanze tra x e i punti di \mathbb{R} equivalenti a y); in modo analogo si definisce la distanza su \mathbb{T}^n .

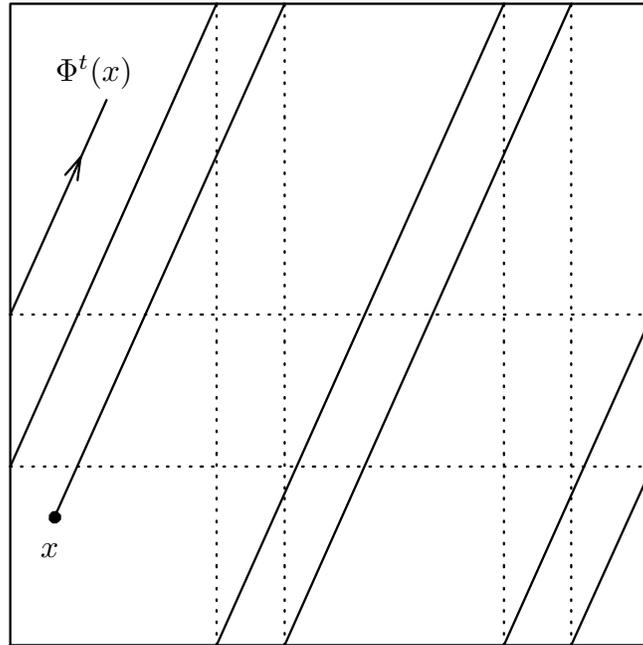


Figura 2.1. Il moto quasi periodico su \mathbb{T}^2 .

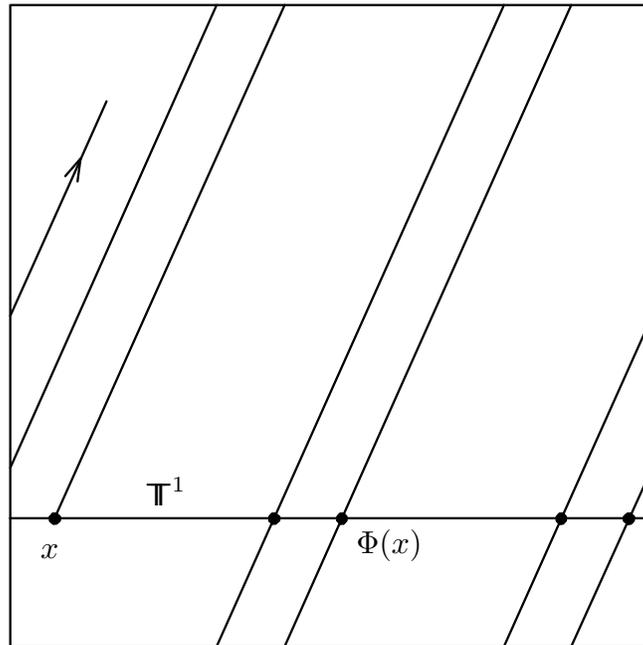


Figura 2.2. La traslazione su \mathbb{T}^1 vista come sezione di Poincaré del moto quasi periodico su \mathbb{T}^2 .

perciò entrano in qualunque intervallo di lunghezza ε ; per l'arbitrarietà di ε la conclusione è immediata. ■

I due esempi si generalizzano a più dimensioni, in particolare

Esempio 2.3. Prendiamo

$$\begin{aligned} M &:= \mathbb{T}^n := \mathbb{R}^n / \mathbb{Z}^n \\ \mu &= \text{misura di Lebesgue} \\ \Phi^t(x_1, \dots, x_n) &= (x_1 + v_1 t, \dots, x_n + v_n t) \bmod 1 . \end{aligned} \tag{2.6}$$

Per studiare, sia pure a livello preliminare, questo sistema, premettiamo la seguente

Definizione. Il vettore $v = (v_1, \dots, v_n)$ si dice *risonante* con il vettore intero $k \in \mathbb{Z}^n$, se risulta

$$k \cdot v = 0 ; \tag{2.7}$$

si dice *non risonante* se risona col solo vettore nullo. Il vettore intero k a sua volta è detto *vettore risonante* per v ; la (2.7), con $k \neq 0$, si dice *relazione di risonanza*.

L'insieme dei k che risonano con un fissato v costituisce un sottospazio vettoriale \mathcal{R} di \mathbb{Z}^n , detto *reticolo risonante*, o *modulo risonante*²⁰⁾ di v ; la sua dimensione, cioè il numero r di vettori interi indipendenti che risonano con v , è detto *molteplicità* della risonanza.

Proposizione 2.2 (Jacobi, 1835). *Se v è non risonante, allora ogni orbita del sistema (2.6) è densa in \mathbb{T}^n .*

La dimostrazione si vedrà più avanti, quando si dimostrerà l'ergodicità di questa trasformazione.

Proposizione 2.3. *Se v ammette r relazioni indipendenti di risonanza, allora ogni orbita è confinata a una varietà $(n - r)$ -dimensionale N , diffeomorfa al toro \mathbb{T}^{n-r} . Più precisamente, esiste una matrice intera J , $\det J = 1$, tale che il cambio di coordinate*

$$y = Jx , \quad x = J^{-1}y , \tag{2.8}$$

muta v in

$$u = Jv = (0, \dots, 0, u_{r+1}, \dots, u_n) .$$

- **Osservazione.** Grazie al fatto che J è intera e $|\det J| = 1$, si ha che J e J^{-1} sono entrambe intere (la condizione è necessaria e sufficiente). Allora la trasformazione (2.8) è ben definita e regolare sul toro (cioè la (2.8), pensata come trasformazione lineare in \mathbb{R}^n , è coerente con l'operazione di modulo: punti equivalenti sono inviati in punti equivalenti, e viceversa).

²⁰⁾ Modulo, in algebra, è uno spazio vettoriale su un anello, anziché su un campo (in questo caso su \mathbb{Z} , anziché su \mathbb{R}).

Dimostrazione. La dimostrazione è basata sul seguente lemma, dimostrato nell'appendice A:

Lemma 2.4. *Assegnato un reticolo $\mathcal{R} \subset \mathbb{Z}^n$ di dimensione r , $1 \leq r \leq n$, esiste una matrice intera J , $\det J = 1$, le cui prime r righe appartengono a \mathcal{R} : ovvero si ha $k^{(i)} := (J_{i1}, \dots, J_{in}) \in \mathcal{R}$, per $1 \leq i \leq r$.*

Ammesso il lemma, la dimostrazione della proposizione è immediata: infatti, posto $y = Jx$ si ha

$$y \mapsto y + ut, \quad u = Jv,$$

e per la (2.7) si trova $u_i = k^{(i)} \cdot v = 0$, $i = 1, \dots, r$, come richiesto. Il moto con dato iniziale si svolge pertanto su una varietà V di equazione $y_i = y_i^0$, $i = 1, \dots, r$, che è banalmente isomorfa al toro \mathbb{T}^{n-r} (le prime r coordinate sono bloccate, le altre corrono su \mathbb{T}^{n-r}). Corrispondentemente la variabile x è confinata a $N = J^{-1}(V)$. ■

- **Osservazione.** Per la proposizione precedente, se non vi sono ulteriori relazioni di risonanza, ovvero se r è l'ordine, allora l'orbita è densa su N (su V , nelle variabili y introdotte nella dimostrazione). *Il toro \mathbb{T}^n risulta così decomposto (foliato) in sotto-varietà invarianti* (i tori \mathbb{T}^{n-r} che si ottengono al variare del dato iniziale, ovvero di y_1^0, \dots, y_n^0), *non ulteriormente decomponibili*.
- **Osservazione.** Questo problema si può riguardare come il problema del moto di n oscillatori armonici indipendenti, di hamiltoniana

$$H = \sum_{i=1}^n \omega_i I_i, \quad I = (I_1, \dots, I_n) \in \mathbb{R}^n, \quad \varphi = (\varphi_1, \dots, \varphi_n) \in \mathbb{T}^n.$$

Le azioni infatti sono costanti, dunque il moto si svolge sul toro \mathbb{T}^n (qui è naturale prendere le coordinate φ sul toro tra 0 e 2π anziché tra 0 e 1, ma ovviamente non fa differenza; la corrispondenza è $\varphi = 2\pi x$, $\omega = 2\pi v$). A seconda delle proprietà di risonanza di ω , le orbite sono dense in \mathbb{T}^n , o su tori di dimensione inferiore.

L'esempio si generalizza al caso di un qualsiasi sistema dinamico integrabile:

$$H(I, \varphi) = h(I), \quad \dot{\varphi} = \omega(I) = \frac{\partial h}{\partial I}(I),$$

e la cosa è (molto) interessante perchè le proprietà di risonanza di ω , che ora dipende da I , sono adesso diverse da toro a toro. Il gioco delle risonanze è un po' l'anima della teoria delle perturbazioni per sistemi prossimi a sistemi integrabili, e dunque l'esempio che stiamo seguendo, e sul quale avremo modo di ritornare, è particolarmente significativo.

Esempio 2.4. I sistemi hamiltoniani.

Per il teorema di Liouville²¹⁾ il volume euclideo (in arbitrarie coordinate canoniche: il volume non ne dipende) si conserva, perciò si potrebbe definire il sistema dinamico prendendo come M una corona compresa tra due superfici di energia costante (compatte, per

²¹⁾ In breve: sia $\dot{x} = f(x)$ un'equazione differenziale in \mathbb{R}^m (o su di una varietà m -dimensionale), e sia $x \mapsto \Phi^t(x)$, $t \in \mathbb{R}$, il corrispondente flusso. Condizione necessaria e sufficiente perchè Φ^t preservi il

la normalizzazione), e come μ il volume stesso. Ma non è interessante, perchè la dinamica preserva l'energia, e di fatto si svolge sulle superfici Σ_E di energia costante. Vale ora la

Proposizione 2.5. *Su ciascuna superficie di energia costante Σ_E compatta esiste una misura invariante μ , che a ogni disco $A \subset \Sigma_E$ associa la misura*

$$\begin{aligned} \mu(A) &= C \frac{d}{d\varepsilon} \text{Vol}(A, \varepsilon) \Big|_{\varepsilon=0} \\ &= C \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \text{Vol}(A, \varepsilon) \end{aligned} \quad (2.9)$$

ove $\text{Vol}(A, \varepsilon)$ indica il volume (euclideo) del "cilindro" di base A , compreso tra Σ_E e $\Sigma_{E+\varepsilon}$, con pareti laterali qualsiasi purché trasverse a Σ_E ; la costante C è determinata dalla normalizzazione. In coordinate canoniche la misura dell'elemento di area $d\Sigma$ si scrive

$$d\mu = \mu(d\Sigma) = C \frac{d\Sigma}{\|\nabla H\|}, \quad (2.10)$$

ove $d\Sigma$ e $\|\cdot\|$ indicano area e norma euclidea in coordinate canoniche arbitrarie; l'espressione risulta indipendente dalla scelta delle coordinate canoniche.

Dimostrazione. Innanzitutto osserviamo che $\mu(A)$ è ben definita: scelte diverse delle pareti laterali, purché trasverse a Σ_E , determinano differenze $\mathcal{O}(\varepsilon^2)$ in $\text{Vol}(A, \varepsilon)$, che dunque sono irrilevanti al limite $\varepsilon \rightarrow 0$. Per effetto della dinamica, sia $\text{Vol}(A, \varepsilon)$ sia ε sono preservati dalla dinamica, e l'invarianza di μ è immediata. La (2.10) segue da "volume = base per altezza", con altezza $h = \varepsilon/\|\nabla H\| + \mathcal{O}(\varepsilon^2)$. L'indipendenza di $d\mu$ dalle

volume è che la matrice jacobiana

$$J^t := \frac{\partial \Phi^t}{\partial x} = \left(\frac{\partial \Phi_i^t}{\partial x_j} \right)$$

abbia $|\det J^t(x)| = 1$ per ogni t e ogni x . Ora non è difficile dedurre per il determinante l'equazione di evoluzione

$$\frac{d}{dt} \det J^t(x) = \det J^t(x) (\nabla \cdot f)(\Phi^t(x));$$

nel caso hamiltoniano, con $m = 2n$ e $x = (p, q)$, si trova immediatamente $\nabla \cdot f = 0$ identicamente, e poiché J^0 è l'identità la conclusione è immediata. A sua volta l'equazione per $\det J^t$ si deduce così: innanzitutto, usando la proprietà grupitale del flusso $\Phi^{t+s}(x) = \Phi^s(\Phi^t(x))$ e la conseguente proprietà

$$J^{t+s}(x) = J^s(\Phi^t(x)) J^t(x)$$

per la matrice Jacobiana, la verifica si riduce a

$$\frac{d}{ds} \Big|_{s=0} \det J^s(x) = (\nabla \cdot f)(x);$$

d'altra parte, $J^s(x)$ ha la forma

$$J^s(x) = \begin{pmatrix} 1 + sf_1(x) & & & & \\ & 1 + sf_2(x) & & & \\ & & \cdots & & \\ \mathcal{O}(s) & & & \cdots & \mathcal{O}(s) \\ & & & & \cdots \\ & & & & & 1 + sf_n \end{pmatrix} + \mathcal{O}(s^2),$$

da cui segue immediatamente $\det J^s = 1 + s \nabla \cdot f + \mathcal{O}(s^2)$, e questo basta.

coordinate *canoniche* prescelte è automatica (il cambio di coordinate, come l'evoluzione hamiltoniana, preserva sia il volume sia ε).

- **Esercizio.** Si valuti esplicitamente $d\mu$ per l'oscillatore armonico di hamiltoniana $H = \frac{1}{2m}p^2 + \frac{k}{2}q^2$; si esegua il riscaldamento canonico $p = \alpha\tilde{p}$, $q = \alpha^{-1}\tilde{q}$, e si confrontino i risultati [*Risp.*: si trova $d\Sigma = \frac{1}{p}\sqrt{p^2 + m^2k^2q^2}dq$, $\|\nabla H\| = \frac{1}{m}\sqrt{p^2 + m^2k^2q^2}$, $d\mu = Cm\frac{dq}{p}$; il riscaldamento muta $\frac{dq}{p}$ in $\alpha^{-2}\frac{d\tilde{q}}{\tilde{p}}$, ma si ha anche $m = \alpha^2\tilde{m}$, dunque $d\mu$ è invariante].

A un sistema hamiltoniano corrisponde dunque un sistema dinamico classico (M, μ, Φ) , continuo e invertibile, su ogni superficie di energia costante compatta Σ_E , con $M = \Sigma_E$, μ definita come sopra, e Φ soluzione delle equazioni di Hamilton (*flusso hamiltoniano*).

La proposizione 2.5 si generalizza al caso in cui vi siano $k \geq 1$ integrali del moto F_1, \dots, F_k linearmente indipendenti, cioè tali che $\nabla F_1, \dots, \nabla F_k$ siano linearmente indipendenti su una superficie²²⁾ di livello Σ_c , $c = (c_1, \dots, c_k)$, definita da $F_i(p, q) = c_i$. La misura conservata risulta essere

$$d\mu = \frac{d\Sigma}{\text{Vol}(\nabla F_1, \dots, \nabla F_k)},$$

ove $d\Sigma$ è l'area euclidea su Σ_c , mentre al denominatore compare il volume euclideo del parallelepipedo generato da $\nabla F_1, \dots, \nabla F_k$.

Il seguente esempio è particolarmente importante:

Esempio 2.5 (Automorfismo algebrico di \mathbb{T}^2 , meglio noto come²³⁾ *Gatto di Arnol'd*). Prendiamo:

$$\begin{aligned} M &= \mathbb{T}^2 = \{(x, y) \bmod 1\} \\ \mu &= \text{misura di Lebesgue} \\ \Phi(x, y) &= (x + y, x + 2y) \bmod 1. \end{aligned} \tag{2.11}$$

Sotto l'azione di Φ si ha $(x, y) \mapsto (x', y') = \Phi(x, y)$, con

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} \bmod 1, \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}; \tag{2.12}$$

si veda la figura 2.3.

Con l'espressione "automorfismo algebrico di \mathbb{T}^n " si intende, in generale, una trasformazione del tipo (2.12), ove A è una qualunque matrice intera con $|\det A| = 1$ (cosa questa indispensabile perchè la trasformazione sia regolare e invertibile sul toro).

- **Osservazione.** Il sistema *non* è lineare, per via della riduzione al toro. Si osservi anche che vale la relazione

$$\Phi^t(x, y) = A^t(x, y) \bmod 1$$

(ovvero la riduzione al toro "si può rinviare alla fine").

²²⁾ L'insieme di livello definito da $F_i = c_i$ è una varietà proprio grazie all'indipendenza lineare.

²³⁾ Nelle figure di Arnol'd, alle quali ci uniformiamo, si seguono le deformazioni di un gatto tracciato all'interno. Prima di Arnol'd i gatti hanno ispirato Schrödinger e Einstein. "Un qualunque gatto è più geniale di qualsiasi genio umano" (H. Hesse).

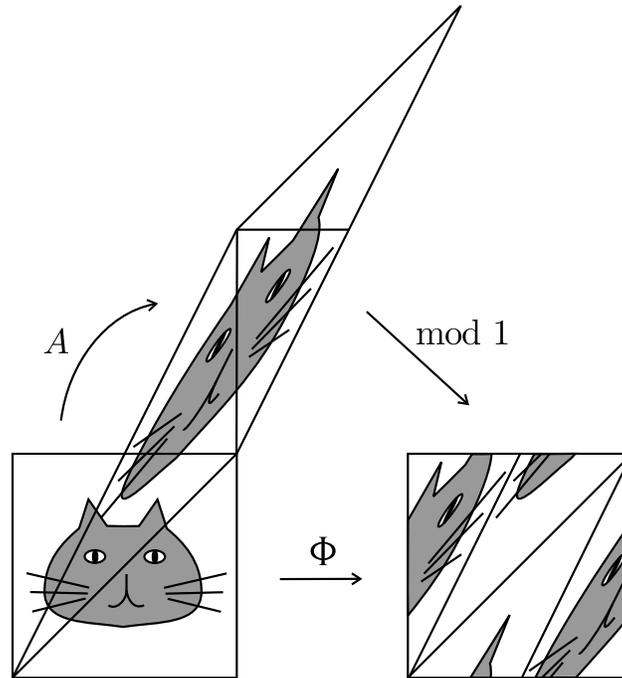


Figura 2.3. Il “Gatto di Arnol’d (esempio 2.5)."

Approfondiamo un po’ l’esempio, che è molto significativo. La matrice A ha autovalori non razionali

$$\lambda_1 = \lambda := \frac{3 + \sqrt{5}}{2} > 1, \quad \lambda_2 = \lambda^{-1} = \frac{3 - \sqrt{5}}{2} < 1;$$

gli autovettori corrispondenti sono

$$e_1 = (1, \lambda - 1), \quad e_2 = (1, \lambda^{-1} - 1),$$

perciò hanno pendenza irrazionale, e le rette E_1 e E_2 corrispondenti, riportate al toro, diventano curve dense sul toro (come segue dall’esempio 2.1); si veda la figura 2.4.

Ora, un qualunque insieme $A \subset \mathbb{T}^2$, ad esempio un dischetto di diametro d (è più facile immaginarlo per un intorno dell’origine), per effetto dell’applicazione di A^t , $t > 0$, diviene un “nastrino” (un’ellisse sottile) di larghezza $\lambda^{-t}d$ e lunghezza $\lambda^t d$, schiacciato lungo E_1 ; la riduzione al toro lo porta allora a invadere densamente il toro (per $t < 0$ il ruolo di E_1 e E_2 si scambia). La figura 2.5 mostra l’azione della mappa Φ su 20 000 punti inizialmente presi a caso in un quadrato vicino all’origine. Questo è il meccanismo tipico che, pur con una dinamica reversibile, consente un comportamento irreversibile per gli insiemi di punti.

Questo sistema è anche il prototipo dei cosiddetti *sistemi espansivi*: punti vicini si separano esponenzialmente (nel passato o nel futuro; genericamente, nel passato e nel futuro), e *non esistono coppie di orbite $\Phi^t(x)$ e $\Phi^t(y)$ che restino vicine per ogni $t \in \mathbb{Z}$, se $x \neq y$.*

- **Esercizio.** Si mostri che il sistema possiede infinite orbite periodiche dense in \mathbb{T}^2 (suggerimento: tutti i punti con coordinate razionali sono periodici).

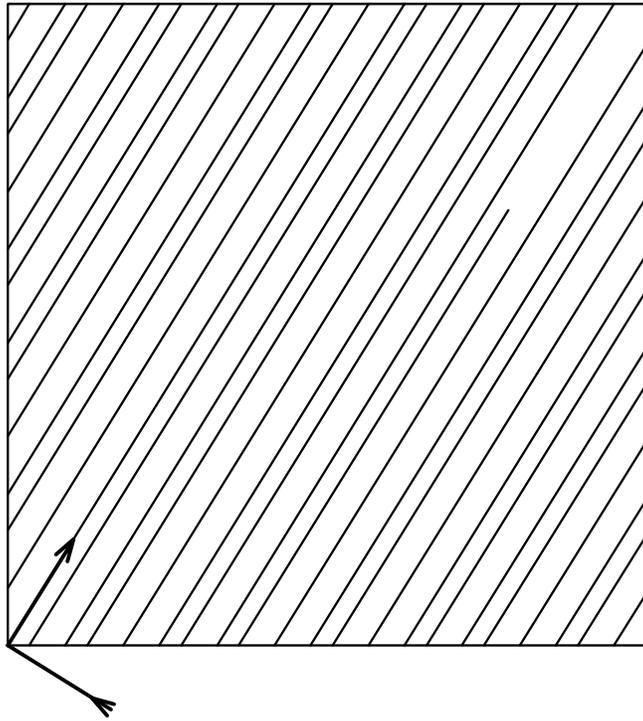


Figura 2.4. Gli autovettori e_1 e e_2 ; la retta E_1 , riportata al toro.

Un endomorfismo (un sistema non invertibile) classico, con proprietà analoghe agli automorfismi algebrici del toro, è dato dal seguente

Esempio 2.6. Prendiamo

$$M := \mathbb{T}^1 := \mathbb{R}/\mathbb{Z}$$

$\mu =$ misura di Lebesgue

$$\Phi(x) = 2x \bmod 1 .$$

La mappa agisce su \mathbb{T}^1 come indicato in figura 2.6 (cioè nel modo in cui abitualmente si avvolge su se stesso un elastico); si osservi che è regolare sul toro, e che preserva la misura (l'immagine *inversa* di un insieme mantiene la misura). Una trasformazione con proprietà simili, ma regolare solo a tratti, è la trasformazione “a tenda” (si faccia il disegno) dell'intervallo $[0, 1]$ in sé, definita da

$$\Phi(x) = \begin{cases} 2x & \text{per } x \leq 1/2 \\ 2 - 2x & \text{per } x > 1/2 . \end{cases}$$

L'esempio 2.6 è analogo al seguente più importante esempio, in cui con l'aggiunta di una coordinata la trasformazione è resa invertibile:

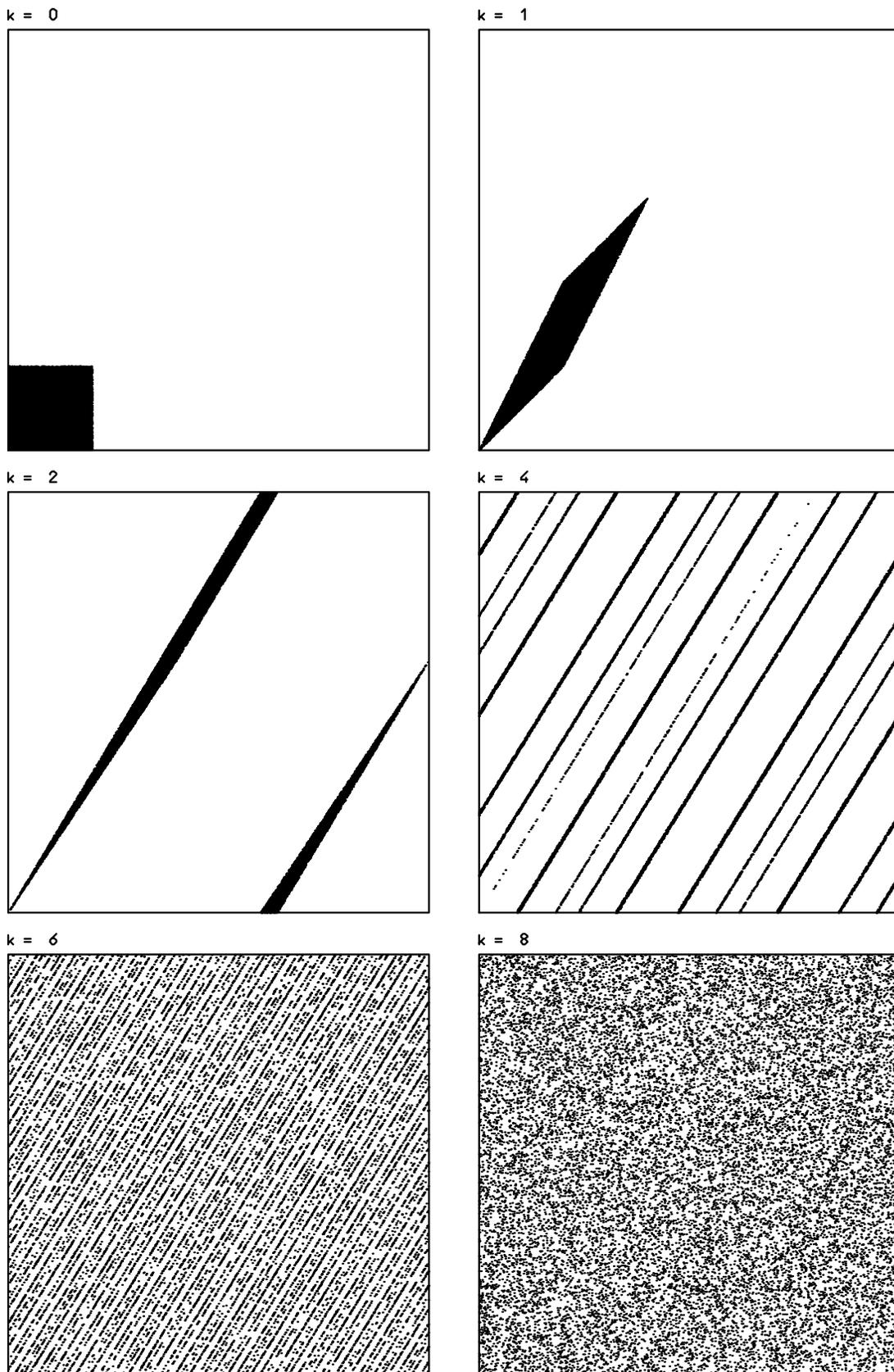


Figura 2.5. Evoluzione di 20 000 punti nel Gatto di Arnol'd.

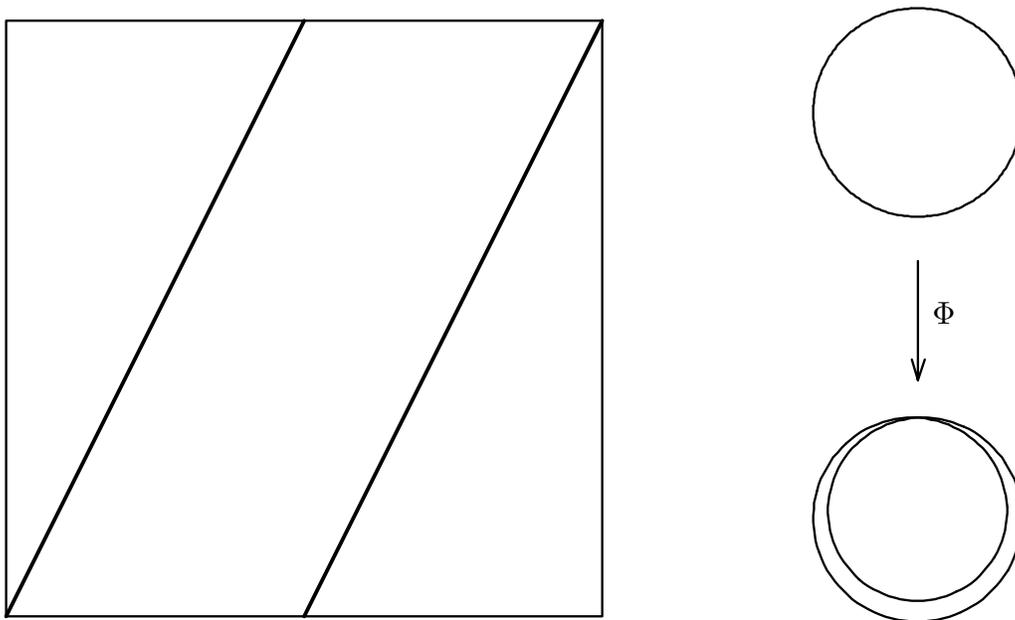


Figura 2.6. L'esempio 2.6

Esempio 2.7 (Trasformazione del panettiere). Prendiamo

$$\begin{aligned}
 M &= [0, 1) \times [0, 1) \\
 \mu &= \text{misura di Lebesgue} \\
 \Phi(x, y) &= \begin{cases} (2x, \frac{1}{2}y) & \text{per } x < \frac{1}{2} \\ (2x - 1, \frac{1}{2}y + \frac{1}{2}) & \text{per } x \geq \frac{1}{2} \end{cases} .
 \end{aligned}
 \tag{2.13}$$

“la pasta è schiacciata fino a dimezzare l'altezza e raddoppiare la lunghezza; poi è tagliata e sovrapposta” (si veda la figura 2.7).

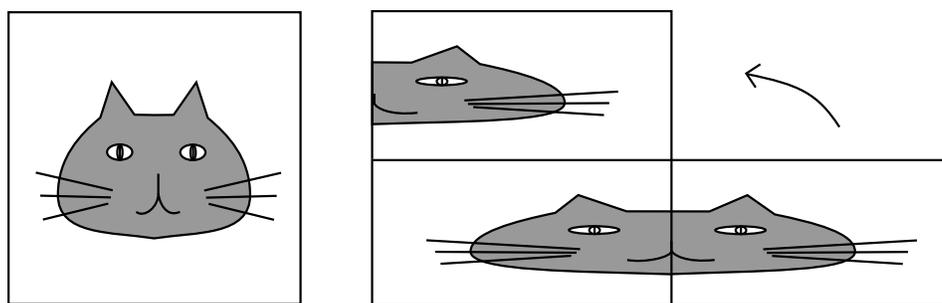


Figura 2.7. La trasformazione del panettiere.

Il prossimo esempio, di notevole importanza, è un sistema astratto.

Esempio 2.8 (Schemi di Bernoulli). Consideriamo un insieme finito I di simboli (*alfabeto*); per essere definiti, poniamo

$$I = \{0, 1, \dots, n - 1\} .$$

Consideriamo poi lo spazio $M = I^{\mathbb{Z}}$ delle sequenze doppiamente infinite di simboli dell'alfabeto:

$$M = \{x = (\dots, x_{-1}, x_0, x_1, \dots), x_i \in I\} \quad (2.14)$$

(l'idea da seguire è quella dei possibili esiti di infiniti lanci di un dado a n facce), e diamo a M la struttura di spazio di Borel, cioè costruiamo su M una σ -algebra di insiemi misurabili, prendendo come generatori²⁴⁾ i *cilindri elementari*

$$\mathcal{C}_k^l := \{x \in M : x_k = l\}, \quad k \in \mathbb{Z}, \quad l \in I \quad (2.15)$$

(insieme delle sequenze con esito assegnato l al k -esimo lancio). Per costruire la misura, prendiamo n numeri reali positivi a somma 1:

$$p_0, \dots, p_{n-1} > 0, \quad \sum_{l \in I} p_l = 1 \quad (2.16)$$

(probabilità *a priori* delle n facce del dado), e poniamo

$$\mu(\mathcal{C}_k^l) = p_l \quad (2.17)$$

(la probabilità di aver esito l al k -esimo lancio dipende solo da l e non da k). Si stabilisce poi che le intersezioni dei generatori, ovvero i *cilindri*

$$\mathcal{C}_{k_1, \dots, k_m}^{l_1, \dots, l_m} := \bigcap_{j=1}^m \mathcal{C}_{k_j}^{l_j},$$

abbiano misura fattorizzata

$$\mu(\mathcal{C}_{k_1, \dots, k_m}^{l_1, \dots, l_m}) = p_{l_1} \cdots p_{l_m}$$

(ovvero, i diversi lanci sono eventi casuali indipendenti). Alle unioni disgiunte di cilindri si assegna quindi (obbligatoriamente) come misura la somma delle misure, e con ciò la misura resta assegnata su di un anello; un teorema garantisce allora l'estensione univoca all'intera σ -algebra degli insiemi misurabili. Per la (2.16), la misura risulta normalizzata:

$$M = \bigcup_{l \in I} \mathcal{C}_k^l, \quad \mu(M) = \sum_{l \in I} p_l = 1.$$

Infine, definiamo l'automorfismo Φ su M come traslazione di un posto, o *shift*, verso sinistra: precisamente, a

$$x = (\dots, x_{-2}, x_{-1}, x_0, x_1, x_2, \dots)$$

associamo $x' := \Phi(x)$ definito da

$$x'_k = x_{k+1}. \quad (2.18)$$

²⁴⁾ Ricordiamo che la σ -algebra generata da una famiglia di insiemi è la σ -algebra minimale che li contiene.

Il sistema dinamico astratto (M, μ, Φ) così costruito si chiama *schema di Bernoulli*, e si denota, ricordando anche le probabilità con cui è costruito, $B_{p_0, \dots, p_{n-1}}$. Gli schemi di Bernoulli sono modelli adatti a descrivere successioni di eventi indipendenti, ciascuno con una assegnata probabilità *a priori* (il più semplice schema di Bernoulli, $B_{\frac{1}{2}, \frac{1}{2}}$, corrisponde al gioco di testa e croce; $B_{\frac{1}{6}, \dots, \frac{1}{6}}$ corrisponde al lancio di un dado onesto).²⁵⁾

Esempio 2.9 (Flusso geodetico). Data una varietà Riemanniana²⁶⁾ \mathcal{Q} , di dimensione n , poniamo

$$M = \{(q, v) : q \in \mathcal{Q}, v \in T_q \mathcal{Q}, \|v\| = 1\};$$

si potrebbe vedere che M ha la struttura di varietà, precisamente è la sottovarietà del *fibrato tangente* $T\mathcal{Q}$ corrispondente a $\|v\| = 1$ (fibrato tangente unitario). Per ogni punto $(q, v) \in M$ è ben definita la geodetica di \mathcal{Q} passante per q e tangente a v . Il flusso $\{\Phi^t, t \in \mathbb{R}\}$ su M si definisce allora ponendo $\Phi^t(q, v) =$ “trasporto lungo la geodetica, per una distanza t .” La misura conservata $d\mu$ risulta essere il prodotto $dv d\sigma$ della misura dv indotta dalla metrica su \mathcal{Q} e dell’area euclidea $d\sigma$ sulla sfera unitaria S^{n-1} .

Gli esempi da tenere a mente sono il moto libero con velocità unitaria di un punto materiale su una ordinaria superficie bidimensionale liscia, o più in generale il moto di un sistema lagrangiano a n gradi di libertà, in assenza di forze attive: \mathcal{Q} è allora la varietà vincolare, e la matrice cinetica fornisce la metrica. Si ricordi che il principio variazionale di Maupertuis garantisce proprio che il moto avviene lungo le geodetiche di \mathcal{Q} , con la metrica data dall’energia cinetica.

- **Esercizio.** Si studi con il formalismo lagrangiano (coordinate ignorabili) il flusso geodetico sul toro bidimensionale immerso in \mathbb{R}^3 ; le equazioni parametriche della superficie sono

$$x = (R + r \cos \theta) \cos \varphi, \quad y = (R + r \cos \theta) \sin \varphi, \quad z = r \sin \theta.$$

²⁵⁾ La fattorizzazione della misura è una caratteristica peculiare degli schemi di Bernoulli. Sullo stesso spazio M , con la stessa σ -algebra (generata dai cilindri), e la stessa dinamica (lo *shift*), si possono introdurre altre misure invarianti, tra cui le *misure di Markov*, nelle quali sono assegnate, oltre alle probabilità p_l di ciascun simbolo l (con $\sum_l p_l = 1$) le “probabilità di transizione” $\mathcal{P}_{jl} \geq 0$ dal simbolo j al simbolo l , con il requisito di normalizzazione $\sum_l \mathcal{P}_{jl} = 1$ e di compatibilità $\sum_j p_j \mathcal{P}_{jl} = p_l$ (negli schemi di Bernoulli si ha $\mathcal{P}_{jl} = p_l$, e le condizioni sono soddisfatte). Si potrebbe vedere che per ogni matrice \mathcal{P} positiva con $\sum_l \mathcal{P}_{jl} = 1$ esiste sempre (e in casi rilevanti è unica) una scelta compatibile delle p_l . Le probabilità p_l e le probabilità di transizione \mathcal{P}_{jl} consentono di definire la misura di tutti i cilindri (si pone $\mu(C_{k, k+1}^{j, l}) = p_j \mathcal{P}_{jl}$, e così via), e con esse la misura μ su M . I sistemi dinamici così definiti sono detti *sistemi di Markov*; $\mathcal{P} := (\mathcal{P}_{ij})$ è detta *matrice di transizione* del sistema. Mentre nei sistemi di Bernoulli c’è totale scorrelazione tra eventi successivi (l’esito di un lancio di dado non è influenzato dalla storia passata), nei sistemi di Markov ogni evento è correlato all’evento immediatamente precedente, ma solo ad esso. Si pensi a una passeggiata su una scacchiera, ad esempio con “passo di Re” e uguale probabilità di transire da una casa a tutte le case adiacenti. Detto m_j il numero di case adiacenti alla casa j , si ha allora $\mathcal{P}_{jl} = 1/m_j$ se l è adiacente a j , e zero altrimenti; la scelta buona delle p_l si verifica facilmente essere $p_l = m_l / \sum_j m_j$.

²⁶⁾ Ricordiamo che una varietà riemanniana è una varietà \mathcal{Q} , su ciascun punto della quale è assegnata una forma bilineare g (regolare) simmetrica e definita positiva, detta metrica, che agisce su coppie di vettori tangenti. In coordinate la forma si scrive $g = \sum_{ij} g_{ij}(x) dx_i dx_j$, con (g_{ij}) matrice simmetrica definita positiva. Su ciascuno spazio tangente resta definito il prodotto scalare $\xi \cdot \eta := g(\xi, \eta)$, in coordinate $\sum_{ij} g_{ij}(x) \xi_i \eta_j$. Restano poi definite lunghezze di curve e aree (volumi) di ogni dimensione su \mathcal{Q} ; sono ben definite le geodetiche su \mathcal{Q} .

Concludiamo questa breve rassegna di esempi illustrando una classe notevole di sistemi dinamici, i cosiddetti *biliardi*; introdotti già da Birkhoff negli anni '30, questi sistemi dinamici hanno avuto (e hanno ancora) un ruolo notevole nella teoria ergodica.

Esempio 2.10 (Biliardi). Sia Q una regione limitata e connessa di \mathbb{R}^2 , delimitata da una curva Γ regolare a tratti (Γ può essere composta di più tratti connessi Γ_i , se Q è moltiplicemente connessa: “biliardo con ostacoli”). Il sistema che vogliamo definire corrisponde al moto libero di un punto materiale (o di un raggio di luce) in Q , con velocità costante di modulo uno all'interno, e riflessione elastica sul bordo Γ (la componente tangente della velocità si conserva, la componente perpendicolare si inverte); si tratta di una estrapolazione, in un certo senso, del flusso geodetico.²⁷⁾ Poniamo allora $M = Q \times S^1$, e definiamo Φ^t , continuo a tratti, secondo quanto detto sopra: in linea retta fino al bordo, poi riflessione, poi in linea retta e così via, per una lunghezza complessiva t (figura 2.8). Si vede facilmente che è invariante la misura $d\mu = dx dy d\theta$, con x, y coordinate cartesiane in Q e θ angolo tra il vettore velocità e una qualsiasi direzione prefissata (μ è infatti invariante sia durante la traslazione, sia per la mappa corrispondente alla riflessione); resta così definito un sistema dinamico classico (M, μ, Φ) , regolare a tratti.

A (M, μ, Φ) conviene tuttavia associare un secondo sistema dinamico discreto, fissando l'attenzione sulle collisioni (che sono l'unica cosa interessante: tra una collisione e l'altra il moto è banale). Per questo introduciamo una coordinata curvilinea γ sul bordo Γ (o su ciascuna Γ_i), con origine fissata e verso antiorario, e identifichiamo una collisione con le coordinate γ e $\lambda = \cos \alpha$, ove α è l'angolo tra la tangente orientata a Γ in γ e la direzione della velocità in uscita dalla collisione (aggiungiamo l'indice discreto i , se Γ è composta di più parti connesse). Poniamo quindi $N = \Gamma \times [-1, 1]$, e sia $\Psi : N \rightarrow N$ la mappa che associa a ogni collisione la collisione successiva. Con un po' di pazienza si calcola la matrice jacobiana $J = \frac{\partial \Psi(\gamma, \lambda)}{\partial(\gamma, \lambda)}$ associata alla mappa, e si trova $\det J = 1$; l'area di N pertanto si conserva, e la misura normalizzata $d\nu = \frac{1}{2L} d\gamma d\lambda$ risulta invariante. Abbiamo così costruito un sistema dinamico classico discreto (N, ν, Ψ) , chiamato anch'esso biliardo, o mappa del biliardo.

- **Esercizio.** Si verifichi che si ha

$$J = \begin{pmatrix} (cl - \sigma)/\sigma' & l/(\sigma\sigma') \\ cc'l - c\sigma' - c'\sigma & (c'l - \sigma')/\sigma \end{pmatrix},$$

ove l è la lunghezza del tratto rettilineo tra una collisione e la successiva, c, c' indicano la curvatura del bordo nel punto di partenza e di arrivo (positivi per un biliardo convesso), mentre $\sigma = \sin \alpha$, $\sigma' = \sin \alpha'$.

L'interesse per i biliardi deriva dalla loro semplicità, e dal fatto che al variare della forma di Γ si ottengono le proprietà ergodiche più diverse. Una variante alla definizione data sopra è quella in cui la tavola del biliardo è il toro \mathbb{T}^2 , con uno o più ostacoli, o equivalentemente il piano, con ostacoli disposti in modo periodico.²⁸⁾

²⁷⁾ Ciò risulta più chiaro se pensiamo la superficie Q a due facce, con cambio di faccia a ogni riflessione: il punto materiale “passa dietro.”

²⁸⁾ Si studia anche il cosiddetto *Gas di Lorenz*, cioè il moto di un punto in un piano con ostacoli disposti in modo casuale.

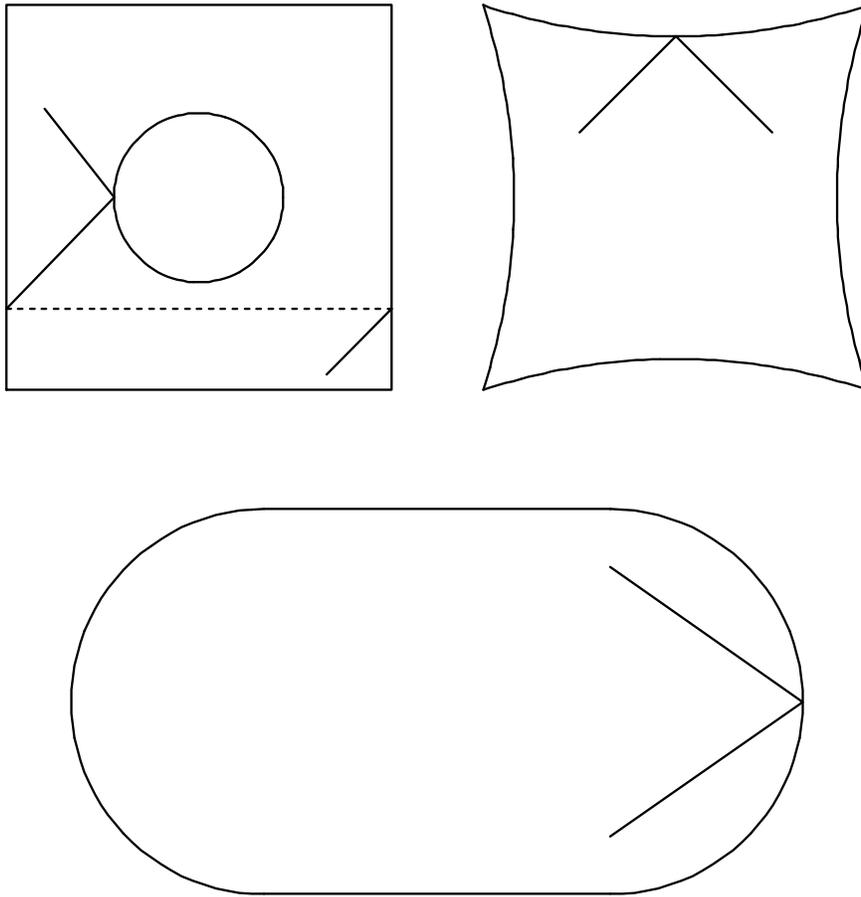


Figura 2.8. Esempi significativi di biliardo.

- **Esercizio.** Si studino le singolarità di Ψ , per biliardi di varia forma (si osservi che le singolarità possono essere presenti anche se il bordo è regolare e connesso).

Chiudiamo questo paragrafo introduttivo con la nozione di *sezione di Poincaré*, cui faremo riferimento di frequente nel seguito. Siano dunque $\Phi = \{\Phi^t, t \in \mathbb{R}\}$ un flusso su una varietà riemanniana M , e N una *sezione globale* di M , ovvero una sottovarietà orientata di M tale che tutte le orbite di N prima o poi la attraversino in un verso prefissato. Si vede subito che il flusso Φ induce una mappa $\Psi : N \rightarrow N$, precisamente la mappa che associa a $x \in N$ la prima intersezione con N (nel verso prefissato) dell'orbita $\Phi^t(x)$; la mappa $\Psi : N \rightarrow N$ è detta *sezione di Poincaré* (globale) del flusso Φ su M . Se Φ preserva una misura μ su M di densità $\rho(x)$, ovvero $d\mu = \rho dV$, si dimostra facilmente che Ψ a sua volta preserva la misura $d\nu = C\rho|v_\perp|d\sigma$, ove $d\sigma$ indica l'area su N , mentre $|v_\perp|$ denota la componente di $v = \dot{x}$ perpendicolare a N , e la costante C serve alla normalizzazione.

- **Esercizio.** Si faccia la dimostrazione.

Al sistema dinamico classico continuo (M, μ, Φ) resta allora associato il sistema dinamico classico discreto (N, ν, Ψ) , ad esso sostanzialmente equivalente, detto a sua volta *sezione di Poincaré* di (M, μ, Φ) . Nel caso dei biliardi, la mappa del biliardo sopra definita è sostanzialmente la sezione di Poincaré del flusso.

2.2. Isomorfismo tra sistemi dinamici

Concludiamo questa introduzione generale ai sistemi dinamici con la nozione di *isomorfismo* tra sistemi dinamici, classici e/o astratti.

Definizione. Due sistemi dinamici (M, μ, Φ) e (N, ν, Ψ) si dicono *isomorfi* se esiste una mappa $h : M \rightarrow N$, definita e invertibile quasi ovunque, che: (1) preserva la misura, ovvero si ha

$$\nu(h(A)) = \mu(A) , \quad \mu(h^{-1}(B)) = \nu(B) , \quad (2.19)$$

per tutti gli insiemi misurabili $A \subset M$ e $B \subset N$; (2) commuta con la dinamica, ovvero per ogni t si ha

$$\Psi^t \circ h = h \circ \Phi^t . \quad (2.20)$$

La nozione di isomorfismo è importante, perchè pone le basi per una classificazione dei sistemi dinamici (si osservi che è una relazione di equivalenza). Tutte le proprietà che andremo a introdurre sono invarianti per isomorfismo. Sono ad esempio isomorfi due sistemi hamiltoniani coniugati da una trasformazione canonica. Non è difficile trovare esempi di sistemi dinamici classici isomorfi a sistemi astratti (il che mostra che la distinzione tra sistemi classici e astratti non è profonda). Precisamente,

Proposizione 2.6. *La trasformazione del panettiere e lo schema di Bernoulli $B_{\frac{1}{2}, \frac{1}{2}}$ sono isomorfi.*

Dimostrazione. Indichiamo con (M, μ, Φ) e rispettivamente (N, ν, Ψ) i due sistemi, e denotiamo con $x = (a, b)$ i punti di M , con y le sequenze di N . Se in scrittura binaria si ha

$$a = 0.a_0 a_1 a_2 \dots , \quad b = 0.b_0 b_1 b_2 \dots , \quad a_i, b_i \in I = \{0, 1\} ,$$

allora si definisce h associando a $x \in M$ il punto $y = h(x) \in N$, dato da

$$\begin{aligned} y &= (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots) \\ &= (\dots, b_1, b_0, a_0, a_1, a_2, \dots) , \end{aligned}$$

ovvero si pone

$$y_i = a_i \quad \text{per } i \geq 0 , \quad y_i = b_{-i-1} \quad \text{per } i < 0 .$$

La corrispondenza è definita e biunivoca quasi ovunque;²⁹⁾ si verifica facilmente che h commuta con la dinamica, e che preserva la misura (come sono fatte le immagini dei cilindri?). ■

- **Osservazione.** L'isomorfismo h sopra introdotto si può definire anche nel modo seguente, che risulta particolarmente significativo: introduciamo in M una partizione in due insiemi A_0 e A_1 “dividendo verticalmente a metà il quadrato”, precisamente poniamo

$$A_0 = \{x = (a, b) \in M : a < \frac{1}{2}\} , \quad A_1 = M \setminus A_0 .$$

²⁹⁾ Non ovunque: si ricordi l'ambiguità della scrittura decimale o binaria.

Consideriamo poi $\Phi^t(x)$, $t \in \mathbb{Z}$, e associamo a x la sequenza (“storia di x ”)

$$y = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots), \quad y_t = l \quad \text{se} \quad \Phi^t(x) \in A_l. \quad (2.21)$$

Non è difficile convincersi che la corrispondenza così realizzata è proprio l’isomorfismo h definito sopra (si ricordi che $a < \frac{1}{2}$ se la prima cifra dello sviluppo binario di a è 0). Una conseguenza dell’isomorfismo è che comunque ci si inventi una stringa y di simboli 0 e 1, esiste sempre un punto $x \in M$ che la realizza, nel senso che $\Phi^t(x)$ sta in A_0 o in A_1 a seconda che in posizione t della stringa si sia scelto il simbolo 0 o 1. La generalizzazione di questo esempio semplice conduce alla nozione, che ha grande importanza nella teoria dei sistemi dinamici e sarà ripresa più avanti, di *dinamica simbolica*.

- **Esercizio.** Si consideri la quadrettatura di M in celle di lato 2^{-k} , $k > 0$, e si ponga $\tilde{\Phi} = \Phi^{2^k}$. Si dimostri che scelta a piacere una passeggiata sulle celle (ovvero una arbitraria successione di celle) esiste $x \in M$ tale che $\tilde{\Phi}^t(x)$, $t \in \mathbb{Z}$, la esegua fedelmente. ³⁰⁾

3. Due teoremi generali

Vi sono due teoremi di carattere del tutto generale, che valgono per sistemi dinamici qualsiasi classici o astratti, continui o discreti, anche non invertibili: il *teorema della ricorrenza* di Poincaré, e il *teorema ergodico* di Birkhoff–Kinchin.

3.1. Il teorema della ricorrenza

Per introdurre il teorema della ricorrenza, premettiamo la seguente

Definizione. Per $A \subset M$ misurabile, diciamo che $x \in A$ è *ricorrente*³¹⁾ in A se per ogni $T > 0$ esiste $t \geq T$ tale che $\Phi^t(x) \in A$.

L’insieme dei punti ricorrenti in A sarà denotato R_A ; il complementare $A \setminus R_A$, denotato V_A e detto insieme dei punti *vaganti* di A , è allora dato da

$$V_A = \{x \in A : \exists T > 0 : \Phi^t(x) \notin A \forall t \geq T\}.$$

³⁰⁾ Se M è un foglio con il lato di un metro, e $k = 15$, si ottiene l’invidiabile risoluzione di 32 768 punti per metro (oltre 800 punti per pollice). Scelto a piacere il disegno preferito, ad esempio la pianta di Padova oppure Dante e Virgilio in una illustrazione del Doré, esiste $x \in M$ tale che gli iterati $\Phi^{30t}(x)$, $t = 0, 1, 2, \dots$ lo riproducono esattamente entro la risoluzione. Anzi: possiamo anche prefissare l’ordine in cui i puntini vanno a comporre il disegno. Ma con quanta precisione va determinato il dato iniziale?

³¹⁾ Si trova frequentemente in letteratura una definizione leggermente più ampia: sono chiamati ricorrenti i punti appartenenti alla chiusura di R_A qui definito, e corrispondentemente vaganti i punti interni a V_A ; il teorema della ricorrenza afferma allora che per i sistemi che conservano una misura l’insieme dei punti vaganti è vuoto.

Proposizione 3.1 (Teorema della ricorrenza). *Sia (M, μ, Φ) un qualsiasi sistema dinamico. Per ogni $A \subset M$ misurabile l'insieme V_A è misurabile e ha misura nulla.*

Dimostrazione. Possiamo limitarci al caso discreto; il caso continuo si riporta banalmente al caso discreto considerando la mappa al tempo uno (l'insieme dei punti vaganti, restringendo t agli interi, al più si allarga). Per ogni $T \in \mathbb{N}^+$ poniamo

$$V_{T,A} = \{x \in A : \Phi^t(x) \notin A \forall t \geq T\}$$

(punti “ T -vaganti” di A); si ha evidentemente

$$V_A = \bigcup_{T \in \mathbb{N}^+} V_{T,A} ,$$

dunque basta dimostrare che per ogni $A \subset M$ misurabile e ogni $T \in \mathbb{N}^+$ l'insieme $V_{T,A}$ è misurabile e ha misura nulla. La misurabilità di $V_{T,A}$ segue dalla definizione, in base alla quale si può scrivere

$$V_{T,A} = A \bigcap \left[\bigcap_{t \geq T} \Phi^{-t}(M \setminus A) \right] .$$

Consideriamo allora gli insiemi

$$V_{T,A} , \Phi^{-T}(V_{T,A}) , \Phi^{-2T}(V_{T,A}) , \dots$$

Essi sono necessariamente disgiunti: infatti se

$$x \in \Phi^{-kT}(V_{T,A}) \bigcap \Phi^{-lT}(V_{T,A}) \neq \emptyset ,$$

ad esempio con $k < l$, allora

$$\Phi^{kT}(x) \in V_{T,A} \bigcap \Phi^{-(l-k)T}(V_{T,A}) ,$$

contro la definizione stessa di $V_{T,A}$. Per la conservazione della misura, essendo finita la misura dell'intero spazio M , risulta $\mu(V_{T,A}) = 0$. ■

In base al teorema della ricorrenza, tutti i punti a meno di un insieme di misura nulla ritornano infinite volte in un qualunque fissato insieme A di misura positiva che li contenga (ad esempio una palla di raggio ε arbitrario).

- **Osservazione.** I tempi di ricorrenza, stimati sulla base della dimostrazione, sono tuttavia assurdamente grandi, per sistemi anche molto semplici. Si consideri un gas di N punti materiali in un cubo di lato L ; supponiamo (ad esempio) che i punti si muovano indipendentemente, e che ciascuno rimbalzi sulle pareti spostandosi con velocità costante v . Sia A l'insieme ottenuto specificando la sola posizione di ciascun punto, con precisione εL ; si prenda $T =$ tempo tipico di uscita da A , $T = \varepsilon L / (Nv)$ (è il più piccolo tempo in cui “succede qualche cosa”). Allora la dimostrazione “obbliga” il sistema alla ricorrenza, dopo che $A, \Phi^{-T}(A), \Phi^{-2T}(A), \dots$ “non ci stanno più”, dunque dopo il tempo $t = kT$ con $k = 1/\mu(A) = \varepsilon^{-3N}$. Si trova $t = \varepsilon^{-3N+1} L / (Nv)$,

e ad esempio con $N = 10$, $L = 1$ m, $\varepsilon = 10^{-3}$, $v = 100$ m/sec, si ha $t = 10^{84}$ sec (l'età dell'universo, secondo le stime attuali, non supera 10^{18} sec).

- **Osservazione.** A parte questa considerazione “fisica”, è bene sottolineare che il comportamento ricorrente dei singoli moti non preclude il comportamento irreversibile degli insiemi di punti (o di “stati macroscopici” nel senso di Gibbs) tipico dei sistemi mescolanti di cui si dirà tra breve.

3.2. Il teorema ergodico di Birkhoff–Kinchin

Veniamo ora al teorema ergodico di Birkhoff–Kinchin, premettendo anche qui qualche definizione.

Definizione. Per ogni funzione $f : M \rightarrow \mathbb{R}$, la funzione $\bar{f} : M \rightarrow \mathbb{R}$ definita da

$$\bar{f}(x) := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) , \quad (3.1)$$

o nel caso continuo

$$\bar{f}(x) := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\Phi^s(x)) ds , \quad (3.2)$$

è detta, se esiste, *media temporale* della funzione f .

Ad esempio, il tempo medio di soggiorno di un'orbita in un insieme misurabile A ,

$$\tau_A(x) := \lim_{t \rightarrow \infty} \frac{1}{t} \mathcal{T}_A(x, t) , \quad \mathcal{T}_A(x, t) := \int_0^t \chi_A(\Phi^s(x)) ds , \quad (3.3)$$

ove χ_A denota la funzione caratteristica di A :

$$\chi_A(x) = \begin{cases} 1 & \text{per } x \in A \\ 0 & \text{per } x \notin A , \end{cases} \quad (3.4)$$

è precisamente la media temporale di χ_A .

La media temporale di una funzione, oltre ad avere interesse per la meccanica statistica, gioca un ruolo importante ogniqualevolta, in un sistema dinamico, convivono variabili “lente” e “veloci”: ad esempio in Meccanica Celeste sono veloci le fasi dei pianeti sulle ellissi kepleriane, sono lente le variabili associate agli elementi orbitali (semiassi delle ellissi, eccentricità, inclinazione); oppure, veloce è la rotazione della terra su se stessa, lenta la variazione dell'inclinazione dell'asse di rotazione. Come avremo modo di discutere più avanti, in prima approssimazione (“a meno di effetti del secondo ordine”) l'effetto delle variabili veloci sulle variabili lente coincide con un effetto medio, opportunamente definito per mezzo di una media temporale. Ad esempio, l'effetto della perturbazione di un pianeta sugli elementi orbitali degli altri è quello che si avrebbe se la massa del pianeta fosse distribuita lungo l'orbita, in proporzione al tempo medio di soggiorno in ogni tratto.

La media temporale \bar{f} di una funzione f è a sua volta una funzione; è invece un numero la *media in fase* $\langle f \rangle$ di f , definita per ogni f sommabile, ovvero $f \in L_1(M, \mu)$, da

$$\langle f \rangle := \int_M f d\mu . \quad (3.5)$$

A differenza della media temporale, in linea di principio assai difficile da calcolare perché presuppone la conoscenza del moto, la media in fase è “soltanto” un integrale, che in linea di principio si deve pensare di saper calcolare, e che comunque non è mai³²⁾ difficile, in qualche modo, valutare almeno approssimativamente, eventualmente ricorrendo a metodi numerici.

E’ spontaneo chiedersi per quali sistemi dinamici, per quali funzioni, e quali dati iniziali, la media temporale esista. La risposta, tutt’altro che scontata, è che essa esiste per ogni sistema dinamico, ogni funzione sommabile, e quasi ovunque; ciò è assicurato dal fondamentale teorema di Birkhoff–Kinchin, che qui enunciamo nel linguaggio del caso discreto. La dimostrazione (in realtà non molto istruttiva) è rinviata all’appendice B.

Proposizione 3.2 (Teorema ergodico di Birkhoff–Kinchin). *Sia (M, μ, Φ) un sistema dinamico astratto discreto, non necessariamente invertibile; sia $f : M \rightarrow \mathbb{R}$ sommabile. Allora esiste quasi ovunque in M il limite*

$$\bar{f}(x) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) , \quad (3.6)$$

e inoltre si ha

$$\bar{f}(\Phi(x)) = \bar{f}(x) , \quad \langle \bar{f} \rangle = \langle f \rangle . \quad (3.7)$$

Se il sistema è invertibile, allora anche il limite

$$\bar{f}_-(x) := \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^{-s}(x))$$

esiste quasi ovunque, e quasi ovunque coincide con $\bar{f}(x)$.

Vi sono casi in cui la media temporale esiste ovunque, anziché quasi ovunque: ad esempio nella traslazione di \mathbb{T}^1 (esempio 2.2), o nel moto uniforme su \mathbb{T}^n , purché la funzione f abbia un minimo di regolarità (basta che sia integrabile secondo Riemann; si veda il libro sopra citato di V.I. Arnol’d e A. Avez, appendice 9). Ma in generale, l’esistenza della media temporale è assicurata solo quasi ovunque.

Ad esempio, si consideri lo schema di Bernoulli $B_{\frac{1}{2}, \frac{1}{2}}$, e si prenda la (semplicissima) funzione caratteristica del cilindro \mathcal{C}_0^1 (ovvero: $f(x) = x_0$ per ogni $x = \{x_k, k \in \mathbb{Z}\}$), per la quale è immediato verificare che si ha

$$\frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) = \frac{1}{t} \sum_{s=0}^{t-1} x_s .$$

Si vede facilmente che per il dato iniziale “eccezionale”

$$x = (\dots, x_{-2}, x_{-1}, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, (16 \times 0), (32 \times 1), \dots) .$$

³²⁾ Quasi mai

il limite non esiste: infatti per t al termine di un blocco “1” la somma vale $\frac{2}{3}t$, mentre al termine del blocco “0” è minore di $\frac{1}{3}t$. La medesima situazione si ha, evidentemente, per la trasformazione del panettiere, prendendo il dato iniziale corrispondente.

- **Osservazione.** Dati iniziali “eccezionali” simili a quello sopra messo in evidenza, precisamente dati in cui la sequenza speciale di simboli comincia dalla posizione k (arbitrariamente grande) anziché da zero, e i rimanenti simboli sono qualsiasi, sono densi in una topologia naturale degli schemi di Bernoulli, di cui si dirà più avanti, equivalente per $B_{\frac{1}{2}, \frac{1}{2}}$ alla consueta topologia del quadrato nella trasformazione del panettiere. Per tali sistemi dunque la media temporale *non* è definita in nessun aperto. Da questa semplice osservazione si capisce come sia importante, e delicato, il ruolo della misura nello studio dei sistemi dinamici.

4. Ergodicità

L’ergodicità è la prima fondamentale e non banale proprietà che si incontra nella teoria ergodica; la si può vedere come una prima possibile formalizzazione del “problema ergodico” posto da Boltzmann e Gibbs (paragrafo 1.2 e 1.3) per giustificare l’uso della meccanica statistica nello studio dei sistemi termodinamici, benché la sua formulazione e il suo studio si pongano già per sistemi dinamici semplicissimi, che poco hanno a che fare con la termodinamica.

4.1. Definizione di sistema ergodico

Vi sono quattro proprietà esattamente equivalenti che si possono usare come definizione di sistema ergodico, e che vale la pena di prendere in considerazione (l’abbreviazione “q. o.”, qui e nel seguito, sta per “quasi ovunque”).

Proprietà E1. Per ogni funzione sommabile $f : M \rightarrow \mathbb{R}$, media temporale e media in fase coincidono:

$$\bar{f}(x) = \langle f \rangle \quad \text{q. o. in } M . \quad (4.1)$$

Proprietà E2. Per ogni insieme $A \subset M$ misurabile il tempo medio di soggiorno in A è uguale alla misura di A :

$$\tau_A(x) = \mu(A) \quad \text{q. o. in } M , \quad (4.2)$$

con $\tau_A(x)$ definito nella (3.3).

Proprietà E3. Non esistono integrali del moto sommabili non banali:

$$f(\Phi^t(x)) = f(x) \quad \forall t \quad \text{q. o. in } M \quad \implies \quad f \text{ costante q. o. in } M \quad (4.3)$$

per ogni $f : M \rightarrow \mathbb{R}$ sommabile.

Proprietà E4. Il sistema dinamico è *metricamente indecomponibile*, ovvero

$$\Phi^{-t}(A) = A \implies \mu(A) = 0 \text{ oppure } 1 \quad (4.4)$$

per ogni $A \subset M$ misurabile (detto altrimenti: ogni decomposizione $M = A \cup (M \setminus A)$ misurabile e invariante è metricamente banale).

- **Osservazione.** La nozione E1 è tra le più classiche, ed è alla base della definizione di ergodicità in diversi testi di meccanica statistica; risponde all'idea "pratica" di ergodicità, come proprietà che consente di sostituire alle medie temporali, difficili da calcolare, le medie in fase.

La proprietà E2 risponde all'idea (Boltzmann) che a un'osservazione fatta a un istante casuale corrisponda una probabilità di trovare lo stato microscopico del sistema in A , pari alla misura di A ; in questo senso, per un sistema che goda della proprietà E2, il volume assume il significato di probabilità. Nei sistemi dinamici classici tutti gli aperti sono misurabili e hanno misura positiva: la proprietà E2 implica pertanto che in tali sistemi le traiettorie generiche siano dense.

La proprietà E3 corrisponde invece all'unicità dell'equilibrio nel senso di Gibbs: se lo stato macroscopico ρ_t evolve con la legge

$$\rho_t(x) = \rho_0(\Phi^{-t}(x)) ,$$

allora l'unico stato di equilibrio, cioè tale che sia $\rho_t = \rho_0$ per ogni t , è quello uniforme, $\rho(x) = 1$ quasi ovunque (il che tuttavia non implica che lo stato di equilibrio sia in qualche senso raggiunto, a partire da ρ_0 qualsiasi).

Infine, la proprietà E4, introdotta da Birkhoff, è interessante per la sua semplicità geometrica, e torna utile in molte dimostrazioni.

- **Osservazione.** Ciascuna delle quattro proprietà è con evidenza invariante per isomorfismo.

Proposizione 4.1. *Le proprietà E1, ..., E4 sono equivalenti.*

Dimostrazione. Mostriamo qui che:

- | | | | |
|-----|----|------------|------|
| (a) | E1 | \implies | E2 |
| (b) | E2 | \implies | E4 |
| (c) | E4 | \implies | E1 |
| (d) | E4 | \iff | E3 . |

L'implicazione (a) è banale, quando si prenda $f = \chi_A$. L'implicazione (b) segue dal fatto che se, per assurdo, esiste A invariante e non banale, allora per $x \in A$ si ha $\tau_A(x, t) = t$, e dunque il tempo medio di soggiorno è 1, anziché $\mu(A)$. L'implicazione (c) si dimostra così: se, per assurdo, risulta (ad esempio) $\bar{f}(x) > \langle f \rangle$ in un insieme di misura positiva, allora per il teorema di Birkhoff–Kinchin l'insieme

$$A := \{x \in M : \bar{f}(x) > \langle f \rangle\}$$

è invariante (si usa la prima delle (3.7)) e non banale (si usa la seconda, per garantire che A è misurabile e $\mu(A) \neq 1$), contro l'assunzione E4. Infine, l'implicazione (d) segue in un verso perchè, se esistesse A invariante (non banale), allora la sua funzione caratteristica χ_A sarebbe integrale del moto (non costante quasi ovunque); nel verso opposto perchè, se esistesse un integrale del moto f (non costante quasi ovunque), allora con opportuna scelta di $c \in \mathbb{R}$ l'insieme

$$A := \{x \in M : f(x) \leq c\}$$

risulterebbe invariante (non banale). ■

- **Esercizio.** Si dimostrino altre implicazioni dirette tra le E1 – E4, ad esempio l'implicazione E1 \Leftrightarrow E3.

Definizione. Il sistema dinamico (M, μ, Φ) si dice *ergodico* se soddisfa una qualsiasi delle proprietà E1–E4.

- **Osservazione.** Come già si è detto, E2 è un caso particolare di E1, relativo alle funzioni caratteristiche. D'altra parte tali funzioni, con evidenza, appartengono allo spazio $L_2(M, \mu)$ delle funzioni a quadrato sommabile su M ; la verifica della proprietà E1 per le funzioni di $L_2(M, \mu)$ sarà dunque ampiamente sufficiente a garantire l'ergodicità. La stessa cosa si può dire per la proprietà E3.
- **Esercizio.** Sfruttando il fatto che le funzioni caratteristiche si approssimano a piacere in misura con funzioni continue, si dimostri che la proprietà E1 (o E3) per le sole funzioni continue implica l'ergodicità.
- **Osservazione.** Dal punto di vista del significato fisico, in particolare nell'ambito della meccanica statistica, l'ergodicità è una proprietà sicuramente importante, e tuttavia, si intuisce, allo stesso tempo troppo forte e troppo debole. Troppo forte, perchè riguarda la classe, enorme, delle funzioni sommabili: comprese quelle che ad esempio, in un gas di molecole identiche, differiscono in modo significativo per scambio di due molecole tra di loro. Troppo debole, perchè ancora non contiene né l'idea del raggiungimento dell'equilibrio, né tantomeno l'idea che la dinamica sia in qualche modo irregolare o, con una terminologia entrata recentemente nella letteratura scientifica, caotica (vedremo in un momento che, ad esempio, la traslazione di \mathbb{T}^1 , o il moto di un qualunque sistema hamiltoniano a un grado di libertà, sono sistemi ergodici). Per quanto riguarda la prima difficoltà, va segnalato il tentativo di Kinchin (si veda il suo libretto di meccanica statistica) di definire l'ergodicità “condizionata” a una classe di funzioni “fisicamente significative”. L'idea è stata occasionalmente ripresa, ma non ha mai avuto sviluppi veramente significativi.
- **Osservazione.** Non si creda tuttavia che insiemi misurabili “mal fatti” topologicamente, ad esempio insiemi chiusi con interno vuoto ma di misura positiva (la frontiera coincide con l'insieme, e ha misura positiva; la funzione caratteristica è discontinua su un insieme di misura positiva) siano privi di interesse fisico. Tali insiemi giocano anzi un ruolo molto importante nella teoria delle perturbazioni (teorema KAM), e si presentano naturalmente in connessione a sistemi hamiltoniani anche semplicissimi.³³⁾

³³⁾ Fermi stesso, in un lavoro peraltro molto interessante del 1921, ha sottovalutato questo problema,

Come tipico esempio di insiemi di questo tipo, si prenda l'intervallo $I = [0, 1]$, e numerati i razionali, si tolga un intervallo di lunghezza $2^{-1}\varepsilon$ attorno al primo, $2^{-2}\varepsilon$ attorno al secondo, e così via. L'unione di quello che si toglie ha misura inferiore a ε , ed è aperto denso; quello che resta ha misura superiore a $1 - \varepsilon$, ma ha interno vuoto. Gli insiemi che intervengono nel teorema KAM hanno proprio questa struttura (la mancanza di regolarità degli insiemi è dovuta al fatto che si guarda a proprietà asintotiche).

4.2. Esempi elementari

Vediamo allora qualche esempio elementare di sistema ergodico.

Proposizione 4.2. *Qualunque sistema hamiltoniano a un sol grado di libertà, su una superficie di energia costante Σ_E compatta connessa e priva di punti singolari, è ergodico.*

Dimostrazione. Con evidenza, per ogni dato iniziale in Σ_E l'orbita passa per ogni punto di Σ_E ; la proprietà E4 è soddisfatta e dunque il sistema è ergodico. ■

- **Esercizio.** Si dimostri l'ergodicità facendo riferimento alla proprietà E2 [*Suggerimento:* si scriva il tempo di soggiorno in un intervallo I come $\int_I dl/\|v\|$, ove dl e v denotano la lunghezza d'arco e la velocità nel piano di fase].

E' dunque ergodico, in particolare, il singolo oscillatore armonico, $H(p, q) = \frac{1}{2}(p^2 + \omega^2 q^2)$, su ciascuna delle curve di energia costante. Non è invece ergodico, sulla superficie di energia costante, un sistema di due o più oscillatori armonici, $H = \frac{1}{2} \sum_{i=1}^n (p_i^2 + \omega_i^2 q_i^2)$, o più in generale un sistema del tipo

$$H(p, q) = \sum_{i=1}^n h_i(p_i, q_i) , \quad (4.5)$$

in cui ognuna delle h_i dipende solo dalle sue variabili: infatti, le energie delle singole componenti sono integrali del moto, e si va contro la proprietà E3. Non è ergodico, evidentemente, un sistema integrabile a $n \geq 2$ gradi di libertà, dal momento che possiede n integrali del moto non banali.³⁴⁾ Per sistemi di questo tipo si può discutere di ergodicità solo restringendosi ai fogli in cui tutte le azioni sono conservate; come già osservato, ci si riporta allora allo studio del moto uniforme su \mathbb{T}^n , le cui proprietà ergodiche studieremo tra un momento. Sistemi del tipo (4.5) pongono tuttavia una domanda cruciale: che cosa succede se si aggiunge un "piccolo" accoppiamento tra gli oscillatori? E' un problema importante, sia concettualmente sia per le implicazioni fisiche (molti sistemi fisici si presentano in questo modo), ma difficile, sul quale, in definitiva, si sa ancora poco. Ritourneremo più volte sulla questione. Continuiamo intanto lo studio dei sistemi elementari, mostrando che

Proposizione 4.3. *La traslazione di α sul toro \mathbb{T}^1 (esempio 2.2) è ergodica se e solo se il parametro α è irrazionale.*

divenuto chiaro solo dopo lo sviluppo della teoria della misura, e soprattutto dopo il contributo di Kolmogorov del 1954.

³⁴⁾ La nozione di sistema integrabile sarà ampiamente discussa nel terzo capitolo.

Dimostrazione. Se α è razionale, $\alpha = p/q$ con $p, q \in \mathbb{Z}$, allora l'orbita non è densa, perciò la proprietà E2 non è soddisfatta e il sistema non è ergodico (in altro modo: se $\alpha = p/q$, allora la funzione $f(x) = \cos 2\pi qx$ è un integrale del moto non banale). Sia ora α irrazionale. Presa $f \in L_2(M, \mu)$, possiamo scrivere

$$f(x) = \sum_{k \in \mathbb{Z}} \hat{f}_k e^{2\pi i k x}$$

$$f(\Phi(x)) = \sum_{k \in \mathbb{Z}} [\hat{f}_k e^{2\pi i k \alpha}] e^{2\pi i k x},$$

e si vede subito che f è un integrale del moto se e solo se

$$\hat{f}_k (e^{2\pi i k \alpha} - 1) = 0 \quad \forall k \in \mathbb{Z}.$$

Ma per α irrazionale, e $k \neq 0$, risulta $e^{2\pi i k \alpha} \neq 1$, e dunque necessariamente $\hat{f}_k = 0$. Pertanto f è quasi ovunque costante, e il sistema è ergodico. ■

Proposizione 4.4. *Il moto quasi periodico su \mathbb{T}^n (esempio 2.3) con velocità $v = (v_1, \dots, v_n)$ è ergodico, se e solo se v è non risonante.*

Dimostrazione. Presa $f \in L_2(M, \mu)$, possiamo scrivere

$$f(x) = \sum_{k \in \mathbb{Z}^n} \hat{f}_k e^{2\pi i (k \cdot x)}$$

$$f(\Phi^t(x)) = \sum_{k \in \mathbb{Z}^n} [\hat{f}_k e^{2\pi i (k \cdot v)t}] e^{2\pi i (k \cdot x)},$$

perciò f è integrale del moto se e solo se

$$\hat{f}_k (e^{2\pi i (k \cdot v)t} - 1) = 0 \quad \forall k \in \mathbb{Z}^n, t \in \mathbb{R},$$

e si trova soluzione non banale se e solo se esiste $k \neq 0$, tale che risulti $k \cdot v = 0$. ■

- **Esercizio.** Si formuli e si dimostri l'analogia proposizione per la traslazione su \mathbb{T}^n , definita da $\Phi(x) = x + \alpha \pmod{1}$, $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. Come si scrive la condizione di non risonanza di α ? [Risposta: $k \cdot \alpha \in \mathbb{Z}$].
- **Osservazione.** Per la traslazione o il moto quasi periodico su \mathbb{T}^n (nel caso non risonante) si può dimostrare che la proprietà E1 risulta verificata *per ogni* dato iniziale, e non solo quasi ovunque, non appena f è misurabile secondo Riemann³⁵⁾ (si veda il libro citato V.I. Arnol'd e A. Avez, appendice 9).

³⁵⁾ Per funzioni generiche non sarebbe mai possibile: si prenda la traslazione di \mathbb{T}^1 , e si ponga $f = 0$ ovunque tranne che sulle immagini $\Phi^t(x^*)$, $t \in \mathbb{Z}$, del dato iniziale x^* , ove si pone $f = 1$. Si ha $\langle f \rangle = 0$, ma $\bar{f}(x^*) = 1$.

k	2^k	k	2^k	k	2^k
0	1	15	32768	30	1073741824
1	2	16	65536	31	2147483648
2	4	17	131072	32	4294967296
3	8	18	262144	33	8589934592
4	16	19	524288	34	17179869184
5	32	20	1048576	35	34359738368
6	64	21	2097152	36	68719476736
7	128	22	4194304	37	137438953472
8	256	23	8388608	38	274877906944
9	512	24	16777216	39	549755813888
10	1024	25	33554432	40	1099511627776
11	2048	26	67108864	41	2199023255552
12	4096	27	134217728	42	4398046511104
13	8192	28	268435456	43	8796093022208
14	16384	29	536870912	44	17592186044416

- **Esercizio.** Nella tabella appaiono le prime quarantacinque potenze di due; nessuno dei numeri comincia per 7. Esisterà, asintoticamente, una frequenza p_l delle potenze di due che cominciano con la cifra l , $1 \leq l \leq 9$? Sarà più grande p_7 o p_8 ?
- **Osservazione.** Dimostreremo più avanti che gli schemi di Bernoulli (e dunque la trasformazione del panettiere) e gli automorfismi algebrici del toro del tipo dell'esempio 2.5 sono sistemi ergodici. D'altra parte si vede facilmente che sono dense, in questi sistemi, le orbite periodiche, come anche i dati iniziali che per $t \rightarrow \infty$ (o $t \rightarrow -\infty$) convergono a un punto fisso dello spazio ove il sistema dinamico è definito; corrispondentemente è molto facile trovare funzioni per le quali $\bar{f}(x)$ e $\langle f \rangle$ sono diverse per un insieme denso di dati iniziali.
- **Esercizio.** Si costruiscano esempi concreti a supporto di questa affermazione.
- **Esercizio.** Dato per buono che gli schemi di Bernoulli siano sistemi ergodici, si dimostri che in $B_{p,1-p}$ la frequenza di apparizione del simbolo "0" è esattamente p per tutte le stringhe, a meno di un insieme di misura nulla.

5. Sistemi mescolanti

Dopo la nozione di ergodicità, la successiva nozione rilevante nella teoria ergodica è quella di sistema *mescolante*, o *mixing*. L'idea, come già si è accennato, è quella di richiedere, pur con una dinamica microscopica reversibile, un comportamento irreversibile per gli insiemi di punti.

5.1. Definizione di sistema mescolante.

La nozione di sistema mescolante si può introdurre attraverso due proprietà esattamente equivalenti:

Proprietà M1. Per ogni coppia di insiemi misurabili $A, B \subset M$ risulta

$$\lim_{t \rightarrow \infty} \mu(\Phi^{-t}(A) \cap B) = \mu(A)\mu(B) . \quad (5.1)$$

Proprietà M2. Per ogni coppia di funzioni $f, g \in L_2(M, \mu)$ risulta

$$\lim_{t \rightarrow \infty} \int_M (f \circ \Phi^t) g \, d\mu = \int_M f \, d\mu \int_M g \, d\mu , \quad (5.2)$$

o in altra notazione

$$\lim_{t \rightarrow \infty} \langle (f \circ \Phi^t) g \rangle = \langle f \rangle \langle g \rangle . \quad (5.3)$$

Il significato di tali proprietà è chiaro: in base a M1, l'insieme $\Phi^{-t}(A)$, pur conservando la misura, va “diluendosi” uniformemente in M ; ciò è possibile se, per così dire, A si “filamenta” per effetto della dinamica, e (nel senso debole indicato dalla M1) invade uniformemente M ; si riveda la figura 2.5. L'insieme B va pensato come “finestra di osservazione”. La proprietà M2 esprime invece la perdita delle correlazioni tra osservabili qualsiasi. In generale, la funzione G definita da

$$G(t) = \langle (f \circ \Phi^t) g \rangle - \langle f \rangle \langle g \rangle \quad (5.4)$$

è chiamata *funzione di correlazione* di f e g ; se $g = f$, G è chiamata *funzione di autocorrelazione* di f . Se $G(t) \neq 0$, allora i valori (la misurazione) di f e g , a distanza di tempo t , non sono indipendenti.

- **Osservazione.** Sia ρ_t la distribuzione di probabilità associata allo stato macroscopico di un sistema statistico, nel senso di Gibbs; ricordiamo (paragrafo 1.3) che ρ_t evolve secondo la legge

$$\rho_t = \rho_0 \circ \Phi^{-t} .$$

Diremo che ρ_t tende in senso debole alla distribuzione ρ_∞ , se per ogni $f \in L_2(M, \mu)$ risulta

$$\lim_{t \rightarrow \infty} \int_M f \rho_t \, d\mu = \int_M f \rho_\infty \, d\mu .$$

Si vede allora immediatamente che ρ_t tende debolmente alla distribuzione uniforme, $\rho_\infty(x) = 1 \, \forall x \in M$, per ogni scelta di ρ_0 (in $L_2(M, \mu)$), se e solo se è soddisfatta M2. Infatti, si ha

$$\int_M f \rho_t \, d\mu = \int_M f (\rho_0 \circ \Phi^{-t}) \, d\mu = \int_M (f \circ \Phi^t) \rho_0 \, d\mu ,$$

(l'ultimo passaggio è il cambiamento di variabile $x = \Phi^t(x')$, il cui determinante jacobiano vale 1), e la conclusione è immediata non appena si ricordi la condizione

di normalizzazione $\langle \rho_0 \rangle = 1$ (il fatto che ρ_0 non sia completamente arbitraria ma normalizzata è chiaramente irrilevante).

Si vede assai facilmente che

Proposizione 5.1. *Le proprietà M1 e M2 sono equivalenti.*

Dimostrazione. Per vedere che M2 implica M1, basta prendere $f = \chi_A$, $g = \chi_B$; poiché $\chi_A \circ \Phi^t = \chi_{\Phi^{-t}(A)}$, si ha

$$\int_M (f \circ \Phi^t) g d\mu = \mu(\Phi^{-t}(A) \cap B) ,$$

e la conclusione è immediata.

Viceversa, per mostrare che M1 implica M2, si considerano dapprima f e g che siano funzioni semplici, ovvero somma finita di funzioni caratteristiche:

$$f = \sum_i f_i \chi_{A_i} , \quad g = \sum_j g_j \chi_{B_j} ;$$

per tali funzioni la verifica è immediata: poiché $\chi_{A_i} \circ \Phi^t = \chi_{\Phi^{-t}(A_i)}$, si ha

$$\begin{aligned} \langle (f \circ \Phi^t)g \rangle &= \sum_{ij} f_i g_j \langle \chi_{\Phi^{-t}(A_i)} \chi_{B_j} \rangle = \sum_{ij} f_i g_j \mu(\Phi^{-t}(A_i) \cap B_j) \\ &\longrightarrow \sum_{ij} f_i g_j \mu(A_i) \mu(B_j) = \sum_{ij} f_i g_j \langle \chi_{A_i} \rangle \langle \chi_{B_j} \rangle = \langle f \rangle \langle g \rangle . \end{aligned}$$

Per funzioni generiche, si sfrutta il fatto che per f, g in $L_2(M, \mu)$, comunque si prenda $\varepsilon > 0$, si può sempre scrivere $f = \tilde{f} + f'$, $g = \tilde{g} + g'$, con \tilde{f}, \tilde{g} semplici, e $\|f'\|_2, \|g'\|_2 < \varepsilon$. Seguono facilmente disuguaglianze del tipo

$$|\langle (f \circ \Phi^t)g \rangle - \langle (\tilde{f} \circ \Phi^t)\tilde{g} \rangle| < (\text{cost})\varepsilon , \quad |\langle f \rangle \langle g \rangle - \langle \tilde{f} \rangle \langle \tilde{g} \rangle| < (\text{cost})\varepsilon$$

(la prima uniforme in t); usando per \tilde{f} e \tilde{g} il risultato precedente, e tenendo conto dell'arbitrarietà di ε , la conclusione è immediata. ■

A questo punto ha senso dare la seguente

Definizione. Il sistema dinamico (M, μ, Φ) si dice *mescolante*, se le proprietà M1 o M2 sono soddisfatte.

Si verifica immediatamente che la nozione di sistema mescolante è invariante per isomorfismo. Inoltre:

Proposizione 5.2. *Ogni sistema mescolante è ergodico.*

Dimostrazione. Se $A \subset M$ misurabile è invariante, allora per la proprietà M1, con $B = A$, si ha

$$\mu(A) = \mu(\Phi^{-t}(A) \cap A) \longrightarrow \mu(A)^2 ,$$

ovvero A è banale. La E4 è allora soddisfatta, e questo basta. ■

- **Osservazione.** Si vede immediatamente che il moto quasi periodico (o la traslazione) sul toro, come tutte le isometrie, non sono mai sistemi mescolanti. E' ugualmente evidente che i sistemi hamiltoniani a un sol grado di libertà non sono mescolanti. Perciò la proprietà di mescolamento è effettivamente più forte dell'ergodicità.

5.2. Esempi elementari

Dalle considerazioni svolte nel paragrafo 2.1 (esempio 2.5) si intuisce che

Proposizione 5.3. *L'automorfismo algebrico di \mathbb{T}^2 definito nell'esempio 2.5 è un sistema mescolante.*

Dimostrazione. Le funzioni $u_k(x) = e^{2\pi i k \cdot x}$, $k \in \mathbb{Z}^2$, sono una base ortogonale per $L_2(M, \mu)$, ovvero si ha

$$\langle u_k u_l \rangle = 0 \quad \text{per } k \neq l ;$$

dalla definizione di Φ si vede poi subito che risulta

$$u_k \circ \Phi = u_{\tilde{A}k} , \quad u_k \circ \Phi^t = u_{\tilde{A}^t k} , \quad (5.5)$$

ove \tilde{A} indica la trasposta di A . Ora (ed è questa l'idea centrale della dimostrazione), per $k \neq 0$ l'orbita di k , precisamente $\{\tilde{A}^t k, t \in \mathbb{Z}\}$, è aperta, e in particolare, posto ad esempio $|k| = |k_1| + |k_2|$, risulta

$$|\tilde{A}^t k| \rightarrow \infty \quad \text{per } t \rightarrow \infty , \quad \forall k \in \mathbb{Z}^2, k \neq 0 ; \quad (5.6)$$

ciò è conseguenza immediata del fatto che la matrice \tilde{A} , applicata ripetutamente, espande tutti i vettori, ad eccezione di quelli appartenenti alla retta contraente: che però ha pendenza irrazionale, e dunque non contiene vettori interi.

Da questa proprietà si deduce facilmente che il sistema è mescolante. Prendiamo infatti $f, g \in L_2(M, \mu)$; le funzioni si approssimano a piacere con somme finite di funzioni di base: precisamente, preso $\varepsilon > 0$ arbitrario, esiste K tale che, posto³⁶⁾

$$f = f^{\leq K} + f^{> K} , \quad f^{\leq K} := \sum_{|k| \leq K} \hat{f}_k u_k ,$$

risulta $\|f^{> K}\|_2 < \varepsilon$, e analogamente per g (si può prendere il medesimo K per le due funzioni). Si vede ora immediatamente che per t sufficientemente grande risulta

$$\langle (f^{\leq K} \circ \Phi^t) g^{\leq K} \rangle - \langle f \rangle \langle g \rangle = 0 ;$$

infatti, si ha evidentemente

$$\langle (f^{\leq K} \circ \Phi^t) g^{\leq K} \rangle = \sum_{|k|, |l| \leq K} \hat{f}_k \hat{g}_l \langle u_{\tilde{A}^t k} u_l \rangle ,$$

³⁶⁾ Le funzioni $f^{\leq K}$ e $f^{> K}$ sono a volte chiamate parte infrarossa e parte ultravioletta di f ; decomposizioni di questo tipo hanno grande interesse in teoria delle perturbazioni, perchè se f è regolare, allora la norma della parte ultravioletta decresce molto rapidamente con K (addirittura esponenzialmente, per funzioni analitiche), e in una certa misura può essere trascurata nel procedimento perturbativo.

ma in base alla (5.6), per t abbastanza grande e $k \neq 0$ risulta $|\tilde{A}^t k| > K$: e dunque (per l'ortogonalità delle funzioni di base) l'unico termine non nullo della sommatoria è quello corrispondente a $k = l = 0$, che è proprio $\langle f \rangle \langle g \rangle$. La conclusione ora è immediata: per ogni t si ha $|\langle (f \circ \Phi^t) g \rangle - \langle (f^{\leq K} \circ \Phi^t) g^{\leq K} \rangle| < (\text{cost})\varepsilon$, e dunque per t abbastanza grande si trova

$$|\langle (f \circ \Phi^t) g \rangle - \langle f \rangle \langle g \rangle| < (\text{cost})\varepsilon ;$$

per l'arbitrarietà di ε la proprietà $M2$ è soddisfatta, e dunque il sistema è mescolante.

- **Osservazione.** Insistiamo ancora sul fatto che il meccanismo che conduce al mescolamento è l'esistenza di una base ortogonale in $L_2(M, \mu)$, chiusa per composizione con il flusso secondo la (5.5), tale che tutte le funzioni di base, ad eccezione della funzione costante, hanno comportamento irreversibile (l'indice k della funzione di base ha un'orbita aperta).
- **Esercizio.** Si determini una condizione necessaria e sufficiente sulla matrice A , perchè il sistema sia mescolante. Si studi il problema analogo per gli automorfismi algebrici di \mathbb{T}^n .

Occupiamoci ora della classe, assai rilevante, degli schemi di Bernoulli, e dimostriamo che

Proposizione 5.4. *Gli schemi di Bernoulli sono sistemi dinamici mescolanti.*

Dimostrazione. L'idea della dimostrazione non è molto diversa dalla precedente. Presi A e B misurabili, e fissato $\varepsilon > 0$, esistono sicuramente \tilde{A} e \tilde{B} che approssimano A e B , precisamente tali che³⁷⁾

$$\mu(A - \tilde{A}) < \varepsilon, \quad \mu(B - \tilde{B}) < \varepsilon, \quad (5.7)$$

e che inoltre si ottengono da un insieme finito di generatori, con le operazioni consuete di unione e intersezione.³⁸⁾ Nel nostro caso, i generatori sono i cilindri \mathcal{C}_k^l , con $k \in \mathbb{Z}$, e l appartenente all'alfabeto I ; \tilde{A} e \tilde{B} si otterranno dunque a partire dall'insieme dei cilindri \mathcal{C}_k^l , con $|k| \leq K$, per un opportuno $K > 0$ (dipendente da ε); in particolare, si potranno scrivere come opportuna unione disgiunta degli insiemi più piccoli che si possono ottenere con questi generatori, che sono i cilindri $\mathcal{C}_{-K, \dots, K}^{l-K, \dots, lK}$ (tutti i simboli tra $-K$ e K sono specificati).

Ora, dalla definizione stessa di Φ , si ha che $\Phi^{-t}(\mathcal{C}_k^l) = \mathcal{C}_{k+t}^l$; se ne deduce immediatamente che, per $t > 2K + 1$, l'insieme evoluto $\Phi^{-t}(\tilde{A})$ è costruito a partire da un insieme di generatori interamente diverso da quelli di \tilde{B} : di conseguenza, è facile vedere, la misura si fattorizza, ovvero si ha

$$\mu(\Phi^{-t}(\tilde{A}) \cap \tilde{B}) = \mu(\Phi^{-t}(\tilde{A})) \mu(\tilde{B}) = \mu(\tilde{A}) \mu(\tilde{B}) \quad \forall t > 2K + 1 .$$

La conclusione è immediata: dalla (5.7) segue

$$\begin{aligned} |\mu(\Phi^{-t}(A) \cap B) - \mu(\Phi^{-t}(\tilde{A}) \cap \tilde{B})| &< (\text{cost})\varepsilon \\ |\mu(A) \mu(B) - \mu(\tilde{A}) \mu(\tilde{B})| &< (\text{cost})\varepsilon , \end{aligned}$$

³⁷⁾ Il simbolo “ $-$ ” indica la differenza simmetrica: $A - \tilde{A} = (A \setminus \tilde{A}) \cup (\tilde{A} \setminus A)$.

³⁸⁾ Questo è un fatto generale della teoria della misura; in modo analogo qualunque insieme del piano si può approssimare con rettangoli.

perciò per t sufficientemente grande ($t > 2K$) si ha

$$|\mu(\Phi^{-t}(A) \cap B) - \mu(A)\mu(B)| < (\text{cost})\varepsilon ;$$

si conclude che il sistema soddisfa la proprietà M1, e dunque è mescolante. ■

- **Osservazione.** Si osservi l'analogia tra \tilde{A} e \tilde{B} , e la parte infrarossa di f e g nel caso dell'automorfismo algebrico di \mathbb{T}^2 sopra studiato. In entrambi i casi, sono oggetti che approssimano a piacere gli oggetti che ci interessano, e allo stesso tempo diventano indipendenti in un numero finito di passi (la misura degli insiemi si fattorizza; la correlazione tra funzioni si annulla). In entrambi i casi l'indipendenza si realizza perché la dinamica sulle funzioni di base / sui cilindri si realizza con traiettorie aperte: è questo l'elemento tecnico che produce la non ricorrenza, neanche in tempi infiniti, nella dinamica delle funzioni / degli insiemi, a dispetto della ricorrenza nella dinamica dei punti.
- **Esercizio.** Si dimostri che l'endomorfismo di \mathbb{T}^1 , definito da $\Phi(x) = 2x \bmod 1$ (esempio 2.6) è mescolante.
- **Esercizio.** Si dimostri che (M, μ, Φ^2) è mescolante, se e solo se (M, μ, Φ) è mescolante. Si dimostri che se (M, μ, Φ^2) è ergodico, allora (M, μ, Φ) è ergodico, ma non viceversa (si costruisca un controesempio).

Al di là degli esempi elementari, è *molto* difficile dimostrare che un sistema è mescolante (o ergodico). Gli esempi noti più importanti sono i seguenti:

- i. Alcune famiglie di biliardi. In particolare: i *biliardi di Sinai* (1962), definiti dall'avere il bordo convesso verso l'interno (il più semplice è costituito da un toro \mathbb{T}^2 , con un ostacolo circolare, oppure si può prendere un poligono, e curvare verso l'interno i lati) oppure, i *biliardi di Bunimovich–Sinai* (1974): tavole piane semplicemente connesse, il cui bordo è costituito da archi di circolo e segmenti, con la prescrizione che se un arco di circolo appartiene al bordo, l'intero cerchio è contenuto nella tavola (il più semplice ha la forma di uno stadio di atletica: un rettangolo, con due lunette semicircolari su due lati opposti). Si veda la figura 2.8. Il toro con ostacolo si mostra essere equivalente a un gas di due dischi duri che si urtano elasticamente; in un primo momento sembrava che la dimostrazione di Sinai fosse valida per n dischi (così è ancora riportato il risultato di Sinai in molti libri o articoli di rassegna), ma poi si sono trovate difficoltà. Sono seguite dimostrazioni particolari per $n = 3$ e $n = 4$, e molto di recente (Szász, 1995) è stata annunciata una dimostrazione valida per n qualsiasi (purchè il volume a disposizione dei dischi sia abbastanza grande, ovvero il gas non sia troppo denso). Se confermato, questo risultato — inseguito per oltre vent'anni — sarebbe di grande interesse fisico.
- ii. I *Sistemi di Anosov*, di cui parleremo più avanti: sono sistemi molto speciali, caratterizzati da divergenza forte (uniforme) delle traiettorie vicine. Il Gatto di Arnol'd è un sistema di Anosov; un esempio non banale è il flusso geodetico su superfici compatte a curvatura negativa (ogni punto è una sella); superfici di questo genere, difficilmente immaginabili, si ottengono dal cerchio di Poincaré–Lobatchewskii con un procedimento di compattizzazione, simile all'operazione di modulo con cui si passa da \mathbb{R}^2 a \mathbb{T}^2 .

- iii. Il moto di un punto su \mathbb{T}^2 , con un potenziale che abbia singolarità attrattive di tipo coulombiano (purché l'energia sia abbastanza alta); o anche, il biliardo sul toro, quando all'ostacolo (o agli ostacoli) si sostituisca un potenziale a simmetria circolare, a supporto in un disco, con derivata discontinua sul bordo del disco (purché l'energia sia abbastanza bassa). Sono lavori abbastanza recenti, anche questi abbastanza interessanti dal punto di vista fisico.

Per tutti questi sistemi si dimostrano in realtà proprietà più forti del mescolamento, cui accenneremo più avanti.

5.3. Ergodicità e mescolamento; mescolamento debole.

Vale la pena di discutere un po' più a fondo sulla relazione tra l'ergodicità e la proprietà di mescolamento. Sia (M, μ, Φ) un sistema dinamico ergodico, che per essere definiti supporremo discreto. In base alla nozione E1 di ergodicità segue che quasi ovunque, per $t \rightarrow \infty$, si ha

$$\frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) \longrightarrow \langle f \rangle, \quad \frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) g(x) \longrightarrow \langle f \rangle \langle g \rangle,$$

per ogni scelta di f, g in $L_2(M, \mu)$. Integrando e scambiando il limite con l'integrazione (lo si può certamente fare se f è limitata, per il teorema della convergenza dominata; ma anche se non lo è, si può vedere che tutto funziona ugualmente³⁹⁾) si ricava una proprietà simile al mescolamento:

$$\frac{1}{t} \sum_{s=0}^{t-1} [\langle (f \circ \Phi^s) g \rangle - \langle f \rangle \langle g \rangle] \longrightarrow 0,$$

o in altra notazione

$$\frac{1}{t} \sum_{s=0}^{t-1} G(s) \longrightarrow 0,$$

ove G è la funzione di correlazione definita in (5.4). Equivalentemente (si prendano $f = \chi_A$ e $g = \chi_B$, che per di più sono limitate e non danno problemi per lo scambio dei limiti), si ottiene, dalla sola ergodicità,⁴⁰⁾

$$\frac{1}{t} \sum_{s=0}^{t-1} [\mu(\Phi^{-s}(A) \cap B) - \mu(A)\mu(B)] \longrightarrow 0$$

per ogni coppia di insiemi misurabili A e B .

L'ergodicità corrisponde dunque a una proprietà di “decadimento in media” delle correlazioni (o di “mescolamento in media”): occorre in più una media temporale. Se il

³⁹⁾ Si approssima f con una funzione limitata, e a quel punto si scambiano limite e integrazione; l'errore che si commette risulta essere arbitrariamente piccolo.

⁴⁰⁾ Oltre a essere proprietà dedotte dall'ergodicità, queste proprietà a loro volta la implicano; ad esempio, per dedurre l'ergodicità dalla (5.6), si procede come nella proposizione 5.2).

mescolamento ha il significato che le operazioni di misura di osservabili qualsiasi (f e g) a distanza di tempo t diventano indipendenti per t grande, l'ergodicità da sola dice che la loro funzione di correlazione G , pur non decadendo, è però mediamente nulla.⁴¹⁾

- **Osservazione.** Accanto alla funzione di correlazione temporale $G(t)$ introdotta sopra, si può definire una funzione in un certo senso analoga,

$$\mathcal{G}(t, x) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=0}^{T-1} f(\Phi^{t+s}(x))g(\Phi^s(x)) - \bar{f}(x)\bar{g}(x) , \quad (5.8)$$

nella quale le medie temporali lungo l'orbita di origine x sostituiscono la media in fase. Se il sistema è ergodico, evidentemente risulta $\mathcal{G}(t, x) = G(t)$ quasi ovunque; la (5.8) è interessante, perché consente di discutere del decadimento delle correlazioni lungo una singola orbita, indipendentemente dalle proprietà globali del sistema.

Concludiamo accennando a una proprietà intermedia tra ergodicità e mescolamento, chiamata mescolamento debole (*weak mixing*): precisamente,

Definizione. Il sistema dinamico (M, μ, Φ) si dice *debolmente mescolante*, se per ogni coppia di insiemi misurabili A e B risulta

$$\frac{1}{t} \sum_{s=0}^{t-1} |\mu(\Phi^{-s}(A) \cap B) - \mu(A)\mu(B)| \longrightarrow 0$$

(in più dell'ergodicità si chiede la convergenza assoluta).

6. Misure ergodiche

Consideriamo un sistema dinamico, classico o astratto, (M, μ, Φ) . E' in qualche modo naturale pensare che M e Φ siano assegnati, e la misura μ su M sia invece, per così dire, una nostra scelta.

Si pone pertanto il problema seguente: fissata la coppia (M, Φ) , ed anche la σ -algebra di insiemi misurabili su M (altrimenti nessun confronto tra misure sarà mai possibile), studiare l'insieme delle misure invarianti (ovvero quelle per cui (M, μ, Φ) è un sistema dinamico), e in particolare il sottoinsieme delle *misure ergodiche* (ovvero quelle per cui

⁴¹⁾ Nel linguaggio della distribuzione di probabilità ρ_t di Gibbs: in un sistema ergodico si ha

$$\frac{1}{t} \sum_{s=0}^{t-1} \int_M f \rho_t d\mu \longrightarrow \int_M f d\mu ,$$

ovvero ρ_t non tende necessariamente alla distribuzione costante $\rho = 1$, tuttavia oscilla intorno al valor medio $\rho = 1$ (non è sistematicamente né sopra né sotto questo valore).

(M, μ, Φ) è ergodico). E' una prospettiva in un certo senso più ampia di quella che avevamo assunto fino ad ora assegnando a priori, oltre alla dinamica, anche la misura. E si tratta, si badi, di una questione piuttosto interessante anche da un punto di vista fisico: a ogni misura conservata corrisponde infatti, nel senso di Gibbs (paragrafo 1.3), un possibile stato di equilibrio del sistema. In effetti, le misure invarianti sono anche dette *misure di equilibrio* compatibili con una assegnata dinamica.

La discussione svolta a proposito del punto di vista di Gibbs, poi ripresa studiando la nozione di sistema ergodico, fornisce una parziale risposta al problema qui posto: se il sistema dinamico (M, μ, Φ) è ergodico, e si restringe la ricerca a misure ν regolari, più precisamente assolutamente continue rispetto a μ ,⁴²⁾ cosicché si può scrivere $d\nu = \rho d\mu$ con ρ misurabile, allora l'unica misura invariante è μ stessa (altrimenti ρ sarebbe funzione misurabile invariante non banale), e dunque, con questa restrizione, lo stato di equilibrio è unico.

Più in generale, si dimostra senza difficoltà la seguente

Proposizione 6.1.

- a) Se il sistema dinamico (M, μ, Φ) è ergodico, e la misura ν , diversa da μ , è invariante, allora ν è singolare⁴³⁾ rispetto a μ .
- b) Se anche (M, ν, Φ) è ergodico, esistono M_μ e M_ν disgiunti tali che

$$\mu(M_\mu) = 1, \quad \mu(M_\nu) = 0, \quad \nu(M_\mu) = 0, \quad \nu(M_\nu) = 1$$

(ciascuna misura ha supporto in un insieme di misura nulla dell'altra).

- c) Se infine (M, μ, Φ) non è ergodico, allora esistono infinite misure invarianti, assolutamente continue rispetto a μ .

Dimostrazione. Proviamo innanzitutto b), e per questo denotiamo con $\langle \cdot \rangle_\mu, \langle \cdot \rangle_\nu$ le medie in fase relative alle due misure; poiché le due misure sono diverse, certamente esiste $f : M \rightarrow \mathbb{R}$ sommabile, tale che $\langle f \rangle_\mu \neq \langle f \rangle_\nu$. D'altra parte, per la nozione E1 di ergodicità, se i due sistemi sono ergodici allora sicuramente esistono M_μ e M_ν di misura uno per le rispettive misure,

$$\mu(M_\mu) = \nu(M_\nu) = 1,$$

tali che

$$\begin{aligned} \bar{f}(x) &= \langle f \rangle_\mu && \text{per } x \in M_\mu \\ \bar{f}(x) &= \langle f \rangle_\nu && \text{per } x \in M_\nu; \end{aligned}$$

i due insiemi sono allora disgiunti, e quindi, come richiesto,

$$\mu(M_\nu) = \nu(M_\mu) = 0.$$

Per provare a) ci basta ora considerare il caso in cui ν è invariante, ma (M, ν, Φ) non è ergodico (altrimenti a) si riduce banalmente a b)). Lo si vede immediatamente: se

⁴²⁾ Ricordiamo che ν si dice assolutamente continua rispetto a μ , se $\mu(A) = 0$ implica $\nu(A) = 0$; in tal caso esiste sempre ρ misurabile tale che $d\nu = \rho d\mu$.

⁴³⁾ Ovvero esiste $A \subset M$, tale che $\mu(A) = 0$, e $\nu(A) \neq 0$.

(M, ν, Φ) non è ergodico, in base alla nozione E4 esiste A invariante, tale che sia $\nu(A)$ sia $\nu(A^c)$ sono positive, $A^c := M \setminus A$. Viceversa, per l'ergodicità di (M, μ, Φ) , uno almeno di questi insiemi ha misura nulla, e questo basta.

Infine, per provare c), supponiamo che (M, μ, Φ) non sia ergodico, e dunque esista A invariante, tale che $\mu(A), \mu(A^c)$ siano entrambe diverse da zero. Allora sono invarianti le misure ν' e ν'' , che a un generico insieme misurabile B attribuiscono misura

$$\nu'(B) = \frac{\mu(B \cap A)}{\mu(A)}, \quad \nu''(B) = \frac{\mu(B \cap A^c)}{\mu(A^c)}$$

(in sostanza: μ stessa, ristretta agli insiemi A e A^c , e opportunamente normalizzata), e sono evidentemente invarianti tutte le infinite combinazioni lineari

$$\nu = \alpha \nu' + (1 - \alpha) \nu'', \quad 0 \leq \alpha \leq 1. \tag{6.1}$$

■

- **Osservazione.** In quest'ultima parte, precisamente nella (6.1), abbiamo sfruttato una proprietà delle misure conservate, di immediata verifica, che vale la pena di enucleare: *se $\Phi : M \rightarrow M$ preserva le misure μ e ν , preserva ogni loro combinazione lineare $\alpha\mu + (1 - \alpha)\nu$, $0 \leq \alpha \leq 1$.* Con immagine geometrica, possiamo dire che l'insieme delle misure conservate da una assegnata trasformazione è convesso.
- **Osservazione.** Si dimostra anche che, in un senso opportuno, le misure rispetto alle quali il sistema risulta ergodico costituiscono il “bordo” di questo insieme convesso, ovvero che ogni misura conservata si può sempre scrivere, ancora in un senso opportuno, come combinazione di misure ergodiche (in generale ne occorrono infinite). Un esempio elementare è dato dal moto su \mathbb{T}^2 con velocità $v = (v_x, v_y) = (f(y), 0)$, con $f(y)$ qualsiasi non nulla: per ogni y esiste con evidenza una misura invariante μ_y con supporto sulla sezione “orizzontale” \mathbb{T}_y^1 di ascissa y , uniforme su di essa, e ancora con evidenza ciascuna di queste misure è ergodica. Si ottiene poi una misura invariante ν' prendendo una successione (eventualmente infinita) $\{y_j\}$, e ponendo

$$\nu' = \sum_j c_j \mu_{y_j}, \quad c_j > 0 \quad \forall j, \quad \sum_j c_j = 1;$$

similmente si ottiene una misura invariante ν'' ponendo

$$\nu'' = \int_0^1 \rho(y) \mu_y dy, \quad \rho > 0, \quad \int_0^1 \rho(y) dy = 1.$$

Si intuisce, e si potrebbe dimostrare, che ogni misura invariante ν ammette la decomposizione

$$\nu = \alpha \nu' + (1 - \alpha) \nu'',$$

con ν' e ν'' come sopra. In generale, assegnato un sistema dinamico (M, μ, Φ) , è sempre possibile decomporre M in sottoinsiemi invarianti M_β , con indice discreto e/o continuo, ciascuno munito di una misura ν_β tale che $(M_\beta, \nu_\beta, \Phi)$ sia ergodico, e

contemporaneamente scrivere la misura μ come sovrapposizione delle ν_β . Tale operazione prende il nome di *decomposizione ergodica* del sistema (o della misura); si veda ad esempio il testo citato di Cornfeld *et. al.*

- **Osservazione.** Per sistemi con un punto fisso x globalmente attrattivo (oscillatore armonico smorzato o sistemi simili) l'unica misura conservata è quella banale, che assegna misura uno a ogni insieme contenente x , zero a ogni insieme che non contiene x . Per sistemi come l'equazione di Van der Pol, che hanno un'orbita periodica attrattiva, o sistemi con attrattori più complicati (si veda il secondo capitolo), le uniche misure conservate hanno necessariamente supporto sull'attrattore, la cui misura di Lebesgue è nulla. Sistemi di questo tipo sono detti *dissipativi* perchè non si conserva nessuna misura continua rispetto alla misura di Lebesgue. Per sistemi con attrattori si pone il problema, *molto* interessante, di determinare le misure conservate e le misure ergodiche sull'attrattore, e anche di connettere in qualche modo tali misure alla misura di Lebesgue nel bacino di attrazione (capitolo secondo, paragrafo 6).

Un esempio di sistemi dinamici in cui coesistono misure ergodiche diverse è offerto dagli schemi di Bernoulli: ad esempio, gli schemi di Bernoulli a due simboli $B_{p,1-p}$, hanno tutti il medesimo spazio M (le sequenze qualsiasi di simboli), la medesima σ -algebra (generata dai cilindri), e il medesimo automorfismo Φ (lo slittamento a sinistra), ma differiscono nella misura. D'altra parte, tutti questi sistemi sono ergodici: dunque, *al variare di p abbiamo infinite misure ergodiche*, ciascuna delle quali ha supporto in un insieme che per tutte le altre ha misura nulla. Del resto si era già visto (esercizio alla fine del paragrafo 4.2) che, coerentemente con le considerazioni qui svolte, la misura dello schema di Bernoulli $B_{p,1-p}$ ha supporto nell'insieme M_p delle stringhe in cui il simbolo "0" ha frequenza p . Tutti gli M_p sono a due a due disgiunti.

7. Intermezzo: il problema di Fermi–Pasta–Ulam; il problema di Hénon e Heiles.

7.1. Il lavoro numerico di Fermi, Pasta e Ulam

Nel 1954 Fermi, Pasta e Ulam impiegarono uno dei primi calcolatori elettronici mai costruiti, il "MANIAC I" dei Laboratori di Los Alamos, per simulare il comportamento di un sistema hamiltoniano, interessante per la meccanica statistica, tramite la risoluzione numerica delle sue equazioni microscopiche del moto. Ne uscì un lavoro rivoluzionario,⁴⁴⁾ che dava il via a un metodo nuovo di ricerca per il quale fu presto coniato il nome, assai espressivo, di "esperimenti numerici". In effetti, a partire da questo lavoro si cominciò a capire che il calcolatore non serve solo a far di conto velocemente, ma⁴⁵⁾ può essere stru-

⁴⁴⁾ Fu questo l'ultimo lavoro di Fermi. Il lavoro apparve anzi postumo, nel 1955, nella forma di rapporto interno dei Laboratori di Los Alamos.

⁴⁵⁾ Se usato con intelligenza, siamo costretti ad aggiungere oggi.

mento di compressione anche profonda, perchè può mettere in evidenza nuovi fenomeni, mettere alla prova idee e congetture, e in definitiva aprire spazi alla ricerca, al pari di ogni nuova potente tecnica sperimentale.

Il modello di Fermi, Pasta e Ulam (FPU) è costituito da una catena di $n + 2$ particelle identiche su una retta (n particelle mobili, due fisse agli estremi) interagenti con forze debolmente non lineari; si tratta di uno dei piú semplici sistemi, rilevanti per la meccanica statistica, per i quali sia interessante il problema ergodico. Se indichiamo con q_1, \dots, q_n lo spostamento delle particelle mobili dalla posizione di equilibrio, con p_1, \dots, p_n i loro momenti, con m la loro massa, e con $m\Omega^2$ la costante elastica lineare della forza, l'hamiltoniana di Fermi, Pasta e Ulam si scrive

$$H = \sum_{l=1}^n \frac{p_l^2}{2m} + m\Omega^2 \sum_{l=1}^{n+1} \left[\frac{1}{2}(q_l - q_{l-1})^2 + \frac{\mu}{r}(q_l - q_{l-1})^r \right], \quad q_0 = q_{n+1} = 0, \quad (7.1)$$

con $r = 3$ oppure $r = 4$; le corrispondenti equazioni del moto, in forma del secondo ordine, sono

$$\ddot{q}_l = -\Omega^2(2q_l - q_{l-1} - q_{l+1}) - \mu\Omega^2[(q_{l+1} - q_l)^{r-1} + (q_l - q_{l-1})^{r-1}], \quad l = 1, \dots, n. \quad (7.2)$$

Il parametro μ misura la nonlinearità delle forze, ma è chiaro che la nonlinearità del sistema dipende, oltre che da μ , dall'energia: infatti, al crescere dell'energia le ampiezze dei moti crescono, e con esse cresce il contributo relativo del termine non lineare. Formalmente è facile vedere che, grazie all'omogeneità della parte non lineare del potenziale, la nonlinearità del sistema è ben caratterizzata dal solo parametro adimensionale $\varepsilon = \mu(\frac{\mathcal{E}}{m\Omega^2})^{\frac{2}{r}-1}$, ove $\mathcal{E} = E/n$ è l'energia specifica, nel senso che al variare di m , Ω , μ e E , in modo che ε resti invariata, si ottengono moti del sistema esattamente identici, a meno di un banale cambiamento di scala o ridefinizione delle unità di misura (l'uso di \mathcal{E} al posto di E da questo punto di vista non è rilevante, ma è spontaneo se si pensa a un limite di n grande). Infatti, il riscalamento

$$p = \alpha\tilde{p}, \quad q = \beta\tilde{q}, \quad t = \gamma\tilde{t}$$

muta l'hamiltoniana H in

$$\tilde{H}(\tilde{p}, \tilde{q}) = \frac{\gamma}{\alpha\beta} H(\alpha\tilde{p}, \beta\tilde{q}),$$

e si vede immediatamente che \tilde{H} ha ancora la forma (7.1), con nuove costanti

$$\tilde{m} = \frac{\beta}{\alpha\gamma} m, \quad \tilde{\Omega} = \gamma\Omega, \quad \tilde{\mu} = \beta^{r-2}\mu,$$

mentre la nuova energia è evidentemente $\tilde{E} = \frac{\gamma}{\alpha\beta} E$; al variare di α , β e γ si ottengono così valori arbitrari di m , ω , μ e E , col solo vincolo che resti costante la quantità adimensionale ε introdotta sopra.

Per piccola nonlinearità è sensato introdurre i modi normali armonici del sistema, che risultano essere dati dalle "somme di Fourier" finite

$$\xi_k = \sum_{l=1}^n U_{kl} q_l, \quad U_{kl} = \sqrt{\frac{2}{n+1}} \sin \frac{kl\pi}{n+1},$$

$k = 1, \dots, n$. Non è difficile verificare che in queste nuove variabili le equazioni del moto del sistema assumono la forma

$$\ddot{\xi}_k + \omega_k^2 \xi_k = \mu F_k(\xi_1, \dots, \xi_n), \quad \omega_k = 2\Omega \sin \frac{k\pi}{2(n+1)}, \quad (7.3)$$

ove F_k è un (complicato) polinomio omogeneo di grado $r-1$ in ξ_1, \dots, ξ_n . Si verifica anche senza difficoltà che risulta

$$\sum_{j=1}^n U_{kj} U_{jl} = \delta_{kl},$$

ovvero che la matrice $U = (U_{kl})$ coincide con l'inversa (oltre che, ovviamente, con la trasposta), cosicché la trasformazione inversa è la medesima (la trasformazione è involutoria), e anche la trasformazione sulle variabili coniugate, che chiameremo η_1, \dots, η_n , è la medesima. Infine, l'hamiltoniana nelle nuove variabili prende la forma

$$K(\eta, \xi) = \sum_{k=1}^n E_k + \varepsilon V(\xi), \quad E_k = \frac{\eta^2}{2m} + \frac{m\omega_k^2}{2} \xi_k^2,$$

ove V è un (complicato) polinomio di grado r nelle ξ_k .

- **Esercizio.** Si verifichino esplicitamente queste affermazioni.

Per $\varepsilon = 0$ i modi normali si comportano come oscillatori armonici disaccoppiati, cosicché le loro energie E_k sono integrali primi del moto; in queste condizioni, ovviamente, il sistema è decisamente non ergodico. La domanda spontanea allora è se per ε piccolo (piccolo μ , o basse energie) l'accoppiamento sia in grado di rendere ergodico il sistema, rendendo sensata una trattazione statistica. Se così fosse, in particolare, ci si dovrebbe aspettare la validità del principio di equipartizione dell'energia (basato in modo essenziale sull'ipotesi ergodica), in base al quale le medie temporali

$$\overline{E}_k(T) = \frac{1}{T} \int_0^T E_k(t) dt$$

convergono (approssimativamente, per via del contributo energetico dei termini nonlineari che tuttavia è piccolo per bassa nonlinearity) al medesimo valore E/n .

Fermi, Pasta e Ulam si misero in condizioni di piccola nonlinearity, e posero il sistema fortemente fuori dall'equilibrio, assegnando una condizione iniziale in cui solo uno o pochi modi normali (con k piccolo) fossero inizialmente eccitati. Essi condussero numerosi esperimenti numerici, in modelli con $n = 32$ o 64 , variando un po' (ma non di molto) le condizioni iniziali e la nonlinearity. Si aspettavano di osservare un rapido flusso di energia dai modi eccitati a tutti gli altri modi, che conducesse abbastanza in fretta il sistema all'equipartizione: anzi, scopo dichiarato del lavoro era proprio la stima del tempo necessario al raggiungimento dell'equilibrio statistico. Essi si trovarono invece di fronte a tutt'altro genere di fenomeno:⁴⁶⁾

⁴⁶⁾ Let us here say that the results of our numerical computations show features which were, from the beginning, surprising to us. Instead of a gradual, continuous flow of energy from the first mode to the higher modes, all of the problems show an entirely different behavior. (...) Instead of a gradual increase of all the higher modes, the energy is exchanged, essentially, among only a certain few. It is, therefore, very hard to observe the rate of 'thermalization' or mixing in our problem, and this was the initial purpose of the calculation.

Bisogna dire che i risultati dei nostri calcoli numerici presentano caratteristiche che ci sorpresero fin dal principio. Invece di un continuo, progressivo flusso di energia dal primo modo ai modi piú alti, tutto indica un comportamento completamente diverso. (...) Invece di un progressivo aumento di energia di tutti i modi alti, l'energia è scambiata, essenzialmente, solo da pochi di essi. E' pertanto ben difficile osservare il ritmo di 'termalizzazione', o mescolamento nel nostro sistema, che pure era lo scopo iniziale del calcolo.

La mancanza di equipartizione è evidente in figura 7.1, che riporta le energie medie dei modi normali, più precisamente i rapporti $\bar{E}_k(T)/E$, in funzione di T (in scala logaritmica), per un modello con $\varepsilon \simeq 0.026$, precisamente $n = 32$, $r = 3$, $m = 1$, $\Omega = 1$, $\mu = 0.1$, $\mathcal{E} = 0.07$; il tempo massimo di integrazione è qui⁴⁷⁾ $T_M \simeq 2.2 \times 10^6$. Le prime quattro curve, dall'alto in basso, rappresentano $\bar{E}_1(T), \dots, \bar{E}_4(T)$, mentre l'ultima rappresenta la somma $\bar{E}_5(T) + \dots + \bar{E}_{32}(T)$. Come si vede, l'energia (data tutta inizialmente al primo modo) è condivisa sostanzialmente da pochi modi normali, e certo non in egual misura.

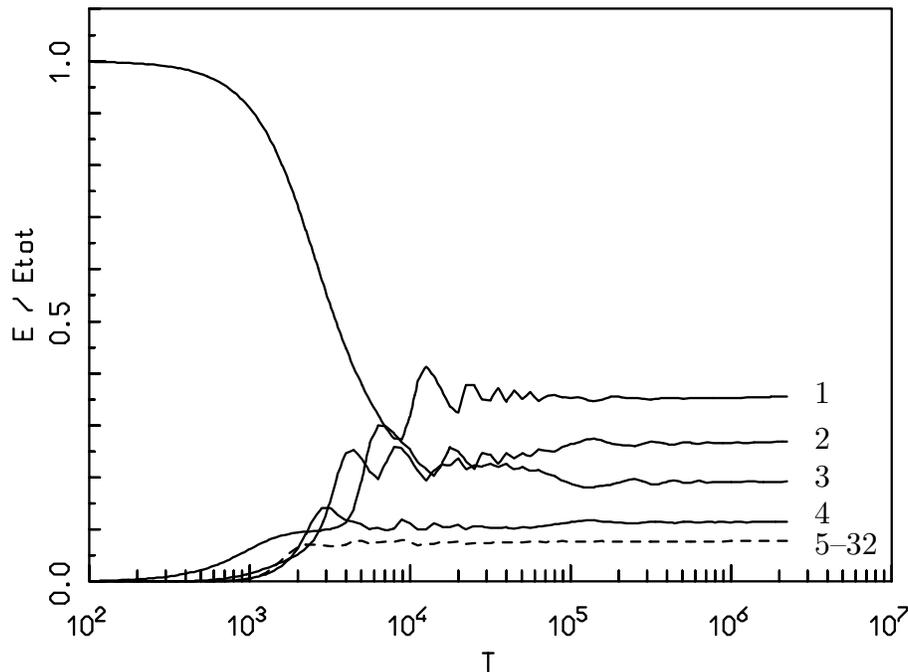


Figura 7.1. I rapporti $\bar{E}_k(T)/E$, $k = 1, \dots, 4$ (linea continua), e la somma $\sum_{k=5}^{32} \bar{E}_k(t)/E$ (linea tratteggiata), in funzione di T , per il modello di FPU con $n = 32$, $r = 3$, $m = 1$, $\Omega = 1$, $\mu = 0.1$ e $\mathcal{E} = 0.07$.

Un'immagine piú precisa del carattere della dinamica in queste condizioni si può avere guardando al comportamento delle energie istantanee $E_k(t)$ dei modi normali maggiormente coinvolti nella dinamica. La figura 7.2a rappresenta $E_k(t)$ in funzione di t , per $k=1,2$ e 3 . E' evidente la presenza di una "ricorrenza" nel sistema, per tempi molto superiori al tempo proprio microscopico $\Omega^{-1} = 1$: ad esempio, la ripartizione dell'energia riappare simile a quella iniziale, anche se non identica, per $T \simeq 10^4$ e $T \simeq 2 \times 10^4$, mentre

⁴⁷⁾ Tempi di integrazione così lunghi erano inaccessibili ai calcolatori degli anni '50, mentre oggi si raggiungono facilmente con pochi minuti di calcolo su macchine di media potenza.

ricorrenze più precise (Tuck, 1968) si trovano su tempi più lunghi, come si può vedere dalla figura 7.2b, che riporta $E_1(t)$ in funzione di t , esattamente come nella figura precedente, ma su di un arco di tempo circa 20 volte più grande.

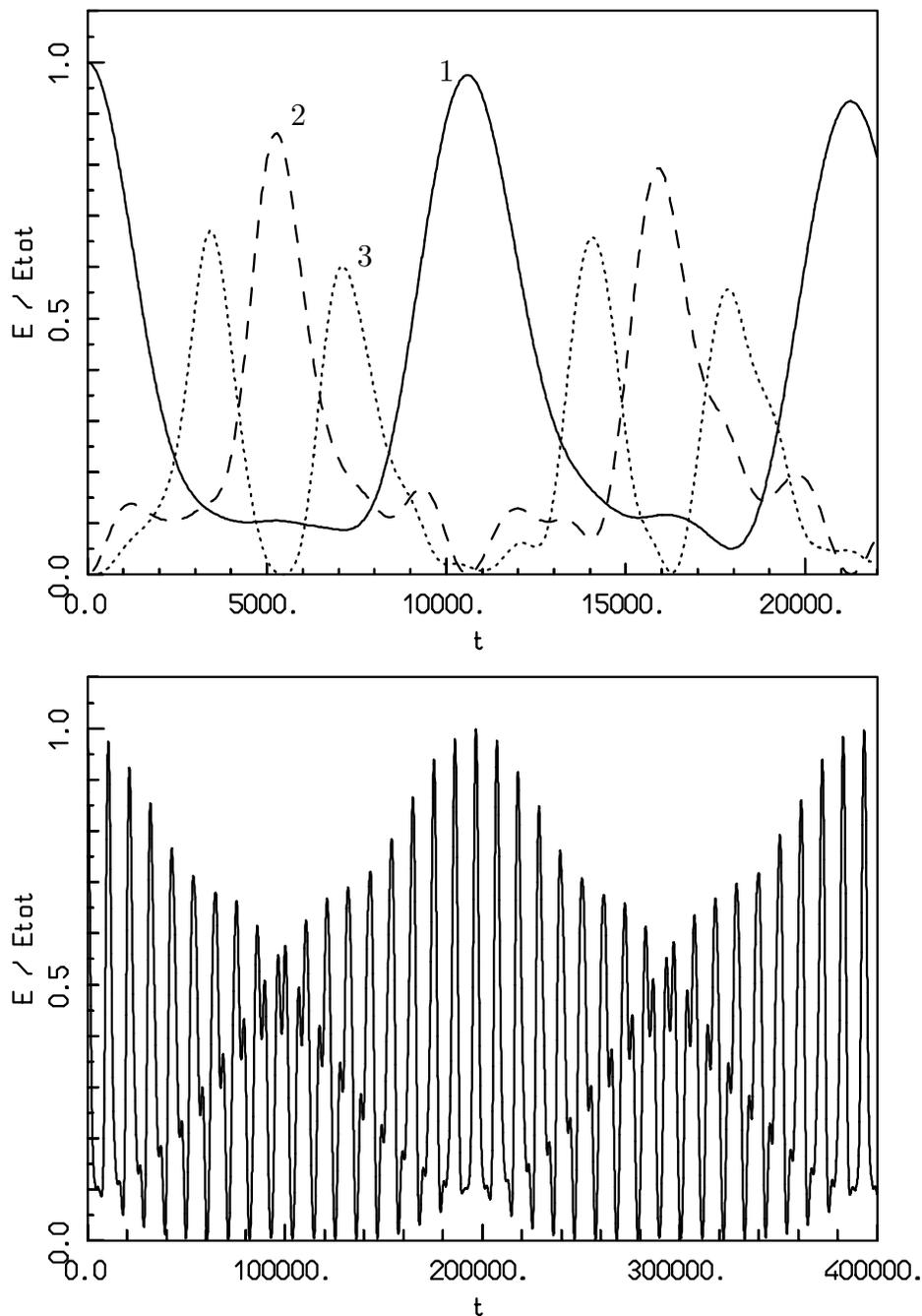


Figura 7.2. Il fenomeno della ricorrenza. In alto: energie istantanee $E_k(t)$, per $k = 1, 2, 3$, in funzione di t , nelle stesse condizioni della figura precedente. In basso: energia istantanea $E_1(t)$, ancora nelle stesse condizioni, su una scala di tempo più lunga.

Dall'insieme di questi risultati numerici si ha l'impressione che il problema ergodico, in un sistema di interesse fisico come il modello di FPU, sia tutt'altro che banale, e le proprietà

fisicamente più fondamentali, come la buona circolazione dell'energia tra i modi normali del sistema, siano tutt'altro che scontate: almeno sui tempi esplorati in questi esperimenti numerici, non si trova alcun segno che il sistema stia in qualche modo evolvendo verso una situazione di equilibrio, o perdendo in qualche senso la memoria della speciale condizione iniziale assegnata.

Che ciò fosse invece, per così dire, a priori possibile per un simile modello, su questa stessa scala dei tempi, si può vedere cambiando μ o l'energia, in modo da aumentare a sufficienza la nonlinearità.⁴⁸⁾ Per ε abbastanza grande, infatti, il comportamento del modello progressivamente cambia, fino a che, per $\varepsilon \simeq 0.1$, si ottiene (sulla medesima scala di tempo) una situazione di buona equipartizione dell'energia tra tutti i modi normali. Questa situazione è illustrata in figura 7.3, che si riferisce allo stesso modello con $\mathcal{E} = 1.2$ (e tutti gli altri parametri invariati), e riporta $\bar{E}_k(T)$ in funzione di T , per $k = 1, \dots, 32$ (per dare maggiore chiarezza alla figura, anche la scala delle energie è qui logaritmica).

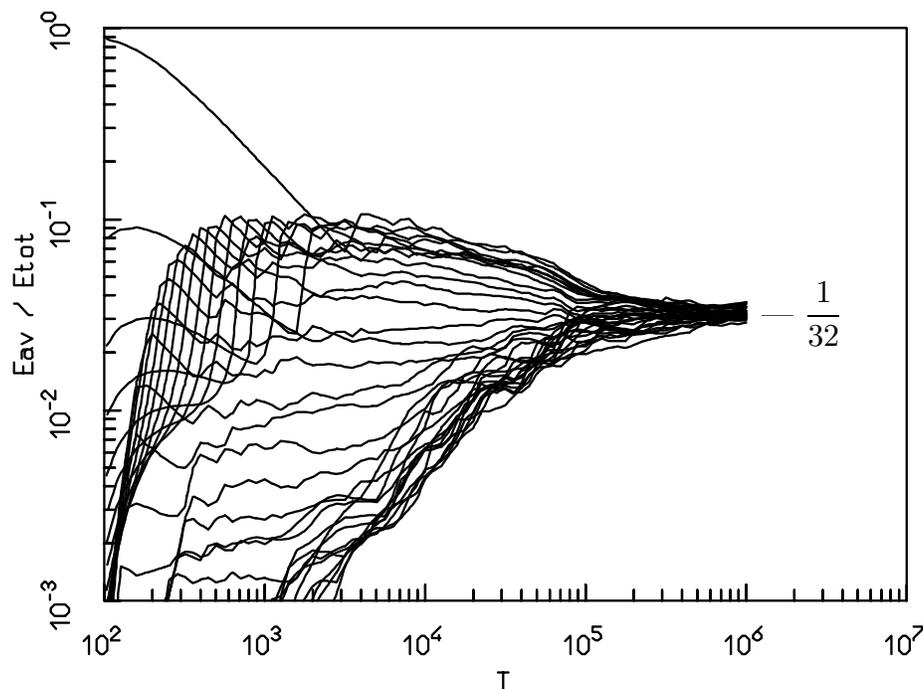


Figura 7.3. Il comportamento di $\bar{E}_k(T)/E$, $k = 1, \dots, 32$, per nonlinearità più elevata, precisamente $\mathcal{E} = 1.2$ e n, r, m, Ω e μ come sopra.

Questo strano comportamento, rivelato con tale chiarezza dal calcolo numerico, apre la porta a diverse questioni, non facili ed anzi piuttosto sottili. La prima di esse riguarda l'affidabilità del calcolo numerico su tempi di integrazione così lunghi. A questo problema, veramente delicato, non ci possiamo qui dedicare come esso meriterebbe, e pertanto ci limiteremo ad una sola considerazione: purchè il calcolo sia abbastanza accurato (piccoli passi di integrazione, buon algoritmo), i risultati si presentano ben stabili al variare della macchina, dell'algoritmo, dei parametri in esso introdotti. Questo elementare criterio di affidabilità — osservare la stabilità dei risultati al variare degli “accidenti” introdotti nella procedura di calcolo numerico — è di regola assai significativo.

⁴⁸⁾ Stranamente, gli autori lavorarono sempre a basse nonlinearità, e non si accorsero del fenomeno che andiamo ora a descrivere, che fu osservato solo nel 1966 da Chirikov e collaboratori.

Una seconda naturale questione riguarda la rilevanza dei fenomeni osservati per la meccanica statistica: 32 o 64 particelle sono ben poche rispetto al numero di molecole contenute in una infima porzione di materia ordinaria, e si potrebbe congetturare che la mancanza di ergodicità scompaia aumentando n (a fissato E/n). Va detto che questa possibilità si presenta così semplice e naturale, che fu la prima ad essere prospettata dai fisici intervenuti sul problema. Fino ad oggi, tuttavia, non si sono mai trovate indicazioni chiare che la crescita di n muti in modo significativo il comportamento del sistema: variazioni di n di un fattore 10 o 100 apparentemente non introducono novità sostanziali.

Un ultimo problema riguarda la scala di tempo, anch'essa ovviamente assai limitata, accessibile alle esperienze numeriche. E' infatti a priori possibile interpretare tutti i risultati sopra riportati, pensando che l'equilibrio statistico si raggiunga comunque, ma su scale di tempo che, per bassa non linearità, sono così lunghe da non poter essere viste nell'ambito delle possibilità attuali del calcolo numerico. Diversi tentativi sono stati fatti anche in questo senso, e per molti anni, nonostante la crescita della potenza di calcolo delle macchine, al crescere del tempo di integrazione non si è trovato, si può dire, niente di nuovo. Di recente però (Kantz *et. al.*, 1994) si sono trovati nuovi risultati (da considerarsi ancora preliminari) che rendono nuovamente incerta la situazione: sembra che una lenta evoluzione verso l'equipartizione avvenga anche a bassa non linearità, ma i tempi a cui si ottiene (in un senso precisabile) un fissato "grado di equipartizione" non solo possono diventare anche molto grandi al calare della non linearità, ma apparentemente crescono con il numero di gradi di libertà, rendendo assai incerta l'interpretazione fisica del fenomeno.⁴⁹⁾

In conclusione: la situazione — a oltre quarant'anni dai primi risultati numerici, e a dispetto della crescita enorme della potenza di calcolo che si è avuta in quest'arco di tempo — ancora non è chiara, e il minimo che si può dire è che il problema posto allora da Fermi, Pasta e Ulam (con coraggio non indifferente, a fronte di risultati tutto sommato ancora poveri) ancor oggi resiste non abbastanza capito.

7.2. Il modello di Hénon e Heiles

Vediamo qui un altro studio numerico diventato famoso, precisamente lo studio eseguito nel 1964 da Hénon e Heiles su un modello (apparentemente) elementare a due soli gradi di libertà, costituito da due oscillatori armonici accoppiati da un termine non lineare.

Il modello di Hénon e Heiles nasce da una motivazione di carattere astronomico, lo studio delle proprietà statistiche del moto di una stella in una galassia a simmetria assiale (il moto di una stella nel campo medio di tutte le altre, che si suppone assegnato e a simmetria assiale). Il problema ha tre gradi di libertà, ma la conservazione del momento angolare consente di ridurre il numero effettivo di gradi di libertà a due soli. Ci si mette allora in prossimità di un punto di equilibrio stabile del sistema ridotto (un'orbita circolare del problema completo) e lì attorno si trova, evidentemente, una coppia di oscillatori armonici (i due modi normali del sistema) accoppiati da un termine non lineare.

Con questa motivazione alle spalle (motivazione peraltro abbastanza irrilevante dal nostro punto di vista) gli autori decisero di studiare l'hamiltoniana

$$H = \frac{1}{2}(p_1^2 + q_1^2) + \frac{1}{2}(p_2^2 + q_2^2) + q_1^2 q_2 - \frac{1}{3}q_2^3 ;$$

⁴⁹⁾ Sul problema della possibile esistenza di scale di tempo lunghe nella dinamica di un sistema hamiltoniano debolmente accoppiato avremo occasione di tornare nel terzo capitolo.

come si vede, massa, pulsazioni e costante di accoppiamento sono poste uguali a uno, cosicché la non-linearità (in accordo a quanto sopra visto per il modello di Fermi, Pasta e Ulam) è misurata dalla stessa energia.

Il problema è di capire se, oltre all'energia (e al momento angolare, di cui però si è già tenuto conto al momento di ridurre il sistema), esiste una ulteriore quantità regolare⁵⁰⁾ conservata, nel qual caso il moto è necessariamente ristretto a una superficie a $2n - 2 = 2$ dimensioni, immersa nella superficie di energia costante a $2n - 1 = 3$ dimensioni, ed è ovviamente non ergodico; o se invece, in assenza di altri integrali del moto (regolari) oltre all'energia, l'orbita può invadere liberamente la superficie di energia costante.

Il comportamento del sistema alle diverse energie si studia bene con il metodo della *sezione di Poincaré*, cui si è già accennato nel paragrafo 2.1. In questo caso è comodo prendere come sezione l'iperpiano Π di equazione $q_1 = 0$, orientato nel verso delle q_1 crescenti, e considerare la mappa Ψ che a ogni punto $x \in \Pi$ associa la prima intersezione con Π stesso (nel verso positivo, ovvero con $\dot{q}_1 = p_1 > 0$) dell'orbita uscente da x . In pratica, a ogni orbita continua del sistema resta associata una successione discreta di punti, la successione delle sue intersezioni con Π (con $p_1 > 0$). Ora, per la conservazione dell'energia, i punti di intersezione appartengono sicuramente ad una superficie bidimensionale Σ , intersezione di Π con la superficie di energia costante, e dunque si possono rappresentare con due coordinate, ad esempio p_2 e q_2 (fissate tali coordinate, lo stato del sistema nello spazio delle fasi è ben definito, perchè si ha $q_1 = 0$, mentre $p_1 > 0$ è univocamente determinato dalla conservazione di E). Dovendo risultare $p_1^2 = 2E - p_2^2 - q_2^2 + \frac{2}{3}q_2^3 > 0$, segue immediatamente che le successive intersezioni sono limitate a priori alla regione del piano q_2, p_2 compresa tra le curve $p_2 = \pm \sqrt{2E - q_2^2 + \frac{2}{3}q_2^3}$. Tornando alla domanda che ci siamo posti, ovvero se esista o meno l'ulteriore integrale del moto, è chiaro che la sua eventuale esistenza costringe le intersezioni generate da una stessa orbita a soddisfare una equazione in più, e dunque a stare su di una curva regolare contenuta in Σ , mentre se l'integrale manca le successive intersezioni possono invadere l'intera zona accessibile.

La figura 7.4a mostra il risultato che si ottiene per diverse condizioni iniziali a energia sufficientemente bassa, in questo caso $E = 0.08$; come si vede, le intersezioni delle diverse orbite vanno a formare curve ben definite contenute in Σ , cioè il secondo integrale del moto sembra esistere, e il sistema sembra essere decisamente non ergodico. A energie di poco superiori tuttavia la situazione cambia drasticamente: come mostra la figura 7.4b, che si riferisce a $E = 0.125$, per alcune condizioni iniziali ritroviamo la situazione precedente, mentre per altre appare un comportamento nuovo: i punti di intersezione, non più allineati in curve, vanno spargendosi su di un'area bidimensionale (tutti quelli tracciati in figura provengono da una stessa orbita). Questa regione ove le curve sono scomparse è chiamata "regione caotica": praticamente inesistente fino a $E \simeq 0.1$, essa cresce rapidamente in misura al crescere di E , fino a diventare dominante per $E \simeq 0.166666$, come si vede in figura 7.4c (per valori di E superiori a $1/6$ si vede facilmente che le superfici di energia costante non sono più compatte, e l'indagine perde di senso). La figura 7.4d mostra una rozza stima dell'area ricoperta dalle curve regolari, al variare di E . Vediamo così il progressivo instaurarsi di un comportamento di tipo ergodico al crescere della non-linearità;

⁵⁰⁾ Ovviamente l'analisi numerica deve restringersi a funzioni regolari, mentre di per sé la classe delle funzioni rilevanti per la teoria ergodica è la classe, immensa e orribile, delle funzioni misurabili. Va detto che qui si tocca un punto delicato nell'ambito della rilevanza fisica della teoria ergodica.

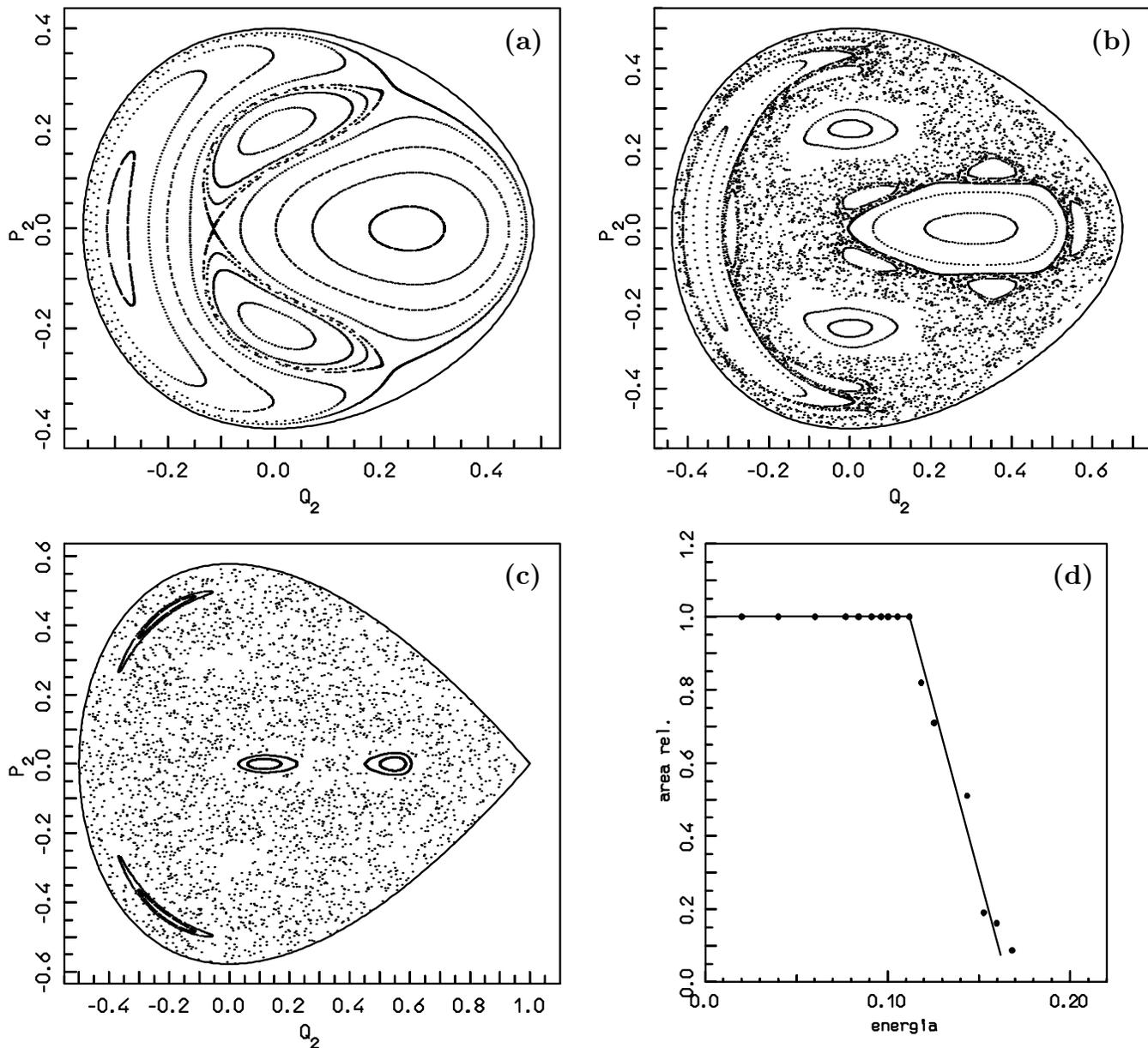


Figura 7.4. La mappa di Poincaré per il sistema di Hénon e Heiles: (a–c) sezioni a $E = 0.08, 0.125, 0.166666$; (d) proporzione della regione caotica al variare di E .

per molti aspetti è lo stesso fenomeno che abbiamo visto precedentemente a proposito del modello di FPU. Va anche detto che il comportamento del modello di Hénon e Heiles non è eccezionale, ma tipico dei sistemi dinamici a due gradi di libertà in condizioni di debole accoppiamento (più precisamente sistemi quasi integrabili, si veda il terzo capitolo).

Il modello di Hénon e Heiles ha avuto un grande impatto nella ricerca sui sistemi dinamici, almeno in ambito fisico: ha mostrato infatti con disarmante chiarezza che sistemi dinamici dall'aria semplice e innocente, come due oscillatori armonici accoppiati da un termine cubico, possono avere un comportamento inaspettatamente complicato, e ha così contribuito a far capire che l'immagine della meccanica classica quale era tramandata

da una generazione di fisici all'altra era nel complesso povera. Come avremo modo di discutere nel secondo e terzo capitolo, negli stessi anni la ricerca matematica sui sistemi dinamici faceva passi da gigante, consentendo una comprensione assai più profonda di quella accessibile tramite il calcolo numerico (e anzi, come vedremo, rivelando una ricchezza e una complicazione ben maggiore di quella che appare dagli studi numerici). Ma il minimo che si possa dire è che studi numerici come quello di Hénon e Heiles, facendo in qualche modo vedere e toccare i moti caotici,⁵¹⁾ hanno avuto il merito di far capire, soprattutto tra i fisici, la rilevanza dello studio dei sistemi dinamici non lineari.

Per un'esposizione più ampia di questi e altri risultati numerici si veda ad esempio M. Hénon, *Numerical exploration of dynamical systems*, lezioni tenute alla Scuola di Les Houches 1980: *Chaotic behavior of deterministic systems* (North-Holland, Amsterdam 1982).

8. Introduzione alla dinamica simbolica

8.1. Partizioni e dinamica simbolica

Abbiamo considerato fino ad ora mappe $\Phi^t : M \rightarrow M$, ove lo spazio M ha come struttura primaria di riferimento la misura. In questo breve paragrafo di introduzione alla dinamica simbolica faremo invece riferimento solo occasionalmente alla presenza di una misura invariante, attribuendo invece a M la semplice struttura di spazio metrico. Scopo del paragrafo è solo quello di prender familiarità con la nozione di dinamica simbolica (che prescinde per sé da quale sia la struttura di riferimento su M) e così preparare la strada alla successiva importante nozione di *entropia*, nella quale invece la misura gioca nuovamente un ruolo essenziale.⁵²⁾

Definizione. La coppia (M, Φ) , con M spazio metrico separabile completo e Φ homeomorfismo: $M \rightarrow M$ (ovvero: gruppo a un parametro di homeomorfismi $\Phi^t : M \rightarrow M$, $t \in \mathbb{R}$), è detta *sistema dinamico topologico* discreto (ovvero continuo, detto anche flusso topologico).

Il caso tipico da avere in mente è quello di una varietà compatta dotata di metrica (varietà riemanniana), ma non è l'unico; in particolare, anche l'insieme astratto di stringhe a n simboli, che per evitare conflitti di notazione indicheremo qui

$$\Sigma = \{ \sigma = (\dots, \sigma_{-1}, \sigma_0, \sigma_1, \dots), \sigma_t \in I = \{0, \dots, n-1\} \},$$

⁵¹⁾ C'è chi crede senza bisogno di vedere, e chi per credere ha bisogno di vedere. L'autore di queste note ha deciso di dedicarsi ai sistemi dinamici dopo aver visto.

⁵²⁾ Ma vi sono varianti interamente topologiche, e assai interessanti anche se forse più astratte, dell'idea di entropia.

si può dotare di una metrica, ponendo ad esempio

$$\text{dist}(\sigma, \sigma') = \sum_{t \in \mathbb{Z}} 2^{-|t|} \delta(\sigma_t, \sigma'_t), \quad \delta(\sigma_t, \sigma'_t) = \begin{cases} 0 & \text{per } \sigma_t = \sigma'_t \\ 1 & \text{per } \sigma_t \neq \sigma'_t \end{cases} \quad (8.1)$$

Si osservi che con tale metrica l'applicazione $\Psi : \Sigma \rightarrow \Sigma$ introdotta a proposito degli schemi di Bernoulli, precisamente la traslazione a sinistra definita da

$$(\Psi(\sigma))_t = \sigma_{t+1},$$

è un homeomorfismo. Ci restringiamo di qui in avanti al caso discreto; dal caso continuo ci si può riportare al caso discreto semplicemente considerando la mappa al tempo uno, $\Phi = \Phi^1$.

Consideriamo allora una partizione finita di M , ovvero una collezione

$$\alpha = \{A_0, \dots, A_{n-1}\}, \quad A_i \subset M,$$

tale che

$$A_i \cap A_j = \emptyset, \quad \bigcup_{0 \leq i < n} A_i = M;$$

gli elementi A_0, \dots, A_{n-1} sono anche detti *atomi* della partizione. Possiamo pensare alla partizione come a una "osservazione discreta" del sistema, ovvero a una operazione di misura con un insieme finito di esiti *a priori* possibili. Gli atomi A_i dovranno essere "ben fatti" con riferimento alla struttura importante di M : così, nel caso della misura, dovranno essere insiemi misurabili; nel caso topologico che qui abbiamo introdotto non possiamo chiedere che siano aperti (perchè non si può fare una partizione in aperti), ma chiederemo che siano prossimi ad aperti, diciamo contenuti nella chiusura del loro interno;⁵³⁾ nell'eventualità poi che M , come nei sistemi classici, abbia struttura di varietà, si chiederà in aggiunta che il bordo degli atomi sia unione di sottovarietà.

Fissata dunque una partizione α , per ogni dato iniziale $x \in M$ resta univocamente individuata la *storia* di x , precisamente la stringa

$$\sigma(x) = \{\sigma_t, t \in \mathbb{Z}\}, \quad \sigma_t = \sigma_t(x) \in I,$$

definita da

$$\sigma_t(x) = l \in I \quad \text{se} \quad \Phi^t(x) \in A_l. \quad (8.2)$$

Al variare di x in M si ottiene così un sottoinsieme Σ_M dell'insieme Σ di tutte le stringhe a n simboli. La corrispondenza (8.2) tra x e $\sigma(x)$, cioè la definizione di $\sigma : M \rightarrow \Sigma_M \subset \Sigma$, è chiaramente fatta in modo che all'applicazione $\Phi : M \rightarrow M$ resti *coniugata* l'applicazione $\Psi : \Sigma \rightarrow \Sigma$ sopra considerata:

$$\Psi \circ \sigma = \sigma \circ \Phi.$$

⁵³⁾ Incontreremo più avanti, nei prossimi capitoli, insiemi ben diversi: insiemi di Cantor, con interno vuoto, ma non numerabili e anche di misura positiva (e grande). Li incontreremo non come bizzarrie, ma come oggetti naturali della teoria, in connessione ad esempio a sistemi hamiltoniani (apparentemente) semplicissimi come il modello di Hénon e Heiles precedentemente considerato.

Questa dinamica in uno spazio di stringhe di simboli, che resta associata al sistema dinamico considerato tramite la partizione prescelta, prende il nome di *dinamica simbolica*.

Si pongono allora naturalmente due problemi:

- i. preliminarmente, studiare la biunivocità della corrispondenza tra M e la sua immagine Σ_M in Σ (punti diversi potrebbero avere la stessa storia);
- ii. studiare in qualche modo le proprietà delle stringhe effettivamente realizzate ($\sigma \in \Sigma_M$), e per questa via caratterizzare i moti realizzati dal sistema (M, Φ) , osservati attraverso α .

Per formalizzare il primo problema, diamo la seguente

Definizione. La partizione α si dice *separante* per il sistema (M, Φ) , se punti diversi hanno necessariamente storie diverse:

$$\sigma(x) = \sigma(y) \implies x = y .$$

Le partizioni separanti sono dunque, per definizione, quelle per cui il primo problema non si pone (l'applicazione σ è iniettiva). Una tipica condizione sufficiente perché un sistema dinamico topologico (M, Φ) ammetta partizioni separanti, è che il sistema sia *espansivo*:

Definizione. Il sistema dinamico topologico discreto (M, Φ) si dice *espansivo*, se esiste $\varepsilon > 0$ tale che

$$\text{dist}(\Phi^t(x), \Phi^t(y)) < \varepsilon \quad \forall t \in \mathbb{Z} \implies x = y .$$

E' immediato vedere che se (M, Φ) è espansivo allora esistono partizioni separanti (basta prendere atomi con diametro minore di ε). E' espansivo ad esempio il "Gatto di Arnol'd"; è espansivo lo stesso sistema dinamico (Σ, Ψ) , se si fa riferimento alla metrica (8.1): infatti, se σ e σ' sono diverse, esiste $t \in \mathbb{Z}$, tale che $\sigma_t \neq \sigma'_t$; ma allora $\text{dist}(\Psi^t(\sigma), \Psi^t(\sigma')) \geq 1$, e questo basta.

8.2. Stringhe statisticamente regolari; complessità di una stringa

Veniamo ora al secondo problema, assai interessante. Questo può essere affrontato in modi e con punti di vista diversi. Qui cercheremo soltanto di caratterizzare in qualche modo⁵⁴⁾ le singole stringhe di caratteri associate ai suoi moti, più che altro per prendere dimestichezza con la dinamica simbolica; lo studio globale del sistema nel suo complesso è demandato al prossimo paragrafo (ove faremo di nuovo riferimento pieno alla misura, e introdurremo allo scopo la nozione di entropia).⁵⁵⁾

⁵⁴⁾ In modo in realtà non molto raffinato; per una trattazione più approfondita, si veda G. Gallavotti, *Aspetti della teoria ergodica, qualitativa e statistica del moto* (Pitagora, Bologna 1981).

⁵⁵⁾ Può essere utile, per cogliere la logica della parte restante del paragrafo, pensare alla stringa in esame come a un (lungo!) testo, cui applichiamo una primitiva analisi linguistica.

Preso allora $x \in M$, sia $\sigma(x) = (\dots, \sigma_{-1}, \sigma_0, \sigma_1, \dots)$ la sua storia. Diremo che $\sigma(x)$ contiene il *tratto* finito di base k_1, \dots, k_m e specificazione l_1, \dots, l_m , denotato

$$\begin{bmatrix} l_1, \dots, l_m \\ k_1, \dots, k_m \end{bmatrix},$$

se $\sigma_{k_1} = l_1, \dots, \sigma_{k_m} = l_m$, ovvero se $\sigma(x) \in \mathcal{C}_{k_1, \dots, k_m}^{l_1, \dots, l_m}$, ovvero ancora se

$$x \in \bigcap_{1 \leq i \leq m} \Phi^{-k_i}(A_{l_i}).$$

Due tratti di ugual lunghezza si diranno *omologhi*, se hanno identica specificazione, e basi k_1, \dots, k_m e k'_1, \dots, k'_m traslate ($k'_i = k_i + s$ per ogni i , con opportuno s).

Prendiamo allora una stringa σ ; una prima idea per studiarla può essere quella di guardare alla frequenza ν con cui appaiono tratti omologhi a un tratto assegnato: per questo poniamo (se esiste)

$$\nu\left(\begin{bmatrix} l_1, \dots, l_m \\ k_1, \dots, k_m \end{bmatrix}\right) = \lim_{t \rightarrow \infty} \frac{N_t}{t},$$

con N_t = numero di indici s tra 0 e $t-1$, tali che $\sigma_{s+k_i} = l_i$, $i = 1, \dots, m$. Diremo che la stringa σ ha frequenze definite, o anche che è *statisticamente regolare*, se il limite esiste per ogni scelta del tratto.

- **Osservazione.** Si vede subito che per la stringa $\sigma(x)$ si ha

$$\nu\left(\begin{bmatrix} l_1, \dots, l_m \\ k_1, \dots, k_m \end{bmatrix}\right) = \text{tempo di soggiorno di } \Phi^t(x) \text{ in } \bigcap_{1 \leq i \leq m} \Phi^{-k_i}(A_{l_i});$$

perciò (teorema di Birkhoff–Kinchin) se (M, Φ) ammette una misura conservata μ , allora per quasi ogni x la stringa $\sigma(x)$ ha frequenze definite. Se poi (M, Φ) , dotato della misura μ , è ergodico, allora per quasi ogni x tali frequenze sono indipendenti da x , e uguali alla misura di $\bigcap_{1 \leq i \leq m} \Phi^{-k_i}(A_{l_i})$. Viceversa, si potrebbe vedere che ogni stringa $\sigma(x)$ a frequenze definite induce una misura μ_x su M , tale che al generico insieme $\bigcap_{1 \leq i \leq m} \Phi^{-k_i}(A_{l_i})$ resta assegnata una misura uguale alla frequenza del tratto corrispondente (la σ -algebra degli insiemi misurabili è quella generata dagli atomi A_i e dai loro iterati $\Phi^{-t}(A_i)$, $i = 0, \dots, n-1$, $t \in \mathbb{Z}$).

Una volta stabilite le frequenze con cui appaiono i singoli tratti, queste si possono usare naturalmente come probabilità *a priori* dei tratti, senza bisogno che sia assegnata in precedenza una misura su M o su Σ . In particolare si può studiare la funzione di correlazione temporale di due tratti qualsiasi $\begin{bmatrix} l_1, \dots, l_m \\ k_1, \dots, k_m \end{bmatrix}$ e $\begin{bmatrix} l'_1, \dots, l'_{m'} \\ k'_1, \dots, k'_{m'} \end{bmatrix}$, definita nel modo naturale:

$$\mathcal{G}(t) = \nu\left(\begin{bmatrix} l_1, \dots, l_m, & l'_1, \dots, l'_{m'} \\ k_1, \dots, k_m, t+k'_1, \dots, t+k'_{m'} \end{bmatrix}\right) - \nu\left(\begin{bmatrix} l_1, \dots, l_m \\ k_1, \dots, k_m \end{bmatrix}\right) \nu\left(\begin{bmatrix} l'_1, \dots, l'_{m'} \\ k'_1, \dots, k'_{m'} \end{bmatrix}\right)$$

(si confronti \mathcal{G} con l'analogia funzione introdotta nel paragrafo 5.3). Se la correlazione decade, ovvero se

$$\lim_{t \rightarrow \infty} \mathcal{G}(t) = 0$$

per ogni scelta dei tratti, la stringa si dirà *mescolante*.⁵⁶⁾

Una stringa può essere molto semplice, ad esempio essere periodica, o essere più complessa. Si può introdurre una nozione formale di *complessità*⁵⁷⁾ di una stringa (o se vogliamo del corrispondente moto di (M, Φ) visto attraverso α), cercando di caratterizzare in qualche modo la “varietà” dei tratti finiti in essa presenti. Possiamo limitarci ai tratti con indici di base consecutivi, ovvero tratti omologhi a

$$\left[\begin{array}{c} l_0, \dots, l_{t-1} \\ 0, \dots, t-1 \end{array} \right], \quad (8.3)$$

detti *t*-*storie*. Sia N_t il numero di *t*-*storie* presenti nella stringa con frequenza non nulla, distinte tra loro, cioè non omologhe; tale numero è soggetto all’ovvia limitazione

$$N_t \leq n^t \quad (8.4)$$

(l’uguaglianza si ha quando tutti i tratti di lunghezza *t* che si possono scrivere con gli *n* simboli a disposizione sono effettivamente presenti nella stringa), e anche

$$N_{t+s} \leq N_t N_s \quad (8.5)$$

(l’uguaglianza si ha quando ciascuno dei tratti di lunghezza *t* presenti nella stringa può essere seguito da uno qualunque dei tratti di lunghezza *s*). Avendo in mente la crescita massima (8.4), possiamo definire la complessità *H* della stringa ponendo⁵⁸⁾

$$H(\sigma) = \lim_{t \rightarrow \infty} \frac{1}{t} \text{Log } N_t, \quad \text{Log} := \log_2; \quad (8.6)$$

si dimostrerà che il limite esiste sempre. La scelta di \log_2 al posto del logaritmo naturale, ad esso proporzionale, è del tutto irrilevante: è naturale in teoria dell’informazione, nell’ambito della quale tutte queste cose sono nate, ma in letteratura si trova, per problemi di questo genere, indifferentemente l’una o l’altra scelta. Per la (8.4), *H* è limitata da

$$H(\sigma) \leq \text{Log } n.$$

Poniamo

$$h_t = \frac{1}{t} \text{Log } N_t;$$

⁵⁶⁾ Ha anche senso dire che la singola stringa σ è *ergodica*, chiedendo per questo che risulti

$$\lim_{t \rightarrow \infty} t^{-1} \sum_{s=0}^{t-1} \mathcal{G}(s) = 0.$$

Si potrebbe vedere che se (M, Φ) , munito *a priori* della misura μ , è ergodico (mescolante), e la partizione α è misurabile, allora per quasi ogni x la stringa $\sigma(x)$ è ergodica (mescolante).

⁵⁷⁾ Complessità è un termine dal multiforme significato (non sempre ben definito), che varia a seconda del contesto in cui viene usato. Oggi è frequentemente associato, ma in modo vago, ai cosiddetti fenomeni caotici (essi stessi definiti in modo vago). Vago non significa poco interessante (H.A. Kramers: *Nel regno del pensiero umano in generale, e nella scienza fisica in particolare, i concetti più fecondi sono quelli a cui è impossibile associare un significato ben definito*). Significa però che bisogna esercitare una certa vigilanza.

⁵⁸⁾ Vi sono varianti a questa definizione, in realtà piuttosto ingenua; si veda il testo sopra citato di G. Gallavotti.

la (8.5) si traduce allora nella relazione di *sub-additività*

$$h_{t+s} \leq \frac{t}{t+s} h_t + \frac{s}{t+s} h_s . \quad (8.7)$$

La sub-additività è una proprietà che in un certo senso generalizza la monotonicità: è immediato verificare che ogni successione monotona non crescente è sub-additiva, mentre vi sono successioni sub-additive non monotone (ad esempio $h_t = 0$ per t pari, $h_t = t^{-1}$ per t dispari). La dimostrazione dell'esistenza del limite (8.6) si riporta al seguente facile lemma, che estende alle successioni sub-additive una ben nota proprietà delle successioni monotone:

Lemma 8.1. *Preso una qualunque successione sub-additiva inferiormente limitata h_1, h_2, \dots , esiste il limite*

$$h = \lim_{t \rightarrow \infty} h_t ,$$

e tale limite coincide con l'estremo inferiore $h = \inf\{h_t, t \in \mathbb{Z}^+\}$.

Dimostrazione. Bisogna dimostrare che per t sufficientemente grande la differenza $h_t - h$ è arbitrariamente piccola. Dalla definizione di h come inferiore della successione si sa che preso comunque $\varepsilon > 0$ sicuramente esiste $s = s(\varepsilon)$ tale che

$$h_s < h + \varepsilon ;$$

dalla sub-additività (8.7) segue subito $h_{js} \leq h_s$, e quindi

$$h_t < h + \varepsilon \quad \text{per } t = js, \quad j > 0 .$$

Sia ora t compreso tra js e $(j+1)s$, diciamo $t = js + k$ con $0 < k < s$. Dalla (8.7) segue immediatamente

$$h_t \leq \frac{js}{js+k} h_{js} + \frac{k}{js+k} h_k < h + \varepsilon + \frac{1}{j+1} h_k ,$$

e d'altra parte, ancora dalla (8.7), risulta $h_k \leq h_1$: pertanto

$$h_t < h + 2\varepsilon$$

non appena $j \geq h_1/\varepsilon$, e dunque per ogni $t \geq T(\varepsilon) := h_1 s(\varepsilon)/\varepsilon$. L'esistenza del limite è così dimostrata. ■

Prese due qualsiasi partizioni $\alpha = \{A_0, \dots, A_{n-1}\}$ e $\alpha' = \{A'_0, \dots, A'_{n'-1}\}$, si definisce l'operazione di *composizione* “ \vee ”, ponendo

$$\alpha \vee \alpha' = \{A_i \cap A'_j \neq \emptyset, 0 \leq i < n, 0 \leq j < n'\}$$

(l'operazione è con evidenza simmetrica e associativa). Sia allora

$$\beta_t = \alpha \vee \Phi^{-1}(\alpha) \vee \dots \vee \Phi^{-t+1}(\alpha) = \bigvee_{0 \leq s \leq t-1} \Phi^{-s}(\alpha) ,$$

avendo posto

$$\Phi^{-t}(\alpha) = \{\Phi^{-t}(A_0), \dots, \Phi^{-t}(A_{n-1})\}.$$

Poniamo poi $n_t =$ cardinalità di β_t . E' evidente che vale la limitazione $N_t \leq n_t$; nel caso ergodico (e con atomi misurabili) $N_t = n_t$ per quasi ogni x . Questo fatto può essere utile in qualche caso per calcolare H .

- **Esercizio.** Si dimostri che per la rotazione di \mathbb{T}^1 , $x \mapsto \Phi(x) = x + \delta \pmod{1}$, la complessità di ogni orbita è nulla per ogni partizione α in intervalli [*suggerimento*: se δ è razionale, le orbite sono periodiche, e la complessità è certamente nulla. Se δ è irrazionale, si deduce facilmente $n_t \leq tn$, ove n è il numero di intervalli della partizione iniziale].
- **Esercizio.** Si studi la complessità delle orbite per la trasformazione di \mathbb{T}^1 definita da $x \mapsto 2x \pmod{1}$, con partizione iniziale di \mathbb{T}^1 in $[0, \frac{1}{2})$, $[\frac{1}{2}, 1)$.

Si vede immediatamente che n_t soddisfa le limitazioni, analoghe alle (8.4) e (8.5), $n_t \leq n^t$ e $n_{t+s} \leq n_t n_s$; si può essere allora tentati di introdurre una quantità del tipo

$$\tilde{H} = \lim_{t \rightarrow \infty} \frac{1}{t} \text{Log } n_t$$

per misurare la frammentazione della partizione β_t al crescere di t , e così caratterizzare in qualche modo non più la singola stringa o la singola orbita, ma la dinamica nel suo complesso (il limite sicuramente esiste perché, come è immediato verificare, $\tilde{H}_t = \frac{1}{t} \log n_t$ è sub-additiva, e dunque si applica il lemma 8.1). Ma è un'idea troppo ingenua, che non porta lontano;⁵⁹⁾ per avere sviluppi interessanti occorre qualche cosa di più raffinato. La via che seguiremo, spontanea nell'ambito della teoria ergodica, sarà quella di reintrodurre nel problema la misura, e tramite la misura valutare la "frammentazione" di una partizione in modo più elaborato del puro conteggio degli insiemi.⁶⁰⁾

⁵⁹⁾ In particolare non è facile liberarsi dalla dipendenza, assai pesante, dalla partizione iniziale α , e così arrivare a una vera proprietà intrinseca del sistema (invariante per isomorfismo). Facilmente infatti avviene che, raffinando α , \tilde{H} diverge, per il formarsi di un gran numero di insiemi molto piccoli.

⁶⁰⁾ Come già si è accennato, esistono formulazioni interamente topologiche: ma comunque occorrono metodi più sofisticati del puro conteggio. Si veda, in un testo di sistemi dinamici, la nozione di *entropia topologica*.

9. L'entropia di Kolmogorov–Sinai

9.1. Informazione media di una operazione di misura

Riprendiamo qui i sistemi dinamici (M, μ, Φ) muniti di misura, classici o astratti, restringendoci però ancora, provvisoriamente, al caso discreto. Prenderemo in considerazione partizioni

$$\alpha = \{A_0, \dots, A_{n-1}\}$$

misurabili; per evitare piccoli fastidi è bene considerare equivalenti e identificare due partizioni $\alpha = \{A_0, \dots, A_{n-1}\}$ e $\beta = \{B_0, \dots, B_{n-1}\}$ se (con opportuno ordinamento) gli atomi A_i e B_i , $i = 0, \dots, n-1$, sono identici a meno di insiemi di misura nulla: $\mu(A_i - B_i) = 0$ (si tratta con evidenza di una relazione di equivalenza). In teoria dell'informazione è naturale associare a α il numero

$$\eta(\alpha) = - \sum_{i=0}^{n-1} p_i \operatorname{Log} p_i, \quad p_i = \mu(A_i), \quad 0 \operatorname{Log} 0 := 0,$$

chiamato *entropia* della partizione. E' utile anche la scrittura

$$\eta = \sum_{i=0}^{n-1} \rho(p_i), \quad \rho(p) := -p \operatorname{Log} p;$$

il grafico di ρ è tracciato schematicamente in figura 9.1.

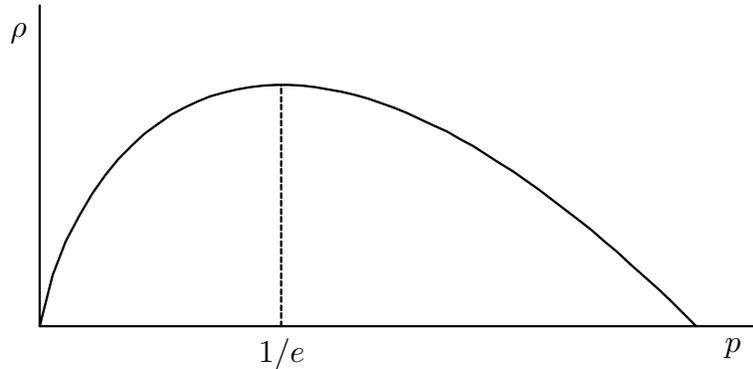


Figura 9.1. Il grafico della funzione $\rho(p) = -p \operatorname{Log} p$.

Cerchiamo di interpretare la definizione. L'idea è quella di valutare la “incertezza *a priori*” di una operazione di misura che abbia n esiti A_0, \dots, A_{n-1} possibili, mutuamente escludentisi, con probabilità *a priori* p_0, \dots, p_{n-1} ; a sua volta, tale incertezza si può interpretare come informazione media, o informazione attesa, fornita dall'operazione di misura una volta che questa venga effettuata. Così ad esempio, l'incertezza η dovrà essere nulla quando uno degli eventi è certo, massima in condizioni di totale uniformità; per $n = 2$, l'incertezza dovrà essere crescente per

$$(p_0, p_1) = (0, 1), \quad (0.01, 0.99), \quad \left(\frac{1}{3}, \frac{2}{3}\right), \quad \left(\frac{1}{2}, \frac{1}{2}\right).$$

Il significato dell'entropia η sopra definita traspare anche da alcune sue proprietà elementari, che qui introduciamo dopo aver dato la seguente

Definizione. Due partizioni

$$\alpha = \{A_0, \dots, A_{n-1}\}, \quad \beta = \{B_0, \dots, B_{m-1}\}$$

si dicono *independenti*⁶¹⁾ se per ogni coppia A_i, B_j di atomi risulta $\mu(A_i \cap B_j) = \mu(A_i)\mu(B_j)$. Si dice poi che β è più fine di α , o che è un *raffinamento* di α , e si denota

$$\beta \succeq \alpha,$$

se ciascun atomo di β è interamente contenuto in un solo atomo di α (sempre a meno di insiemi di misura nulla): ovvero se per ogni j esiste i tale che $\mu(B_j \cap A_i) = \mu(B_j)$ (e di conseguenza $\mu(B_j \cap A_{i'}) = 0$ per $i' \neq i$).

Come esempio tipico di partizioni indipendenti si pensi a due partizioni di un rettangolo in bande orizzontali e verticali, oppure, per gli schemi di Bernoulli, alle partizioni $\alpha_k = \{C_k^0, \dots, C_k^{n-1}\}$ con diversi k .

Proposizione 9.1. *La funzione η ha le seguenti proprietà:*

- i. η si annulla se e solo se uno dei p_i vale uno;
- ii. η è massima per $p_0 = \dots = p_{n-1} = 1/n$, e vale $\eta_{\max} = \text{Log } n$;
- iii. per partizioni α e β indipendenti, η si somma.⁶²⁾

$$\eta(\alpha \vee \beta) = \eta(\alpha) + \eta(\beta);$$

- iv. Infine, se $\beta \succeq \alpha$, allora $\eta(\beta) \geq \eta(\alpha)$.

Interviene nella dimostrazione una elementare proprietà delle funzioni concave,⁶³⁾ detta *disuguaglianza di Jensen*, che estraiamo in un lemma perché sarà utile anche nel seguito.

Lemma 9.2. *Se $f : \mathbb{R} \rightarrow \mathbb{R}$ è concava, allora si ha*

$$f\left(\sum_{i=1}^n c_i x_i\right) \geq \sum_{i=1}^n c_i f(x_i) \quad \text{per} \quad \sum_{i=1}^n c_i = 1. \quad (9.1)$$

Dimostrazione. Per $n = 2$ la disuguaglianza coincide con la definizione stessa di funzione concava. Si trasporta facilmente a $n > 2$ per induzione. ■

Veniamo allora alla dimostrazione della proposizione.

⁶¹⁾ Si pensi a due operazioni di misura indipendenti: ciascuna non modifica le aspettative dell'altra.

⁶²⁾ Vedremo che si generalizza in $\eta(\alpha \vee \beta) \leq \eta(\alpha) + \eta(\beta)$, per partizioni non necessariamente indipendenti.

⁶³⁾ Ricordiamo che $f : \mathbb{R} \rightarrow \mathbb{R}$ si dice *convessa* nell'intervallo $I \subset \mathbb{R}$ se l'insieme $\{(x, y) \in \mathbb{R}^2 : x \in I, y \geq f(x)\}$ è convesso, ovvero se per ogni $x, x' \in I$ e ogni $c, 0 \leq c \leq 1$, si ha $f(cx + (1-c)x') \leq cf(x) + (1-c)f(x')$; si dice *concava* se $-f$ è convessa.

Dimostrazione. Il primo punto segue dal solo fatto che la funzione ρ sopra definita è nulla per $p = 0, 1$ e positiva per $0 < p < 1$. Il secondo punto segue dalla sola concavità di ρ : basta applicare la disuguaglianza di Jensen (9.1) con $c_i = 1/n$, $x_i = p_i$, e ricordare poi $\sum_i p_i = 1$. Il terzo punto è un banale calcolo, nel quale però si sfrutta la forma esplicita di ρ : se p_0, \dots, p_{n-1} e q_0, \dots, q_{m-1} sono le misure degli atomi di α e β , nell'ipotesi di indipendenza si scrive subito

$$\begin{aligned}\eta(\alpha \vee \beta) &= -\sum_{i,j} p_i q_j (\text{Log } p_i + \text{Log } q_j) = \eta(\alpha) \sum_j q_j + \eta(\beta) \sum_i p_i \\ &= \eta(\alpha) + \eta(\beta) .\end{aligned}$$

Infine, per il quarto punto, basta vedere (procedendo per raffinamenti successivi) che se un atomo A di α , di misura p , è unione di due atomi B e B' di β , di misura cp e $(1-c)p$, allora $\rho(cp) + \rho((1-c)p) \geq \rho(p)$. La disuguaglianza segue dalla sola concavità di ρ , assieme a $\rho(0) = 0$: infatti, in base a queste proprietà si ha

$$\rho(cp) = \rho((1-c)0 + cp) \geq (1-c)\rho(0) + c\rho(p) = c\rho(p) ,$$

e allo stesso modo $\rho((1-c)p) = (1-c)\rho(p)$; la conclusione è immediata. ■

- **Osservazione.** Si può dimostrare⁶⁴⁾ che queste quattro proprietà caratterizzano la funzione η , precisamente *individuano completamente η , a meno di una costante moltiplicativa* (corrispondente all'arbitrarietà della scelta della base del logaritmo).
- **Osservazione.** Se vi sono due soli esiti possibili, *a priori* equiprobabili (testa o croce), allora $\eta = 1$: la misura fornisce una *informazione elementare*, comunemente presa come unità di informazione e chiamata *bit*. Quattro, otto... esiti equiprobabili (ovvero due, tre... misure indipendenti) forniscono due, tre... *bits* di informazione. Se vi sono tre esiti A , B e C , con probabilità $\frac{1}{2}$, $\frac{1}{4}$ e $\frac{1}{4}$, allora con probabilità $\frac{1}{2}$ la misura fornisce un *bit* (esito A), con probabilità $\frac{1}{4}$ due *bits* (esito B), e ancora con probabilità $\frac{1}{4}$ due *bits* (esito C); in tutto $\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = -\sum p_i \text{Log } p_i = \frac{3}{2}$ *bits* (la probabilità uniforme avrebbe fornito $\text{Log } 3 \simeq 1.58$ *bits*).
- **Osservazione.** L'espressione di η corrisponde alla definizione microscopica di entropia termodinamica introdotta da Boltzmann (si veda il paragrafo 1.2, nota 8), se la densità di occupazione f_j di ciascuna celletta dello spazio μ si interpreta come probabilità *a priori* della celletta.

9.2. Entropia di Kolmogorov–Sinai

A partire dalla quantità η sopra introdotta, si definisce una grandezza h chiamata *entropia di Kolmogorov–Sinai* del sistema, o semplicemente *entropia*.⁶⁵⁾ La nozione di entropia ha grande importanza nell'ambito della teoria ergodica.

- La nozione è invariante per isomorfismo, dunque riflette proprietà intrinseche, e consente una classificazione “quantitativa” dei sistemi dinamici.

⁶⁴⁾ Si veda ad esempio A.I. Kinchin, *Mathematical foundations of information theory*, Dover (New York, 1957).

⁶⁵⁾ E' frequente anche l'espressione *entropia metrica*: ma il termine “metrica” si riferisce alla misura, non alla metrica, e si contrappone alla *entropia topologica*, nella quale la misura non interviene.

- E' legata alla teoria dell'informazione, e come vedremo si può anche interpretare come indicatore della “varietà di orbite”, o della ricchezza della dinamica di un sistema.
- E' legata a quantità di significato immediato e ben misurabili sia numericamente sia sperimentalmente, precisamente ai cosiddetti *esponenti caratteristici*, o *esponenti di Lyapunov* del sistema (indicatori della divergenza esponenziale di traiettorie vicine, di cui si dirà più avanti).

Sia dato allora un sistema dinamico discreto (M, μ, Φ) ; sia $\alpha = \{A_0, \dots, A_{n-1}\}$ una partizione misurabile di M , e sia

$$\beta_t = \alpha \vee \Phi^{-1}(\alpha) \vee \dots \vee \Phi^{-t+1}(\alpha)$$

la partizione composta definita nel paragrafo precedente. Chiameremo *entropia del sistema relativa alla partizione α* iniziale, il limite (che si dimostra esistere)

$$h(\Phi, \alpha) = \lim_{t \rightarrow \infty} \frac{1}{t} \eta(\beta_t) . \quad (9.2)$$

- **Osservazione.** La quantità $\eta(\beta_t)$ si può interpretare come informazione attesa da t misure successive della stessa quantità, effettuate a intervalli regolari di tempo. L'entropia $h(\Phi, \alpha)$ assume allora il significato di informazione media prodotta dalla singola misura.
- **Osservazione.** Un valore positivo di $h(\Phi, \alpha)$ indica che $\beta_t = \alpha \vee \dots \vee \Phi^{-t+1}(\alpha)$ va rapidamente frammentandosi con t , a ritmo “mediamente esponenziale”: precisamente, se denotiamo con $B(t, x)$ l'atomo di β_t che contiene x , e poniamo

$$g_t(x) = -\frac{1}{t} \text{Log } \mu(B(t, x)) , \quad (9.3)$$

allora (tenendo conto che g è costante a tratti) la definizione di h si può anche riscrivere

$$h = \lim_{t \rightarrow \infty} \int_M g_t d\mu . \quad (9.4)$$

In questo senso (piuttosto debole) si può dire che, mediamente, c'è contrazione esponenziale della misura degli atomi di β_t . Vedremo più avanti un risultato assai più forte di questo tipo (teorema di Shannon–McMillan–Brian).

- **Osservazione.** Per quanto visto nel paragrafo precedente, a ogni atomo di β_t corrisponde una diversa t -storia effettivamente realizzata dal sistema (la t -storia dei dati iniziali x in quell'atomo). Si può allora dire che $h(\Phi, \alpha)$ rappresenta il “ritmo di crescita” con t dell'insieme delle t -storie del sistema, non però semplicemente contate in numero, ma in qualche modo pesate con la misura degli insiemi che le producono. Nel caso (assolutamente eccezionale) di insiemi sempre di ugual misura per ogni t ci si riporta al conteggio, perché allora il limite (9.2) si riduce a $h = \lim_{t \rightarrow \infty} \frac{1}{t} \text{Log } n_t$, con $n_t =$ cardinalità di $\beta_t =$ numero di t -storie distinte.

Per dimostrare che il limite (9.2) esiste, ci è utile il seguente

Lemma 9.3. *Comunque si prendano due partizioni misurabili*

$$\alpha = \{A_0, \dots, A_{n-1}\}, \quad \beta = \{B_0, \dots, B_{m-1}\},$$

risulta

$$\eta(\alpha \vee \beta) \leq \eta(\alpha) + \eta(\beta), \quad (9.5)$$

e l'uguaglianza si ha se e solo se le partizioni sono indipendenti.

Interviene nella dimostrazione una naturale definizione, che vale la pena di mettere in evidenza:

Definizione. La quantità

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)}$$

si dice *misura di A condizionata a B*.

Non è altro che la frazione di B ricoperta da A ; $\mu(A|B) = \mu(A)$ per atomi qualsiasi di partizioni indipendenti.

Dimostrazione. Dalla definizione di η , facendo riferimento alla misura condizionata $\mu(A|B)$, si scrive subito

$$\begin{aligned} \eta(\alpha \vee \beta) &= -\sum_{ij} \mu(A_i \cap B_j) \text{Log} (\mu(B_j) \mu(A_i|B_j)) \\ &= -\sum_{ij} \mu(A_i \cap B_j) \text{Log} \mu(B_j) - \sum_{ij} \mu(A_i \cap B_j) \text{Log} \mu(A_i|B_j) \\ &= \eta(\beta) - \sum_{ij} \mu(B_j) \mu(A_i|B_j) \text{Log} \mu(A_i|B_j) \end{aligned} \quad (9.6)$$

(si è usato $\sum_i \mu(A_i \cap B_j) = \mu(B_j)$). Il secondo termine a destra ha la forma

$$\sum_i \left[\sum_j c_j \rho(p_{ij}) \right],$$

con

$$c_j = \mu(B_j), \quad p_{ij} = \mu(A_i|B_j), \quad \rho(p) = -p \text{Log} p;$$

applicando la disuguaglianza di Jensen (9.1) per ogni indice i e sommando su i si vede subito che tale termine non supera complessivamente $\eta(\alpha)$, e dunque è verificata la (9.5). D'altra parte (per la convessità stretta di ρ), la disuguaglianza di Jensen si traduce in uguaglianza se e solo se tutti i p_{ij} per ogni i non dipendono da j , ovvero se $\mu(A_i|B_j)$ non dipende da j , e ciò a sua volta si verifica se e solo se le partizioni sono indipendenti. Ciò conclude la dimostrazione del lemma. ■

Usando questo lemma, assieme al lemma 8.1 già provato nel paragrafo precedente, si dimostra assai facilmente che

Proposizione 9.4. *Il limite (9.2) esiste.*

Dimostrazione. Dalla definizione di β_t e dal lemma 9.3 segue senza difficoltà

$$\eta(\beta_{t+s}) \leq \eta(\beta_t) + \eta(\Phi^{-t}(\beta_s)) = \eta(\beta_t) + \eta(\beta_s).$$

Posto allora $\tilde{h}_t = \frac{1}{t}\eta(\beta_t)$, si ha

$$\tilde{h}_{t+s} \leq \frac{t}{t+s} \tilde{h}_t + \frac{s}{t+s} \tilde{h}_s ,$$

e per il lemma 8.1 il limite esiste. ■

- **Osservazione.** Si potrebbe in realtà dimostrare qualche cosa di più, precisamente che la successione $\frac{1}{t}\eta(\beta_t)$ è monotona non crescente.

Proposizione 9.5. *Si consideri lo schema di Bernoulli $B_{p_0, \dots, p_{n-1}}$, e sia*

$$\alpha = \{\mathcal{C}_0^0, \dots, \mathcal{C}_0^{n-1}\}$$

una partizione dell'insieme Σ delle stringhe in cilindri elementari. Allora risulta

$$h(\Phi, \alpha) = \eta(\alpha) = - \sum_{i=0}^{n-1} p_i \text{Log } p_i .$$

Dimostrazione. Poiché $\Phi^{-t}(\mathcal{C}_0^l) = \mathcal{C}_t^l$, le partizioni $\alpha, \Phi^{-1}(\alpha), \dots, \Phi^{-t+1}(\alpha)$ sono mutuamente indipendenti. Allora, posto $\beta_t = \alpha \vee \dots \vee \Phi^{-t+1}(\alpha)$, si ha

$$\eta(\beta_t) = \sum_{s=0}^{t-1} \eta(\Phi^{-s}(\alpha)) = t \eta(\alpha) ,$$

e la conclusione è immediata. ■

- **Esercizio.** Si dimostri che per uno schema di Bernoulli le partizioni $\alpha = \{\mathcal{C}_0^l, l \in I\}$ e $\beta = \{\mathcal{C}_{0, \dots, s-1}^{l_0, \dots, l_{s-1}}, l_0, \dots, l_{s-1} \in I\}$, $s > 1$, danno la stessa entropia (ovvero: può essere inutile prendere partizioni troppo raffinate). Più in generale, per sistemi qualsiasi, α e $\beta = \alpha \vee \dots \vee \Phi^{-s+1}(\alpha)$, $s > 1$, danno la stessa entropia.

Infine, l'entropia del sistema dinamico $h(\Phi)$ si definisce prendendo il superiore sulle partizioni misurabili:

Definizione. Si dice *entropia di Kolmogorov–Sinai* del sistema dinamico discreto (M, μ, Φ) il numero (eventualmente infinito)

$$h(\Phi) = \sup_{\alpha \text{ misurabile}} h(\Phi, \alpha) .$$

Si verifica immediatamente che

Proposizione 9.6. *Sistemi dinamici isomorfi hanno la stessa entropia.*

Dimostrazione. Partizioni corrispondenti hanno la medesima entropia, e la conclusione è immediata. ■

- **Esercizio.** Si dimostri che, nel caso invertibile, $h(\Phi) = h(\Phi^{-1})$ [suggerimento: $\eta(\beta_t)$ e $\eta(\Phi^{t-1}(\beta_t))$ sono uguali].
- **Esercizio.** Si dimostri che $h(\Phi^s) = |s|h(\Phi)$ [suggerimento: per ogni α , posto $\tilde{\alpha} = \alpha \vee \dots \vee \Phi^{-s+1}(\alpha)$, risulta $h(\Phi^s, \tilde{\alpha}) = |s|h(\Phi, \alpha)$. D'altra parte, la ricerca del superiore sulle partizioni si può certamente restringere alle partizioni del tipo $\tilde{\alpha}$].

Riprendiamo i sistemi dinamici continui, che fino ad ora abbiamo lasciato da parte. Si può dimostrare che per ogni t reale fissato il sistema dinamico discreto (M, μ, Φ^t) ha entropia $h(\Phi^t) = |t|h(\Phi^1)$; questo giustifica la

Definizione. Si definisce entropia di Kolmogorov–Sinai $h(\Phi)$ del sistema dinamico continuo (M, μ, Φ) la quantità $h(\Phi^1)$ relativa al corrispondente sistema dinamico discreto (M, μ, Φ^1) .

9.3. Alcuni teoremi sull'entropia di Kolmogorov–Sinai

In questo paragrafo esponiamo alcuni dei risultati più significativi sull'entropia di Kolmogorov–Sinai; per ulteriori risultati, e per la prova di quanto non dimostrato, si rinvia ai testi specializzati. Premettiamo alcune nozioni e due utili lemmi, dimostrati nell'appendice C.

Definizione. Date due partizioni misurabili

$$\alpha = \{A_0, \dots, A_{n-1}\}, \quad \beta = \{B_0, \dots, B_{m-1}\},$$

si dice *entropia relativa* di α rispetto a β la quantità

$$\begin{aligned} \eta(\alpha | \beta) &= \sum_j \mu(B_j) \sum_i \rho(\mu(A_i | B_j)) \\ &= -\sum_j \mu(B_j) \sum_i \mu(A_i | B_j) \text{Log} \mu(A_i | B_j). \end{aligned}$$

Rileggendo la (9.6) si vede bene che si ha

$$\eta(\alpha | \beta) = \eta(\alpha \vee \beta) - \eta(\beta); \quad (9.7)$$

questa espressione, che si sarebbe anche potuta usare come definizione di entropia relativa, invita a interpretare l'entropia relativa come “guadagno di informazione” della misura combinata $\alpha \vee \beta$ rispetto alla sola misura β . Il lemma 9.3 a sua volta si riscrive nella forma, assai trasparente,

$$\eta(\alpha | \beta) \leq \eta(\alpha). \quad (9.8)$$

L'interesse principale per l'entropia relativa è che essa consente di introdurre la nozione di *distanza* tra partizioni,

$$\text{dist}(\alpha, \beta) = \eta(\alpha | \beta) + \eta(\beta | \alpha),$$

e con ciò di dare all'insieme \mathcal{P} delle partizioni misurabili la struttura di spazio metrico. Precisamente si dimostra (appendice C) che

Lemma 9.7.

- i. La quantità $\text{dist}(\alpha, \beta)$ sopra definita ha le proprietà di una distanza nello spazio \mathcal{P} delle partizioni misurabili di M .
- ii. Per ogni sistema dinamico (M, μ, Φ) , l'entropia $h(\Phi, \alpha)$ è funzione continua⁶⁶⁾ di α , precisamente si ha $|h(\Phi, \alpha) - h(\Phi, \beta)| \leq \text{dist}(\alpha, \beta)$.

La presenza di una metrica in \mathcal{P} e la continuità di h consentono di calcolare o stimare l'entropia sostituendo a partizioni complicate, quali possono essere quelle in atomi soltanto misurabili, partizioni più semplici ad esse prossime.

Definizione. Sia M una varietà differenziabile. Una partizione α di M si dice *classica* se ciascun atomo della partizione è unione finita di *poliedri* (o *complessi*) semiaperti, ovvero di insiemi A semplicemente connessi, tali che $A \subset \overline{A}$, e con il bordo ∂A regolare a tratti.

Si dimostra facilmente (appendice C) che

Lemma 9.8. Per ogni varietà compatta M , munita di una misura regolare μ , le partizioni classiche sono dense nello spazio \mathcal{P} delle partizioni misurabili, ovvero per ogni partizione misurabile α di M e ogni $\varepsilon > 0$ esiste una partizione classica β , tale che $\text{dist}(\alpha, \beta) < \varepsilon$.

Possiamo ora introdurre e dimostrare l'importante

Proposizione 9.9 (Teorema di Kouchnirenko). L'entropia dei sistemi dinamici classici è finita.

Dimostrazione. Ci limitiamo per semplicità al caso in cui la mappa Φ è un diffeomorfismo $M \rightarrow M$ (caso invertibile e regolare ovunque, non solo a tratti). Innanzitutto, poiché le partizioni classiche sono dense in \mathcal{P} e h è continua, nel prendere il superiore sulle partizioni ci si può limitare alle partizioni classiche. L'idea, da formalizzare opportunamente, è poi abbastanza semplice: si dimostra che

- i. la superficie complessiva degli atomi della partizione composta

$$\beta_t = \alpha \vee \dots \vee \Phi^{-t+1}(\alpha) = \{B_0^t, \dots, B_{n_t-1}^t\},$$

definita come unione delle superfici dei singoli atomi, cresce al più esponenzialmente⁶⁷⁾ con t : $\mathcal{S}(\beta_t) \leq (\text{cost})\lambda^t$;

- ii. a sua volta, una maggiorazione sulla superficie si traduce in una maggiorazione sull'entropia di una partizione, del tipo $\eta(\beta) \leq (\text{cost}) + m \text{Log } \mathcal{S}(\beta)$, $m = \dim M$ (ovvero: frammentare M in modo da far crescere η ha un "costo" minimo in superficie). Segue immediatamente una maggiorazione del tipo $\eta(\beta_t) \leq (\text{cost}) + tm \text{Log } \lambda$, da cui $h(\Phi, \alpha) \leq m \text{Log } \lambda$ per ogni α . Si osservi che il numero di atomi (cui ci siamo occasionalmente riferiti a proposito della complessità) non è invece soggetto a una simile limitazione geometrica; la misura perciò gioca un ruolo decisivo in questo teorema.

⁶⁶⁾ In effetti lipschitziana.

⁶⁷⁾ *Esercizio*: si prenda la trasformazione del panettiere, si consideri la partizione di M in due bande verticali di larghezza $\frac{1}{2}$, e si valuti $\mathcal{S}(\beta_t)$.

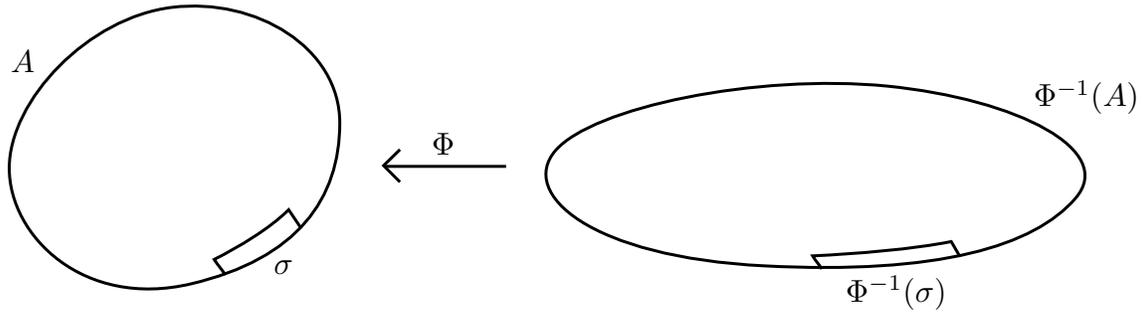


Figura 9.2. L'atomo A e la sua immagine inversa, $\Phi^{-t}(A)$. Per la conservazione del volume, la porzione σ del bordo di A non può crescere, in $\Phi^{-1}(A)$, più di quanto un sottile strato appoggiato su A non cali di spessore per effetto di Φ^{-1} , ovvero più di quanto Φ non dilati le lunghezze.

Per il primo punto, innanzitutto è sempre possibile introdurre su M una metrica “adattata” alla misura, cioè tale che il volume di un insieme coincida con la sua misura.⁶⁸⁾ Sia ora λ il massimo coefficiente di dilatazione lineare⁶⁹⁾ di Φ in tale metrica. Allora per ogni atomo A di α si ha $\mathcal{S}(\Phi^{-1}(A)) \leq \lambda \mathcal{S}(A)$, ove \mathcal{S} denota l'area; infatti, il volume si conserva, e “la base non si può dilatare più di quanto si contragga l'altezza” (si veda la figura 9.2). Pertanto,

$$\mathcal{S}(\Phi^{-t}(A)) \leq \lambda^t \mathcal{S}(A) .$$

Per una qualsiasi partizione classica α , denotiamo con $\mathcal{S}(\alpha)$ la somma delle aree dei suoi atomi (ciascuna faccia sarà contata due volte). Si riconosce allora immediatamente che risulta $\mathcal{S}(\alpha \vee \beta) \leq \mathcal{S}(\alpha) + \mathcal{S}(\beta)$, e iterando

$$\mathcal{S}(\alpha \vee \Phi^{-1}(\alpha) \vee \dots \vee \Phi^{-t+1}(\alpha)) \leq (1 + \lambda + \dots + \lambda^{t-1}) \mathcal{S}(\alpha) = \frac{\lambda^t - 1}{\lambda - 1} \mathcal{S}(\alpha) \quad (9.9)$$

(vale anche per $\lambda = 1$, corrispondente alle traslazioni che nella metrica adattata appaiono rigide, pur di intendere in questo caso per continuità $(\lambda^t - 1)/(\lambda - 1) = t$). Questa è la maggiorazione cercata alla crescita dell'area.

Passiamo ora al secondo punto, ovvero dalla maggiorazione sull'area deduciamo una maggiorazione su η . E' noto che per ogni varietà m -dimensionale, e per ogni scelta della metrica riemanniana, vale una *disuguaglianza isoperimetrica*: ovvero esistono costanti C e \bar{V} , tali che se $\text{Vol}(A) \leq \bar{V}$ allora⁷⁰⁾

$$\text{Vol}(A) \leq C \mathcal{S}(A)^{\frac{m}{m-1}} .$$

⁶⁸⁾ Su di una varietà regolare si può sempre introdurre una metrica riemanniana, e con essa un volume; se ρ (regolare e positiva) è la densità della misura conservata in tale metrica, allora con un riscaldamento locale del tensore metrico, che a g sostituisce $\tilde{g} = \rho^{1/m} g$, la nuova densità diventa uno.

⁶⁹⁾ Se $D\Phi_x$ denota l'applicazione tangente $T_x M \rightarrow T_{\Phi(x)} M$, il coefficiente di dilatazione lineare di Φ in x si può definire come la norma (operatoriale) di $D\Phi_x$.

⁷⁰⁾ In \mathbb{R}^3 il solido di volume massimo, a parità di superficie, è la sfera, e $V = (4/3)\pi r^3$, $\mathcal{S} = 4\pi r^2$, eccetera; su una superficie sferica la figura di massima area è la calotta circolare, e tra il volume V (l'area della calotta) e la superficie \mathcal{S} (la lunghezza del bordo) si trova facilmente la relazione $V \leq (2\pi)^{-1} \mathcal{S}^2$, purché però la calotta non sia più grande di mezza sfera.

Sia ora α una qualunque partizione; ai fini del calcolo di h non è restrittivo supporre $\text{Vol}(A_i) \leq \bar{V}$ per ogni suo atomo A_i . Con una successione di passaggi banali (si usa, tra il secondo e il terzo, la concavità del logaritmo), per una generica partizione α si ottiene [notazione: $p_i = \mu(A_i) = \text{Vol}(A_i)$]

$$\begin{aligned} \eta(\alpha) &= -\sum_i p_i \text{Log } p_i = m \sum_i p_i \text{Log } p_i^{-\frac{1}{m}} \\ &\leq m \text{Log} \left(\sum_i p_i p_i^{-\frac{1}{m}} \right) = m \text{Log} \left(\sum_i p_i^{\frac{m-1}{m}} \right) \\ &\leq m \text{Log} \left(C^{\frac{m-1}{m}} \sum_i \mathcal{S}(A_i) \right) = m \text{Log} (\mathcal{S}(\alpha)) + (m-1) \text{Log } C . \end{aligned}$$

Questo è il legame cercato tra $\eta(\alpha)$ e $\mathcal{S}(\alpha)$. Riprendendo la (9.9) si trova allora

$$\begin{aligned} \eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha)) &\leq m \text{Log} (\mathcal{S}(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha))) + (m-1) \text{Log } C \\ &\leq m \text{Log} \frac{\lambda^t - 1}{\lambda - 1} + m \text{Log} \mathcal{S}(\alpha) + (m-1) \text{Log } C ; \end{aligned}$$

di qui, dividendo per t e passando al limite per $t \rightarrow \infty$, si ottiene la maggiorazione, uniforme in α , $h(\Phi, \alpha) \leq m \text{Log } \lambda$ (buona anche per $\lambda = 1$). L'entropia del sistema $h(\Phi)$ è pertanto limitata, e anzi maggiorata anch'essa da

$$h(\Phi) \leq m \text{Log } \lambda . \quad (9.10)$$

■

- **Osservazione.** La (9.10), con cui abbiamo concluso la dimostrazione, mostra che il coefficiente di dilatazione lineare λ fornisce una stima da sopra all'entropia. Il legame tra entropia e dilatazione lineare non è casuale: incontreremo più avanti una relazione più precisa (*formula di Pesin*), che lega h agli *esponenti di Lyapunov* del sistema dinamico classico. Per il "Gatto di Arnol'd" la (9.10) afferma che

$$h \leq 2 \text{Log } \lambda ,$$

ove $\lambda = \frac{1}{2}(3 + \sqrt{5})$ è l'autovalore massimo di $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$; il risultato esatto, basato sulla formula di Pesin, è $h = \text{Log } \lambda$.

La disuguaglianza (9.10) implica in particolare che

Proposizione 9.10. *La traslazione e il moto uniforme sul toro \mathbb{T}^n hanno entropia nulla.*

Dimostrazione. Per questi sistemi si ha $\lambda = 1$. ■

Esistono sistemi dinamici con entropia infinita (e dunque non isomorfi a sistemi classici), come è illustrato nel seguente

- **Esercizio.** Si costruisca uno schema di Bernoulli a infiniti simboli $B_{p_0, p_1, \dots}$, con la stringa delle p_i tale che $h(\Phi) = \infty$. [Suggerimento: si parte da una successione di probabilità q_k del tipo $1/n_k$ (n_k intero), $k = 1, 2, \dots$, tale che $\sum_k q_k = 1$: ad esempio $n_k = 2^k$. Poi

la probabilità q_k si suddivide in 2^{n_k} parti uguali, ciascuna grande $2^{-n_k}/n_k$, e si allineano i numeri così ottenuti. Si considera la partizione (finita) α_s costituita dagli s cilindri elementari $\{\mathcal{C}_0^l, l = 1, \dots, s-1\}$ assieme a $\tilde{\mathcal{C}} = \{\sigma \in \Sigma : \sigma_0 \geq s\}$. Si vede facilmente che $\eta(\alpha_s)$ diverge per $s \rightarrow \infty$; se ne deduce che anche $h(\Phi, \alpha_s)$ diverge.]

Proseguiamo con una definizione, preliminare a un teorema importante:

Definizione. Per un sistema dinamico (M, μ, Φ) la partizione α si dice *generante*, se gli atomi della partizione assieme ai loro iterati (da $-\infty$ a $+\infty$, nel caso invertibile) sono generatori dell'intera σ -algebra su cui è definita la misura μ .

Ad esempio, la partizione in cilindri elementari degli schemi di Bernoulli è generante. Vale ora la seguente proposizione, dovuta a Kolmogorov e Sinai:

Proposizione 9.11 (Teorema del generatore). *Se il sistema dinamico (M, μ, Φ) ammette una partizione generante finita α , allora*

$$h(\Phi) = h(\Phi, \alpha) .$$

La dimostrazione della proposizione è riportata nell'appendice D. Tenendo conto di quanto visto sopra, si conclude che

Corollario 9.12. *Lo schema di Bernoulli $B_{p_0, \dots, p_{n-1}}$ ha entropia $h = -\sum_i p_i \text{Log } p_i$.*

Dimostrazione. La partizione α in cilindri elementari è generante, e come si è visto sopra, l'entropia $h(\Phi, \alpha)$ vale proprio $-\sum_i p_i \text{Log } p_i$. ■

- **Osservazione.** Così ad esempio i due schemi di Bernoulli $B_{\frac{1}{3}, \frac{2}{3}}$ e $B_{\frac{1}{2}, \frac{1}{2}}$ certamente non sono isomorfi. Avere la medesima entropia è tuttavia condizione necessaria per l'isomorfismo, ma certo non sufficiente (ad esempio, i sistemi di entropia nulla non sono necessariamente tra loro isomorfi). Fanno eccezione gli schemi di Bernoulli, per i quali si è dimostrato (Ornstein, 1970) che la condizione è anche sufficiente (così, gli schemi $B_{\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}}$ e $B_{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}}$ sono isomorfi).

Si può anche dimostrare che, per i sistemi classici, ogni partizione separante è generante: più precisamente, che

Proposizione 9.13. *Condizione necessaria e sufficiente perché in un sistema dinamico classico (M, μ, Φ) la partizione classica α sia generante, è che essa sia separante su un sottoinsieme di M di misura uno.*

In particolare dunque per i sistemi classici espansivi è facile trovare partizioni generanti (ma non è altrettanto facile calcolare l'entropia).

Chiudiamo il paragrafo citando un ultimo risultato, conosciuto come *Teorema di Shannon–McMillan–Brian*.

Proposizione 9.14. *Sia (M, μ, Φ) un sistema dinamico discreto, e α una partizione misurabile di M . Si denoti con $B(t, x)$ l'atomo di $\beta_t = \alpha \vee \dots \vee \Phi^{-t}(\alpha)$ che contiene x . Allora:*

i. esiste quasi ovunque il limite

$$g(\Phi, \alpha, x) = \lim_{t \rightarrow \infty} -\frac{1}{t} \text{Log} \mu(B(t, x)) ; \quad (9.11)$$

ii. risulta

$$h(\Phi, \alpha) = \int_M g(\Phi, \alpha, x) d\mu . \quad (9.12)$$

Si può dimostrare che $g(\Phi, \alpha, x)$ è un integrale del moto del sistema; se ne deduce che

Corollario 9.15. *Se il sistema è ergodico, $g(\Phi, \alpha, x)$ è costante quasi ovunque, e coincide con $h(\Phi, \alpha)$.*

- **Osservazione.** Si confrontino le (9.11), (9.12) con le (9.3), (9.4). Quello che dice il teorema di Shannon–McMillan–Brian è, in sostanza, che il limite dell'integrando esiste quasi ovunque, e inoltre l'operazione di limite commuta con l'integrazione. La differenza non è da poco, e in particolare l'idea di h come misura del ritmo di frammentazione di β_t è assai più chiara, soprattutto nel caso ergodico.

10. Introduzione agli esponenti caratteristici di Lyapunov

Lo studio degli esponenti caratteristici di Lyapunov (ECL) come “indicatori di stabilità” di un punto singolare o di un'orbita periodica risale ai lavori di Lyapunov di inizio secolo; l'introduzione di tali quantità nella teoria ergodica è invece più recente: il lavoro fondamentale di Oseledec, che stabilisce l'esistenza degli ECL per tutte le orbite a meno di un insieme di dati iniziali di misura nulla, è del 1968, mentre il teorema di Pesin, che connette gli ECL all'entropia di Kolmogorov–Sinai, è del 1975.

Gli ECL, si può dire, sono un modo naturale di introdurre formalmente la nozione di divergenza esponenziale delle traiettorie che circondano una traiettoria assegnata; come abbiamo visto in diverse occasioni (Gatto di Arnol'd; teorema di Kouchnirenko...), tale nozione gioca un ruolo assai importante nell'ambito della teoria ergodica. In effetti, ben al di là della teoria ergodica, l'interesse per gli ECL è andato crescendo negli anni, sia in connessione ai più diversi problemi fisici (meccanica statistica, meccanica celeste, turbolenza, fisica del plasma, fisica degli acceleratori...), sia anche nell'ambito di discipline più applicative, come l'ecologia o l'economia o diversi rami della fisica medica (in particolare analisi di cardiogrammi e encefalogrammi). Tale interesse è dovuto in buona parte al fatto che da un lato si tratta di quantità dal significato matematico chiaro, che caratterizzano bene la

dinamica di un sistema, dall'altro che essi, come vedremo, sono anche calcolabili numericamente. Ancor oggi gli ECL sono forse le quantità cui più di frequente ci si richiama nello studio, e nella definizione stessa, dei cosiddetti sistemi dinamici "caotici".

10.1. Divergenza esponenziale delle traiettorie vicine

Sia (M, Φ) un sistema dinamico topologico,⁷¹⁾ continuo o discreto, con M varietà riemanniana. Con la metrica restano definite la norma $\|\xi\|$ di un qualsiasi vettore tangente $\xi \in T_x M$ ($T_x M$ denota lo spazio tangente, o piano tangente, a M in $x \in M$), e la distanza $\text{dist}(x, y)$ tra coppie qualsiasi di punti $x, y \in M$. Come vedremo, la metrica interviene nella definizione degli ECL, ma la definizione in realtà non ne dipende (norme equivalenti in ciascuno spazio tangente conducono agli stessi ECL).

Preso un punto x su M , sia $I = (-\varepsilon, \varepsilon) \subset \mathbb{R}$ un intervallo attorno all'origine, e $y : I \rightarrow M$ una curva regolare su M , tale che $y(0) = x$. Ogni dato iniziale $y(s)$, $s \in I$, viene inviato al tempo t in $\Phi^t(y(s))$; la curva y ha allora per immagine al tempo t la curva $y_t = \Phi^t \circ y$, ovvero $y_t(s) = \Phi^t(y(s))$, e si ha evidentemente $y_t(0) = x_t := \Phi^t(x)$. E' spontaneo allora introdurre il "coefficiente di dilatazione" γ relativo al punto x , al tempo t e alla curva prescelta y , definito da

$$\gamma = \lim_{s \rightarrow 0} \frac{\text{dist}(y_t(s), x_t)}{\text{dist}(y(s), x)}$$

(si veda la figura 10.1). Il coefficiente è chiaramente lo stesso per tutte le curve y tra loro tangenti in x ; se allora $\xi \in T_x M$ è il vettore tangente corrispondente alla curva y , e $D\Phi_x^t : T_x M \rightarrow T_{\Phi^t(x)} M$ denota l'applicazione tangente a Φ^t in $x \in M$, il coefficiente di dilatazione risulta essere funzione di t , x e ξ , ed è dato da

$$\gamma(t, x, \xi) = \frac{\|D\Phi_x^t \xi\|}{\|\xi\|}.$$

E' naturale dire che si ha divergenza esponenziale delle traiettorie vicine a $\Phi^t(x)$, con dato iniziale spostato nella direzione di ξ , se per t grande γ cresce esponenzialmente con t , $\gamma \sim e^{\chi t}$, con χ positivo. Un modo formale di definire l'esponente χ è il seguente:

Definizione. Si dice *esponente caratteristico di Lyapunov* del vettore tangente $\xi \in T_x M$ il limite,⁷²⁾ se esiste,

$$\chi(x, \xi) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\|D\Phi_x^t \xi\|}{\|\xi\|}. \quad (10.1)$$

⁷¹⁾ Faremo riferimento alla misura conservata solo quando necessario, precisamente in connessione al teorema di Oseledec e al teorema di Pesin.

⁷²⁾ Questo limite si può anche riscrivere nella forma di media temporale: infatti, posto

$$\gamma_1(x, \xi) := \|D\Phi_x^1 \xi\| / \|\xi\|,$$

(coefficiente di dilatazione per $t = 1$, nella direzione di ξ), si ha

$$\chi(x, \xi) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \log \gamma_1(\Phi^s(x), \xi_s), \quad \xi_s := D\Phi_x^s \xi$$

(media logaritmica della dilatazione a ogni unità di tempo). Si osservi però che γ_1 non si può scrivere come funzione della sola x , e dunque $\chi(x, \xi)$ non è propriamente la media temporale di una funzione; se così fosse, l'esistenza del limite quasi ovunque sarebbe garantita dal teorema di Birkhoff–Kinchin.

Il denominatore $\|\xi\|$ si potrebbe evidentemente omettere; dalla definizione stessa segue banalmente che metriche equivalenti danno gli stessi valori di $\chi(x, \xi)$.

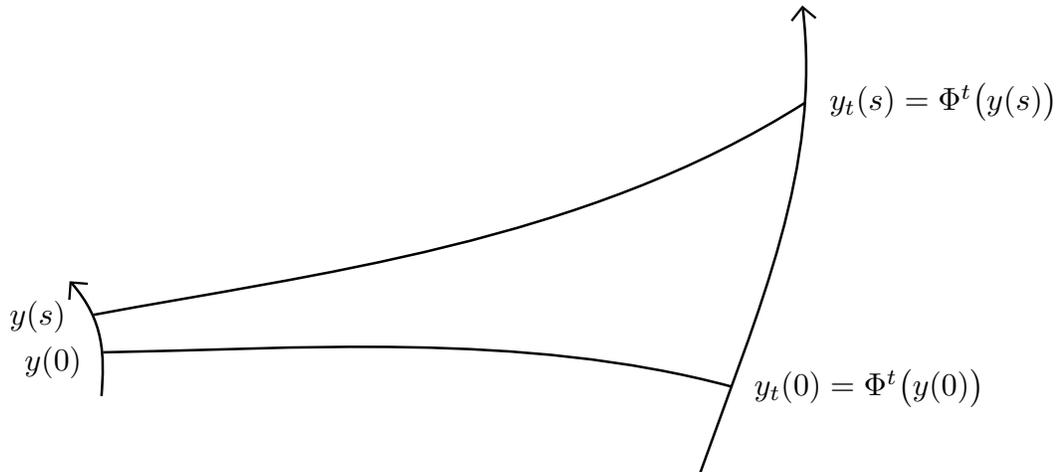


Figura 10.1. Divergenza esponenziale delle traiettorie vicine.

10.2. Un esempio elementare

Consideriamo il caso semplice di un moto periodico di periodo τ , $\Phi^\tau(x) = x$; allora l'applicazione tangente $D\Phi_x^\tau$ è un ordinario operatore lineare: $T_xM \rightarrow T_xM$, e ha senso discutere dei suoi autovalori e autovettori. Supponiamo per semplicità che $D\Phi_x^\tau$ abbia $n = \dim M$ autovalori reali distinti, e anzi di modulo diverso, $\lambda_1, \dots, \lambda_n$, che possiamo pensare ordinati in modo che sia

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| ;$$

siano e_1, \dots, e_n i corrispondenti autovettori. Allora:

a) risulta

$$\chi(x, e_i) = \tau^{-1} \log |\lambda_i| .$$

Infatti, si ha $D\Phi_x^{k\tau} e_i = \lambda_i^k e_i$, e la conclusione è banale se si prende il limite sulla sottosuccessione $t = k\tau$, $k \in \mathbb{N}$; ma è immediato vedere che la restrizione alla sottosuccessione non è rilevante.

Si potrebbe vedere che a tale teorema ci si riduce solo nel caso eccezionale in cui, con scelta opportuna delle coordinate, le matrici $A(x) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ rappresentative di $D\Phi_x^1$, $x \in M$, commutano tutte tra loro. Il teorema di Oseledec di cui si dirà tra poco, che garantisce l'esistenza quasi ovunque del limite per qualunque misura conservata, generalizza in un certo senso il teorema di Birkhoff–Kinchin (e per questo è anche chiamato *teorema ergodico non commutativo*).

- b) Preso un vettore $\xi = \sum_{i \geq r} c_i e_i$, con il coefficiente c_r non nullo, la dilatazione di e_r domina, e si ha

$$\chi(x, \xi) = \chi(x, e_r) .$$

Infatti, si ha

$$\|D\Phi_x^{k\tau} \xi\| = \lambda_r^k \left\| c_r e_r + c_{r+1} \left(\frac{\lambda_{r+1}}{\lambda_r} \right)^k e_{r+1} + \dots + c_n \left(\frac{\lambda_n}{\lambda_r} \right)^k e_n \right\| ,$$

e per k abbastanza grande il membro di destra è certamente compreso, ad esempio, tra $\frac{1}{2} \lambda_r^k \|c_r e_r\|$ e $2 \lambda_r^k \|c_r e_r\|$, cosicché la conclusione è immediata. Se ne conclude che al variare di ξ in $T_x M$ i possibili valori di $\chi(x, \xi)$ sono in tutto $n = \dim M$.

- c) Denotiamo con $[a, b, c, \dots]$ il sottospazio lineare di $T_x M$ generato dai vettori $a, b, c, \dots \in T_x M$, e poniamo

$$\begin{aligned} L_1 &= [e_1, \dots, e_n] = T_x M \\ L_2 &= [e_2, \dots, e_n] \\ &\vdots \\ L_n &= [e_n] ; \end{aligned}$$

risulta evidentemente

$$T_x M = L_1 \supset L_2 \supset \dots \supset L_n , \quad (10.2)$$

e per quanto visto sopra

$$\xi \in L_i \setminus L_{i+1} \implies \chi(x, \xi) = \chi(x, e_i) , \quad i = 1, \dots, n \quad (10.3)$$

(si ponga, per coerenza di notazione, $L_{n+1} = \{0\}$).

Una successione di sottospazi lineari

$$E = L_1 \supset \dots \supset L_m , \quad m \leq n = \dim E ,$$

di dimensione decrescente, è detta *filtrazione* dello spazio vettoriale E ; la (10.2) è dunque una particolare filtrazione di $T_x M$.

Non sarebbe difficile vedere che le particolari ipotesi sopra introdotte si possono rimuovere: per una generica orbita periodica, se anche $D\Phi_x^\tau$ ha autovalori λ_i multipli (o coincidenti in modulo), eventualmente complessi, ciascuno con un numero di autovettori indipendenti anche inferiore⁷³⁾ alla sua molteplicità ν_i , si trova comunque una filtrazione di $T_x M$

$$T_x M = L_1 \supset \dots \supset L_m , \quad m \leq n , \quad \dim L_i \setminus L_{i+1} = \nu_i ,$$

tale che $\chi(x, \xi) = \tau^{-1} \log |\lambda_i|$ non appena $\xi \in L_i \setminus L_{i+1}$.

⁷³⁾ Anche se $D\Phi_x^\tau$ non è diagonalizzabile, gli si può comunque associare una matrice in forma di Jordan; resta allora comunque individuato, per ogni autovalore λ_i di molteplicità ν_i , un sottospazio di dimensione ν_i , ed è facile verificare, usando la forma di Jordan, che ogni vettore del sottospazio, e non solo l'autovettore, dà come esponente di Lyapunov $\tau^{-1} \log |\lambda_i|$.

10.3. Il teorema di Oseledec e il teorema di Pesin

Nel caso generico di un moto non periodico, la nozione di autovettore e autovalore perde senso: per ogni t infatti $D\Phi_x^t$ manda uno spazio tangente in un altro, e non essendoci alcun modo naturale di identificare tra loro gli spazi tangenti in punti diversi, non ha senso (non è intrinseca) la nozione di autovettore. Viceversa, dalla definizione stessa di $\chi(x, \xi)$, e anzi dalla definizione provvisoria più debole

$$\chi(x, \xi) := \limsup_{t \rightarrow \infty} \frac{1}{t} \log \|D\Phi_x^t \xi\| ,$$

che ci permette di rinviare il problema dell'esistenza del limite (si è anche omesso l'irrilevante denominatore $\|\xi\|$), segue senza difficoltà l'esistenza intrinseca della filtrazione in ogni punto $x \in M$. Precisamente:

Proposizione 10.1. *Per ogni $x \in M$,*

i. al variare di ξ in $T_x M$ la quantità $\chi(x, \xi)$ assume un numero finito $m \leq n$ di valori distinti

$$\chi_1^*(x) > \chi_2^*(x) > \cdots > \chi_m^*(x) ;$$

ii. esiste una filtrazione di $T_x M$ in m sottospazi,

$$T_x M \equiv L_1 \supset L_2 \supset \cdots \supset L_m ,$$

tale che

$$\xi \in L_i \setminus L_{i+1} \implies \chi(x, \xi) = \chi_i^*(x) , \quad i = 1, \dots, m ;$$

iii. sia (e_1, \dots, e_n) una base (base normale) di $T_x M$ ottenuta prendendo in $L_i \setminus L_{i+1}$ un numero di vettori indipendenti pari a $\nu_i = \dim L_i - \dim L_{i+1}$, e (f_1, \dots, f_n) una base qualsiasi di $T_x M$. Allora si ha

$$\sum_{i=1}^n \chi(x, e_i) \leq \sum_{i=1}^n \chi(x, f_i) ,$$

e l'uguaglianza vale se e solo se anche (f_1, \dots, f_n) è normale.

Definizione. Il numero $\nu_i = \dim L_i - \dim L_{i+1}$ è detto *molteplicità* di χ_i^* . L'insieme

$$\text{Sp}(x) := \{\chi_1(x), \dots, \chi_n(x)\} ,$$

che si ottiene prendendo in ordine ciascuno dei valori distinti χ_i^* , $i = 1, \dots, m$, e ripetendolo un numero di volte pari alla sua molteplicità ν_i , o equivalentemente ponendo $\chi_i(x) = \chi(x, e_i)$ con (e_1, \dots, e_n) base normale, è detto⁷⁴⁾ *spettro* degli ECL nel punto $x \in M$.

⁷⁴⁾ E' chiamato a volte spettro anche l'insieme $\text{Sp}^*(x) = \{\chi_1^*(x), \nu_1, \dots, \chi_m^*(x), \nu_m\}$.

Dimostrazione. Preso $x \in M$, mostriamo innanzitutto che per ogni $c \neq 0$, e per ogni scelta di $\xi, \xi' \in T_x M$, si ha⁷⁵⁾

$$\chi(x, c\xi) = \chi(x, \xi), \quad \chi(x, \xi + \xi') \leq \max(\chi(x, \xi), \chi(x, \xi')). \quad (10.4)$$

La prima relazione è banale; per la seconda basta usare la definizione di \limsup , secondo la quale per ogni $\varepsilon > 0$ esiste $T_\varepsilon > 0$ tale che

$$t^{-1} \log \|D\Phi_x^t \xi\| < \chi(x, \xi) + \varepsilon$$

per ogni $t > T_\varepsilon$, e analogamente per ξ' (si può prendere il medesimo T_ε): da questa disuguaglianza segue infatti, per $t > T_\varepsilon$,

$$\begin{aligned} t^{-1} \log \|D\Phi_x^t(\xi + \xi')\| &\leq t^{-1} \log(\|D\Phi_x^t \xi\| + \|D\Phi_x^t \xi'\|) \\ &\leq t^{-1} \log 2 \max(\|D\Phi_x^t \xi\|, \|D\Phi_x^t \xi'\|) \\ &= t^{-1} [\max(\log \|D\Phi_x^t \xi\|, \log \|D\Phi_x^t \xi'\|) + \log 2] \\ &\leq \max(\chi(x, \xi), \chi(x, \xi')) + \varepsilon + t^{-1} \log 2, \end{aligned}$$

e la seconda delle (10.4) è immediata. Le (10.4) mostrano che per ogni $\theta \in \mathbb{R}$ l'insieme

$$L(\theta) := \{\xi \in T_x M : \chi(x, \xi) \leq \theta\}$$

è un sottospazio lineare di $T_x M$ (si ponga formalmente $\chi(x, 0) = -\infty$, per includere in $L(\theta)$ il vettore nullo). Dalla definizione stessa di $L(\theta)$ segue, evidentemente, $L(\theta') \subset L(\theta)$ se $\theta' < \theta$; anzi, se esiste $\xi \in L(\theta)$ tale che $\chi(x, \xi) = \theta$, allora $\xi \notin L(\theta')$, ovvero l'inclusione è propria e $\dim L(\theta) > \dim L(\theta')$ strettamente. Se ne conclude che al variare di $\xi \neq 0$ in $T_x M$ la quantità $\chi(x, \xi)$ assume $m \leq n$ valori distinti, che denotiamo $\chi_1^* > \dots > \chi_m^*$. I sottospazi $L_k := L(\chi_k^*)$, $1 \leq k \leq m$, costituiscono la filtrazione cercata. L'ultimo punto segue dal fatto, del tutto evidente, che la base normale ha per costruzione il maggior numero possibile di vettori negli spazi con indice alto, dunque con χ basso. ■

La proprietà banale usata nel corso della dimostrazione, secondo cui $\chi(x, c\xi) = \chi(x, \xi)$, mostra che χ è in realtà funzione, più che del vettore ξ , del sottospazio lineare $E \subset T_x M$, $\dim E = 1$, cui ξ appartiene. E' spontanea allora la generalizzazione a sottospazi lineari di dimensione qualsiasi:

Definizione. Si dice *esponente caratteristico di Lyapunov di ordine p* del sottospazio lineare $E \subset T_x M$, $\dim E = p \leq n$, il limite, se esiste,

$$\chi^{(p)}(x, E) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\text{Vol}^p(D\Phi_x^t(\xi_1), \dots, D\Phi_x^t(\xi_p))}{\text{Vol}^p(\xi_1, \dots, \xi_p)}, \quad (10.5)$$

ove (ξ_1, \dots, ξ_p) è una qualsiasi base di E , mentre $\text{Vol}^p(\dots)$ denota il volume p -dimensionale del parallelepipedo generato dai vettori tangenti che compaiono ad argomento.

⁷⁵⁾ Nella seconda relazione si ha la disuguaglianza anziché l'uguaglianza perché (si pensi all'esempio del paragrafo precedente) ξ e ξ' potrebbero avere componente dominante (componente lungo e_r) esattamente opposta.

Gli ECL precedentemente introdotti coincidono con gli ECL di ordine uno qui definiti. La scelta della base (ξ_1, \dots, ξ_p) è con evidenza irrilevante, e neanche la scelta particolare della metrica gioca alcun ruolo.

L'esistenza degli ECL di ogni ordine come limite esatto è garantita da un fondamentale teorema, dovuto a Oseledec (1968), nel quale gioca un ruolo fondamentale la misura.

Proposizione 10.2. *Sia (M, μ, Φ) un sistema dinamico (classico o astratto; discreto o continuo; invertibile o non invertibile), con M varietà riemanniana.⁷⁶⁾*

- i. *Per quasi ogni $x \in M$, e ogni sottospazio $E \subset T_x M$, $\dim E = p \leq n$, esiste finito il limite esatto (10.5); in particolare, per $p = 1$, esiste finito il limite esatto (10.1) in corrispondenza a ogni vettore tangente $\xi \in T_x M$.*
- ii. *Lo spettro $Sp(x)$ sopra definito è funzione misurabile di x .*
- iii. *Per ogni $E \subset T_x M$ esiste una base normale (e_1, \dots, e_p) , tale cioè che*

$$\sum_{i=1}^p \chi(x, e_i) \leq \sum_{i=1}^p \chi(x, f_i) ,$$

ove (f_1, \dots, f_p) è una qualunque altra base di E , e in corrispondenza a tale base si ha

$$\chi^{(p)}(x, E) = \sum_{i=1}^p \chi(x, e_i) .$$

L'ultimo punto, non sorprendente e del tutto ovvio nell'esempio elementare dell'orbita periodica⁷⁷⁾ chiarisce la relazione tra gli ECL di ordine p generico e quelli di ordine uno.

Il teorema non è elementare, e ne omettiamo la dimostrazione. Con un po' di pazienza, ma senza vere difficoltà, si verificano invece alcune rilevanti proprietà degli ECL (la verifica è lasciata come esercizio).

- a) Gli ECL $\chi_i(x)$ sono (con evidenza) costanti del moto: $Sp(\Phi^t(x)) = Sp(x)$. Pertanto, *in un sistema ergodico gli esponenti di Lyapunov sono quasi ovunque costanti.*
- b) Per un sistema invertibile,⁷⁸⁾ se la misura conservata μ è equivalente al volume n -dimensionale, allora l'esponente di Lyapunov di tutto $T_x M$, corrispondente alla dilatazione del volume n -dimensionale, si annulla, e con esso si annulla la somma di tutti gli ECL:

$$\chi^{(n)}(x, T_x M) = \sum_{i=1}^n \chi_i(x) = 0 .$$

⁷⁶⁾ Benché M sia una varietà differenziabile, il sistema può non essere classico, perchè la misura invariante μ può non essere regolare; in particolare, il supporto di μ può essere sostanzialmente più piccolo di M , ed essere un insieme assai complicato: è il caso dei sistemi con un attrattore, di cui si parlerà nel prossimo capitolo (paragrafi 4 e 6).

⁷⁷⁾ Ma non in generale: gli angoli tra i diversi vettori tangenti, nel corso del moto, in generale non sono limitati fuori dallo zero, e dunque non si può escludere a priori $\chi^{(p)}(x, E) < \sum_i \chi(x, e_i)$.

⁷⁸⁾ Perché si fa questa ipotesi?

- c) Per un sistema continuo retto dall'equazione differenziale $\dot{x} = f(x)$ su M , se $x \in M$ è tale che la traiettoria $\Phi^t(x)$ non ha limite a un punto fisso per $t \rightarrow \infty$, allora

$$\chi(x, f) = 0 .$$

- d) Per un sistema dinamico hamiltoniano (M, μ, Φ) a n gradi di libertà, con $M =$ superficie di energia costante ($\dim M = 2n - 1$), lo spettro di ogni punto è simmetrico:⁷⁹⁾

$$\text{Sp}(x) = \{\chi_1(x), \dots, \chi_{n-1}(x), \chi_n(x), -\chi_{n-1}(x), \dots, -\chi_1(x)\} ;$$

per il punto precedente, se $\Phi^t(x)$ non ha limite a un punto fisso, il valore centrale χ_n è nullo. Per il corrispondente sistema hamiltoniano con $M =$ guscio tra due superfici di energia costante ($\dim M = 2n$) lo spettro è ancora simmetrico:

$$\text{Sp}(x) = \{\chi_1(x), \dots, \chi_{n-1}(x), \chi_n(x), -\chi_n(x), -\chi_{n-1}(x), \dots, -\chi_1(x)\} ,$$

e differisce dal precedente solo per l'inserimento di $-\chi_n$ al centro; genericamente vi è dunque una coppia di zeri al centro.⁸⁰⁾ E' infine simmetrico lo spettro di un diffeomorfismo simplettico (una trasformazione canonica) $\Phi : M \rightarrow M$, $M =$ varietà simplettica (il consueto spazio delle fasi di un sistema hamiltoniano).

- e) Per un vettore generico $\xi \in T_x M$ (generico nel senso che resta escluso il solo sottospazio L_2 della filtrazione, $\dim L_2 < \dim L_1 = T_x M$) si trova il valore massimo $\chi(x, \xi) = \chi_1$; allo stesso modo, per un generico sottospazio $E \in T_x M$, $\dim E = p$, si trova il valore massimo $\chi(x, E) = \sum_{i=1}^p \chi_i(x)$. Queste proprietà sono importanti soprattutto perché sono alla base del metodo numerico di calcolo degli ECL.

- **Esercizio.** Si ritorni qui dopo lo studio dei sistemi hamiltoniani integrabili (capitolo III), e si dimostri che per tali sistemi tutti gli ECL sono nulli.
- **Osservazione.** Mettiamoci per semplicità nel caso discreto; poniamo $A(x, t) = D\Phi_x^t : T_x M \rightarrow T_{\Phi^t(x)} M$, e sia $A^*(x, t)$ l'operatore aggiunto, $A^*(x, t) : T_{\Phi^t(x)} M \rightarrow T_x M$, cosicché $A^*(x, t)A(x, t)$ è un operatore lineare su $T_x M$ (simmetrico e definito positivo). Si dimostra che, nelle stesse ipotesi del teorema di Oseledec, esiste quasi ovunque l'operatore limite

$$B(x) = \lim_{t \rightarrow \infty} [A^*(x, t)A(x, t)]^{\frac{1}{2t}} ,$$

che gli autovettori $e_1(x), \dots, e_n(x)$ di $B(x)$ costituiscono una base normale in x , e che se $\lambda_i(x)$ è l'autovalore corrispondente a $e_i(x)$ risulta $\chi_i(x) = \log \lambda_i(x)$. *Esercizio:* si costruisca B nel caso di un'orbita periodica, supponendo $n = 2$ e $D\Phi_x^\tau$ rappresentato dalla matrice $\begin{pmatrix} a & 1 \\ 0 & b \end{pmatrix}$.

⁷⁹⁾ Qui ci vuole effettivamente un po' di lavoro: vanno paragonati $D\Phi_x^t$, l'inverso e l'aggiunto. Si veda anche l'osservazione subito sotto.

⁸⁰⁾ Il secondo zero sta ad indicare che, con l'eccezione dei punti di equilibrio, non vi è dilatazione esponenziale in una opportuna direzione trasversa alla superficie di energia costante.

Come si è visto, il teorema di Oseledec vale per sistemi dinamici qualsiasi, non necessariamente classici. Nel caso classico vale l'ulteriore importantissimo teorema di Pesin, che connette gli ECL all'entropia:

Proposizione 10.3. *Per ogni sistema dinamico classico (M, μ, Φ) si ha*

$$h(\Phi) = C \int_M \sum_i^+ \chi_i(x) d\mu, \quad C = 1/\log 2,$$

ove \sum^+ denota la somma ristretta ai soli χ_i positivi.⁸¹⁾

La costante C è dovuta all'uso un po' sconsiderato del logaritmo in base due nella definizione dell'entropia, e scompare se in tale definizione il logaritmo in base due è sostituito dal logaritmo naturale. Dunque, in base al teorema di Pesin, l'entropia ammette una densità, e tale densità coincide con la somma degli ECL positivi. Per sistemi ergodici i χ_i sono costanti quasi ovunque su M , e si ha semplicemente

$$h(\Phi) = C \sum_i^+ \chi_i.$$

Per il Gatto di Arnol'd si trova immediatamente $\chi_{1,2} = \pm \log \lambda$, ove $\lambda = (3 + \sqrt{5})/2$ è il più grande dei due autovalori della matrice A , e dunque $h(\Phi) = \text{Log } \lambda$ (si confronti questo risultato esatto con la stima introdotta a proposito del teorema di Kouchnirenko).

Come si è sopra ricordato, uno dei motivi di interesse degli ECL è la possibilità di calcolarli numericamente. Una breve indicazione sul metodo di calcolo più comunemente usato è riportata nell'Appendice E.

⁸¹⁾ Accenniamo per completezza a una recente generalizzazione del teorema di Pesin al caso di sistemi non classici, precisamente con la misura μ a supporto su un attrattore X (cap. II, par. 4) di misura di Lebesgue zero, anche non liscio (cioè privo della struttura di varietà). La sommatoria che compare nell'enunciato del teorema si può evidentemente riscrivere

$$\sum_j^+ \nu_j \chi_j^*, \quad \nu_j = \dim E_j - \dim E_{j-1};$$

in opportune ipotesi il teorema si estende sostituendo a $\dim E_j$ la cosiddetta dimensione frattale (dimensione di Hausdorff) dell'intersezione $E_j \cap X$.

APPENDICI

A. Prova del Lemma 2.4

Dimostrazione. Procedendo in modo un po' indiretto,⁸²⁾ mostriamo che per ogni reticolo $\mathcal{R} \in \mathbb{Z}^n$, $\dim \mathcal{R} = r$, esiste una matrice intera L , con $\det L = 1$, le cui ultime $n - r$ righe siano vettori ortogonali a \mathcal{R} : precisamente tale che, posto

$$l^{(i)} = (L_{i1}, \dots, L_{in}) , \quad (\text{A.1})$$

risulti

$$l^{(i)} \cdot k = 0 , \quad r < i \leq n , \quad (\text{A.2})$$

per tutti e soli i vettori interi $k \in \mathcal{R}$. Questo basta perché la matrice J cercata allora è $J = L^{-T}$ (inversa della trasposta): infatti, l'inversa L^{-1} è sicuramente intera con $\det L^{-1} = 1$, e denotate con

$$k^{(j)} = (L_{1j}^{-1}, \dots, L_{nj}^{-1})$$

le sue colonne, si ha evidentemente $l^{(i)} \cdot k^{(j)} = 0$ per $i > r$ e $j \leq r$. Ciascuno dei vettori interi $k^{(j)}$, $j \leq r$, è allora ortogonale a $n - r$ vettori indipendenti ortogonali a \mathcal{R} , e dunque appartiene a \mathcal{R} ; equivalentemente, come chiesto nel lemma, le prime r righe di $J = L^{-T}$ appartengono a \mathcal{R} .

L'esistenza di L con le proprietà indicate si dimostra in tre passaggi, di generalità crescente.

- a) *Il caso $n = 2$, $r = 1$.* Preso un qualsiasi vettore $k = (k_1, k_2) \in \mathcal{R}$ (il reticolo è ora una retta in \mathbb{Z}^2), sia $m = (m_1, m_2)$ parallelo a k , con m_1 e m_2 primi tra loro. Basta mostrare che si trovano interi α e β , tali che la matrice

$$L = \begin{pmatrix} \alpha & -\beta \\ -m_2 & m_1 \end{pmatrix}$$

abbia determinante uno:

$$d := \alpha m_1 - \beta m_2 = 1 ;$$

basta anzi considerare il caso $m_2 \neq 0$, perché altrimenti si ha $m_1 = 1$, e la conclusione è banale. A questo scopo, per ogni α intero poniamo $\beta =$ quoziente della divisione intera $\alpha m_1 / m_2$; allora il numero d definito sopra è il resto della divisione, cosicché $0 \leq d < m_2$. Al variare di α tra 0 e $m_2 - 1$, i valori che si ottengono per d sono necessariamente tra loro diversi: se infatti per α e α' diversi si trova resto uguale, allora, per sottrazione, segue

$$(\alpha - \alpha')m_1 - (\beta - \beta')m_2 = 0 , \quad \frac{m_1}{m_2} = \frac{\beta - \beta'}{\alpha - \alpha'} ,$$

⁸²⁾ Questo per trattare il caso generico. La verifica diretta è invece la più semplice nel caso $n = r = 2$ (si veda l'osservazione alla fine della dimostrazione).

contro l'ipotesi che m_1, m_2 siano primi tra loro. In particolare dunque esistono α, β tali che $d = 1$.

- b) *Estensione a n arbitrario, $r = 1$.* Dobbiamo dimostrare che fissato $k \in \mathcal{R}$ esiste una matrice L , $\det L = 1$, tale che tutte le sue righe tranne la prima siano ortogonali a k , ovvero tale che il vettore Lk abbia tutte le componenti nulle tranne la prima. La matrice si costruisce come prodotto di $n - 1$ matrici diagonali a blocchi:

$$L = L_n \cdots L_3 L_2 ;$$

ciascuna matrice L_i ha un blocco di ordine due in corrispondenza agli indici 1 e i , e per il resto è l'identità. Per quanto sopra dimostrato, L_2 si può prendere in modo che il vettore $k^{(2)} = L_2 k$ abbia la seconda componente nulla; la prima componente cambia, le altre evidentemente restano invariate. Analogamente, L_3 si può prendere in modo che $k^{(3)} = L_3 k^{(2)}$ abbia la terza componente nulla, e anzi, poichè la seconda resta invariata, abbia sia la seconda che la terza componente nulla. Procedendo, si ottiene che il vettore Lk ha tutte le componenti nulle, tranne la prima, come richiesto. D'altra parte, L è prodotto di matrici intere con determinante uno, dunque è proprio la matrice cercata.

- c) *Estensione a n, r qualsiasi.* Siano $k^{(1)}, \dots, k^{(r)}$ vettori indipendenti appartenenti a \mathcal{R} ; mostriamo che esiste una matrice intera L , $\det L = 1$, tale che il vettore intero $Lk^{(j)}$, $1 \leq j \leq r$, abbia tutte le componenti nulle da $j + 1$ in poi:

$$(Lk^{(j)})_i = 0 \quad \text{per } 1 \leq j \leq r, j < i \leq n . \quad (\text{A.3})$$

Ciò basta alla conclusione del lemma, perché allora la (A.2) resta soddisfatta per r vettori indipendenti di \mathcal{R} , dunque per tutto \mathcal{R} , né possono esservi altri vettori fuori di \mathcal{R} per cui tale relazione è soddisfatta, altrimenti le ultime $n - r$ righe di L non sarebbero linearmente indipendenti.

La (A.3) si dimostra per induzione su j . Per $j = 1$ la matrice certamente esiste, ed è proprio quella costruita al punto b). Supponiamo allora di aver trovato L_s tale che la (A.3) sia soddisfatta per $j \leq s < r$, e mostriamo che esiste L_{s+1} tale che la medesima relazione sia soddisfatta anche per $j = s + 1$. Per questo poniamo

$$\tilde{k}^{(j)} = L_s k^{(j)} , \quad \tilde{k}^{(j)} = (\tilde{k}_1^{(j)}, \dots, \tilde{k}_j^{(j)}, 0, \dots, 0) \quad \text{per } j \leq s ,$$

e cerchiamo $L_{s+1} = \Lambda L_s$, con Λ che lasci inalterati i $\tilde{k}^{(j)}$ con $j \leq s$, e lavorando su $\tilde{k}^{(s+1)}$ lo muti in un vettore con tutte le componenti nulle dalla $s + 1$ in poi. Si vede senza difficoltà che basta prendere Λ diagonale a blocchi,

$$\Lambda = \begin{pmatrix} \text{Id} & 0 \\ 0 & \hat{L} \end{pmatrix} , \quad \text{Id} = \text{identità } s \times s ,$$

con \hat{L} matrice $(n - s) \times (n - s)$ tale che il vettore intero $(\tilde{k}_{s+1}^{(s+1)}, \dots, \tilde{k}_n^{(s+1)}) \in \mathbb{Z}^{n-s}$ sia mutato in un vettore intero con tutte le componenti nulle tranne la prima (di indice $s + 1$); \tilde{L} esiste proprio in virtù del punto b) sopra dimostrato. La dimostrazione del lemma è così conclusa. ■

- **Osservazione.** La dimostrazione, come si è visto, è lunga e un po' fastidiosa, ma (tranne forse per il primo punto, che richiede un po' di ispirazione) sostanzialmente banale. Il primo caso interessante del lemma, precisamente $n = 2$ e $r = 1$, si dimostra in modo molto semplice: senza bisogno di passare attraverso L , si costruisce direttamente J della forma

$$J = \begin{pmatrix} k_1 & k_2 \\ \alpha & \beta \end{pmatrix},$$

e α, β si trovano come nel punto a).

B. Prova del teorema di Birkhoff–Kinchin

La dimostrazione si basa sul seguente lemma tecnico, chiamato talvolta *teorema ergodico massimale*:

Lemma B.1. *Preso $f : M \rightarrow \mathbb{R}$ sommabile, sia*

$$F_t(x) := f(x) + f(\Phi(x)) + \dots + f(\Phi^{t-1}(x)), \quad x \in M, t > 0,$$

e poniamo

$$A := \{x \in M : \sup_{t \geq 1} F_t(x) \geq 0\}.$$

Allora A è misurabile, e

$$\int_A f \, d\mu \geq 0.$$

Si osservi che $t^{-1}F_t$ è la media temporale di f , arrestata al tempo t ; in sostanza dunque il lemma afferma che, se si prende *tutto* l'insieme A dove la media temporale di f arrestata al tempo t è non negativa almeno per un valore di t , allora l'integrale di f esteso a A è a sua volta non negativo.

Mostriamo qui di seguito come da questo lemma segua il teorema di Birkhoff–Kinchin; la dimostrazione del lemma è riportata subito sotto.

Dimostrazione (teorema di Birkhoff–Kinchin).

(a) *Esistenza del limite quasi ovunque.* Presi $a, b \in \mathbb{R}$, $a < b$, poniamo

$$E_{a,b} = \left\{ x \in M : \liminf_{t \rightarrow \infty} \frac{1}{t} F_t(x) < a < b < \limsup_{t \rightarrow \infty} \frac{1}{t} F_t(x) \right\}; \quad (\text{B.1})$$

si osservi che $E_{a,b}$ è misurabile (lim inf e lim sup di successioni di funzioni misurabili sono misurabili) e con evidenza anche invariante. L'insieme per il quale la media temporale di f *non* esiste è allora

$$E = \bigcup_{a,b} E_{a,b} ,$$

e siccome basta un'unione numerabile (ad esempio su a, b razionali), è sufficiente mostrare che ciascuno degli $E_{a,b}$ ha misura nulla.

A questo scopo facciamo vedere che in virtù del lemma vale la disuguaglianza “rovescia”

$$b \mu(E_{a,b}) \leq \int_{E_{a,b}} f \, d\mu \leq a \mu(E_{a,b}) , \quad (\text{B.2})$$

dalla quale segue necessariamente $\mu(E_{a,b}) = 0$. Per provare, ad esempio, la disuguaglianza di sinistra, si applica il lemma a

$$g(x) = \begin{cases} f(x) - b & \text{per } x \in E_{a,b} \\ -1 & \text{per } x \notin E_{a,b} \end{cases} .$$

Si vede facilmente che l'insieme A cui fa riferimento il lemma, cioè l'insieme in cui $G_t := \sum_{s=0}^t g \circ \Phi^s$ è non negativa per almeno un valore $t > 0$, è proprio $E_{a,b}$: infatti se $x \in E_{a,b}$ allora, in base alla definizione (B.1) di $E_{a,b}$, certamente esiste t per cui $t^{-1} F_t(x) > b$, e dunque $G_t(x) > 0$. D'altra parte, se invece $x \notin E_{a,b}$, allora $\Phi^t(x) \notin E_{a,b}$ per ogni $t > 0$ ($E_{a,b}$ è invariante), e $G_t(x)$ è sempre negativa. Per il lemma risulta allora

$$\int_{E_{a,b}} g \, d\mu = \int_{E_{a,b}} f \, d\mu - b \mu(E_{a,b}) \geq 0 ,$$

e la disuguaglianza sinistra della (B.2) è verificata. In modo analogo, prendendo

$$g(x) = \begin{cases} a - f(x) & \text{per } x \in E_{a,b} \\ -1 & \text{per } x \notin E_{a,b} \end{cases} ,$$

si prova la disuguaglianza destra.

(b) *Prova che risulta $\bar{f}(\Phi^t(x)) = \bar{f}(x)$.* Questo punto segue banalmente dalla definizione di $\bar{f}(x)$, ovunque il limite esista.

(c) *Prova che risulta $\langle \bar{f} \rangle = \langle f \rangle$.* Osserviamo innanzitutto che \bar{f} è sommabile: infatti,

$$\int_M \left| \frac{1}{t} \sum_{s=0}^{t-1} f(\Phi^s(x)) \right| \, d\mu(x) \leq \frac{1}{t} \sum_{s=0}^{t-1} \int_M |f(\Phi^s(x))| \, d\mu(x) = \int_M |f| \, d\mu ,$$

e questo basta. Poniamo ora

$$C_{a,b} = (\bar{f})^{-1}[a, b) = \{x \in M : a \leq \bar{f}(x) < b\} ;$$

dalla definizione segue banalmente

$$a \mu(C_{a,b}) \leq \int_{C_{a,b}} \bar{f} \, d\mu \leq b \mu(C_{a,b}) ,$$

ma anche, usando come sopra il lemma (con una opportuna g ecc.; la disuguaglianza debole a sinistra non è rilevante: basta sostituire a con $a - \varepsilon$),

$$a \mu(C_{a,b}) \leq \int_{C_{a,b}} f \, d\mu \leq b \mu(C_{a,b}) ,$$

e dunque

$$\left| \int_{C_{a,b}} \bar{f} \, d\mu - \int_{C_{a,b}} f \, d\mu \right| \leq (b - a) \mu(C_{a,b}) .$$

Da questa disuguaglianza, sfruttando l'arbitrarietà di a e b , segue senza difficoltà la conclusione, ovvero

$$\int_M \bar{f} \, d\mu - \int_M f \, d\mu = 0 . \quad (\text{B.3})$$

Si procede così: preso $\varepsilon > 0$ arbitrario, si divide \mathbb{R} in intervalli $[k\varepsilon, (k+1)\varepsilon)$, $k \in \mathbb{Z}$; risulta evidentemente $\cup_k C_{k\varepsilon, (k+1)\varepsilon} = M$, e

$$\begin{aligned} \left| \int_M \bar{f} \, d\mu - \int_M f \, d\mu \right| &\leq \sum_{k \in \mathbb{Z}} \left| \int_{C_{k\varepsilon, (k+1)\varepsilon}} \bar{f} \, d\mu - \int_{C_{k\varepsilon, (k+1)\varepsilon}} f \, d\mu \right| \\ &\leq \varepsilon \sum_{k \in \mathbb{Z}} \mu(C_{k\varepsilon, (k+1)\varepsilon}) = \varepsilon ; \end{aligned}$$

la (B.3) pertanto è verificata.

(d) *Prova che nel caso invertibile \bar{f}_- esiste e coincide con \bar{f} quasi ovunque.* L'esistenza di \bar{f}_- quasi ovunque è ovvia (basta considerare il sistema dinamico che si ottiene sostituendo Φ^{-1} a Φ). Denotiamo allora $\bar{f}_+ = \bar{f}$, e supponiamo che risulti, ad esempio, $\bar{f}_+ > \bar{f}_-$ in un insieme di misura positiva; allora posto

$$A := \{x \in M : \bar{f}_+ - \bar{f}_- > 0\} ,$$

si ha certamente

$$\int_A (\bar{f}_+ - \bar{f}_-) \, d\mu > 0 .$$

Ma poiché \bar{f}_+ e \bar{f}_- , come si è visto al punto (b), sono costanti del moto, l'insieme A è invariante: perciò se poniamo $g(x) = \chi_A(x)f(x)$, con χ_A funzione caratteristica di A , si ha $\bar{g}_\pm(x) = \bar{f}_\pm(x)$ per $x \in A$, $\bar{g}_\pm(x) = 0$ altrimenti, e dunque

$$\langle \bar{g}_+ \rangle - \langle \bar{g}_- \rangle = \int_A (\bar{f}_+ - \bar{f}_-) \, d\mu > 0 .$$

Ma questa è una contraddizione perché, per quanto già dimostrato al punto (c), deve risultare $\langle \bar{g}^+ \rangle = \langle \bar{g}^- \rangle = \langle g \rangle$. ■

Vediamo infine la dimostrazione del lemma.

Dimostrazione (lemma). Poniamo

$$\mathcal{F}_t = \max_{1 \leq s \leq t} F_s, \quad A_t = \{x \in M : \mathcal{F}_t \geq 0\},$$

cosicché $\mathcal{F}_{t+1} \geq \mathcal{F}_t$, e

$$A_t \subset A_{t+1}, \quad A = \bigcup_{1 \leq t < \infty} A_t.$$

Allora

$$\int_A f \, d\mu = \lim_{t \rightarrow \infty} \int_{A_t} f \, d\mu,$$

dunque per dimostrare il lemma basta provare che per ogni $t > 0$ risulta

$$\int_{A_t} f \, d\mu \geq 0.$$

A questo scopo, poniamo $\mathcal{F}_t^+ = \max(0, \mathcal{F}_t)$. Un minimo di riflessione mostra che per $t \geq 2$ si può scrivere

$$f(x) = \mathcal{F}_t(x) - \mathcal{F}_{t-1}^+(\Phi(x)) \geq \mathcal{F}_t(x) - \mathcal{F}_t^+(\Phi(x)),$$

da cui segue

$$\begin{aligned} \int_{A_t} f \, d\mu &\geq \int_{A_t} \mathcal{F}_t \, d\mu - \int_{A_t} \mathcal{F}_t^+ \circ \Phi \, d\mu \\ &\geq \int_M \mathcal{F}_t^+ \, d\mu - \int_M \mathcal{F}_t^+ \circ \Phi \, d\mu = 0 \end{aligned}$$

(la seconda disuguaglianza si spiega osservando che \mathcal{F}_t^+ coincide con \mathcal{F}_t in A_t , ed è nullo in $M \setminus A_t$). ■

C. Prova dei Lemmi 9.7 e 9.8

Dimostriamo preliminarmente alcune proprietà elementari dell'entropia relativa, precisamente:

Lemma C.1. *L'entropia relativa soddisfa le seguenti proprietà:*

- i. $\beta \succeq \alpha \iff \eta(\alpha | \beta) = 0.$
- ii. $\eta(\alpha | \beta) \leq \eta(\alpha),$ e l'uguaglianza vale se e solo se le partizioni sono indipendenti.
- iii. $\beta \succeq \alpha \implies \eta(\beta | \gamma) \geq \eta(\alpha | \gamma)$ e viceversa $\eta(\gamma | \beta) \leq \eta(\gamma | \alpha).$
- iv. $\eta(\alpha \vee \beta | \gamma) \leq \eta(\alpha | \gamma) + \eta(\beta | \gamma).$

Dimostrazione.

- i. Risulta $\eta(\alpha | \beta) = 0$ se e solo se se ciascuno dei termini $\mu(A_i | B_j)$ vale 0 oppure 1, e dunque se e solo se $\mu(A_i \cap B_j) = \mu(B_j)$ oppure 0, che vuol dire $\beta \succeq \alpha.$
- ii. Non è altro che la riscrittura (9.8) del Lemma 9.3.
- iii. Per la prima parte è utile la riscrittura

$$\eta(\alpha | \beta) = \sum_j \mu(B_j) \eta(\alpha | B_j), \quad \eta(\alpha | B_j) := -\sum_i \mu(A_i | B_j) \text{Log} \mu(A_i | B_j);$$

$\eta(\alpha | B_j)$ altro non è che l'entropia della partizione α , calcolata usando per i suoi atomi la misura relativa $\mu_j(A_i) := \mu(A_i | B_j).$ La conclusione allora è immediata, non appena si ricordi che raffinando una partizione l'entropia non diminuisce. Per la seconda parte, sia $\gamma = \{C_k\}_{0 \leq k < r-1},$ cosicchè

$$\eta(\gamma | \beta) = \sum_j \mu(B_j) \sum_k \mu(C_k | B_j) \text{Log} \mu(C_k | B_j) .$$

Per ogni $i,$ sia J_i l'insieme degli indici $j,$ tali che $B_j \cap A_i \neq \emptyset,$ cosicchè $\sum_j (\cdot) = \sum_i \sum_{j \in J_i} (\cdot).$ Sommando su $j \in J_i,$ ed usando ancora una volta la disuguaglianza di Jensen con pesi $c_j = \frac{\mu(B_j)}{\mu(A_i)},$ la conclusione segue facilmente.

- iv. Si ha

$$\begin{aligned} \eta(\alpha \vee \beta | \gamma) &= \eta(\alpha \vee \beta \vee \gamma) - \eta(\gamma) \\ &= \eta(\alpha \vee \beta \vee \gamma) - \eta(\beta \vee \gamma) + \eta(\beta \vee \gamma) - \eta(\gamma) \\ &= \eta(\alpha | \beta \vee \gamma) + \eta(\beta | \gamma) \leq \eta(\alpha | \gamma) + \eta(\beta | \gamma) . \end{aligned}$$

■

Veniamo ora alla dimostrazione del Lemma 9.7.

Dimostrazione. Primo punto: risulta sempre, evidentemente, $\text{dist}(\alpha, \beta) \geq 0$ e $\text{dist}(\alpha, \alpha) = 0;$ usando poi il Lemma C.1 si vede subito che

$$\text{dist}(\alpha, \beta) = 0 \implies \eta(\alpha | \beta) = \eta(\beta | \alpha) = 0 \implies \alpha = \beta .$$

La simmetria di $\text{dist}(\alpha, \beta)$ è evidente. Per provare la disuguaglianza triangolare, usiamo i Lemmi C.1 e 9.3 per scrivere

$$\begin{aligned} \eta(\alpha | \gamma) &= \eta(\alpha \vee \gamma) - \eta(\gamma) \\ &\leq \eta(\alpha \vee \beta \vee \gamma) - \eta(\beta \vee \gamma) + \eta(\beta \vee \gamma) - \eta(\gamma) = \eta(\alpha | \beta \vee \gamma) + \eta(\beta | \gamma) \\ &\leq \eta(\alpha | \beta) + \eta(\beta | \gamma) . \end{aligned}$$

Allo stesso modo si ha $\eta(\gamma | \alpha) \leq \eta(\gamma | \beta) + \eta(\beta | \alpha)$, e la conclusione è immediata.

Per il secondo punto basta mostrare che per ogni t si ha

$$|\eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha)) - \eta(\beta \vee \dots \vee \Phi^{-t+1}(\beta))| \leq t \text{dist}(\alpha, \beta) .$$

Per questo supponiamo ad esempio che per un certo t risulti $\eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha)) > \eta(\beta \vee \dots \vee \Phi^{-t+1}(\beta))$; segue facilmente

$$\begin{aligned} &\eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha)) - \eta(\beta \vee \dots \vee \Phi^{-t+1}(\beta)) \\ &\leq \eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha) \vee \beta \vee \dots \vee \Phi^{-t+1}(\beta)) - \eta(\beta \vee \dots \vee \Phi^{-t+1}(\beta)) \\ &= \eta(\alpha \vee \dots \vee \Phi^{-t+1}(\alpha) | \beta \vee \dots \vee \Phi^{-t+1}(\beta)) \\ &\leq \sum_s \eta(\Phi^{-s}(\alpha) | \beta \vee \dots \vee \Phi^{-t+1}(\beta)) \\ &\leq \sum_s \eta(\Phi^{-s}(\alpha) | \Phi^{-s}(\beta)) = \sum_s \eta(\alpha | \beta) = t \eta(\alpha | \beta) , \end{aligned}$$

e questo basta (si sono usati il quarto punto della Proposizione 9.1, la (9.7), e il quarto e terzo punto del Lemma C.1). ■

Infine, dimostriamo il Lemma 9.8.

Dimostrazione. Qualunque insieme misurabile A è approssimabile esternamente con un'unione di rettangoli, dunque con un insieme classico $\tilde{A} \supset A$, tale che la differenza $\tilde{A} \setminus A$ sia di misura piccola a piacere. Data allora $\alpha = \{A_0, \dots, A_{n-1}\}$, e preso ad arbitrio $\varepsilon > 0$, siano $\tilde{A}_0, \dots, \tilde{A}_{n-1}$ insiemi classici tali che $\tilde{A}_i \supset A_i$, $\mu(\tilde{A}_i \setminus A_i) < \varepsilon$; ponendo ricorsivamente

$$B_0 = \tilde{A}_0 , \quad B_i = \tilde{A}_i \setminus \bigcup_{0 \leq j < i} B_j , \quad 1 \leq i \leq n-1 ,$$

si ottiene allora una partizione classica $\beta = \{B_0, \dots, B_{n-1}\}$, tale che ogni atomo B_i di B approssima bene il corrispondente atomo A_i di A : precisamente, $|\mu(A_i - B_i)| < (\text{cost}) \varepsilon$, e $\mu(A_i \cap B_j) < (\text{cost}) \varepsilon$ per $i \neq j$. Segue subito

$$|\mu(A_i | B_j) - \delta_{ij}| < (\text{cost}) \varepsilon , \quad |\mu(B_i | A_j) - \delta_{ij}| < (\text{cost}) \varepsilon ,$$

e la conclusione è immediata. ■

D. Prova del teorema del generatore

Definizione. Si dice che la successione crescente di partizioni misurabili

$$\gamma_0 \preceq \gamma_1 \preceq \cdots \preceq \gamma_k \preceq \cdots$$

tende alla *partizione in punti*⁸³⁾ $\widehat{\varepsilon}$, se comunque si prendano un numero $\varepsilon > 0$ e un insieme A misurabile esistono $k \geq 0$ e $C_1, \dots, C_m \in \gamma_k$, tali che

$$\mu(A - \bigcup_i C_i) < \varepsilon .$$

Ad esempio, per $M =$ quadrato di lato L si ponga $\gamma_k =$ partizione in quadretti di lato $2^{-k}L$; la successione delle partizioni tende con evidenza alla partizione in punti (qualunque rettangolo, e dunque qualunque insieme misurabile secondo Lebesgue, si può approssimare a piacere con un numero finito di quadretti sufficientemente piccoli). Dalla definizione stessa di partizione generante segue che se α è generante per il sistema dinamico (M, μ, Φ) , supposto ad esempio invertibile, allora la successione di partizioni

$$\gamma_k = \Phi^k(\alpha) \vee \cdots \vee \alpha \vee \cdots \vee \Phi^{-k}(\alpha) , \quad k \in \mathbf{N} ,$$

tende alla partizione in punti.

Si verifica senza difficoltà il seguente lemma:

Lemma D.1. *Sia \mathcal{P} l'insieme delle partizioni misurabili di M . Se la sequenza di partizioni*

$$\gamma_0 \preceq \gamma_1 \preceq \cdots \preceq \gamma_k \preceq \cdots$$

converge a $\widehat{\varepsilon}$, allora l'insieme

$$\mathcal{P}_0 = \{ \beta \in \mathcal{P} ; \exists k \geq 0 : \beta \preceq \gamma_k \}$$

è denso in \mathcal{P} .

Nell'esempio: le partizioni $\beta \preceq \gamma_k$ sono quelle i cui atomi hanno tutti per bordo una spezzata che segue la quadrettatura di γ_k .

Dimostrazione. Mostriamo che comunque si prenda $\alpha = \{A_0, \dots, A_{n-1}\} \in \mathcal{P}$, si può costruire $\beta \in \mathcal{P}_0$ con $\text{dist}(\alpha, \beta)$ arbitrariamente piccola. Poichè $\gamma_k \rightarrow \widehat{\varepsilon}$, per ogni scelta di ε e dell'atomo $A_i \in \alpha$, esistono k e A'_i tali che

$$A'_i = \text{unione di atomi di } \gamma_k , \quad \mu(A_i - A'_i) \leq \varepsilon ;$$

⁸³⁾ L'espressione sottolinea il fatto che al crescere di k la partizione γ_k diventa "infinitamente fine".

si può evidentemente prendere il medesimo k per tutti gli A_i . Si ottiene allora una partizione $\beta = \{B_0, \dots, B_{n-1}\}$ di M ponendo

$$B_0 = A'_0, \quad B_i = A'_i \setminus \bigcup_{0 \leq j < i} B_j \quad \text{per } 0 < i < n-1, \quad B_{n-1} = M \setminus \bigcup_{0 \leq j < n-1} B_j;$$

non è difficile convincersi che $\beta \in \mathcal{P}_0$, e inoltre che $\text{dist}(\alpha, \beta) < (\text{cost})\varepsilon$. ■

Veniamo ora alla dimostrazione del teorema del generatore (per fissare le idee consideriamo il caso invertibile).

Dimostrazione. Basta mostrare che se α è generante, e β qualsiasi, allora si ha

$$h(\Phi, \alpha) \geq h(\Phi, \beta). \quad (\text{D.1})$$

Per questo, presa α generante, poniamo

$$\gamma_k = \Phi^k(\alpha) \vee \dots \vee \alpha \vee \dots \vee \Phi^{-k}(\alpha).$$

La successione che si ottiene è crescente e converge a $\widehat{\varepsilon}$, perciò si può applicare il lemma. La verifica della disuguaglianza (D.1), per β in \mathcal{P}_0 , è immediata: infatti, se $\beta \preceq \gamma_k$, allora

$$\begin{aligned} h(\Phi, \beta) &\leq h(\Phi, \gamma_k) = \lim_{t \rightarrow \infty} \frac{1}{t} \eta(\gamma_k \vee \dots \vee \Phi^{-t+1}(\gamma_k)) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \eta(\Phi^k(\alpha) \vee \dots \vee \Phi^{-k-t+1}(\alpha)) \\ &= \lim_{t \rightarrow \infty} \frac{2k+t}{t} \frac{1}{2k+t} \eta(\alpha \vee \dots \vee \Phi^{-2k-t+1}(\alpha)) \\ &= h(\Phi, \alpha). \end{aligned}$$

D'altra parte, poichè \mathcal{P}_0 è denso e h continua nel secondo argomento, la disuguaglianza si trasporta a tutto \mathcal{P} . ■

E. Un algoritmo per il calcolo numerico degli ECL

Ci mettiamo qui nel caso (più laborioso ma più interessante) di un sistema dinamico continuo (M, μ, Φ) , e supponiamo per semplicità (è il caso dei sistemi hamiltoniani) di poter ricoprire M con un solo sistema di coordinate x_1, \dots, x_n ; in pratica, supponiamo di avere una equazione differenziale in \mathbb{R}^n del tipo

$$\dot{x} = f(x). \quad (\text{E.1})$$

Tutti gli spazi tangenti si identificano evidentemente con \mathbb{R}^n , e l'evoluzione del vettore tangente ξ in \mathbb{R}^n è data dall'equazione alle variazioni (lineare non autonoma)

$$\dot{\xi} = A(\Phi^t(x))\xi . \quad (\text{E.2})$$

Vi sono algoritmi che consentono di integrare numericamente le (E.1), (E.2) con ragionevole approssimazione⁸⁴⁾ e dunque di determinare $x(t) := \Phi_t(x_0)$ e $\xi(t) := D\Phi_{x_0}^t \xi_0$ per ogni t , in corrispondenza ad assegnati dati iniziali x_0 e ξ_0 .

Per calcolare il solo esponente massimo $\chi_1 = \chi_1(x_0)$ sarebbe sufficiente, in linea di principio, prendere a caso⁸⁵⁾ un vettore tangente iniziale ξ_0 , per fissare le idee di norma 1, e applicare la definizione, ovvero calcolare

$$\chi_1(t) = \frac{1}{t} \|\xi(t)\| ,$$

proseguendo l'integrazione finché questa quantità non appaia stabilizzarsi a un valore limite sufficientemente definito. In pratica, questa procedura si scontra con una difficoltà tecnica: proprio a causa della divergenza esponenziale, il vettore $\xi(t)$ cresce in generale rapidamente, e supera presto il limite dei numeri rappresentabili dalla macchina: si ha il cosiddetto *overflow* di macchina.

La via d'uscita tuttavia è abbastanza semplice: fissato τ non troppo grande, poniamo

$$\xi_1 = D\Phi_{x_0}^\tau \xi_0 , \quad \alpha_1 = \|\xi_1\| , \quad \tilde{\xi}_1 = \xi_1/\alpha_1 ;$$

poniamo poi, iterativamente, $x_k = \Phi^\tau(x_{k-1})$, e

$$\xi_k = D\Phi_{x_{k-1}}^\tau \tilde{\xi}_{k-1} , \quad \alpha_k = \|\xi_k\| , \quad \tilde{\xi}_k = \xi_k/\alpha_k .$$

Un minimo di riflessione mostra che si ha

$$D\Phi_{x_0}^{k\tau} \xi_0 = \alpha_k \cdots \alpha_1 \tilde{\xi}_k ,$$

e dunque

$$\chi_1(k\tau) = \frac{1}{k\tau} \sum_{j=1}^k \log \alpha_j .$$

Il problema dell'*overflow* in questo modo scompare.

La figura E.1 mostra l'andamento di $\chi_1(t)$ in funzione di t , per il modello di Hénon e Heiles a energia $E = 0.125$. Le diverse curve si riferiscono a diversi dati iniziali: tre di esse sono prese nella regione caotica, e sembrano convergere a un limite comune per $t \rightarrow \infty$,

⁸⁴⁾ Il problema dell'approssimazione degli integratori numerici è spinosissimo, e non vi entriamo. Ricordiamo solo che un metodo buono in assoluto non esiste, e che per quanto possibile bisogna adattare l'algoritmo di calcolo al problema in esame. Ad esempio, per i sistemi hamiltoniani è opportuno usare i cosiddetti *algoritmi simplettici* (ogni passo di integrazione è realizzato con una trasformazione canonica prossima all'identità).

⁸⁵⁾ Un vettore scelto a caso fornisce, evidentemente, il massimo esponente χ_1 . Ma anche nell'ipotesi in cui ξ_0 coincidesse con e_2 , basterebbero i primi errori numerici a fornire una componente non nulla lungo e_1 , cosicché in ogni caso si arriva a calcolare χ_1 .

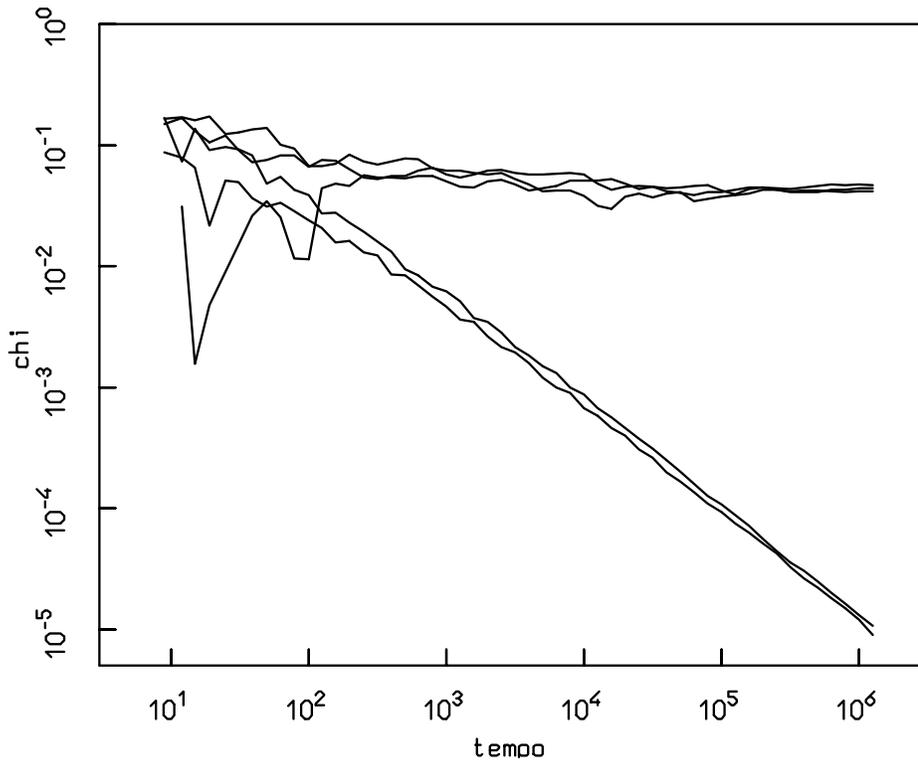


Figura E.1. $\chi_1(t)$ in funzione di t , per il modello di Hénon e Heiles a energia $E = 0.125$. Tre curve con dato iniziale nella regione caotica convergono a un medesimo limite; due altre curve con dato iniziale nella regione ordinata mostrano $\chi_1(t)$ che tende a zero per $t \rightarrow \infty$.

coerentemente con l'idea che il sistema, ristretto alla regione caotica, sia ergodico. Due curve sono invece prese in due diverse regioni ordinate, e l'andamento, che risulta essere grosso modo $\chi_1(t) \sim t^{-1}$ (la scala è log-log), indica che $\chi_1(t)$ tende a zero per $t \rightarrow \infty$, come avviene nei sistemi integrabili.

Veniamo ora al calcolo di χ_2 ; l'estensione al calcolo dei primi p ECL è poi immediata. L'idea di base è quella di prendere a caso 2 vettori iniziali ξ_0, η_0 , e posto $[\xi, \eta] =$ parallelogramma generato dai vettori ξ, η , guardare alla crescita dell'area di $[\xi(t), \eta(t)]$: in linea di principio, in base alla teoria, la somma $\chi_1(x_0) + \chi_2(x_0)$ è il limite della quantità

$$\frac{1}{t} \log \text{Area} [\xi(t), \eta(t)] .$$

In pratica, alla precedente difficoltà dovuta alla rapida crescita di uno o entrambi i vettori, si aggiunge qui una difficoltà ulteriore: in generale i due vettori $\xi(t), \eta(t)$ collasano in una medesima direzione (la più dilatante), e l'angolo tra i due, essenziale per il calcolo dell'area, si perde. Si procede allora così: si parte con ξ_0, η_0 ortogonali, precisamente tali da formare un quadrato di lato uno, e si calcolano $\xi_1 = D\Phi_{x_0}^T \xi_0, \eta_1 = D\Phi_{x_0}^T \eta_0$; per il primo vettore si procede come sopra, definendo α_1 e $\tilde{\xi}_1$, mentre per il secondo non solo lo si riporta a norma uno, ma lo si riporta anche a essere ortogonale al primo, *mantenendolo però nel piano di ξ_1 e η_1* . Precisamente, si pone

$$\eta_{\perp} = \eta_1 - (\eta_1 \cdot \tilde{\xi}_1) \tilde{\xi}_1 , \quad \beta_1 = \|\eta_{\perp}\| , \quad \tilde{\eta}_1 = \eta_{\perp} / \beta_1 .$$

Si osservi che l'area del parallelogramma è stata ridotta complessivamente del fattore $\alpha_1\beta_1$. Si definiscono poi iterativamente, in modo ovvio, $\xi_k, \eta_k, \alpha_k, \beta_k, \tilde{\xi}_k, \tilde{\eta}_k$, e un minimo di riflessione mostra che si ha

$$\text{Area}[D\Phi_{x_0}^{k\tau}\xi_0, D\Phi_{x_0}^{k\tau}\eta_0] = \prod_{1 \leq j \leq k} \alpha_j \beta_j$$

$$\chi_1(x_0) = \lim_{k \rightarrow \infty} \frac{1}{kt} \sum_{j=1}^k \log \alpha_j, \quad \chi_2(x_0) = \lim_{k \rightarrow \infty} \frac{1}{kt} \sum_{j=1}^k \log \beta_j.$$

Per questa via il calcolo simultaneo di χ_1 e χ_2 diventa agevole. Si osservi che per ogni k il passaggio da ξ_k, η_k a $\tilde{\xi}_k, \tilde{\eta}_k$ altro non è che la procedura di ortonormalizzazione di Gram–Schmidt.

La generalizzazione al calcolo dei primi p esponenti di Lyapunov è semplice, e lasciata come esercizio [*Suggerimento*: si segue l'evoluzione di p vettori tangenti, e a intervalli regolari di tempo si esegue l'ortonormalizzazione di Gram–Schmidt. Al passo k intervengono in modo naturale nell'algorithmo di ortonormalizzazione p costanti $\alpha_k^{(i)}$, $i = 1, \dots, p$, e si ha $\chi_i(x_0) = \lim_{k \rightarrow \infty} (kt)^{-1} \sum_{j=1}^k \log \alpha_j^{(i)}$].

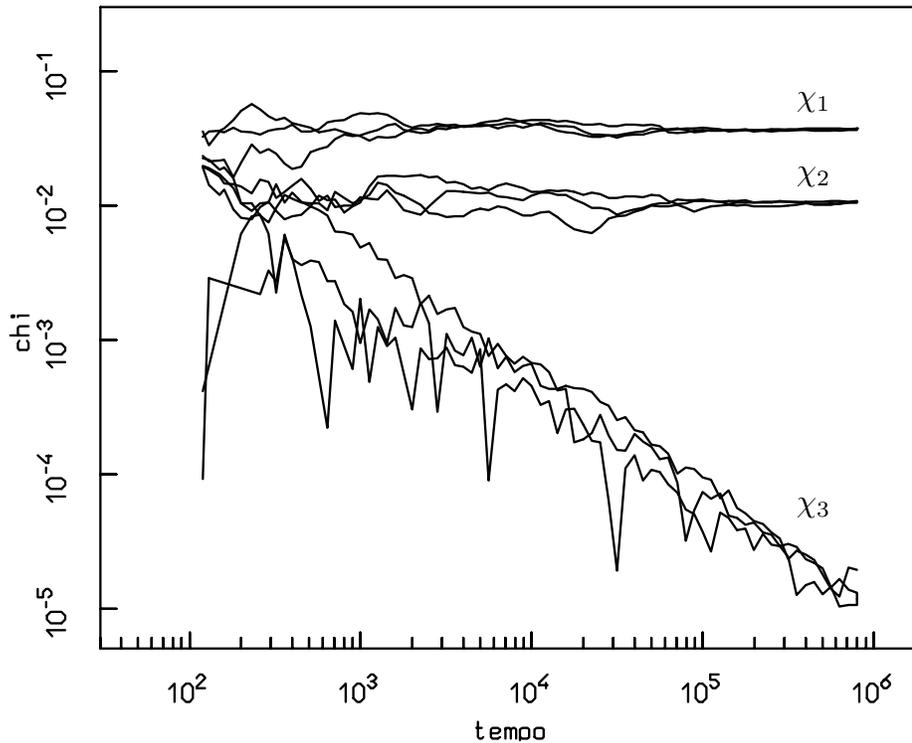


Figura E.2. Il calcolo dei primi tre ECL per un sistema hamiltoniano a tre gradi di libertà (tre diverse orbite caotiche). Le curve corrispondenti a χ_1 e χ_2 sembrano convergere a ben definiti limiti, mentre quelle corrispondenti a χ_3 sembrano tendere a zero per $t \rightarrow \infty$.

La figura E.2 illustra il calcolo di $\chi_i(t)$, $i = 1, 2, 3$, per un sistema hamiltoniano a tre gradi di libertà, introdotto da G. Contopoulos in connessione a problemi di dinamica

stellare; l'hamiltoniana è

$$H(p, q) = \frac{1}{2} \sum_{i=1}^3 (p_i^2 + \omega_i^2 q_i^2) + q_1^2 (q_2 + q_3) ,$$

con $\omega_1 = 1$, $\omega_2 = \sqrt{2}$, $\omega_3 = \sqrt{3}$. Sono stati presi tre dati iniziali nella regione caotica a energia $E = 0.15$, e per ciascuno di essi sono stati calcolati i primi tre ECL; come si vede, le curve corrispondenti ai primi due sembrano convergere a ben definiti limiti indipendenti dal dato iniziale, mentre, in accordo con la teoria, quelle corrispondenti al terzo sembrano tendere a zero (grosso modo come t^{-1}) per $t \rightarrow \infty$.

Per maggiori informazioni, si veda: G. Benettin *et al.*, *Lyapunov characteristic exponents for smooth dynamical systems; a method for computing all of them*. Part I: theory. Part II: numerical application. *Meccanica* **15**, pag. 9 e 21 (1980). Anche: G. Benettin e L. Galgani, *Lyapunov characteristic numbers and stochasticity*, in *Intrinsic stochasticity in plasmas*, edito da G. Laval e D. Gresillon (Ed. de Physique, Orsay 1979).

- **Osservazione.** Il metodo di calcolo qui proposto presuppone evidentemente la conoscenza delle equazioni del moto del sistema. Per sistemi accessibili solo sperimentalmente (andamenti economici, tracciati encefalografici) gli ECL si possono ugualmente calcolare, ma la faccenda si fa più complicata. Si veda, a questo proposito, Wolf *et al.*, *Determining Lyapunov exponents from a time series*, *Physica* 16D, 285–317 (1985).