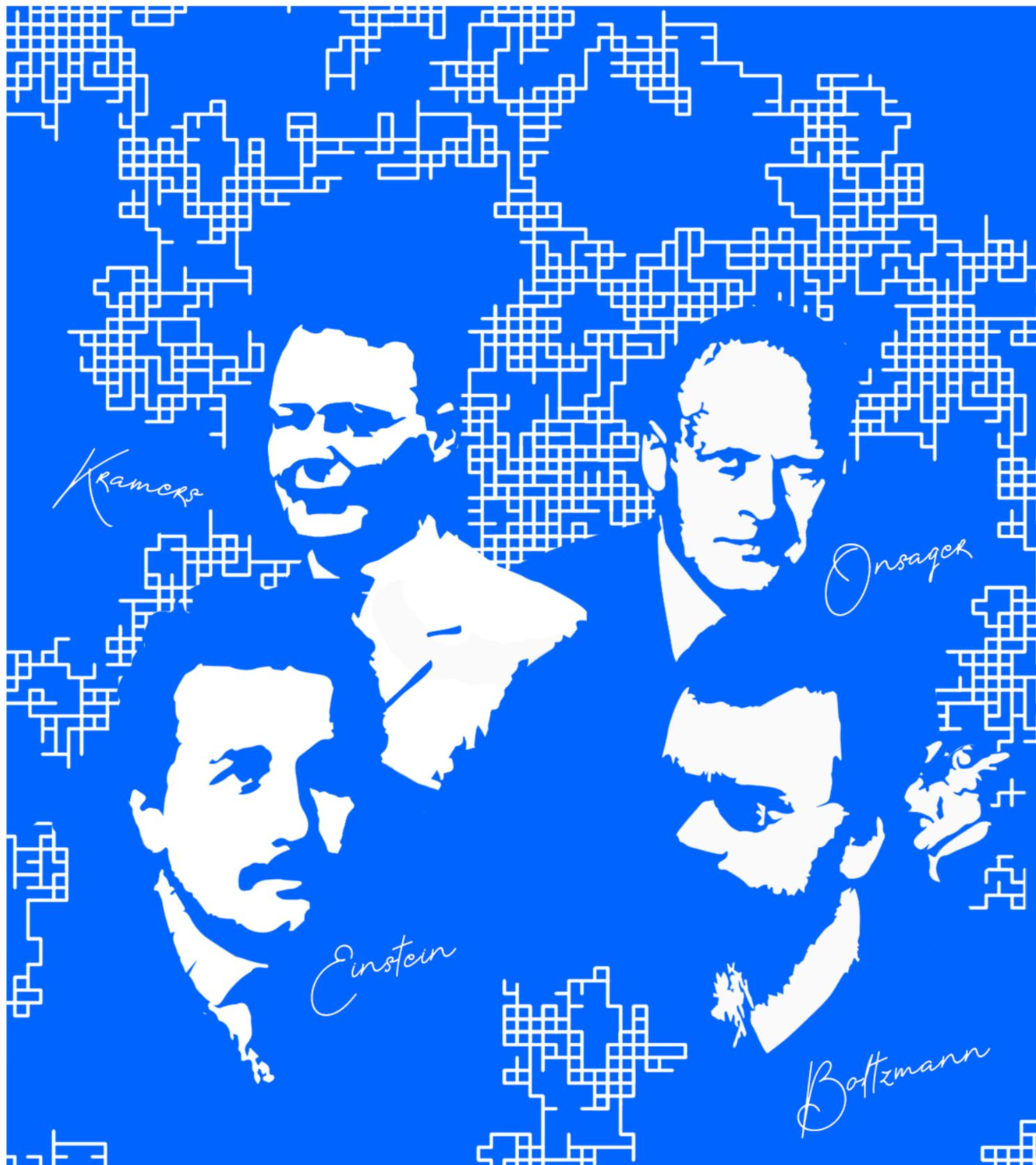


Appunti del corso di A. Vulpiani

Meccanica Statistica del Non Equilibrio



Gli appunti raccolti sono il risultato dell'elaborazione personale delle lezioni del corso "Meccanica Statistica del Non Equilibrio" del professor Angelo Vulpiani, A.A. 2017/18 presso il Dipartimento di Fisica dell'università di Roma "La Sapienza". Si ringrazia vivamente il professor Angelo Vulpiani per essersi reso disponibile alla rilettura e correzione della versione finale.

Qualunque errore è da imputarsi agli autori degli appunti.

Hanno contribuito alla stesura:

D. Arella, A. Caligiuri, F. Camaglia, M. Celotto, L. Correale, C. De Luca, F. Giuffrida, A. Maiuro, P. Muratore, G. Petroni, L. Pinto, L. Sesta, R. Valencia e D. Venturelli.

Si ringrazia per la rilettura P. Bilotto.

Redazione e veste grafica a cura di F. Camaglia.

REALIZZATO CON OVERLEAF

Ultima modifica: 17 settembre 2019

Indice

1	Moto Browniano e Mondo Microscopico	7
1.1	Un po' di fenomenologia	7
1.1.1	Equazione di Einstein	8
1.1.2	Sistemi debolmente interagenti	8
1.2	Moto Browniano	10
1.3	Equazione di Langevin e coefficiente di diffusione	11
1.3.1	Il modello a tempi discreti di Lorentz	12
1.3.2	Un approccio veloce	16
1.3.3	Il modello di Lorentz generalizzato	17
1.4	Ulteriori legami tra il mondo microscopico e quello macroscopico	18
1.4.1	La viscosità	19
1.4.2	La velocità del suono	20
2	Equazione di Boltzmann e Teoria Cinetica	23
2.1	Il ruolo dell'interazione	23
2.2	Teoria cinetica	24
2.3	Teorema H di Boltzmann	27
2.3.1	Obiezioni al teorema H	29
2.3.2	Considerazione sulla forma funzionale di H	29
2.4	Il paradosso della ricorrenza	30
3	Processi markoviani	33
3.1	Le catene di Markov	33
3.1.1	Invarianza per numero finito di memorie	34
3.1.2	Matrice di transizione	35

3.2	Convergenza della probabilità	37
3.2.1	Bilancio dettagliato	38
3.3	Modello degli Ehrenfest	39
3.3.1	Risposta all'obiezione di Zermelo	41
3.3.2	Risposta all'obiezione di Loschmidt	41
4	Master equation e Fokker-Planck	45
4.1	Master equation	45
4.1.1	Teorema H per la master equation	46
4.2	Equazione di Fokker-Planck	48
4.2.1	Derivazione dell'equazione	49
4.3	Fokker-Planck e l'equazione di Langevin	50
4.3.1	Fokker-Planck e l'equazione di diffusione	51
4.3.2	Rilassamento ad una gaussiana	52
4.4	Fokker-Planck in più dimensioni	53
4.4.1	Soluzione stazionaria	54
4.4.2	Osservazioni sull'equazione di Fokker-Planck	55
4.5	Legame tra la fisica e i modelli matematici stocastici	55
4.6	Formula di Kramers-Arrhenius	57
5	Sistemi dinamici	61
5.1	Ergodicità	65
5.2	Mixing	66
5.2.1	Calcolo della misura invariante	67
5.2.2	Sistemi ergodici non in mixing	69
5.3	Teoremi Ergodici	71
5.4	Ergodicità e frequenza	72
6	Termodinamica di non equilibrio	75
6.1	Teoria delle fluttuazioni di Einstein	76
6.1.1	Dimostrazione della relazione per le fluttuazioni	78
6.2	Teoria delle perturbazioni lineari	80
6.3	Relazioni di Onsager	82
6.3.1	Connessione tra Onsager e Bilancio Dettagliato	84
6.3.2	Teoria di Onsager-Machlup	86
7	Teoremi di fluttuazione	89
7.1	Sistemi lineari	89
7.1.1	Caso in più dimensioni	91
7.2	Sistemi hamiltoniani	92
7.2.1	Obiezione di Van Kampen	94

7.3	Teorema di fluttuazione-dissipazione generalizzato	94
7.4	Applicazioni della teoria della risposta lineare	96
7.4.1	Sistema Hamiltoniano	96
7.4.2	Sistema caotico	97
7.4.3	Sistemi gradiente	98
7.5	Conclusioni	98
8	Grandi deviazioni	99
8.1	Approssimazione gaussiana e secondo principio	99
8.2	La funzione di Cramér	101
8.2.1	I limiti del TLC: la distribuzione log-normale	102
8.2.2	Ancora sulla distribuzione dell'energia	103
8.3	Il vantaggio del metodo di Cramér	104
8.3.1	Calcolo della funzione di Cramér	105
8.3.2	Simmetria di Gallavotti-Cohen	106
8.4	Teoria dell'informazione	107
8.4.1	L'entropia di Shannon	109
8.4.2	Comprimibilità dell'informazione	116
8.4.3	Entropia topologica	117
8.5	Collegamento tra teoria dell'informazione e non equilibrio	120
8.6	Relazione di Gallavotti-Cohen	122
8.6.1	Considerazioni sulla relazione di Gallavotti	123

1. Moto Browniano e Mondo Microscopico

1.1 Un po' di fenomenologia

Nello studio di problemi all'equilibrio la variabile temporale diventa ininfluente ai fini della trattazione. Quanto adesso è di nostro interesse è di estendere i metodi propri della meccanica statistica allo studio dei sistemi che sono fuori dall'equilibrio.

Definizione 1.1 — Non-Equilibrio. Distinguiamo due diversi tipi di non equilibrio:

- **non equilibrio transiente**, in cui abbiamo che il sistema si sta avvicinando all'equilibrio se si aspetta un tempo sufficientemente lungo;
- **non equilibrio stazionario**, dovuto ad una forzante esterna che fornisce al sistema energia, rendendolo un sistema aperto.

Un problema importante non affrontato nel corso di meccanica statistica all'equilibrio risiede nel potenziale di interazione. Quest'ultimo risulta centrale nella trattazione dei vari problemi e quanto sorge spontaneo domandarsi è il come si ricavi.

La procedura usualmente seguita è la seguente: si ipotizza una forma e si ottengono i vari parametri a partire dalle quantità termodinamiche misurabili. In sostanza si segue un approccio opposto rispetto ai tipici esercizi della meccanica statistica, dato che si risale al potenziale a partire dalla termodinamica e non viceversa. Tipici potenziali sono ad esempio il **potenziale di Morse** e il **potenziale di Lennard-Jones**, molto utilizzati in problemi di struttura della materia.

■ **Esempio 1.1** Potenziale di tipo Lennard-Jones:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Il termine con la potenza sesta risulta ottenibile dall'interazione dipolo-dipolo, mentre l'altro risulta inventato ad hoc per fittare i dati, dove ϵ è l'energia al minimo. ■

Questa procedura risulta particolarmente efficace rispetto al ricavare il potenziale di interazione a partire da una trattazione quantistica del sistema. Si noti infatti che una trattazione classica, basata sul considerare il potenziale risultante come semplice somma di potenziali Coulombiani, non porta alle

forme osservate. In particolare la somma di potenziali del tipo $1/r$ non consente di ottenere un minimo del potenziale in corrispondenza della distanza di legame R .

1.1.1 Equazione di Einstein

Quanto è di nostro interesse è studiare come una densità di probabilità arbitraria $\rho(t)$ approssimi la densità di probabilità di equilibrio ρ_{eq} .

■ **Esempio 1.2** Dato un sistema all'equilibrio termodinamico si ha che la distribuzione delle velocità è una Maxwell-Boltzmann:

$$\rho_{\text{eq}}(\underline{v}) \propto \exp\left\{\left(-\frac{mv^2}{2k_B T}\right)\right\}$$

come facilmente ottenibile dal canonico. Tuttavia, se prepariamo il sistema con una distribuzione $\rho(\underline{v}, t=0) \neq \rho_{\text{eq}}(\underline{v})$ cosa assicura che la distribuzione converga all'equilibrio?

$$\lim_{t \rightarrow \infty} \rho(\underline{v}, t) \stackrel{?}{=} \rho_{\text{eq}}(\underline{v})$$

Per poterlo dimostrare la metodologia rigorosa si basa sull'applicazione dallo schema hamiltoniano, il quale ha come risultato l'ottenimento di una equazione differenziale integrale non lineare di difficile risoluzione. La situazione si semplifica abbandonando l'approccio deterministico a favore di un approccio di tipo stocastico Markoviano: nei sistemi fuori dall'equilibrio l'approccio hamiltoniano non è applicabile. Un primo tentativo è dato dall'**equazione di Einstein**:

$$w = e^{S/k_B} \quad (1.1)$$

in cui w è il numero di configurazioni, legato alla probabilità, mentre S è l'entropia del sistema. All'equilibrio si ha che S , così come w , è un numero. Questo risulta scomodo perchè a noi interessa ottenere una relazione tra funzioni. Cosa fare? Possiamo considerare delle fluttuazioni di S dall'equilibrio.

■ **Esempio 1.3** Consideriamo un sistema costituito da una scatola con al suo interno N particelle. Suddividiamola idealmente in due regioni, che chiamiamo 1 e 2. In entrambe definiamo la densità di particelle ρ_i , tale che l'entropia del sistema risulta $S = S(\rho_1, \rho_2)$. All'equilibrio si ha che $\rho_1 = \rho_2$ e dunque le fluttuazioni di $S(\rho_1, \rho_2)$ rispetto a tale condizione sono

$$S(\rho_1, \rho_2) = S(\rho_1, \rho_2) \Big|_{\rho_1=\rho_2} + \frac{\partial S}{\partial \rho_1} \Big|_{\rho_1=\rho_2} \delta \rho_1 + \frac{\partial S}{\partial \rho_2} \Big|_{\rho_1=\rho_2} \delta \rho_2 + \dots$$

Un problema interessante è studiare come il sistema approssimi la condizione di equilibrio a partire da condizioni $\rho_1 \neq \rho_2$. ■

I relativi modelli stocastici, devono avere come limite l'equazione di Einstein e devono rispettare alcune simmetrie e strutture che riflettono l'origine hamiltoniana del problema.

1.1.2 Sistemi debolmente interagenti

Consideriamo la seguente hamiltoniana:

$$\mathcal{H} = \sum_n \frac{p_n^2}{2m} + \sum_{i,j} U(\underline{q}_i - \underline{q}_j) \quad (1.2)$$

di cui calcoliamo la funzione di partizione

$$Z = \int e^{-\beta H(p,q)} d^N \underline{p} d^N \underline{q} = (2\pi m k_B T)^{3N/2} \underbrace{\int e^{-\beta \sum_{i,j} U_{ij}} d^N \underline{q}}_Q \quad (1.3)$$

L'energia libera del sistema risulta

$$F = -k_B T \ln Z = -k_B T \ln (2\pi m k_B T)^{3N/2} - k_B T \ln Q(T, V), \quad (1.4)$$

mentre la pressione

$$P = -\frac{\partial F}{\partial V} = k_B T \frac{\partial}{\partial V} \ln Q(T, V), \quad (1.5)$$

la quale dipende dalla sola forma del potenziale di interazione $U(r)$. Consideriamo un potenziale $U(r)$ che risulta divergente per $r \rightarrow 0$ mentre nullo per $r \gg \sigma$, dove σ è la distanza di equilibrio. Se la distanza media tra particelle $l \sim (V/N)^{1/3} \gg \sigma$, allora il gas è diluito. In tale circostanza possiamo, all'ordine zero, approssimare $U(r) = 0$ che è il potenziale di un gas perfetto.

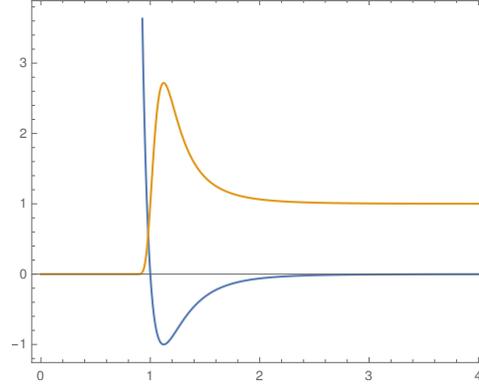


Figura 1.1: potenziale di Lennard-Jones (curva blu) con $\epsilon = 1$ e $\sigma = 1$ con sovrapposto $\exp(-U(r))$ (curva arancione). Si noti che per $r \gg 1$ abbiamo $\exp(-U(r)) = 1$, che coincide con l'approssimazione di gas ideale.

Per procedere al calcolo di Q compiamo la seguente approssimazione

$$e^{-\beta U_{ij}} = \underbrace{1}_{\text{gas perfetto}} + \underbrace{f_{ij}}_{\text{correzioni}}$$

da cui

$$Q(T, V) = \int \prod_{i < j} (1 + f_{ij}) d^N \underline{q}$$

La produttoria scritta per esteso è

$$\begin{aligned} \prod_{i < j} (1 + f_{ij}) &= (1 + f_{12})(1 + f_{13})(1 + f_{14}) \dots (1 + f_{23})(1 + f_{24}) \dots \\ &= 1 + \sum_{i < j} f_{ij} + \sum_{i < j < k < l} f_{ij} f_{kl} + O(f^3). \end{aligned} \quad (1.6)$$

Tenendo i termini solo al primo ordine in f si ha

$$\begin{aligned} Q(T, V) &= V^N + \sum_{i < j} \int f_{ij} d\underline{q}_i d\underline{q}_j d^{N-2} \underline{q} = V^N + \frac{N(N-1)}{2} \int f_{12} d\underline{q}_1 d\underline{q}_2 d^{N-2} \underline{q} \\ &= V^N + \frac{N(N-1)}{2} V^{N-2} \int f_{12} \underline{q}_1 d\underline{q}_2, \end{aligned}$$

in cui al secondo passaggio si è sfruttato il fatto che la somma degli integrali risulta essere la somma di un numero di termini uguali pari al numero di coppie tra N elementi, cioè $N(N-1)/2$. L'ultimo

integrale, ricordando che $f_{12} = e^{-\beta U(r)} - 1$ dove $r = |\underline{q}_1 - \underline{q}_2|$, lo si può risolvere passando alle coordinate $d\underline{q}_1 d\underline{q}_2 \rightarrow d\underline{r} d\underline{q}_1$

$$\begin{aligned} Q(T, V) &= V^N + \frac{N^2}{2} V^{N-2} \int (e^{-\beta U(r)} - 1) d\underline{r} d\underline{q}_1 = V^N + \frac{N^2}{2} V^{N-1} \int (e^{-\beta U(r)} - 1) d\underline{r} \\ &= V^N + N^2 V^{N-1} \underbrace{2\pi \int_0^\infty r^2 (e^{-\beta U(r)} - 1) dr}_{-b_2(T)} = V^N \left(1 - \frac{N^2}{V} b_2(T) + \dots \right), \end{aligned}$$

dove si è integrato in $d\underline{q}_1$ dato che la funzione integranda ne risulta indipendente, dando un ulteriore V , per poi passare in sferiche rispetto a r . Calcoliamo

$$\ln Q = N \ln V + \ln \left(1 - N \rho b_2(T) \right) \approx N \ln V - N \left(\frac{N}{V} \right) b_2(T),$$

in cui si è sfruttato il fatto che il gas è diluito, per poi

$$P = k_B T \frac{\partial}{\partial V} \ln Q(T, V) \simeq k_B T \left(\rho + \rho^2 b_2(T) \right), \quad (1.7)$$

che costituisce lo **sviluppo del viriale**, cioè le correzioni al gas perfetto. Abbiamo ottenuto una relazione $P = f(\rho, T)$ che possiamo provare a invertire per ottenere il potenziale di interazione, una volta ipotizzata la sua forma. Il risultato di sopra era stato predetto fenomenologicamente da Van der Waals da ragionamenti sul volume escluso e la pressione in eccesso del gas, ottenendo:

$$\frac{P + A\rho^2}{k_B T} \left(\frac{1}{\rho} - B \right) = 1. \quad (1.8)$$

Questo risultato lo si può ottenere a partire da eq. (1.7) considerando un potenziale di tipo square-well:

$$U(r) = \begin{cases} \infty & r < r_0 \\ -\epsilon & r_0 < r < r_1 \\ 0 & r > r_1. \end{cases}$$

con approcci legati alla teoria dei liquidi, in cui le ipotesi di interazione dominante a coppie in un sistema isotropo ed omogeneo si traducono nella funzione di distribuzione radiale.

1.2 Moto Browniano

Il moto browniano non è altri che l'osservazione del moto casuale di oggetti di dimensione dell'ordine del μm in una soluzione. La peculiarità di questo fenomeno è che ha consentito l'estensione delle tecniche proprie della meccanica statistica a fenomeni che hanno luogo a scale molto più grandi di quelle atomiche. Infatti, mediante il suo utilizzo, Einstein dimostrò che il moto lungo una direzione x è tale che per $t \gg 1$:

$$\langle (x(t) - x(0))^2 \rangle \approx 2Dt, \quad (1.9)$$

Qui D è il **coefficiente di diffusione** (Einstein-Smoluchowski) risulta pari a

$$D = \frac{T}{6\pi\eta r} \frac{\mathcal{R}}{N_A} \quad (1.10)$$

Nota D dall'osservazione del moto (eq. (1.9)), è possibile quindi ricavare una stima sperimentale del numero di Avogadro N_A usando la eq. (1.10), la viscosità η , il raggio tipico r degli oggetti, la temperatura T e la costante dei gas ideali \mathcal{R} . Per la determinazione del coefficiente di diffusione D è necessario compiere uno studio del moto di queste palline in una soluzione, con la speranza che queste vadano all'equilibrio come degli atomi.

1.3 Equazione di Langevin e coefficiente di diffusione

Nel 1908 Langevin, propose una riformulazione matematica del lavoro di Einstein, mediante l'utilizzo di **equazioni differenziali stocastiche**. Per semplicità illustriamo il caso in una dimensione. Il nostro scopo è quello di scrivere un'equazione per una particella colloidale di massa m sospesa in un fluido. Il moto della particella sarà governato dall'equazione

$$m\dot{v} = f_T \quad (1.11)$$

dove f_T è la forza totale agente sulla particella.

Supponiamo di aver immerso la nostra particella nel fluido: le uniche forze agenti sulla particella saranno quelle esercitate dai continui urti con le molecole del mezzo. Se la particella è in movimento, subirà più urti in direzione opposta al moto (frontalmente), così che l'effetto medio sarà di frenamento; questo si può schematizzare con la **forza di Stokes**:

$$f_{Stokes} = -6\pi\eta r v \quad (1.12)$$

dove η è la viscosità, r il raggio della particella e v la velocità. Inoltre può accadere che, se la particella si trova inizialmente ferma, venga messa in moto a causa degli urti subiti dalle molecole del fluido. Schematizziamo questo secondo fenomeno con una **forza fluttuante** $F_R(t)$ ("random") di natura stocastica dipendente dal tempo. Possiamo dunque scrivere l'equazione del moto come

$$m\dot{v} = -6\pi\eta r v + F_R(t) \quad (1.13)$$

Notiamo che senza il termine fluttuante otterremmo

$$\dot{v} = -\frac{6\pi\eta r}{m}v$$

che integrata fornisce

$$v(t) = v(0)e^{-\frac{6\pi\eta r}{m}t} \quad (1.14)$$

da cui si vede che la quantità $\frac{m}{6\pi\eta r}$ ha le dimensioni di un tempo che chiamiamo τ , **tempo tipico di rilassamento**¹. Dall'eq. (1.14) la velocità tenderebbe a 0 per tempi lunghi, che non è ciò che si osserva sperimentalmente, per cui la forza fluttuante è necessaria.

Possiamo scrivere l'equazione completa, che tiene conto degli urti tra le molecole, come

$$\dot{v} = -\frac{1}{\tau}v + \frac{1}{m}F_R(t) \quad (1.15)$$

Partendo da questa vogliamo trovare l'espressione del coefficiente di diffusione D già trovata da Einstein. Scriviamo l'eq. (1.15) come

$$\ddot{x} = -\frac{1}{\tau}\dot{x} + \frac{1}{m}F_R(t)$$

moltiplichiamo per x

$$x\ddot{x} = -\frac{1}{\tau}\dot{x}x + \frac{1}{m}xF_R(t)$$

A questo punto usiamo le seguenti identità

$$\begin{aligned} x\dot{x} &= \frac{1}{2} \frac{d}{dt} x^2 \\ x\ddot{x} + \dot{x}^2 &= \frac{1}{2} \frac{d^2}{dt^2} x^2 \end{aligned}$$

¹In sistemi colloidali di una particella micrometrica immersa in un fluido $\tau \sim 10^{-6} - 10^{-7}$ s.

$$\frac{d^2}{dt^2}x^2 = -\frac{1}{\tau} \frac{d}{dt}x^2 + \frac{2}{m}xF_R(t) + 2\dot{x}^2 \quad (1.16)$$

Facciamone ora la media (dobbiamo immaginare di avere tante particelle con cui fare l'esperimento e per ognuna calcoliamo la posizione ecc..)

$$\frac{d^2}{dt^2}\langle x^2 \rangle = -\frac{1}{\tau} \frac{d}{dt}\langle x^2 \rangle + \frac{2}{m}\langle xF_R(t) \rangle + 2\langle \dot{x}^2 \rangle \quad (1.17)$$

Il termine $\langle xF_R(t) \rangle$ rappresenta gli urti delle molecole d'acqua sulla particella, per cui viene naturale mettere questo termine a zero (subisce tanti urti da una parte quante dall'altra). Il termine $\langle \dot{x}^2 \rangle$ invece dipenderà dal tempo e dalla condizione iniziale sulla velocità. Essendo interessati al comportamento a tempi lunghi, dopo una serie di urti con le molecole, la particella andrà all'equilibrio con loro e quindi vale il **teorema di equipartizione dell'energia**:

$$\frac{1}{2}m\langle \dot{x}^2 \rangle = \frac{1}{2}k_B T$$

Quindi abbiamo ottenuto

$$\frac{d^2}{dt^2}\langle x^2 \rangle = -\frac{1}{\tau} \frac{d}{dt}\langle x^2 \rangle + 2\frac{k_B T}{m}$$

Questa equazione è facilmente integrabile e il risultato ($x(0) = 0$, $v(0) = v_0$) è

$$\langle x^2 \rangle = 2\frac{K_B T \tau}{m}t + v_0 \tau e^{-\frac{t}{\tau}} \quad (1.18)$$

Che per tempi lunghi possiamo scrivere come

$$\langle x^2 \rangle \stackrel{t \gg 1}{\approx} 2\frac{K_B T \tau}{m}t \quad (1.19)$$

Utilizzando la relazione

$$\langle x^2 \rangle \simeq 2Dt \quad (1.20)$$

si vede che $D = \frac{K_B T \tau}{m}$. Osserviamo che il \simeq in eq. (1.20) sta ad indicare che quello è ciò che succede dopo un tempo di ordine τ . Sostituendo τ si ha $D = \frac{K_B T}{6\pi\eta r}$, che è l'espressione di Einstein-Smoluchowski. Osserviamo che è sparita la massa: per grandi tempi particelle massive e particelle leggere si muovono allo stesso modo. Per capire meglio quello che sta succedendo possiamo fare un modello molto semplificato a tempi discreti. Questo farà vedere esplicitamente che le assunzioni fatte sono effettivamente sensate.

1.3.1 Il modello a tempi discreti di Lorentz

Per capire come discretizzare l'equazione

$$\dot{v} = -\frac{1}{\tau}v + \frac{1}{m}F_R(t)$$

se ne può calcolare la velocità al tempo $t + \Delta t$ nota la velocità al tempo t (vedremo in seguito come), ottenendo:

$$v(t + \Delta t) = v(t)e^{-\frac{\Delta t}{\tau}} + \frac{1}{m} \int_t^{t+\Delta t} e^{-\frac{\Delta t}{\tau}} e^{-\frac{(t-t')}{\tau}} F_R(t') dt'$$

Se adesso assumiamo che Δt sia piccolo, ossia $\frac{\Delta t}{\tau} \ll 1$, con Δt di almeno un ordine di grandezza inferiore a $\frac{1}{\tau}$, allora possiamo espandere in serie l'esponenziale arrivando ad un'espressione per la velocità

$$v(t + \Delta t) \simeq v(t)\left(1 - \frac{\Delta t}{\tau}\right) + \frac{1}{m}F_R(t)\Delta t \quad (1.21)$$

dove sono stati considerati solo i termini al primo ordine in Δt . Per la posizione invece si ha che:

$$x(t + \Delta t) \simeq x(t) + v(t)\Delta t \quad (1.22)$$

Ponendo $t_n = n\Delta t$, il moto discretizzato della nostra particella sarà descritto dal **modello di Lorentz**:

$$v_{n+1} = av_n + bw_n \quad (1.23)$$

$$x_{n+1} = x_n + cv_n \quad (1.24)$$

dove $c = \Delta t$, $a = 1 - \frac{\Delta t}{\tau}$ con $0 < a < 1$ e b una costante da scegliere in modo opportuno. Il termine bw_n schematizza la forza stocastica $F_R(t)$, per cui, se supponiamo che la $F_R(t)$ sia distribuita gaussianamente con una distribuzione del tipo $\mathcal{N}(0, F_{R0}^2)$, possiamo scrivere:

$$F_R(n\Delta t) = F_{R0}w_n \quad (1.25)$$

dove le w_n sono identicamente distribuite secondo una gaussiana a media nulla e varianza unitaria $\mathcal{N}(0, 1)$ e indipendenti (i.i.d.). Assumiamo che la distribuzione della velocità iniziale v_0 sia $\mathcal{N}(\langle v_0 \rangle, \sigma_0^2)$ e indipendente da $\{w_n\}$: le posizioni e le velocità saranno distribuite gaussianamente in quanto combinazioni di variabili gaussiane. Possiamo riscrivere le eq. (1.23) e eq. (1.24) esplicitando i termini in modo iterativo, e mediando otteniamo:

$$\langle v_1 \rangle = a\langle v_0 \rangle$$

$$\langle v_1^2 \rangle = a^2\langle v_0^2 \rangle + b^2$$

⋮

$$\langle v_n \rangle = a^n\langle v_0 \rangle \quad (1.26)$$

$$\langle v_{n+1}^2 \rangle = a^2\langle v_n^2 \rangle + b^2 \quad (1.27)$$

Osserviamo che poiché $a < 1$, $\langle v_n \rangle = a^n\langle v_0 \rangle$ tende a zero per $n \rightarrow \infty$ (ossia per grandi tempi). Questo fenomeno dal punto di vista fisico corrisponde all'attrito.

Dalle precedenti possiamo calcolare la varianza della distribuzione definita come $\sigma_n^2 = \langle v_n^2 \rangle - \langle v_n \rangle^2$. Ricordando che $a\langle v_n \rangle = \langle v_{n+1} \rangle$ si ottiene:

$$\sigma_{n+1}^2 = a^2\sigma_n^2 + b^2 \quad (1.28)$$

Essendo interessati al limite $n \rightarrow \infty$, osserviamo in fig. 1.2 che un valore limite (che corrisponde ad avere $\sigma_{n+1}^2 = \sigma_n^2$) esiste grazie al fatto che $a < 1$:

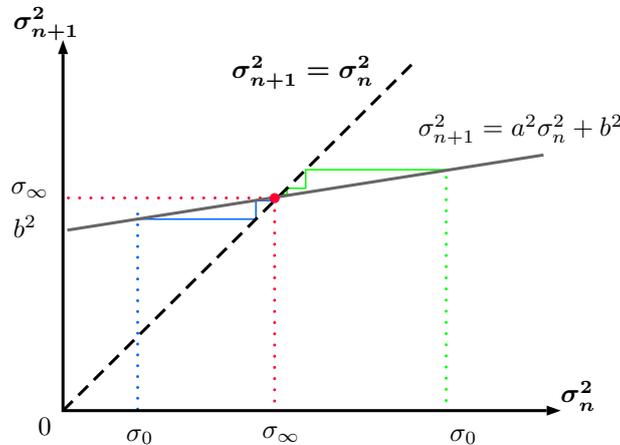


Figura 1.2: Schema per l'esistenza di un valore limite per σ . Qualunque sia il valore iniziale di σ_0^2 , dalla legge di ricorsività per cui σ_{n+1} del passo precedente diventa σ_n per il passo successivo, osserviamo schematicamente che deve esistere σ_∞^2 .

L'eq. (1.28) è una serie convergente e all'infinito con:

$$\lim_{n \rightarrow +\infty} \sigma_n^2 = \frac{b^2}{1-a^2} \quad (1.29)$$

quindi, considerando che $\langle v_n \rangle \xrightarrow{n \rightarrow \infty} 0$, troviamo

$$\langle v^2 \rangle = \frac{b^2}{1-a^2} \quad (1.30)$$

Pertanto la velocità v_n della particella al tempo n è distribuita come una gaussiana (come accennato in precedenza) e tende rapidamente² alla distribuzione asintotica $\mathcal{N}(0, \langle v^2 \rangle)$. Si tratta della **distribuzione di Maxwell-Boltzmann**. Analizzando l'eq. (1.28) in termini di distanza dal valore asintotico $\delta_n \equiv \sigma_n^2 - \langle v^2 \rangle$ abbiamo che

$$\delta_{n+1} = \sigma_{n+1}^2 - \langle v^2 \rangle = a^2 \delta_n + \underbrace{a^2 \langle v^2 \rangle + b^2 - \langle v^2 \rangle}_{=0} = a^2 \delta_n$$

e iterando si ottiene $\delta_n = a^{2n} \delta_0$ da cui

$$\sigma_n^2 = \langle v^2 \rangle + a^{2n} (\sigma_0^2 - \langle v^2 \rangle) \quad (1.31)$$

dove si vede chiaramente che per $n \rightarrow \infty$, $\sigma_n \rightarrow \sigma_\infty = \sqrt{\langle v^2 \rangle} = \frac{b}{\sqrt{1-a^2}}$. In termini "fisici" il valore di a definisce un tempo caratteristico dopo il quale la distribuzione diventa pressoché stazionaria.

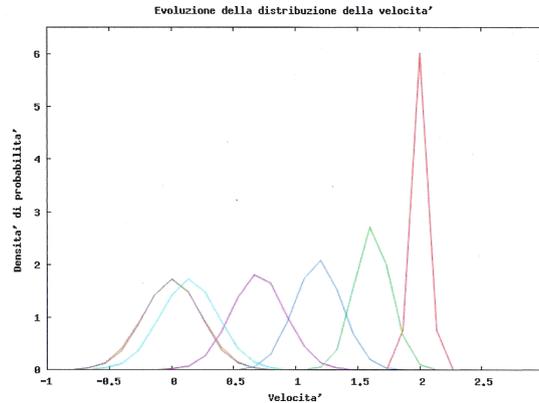


Figura 1.3: Il grafico mostra l'evoluzione della densità di probabilità della velocità a partire da una $\mathcal{N}(v_0, \sigma_0^2)$ (in rosso) a $\mathcal{N}(v_0, \sigma_\infty^2)$ (in marrone).

Morale: se partiamo da una distribuzione gaussiana stretta otteniamo una distribuzione gaussiana "più allargata". Si può dimostrare che se la distribuzione di partenza non è una gaussiana ma la distribuzione di w lo è, allora finiamo sempre su una gaussiana³. Quindi in un certo senso questo tipo di processo "dimentica" da dove è partito. Per avere un'idea dei tempi di rilassamento della gaussiana basta considerare $\delta_n = a^{2n} \delta_0$, dove a^{2n} per tempi lunghi possiamo scriverlo come

$$a^{2n} = \left(1 - \frac{\Delta t}{\tau}\right)^{2n} \approx e^{-\frac{2n\Delta t}{\tau}}$$

allora

$$\delta(t) = \delta(0) e^{-\frac{2t}{\tau}}$$

²La rapidità del processo dipende dal tempo caratteristico τ .

³Questa è una cosa positiva perché tramite processi fisici arriviamo alla distribuzione di M-B.

cioè il tempo per cui la velocità termalizza è $\sim \tau$ proprio come ci si aspetta. Torniamo un'istante su σ_∞ perché da questo possiamo capire cosa è b , infatti

$$\sigma_\infty = \langle v^2 \rangle = \frac{b^2}{1-a^2} = \frac{b^2}{(1-a)(1+a)} \simeq \frac{b^2}{2c}\tau$$

avendo usato che $(1+a) \simeq 2$ e $(1-a) = \frac{c}{\tau}$. A questo punto possiamo ricavare b

$$b \simeq \sqrt{\frac{2\langle v^2 \rangle}{\tau}}c = \sqrt{\frac{2\langle v^2 \rangle}{\tau}}\sqrt{\Delta t} \quad (1.32)$$

Questo ci fa capire perché abbiamo preferito fare i conti in maniera discreta e non lavorare nel continuo. Sostituendo i risultati ottenuti nell'eq. (1.23) abbiamo

$$v_{n+1} - v_n = -\frac{\Delta t}{\tau}v_n + \sqrt{\frac{2\langle v^2 \rangle}{\tau}}\sqrt{\Delta t}w_n$$

dove w_n è un numero estratto gaussianamente, diciamo $O(1)$. Vista così uno sarebbe tentato a fare la derivata: dividiamo per Δt

$$\frac{v_{n+1} - v_n}{\Delta t} = -\frac{1}{\tau}v_n + \sqrt{\frac{2\langle v^2 \rangle}{\tau}}\frac{1}{\sqrt{\Delta t}}w_n$$

Mandando $\Delta t \rightarrow 0$ otteniamo la scrittura formale

$$\dot{v} = -\frac{v}{\tau} + A\xi \quad (1.33)$$

dove A è una certa costante e ξ diverge poiché è proporzionale ad $\frac{1}{\sqrt{\Delta t}}$. Tale scrittura è solamente formale perché la derivata non esiste! La scrittura corretta è quindi

$$dv = -\frac{v}{\tau}dt + Adw \quad (1.34)$$

Da questo modello cerchiamo ora di ottenere la formula di Einstein-Smoluchowski.

Poiché siamo interessati al moto per tempi "lunghi" è ragionevole fare l'assunzione che la distribuzione delle velocità sia data da quella asintotica $\mathcal{N}(0, \langle v^2 \rangle)$. Pertanto possiamo studiare solo il cammino casuale, o **random walk**, dato dall'eq. (1.24). Ad n passi lo spostamento della particella sarà dato da

$$\Delta x_n = x_n - x_0 = c \sum_{j=0}^{n-1} v_j \quad (1.35)$$

quindi la variabile Δx_n è lo spostamento al passo n -esimo. In accordo con l'intuizione fisica, Δx_n è una variabile a media nulla $\langle \Delta x_n \rangle = 0$, mentre lo spostamento quadratico è dato da

$$\langle \Delta x_n^2 \rangle = \left\langle c^2 \left(\sum_{j=0}^{n-1} v_j \right)^2 \right\rangle = c^2 \sum_{j=0}^{n-1} \langle v_j^2 \rangle + 2c^2 \sum_{j=0}^{n-1} \sum_{k=1}^{n-j-1} \langle v_j v_{j+k} \rangle \quad (1.36)$$

Dobbiamo stimare il termine $\langle v_j v_{j+k} \rangle$. Ricordando l'eq. (1.23), possiamo scrivere

$$v_n v_{n+1} = v_n (av_n + bw_n) = av_n^2 + bv_n w_n$$

Mediando e ricordando le proprietà di w si ha

$$\langle v_n v_{n+1} \rangle = a \langle v_n^2 \rangle \simeq a \langle v^2 \rangle$$

Allo stesso modo si ha che $\langle v_{n-1}v_{n+1} \rangle = a^2 \langle v^2 \rangle$. In generale quindi:

$$\langle v_j v_{j+k} \rangle = a^k \langle v^2 \rangle \quad (1.37)$$

In questo modo otteniamo⁴:

$$\begin{aligned} \langle \Delta x_n^2 \rangle &= c^2 n \langle v^2 \rangle + 2c^2 \langle v^2 \rangle \sum_{j=0}^{n-1} \frac{a(1-a^{n-j-1})}{1-a} = \\ &= c^2 n \langle v^2 \rangle + 2c^2 \langle v^2 \rangle \frac{a}{1-a} \left(n - \sum_{j=0}^{n-1} a^{n-j-1} \right) \end{aligned}$$

La sommatoria è sicuramente una quantità limitata per qualsiasi valore di n . A questo punto, poiché il coefficiente di diffusione è definito formalmente da

$$D = \lim_{t \rightarrow +\infty} \frac{1}{2t} \langle [x(t) - x(0)]^2 \rangle$$

e ricordando che $t = n\Delta t$, possiamo scrivere

$$D = \lim_{n \rightarrow +\infty} \frac{1}{2n\Delta t} \langle (x_n - x_0)^2 \rangle$$

Che ci consente di arrivare all'espressione per il coefficiente di diffusione

$$D = \frac{c^2}{\Delta t} \frac{\langle v^2 \rangle}{2} + \frac{a}{1-a} \frac{c^2}{\Delta t} \langle v^2 \rangle$$

Esplicitando a e c otteniamo

$$D = \langle v^2 \rangle \tau - \frac{\langle v^2 \rangle \Delta t}{2}$$

Notiamo infine che nel limite in cui Δt va a zero si riottiene l'espressione precedentemente ottenuta

$$D = \lim_{\Delta t \rightarrow 0} \left(\langle v^2 \rangle \tau - \frac{\langle v^2 \rangle \Delta t}{2} \right) = \langle v^2 \rangle \tau \quad (1.38)$$

E ricordando il teorema di equipartizione dell'energia ritroviamo

$$D = \frac{K_B T}{m} \tau$$

1.3.2 Un approccio veloce

Partiamo direttamente dall'eq. (1.36):

$$\langle \Delta x_n^2 \rangle = c^2 \sum_{j=0}^{n-1} \langle v_j^2 \rangle + 2c^2 \sum_{j=0}^{n-1} \sum_{k=1}^{n-j-1} \langle v_j v_{j+k} \rangle$$

Ricordando quindi che vale l'eq. (1.37), assumendo che la distribuzione vada "velocemente" all'equilibrio, possiamo scrivere:

$$\langle \Delta x_n^2 \rangle = \langle v^2 \rangle n c^2 + 2c^2 n \langle v^2 \rangle \sum_{k=1}^n a^k$$

con $n \gg 1$. Notiamo che, usando i risultati appena ottenuti, tutte le serie che compaiono sono calcolabili (quella su k è una serie geometrica troncata). Inoltre osserviamo che se la serie in k decade

⁴Ricordiamo che per una serie geometrica vale $\sum_{j=1}^n a^j = \frac{a(1-a^n)}{1-a}$ se $a < 1$.

più velocemente di k^{-1} (serie sommabile), perciò possiamo sommare fino a ∞ : è il nostro caso perché gli elementi della serie decadono a potenza, dunque si ha che

$$\langle \Delta x_n^2 \rangle \simeq \langle v^2 \rangle n c^2 + 2c^2 n \langle v^2 \rangle \sum_{k=1}^{\infty} a^k = \langle v^2 \rangle n c^2 + 2c^2 n \langle v^2 \rangle \frac{a}{1-a}$$

Ricordando che $D \simeq \frac{1}{2} \frac{\langle \Delta x_n^2 \rangle}{n \Delta t}$ si ha

$$D \simeq \frac{\langle v^2 \rangle n \Delta t^2}{2n \Delta t} + \frac{2n \Delta t^2}{2n \Delta t} \frac{a}{1-a} \langle v^2 \rangle$$

e sostituendo a , nel limite $\Delta t \rightarrow 0$ si (ri)ottiene

$$D = \langle v^2 \rangle \tau$$

Notare che dimensionalmente questa è l'unica formula che può avere senso (a parte un eventuale fattore numerico): in qualche modo il coefficiente di diffusione ha le dimensioni di una velocità al quadrato per un tempo, dovuto ad un effetto di "scorrelazione" del processo.

1.3.3 Il modello di Lorentz generalizzato

Si può dimostrare per il modello discreto di Lorentz che:

Teorema 1.3.1 presa una distribuzione iniziale arbitraria delle velocità $P_v(v; 0)$, si ha che:

- se $P_w(w)$ gaussiana allora per $n \rightarrow \infty$ si ha che $P(v; n) \rightarrow P^{(\infty)}(v)$ gaussiana ;
- se $P_w(w)$ non è gaussiana (ma sufficientemente regolare da ammettere cumulanti) esiste una distribuzione limite per $P(v; n)$.

Per completezza riportiamo la definizione di cumulante.

Definizione 1.2 — Cumulante. Data la distribuzione di probabilità $P_x(x)$, si definisce la rispettiva **funzione caratteristica** $\Phi_x(t)$

$$\Phi_x(t) = \langle e^{itx} \rangle = 1 + it \langle x \rangle - \frac{t^2}{2} \langle x^2 \rangle + \dots = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \langle x^n \rangle$$

legata ai momenti della distribuzione P dalla relazione

$$\langle x^n \rangle = (-i)^n \left[\frac{d^n}{dt^n} \Phi_x(t) \right]_{t=0}$$

In alternativa si può fare uso della **funzione generatrice dei momenti** $G_x(t) = \langle e^{tx} \rangle$.

La funzione caratteristica associata ai cumulanti si definisce allo stesso modo come

$$\ln \Phi_x(t) = \ln \langle e^{itx} \rangle = it c_1 - \frac{t^2}{2} c_2 + \dots = \sum_{n=1}^{\infty} \frac{(it)^n}{n!} c_n$$

dove c_1, c_2, \dots sono i **cumulanti** della distribuzione, esprimibili in termini dei momenti (e viceversa)

$$\begin{aligned} c_1 &= \langle x \rangle \\ c_2 &= \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \\ &\dots \end{aligned}$$

Dimostrazione. Data la distribuzione di probabilità $P(v; n)$ delle velocità v al tempo n , associo la rispettiva funzione caratteristica:

$$\Phi_{v_n}(t) = \langle e^{iv_n t} \rangle \quad (1.39)$$

analogamente si opera per la distribuzione $P_w(w; n)$:

$$\Phi_{w_n}(t) = \langle e^{iw_n t} \rangle \quad (1.40)$$

E siccome nel modello di Lorentz $v_{n+1} = av_n + w_n$ con $|a| < 1$, ricordando che le v_n e le w_n sono tra loro indipendenti:

$$\Phi_{v_{n+1}}(t) = \langle e^{iv_{n+1}t} \rangle = \langle e^{iav_nt} e^{iw_nt} \rangle = \langle e^{iav_nt} \rangle \langle e^{iw_nt} \rangle = \Phi_{v_n}(at) \Phi_{w_n}(t) \quad (1.41)$$

da cui passando ai logaritmi si ottiene una relazione tra i cumulanti delle distribuzioni:

$$\ln \Phi_{v_{n+1}}(t) = \ln \Phi_{v_n}(at) + \ln \Phi_{w_n}(t) \quad (1.42)$$

Sostituiamo quindi i relativi sviluppi, ponendo C_k^n il cumulante di ordine k di v al tempo n e c_k il cumulante di w (indipendente dal tempo):

$$\sum_{k=1}^{\infty} \frac{(it)^k}{k!} C_k^{(n+1)} = \sum_{k=1}^{\infty} \frac{a^k (it)^k}{k!} C_k^{(n)} + \sum_{k=1}^{\infty} \frac{(it)^k}{k!} c_k \quad (1.43)$$

siccome questa relazione è vera $\forall t$, allora ad ogni ordine deve essere:

$$C_k^{(n+1)} = a^k C_k^{(n)} + c_k \quad (1.44)$$

si tratta di un risultato analogo a quanto ottenuto per σ^2 con l'ipotesi gaussiana, per cui esiste un valore asintotico di $C_k^{(n)}$ ad ogni ordine della forma:

$$C_k^{\infty} = \frac{c_k}{1 - a^k} \quad \text{con } |a| < 1 \quad (1.45)$$

Per cui se $P_w(w; n)$ è gaussiana, anche $P^{(\infty)}(v)$ è gaussiana infatti, essendo $c_1 = \langle w \rangle$, $c_2 = \sigma_w^2$ e $c_k = 0$ per $k \geq 3$, si ottiene che:

$$C_1^{(n)} \rightarrow \frac{\langle w \rangle}{1 - a} \quad C_2^{(n)} \rightarrow \frac{\sigma_w^2}{1 - a^2} \quad C_{k \geq 3}^{(n)} \rightarrow 0$$

■

In questo semplice modello, una distribuzione arbitraria di velocità iniziale, converge ad una distribuzione di equilibrio: ci chiediamo quindi più in generale sotto quali condizioni un sistema reale converga ad una distribuzione di equilibrio, in particolare alla distribuzione di Maxwell-Boltzmann.

1.4 Ulteriori legami tra il mondo microscopico e quello macroscopico

Il moto browniano (come lo sviluppo del viriale) è solo uno dei molti punti di contatto tra il mondo microscopico e quello macroscopico (di fatti ci si riferisce a questa frontiera come il mondo "mesoscopico"). Grazie a questi risultati, ottenuti indagando con la meccanica statistica ambiti di fisica classica, si riesce ad avere un'idea degli ordini di grandezza in analisi, senza l'uso di apparecchiature sperimentali⁵. Vediamo altri due esempi: la viscosità e la velocità del suono.

⁵In effetti questi risultati sono antecedenti all'invenzione dei microscopi ad effetto campo e simili.

1.4.1 La viscosità

Uno dei contributi importanti di Maxwell fu quello di dimostrare il fatto che in un gas rarefatto la viscosità è indipendente dalla densità; cosa abbastanza controintuitiva.

Definiamo la viscosità η attraverso le equazioni di **Navier-Stokes** (per un liquido incompressibile in assenza di gravità):

$$\rho_m(\partial_t \underline{u} + (\underline{u} \cdot \nabla) \underline{u}) = -\nabla p + \eta \Delta \underline{u} \quad (1.46)$$

dove ρ_m è la densità (riferita alla massa), p la pressione e u la velocità.

Maxwell mostrò che se abbiamo un gas di sfere dure abbastanza diluito, cioè se $R \ll L = \sqrt[3]{\frac{V}{N}}$, dove R è il raggio delle molecole e L la distanza tipica tra esse, allora

$$\eta = A \sqrt{\frac{m k_B T}{R^4}} \quad (1.47)$$

Questo risultato è importante perché η la possiamo misurare sperimentalmente; allora, poiché nella formula compaiono $k_B = \frac{R}{N_A}$, m e R , questa collega grandezze macroscopiche e microscopiche. Come possiamo ottenere questo risultato?

Consideriamo l'equazione di Navier-Stokes in prima approssimazione per un fluido con bassi numeri di Reynolds:

$$\partial_t \underline{u} = \frac{\eta}{\rho_m} \Delta \underline{u} \quad (1.48)$$

Notiamo che questa equazione ha esattamente la stessa struttura dell'**equazione di diffusione**:

$$\partial_t P = D \Delta P \quad (1.49)$$

dove possiamo prendere la P come una distribuzione di probabilità tale che, all'istante iniziale, $P(x, 0) = \delta(x)$; allora la densità di probabilità al tempo t risulta essere:

$$P(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \quad (1.50)$$

che ci dice che $\langle x^2 \rangle = 2Dt$. Quel D è proprio quello abbiamo trovato in precedenza.

Dato che l'eq. (1.48) e eq. (1.49) hanno la stessa struttura, siamo in qualche modo autorizzati a dire che $D \propto \frac{\eta}{\rho_m}$ (le equazioni di Navier-Stokes sono in effetti derivabili dalla dinamica molecolare che è descritta da eq. (1.49)). Supponiamo di avere una gas di sfere di raggio R ad una certa temperatura T con densità $\rho = \frac{N}{V}$, la densità di massa sarà quindi $\rho_m = m\rho$. In queste condizioni la velocità tipica delle particelle è la velocità termica

$$v_T \approx \sqrt{\frac{k_B T}{m}} \quad (1.51)$$

Siamo interessati al libero cammino medio l , ossia dopo quale distanza in media una particella che si muove ne incontra una. Una sfera ne colpisce un'altra se, guardando nel sistema del centro di massa della prima, l'altra si trova ad entro $2R$ dalla prima, rispetto la direzione nella quale avviene l'urto (fig. 1.4).

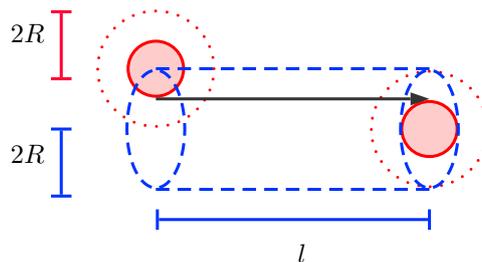


Figura 1.4: Schema dell'urto tra due particelle.

Il volume spazzato, descritto da un cilindro, è pari a $V = 4\pi R^2 l$. Se moltiplichiamo per ρ otteniamo il numero medio di sfere che incontriamo. Quindi, a meno di fattori numerici, abbiamo che $l \sim \frac{1}{R^2 \rho}$ e quindi

$$\tau \sim \frac{1}{R^2 \rho v_T}$$

dove il tempo τ è il libero cammino medio diviso la velocità termica $\tau \sim \frac{l}{v_T}$. Dalla relazione $D = \langle v^2 \rangle \tau$ otteniamo

$$D = \frac{v_T^2}{v_T R^2 \rho} = \sqrt{\frac{K_B T}{m}} \frac{1}{R^2 \rho}$$

e dunque

$$\eta \propto D \rho_m = \sqrt{\frac{m K_B T}{R^4}} \quad (1.52)$$

Che è proprio il risultato di Maxwell. Va notato che in questa formula non compare la densità.

1.4.2 La velocità del suono

Consideriamo un ultimo esempio di legame microscopico/macrosopico dato dalla velocità del suono, per la quale si ha:

$$c_s = \sqrt{\frac{\partial p}{\partial \rho_m}} \quad (1.53)$$

dove p è la pressione. Se adesso consideriamo un gas perfetto adiabatico si ha che $p \rho_m^{-\frac{5}{3}} = \text{cost}$, da cui

$$c_s = \sqrt{\frac{5 k_B T}{3 m}} \quad (1.54)$$

Questo risultato classico è importante per gli stessi motivi precedenti: c_s è misurabile e ci dice che la velocità del suono è la velocità tipica delle molecole⁶. Vediamo come si ottiene questa formula partendo rispettivamente dall'equazione di Navier-Stokes lineare e dall'equazione di continuità:

$$\begin{cases} \rho_m \partial_t \underline{u} = -\nabla p \\ \partial_t \rho_m + \nabla(\rho_m \underline{u}) = 0 \end{cases} \quad (1.55)$$

Dalla condizione di fluido fermo: $\underline{u} = 0$, $p = p_0$ e $\rho_m = \rho_{m0}$, operiamo una perturbazione lineare (onda sonora) del tipo : $p = p_0 + \delta p$ e $\rho_m = \rho_{m0} + \delta \rho_m$, ossia \underline{u} è piccola. Le nuove equazioni diventano:

$$\begin{cases} \rho_m \partial_t \underline{u} = -\nabla \delta p \\ \partial_t \delta \rho_m + \nabla(\delta \rho_m \underline{u}) = 0 \end{cases}$$

dove sono stati trascurati tutti i termini piccoli (es. $\delta \rho_m \partial_t \underline{u}$ e $\delta \rho_m \underline{u}$). Facciamo il gradiente della prima e deriviamo rispetto al tempo la seconda

$$\begin{cases} \nabla(\partial_t(\rho_{m0} \underline{u})) = -\Delta \delta p \\ \partial_t \partial_t \delta \rho_m + \partial_t \nabla(\delta \rho_m \underline{u}) = 0 \end{cases}$$

⁶Per capire meglio, si pensa al meccanismo della voce. Quando si parla, non si stanno "sparando" delle molecole nelle nostre orecchie: le molecole infatti rimangono "intorno la nostra bocca", mentre è il suono a propagarsi fino alle orecchie.

Ora, servendoci di un'equazione di stato, si deve avere che $\delta p = \frac{\partial p}{\partial \rho_m} \delta \rho_m$, per cui si ottiene infine:

$$\partial_{tt}^2 \delta \rho_m = \frac{\partial p}{\partial \rho_m} \Delta \delta \rho_m \quad (1.56)$$

Questa ha la stessa forma dell'equazione delle onde, quindi la possiamo scrivere come

$$\frac{1}{c_s^2} \partial_{tt}^2 \delta \rho_m = \Delta \delta \rho_m \quad (1.57)$$

da cui si ottiene che la velocità del suono come riportata in [eq. \(1.53\)](#).

Riferimenti bibliografici

- [1] M. Falcioni e A. Vulpiani. *Meccanica Statistica Elementare*. Springer-Verlag, 2014.

2. Equazione di Boltzmann e Teoria Cinetica

Dal semplice modello di Lorentz, una distribuzione arbitraria di velocità iniziale, converge ad una distribuzione di equilibrio: ci chiediamo quindi più in generale sotto quali condizioni un sistema reale converga ad una distribuzione di equilibrio, in particolare alla distribuzione di Maxwell-Boltzmann.

2.1 Il ruolo dell'interazione

Dato un volume V contenente N particelle identiche di massa m soggette ad un campo esterno U_{ext} (ma non è rilevante), ci poniamo nel caso in cui non ci sia interazione:

$$\mathcal{H} = \sum_n \frac{p_n^2}{2m} + U_{ext} + \mathcal{U}_I \quad (2.1)$$

osserviamo che in assenza di un campo esterno, la distribuzione di velocità arbitraria $P(v, 0)$ relativa alle particelle, rimane invariata nel tempo:

$$P(v, 0) \longrightarrow P(v, t) = P(v, 0) \quad \text{per } t \rightarrow 0$$

Questo implica che se $P(v, 0)$ è diversa dalla distribuzione di equilibrio, allora non tenderà mai a Maxwell-Boltzmann: per raggiungere quindi una distribuzione di equilibrio è necessaria un'interazione tra le particelle. Di fatti, se le particelle non interagissero non ci sarebbe alcuna statistica. Nei problemi ordinari di meccanica statistica si è implicitamente fatta l'assunzione che i sistemi tendono all'equilibrio grazie ad un'interazione del tipo:

$$U_I = \epsilon \sum_{i,j} U_I^{(2)}(\underline{q}_i - \underline{q}_j)$$

dove ϵ è sufficientemente piccolo di modo che U_I risulti praticamente nullo ai fini dei calcoli e quindi sia trascurabile una volta noto a cosa tende il sistema.

Nel 1872, Boltzmann si pose il problema di mostrare quando i sistemi si rilassano verso l'equilibrio. Nonostante i potenziali d'interazione siano una conseguenza della meccanica quantistica, era già noto da trattazioni come lo sviluppo del viriale, quale potesse essere un modello d'interazione.

2.2 Teoria cinetica

Per semplicità consideriamo un **gas di sfere dure** di raggio R soggette al potenziale:

$$U_I^{(2)}(r_{ij}) = \begin{cases} \infty & \text{se } r_{ij} \leq 2R \\ 0 & \text{altrimenti} \end{cases} \quad (2.2)$$

dove $r_{ij} = |q_i - q_j|$, che equivale a dire che le particelle rimbalzano quando si toccano, altrimenti non accade nulla (vedi fig. 2.1).

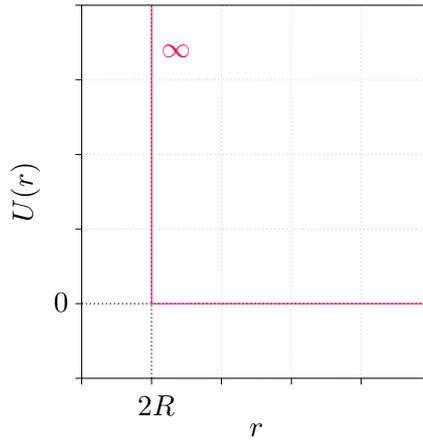


Figura 2.1: Potenziale per un gas di sfere dure.

Affinché l'approssimazione di gas perfetto sia valida ci poniamo nel caso di un **sistema diluito**, ossia:

$$l \sim \sqrt[3]{\frac{V}{N}} \gg R \quad (2.3)$$

Volendo servirsi di una trattazione rigorosa, definito il volume Γ nello spazio delle fasi \mathbb{R}^{6N} , indicando ogni stato come $X = (\underline{q}_1, \dots, \underline{q}_N; \underline{p}_1, \dots, \underline{p}_N)$ siamo interessati alla distribuzione degli stati $\rho(X; t)$ e alla relativa evoluzione temporale data dal **Teorema di Liouville**:

$$\partial_t \rho(X; t) + \{\rho, \mathcal{H}\} = 0 \quad (2.4)$$

Siccome la distribuzione di Maxwell-Boltzmann è relativa alla singola particella $\rho(\underline{q}_1, \underline{p}_1; t)$, ottenuta marginalizzando la densità degli stati:

$$\rho_1(\underline{q}_1, \underline{p}_1; t) = \int \rho(X; t) \underbrace{dq_2 \dots dq_N}_{N-1} \underbrace{dp_2 \dots dp_N}_{N-1}$$

volendo imporre che sia soddisfatta l'evoluzione temporale per la singola particella si otterrebbe un sistema di equazioni non chiuso del tipo

$$\partial_t \rho_1 = \widehat{L}_1 \rho_1 + \widehat{L}_2 \rho_2$$

dove ρ_i è la densità di i particelle e \widehat{L}_i è un opportuno operatore lineare. Questa problematica va sotto il nome di **gerarchia BBGKY** ("Bogoliubov, Born, Green, Kirkwood, Yvon").

L'idea di Boltzmann è di lavorare in uno spazio ridotto $(\underline{q}, \underline{p}) \equiv z \in \mathbb{R}^6$ ed esprimere quindi una relazione per la probabilità di avere una particella in un certo stato:

$$f(\underline{q}, \underline{p}; t) \Delta q \Delta p \propto \text{n. di particelle nel volume } \Delta q \Delta p \text{ dello spazio delle fasi} \quad (2.5)$$

Cominciamo facendo vedere che la presenza di un potenziale d'interazione è fondamentale considerando il caso senza interazione

$$\mathcal{H} = \sum_{n=1}^N \frac{|p_n|^2}{2m} + U_{ext}(\underline{q}_n) \quad (2.6)$$

e ricordando le le equazioni di Hamilton

$$\begin{cases} \dot{\underline{q}} = \frac{\partial \mathcal{H}}{\partial \underline{p}_i} = \frac{\underline{p}}{m} \\ \dot{\underline{p}} = -\frac{\partial \mathcal{H}}{\partial \underline{q}_i} = -\partial_{\underline{q}} U_{ext}(\underline{q}_n) \end{cases} \quad (2.7)$$

che esprimiamo come $\dot{\underline{z}} = \underline{g}(\underline{z})$. Per il teorema di Liouville il volume nello spazio delle fasi si conserva e tutte le particelle che si trovavano in $V(t)$ finiscono in $V(t + \Delta t)$.

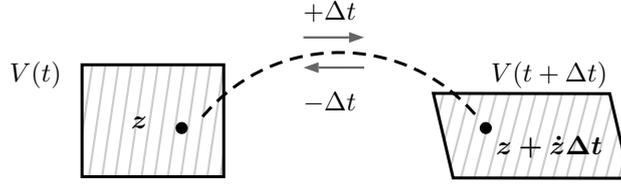


Figura 2.2: Conservazione del volume nello spazio delle fasi

Dall'eq. (2.5) otteniamo quindi la seguente relazione

$$f(z; t) \cancel{V(t)} = f(z + \dot{z}\Delta t; t + \Delta t) \cancel{V(t + \Delta t)} \quad (2.8)$$

da cui sviluppando il secondo membro

$$f(z + \dot{z}\Delta t; t + \Delta t) \simeq f(z; t) + \sum_{i=1}^6 \dot{z}_i \Delta t \frac{\partial f}{\partial z_i} + \partial_t f(z; t) \Delta t \quad (2.9)$$

si ricava sostanzialmente l'equazione di Liouville per una particella non interagente:

$$\partial_t f + (\dot{\underline{z}} \cdot \nabla_{\underline{z}}) f = 0 \quad (2.10)$$

Come precedentemente osservato il sistema non può termalizzare, infatti il **termine di streaming** $(\dot{\underline{z}} \cdot \nabla_{\underline{z}}) f$ semplicemente "sposta" la condizione iniziale per cui non è rilevante per la fisica che vogliamo studiare. Ci poniamo quindi nel caso in cui il termine di streaming è nullo ¹

$$\begin{cases} (\frac{\underline{p}}{m} \cdot \nabla_{\underline{q}}) f = 0 & \text{avendo scelto } f(z; t) \text{ omogeneo rispetto le } \underline{q} \\ (\nabla U_{ext} \cdot \nabla_{\underline{p}}) f = 0 & \text{avendo scelto } U_{ext} = 0 \end{cases} \quad (2.11)$$

Introduciamo quindi la possibilità di fare urti ($\partial_t f =$ "urti") e siccome il sistema è diluito possiamo trascurare la possibilità che si verificano urti con ≥ 3 particelle:

$$\mathcal{H} = \sum_{n=1}^N \left[\frac{|p_n|^2}{2m} + U_{ext}(\underline{q}_n) \right] + \sum_{n, n'} U_I^{(2)}(\underline{q}_n - \underline{q}_{n'}) \quad (2.12)$$

¹Ci stiamo ponendo in un sistema di riferimento in cui la forza e il potenziale esterni sono nulli.

Approcciamo il problema discretizzando gli stati (la trattazione continua è analoga e porta agli stessi risultati):

$$f(\underline{q}, \underline{p}; t) \longrightarrow f_i(t) = \text{probabilità di stare nello stato } i \quad (2.13)$$

Consideriamo che stare nello stato i -esimo corrisponda, ad esempio, avere una certa velocità: fisicamente il modello dell'interazione è dato dagli urti tra coppie di particelle come in fig. 2.3

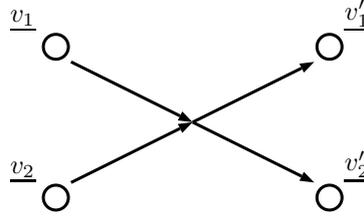


Figura 2.3: Schema degli urti tra 2 particelle

per cui dovranno essere rispettate le leggi di conservazione degli urti²

$$\begin{cases} \underline{v}_1 + \underline{v}_2 = \underline{v}'_1 + \underline{v}'_2 & \text{conservazione dell'impulso} \\ \underline{v}_1^2 + \underline{v}_2^2 = \underline{v}'_1^2 + \underline{v}'_2^2 & \text{conservazione dell'energia cinetica} \end{cases} \quad (2.14)$$

Affinché la probabilità di uno stato aumenti deve succedere che dopo l'urto una (o tutte e due le particelle) finisca nello stato di interesse. Siamo quindi interessati alla variazione nell'unità di tempo della probabilità $\frac{df_i(t)}{dt}$: sarà data dalla differenza tra i termini di *guadagno* G_i che esprimono la probabilità di transire nello stato i e i termini di *perdita* P_i :

$$G_i = \sum_{k,l,j} W_{(k,l) \rightarrow (i,j)} F_{(k,l)}^{(2)} \quad (2.15)$$

dove $W_{(k,l) \rightarrow (i,j)}$ è il rate con cui si transisce dallo stato (k, l) a (i, j) e dipende dalla meccanica del processo, mentre $F_{(k,l)}^{(2)}$ è la probabilità di avere lo stato iniziale della coppia (k, l) .

Viceversa per quanto riguarda la perdita

$$P_i = \sum_{k,l,j} W_{(i,j) \rightarrow (k,l)} F_{(i,j)}^{(2)} \quad (2.16)$$

Dall'invarianza temporale non c'è motivo fisico per cui l'urto avvenga in un verso o nell'altro ($(k, l) \rightarrow (i, j) = (i, j) \rightarrow (k, l)$) perciò si ha che

$$W_{(k,l) \rightarrow (i,j)} = W_{(i,j) \rightarrow (k,l)} \equiv W_{ijkl}$$

e l'eq. (2.16) diventa

$$\frac{df_i(t)}{dt} = G_i - P_i = \sum_{k,l,j} W_{ijkl} [F_{(k,l)}^{(2)} - F_{(i,j)}^{(2)}] \quad (2.17)$$

Tuttavia questa relazione non è chiusa (da una parte ho la probabilità di singola particella, dall'altra la probabilità di una coppia), il che richiede una grossa congettura, costituita dall'**Ipotesi del Caos Molecolare**³:

$$F_{(k,l)}^{(2)} = f_k(t) f_l(t) \quad (2.18)$$

²Si ricorda che le particelle hanno la stessa massa m .

³Il nome è storico e fuorviante dato che l'ipotesi non ha niente a che vedere con il concetto moderno di caos.

corrisponde all'assunzione che le probabilità siano fra loro scorrelate, il che è ovviamente falso, ma è anche il caso più semplice dal punto di vista del calcolo e permette di chiudere l'equazione:

$$\frac{df_i(t)}{dt} = \sum_{k,l,j} W_{ijkl} [f_k(t)f_l(t) - f_i(t)f_j(t)] \quad (2.19)$$

che nel caso continuo di due particelle che urtano, si esprimerebbe come

$$\partial_t f(\underline{v}_1; t) \propto \int (f'_1 f'_2 - f_1 f_2) \delta(\dots) d\underline{v}'_1 d\underline{v}'_2 d\underline{v}_2 d^2 b |v_1' - v_2'| \quad (2.20)$$

dove b è il parametro d'impatto dell'urto e $d^2 b |v_1' - v_2'|$ il flusso di particelle che attraversano l'elemento di volume $d^2 b$. Le leggi di conservazione sono imposte dalla δ :

$$\delta(\dots) = \delta(v_1 + v_2 - v_1' - v_2') \delta(v_1^2 + v_2^2 - v_1'^2 - v_2'^2)$$

L'obiettivo è mostrare che partendo dallo i -esimo stato al tempo $t = 0$ con probabilità $f_i(0)$, nel limite $t \rightarrow \infty$ il sistema evolve $f_i(t) \rightarrow f_i^{(\infty)} = f_{MB}$, cioè nel caso continuo:

$$f_{MB}(\underline{v}) = \sqrt{\frac{m\beta}{2\pi}} e^{-\beta \frac{m}{2} v^2} \quad (2.21)$$

Mostrare che, partendo da una distribuzione arbitraria si arriva dopo un tempo sufficientemente lungo ad una distribuzione di probabilità limite, ha in qualche modo a che vedere con il **secondo principio della termodinamica**, nel senso che si presenta un comportamento irreversibile.

2.3 Teorema H di Boltzmann

Teorema 2.3.1 — Teorema H di Boltzmann. Definito il funzionale $H[t] = \sum_i f_i(t) \ln f_i(t)$ e assumendo che sia valida l'eq. (2.19), allora

$$\frac{dH}{dt} \leq 0 \quad \text{dove} \quad \frac{dH}{dt} = 0 \Leftrightarrow f_i = f_{MB} \quad (2.22)$$

Si noti che che $H[t]$ è un'entropia a meno di un segno: $S = -k_B H[t]$, per cui il teorema equivale a dire che l'entropia aumenta fino ad un massimo quando la distribuzione è quella di Maxwell-Boltzmann.

Dimostrazione.

$$\frac{dH}{dt} = \sum_i \dot{f}_i \ln f_i + \sum_i f_i \frac{1}{f_i} \dot{f}_i = \sum_i \dot{f}_i \ln f_i$$

essendo $\sum_i \dot{f}_i = \frac{d}{dt} \sum_i f_i = 0$ poiché $\sum_i f_i = 1$. Ricordando che

$$\frac{df_i(t)}{dt} = \sum_{k,l,j} W_{ijkl} [f_k(t)f_l(t) - f_i(t)f_j(t)]$$

si ottiene che

$$\begin{aligned} \frac{dH}{dt} &= \sum_i \dot{f}_i \ln f_i = \sum_{i,j,k,l} W_{ijkl} [f_k(t)f_l(t) - f_i(t)f_j(t)] \ln f_i(t) = \\ &= \frac{1}{2} \sum_{i,j,k,l} W_{ijkl} [f_k(t)f_l(t) - f_i(t)f_j(t)] \ln [f_i(t)f_j(t)] \end{aligned}$$

dove abbiamo simmettrizzato $\ln f_i(t) \rightarrow \frac{1}{2} [\ln f_i(t) + \ln f_j(t)] = \frac{1}{2} \ln [f_i(t)f_j(t)]$, infatti gli indici sono muti. Analogamente possiamo simmettrizzare scambiando (k, l) con (i, j) poiché W_{ijkl} rimane invariato:

$$\begin{aligned} \frac{dH}{dt} &= \frac{1}{2} \left\{ \frac{1}{2} \sum_{i,j,k,l} W_{ijkl} [f_k f_l - f_i f_j] \ln [f_i f_j] + \frac{1}{2} \sum_{i,j,k,l} W_{ijkl} [f_i f_j - f_k f_l] \ln [f_k f_l] \right\} = \\ &= -\frac{1}{4} \sum_{i,j,k,l} W_{ijkl} [f_k f_l - f_i f_j] \{ \ln [f_k f_l] - \ln [f_i f_j] \} \end{aligned}$$

Osserviamo che data $g(x) = x \ln x$ con $x > 0$, allora $\frac{dg(x)}{dx} = \ln x + 1$ e $\frac{d^2g(x)}{dx^2} = \frac{1}{x}$ per cui $g(x)$ è convessa. Inoltre, data $h(x, y) = (x - y)(\ln x - \ln y)$ con $x, y > 0$, allora $h(x, y) > 0 \forall x, y$, infatti

1. se $x > y \Rightarrow \ln x > \ln y \Rightarrow h > 0$;
2. se $x < y \Rightarrow \ln x < \ln y \Rightarrow h > 0$;
3. se $x = y \Rightarrow h = 0$.

Ne concludiamo quindi che H è convessa e che quindi $\frac{dH}{dt} \leq 0$.

Poiché H un funzionale per $f_i(t)$, sono soddisfatte le ipotesi del **Teorema di stabilità Lyapunov** (riportato in calce) e di conseguenza se $\dot{f} = 0$ allora $\{f_i(t)\} \rightarrow f_i^{(\infty)}$. Siccome la soluzione stazionaria si ha quando

$$\dot{f} = \sum W [f_k f_l - f_i f_j] = 0$$

ed essendo $W = W_{(k,l) \rightarrow (i,j)} = W_{(i,j) \rightarrow (k,l)} \neq 0$, dalle leggi di conservazione (si veda eq. (2.14)) si ottiene che l'unica soluzione possibile è data dalla probabilità stazionaria \tilde{f}_i t.c.

$$\tilde{f}_k \tilde{f}_l = \tilde{f}_i \tilde{f}_j \quad (2.23)$$

dove gli indici k, l, i , e j non sono arbitrari ma selezionati dalla conservazione, ossia nel continuo

$$\tilde{f}(\underline{v}'_1) \tilde{f}(\underline{v}'_2) = \tilde{f}(\underline{v}_1) \tilde{f}(\underline{v}_2)$$

Passando infine ai logaritmi:

$$\ln \tilde{f}(\underline{v}'_1) + \ln \tilde{f}(\underline{v}'_2) = \ln \tilde{f}(\underline{v}_1) + \ln \tilde{f}(\underline{v}_2) \quad (2.24)$$

si ottiene quindi una legge di conservazione per i logaritmi che devono quindi dipendere dalle grandezze fisiche conservate del sistema (la massa, le 3 componenti di ciascun impulso e l'energia cinetica) e dovranno essere esprimibili come combinazione lineare di queste:

$$\ln \tilde{f}(\underline{v}) = cost + av_x + bv_y + cv_z + dv^2 \quad (2.25)$$

Sfruttando le proprietà di simmetria (invarianza per rotazione) e in assenza di una velocità media $\langle v_i \rangle = 0$ in ogni direzione, aggiustando opportunamente le costanti, si ottiene infine la distribuzione di Maxwell-Boltzmann

$$\tilde{f}(\underline{v}) = \mathcal{N} e^{-Av^2} = f_{MB} \quad (2.26)$$

Si tratta della soluzione dell'eq. (2.20) nel continuo. ■

Teorema 2.3.2 — Teorema di stabilità di Lyapunov. Sia dato il problema differenziale $\dot{\underline{x}} = \underline{f}(\underline{x})$ con $\underline{x}, \underline{f} \in \mathbb{R}^D$ e sia \underline{x}^* t.c. $\underline{f}(\underline{x}^*) = 0$, se \exists un funzionale $\mathcal{V}[\underline{x}]$ t.c. \underline{x}^* punto di min per \mathcal{V} ,

$$\text{se } \left. \frac{d\mathcal{V}}{dt} \right|_{\text{soluzione}} < 0 \text{ per } \underline{x} \neq \underline{x}^* \text{ e se } \left. \frac{d\mathcal{V}}{dt} \right|_{\text{soluzione}} = 0 \text{ per } \underline{x} = \underline{x}^*$$

allora si ha che

$$\underline{x}(t) \rightarrow \underline{x}^* \quad \text{per} \quad t \rightarrow \infty$$

2.3.1 Obiezioni al teorema H

Il teorema H non è esente da obiezioni, in particolare le principali sono il paradosso della reversibilità di Loschmidt e della ricorrenza di Zermelo:

1. Paradosso della ricorrenza di Zermelo.

Dal teorema di Poincaré, dato un sistema meccanico con spazio delle fasi limitato che evolve $X(0) \rightarrow X(t)$ allora $\exists \tau_R$ detto tempo di ritorno t.c. $X(\tau_R) \approx X(0)$. Questo implica che dopo un certo tempo il sistema deve tornare in prossimità della condizione iniziale: $H(\tau_R) \approx H(0)$ in contrasto con quanto stabilito dal teorema H. Tuttavia l'obiezione viene respinta da Boltzmann stesso con l'argomento che i tempi di attesa per sistemi di questo tipo sono ben superiori alla vita dell'universo, per cui non hanno interesse fisico.

2. Paradosso della reversibilità di Loschmidt.

Siccome H dipende da f_i , ossia dallo stato del sistema, nulla vieta di fermarsi in uno stato ad un dato tempo t^* ed invertire i parametri del sistema grazie alla reversibilità temporale $(\underline{q}(t^*), \underline{v}(t^*)) \rightarrow (\underline{q}(t^*), -\underline{v}(t^*))$. Il sistema evolve quindi al contrario fino al tempo 0, tuttavia lo stato al tempo t^* è equivalente a tutti gli altri stati per cui anch'esso dovrebbe evolvere fino all'equilibrio. Questa obiezione è di difficile risposta.

Per poter completare il collegamento tra la meccanica e la termodinamica c'è bisogno di superare le "buone obiezioni", risultato non banale portato avanti da una branca all'avanguardia della matematica con risultati in tempi moderni. Tuttavia i coniugi Ehrenfest fecero vedere che abbandonando il mondo deterministico a favore di uno stocastico, le obiezioni possono essere aggirate.

2.3.2 Considerazione sulla forma funzionale di H

Vogliamo approfondire alcune considerazioni circa la scelta della forma funzionale di H e dei collegamenti di questa grandezza con l'entropia. Ricordiamone dunque la definizione:

$$H[t] = \sum_i f_i(t) \ln f_i(t) \quad (2.27)$$

L'importanza di questo funzionale risiede nella dimostrazione che Boltzmann fece della decrescita monotona nel tempo e della convergenza ad un minimo corrispondente alla distribuzione di Maxwell-Boltzmann: $f_i^{max} = f_{MB}$. Risulta a questo punto naturale l'interpretazione a posteriori di questa grandezza con l'entropia fisica, a meno di una costante e di un segno, in particolare: $S = -k_B H[t]$. Osserviamo a questo punto tuttavia come la scelta della particolare forma di $H[t]$ non risulti tuttavia necessaria nella dimostrazione del Teorema H, in quanto l'applicazione del teorema di stabilità di Lyapunov richiede solo la concavità (o convessità, a seconda del segno con cui si lavora) del funzionale. La questione è tuttavia più profonda.

Dato un sistema descritto dalle coordinate generalizzate $\underline{x} \in \mathbb{R}^{6N}$ e descritto da una distribuzione di probabilità nello spazio delle fasi $\rho(\underline{x}, t)$, possiamo definire le due entropie:

$$S_G(t) = -k_B \int \rho(\underline{x}, t) \ln \rho(\underline{x}, t) d\underline{x} \quad (2.28)$$

$$S_B(t) = -k_B \int \rho_1(\underline{q}, \underline{p}, t) \ln \rho_1(\underline{q}, \underline{p}, t) d\underline{q} d\underline{p} \quad (2.29)$$

Dove con $\rho_1(\underline{q}, \underline{p}, t)$ si è indicata la distribuzione di probabilità della singola particella. È possibile mostrare come entrambe queste grandezze all'equilibrio si comportino come entropie (possiedono quindi

le necessarie proprietà di estensività), inoltre, per particelle non interagenti, all'equilibrio, risulta immediata la scomposizione: $\rho_N(\underline{x}) = \prod_n \rho_1(\underline{q}_n, \underline{p}_n)$ da cui l'importante relazione: $S_G^{eq} = N S_B^{eq}$. Il comportamento all'equilibrio risulta dunque sostanzialmente equivalente in quanto le due grandezze differiscono solamente per una costante di proporzionalità, tuttavia la differente natura delle due grandezze si palesa nello studio delle rispettive proprietà fuori dall'equilibrio. Dal Teorema H sappiamo già che S_B cresce in maniera monotona, mostriamo invece come S_G risulti costante nel tempo e come ciò sia diretta conseguenza del Teorema di Liouville.

Teorema 2.3.3 — Invarianza di S_G nel tempo. Dato un sistema Hamiltoniano descritto dalla funzione di distribuzione $\rho(\underline{x}, t)$ con $\underline{x} \in \mathbb{R}^{6N}$, si consideri la grandezza $S_G(t)$ definita dall'eq. (2.28), allora si ha:

$$S_G(t) = S_G(0) \quad (2.30)$$

Dimostrazione. Dato $\underline{x} \in \mathbb{R}^{6N}$ al tempo $t = 0$, consideriamo al tempo t la grandezza $\underline{y} = \mathcal{S}^t \underline{x}$, dove \mathcal{S}^t è l'operatore di evoluzione temporale fino al tempo t applicato al vettore \underline{x} . Poiché il sistema è Hamiltoniano, l'operatore di evoluzione risulterà unitario, varrà dunque la relazione di trasformazione delle probabilità:

$$\rho(\underline{x}, t = 0) d\underline{x} = \rho(\underline{y}, t) d\underline{y} \quad (2.31)$$

Dove con $d\underline{x}$ e $d\underline{y}$ si sono indicati gli elementi di volume nello spazio delle fasi rispettivamente al tempo $t = 0$ ed al tempo t . Il teorema di Liouville afferma che lo jacobiano della trasformazione temporale è unitario, varrà dunque: $d\underline{x} = d\underline{y} \forall t$. Da questa relazione e dall'uguaglianza eq. (2.31) discende immediatamente: $\rho(\underline{x}, t = 0) = \rho(\underline{y}, t)$.

È dunque valida la seguente catena di uguaglianze $\forall t$:

$$S_G(t) = - \int \rho(\underline{y}, t) \ln \rho(\underline{y}, t) d\underline{y} = - \int \rho(\underline{x}, 0) \ln \rho(\underline{y}, t) d\underline{x} = \quad (2.32)$$

$$= - \int \rho(\underline{x}, t = 0) \ln \rho(\underline{x}, t = 0) d\underline{x} = S_G(t = 0) \quad (2.33)$$

■

Questo teorema dunque mina profondamente l'utilità fisica della grandezza S_G la cui costanza nel tempo non rispecchia dunque l'idea termodinamica dell'aumento costante di entropia codificata nel Secondo Principio, ma non solo: osserviamo come S_G dipenda da $\rho(\underline{x}, t)$ una quantità dipendente da variabili in uno spazio $6N$ dimensionale, con $N \simeq \mathcal{N}_A$. Sostanzialmente dunque è una grandezza sperimentalmente (come anche numericamente) non osservabile. Viceversa S_B dipende da $\rho_1(\underline{q}, \underline{p}, t)$, dunque da uno spazio a 6 dimensioni, certamente più maneggevole ma soprattutto simulabile numericamente.

2.4 Il paradosso della ricorrenza

Affrontiamo il paradosso di Zermelo: la decrescita monotona di H (e dunque dell'aumento monotono di S_B) appare in contrasto con il teorema di ricorrenza di Poincaré.

L'argomentazione di risposta, avanzata dallo stesso Boltzmann, fu semplice: una stima del tempo $\langle \tau_R \rangle$ di ritorno ad una configurazione iniziale $X(0)$ è esponenziale nel numero N di particelle, dunque di gran lunga superiore alla vita dell'universo. Il paradosso dunque, seppur concettualmente corretto, era sostanzialmente di scarso interesse fisico.

La formalizzazione matematica della stima del tempo medio di ritorno $\langle \tau_R \rangle$ è data dal Lemma di Mark Kac, ne forniamo dunque una dimostrazione per il caso di un sistema il cui spazio delle configurazioni sia discreto.

Teorema 2.4.1 — Lemma di Kac. Sia \mathcal{S} un sistema ergodico definito in uno spazio delle configurazioni limitato e caratterizzato dalla distribuzione di probabilità $\rho_{\mathcal{S}}(\underline{x})$ con $\underline{x}(t=0) \in A$. Il tempo medio di ritorno $\langle \tau_R \rangle_A$ alla configurazione iniziale vale:

$$\langle \tau_r \rangle = \frac{\tau_0}{\mathcal{P}(A)} \quad (2.34)$$

Dove $\mathcal{P}(A)$ è la probabilità che il sistema si trovi in una configurazione appartenente all'insieme A ed è data da:

$$\mathcal{P}(A) = \int_A \rho_{\mathcal{S}}(\underline{x}) d\underline{x} \quad (2.35)$$

Dimostrazione. Supponiamo il sistema possa trovarsi in una delle configurazioni

$$\mathcal{S} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_M\}$$

ed evolva secondo una qualche regola in maniera tale che ad ogni tempo t sia definita la relazione:

$$\mathcal{S}_t \longrightarrow \mathcal{S}_{t+1} \quad ; \quad \mathcal{S}_0 = \mathcal{C}_k \quad (2.36)$$

Denominiamo con τ_n il tempo di primo ritorno dopo un tempo trascorso τ_{n-1} , in particolare τ_1 sarà dunque il tempo di primo ritorno, τ_2 il tempo di primo ritorno dopo un tempo trascorso pari a τ_1 , e così via. Varranno dunque le relazioni:

$$\begin{aligned} \mathcal{S}_{\tau_1} &= \mathcal{C}_k \\ \mathcal{S}_{\tau_1+\tau_2} &= \mathcal{C}_k \\ &\vdots \\ \mathcal{S}_{\tau_1+\tau_2+\dots+\tau_n} &= \mathcal{C}_k \end{aligned}$$

Avremo dunque una collezione di tempi di ritorno $\tau = \{\tau_1, \tau_2, \dots, \tau_n\}$. Possiamo valutare il tempo totale trascorso e la frequenza del tempo passata nella configurazione \mathcal{C}_k :

$$t_{\mathcal{N}} = \sum_{n=1}^{\mathcal{N}} \tau_n \quad ; \quad f_{\mathcal{C}_k}(t_{\mathcal{N}}) = \frac{\mathcal{N}}{t_{\mathcal{N}}} = \frac{1}{\frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \tau_n} \quad \forall \mathcal{N} \quad (2.37)$$

Possiamo a questo punto valutare il limite per $\lim \mathcal{N} \rightarrow +\infty$. Sfruttando le ipotesi di ergodicità del sistema considerato otteniamo i due comportamenti limite:

$$f_{\mathcal{C}_k} \longrightarrow \mathcal{P}(\mathcal{C}_k) \quad (2.38)$$

$$\frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \tau_n \longrightarrow \langle \tau_R \rangle_{\mathcal{C}_k} \quad (2.39)$$

Invertendo la seconda delle relazioni eq. (2.37) otteniamo dunque la tesi:

$$\langle \tau_R \rangle_{\mathcal{C}_k} = \frac{1}{\mathcal{P}(\mathcal{C}_k)} \quad (2.40)$$

■

Possiamo a questo punto applicare questo risultato per ottenere la stima esponenziale nel numero N ottenuta da Boltzmann del tempo medio di ritorno.

Consideriamo un sistema a $6N$ dimensioni ed assumiamo per semplicità che le dimensioni tipiche dello spazio in opportune unità adimensionali siano $\mathcal{O}(L)$. All'interno di questo sistema consideriamo dunque un sotto-spazio A di raggio ε . Se assumiamo per semplicità la funzione di distribuzione costante (ad esempio per l'ensemble microcanonico $\rho = \text{cost}$) su tutto lo spazio otteniamo facilmente una stima per la probabilità che il sistema si trovi in una configurazione $\mathcal{C}_k \in A$ come semplice rapporto delle misure degli insiemi:

$$\mathcal{P}(A) = \alpha \left(\frac{\varepsilon}{L} \right)^{6N} \quad (2.41)$$

Dove α è una opportuna costante. Applicando il Lemma di Kac (eq. (2.34)) otteniamo per il tempo medio di ritorno la stima:

$$\langle \tau_R \rangle_A = \tau_0 \left(\frac{L}{\varepsilon} \right)^{6N} \quad (2.42)$$

Dove τ_0 è una qualsiasi scala di tempo il cui valore risulta irrilevante: anche scegliendo la scala di Planck, la più piccola scala temporale fisicamente concepibile e dell'ordine $\tau_0 = \mathcal{O}(10^{-44} \text{s})$, essa risulta irrisoriamente sovrastata dall'ordine del doppio esponenziale che figura nella formula eq. (2.42) e stimabile dell'ordine $\mathcal{O}(c^{10^{23}})$ con $c > 1$. Appara dunque chiaro come anche per un numero sorprendentemente ridotto di particelle (è sufficiente $N \sim 50$) la stima del tempo medio di ritorno risulti talmente elevata da cancellare l'apparente paradosso della ricorrenza.

Per superare veramente le difficoltà dell'unione fra la meccanica, dunque la descrizione al livello microscopico, e la termodinamica, dunque la descrizione al livello macroscopico, Paul Ehrenfest e la moglie Tatiana suggerirono di abbandonare una descrizione deterministica a favore una stocastica, iniziamo per questo una parentesi sui principali processi stocastici e sulle principali tecniche matematiche che ci occorreranno in futuro.

Riferimenti bibliografici

- [1] M. Falcioni e A. Vulpiani. *Meccanica Statistica Elementare*. Springer-Verlag, 2014.
- [2] M. Kardar. *Statistical Physics of Particles*. Cambridge University Press, 2007.

3. Processi markoviani

3.1 Le catene di Markov

Introdotte dal matematico russo Andrej Andreevič Markov per una disputa circa la regolarità statistica dei comportamenti sociali con il matematico moscovita Nekrasov, le Catene di Markov sono un importante processo stocastico dalle numerose e sorprendenti applicazioni, forniamone dunque una prima definizione.

Definizione 3.1 — Processo markoviano. Si definisce processo stocastico markoviano un processo aleatorio nel quale la probabilità di transizione che determina il passaggio ad uno stato del sistema dipende solo dallo stato del sistema immediatamente precedente (proprietà di Markov) e non dal come si è giunti a tale stato, vi è dunque assenza di memoria.

Una catena di Markov può essere dunque formalizzata come un processo markoviano con uno spazio degli stati discreto.

La condizione di assenza di memoria (molto frequente in fisica, ad esempio in tutti i processi che possono essere descritti mediante un'equazione differenziale) può essere formalizzata nella relazione:

$$\begin{aligned} \mathcal{P}\left(\mathcal{S}(t_{n+1}) = c_{n+1} \mid \mathcal{S}(t_n) = c_n, \mathcal{S}(t_{n-1}) = c_{n-1}, \dots, \mathcal{S}(t_0) = c_0\right) = \\ = \mathcal{P}\left(\mathcal{S}(t_{n+1}) = c_{n+1} \mid \mathcal{S}(t_n) = c_n\right) \end{aligned} \quad (3.1)$$

Questa relazione afferma come la probabilità che il sistema al tempo t_{n+1} si trovi nella configurazione c_{n+1} data la sua intera storia, dunque la sequenza dei tempi e delle configurazioni corrispondenti, sia in realtà equivalente alla probabilità che si trovi al tempo t_{n+1} nella configurazione c_{n+1} dato solamente la configurazione c_n al tempo immediatamente precedente. È dunque sufficiente l'informazione dello stato presente per poter inferire la probabilità dello stato futuro, non vi è dunque bisogno di conoscere l'intero sviluppo della storia, vi è quindi assenza di memoria.

3.1.1 Invarianza per numero finito di memorie

È lecito domandarsi se la scelta di considerare solamente lo stato presente non risulti eccessivamente restrittiva o cosa accade se permetto di avere memoria per più di uno stato precedente, o per un numero qualunque seppur finito. La risposta è che di fatto nulla cambia, a patto di estendere la definizione di stato \mathcal{S} . Supponiamo di voler considerare:

$$\mathcal{P}(\mathcal{S}(t_{n+1}) = c_{n+1} \mid \mathcal{S}(t_n) = c_n, \mathcal{S}(t_{n-1}) = c_{n-1}) \quad (3.2)$$

Dunque un sistema con una memoria di due stati precedenti, una sorta di catena di Markov al secondo ordine. Introduco la nuova variabile: $\sigma_t = (\mathcal{S}_t, \mathcal{S}_{t-1})$, la probabilità considerata si riscrive dunque come:

$$\mathcal{P}(\sigma_{t+1} = (c_{n+1}, c_n) \mid \sigma_t = (c_n, c_{n-1})) \quad (3.3)$$

Abbiamo dunque di fatto recuperato la proprietà eq. (3.1). Questo ragionamento può chiaramente essere esteso per un arbitrario, ma finito, numero di stati k , a patto di espandere la definizione dello stato stesso.

Di seguito studieremo il caso in cui le probabilità siano temporalmente omogenee, cioè indipendenti dal tempo. Poniamo dunque:

$$\mathcal{P}(\mathcal{S}(t_{n+1}) = i \mid \mathcal{S}(t_n) = k) = \mathcal{P}_{k \rightarrow i} \quad (3.4)$$

Osserviamo a questo punto come un sistema di M stati sia descritto dall'insieme delle $\mathcal{P}_{k \rightarrow i}$, una matrice $M \times M$, detta matrice di transizione, che contiene di fatto tutte le informazioni sul sistema stessa e di fatto definisca il comportamento della Catena di Markov.

Elenchiamo dunque alcune proprietà immediatamente deducibili dalla natura delle catene di Markov. Studiamo come dalle probabilità al tempo t_n sia possibile valutare le probabilità al tempo successivo t_{n+1} . Indichiamo con $\mathcal{P}_j(t_n)$ la probabilità che il sistema si trovi nello stato j -esimo al tempo t_n . Possiamo dunque scrivere la relazione:

$$\mathcal{P}_j(t_{n+1}) = \sum_k \mathcal{P}_{k \rightarrow j} \mathcal{P}_k(t_n) = \sum_k W_{jk} \mathcal{P}_k(t_n) \quad (3.5)$$

Dove si è indicato per brevità $W_{jk} \equiv \mathcal{P}_{k \rightarrow j}$. Questa relazione non è altro che l'intuitiva riscrittura del concetto: la probabilità di trovarsi nello stato j al tempo t è la somma su tutti gli stati intermedi k della probabilità di transire da k a j per la probabilità di trovarsi al tempo $t - 1$ nello stato k .

Osserviamo a questo punto come il vettore $\underline{P}(t + 1)$ delle probabilità al tempo t_{n+1} si ottenga dal vettore delle probabilità al tempo precedente mediante l'applicazione di un operatore lineare:

$$\underline{P}(t + 1) = \hat{\mathcal{W}} \underline{P}(t) \quad (3.6)$$

L'iterazione di questa relazione implica immediatamente:

$$\underline{P}(t) = [\hat{\mathcal{W}}]^t \underline{P}(0) \quad (3.7)$$

Possiamo inoltre esprimere la probabilità $\mathcal{P}_{i \rightarrow j}^{(n)}$ di transire dallo stato i allo stato j in n passi sfruttando la relazione:

$$\mathcal{P}_{i \rightarrow j}^{(n)} = \left([\hat{\mathcal{W}}]^n \right)_{ji} \quad (3.8)$$

Possiamo mostrarne facilmente la validità nel caso $n = 2$, l'estensione al caso n generico è immediato. Consideriamo:

$$\mathcal{P}_{i \rightarrow j}^{(2)} = \sum_k \mathcal{P}_{i \rightarrow k} \mathcal{P}_{k \rightarrow j} = \sum_k W_{ki} W_{jk} = \sum_k W_{jk} W_{ki} = \left([\hat{W}]^2 \right)_{ji} \quad (3.9)$$

Osserviamo come la matrice di transizione $W_{ij} \equiv \mathcal{P}_{j \rightarrow i}$ non possa essere completamente arbitraria per via della natura stessa dei suoi elementi che rappresentano probabilità. In generale avremo dunque che $W_{ij} \in [0, 1]$ con $W_{ij} \neq W_{ji}$, inoltre, poiché consideriamo nella matrice stessa anche il caso in cui il sistema inizialmente nello stato j permanga in j , e poiché qualcosa deve pur avvenire, deve valere la relazione:

$$\sum_i W_{ij} = 1 \quad (3.10)$$

Data la forma dell'evoluzione del vettore delle probabilità degli stati definita dall'espressione eq. (3.7), possiamo domandarci se esistano e sono unici particolari vettori $\underline{P}^{(inv)}$ aventi la proprietà di risultare invarianti per applicazione dell'operatore di transizione \hat{W} , in formule:

$$\underline{P}^{(inv)} = \hat{W} \underline{P}^{(inv)} \quad (3.11)$$

Inoltre, dato un generico vettore iniziale di probabilità, risulterà questo asintoticamente uguale ad un vettore di probabilità invariante? Sottolineiamo come queste domande siano fortemente collegate con gli interrogativi di Boltzmann circa l'evoluzione delle distribuzioni di probabilità verso un equilibrio.

3.1.2 Matrice di transizione

La risposta a questi interrogativi è nello studio delle proprietà della matrice W_{ij} di transizione. Esiste un teorema che garantisce quale sia una condizione necessaria e sufficiente per l'esistenza e l'unicità delle $\mathcal{P}^{(inv)}$, per la convergenza di un generico vettore iniziale e l'ergodicità del sistema markoviano.

Teorema 3.1.1 — Teorema sulla natura delle Catene di Markov. Data una Catena di Markov definita dalla matrice di transizione \hat{W} , se

$$\exists n : \left([\hat{W}]^n \right)_{ji} > 0 \quad \forall (i, j)$$

allora le seguenti condizioni sono equivalenti:

- $\exists! \mathcal{P}^{(inv)}$;
- $\underline{P}(t) \xrightarrow{t \rightarrow +\infty} \underline{P}^{(inv)} + \mathcal{O}(e^{-\alpha t})$ con la costante α valutabile;
- la Catena di Markov è ergodica;

Sfruttando questo risultato è possibile definire le seguenti relazioni:

$$\langle A \rangle = \sum_{n=1}^M A_n \mathcal{P}_n^{(inv)} \quad ; \quad \frac{1}{T} \sum_{t=1}^T A_{S_t} \xrightarrow{T \rightarrow +\infty} \bar{A} \quad ; \quad \langle A \rangle = \bar{A} \quad (3.12)$$

Dove la prima rappresenta un valor medio valutato sull'ensemble, la seconda definisce invece una media sulla dinamica del sistema e l'ultima relazione sfrutta l'ergodicità del sistema per concludere che queste due grandezze sono uguali. Queste relazioni sono la base del funzionamento degli algoritmi Montecarlo: data una $\mathcal{P}^{(inv)}$ tento di costruirmi una Catena di Markov avente $\mathcal{P}^{(inv)}$ come distribuzione asintotica e valuto i valori medi delle grandezze di interesse come media sulla dinamica d'evoluzione del sistema. Non sempre le catene di Markov possiedono la proprietà necessaria per l'applicazione del teorema enunciato, ad esempio è possibile mostrare come tale teorema non valga in un random walk su un anello con un numero di stati pari.

Studiamo per semplicità il caso della catena di Markov più semplice ma non banale (il caso banale è quello degli eventi indipendenti in cui $\mathcal{P}_{i \rightarrow j} = \mathcal{P}_j$).

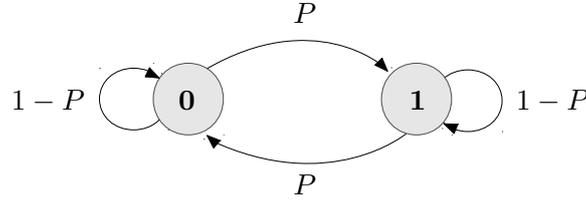


Figura 3.1: Schema di una Catena di Markov a due stati.

Per questo sistema possiamo scrivere la legge di evoluzione per uno dei due stati, ad esempio $\mathcal{P}_0(t)$ in quanto $\mathcal{P}_1(t)$ si ottiene banalmente come: $\mathcal{P}_1(t) = 1 - \mathcal{P}_0(t)$. Abbiamo:

$$\begin{aligned} \mathcal{P}_0(t+1) &= \mathcal{P}_0(t)\mathcal{P}_{0 \rightarrow 0} + \mathcal{P}_1(t)\mathcal{P}_{1 \rightarrow 0} = \\ &= \mathcal{P}_0(t)\mathcal{P}_{0 \rightarrow 0} + (1 - \mathcal{P}_0(t))\mathcal{P}_{1 \rightarrow 0} = \\ &= (1 - P)\mathcal{P}_0(t) + (1 - \mathcal{P}_0(t))P = \\ &= (1 - 2P)\mathcal{P}_0(t) + P \end{aligned}$$

Osserviamo a questo punto come quest'ultima relazione sia formalmente equivalente a quella ottenuta nello studio del sistema d'evoluzione random di Lorentz, possiamo dunque sfruttarne i risultati e concludere che:

$$\mathcal{P}_0^{(inv)} = \frac{1}{2} \quad (3.13)$$

Possiamo inoltre valutare la velocità con cui la probabilità converge a questo valore asintotico sfruttando la decomposizione: $\mathcal{P}_0(t) = \frac{1}{2} + \Delta(t)$. Sostituendo ottengo per $\Delta(t)$ l'espressione:

$$\Delta(t+1) = (1 - 2P)\Delta(t) \quad (3.14)$$

Da cui, risolvendo l'iterazione, otteniamo:

$$\Delta(t) = (1 - 2P)^t \Delta(0) \quad (3.15)$$

Abbiamo dunque ottenuto la forma dell'andamento di $\mathcal{P}_0(t)$ nel tempo, valutando esplicitamente il valore della costante α che figurava nell'enunciato del teorema:

$$\mathcal{P}_0(t) = \frac{1}{2} + e^{\ln(1-2P)t} \Delta(0) \quad (3.16)$$

Nei nostri studi ci interesseranno le catene di Markov che possiedono specifiche proprietà rappresentanti le proprietà e le simmetrie dei sistemi che vorremo andare a modellizzare. La principale fra queste simmetrie è l'inversione temporale che, tradotta nel contesto delle Catene di Markov, è rappresentata dalla relazione del Bilancio Dettagliato; questa risulterà però una proprietà d'equilibrio, mentre noi siamo maggiormente interessati alle dinamiche fuori dall'equilibrio. Cercheremo dunque di studiare cosa accade quando questa proprietà viene a mancare.

3.2 Convergenza della probabilità

Avevamo visto che grazie al teorema sulla natura delle catene di Markov che si applicava alle catene di Markov aperiodiche e irriducibili (dunque tali che $(\hat{W}^n)_{ij} > 0$ per qualche n) potevamo esser sicuri della convergenza della probabilità ad una probabilità invariante, convergenza che avviene con una correzione esponenzialmente piccola:

$$\underline{P}(t) \rightarrow \underline{P}^{inv} + \mathcal{O}(e^{-\frac{t}{\tau}})$$

Tuttavia questo approccio estremamente teorico lascia amaro il palato degli sperimentali che vorrebbero calcolare l'effettiva probabilità di transizione, per far ciò introduciamo le funzioni di correlazione:

Definizione 3.2 — funzione di correlazione. data una successione di stati $i_0, i_1, i_2, \dots, i_T$ definiamo la funzione di correlazione fra due grandezze f e g

$$\langle f(t)g(0) \rangle = \frac{1}{T} \sum_{l=1}^T g_{i_l} f_{i_{l+t}}$$

Se il sistema è ergodico si ha che nel limite per $T \rightarrow \infty$ tale valor medio sul tempo tende al valor medio calcolato come la somma su tutti i possibili stati di partenza e di arrivo della probabilità di spostamento per il prodotto delle funzioni f e g calcolate rispettivamente nello stato finale e iniziale:

$$\langle f(t)g(0) \rangle = \frac{1}{T} \sum_{l=1}^T g_{i_l} f_{i_{l+t}} \quad \rightarrow \quad \sum_{i,j} P(i, 0, j, t) g_i f_j$$

$P(i, 0, j, t)$ è la probabilità di trovarsi al tempo iniziale in i e al tempo t in j allora non è altro che la probabilità di trovarsi in i per la probabilità di transizione da i a j , per la probabilità di trovarsi in i scegliamo quella invariante d'altronde se il sistema è ergodico la definizione di funzione di correlazione è buona se la media per T grandi coincide con quella fatta sugli stati:

$$\sum_{i,j} P_i^{inv} P_{i \rightarrow j}(t) g_i f_j$$

la $P_{i \rightarrow j}(t)$ possiamo stimarla come $P_j^{inv} + \mathcal{O}(e^{-\frac{t}{\tau}})$ avremo infine per $T \rightarrow \infty$:

$$\langle f(t)g(0) \rangle = \frac{1}{T} \sum_{l=1}^T g_{i_l} f_{i_{l+t}} \quad \rightarrow \quad \sum_i \sum_j P_i^{inv} P_j^{inv} g_i f_j + \mathcal{O}(e^{-\frac{t}{\tau}}) = \langle f \rangle \langle g \rangle + \mathcal{O}(e^{-\frac{t}{\tau}})$$

si osserva dunque che le funzioni di correlazione convergono al valore asintotico con la stessa rate delle probabilità di transizione. Questo risultato fornisce agli sperimentali un modo per ricavare i tempi di rilassamento del sistema misurando la funzione di correlazione e sottraendo il valore asintotico si ha una stima dell'esponenziale e dunque di τ . Noi teorici amanti dei calcoli vorremmo tuttavia calcolare τ con certezza, per fortuna alcuni risultati matematici ci permettono di calcolarlo esplicitamente in alcuni casi fortunati. Innanzitutto possiamo dire che se esiste la probabilità invariante allora uno degli autovalori della matrice di transizione deve necessariamente essere 1 e a tale autovalore è ovviamente associata la probabilità invariante

$$\alpha_0 = 1 \quad \text{t.c.} \quad \hat{W} \underline{P}^{inv} = \alpha_0 \underline{P}^{inv}$$

Gli altri autovalori saranno

$$|\alpha_n| < 1 \quad \text{t.c.} \quad \hat{W} \underline{v}_n = \alpha_n \underline{v}_n$$

questo perché? Trascurando il caso in cui $\alpha_0 = 1$ sia degenere e assumendo che gli autovettori formino una base completa di modo che possiamo scrivere:

$$\underline{P}_0 = \sum_n c_n \underline{v}_n$$

bisogna ovviamente star attenti che la probabilità iniziale non diventi negativa. Calcolando la probabilità al tempo t si avrà:

$$\underline{P}(t) = \hat{W}^t \underline{P}(0) = c_0 \underline{P}^{inv} + \sum_{n \geq 1} \alpha_n^t c_n \underline{v}_n$$

nel limite per $t \rightarrow \infty$ sappiamo che $\underline{P}(t)$ deve tendere a \underline{P}^{inv} da ciò ne risulterà che $c_0 = 1$ e $|\alpha_n| < 1$. Abbiamo assunto che l'autovalore principale non sia degenere ad assicurarcelo è il teorema di Perron-Frobenius.

Teorema 3.2.1 — Teorema di Perron-Frobenius. Data una matrice $A_{ij} > 0$ ordinando gli autovalori $|\alpha_0| \geq |\alpha_1| \geq \dots$ si ha che

- α_0 non è degenere;
- $|\alpha_1| < 1$

questo teorema ci assicura dunque che gli autovalori (tranne quello associato alla probabilità invariante) si trovano all'interno della circonferenza unitaria.

Questo teorema apparentemente di basso interesse fisico si traduce in meccanica statistica in: dato un sistema sul reticolo unidimensionale con interazione a range finito esso non presenta transizioni di fase. In termini di processi markoviani invece afferma che per le catene di Markov aperiodiche e irriducibili la convergenza della probabilità a quella invariante avviene esponenzialmente.

Con il teorema di Perron-Frobenius applicato alla \hat{W}^n siamo ora certi che le catene interessanti (quelle che convergono) sono quelle che rispettano la condizione $(\hat{W}^n)_{ij} > 0$. Dai risultati ottenuti possiamo ricavare τ , si ha infatti

$$\underline{P}(t) \rightarrow \underline{P}^{inv} + c_1 \alpha_1^t \underline{v}_1 + \dots$$

con α_1 autovalore più vicino al bordo della sfera unitaria, gli altri termini andranno a 0 ancora più velocemente perché gli autovalori sono appunto ordinati; riscrivendo

$$\alpha_1^t = e^{-t |\ln(\alpha_1)|}$$

si trova immediatamente che

$$\tau = \frac{1}{|\ln(\alpha_1)|}$$

3.2.1 Bilancio dettagliato

Come era stato anticipato le catene di Markov che rispettano il bilancio dettagliato sono di particolare interesse fisico.

Definizione 3.3 — condizione di bilancio dettagliato. La catena di Markov rispetta la condizione di bilancio dettagliato se $\forall(i, j)$ vale che

$$P_i^{inv} P_{i \rightarrow j} = P_j^{inv} P_{j \rightarrow i}$$

interpretando entrambi i membri come delle correnti di probabilità astratte tale condizione suggerisce che la corrente di probabilità nell'andare da i a j è uguale a quella relativa al percorso inverso.

Il concetto di bilancio dettagliato è strettamente legato alla reversibilità; se consideriamo una particella caratterizzata dalle coordinate nello spazio delle fasi al tempo iniziale $(\underline{q}(0), \underline{p}(0))$ facendo evolvere la particella fino al tempo T e a tal punto invertendo il moto si ha che facendo evolvere $(\underline{q}(T), -\underline{p}(T))$ per

un altro tempo T si riottiene la condizione iniziale con il segno della velocità scambiato. Un'analogia proprietà per le catene de Markov è la seguente :

$$P_{i_0}^{inv} P_{i_0 \rightarrow i_1} \dots P_{i_{T-1} \rightarrow i_T} = P_{i_T}^{inv} P_{i_T \rightarrow i_{T-1}} \dots P_{i_1 \rightarrow i_0}$$

la probabilità di andare da i_0 a i_T passando per i vari stati intermedi è dunque uguale alla probabilità di fare il percorso contrario se la catena di Markov rispetta il bilancio dettagliato. Le situazioni in cui vale il bilancio dettagliato sono di interesse per la meccanica statistica dell'equilibrio, nel nostro caso la situazione tipica è quella in cui tale proprietà non è verificata; basti pensare alla corrente che scorre in un filo, è decisamente poco probabile che la corrente inverta il suo moto.

3.3 Modello degli Ehrenfest

Gli Ehrenfest scrissero un libricino in cui introdussero un semplice modello che essendo matematicamente risolvibile permetteva di rispondere facilmente alle obiezioni avanzate da Zermelo e Loschmidt al teorema H di Boltzmann. Boltzmann come abbiamo visto aveva provato a dare risposta alle obiezioni ma le sue motivazioni seppur esatte sembravano poco consistenti. Il fondamento dell'approccio degli Ehrenfest è l'uso di processi stocastici, dunque di un'interpretazione probabilistica della fisica che all'epoca era ancora difficile da digerire.

Si hanno due cani A e B ed N pulci; sul cane A vi sono n pulci e sul B ovviamente $N-n$, le pulci sono numerate, si estrae dunque un numero da 1 a N e, a seconda del cane sul quale si trova, la pulce corrispondente a tale numero salta con probabilità (fig. 3.2):

$$\begin{cases} P_{n \rightarrow n-1} = \frac{n}{N} \\ P_{n \rightarrow n+1} = 1 - \frac{n}{N} \end{cases} \quad (3.17)$$

riconosciamo dunque una catena di Markov. Gli Ehrenfest ebbero l'intuizione che tale modello riuscisse a descrivere la situazione analizzata da Boltzmann.

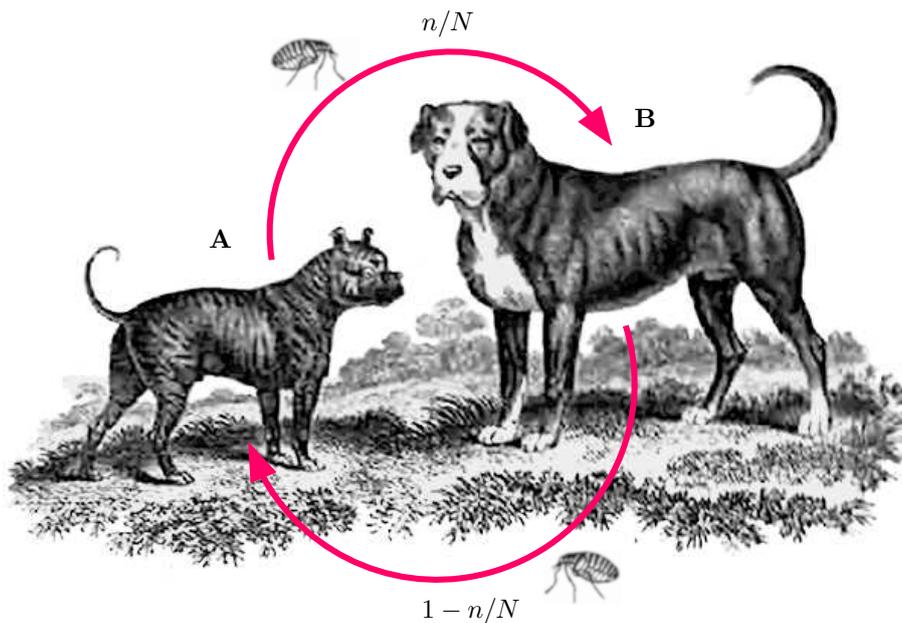


Figura 3.2: Modello cani e pulci.

Ragionevolmente poiché è chiaro dalle probabilità che la situazione con un numero di pulci su A nettamente maggiore di quello su B è sfavorita allora ci aspettiamo che dopo un certo periodo le pulci su A siano $\frac{N}{2}$, ovviamente fluttuazioni potranno sempre essere presenti. Chiaramente se abbiamo una situazione iniziale in cui n_0 è circa N la probabilità per qualsiasi pulce estratta di cambiare scatola è alta: è un comportamento diffusivo delle pulci. Siamo interessati dunque a studiare il passaggio dalla situazione fortemente sbilanciata a quella asintotica, lo si può fare molto semplicemente andando a calcolare il valor medio $\langle n_t \rangle$ e la rispettiva varianza e si vedrà quanto atteso cioè che il primo tende a $\frac{N}{2}$ e la varianza cresce fino ad un valore limite.

Sorge un problema di difficile comprensione (per i chimici) : come relazionare i dati empirici al modello stocastico? Per rispondere a tale domanda risolviamo il modello degli Ehrenfest e ricaviamone le grandezze di interesse. Si ha come legge di evoluzione per n_t la seguente regola di ricorsione:

$$n_{t+1} = n_t + \Delta(t)$$

in cui

$$\Delta(t) = \begin{cases} -1 & \text{con probabilità } \frac{n_t}{N} \\ +1 & \text{con probabilità } 1 - \frac{n_t}{N} \end{cases} \quad (3.18)$$

vorremmo calcolare il valor medio $\langle \Delta(t) \rangle$ per far ciò ricordiamo che, data una funzione $f(x, y)$ di due variabili non indipendenti fra loro, il valor medio di questa funzione può essere come:

$$\langle f(x, y) \rangle = \int f(x, y) P(x, y) dx dy$$

Ma la probabilità composta può essere riscritta come $P(x, y) = P(x|y)P(y)$, allora integrando su x :

$$\langle f(x, y) \rangle = \int \langle f|y \rangle P(y) dy$$

in cui $\langle f|y \rangle$ è il valor medio di f condizionato da y .

Possiamo quindi calcolare il valor medio di $\langle \Delta(t) \rangle$ nei due successivi passaggi :

$$\begin{aligned} \langle \Delta(t) | n_t \rangle &= (-1) \frac{n_t}{N} + (+1) \left(1 - \frac{n_t}{N}\right) \\ \langle \Delta(t) \rangle &= 1 - 2 \frac{\langle n_t \rangle}{N} \end{aligned}$$

da qui dunque :

$$\langle n_{t+1} \rangle = \left(1 - \frac{2}{N}\right) \langle n_t \rangle + 1$$

è ormai evidente che $\langle n_t \rangle$ tende a $\frac{N}{2}$ (lo si dimostra con il metodo grafico utilizzato nelle lezioni passate). Possiamo stabilire in quanto tempo si raggiunge questo limite definendo

$$\delta_t = \langle n_t \rangle - \langle n_\infty \rangle$$

e utilizzando la relazione precedente si ottiene $\delta_{t+1} = \left(1 - \frac{2}{N}\right) \delta_t$ da cui iterando:

$$\delta_t = \left(1 - \frac{2}{N}\right)^t \delta_0$$

allora esplicitando $\langle n_t \rangle$:

$$\langle n_t \rangle = \frac{N}{2} + \left(1 - \frac{2}{N}\right)^t \left(n_0 - \frac{N}{2}\right)$$

Il tempo di convergenza si ricava dal secondo termine a secondo membro della precedente relazione:

$$\left(1 - \frac{2}{N}\right)^t = e^{-t|\ln(1-\frac{2}{N})|} \sim e^{-\frac{2t}{N}} \text{ per } N \gg 1$$

è chiaro dunque che il sistema tende al valor medio in un tempo dell'ordine $\frac{N}{2}$. Per quanto riguarda la varianza, si può dimostrare che cresce fino a raggiungere un valore asintotico pari a $\sigma_\infty^2 = \frac{N}{4}$.

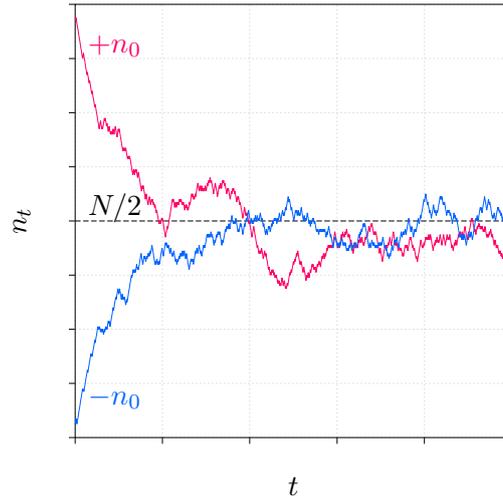


Figura 3.3: Facendo partire un sistema da una condizione iniziale $+n_0$ (o del tutto equivalentemente da $-n_0$), osserviamo come questo si raggiunga la condizione di equilibrio.

Abbiamo dunque completato l'analisi del modello e siamo pronti a mostrare come esso dia una risposta convincente alle obiezioni che sono state poste a Boltzmann.

3.3.1 Risposta all'obiezione di Zermelo

Da tale modello ne risulta che la distribuzione di probabilità invariante è la binomiale

$$P_n^{inv} = \frac{N!}{n!(N-n)!} 2^{-N}$$

per la quale vale il bilancio dettagliato che è la proprietà stocastica per l'invarianza temporale

$$P_n^{inv} P_{n \rightarrow n \pm 1} = P_{n \pm 1}^{inv} P_{n \pm 1 \rightarrow n}$$

per rispondere all'obiezione di Zermelo si può calcolare il tempo di ricorrenza (usando il lemma di Kac) nota la condizione iniziale n_0 , si trova che:

- $n_0 = N$ si ha $\langle \tau_R \rangle \sim 2^N$;
- $n_0 \sim \frac{N}{2}$ si ha $\langle \tau_R \rangle \sim \sqrt{N}$.

quindi se il sistema parte vicino all'equilibrio, il tempo di un possibile ritorno è ragionevole; se invece si parte lontano dall'equilibrio, il tempo di ritorno è esorbitante anche per poche particelle.

3.3.2 Risposta all'obiezione di Loschmidt

Vediamo ora come rispondere all'obiezione sollevata da Loschmidt sulla reversibilità. Il modello degli Ehrenfest vale per ogni valore di N tuttavia siamo interessati al caso in cui il numero di particelle (o pulci) è grande, in particolar modo siamo interessati alla singola realizzazione cioè a:

$$y_t = \frac{n_t}{N}$$

Teorema 3.3.1 — Disuguaglianza di Chebyshev. La disuguaglianza di Chebyshev afferma che

$$Prob(|y - \langle x \rangle| > \epsilon) \leq \frac{\sigma_y^2}{\epsilon^2}$$

applicando la disuguaglianza di Chebyshev alla singola realizzazione si ottiene che

$$Prob\left(\left|\frac{n_t}{N} - \frac{\langle n_t \rangle}{N}\right| > \epsilon\right) \leq \frac{\sigma_{n_t}^2}{N^2 \epsilon^2}$$

scegliendo dunque $\epsilon = N^{-\frac{1}{4}}$ ed utilizzando il valore massimo della varianza si ottiene:

$$Prob\left(\left|\frac{n_t}{N} - \frac{\langle n_t \rangle}{N}\right| > N^{-\frac{1}{4}}\right) \leq \frac{\sigma_{n_t}^2}{N^2 \epsilon^2} \leq \frac{1}{4\sqrt{N}}$$

dunque la probabilità che la singola realizzazione sia molto lontana dal valor medio se N è molto grande risulta essere molto piccola; ci aspettiamo dunque che se si conduce una singola misura in laboratorio il risultato sia circa la media a meno di casi sfortunati e decisamente molto poco probabili in cui il sistema oscilla.

■ **Esempio 3.1** Consideriamo il caso in cui partendo dalla condizione iniziale n_0 il sistema scende, al passo successivo risale, al passo successivo scende ancora e così via, ma la probabilità che ciò avvenga è esponenzialmente piccola:

$$\left[\frac{n_0}{N} \left(1 - \frac{n_0}{N}\right)\right]^t$$

Si introduce dunque il concetto di *tipicità*, ossia per la maggior parte delle condizioni iniziali la singola realizzazione y_t è vicina al valor medio, cioè il valor medio è ciò che ci si aspetta quando si effettua una misura.

Ritornando al paradosso della reversibilità sollevato da Loschmidt, Boltzmann aveva argomentato che se per un qualsiasi tempo la H poteva ritornare indietro doveva esser vero anche al tempo iniziale osservando così intorno a tale punto un picco. Allora, siccome la H non è una curva regolare ma è “piena di punte”, se le punte in su sono in maggioranza, allora il teorema H è vero.

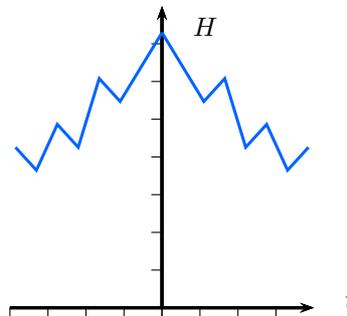


Figura 3.4: Inversione temporale di H all'istante iniziale.

Ovviamente l'argomento esposto in tal modo è molto debole; ci chiediamo dunque se nel modello degli Ehrenfest un tale comportamento sia formalmente giustificato. La richiesta di Boltzmann è che la situazione sia la “A” tra la seguenti:

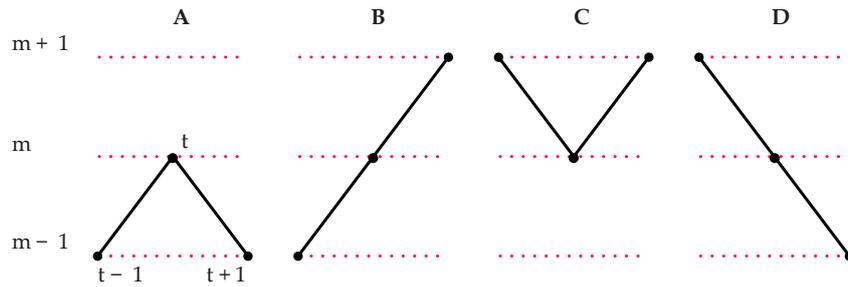


Figura 3.5: situazioni possibili nell'evoluzione di n_t .

cioè la condizione richiesta è: $A = \{n_{t-1} = m - 1, n_t = m, n_{t+1} = m - 1\}$.

La probabilità di tale situazione è:

$$P(A|n_t = m) = \frac{P_{m-1}^{inv} P_{m-1 \rightarrow m} P_{m \rightarrow m-1}}{P_m^{inv}} = \left(\frac{m}{N}\right)^2$$

che tende a 1 se $m \rightarrow N$: se dunque il sistema è fuori dall'equilibrio la probabilità di formazione di punte in su è molto alta ed è maggiore delle altre situazioni. Dunque per tale modello Boltzmann aveva ragione, ma è vero per un gas diluito? Nella seconda metà del '900 vari fisici e matematici¹ dimostrarono il seguente risultato:

Teorema 3.3.2 Dato un sistema con N particelle di raggio σ e massa m , nel **limite di Grad-Boltzmann**:

$$\begin{cases} N \rightarrow \infty \\ \sigma \rightarrow 0 \\ Nm \rightarrow \text{costante} \\ N\sigma^2 \rightarrow \text{costante} \end{cases} \quad (3.19)$$

cioè nel limite in cui le particelle diventano sempre di più ma la massa totale del sistema si conserva così come le dimensioni dello stesso (il tempo medio di collisione delle particelle rimane costante) e se la condizione iniziale è tale che:

$$x(0) \in A \text{ t.c. } \mu(A) \rightarrow 1 \text{ per } N \rightarrow \infty$$

allora la distribuzione di Boltzmann $f_B(\underline{q}, \underline{v}, t)$ è prossima alla vera distribuzione del sistema fino ad un tempo t^* che è pari ad una frazione del tempo di collisione.

Questo teorema dunque afferma che la soluzione dell'equazione di Boltzmann è vicina alla funzione vera fino ad un tempo che è una frazione del tempo di collisione, ma questo solo se la condizione iniziale rispetta alcune proprietà. Questo come ci aiuta a risolvere il problema della reversibilità? Partendo da una "condizione buona" e facendo evolvere il sistema per un tempo che supera quello tipico delle collisioni, la funzione predetta da Boltzmann non corrisponderà più a quella vera ed invertendo il sistema, questo "risalirà" perché la nuova condizione di partenza non è più "buona". In questo ci viene in aiuto la componente casuale, perché se invertiamo esattamente H crescerà, ma se c'è una piccola

¹Hanno partecipato a questo importante risultato, basato su un lavoro di Grad e l'intuizione di Boltzmann, fra i tanti, Illner, Lanford, Shinbrot, DiPerna, Lions, Pulvirenti e Cercignani.

imprecisione iniziale finiamo in quell'insieme di condizioni iniziali per cui vale nuovamente il teorema di Grad (vedi fig. 3.6).

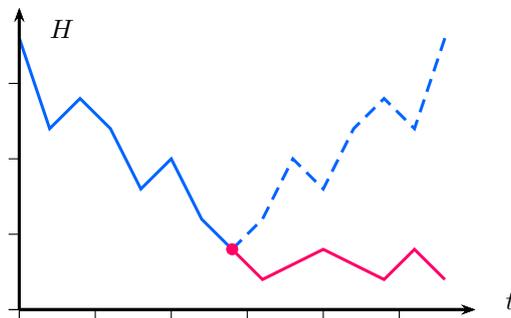


Figura 3.6: Intervento dell'elemento casuale nell'evoluzione di H . La curva azzurra rappresenta H che si inverte in un determinato istante: se l'inversione temporale è esatta si ottiene la curva tratteggiata; inserendo un errore di inversione si ottiene invece la curva magenta.

Il modello degli Ehrenfest contiene tutte queste informazioni e ci fa capire che ciò che è fisicamente importante è che la singola realizzazione non si distanzi troppo dalla media e questo avviene con probabilità vicina ad uno nel caso in cui il numero di particelle è elevato.

Riferimenti bibliografici

- [1] G. Boffetta e A. Vulpiani. *Probabilità in Fisica*. Springer-Verlag, 2012.

4. Master equation e Fokker-Plack

4.1 Master equation

Prendiamo ora in esame i processi stocastici in cui gli stati sono discreti ma il tempo è continuo, $t \in \mathbb{R}$. Per tali processi l'evoluzione della probabilità è data dalla cosiddetta **master equation**.¹ La master equation trova applicazione, ad esempio, nello studio delle dinamiche di popolazioni, dove il carattere discreto è associato al numero degli individui, o delle reazioni chimiche, in cui la variabile discreta rappresenta il numero di particelle.

Al fine di ricavare la master equation, consideriamo una catena di Markov (che ha tempo discreto) nel limite di un intervallo di tempo piccolo, $\Delta t \rightarrow 0$. La probabilità di andare dallo stato j al tempo t nello stato i al tempo $t + \Delta t$ è data da

$$P_i(t + \Delta t) = \sum_j W_{j \rightarrow i}(\Delta t) P_j(t) \quad (4.1)$$

Nel limite di intervalli di tempo piccoli, la probabilità di non cambiare stato è molto maggiore della probabilità di cambiarlo e quindi possiamo assumere che la matrice di transizione sia della forma

$$W_{j \rightarrow i}(\Delta t) = \begin{cases} T_{ij} \Delta t + O(\Delta t^2) & \text{se } i \neq j \\ 1 - \kappa_i \Delta t + O(\Delta t^2) & \text{se } i = j \end{cases} \quad (4.2)$$

La matrice T_{ij} rappresenta il tasso di transizione dallo stato j allo stato i per unità di tempo. Scrivendo la condizione $\sum_i W_{j \rightarrow i} = 1$ come somma dei due contributi

$$\sum_i W_{j \rightarrow i} = \sum_{i \neq j} W_{j \rightarrow i} + W_{j \rightarrow j} = \sum_{i \neq j} T_{ij} \Delta t + 1 - \kappa_j \Delta t = 1 \quad (4.3)$$

si ricava

$$\kappa_j = \sum_{i \neq j} T_{ij} \quad (4.4)$$

¹Il nome *master equation* comparve per la prima volta in un articolo del 1940 (A. Nordsieck *et al.*, Physica 7, 344, 1940) in cui aveva il ruolo di equazione generale da cui discendevano tutti gli altri risultati. Da allora è rimasta in uso tale terminologia nella letteratura scientifica.

Sostituendo nella eq. (4.1) la $W_{j \rightarrow i}$ data dalla eq. (4.2) e sottraendo ad ambo i membri $P_i(t)$ si ha

$$P_i(t + \Delta t) - P_i(t) = \sum_{j \neq i} T_{ij} P_j(t) \Delta t + (1 - \kappa_i \Delta t) P_i(t) - P_i(t) \quad (4.5)$$

Dividendo per Δt e passando al limite $\Delta t \rightarrow 0$ otteniamo

$$\frac{dP_i(t)}{dt} = \sum_{j \neq i} T_{ij} P_j(t) - \kappa_i P_i(t) \quad (4.6)$$

utilizzando la eq. (4.4), notando che ora la somma è su j e quindi bisogna invertire gli indici, si ricava la **master equation**

$$\frac{dP_i(t)}{dt} = \sum_{j \neq i} (T_{ij} P_j(t) - T_{ji} P_i(t)) \quad (4.7)$$

Il primo termine rappresenta il tasso di transizione da tutti gli stati $j \neq i$ allo stato i mentre il secondo rappresenta il tasso di uscita dallo stato i .

L'evoluzione della probabilità descritta dalla master equation è una relazione lineare e ciò permette di ricavare facilmente la probabilità invariante (è un semplice esercizio di algebra lineare).

Osservazione. Il fatto che la master equation sia un sistema lineare costituisce un notevole vantaggio rispetto all'equazione di Boltzmann che invece è bilineare. Anche le catene di Markov e, come vedremo, i processi stocastici a stati e tempi continui (regolati dall'equazione di Fokker-Planck) sono descritti da operatori lineari, e quindi questo è il grande pregio dei processi stocastici, quello di avere una legge lineare di evoluzione della probabilità.

Anche nel caso dei processi stocastici rappresentati da una master equation è possibile definire la condizione di **bilancio dettagliato** una volta trovata la probabilità invariante P_i^{inv} , che assume la seguente espressione

$$P_i^{\text{inv}} P_{i \rightarrow j}(t) = P_j^{\text{inv}} P_{j \rightarrow i}(t) \quad (4.8)$$

In realtà la eq. (4.8) è ridondante perché è sufficiente che valga per un tempo Δt affinché sia sempre verificata, ovvero deve essere verificata la seguente condizione

$$T_{ji} P_i^{\text{inv}} = T_{ij} P_j^{\text{inv}} \quad (4.9)$$

4.1.1 Teorema H per la master equation

Esiste un teorema H anche per i processi descritti da master equation. In particolare questo consente di dare un significato fisico al rilassamento della distribuzione di probabilità alla probabilità invariante legandola all'entropia del sistema.

Teorema 4.1.1 — Teorema H per le master equation. Definita la grandezza $H(t)$ come

$$H(t) = \sum_i P_i(t) \ln \left(\frac{P_i(t)}{P_i^{\text{inv}}} \right) \quad (4.10)$$

dove P_i soddisfa una master equation (eq. (4.7)) con probabilità invariante P_i^{inv} . Allora si ottiene che

$$\frac{dH}{dt} \leq 0$$

dove l'uguaglianza vale quando $P_i = P_i^{\text{inv}}$.

Dimostrazione. Dimostriamo il teorema nel caso caso in cui vale il bilancio dettagliato. Per non appesantire la notazione nel seguito non esplicitiamo la dipendenza temporale indicando $P_i(t)$ semplicemente con P_i . Scegliamo una funzione $f(x)$ limitata inferiormente, definita sull'intervallo $0 \leq x < \infty$ e *convessa*, ovvero $f''(x) > 0$, così definita

$$f(x) = x \ln x \quad (4.11)$$

da cui

$$H = \sum_i P_i \ln \frac{P_i}{P_i^{\text{inv}}} = \sum_i P_i^{\text{inv}} f\left(\frac{P_i}{P_i^{\text{inv}}}\right) = \sum_i P_i^{\text{inv}} f(x_i) \quad (4.12)$$

dove $x_i \equiv \frac{P_i}{P_i^{\text{inv}}}$. Deriviamo H rispetto al tempo

$$\frac{dH}{dt} = \sum_i P_i^{\text{inv}} f'(x_i) \frac{dP_i}{dt} \frac{1}{P_i^{\text{inv}}} \quad (4.13)$$

utilizzando la [eq. \(4.7\)](#) si ottiene

$$\frac{dH}{dt} = \sum_{ij} \left[f'(x_i) T_{ij} P_j - f'(x_i) T_{ji} P_i \right] = \sum_{ij} \left[f'(x_i) T_{ij} P_j - f'(x_j) T_{ij} P_j \right] \quad (4.14)$$

dove nell'ultimo passaggio si sono scambiati gli indici muti nel secondo termine della sommatoria su i e j . Ora usiamo $P_j = P_j^{\text{inv}} x_j$ ottenendo

$$\frac{dH}{dt} = \sum_{ij} T_{ij} P_j^{\text{inv}} \left(f'(x_i) x_j - f'(x_j) x_j \right) \quad (4.15)$$

Dalla condizione di bilancio dettagliato $T_{ji} P_i^{\text{inv}} = T_{ij} P_j^{\text{inv}}$ si ricava per un'arbitraria successione numerica ψ_i

$$\sum_{ij} T_{ij} P_j^{\text{inv}} \psi_i = \sum_{ij} T_{ji} P_i^{\text{inv}} \psi_i = \sum_{ij} T_{ij} P_j^{\text{inv}} \psi_j \quad (4.16)$$

dove nell'ultimo passaggio si sono scambiati gli indici i e j . Dall'uguaglianza tra il primo e l'ultimo membro si ottiene l'identità

$$\sum_{ij} T_{ij} P_j^{\text{inv}} (\psi_i - \psi_j) = 0 \quad (4.17)$$

Scegliamo $\psi_i = f(x_i) - x_i f'(x_i)$ e sommiamo l'identità che ne deriva alla [eq. \(4.15\)](#) ottenendo

$$\begin{aligned} \frac{dH}{dt} &= \sum_{ij} T_{ij} P_j^{\text{inv}} \left(f'(x_i) x_j - f'(x_j) x_j + f(x_i) - x_i f'(x_i) - f(x_j) + x_j f'(x_j) \right) \\ &= \sum_{ij} T_{ij} P_j^{\text{inv}} \left[(x_j - x_i) f'(x_i) + f(x_i) - f(x_j) \right] \end{aligned} \quad (4.18)$$

È evidente che essendo $f(x)$ una funzione convessa il fattore all'interno delle parentesi quadre è negativo a meno che $x_i = x_j$. La conclusione è che $H(t)$ decresce *monotonicamente*. Poiché però $H(t)$ è limitata inferiormente ² si deduce che deve tendere a un limite in corrispondenza del quale la [eq. \(4.18\)](#) si annulla. Ciò è possibile solamente se $x_i = x_j$ per ogni coppia di stati i, j per i quali $T_{ij} \neq 0$, da cui dunque si ottiene:

$$\frac{P_i}{P_i^{\text{inv}}} = \frac{P_j}{P_j^{\text{inv}}} \implies P_i \propto P_i^{\text{inv}} \quad (4.19)$$

che, normalizzando, conduce alla conclusione che la funzione H è minima per $P = P^{\text{inv}}$. ■

²Il limite inferiore di $H(t)$ è dimostrabile attraverso i moltiplicatori di Lagrange con la condizione di normalizzazione.

In conclusione, il teorema H per le master equations, oltre a garantire la convergenza della probabilità alla probabilità stazionaria (che è possibile dimostrare già tramite l'algebra lineare), fornisce anche informazioni sul modo in cui avviene la convergenza, ovvero in modo monotono. Rispetto al teorema H dell'equazione di Boltzmann, il teorema H appena dimostrato risulta meno potente poiché richiede la conoscenza della P_i^{inv} .

4.2 Equazione di Fokker-Planck

Mettiamoci adesso nel caso di un processo markoviano in cui sia i tempi che gli stati sono continui. Per un processo markoviano si può scrivere la probabilità

$$P_i(t+1) = \sum_j P_j(t) P_{j \rightarrow i}$$

Se gli stati sono continui, invece della probabilità si ha una densità di probabilità. Per semplicità facciamo tutto in una dimensione. Chiamando x e y due generici stati del sistema, la precedente relazione si può riscrivere

$$P(x, t + \Delta t) = \int P(y, t) W(y, t \rightarrow x, t + \Delta t) dy \quad (4.20)$$

Possiamo definire

$$\Delta x(\Delta t) = z = x - y$$

Mostreremo ora che, sotto le condizioni

- $\Delta t \rightarrow 0$
- $\Delta x = z$ “varia poco” (in un senso che capiremo meglio a breve)

dall'eq. (4.20) si può ricavare l'equazione di Fokker-Planck:

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial [a(x)P(x, t)]}{\partial x} + \frac{1}{2} \frac{\partial^2 [b(x)P(x, t)]}{\partial x^2} \quad (4.21)$$

Questa equazione è lineare, può cioè essere riscritta nella forma

$$\frac{\partial P(x, t)}{\partial t} = \mathcal{L}P(x, t)$$

dove \mathcal{L} è un operatore differenziale lineare:

$$\mathcal{L}f(x) = -\frac{\partial [a(x)f(x, t)]}{\partial x} + \frac{1}{2} \frac{\partial^2 [b(x)f(x, t)]}{\partial x^2}$$

Sottolineiamo come tutti i processi markoviani descritti fino ad ora sono descritti da equazioni lineari:

MARKOV CHAIN

$$P(t+1) = \widehat{W}P(t)$$

MASTER EQUATION

$$\frac{\partial P}{\partial t} = \widehat{A}P$$

FOKKER-PLANCK

$$\frac{\partial P}{\partial t} = \mathcal{L}P$$

a differenza dell'equazione di Boltzmann che è bilineare.

4.2.1 Derivazione dell'equazione

Consideriamo dunque la eq. (4.20). Assumiamo innanzitutto che W dipenda unicamente dal punto di partenza, dalla variazione z e dal salto nel tempo:

$$W(y, t \rightarrow x, t + \Delta t) = \psi(x - z, z, \Delta t)$$

Quindi la eq. (4.20) può essere riscritta come

$$P(x, t + \Delta t) = \int P(x - z, t) \psi(x - z, z, \Delta t) dz$$

Considerato che ci siamo messi nelle condizioni in cui z è piccolo, possiamo sviluppare $P(x - z, z, \Delta t)$ e $\psi(x - z, z, \Delta t)$ rispetto al primo argomento, intorno ad x :

$$\begin{aligned} P(x, t + \Delta t) &= \int P(x - z, t) \psi(x - z, z, \Delta t) dz = \\ &= \int P(x, t) \psi dz - \frac{\partial}{\partial x} \int P(x, t) \psi z dz + \frac{1}{2} \frac{\partial^2}{\partial x^2} \int P(x, t) \psi z^2 - \dots = \\ &= P(x, t) \int \psi dz - \frac{\partial}{\partial x} \left[P(x, t) \int \psi z dz \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[P(x, t) \int \psi z^2 dz \right] - \dots \end{aligned}$$

dove per snellire la notazione abbiamo ommesso le dipendenze di $\psi = \psi(x, z, \Delta t)$. Notiamo che l'integrale nel primo termine è uguale a 1 (è la probabilità di fare un salto z , integrata su tutti gli z possibili), l'integrale nel secondo termine è uguale al valore di aspettazione, condizionato alla partenza in x , di z (cioè Δx , che a sua volta dipende da Δt), quello nel terzo termine al valore di aspettazione, condizionato alla partenza in x , di z^2 e così via:

$$P(x, t + \Delta t) = P(x, t) - \frac{\partial}{\partial x} [P(x, t) E(\Delta x | x)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [P(x, t) E(\Delta x^2 | x)] - \dots$$

Essendo Δt piccolo, possiamo anche sviluppare al primo ordine rispetto a Δt :

$$P(x, t + \Delta t) = P(x, t) + \frac{\partial}{\partial t} P(x, t) \Delta t$$

e uguagliando le due espressioni trovate, otteniamo

$$\frac{\partial}{\partial t} P(x, t) \Delta t = - \frac{\partial}{\partial x} [P(x, t) E(\Delta x(\Delta t) | x)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [P(x, t) E(\Delta x^2(\Delta t) | x)] - \dots$$

Ora ipotizziamo che

$$\lim_{\Delta t \rightarrow 0} \frac{E(\Delta x | x)}{\Delta t} = a(x) \quad (4.22)$$

$$\lim_{\Delta t \rightarrow 0} \frac{E(\Delta x^2 | x)}{\Delta t} = b(x) \quad (4.23)$$

$$\lim_{\Delta t \rightarrow 0} \frac{E(\Delta x^n | x)}{\Delta t} = 0 \text{ per } n > 2, n \in N \quad (4.24)$$

Osserviamo che questo in realtà non è altro se non un modo formale di affermare che Δx è piccolo se Δt è piccolo. Sotto queste ipotesi troviamo, infine, l'equazione di Fokker-Planck:

$$\frac{\partial P(x, t)}{\partial t} = - \frac{\partial (a(x) P(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2 (b(x) P(x, t))}{\partial x^2}$$

L'equazione è anche chiamata, in ambito matematico, "forward Kolmogorov equation". Il termine *forward* si riferisce al fatto che in questo caso l'equazione descrive un sistema di cui si conosce lo stato ad un tempo precedente rispetto ai tempi di evoluzione considerati (ne esiste una versione *backward* con segni differenti). Esistono versioni generalizzate (note come **equazione di Kramers-Moyal**) dell'equazione di Fokker-Planck, che si ricavano ipotizzando medie non nulle fino a ordini superiori al secondo, e vengono utilizzate soprattutto in ambito chimico.

4.3 Fokker-Planck e l'equazione di Langevin

Esiste una situazione fisicamente realistica in cui le assunzioni fatte sono vere? Sì: tutti i sistemi descritti da un'equazione di Langevin, che è un'equazione differenziale stocastica (mentre la Fokker-Planck è una equazione alle derivate parziali):

$$\frac{dx}{dt} = a(x) + \sqrt{b(x)}\eta(t) \quad (4.25)$$

dove $\eta(t)$ è un processo gaussiano tale che $\langle \eta \rangle = 0$ e $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$. Un altro modo, utilizzato più che altro in ambito matematico, di scrivere l'equazione di Langevin è

$$dx = a(x)dt + \sqrt{b(x)}dW$$

dove W è una variabile gaussiana, tale che $\langle W \rangle = 0$ e $\langle W(t)W(t') \rangle = \min\{t, t'\}$. La relazione (formale)

$$W(t) = \int_0^t \eta(t')dt'$$

lega le due scritte, la prima delle quali è quella più utilizzata dai fisici. Mostriamo adesso che le ipotesi eq. (4.22), eq. (4.23) e eq. (4.24) sono effettivamente verificate per un sistema descritto dall'equazione eq. (4.25). Lo facciamo nel caso, più semplice, in cui $b(x) = b$ è una costante (anche se in effetti lasciare la dipendenza in x non crea grandi problemi, Ndr). Cominciamo col verificare la prima ipotesi.

$$\Delta x(\Delta t) = x(t + \Delta t) - x(t) = \int_t^{t+\Delta t} a(x(t'))dt' + \sqrt{b} \int_t^{t+\Delta t} \eta(t')dt'$$

Vogliamo il valore medio di $\Delta(x)$ condizionato al fatto che il punto di partenza sia x . Ricordando che $\eta(t)$ ha media nulla, e notando che, poiché il punto di partenza x è fissato, $a(x)$ va fuori dall'integrale³, si ha

$$\begin{aligned} E(\Delta x|x) &= \int_t^{t+\Delta t} a(x(t'))dt' = a(x(t))\Delta t \\ \implies \lim_{\Delta t \rightarrow 0} \frac{E(\Delta x|x)}{\Delta t} &= a(x) \end{aligned}$$

Dunque la prima ipotesi è verificata. Anche la seconda lo è, infatti si ha

$$\Delta x^2 = \left(\int_t^{t+\Delta t} a(x(t'))dt' \right)^2 + b \iint_t^{t+\Delta t} \eta(t')\eta(t'')dt'dt'' + 2\sqrt{b} \int_t^{t+\Delta t} a(x(t'))dt' \int_t^{t+\Delta t} \eta(t'')dt''$$

Ma, mediando, il termine $\eta(t')\eta(t'')$ dà una $\delta(t - t')$, quindi rimane uno solo dei due integrali, mentre il terzo integrale è nullo perché lo è la media di η . In definitiva

$$\begin{aligned} E(\Delta x^2|x) &= a^2(x)\Delta t^2 + b\Delta t \\ \implies \lim_{\Delta t \rightarrow 0} \frac{E(\Delta x^2|x)}{\Delta t} &= \lim_{\Delta t \rightarrow 0} a^2(x)\Delta t + b = b \end{aligned}$$

Si mostra facilmente che i termini $\lim_{\Delta t \rightarrow 0} \frac{E(\Delta x^n|x)}{\Delta t}$ sono nulli per $n > 2$.

³o è un'approssimazione per Δt piccolo o qualcos'altro che mi sfugge?

■ **Esempio 4.1** Facciamo il calcolo esplicito per $n = 3$:

$$\begin{aligned}\Delta x^3 &= \left(\int_t^{t+\Delta t} a(x(t')) dt' \right)^3 + \\ &+ b\sqrt{b} \iiint_t^{t+\Delta t} \eta(t_1)\eta(t_2)\eta(t_3) dt_1 dt_2 dt_3 + \\ &+ 3\sqrt{b} \left(\int_t^{t+\Delta t} a(x(t')) dt' \right)^2 \int_t^{t+\Delta t} \eta(t') dt' + \\ &+ 3b \int_t^{t+\Delta t} a(x(t')) dt' \int_t^{t+\Delta t} \eta(t')\eta(t'') dt' dt''\end{aligned}$$

da cui

$$\begin{aligned}E(\Delta x^3|x) &= a^3(x)\Delta t^3 + 3ba(x)\Delta t^2 \\ \implies \lim_{\Delta t \rightarrow 0} \frac{E(\Delta x^3|x)}{\Delta t} &= 0\end{aligned}$$

■

Quindi esiste almeno un processo stocastico, quello di Langevin, a cui è possibile associare l'equazione di Fokker-Planck, che è un'equazione alle derivate parziali. In fluidodinamica ci sono due punti di vista, Euleriano (che studia i campi di velocità) e Lagrangiano (che segue le singole particelle nel loro moto). In un certo senso, le equazioni di Fokker-Planck e Langevin sono l'analogo, rispettivamente, del punto di vista Euleriano e di quello Lagrangiano: due modi diversi di studiare lo stesso sistema.

4.3.1 Fokker-Planck e l'equazione di diffusione

Nel caso banale in cui $a(x)$ (detto termine di “drift”) è identicamente nullo e b è una costante, l'equazione di Fokker-Planck è semplicemente l'**equazione di diffusione**:

$$\frac{\partial P(x, t)}{\partial t} = \frac{b}{2} \frac{\partial^2 P(x, t)}{\partial x^2} = D \frac{\partial^2 P(x, t)}{\partial x^2} \quad (4.26)$$

Mostreremo ora che l'equazione di diffusione e quella di Fokker-Planck si possono entrambe ricavare da un *random-walk* discreto, facendo un opportuno limite al continuo. Consideriamo per ora un random-walk discreto in cui la probabilità che il camminatore vada a destra o a sinistra è uniforme, uguale a $1/2$. Se chiamiamo Δx il salto spaziale che il camminatore fa ad ogni passo, si ha che

$$P(n\Delta x, t + \Delta t) = \frac{1}{2}P(n\Delta x + \Delta x, t) + \frac{1}{2}P(n\Delta x - \Delta x, t)$$

da cui

$$P(n\Delta x, t + \Delta t) - P(n\Delta x, t) = \frac{1}{2} [P(n\Delta x + \Delta x, t) + P(n\Delta x - \Delta x, t) - 2P(n\Delta x, t)]$$

Dividendo tutto per Δt e moltiplicando e dividendo, a destra, per Δx^2 , si ottiene

$$\frac{P(n\Delta x, t + \Delta t) - P(n\Delta x, t)}{\Delta t} = \frac{\Delta x^2}{2\Delta t^2} \frac{[P(n\Delta x + \Delta x, t) + P(n\Delta x - \Delta x, t) - 2P(n\Delta x, t)]}{\Delta x^2}$$

Se adesso facciamo tendere Δt e Δx a 0 in modo che

$$\lim_{\Delta t \rightarrow 0, \Delta x \rightarrow 0} \frac{\Delta x^2}{\Delta t} = 2D$$

otteniamo, riconoscendo nel termine a destra una differenza centrata, proprio l'equazione di diffusione. Quindi l'equazione di diffusione si ottiene dal random-walk facendo un limite al continuo, sia nel tempo che nello spazio, in cui però Δt è dell'ordine di Δx^2 .

Come si fa ad ottenere il termine di drift dal modello random-walk? Basta supporre che la probabilità di andare a destra e a sinistra non sia uniforme:

$$\begin{aligned} P((n-1)\Delta x \rightarrow n\Delta x) &= \frac{1}{2} + f_{n-1}(\Delta x) \\ P((n+1)\Delta x \rightarrow n\Delta x) &= \frac{1}{2} - f_{n+1}(\Delta x) \end{aligned}$$

dove f_n è una funzione che dipende dal punto di partenza e da Δx . In questo caso si ha

$$P(n\Delta x, t + \Delta t) = \left(\frac{1}{2} - f_{n+1}\right)P((n+1)\Delta x, t) + \left(\frac{1}{2} + f_{n-1}\right)P((n-1)\Delta x, t)$$

Il contributo che deriva dai termini che moltiplicano $1/2$ è esattamente quello del caso uniforme. Per quanto riguarda i termini con f , dividiamo per Δt e moltiplichiamo e dividiamo per Δx :

$$\frac{f_{n+1}P((n+1)\Delta x, t) - f_{n-1}P((n-1)\Delta x, t)}{\Delta x} \frac{\Delta x}{\Delta t}$$

Se richiediamo che f dipenda in modo opportuno da Δx e Δt , operando il limite in modo simile a quanto fatto nel caso uniforme ricaviamo un termine proporzionale a $\frac{\partial P}{\partial x}$ e quindi un'equazione di Fokker-Planck.

4.3.2 Rilassamento ad una gaussiana

Consideriamo la particolare equazione di Langevin (che avevamo già incontrato, in forma molto simile):

$$\frac{dv}{dt} = a(v) + \sqrt{b}\eta(t) = -\frac{v}{\tau} + \sqrt{2c}\eta(t) \quad (4.27)$$

in cui $a(v) = -v/\tau$ e $b = 2c$ (quest'ultima è solo una scelta di comodo). Per quanto trovato, $\langle \Delta v^2 | v \rangle = 2c$.

A questa equazione di Langevin possiamo associare l'equazione di Fokker-Planck:

$$\frac{\partial P(v, t)}{\partial t} = -\frac{\partial}{\partial v} \left(-\frac{v}{\tau} P(v, t) \right) + c \frac{\partial^2 P(v, t)}{\partial v^2} \quad (4.28)$$

Poiché l'equazione di Fokker-Planck è lineare, la sua soluzione può essere scritta nella forma

$$P(v, t) = \int P(v', 0) g(v, v', t) dv'$$

dove $g(v, v', t)$ è la **funzione di Green** (o **propagatore**). Sappiamo anche che la soluzione dell'equazione è una probabilità associata ad un processo markoviano, e pertanto

$$P(v, t) = \int P(v', 0) W(v', 0 \rightarrow v, t) dv'$$

Queste due scritte sono effettivamente identiche, e quindi la funzione di Green non è altro se non la probabilità di saltare da v' a v in un tempo t . Poiché il processo è gaussiano, g sarà una gaussiana di media $\langle v(t) | v(0) = v' \rangle$ e varianza σ^2 . Ricaviamo media e varianza esplicitamente. La soluzione della (8) può essere scritta come

$$v(t) = v(0)e^{-\frac{t}{\tau}} + \sqrt{2c} \int_0^t e^{-\frac{t-t'}{\tau}} \eta(t') dt'$$

da cui, mediando, si ricava

$$\langle v(t)|v(0) = v' \rangle = v' e^{-\frac{t}{\tau}}$$

Consideriamo ora il quadrato di $v(t)$:

$$v(t)^2 = v(0)^2 e^{-\frac{2t}{\tau}} + 2\sqrt{2c}v(0)e^{-\frac{t}{\tau}} \int_0^t e^{-\frac{t-t'}{\tau}} \eta(t') dt' + 2c \int_0^t \int_0^t e^{-\frac{2t}{\tau}} e^{\frac{t'+t''}{\tau}} \eta(t') \eta(t'') dt' dt''$$

Mediando, otteniamo

$$\langle v(t)^2|v(0) = v' \rangle = v'^2 e^{-\frac{2t}{\tau}} + 2c \int_0^t e^{-\frac{2t}{\tau}} e^{\frac{2t'}{\tau}} dt' = v'^2 e^{-\frac{2t}{\tau}} + 2ce^{-\frac{2t}{\tau}} (e^{-\frac{2t}{\tau}} - 1) \frac{\tau}{2}$$

Dunque la varianza è:

$$\sigma^2 = \langle v^2 \rangle - \langle v \rangle^2 = c\tau(1 - e^{-\frac{2t}{\tau}})$$

Osservando poi che

$$\begin{aligned} \lim_{t \rightarrow +\infty} \langle v(t)|v(0) = v' \rangle &= 0 \\ \lim_{t \rightarrow +\infty} \sigma^2 &= c\tau \end{aligned}$$

possiamo scrivere

$$\lim_{t \rightarrow +\infty} P(v, t) = \int P(v', 0) \frac{e^{-\frac{v^2}{2c\tau}}}{\sqrt{2\pi\tau}} dv' = \frac{e^{-\frac{v^2}{2c\tau}}}{\sqrt{2\pi\tau}} \int P(v', 0) dv' = \frac{e^{-\frac{v^2}{2c\tau}}}{\sqrt{2\pi\tau}}$$

Inoltre la stessa gaussiana è una soluzione stazionaria dell'equazione di Fokker-Planck con condizioni nulle all'infinito:

$$\begin{aligned} \frac{\partial P_s(v)}{\partial t} &= -\frac{\partial}{\partial v} \left[-\frac{v}{\tau} P_s(v) \right] + c \frac{\partial^2 P_s(v)}{\partial v^2} \equiv 0 \\ -\frac{\partial}{\partial v} \left[-\frac{v}{\tau} - c \frac{\partial P_s(v)}{\partial v} \right] &\equiv 0 \\ -\frac{v}{\tau} - c \frac{\partial P_s(v)}{\partial v} &\equiv \text{costante} = 0 \text{ per le condizioni nulle all'infinito} \\ \frac{\partial P_s(v)}{\partial v} &= -\frac{v}{c\tau} P_s(v) \\ P_s(v) &= \frac{1}{\sqrt{2\pi c\tau}} e^{-\frac{v^2}{2c\tau}} \end{aligned}$$

dove la costante di proporzionalità in $P_s(v)$ si ottiene imponendone la normalizzazione. Notiamo quindi che il sistema ha un rilassamento alla gaussiana (che è soluzione stazionaria) indipendentemente dalla forma funzionale della probabilità iniziale, cioè $P(v', 0)$.

4.4 Fokker-Planck in più dimensioni

Consideriamo il caso a più dimensioni:

$$\underline{x} = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$$

Possiamo riscrivere l'equazione di Fokker-Planck nel seguente modo:

$$\frac{\partial P(\underline{x}, t)}{\partial t} = -\sum_{n=1}^N \frac{\partial}{\partial x_n} [a_n(\underline{x}) P(\underline{x}, t)] + \frac{1}{2} \sum_{n, n'} \frac{\partial^2}{\partial x_n \partial x_{n'}} [b_{nn'}(\underline{x}) P(\underline{x}, t)]$$

con

$$\begin{aligned} a_n(\underline{x}) &= \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta x_n | \underline{x} \rangle}{\Delta t} \\ b_{nn'}(\underline{x}) &= \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta x_n \Delta x_{n'} | \underline{x} \rangle}{\Delta t} \\ \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta x_n^k \Delta x_{n'}^{k'} | \underline{x} \rangle}{\Delta t} &= 0 \text{ se } k + k' > 2 \end{aligned}$$

4.4.1 Soluzione stazionaria

Nelle lezioni precedenti è stata introdotta l'equazione di Fokker-Planck, che descrive l'evoluzione temporale di una densità di probabilità a partire da due funzioni $a(x)$ e $b(x)$, le quali rappresentano rispettivamente il termine di drift e quello di diffusione. Consideriamo nella fattispecie il caso in cui $b = 2C = \text{cost}$, cioè il caso in cui il coefficiente di diffusione è costante. L'equazione assume dunque la forma:

$$\partial_t P(x, t) = -\partial_x (a(x)P(x, t)) + C \partial_{xx}^2 P(x, t) \quad (4.29)$$

All'equazione eq. (4.29) ne corrisponde una stocastica, che è esprimibile in forma differenziale come:

$$dx = a(x) dt + \sqrt{2C} dw \quad (4.30)$$

A tale forma differenziale corrisponde anche una rappresentazione integrale. Nel caso delle equazioni stocastiche però, si pone una questione rispetto alla definizione dell'integrale stesso. Consideriamo a tal proposito il caso generale di un integrale di una generica funzione $f(x)$.

$$\int_a^b f(x(t)) dt = \lim_{\Delta t \rightarrow 0} \sum_{n=1}^N f(x(t_n)) \Delta t \quad (4.31)$$

dove usualmente il particolare tempo t_n contenuto nell'intervallo n-esimo può essere scelto indifferentemente al centro o agli estremi dell'intervallo (solitamente si sceglie l'estremo inferiore). Al contrario per le equazioni stocastiche potrebbe sorgere una dipendenza dalla scelta della posizione del tempo t_n all'interno del particolare intervallo, rendendo così la rappresentazione integrale puramente formale (**rappresentazione alla Ito**). Nel caso in cui $C = \text{cost}$ questo problema fortunatamente non si pone. Cerchiamo adesso una soluzione stazionaria dell'eq. (4.29), tenendo conto che in una dimensione è sempre possibile scrivere la funzione $a(x)$ come $a(x) = -\frac{dV}{dx}$. Da questa ipotesi si ha infatti:

$$\begin{aligned} -\frac{d}{dx} \left(-\frac{dV}{dx} P_s(x) \right) + C \frac{d^2 P_s(x)}{dx^2} &= 0 \\ -\frac{dV}{dx} P_s(x) &= C \frac{dP_s(x)}{dx} + \text{cost} \\ \frac{1}{P_s(x)} \frac{dP_s(x)}{dx} &= -\frac{1}{C} \frac{dV}{dx} \Rightarrow \\ \ln P_s(x) &= \ln \mathcal{N} - \frac{V(x)}{C} \Rightarrow P_s(x) = \mathcal{N} e^{-\frac{V(x)}{C}} \end{aligned} \quad (4.32)$$

dove al primo passaggio si è integrato ambo i membri rispetto a x , e nel secondo si è tenuto conto che se la $P(x)$ è definita tra $-\infty < x < +\infty$ per la normalizzazione della probabilità la costante d'integrazione deve essere posta a zero. Notiamo che nel nostro caso particolare, se x vive su tutta la retta reale, la soluzione stazionaria è facilmente ricavabile nell'ipotesi in cui $V(x) \rightarrow \infty$ per $x \rightarrow \infty$.

Tuttavia in altre situazioni la soluzione stazionaria potrebbe non esistere, basti pensare ad esempio al caso in cui si abbia $a(x) \equiv 0$, quando il fenomeno è puramente diffusivo e la soluzione assume la forma:

$$P(x, t) \sim \frac{1}{\sqrt{4\pi Ct}} e^{-\frac{x^2}{4Ct}} \quad (4.33)$$

cioè una gaussiana che si allarga nel tempo tendendo asintoticamente a zero, che può dunque essere l'unica soluzione stazionaria (banale). Sottolineiamo infine come le condizioni al bordo giochino un ruolo chiave nell'esistenza della distribuzione stazionaria. Se infatti in un problema di diffusione si considera un dominio finito, la soluzione stazionaria è data da un valore costante all'interno del dominio e nullo al di fuori. Questo è il caso in cui si considera la diffusione di particelle all'interno di una scatola chiusa, o equivalentemente di una stanza senza finestre. Nel momento in cui le finestre vengono aperte le particelle possono diffondere in tutto il dominio, ed è chiaro che se questo è infinitamente esteso non vi sarà più alcuna particella.

Vediamo adesso il caso di un'equazione stocastica in più variabili, da cui è possibile ricavare ancora una volta un'equazione di Fokker-Planck.

$$\dot{x}_n = a_n(x) + \sum_j A_{nj} \eta_j \quad (4.34)$$

dove η_j sono le molteplici variabili stocastiche, rappresentanti rumori bianchi gaussiani, che rispettano cioè le condizioni $\langle \eta_i \rangle = 0$ e $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$. Da questo tipo di processi discendono due casi particolari, cioè quello lineare e i sistemi gradiente. Nel primo vale l'uguaglianza $a_n(x) = \sum_j C_{nj} x_j$, dove \hat{C} è una matrice simmetrica, possedente autovalori reali e negativi. I sistemi gradiente sono invece tali, in quanto è possibile esprimere le $a_n(x)$ come $a_n(x) = -\partial V(\underline{x})/\partial x_n$, e $A_{nj} = \sqrt{2C} \delta_{ij}$. In entrambi i casi è facile ricavare la distribuzione di probabilità stazionaria, mentre il propagatore è facilmente ricavabile nel caso lineare, ma meno in quello gradiente.

4.4.2 Osservazioni sull'equazione di Fokker-Planck

Nello scrivere l'equazione di Langevin avevamo messo dentro rumori gaussiani. L'idea che c'è dietro è che quando si lavora con un sistema con tanti gradi di libertà indipendenti, la particella colloidale percepisca un "effetto medio" di questi gdl per cui si possa invocare il *Teorema del limite centrale*. Mettiamoci in una dimensione e studiamo l'equazione:

$$\dot{x} = -\frac{dV}{dx} + \sqrt{2C} \eta$$

con η rumore gaussiano.

Abbiamo due casi:

1. Caso aperto: l'equazione è definita su un insieme non limitato;
2. Caso periodico: l'equazione è definita su un insieme limitato con condizioni periodiche al bordo.

Studiamo il primo. Se la *probabilità stazionaria* esiste, ha una forma del tipo:

$$P_s = \mathcal{N} e^{-V(x)/C}$$

Affinché ciò sia vero, devo avere che V diverga all'infinito sufficientemente rapidamente (altrimenti P_s non è normalizzabile).

4.5 Legame tra la fisica e i modelli matematici stocastici

Nel corso delle precedenti lezioni si sono affrontati diversi modelli matematici di stampo probabilistico, quali le catene di Markov, le master equation e l'equazione di Fokker-Planck. Sebbene in essi sia

contenuto un innegabile fascino intrinseco, tuttavia non è chiaro quale sia il legame tra questi modelli e la meccanica statistica. Per chiarire tale punto consideriamo ancora una volta a titolo esemplificativo il caso di una particella colloidale sospesa in un bagno termico (sia esso liquido o gassoso). Le ipotesi fondanti del nostro problema sono che la massa della particella colloidale sia molto maggiore di quella delle particelle che compongono il bagno, $M \gg m$, e che il numero di quest'ultime sia $N \gg 1$. Detto questo, la particella colloidale sarà identificata da una coppia di coordinate $(\underline{Q}, \underline{P})$ e l'hamiltoniana del sistema sarà data da:

$$H_{tot} = (\underline{Q}, \underline{P}, \{q_n, p_n\}) = H_C(\underline{Q}, \underline{P}) + H_B(\{q_n\}, \{p_n\}) + H_I(\underline{Q}, \{q_n\}) \quad (4.35)$$

dove i singoli addendi possono essere interpretati nel seguente modo.

- $H_C = \frac{P^2}{2M} + U_e(\underline{Q})$ è l'hamiltoniana che contiene solo il contributo dell'energia cinetica della particella colloidale e del potenziale esterno che agisce su di essa.
- $H_B = \sum_n \frac{p_n^2}{2m} + \sum_{n,n'} V_I(q_n - q_{n'})$ è l'hamiltoniana contenente il solo contributo delle particelle del bagno termico, relativamente alla loro energia cinetica e interazione reciproca.
- $H_I = \sum_n U_I(\underline{Q} - q_n)$ tiene infine conto dell'interazione tra la particella colloidale e quelle costituenti la sospensione.

Cerchiamo dunque le equazioni di Hamilton relativamente alla particella colloidale. L'equazione di evoluzione per \underline{Q} risulta essere banale, infatti:

$$\dot{\underline{Q}} = \frac{\partial H}{\partial \underline{P}} = \frac{\underline{P}}{M} \quad (4.36)$$

Per quanto riguarda l'impulso si ha invece:

$$\dot{\underline{P}} = -\frac{\partial H}{\partial \underline{Q}} = -\frac{\partial U_e}{\partial \underline{Q}} - \sum_n \frac{\partial U_I}{\partial \underline{Q}}(\underline{Q} - q_n) \quad (4.37)$$

in questo caso se il termine relativo al potenziale esterno non dà problemi, quello d'interazione dipende invece non solo dalla particella colloidale, ma anche da tutte le particelle della sospensione alle quali non siamo interessati. In quest'ottica l'idea di Langevin è di supporre che l'equazione di evoluzione per la velocità del colloide (uguale a quella dell'impulso a meno di M) sia $\dot{v} = -\frac{v}{\tau} + \sqrt{2C}\eta$. Dato che l'equazione di Langevin ha come fondamento l'evidenza empirica, la nostra speranza è che sia possibile pervenirci a partire dall'equazione di Hamilton. Quindi se consideriamo per semplicità il caso in cui non vi sia un potenziale esterno, vediamo che dal termine di interazione devono emergere un contributo di attrito ed uno di tipo stocastico, cioè $\sum_n \frac{\partial U_I}{\partial \underline{Q}}(\underline{Q} - q_n) \rightarrow -\gamma \underline{P} + \sqrt{2C}\eta$. A questo punto un approccio possibile è quello alla Langevin, in cui i parametri γ e C sono ricavati sperimentalmente. Altrimenti è necessario stimarli a priori partendo da principi primi, cioè usufruendo della teoria cinetica. Il primo fisico che tentò una tale impresa fu Smoluchowski. Il calcolo è approcciabile ad esempio nel caso bidimensionale di dischi duri, cioè in cui il colloide è un disco di raggio R e massa M e le particelle della sospensione sono approssimativamente puntiformi e di massa m (si ottiene in tal caso che $\gamma = 2\rho \frac{\sqrt{2\pi m k_B T}}{M}$). Una volta noto γ , C è determinato dalla condizione che $\langle \underline{P}^2 \rangle$ sia quello derivante dalla teoria cinetica. Questo problema è risolvibile nel caso di un gas diluito, nei solidi e nel caso in cui l'hamiltoniana sia quadratica. Infatti in questi ultimi due casi le equazioni di evoluzione sono lineari. Si procede dunque risolvendo formalmente le equazioni e poi proiettando sulle variabili di interesse. Per concludere riepiloghiamo dicendo che ogni qualvolta che incontriamo un'equazione stocastica, dobbiamo pensarla come derivante da una variabile collettiva di un sistema hamiltoniano in cui si è proiettato su di essa dimenticandosi tutto il resto del sistema, il quale è responsabile della presenza di un termine di attrito e di uno fluttuante. Sottolineiamo inoltre come ogniqualvolta che nel sistema è presente dissipazione deve esserci anche una fluttuazione, altrimenti esso tenderà a fermarsi asintoticamente.

4.6 Formula di Kramers-Arrhenius

Immaginiamo di avere una variabile che evolve stocasticamente con condizione iniziale per la densità di probabilità $P(x, t = 0) = \delta(x - x_0)$. Chiamiamo Ω una regione dello spazio delle fasi a cui x_0 appartiene. Esisterà un tempo τ t.c. $\underline{x}(\tau)$ per la prima volta attraversa la frontiera di Ω . Questo tempo di prima uscita sarà a sua volta una variabile aleatoria $\tau(\Omega, x_0)$. Il problema di determinare $\langle \tau \rangle$ può essere di grande interesse se sulla frontiera di Ω succede qualcosa, ad esempio che la particella muoia. Il problema generale, dal punto di vista prettamente matematico, richiede la soluzione di un'equazione differenziale non semplice (vedi Boffetta-Vulpiani Sez. 6.3.2). Tuttavia nel 1940 Kramers affrontò il calcolo di $\langle \tau \rangle$ nel caso 1D - in cui esiste sempre un potenziale - con un potenziale bistabile. Per farlo Kramers si avvale di alcune importanti osservazioni fisiche.

L'equazione che regola il processo è

$$\dot{x} = -V'(x) + \sqrt{2c}\eta \quad (4.38)$$

dove $V(x)$ è un potenziale come quello, in blu, in fig. 4.1.

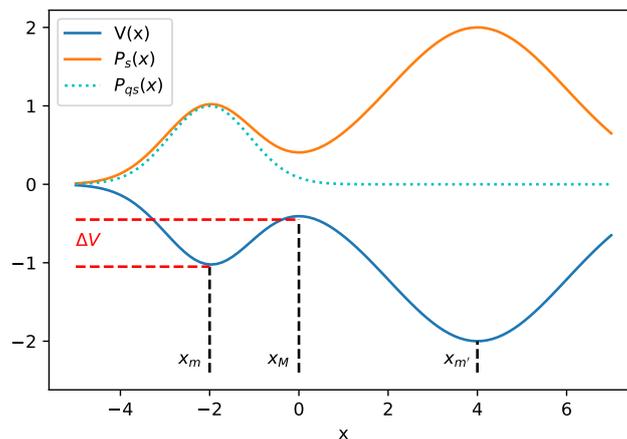


Figura 4.1: In blu: potenziale a due minimi; in arancione: probabilità stazionaria della x ; in azzurro: probabilità quasi-stazionaria della x

Se ora chiamiamo Ω la regione di spazio $(-\infty, x_{M+})$ il tempo di prima uscita τ coincide con l'attraversamento della barriera di potenziale che porta la particella fuori dal minimo di sinistra (meno stabile). Scelgo come condizione iniziale $x_0 = x_m$ cioè la posizione del minimo di sinistra in fig. 4.1. Passando nella variabile $z = x - x_m$ e compiendo un'approssimazione parabolica attorno al minimo x_m t.c. $V'(z) \simeq V''(x_m)z$ (cioè per tempi piccoli, in cui resto confinato attorno al minimo da cui parto), posso approssimare l'eq. (4.38) come:

$$\dot{z} \simeq -V''(x_m)z + \sqrt{2c}\eta \quad (4.39)$$

Seguendo l'evoluzione stocastica della variabile z ho che in un tempo dell'ordine di $\tau_0 = \frac{1}{V''(x_m)}$ la $P(z, t)$ rilassa da $P(z, 0) = \delta(z)$ a una $P_{qs}(z, t)$ quasi stazionaria, ovvero a una gaussiana centrata attorno a $z = 0$ come la curva azzurra in fig. 4.1. Del resto so che questa fase può essere solo transiente e che per $t \rightarrow \infty$ la $P(z, t) \rightarrow P_s(z) = \mathcal{N}e^{-\frac{V(z)}{c}}$ (curva arancione in fig. 4.1).

Dopo che la particella supera la barriera di potenziale per la prima volta (e arriva a x_{M+} , punto appena a destra del massimo), essa cade con probabilità vicina ad uno nel minimo più profondo $V(x_{m'})$, trascinata dal pezzo deterministico dell'eq. (4.38). Perciò il tempo di convergenza verso $P_s(z)$ sarà

dell'ordine $\langle \tau \rangle$ (\gg dei due tempi τ_0 e τ'_0 in cui la $P(x, t)$ si aggiusta gaussianamente attorno ai minimi del potenziale).

Già verso la fine dell'800 Arrhenius (che era un chimico-fisico) era giunto empiricamente alla formula $R \sim e^{-\frac{\Delta E}{k_B T}}$ per il rate di transizioni tra i due minimi di un potenziale bistabile, tipico nella descrizione di reazioni chimiche in cui sono presenti due equilibri (uno da cui la reazione inizia e uno, più stabile, dove finisce).

Cercando di ottenere teoricamente un risultato analogo a quello di Arrhenius, Kramers scrisse l'equazione di Fokker-Planck associata all'eq. (4.38):

$$\frac{\partial P(x, t)}{\partial t} + \frac{\partial}{\partial x}(-V'(x)P(x, t)) - c \frac{\partial^2 P(x, t)}{\partial x^2} = 0 \quad (4.40)$$

Che rinominando $J = -V'P - c \frac{\partial P}{\partial x}$ assume la forma di un'equazione di continuità

$$\frac{\partial P}{\partial t} + \frac{\partial J}{\partial x} = 0 \quad (4.41)$$

J può essere interpretata come una densità di corrente di probabilità. Quindi quando ho $P_s(x)$ trovo $\frac{\partial J}{\partial x} = 0$ e dunque $J = \text{cost}$ nella zona di interesse.

Kramers si riscrisse ⁴ la J come

$$J = -ce^{-\frac{V}{c}} \frac{\partial}{\partial x}(Pe^{V/c}) = -ce^{-\frac{V}{c}} \left[\frac{V'}{c} e^{V/c} + \frac{\partial P}{\partial x}(e^{V/c}) \right] = -V'P - c \frac{\partial P}{\partial x} \quad (4.42)$$

Ricapitolando abbiamo che, partendo con la condizione iniziale $P(x, t = 0) = \delta(x - x_m)$, per tempi $t \gtrsim \tau_0$ la eq. (4.40) ammetterà come soluzione

$$P(x, t) = \begin{cases} P_{qs} & x < x_M \\ 0 & x > x_M \end{cases} \quad (4.43)$$

Cioè abbiamo probabilità circa nulla di aver superato la barriera di potenziale ⁵ e di trovarci a $x > x_M$ mentre per $x < x_M$ si ha la probabilità Gaussiana quasi stazionaria, dipendente poco dal tempo.

Riprendendo la eq. (4.42) abbiamo che vale

$$\frac{J}{c} e^{\frac{V}{c}} = -\frac{\partial}{\partial x}(e^{\frac{V}{c}} P) \quad (4.44)$$

sia a sinistra per $x < x_M$ dove ho $J \sim \text{cost}$ e P_{qs} sia, banalmente, per $x > x_M$ dove $J \sim 0$ e $P(x, t) = 0$. Integrando la eq. (4.44) tra x_m e x_{M+} otteniamo

$$\frac{J}{c} \int_{x_m}^{x_{M+}} e^{\frac{V}{c}} dx = - \int_{x_m}^{x_{M+}} \frac{\partial}{\partial x}(e^{\frac{V}{c}} P) dx = P(x_m, t) e^{\frac{V(x_m)}{c}} - P(x_{M+}, t) e^{\frac{V(x_{M+})}{c}} \quad (4.45)$$

Ma poiché per la eq. (4.43) si ha che $P(x_{M+}, t) \sim 0$ troviamo una formula per J

$$J \simeq c \frac{P(x_m, t) e^{\frac{V(x_m)}{c}}}{\int_{x_m}^{x_{M+}} e^{\frac{V}{c}} dx} \quad (4.46)$$

Per proseguire è necessario assumere che, poiché $P_s(x) = \mathcal{N}e^{-\frac{V(x)}{c}}$, anche per $t \gtrsim \tau_0$ (quando la probabilità si è aggiustata attorno a x_m)

$$P(x, t) \simeq P(x_m, t) e^{-\frac{[V(x) - V(x_m)]}{c}} \quad (4.47)$$

che vale sicuramente per tempi lunghi⁶. Ora Kramers voleva determinare $P = \text{Prob}(x < x_m)$ (che

⁴Riscrittura che forse venne in mente a Kramers notando che all'equilibrio $\frac{\partial}{\partial x}(P_s(x)e^{V/c}) = 0$

⁵Assumiamo che $\frac{\Delta V}{c} \gg 1$ cosicché la coda della Gaussiana si sia praticamente estinta oltre x_M

⁶Per vederlo basta fare il rapporto tra $P_s(x)$ e $P_s(x_m)$.

dipende poco dal tempo). Infatti, nota P , si può calcolare il rate $R \equiv \frac{1}{\langle \tau \rangle}$ con cui una particella dal minimo di sinistra salta nel minimo di destra: proprio la quantità ricavata da Arrhenius. Per far ciò bisogna pensare fisicamente: il rate R è la costante di proporzionalità che intercorre tra la $P(x < x_m)$ e la corrente di probabilità J che va da x_m a $x_{m'}$

$$J = P(x < x_M) * R \quad (4.48)$$

Intuitivamente, se osservo l'evoluzione di molte particelle in contemporanea, J è il flusso netto di particelle che va da x_m a $x_{m'}$. Sapendo dall'eq. (4.47) che $P = P(x_m, t)e^{\left(\frac{V(x_m)}{c}\right)} \int_{-\infty}^{x_M} e^{-\frac{V(x)}{c}} dx$ e usando l'eq. (4.48) otteniamo

$$R = \frac{J}{P} \simeq \frac{cP(x_m, t)e^{\left(\frac{V(x_m)}{c}\right)}}{\int_{x_m}^{x_{M+}} e^{\frac{V}{c}} dx P(x_m, t)e^{\left(\frac{V(x_m)}{c}\right)} \int_{-\infty}^{x_M} e^{-\frac{V(x)}{c}} dx} = \frac{c}{\int_{x_m}^{x_{M+}} e^{\frac{V}{c}} dx \int_{-\infty}^{x_M} e^{-\frac{V(x)}{c}} dx} \quad (4.49)$$

Utilizziamo il metodo del punto di sella (sotto l'ipotesi che $\frac{\Delta V}{c} \gg 1$, $\Delta V \equiv V(x_M) - V(x_m)$) per svolgere gli integrali a denominatore, considerando che in quello dove l'esponente è positivo il contributo maggiore all'integrale si ha attorno al massimo x_M mentre dove l'esponente è negativo sviluppo attorno al minimo x_m . Allora

$$\int_{x_m}^{x_{M+}} e^{\frac{V(x)}{c}} dx = e^{V(x_M)} \sqrt{\frac{2\pi c}{|V''(x_M)|}}$$

e facendo analogamente per il secondo integrale troviamo⁷

$$R \simeq \frac{1}{2\pi} e^{-\frac{\Delta V}{c}} \sqrt{V''(x_m)|V''(x_M)|} = \frac{1}{\langle \tau \rangle} \quad (4.50)$$

Questo è proprio il risultato di Arrhenius, ottenuto questa volta deduttivamente, tenendo conto del fatto che $c = k_B T$ e $\Delta E = \Delta V$. L'eq. (4.50) mostra chiaramente come $\langle \tau \rangle$ cresca esponenzialmente al crescere della barriera di potenziale ΔV e al diminuire di T .

Riferimenti bibliografici

- [1] G. Boffetta e A. Vulpiani. *Probabilità in Fisica*. Springer- Verlag, 2012.

⁷Risolvendo l'eq. (4.40) con l'approssimazione WKB (Wentzel-Kramers-Brillouin) si ottiene lo stesso risultato.

5. Sistemi dinamici

Cominciamo adesso la trattazione dei **sistemi dinamici**, intesi come tutti quei sistemi in cui le variabili che descrivono lo stato del sistema hanno un'evoluzione di carattere deterministico, cioè completamente specificata dalle proprietà del sistema e dalle condizioni iniziali. Vedremo poi in seguito come questo, nella fattispecie relativamente a quei sistemi che presentano un comportamento caotico, si concili con una visione probabilistica dell'evoluzione. Introduciamo ora un esempio importante di sistema dinamico, rappresentato da un'equazione differenziale autonoma in N dimensioni:

$$\dot{\underline{x}}(t) = F(\underline{x}(t)) \quad (5.1)$$

Nei sistemi hamiltoniani, cioè nei sistemi la cui evoluzione sia determinata dalle equazioni di Hamilton, l'evoluzione si svolge su un'ipersuperficie ad energia costante. In altre parole se ad un certo valore iniziale della variabile \underline{x} , che indichiamo con $\underline{x}(0)$, corrisponde una certa energia $E = H(\underline{x}(0))$, questa si mantiene costante durante il moto, cioè gli unici $\underline{x}(t)$ ammissibili sono quelli per cui $\underline{x}(t) \in \{\underline{x} : H(\underline{x}) = E\} = \Omega$. Il valore di $\underline{x}(t)$ e quello all'istante iniziale sono collegati secondo la scrittura simbolica:

$$\underline{x}(t) = S^t \underline{x}(0) \quad (5.2)$$

dove S^t è il semigrupp che rappresenta l'evoluzione in un tempo t . I sistemi hamiltoniani posseggono un'altra interessante proprietà, descritta dal **teorema di Liouville**. Quest'ultimo afferma che il volume di una certa regione dello spazio delle fasi risulta essere costante durante il moto. In termini matematici questo si esprime attraverso l'uguaglianza $\mu(A) = \mu(S^t A)$, dove μ rappresenta una misura associata allo spazio in esame (e dunque una probabilità) e $A \in \Omega$ ¹. I sistemi meccanici godono inoltre dell'invarianza sotto inversione temporale, per cui se A è una certa sottoregione dell'ipersuperficie Ω e $\underline{y} \in S^{-t}(A) \Rightarrow S^t \underline{y} = \underline{x} \in A$. La misura di un insieme può essere espressa nei sistemi hamiltoniani attraverso la densità microcanonica ρ_m , cioè

$$\mu(A) = \int_A d\underline{x} \rho_m(\underline{x}) \quad (5.3)$$

¹La misura $\mu(A)$ può essere interpretata come la probabilità che la traiettoria si trovi in A .

Abbiamo a questo punto introdotto tutti gli elementi che compongono la “sacra terna” che definisce un sistema dinamico:

Definizione 5.1 — Sistema dinamico. Un sistema dinamico è univocamente individuato dalla terna

$$(\Omega, S^t, \mu) \quad (5.4)$$

dove Ω individua la regione in cui si svolge l'evoluzione (**spazio delle fasi**), S^t come questa effettivamente avviene (**semigruppato di evoluzione temporale**) e infine una **misura invariante** μ (collegata quindi alla probabilità).

Sottolineiamo come la misura μ risulta essere invariante, ed in generale la sua determinazione fa parte del problema stesso.

Fino ad adesso abbiamo trattato problemi in cui la variabile temporale risulta essere continua, cioè equazioni di evoluzione della forma eq. (5.1). Tuttavia possono essere equivalentemente considerati sistemi in cui l'evoluzione avviene a tempi discreti, secondo la legge generale:

$$\underline{x}(t+1) = \mathcal{G}(\underline{x}(t)) \quad (5.5)$$

detta **regola ricorsiva** o mappa. Diversi esempi possono essere fatti per convincersi dell'equivalenza tra le mappe e i processi continui.

■ **Esempio 5.1** In primo luogo la risoluzione numerica di un'equazione differenziale al calcolatore sottintende una discretizzazione dei tempi, conseguente al limitato numero di cifre associabile ad un numero float. Il generico algoritmo di integrazione può quindi essere espresso come $\underline{x}(t + \Delta t) = \mathcal{G}(\underline{x}(t), \Delta t)$, in cui si riconosce la stessa forma dell'eq. (5.5). ■

■ **Esempio 5.2** Un secondo esempio è quello delle mappe stroboscopiche, in cui la regola consiste nel considerare \underline{x} solo ad istanti multipli interi di un certo tempo T_0 (periodo) più uno shift arbitrario t^* . Questa tecnica permette di determinare la periodicità del moto, che può essere ad esempio un multiplo del periodo della forzante che agisce sullo stesso. Nel caso in cui si ponga $t^* = 0$ la regola ricorsiva associata è $\underline{x}(nT_0 + T_0) = \mathcal{G}(\underline{x}(nT_0))$. ■

■ **Esempio 5.3 — Mappa di Poincaré.** Un ultimo esempio è dato dalla cosiddetta mappa di Poincaré: consideriamo per semplicità $\underline{x} \in \mathbb{R}^3$, e scegliamo all'interno dello spazio tridimensionale un particolare piano. All'interno di questo spazio consideriamo inoltre l'evoluzione di un sistema autonomo $\dot{\underline{x}} = F(\underline{x})$ e indichiamo con $\Gamma_i \in \mathbb{R}^2$, $i = 1, 2, \dots, n$ la sequenza dei punti corrispondenti alle intersezioni della traiettoria con il suddetto piano, in modo che queste avvengano sempre con lo stesso verso. La mappa che associa due punti consecutivi di questa successione è appunto una mappa di Poincaré. Essa è ben definita, perché direttamente derivante dal processo di evoluzione temporale. In altri termini dato Γ_n , Γ_{n+1} è determinato tramite integrazione dell'equazione differenziale tra i tempi relativi alle due intersezioni. Per lo stesso motivo la mappa è anche invertibile, cioè se $\Gamma_{n+1} = \mathcal{G}(\Gamma_n)$, allo stesso modo $\Gamma_{n-1} = \mathcal{G}^{-1}(\Gamma_n)$. Concludiamo evidenziando che se la traiettoria corrispondente all'equazione vive in d dimensioni, la mappa di Poincaré ad essa associata vivrà in $d - 1$ dimensioni. Questo permette senza dubbio una semplificazione nell'analisi del problema, e al contempo permette di rappresentare in maniera grafica sistemi in cui $d = 4$. In fig. 5.1 un esempio grafico in $d=3$ di una mappa di Poincaré. ■

Convintisi dell'equivalenza tra sistemi dinamici continui e discreti, passiamo ad analizzare le condizioni sotto le quali un sistema possa presentare un comportamento caotico, che vedremo essere strettamente dipendente dalla dimensione dello spazio. Cominciamo col dire che le equazioni differenziali definite da sistemi lineari presentano (in dimensioni finite), solo tre diverse tipologie di comportamento, cioè esponenziale crescente, decrescente o oscillante. Nel caso non lineare possono invece presentarsi

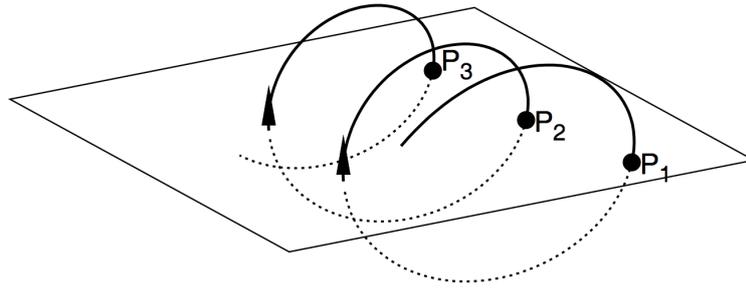


Figura 5.1: Es. di mappa di Poincaré. Si noti che la traiettoria interseca il piano sempre nello stesso verso.

comportamenti di tipo caotico, del quale un celebre esempio è rappresentato dal problema dei tre corpi, al quale Poincaré approcciò per la prima volta con un metodo non perturbativo, dato che quest'ultimo restituiva coefficienti divergenti sin dal primo ordine. Questo è appunto una conseguenza della caoticità deterministica del sistema, che determina la presenza di orbite non periodiche e instabili (ma limitate). Se si considera invece un sistema bidimensionale di hamiltoniana $H = p^2/2 + V(q)$, con $V(q) = q^2 + \alpha q^4 + \beta q^6$, sebbene la soluzione esplicita non sia nota, è possibile affermare a priori la periodicità del moto, che si svolge in corrispondenza dell'orbita chiusa ad energia costante.

Tale proprietà può essere interpretata anche dal punto di vista del **teorema di Poincaré-Bendixon**. Quest'ultimo afferma infatti che, sotto ragionevoli condizioni, le equazioni differenziali autonome in $d = 2$ presentano in una regione limitata dello spazio delle fasi solo due possibili tipi di comportamento:

- le traiettorie tendono/divergono ad orbite chiuse (**cicli limite**);
- le traiettorie tendono/divergono ad un punto nello spazio delle fasi (**punti fissi**).

Dunque per osservare un comportamento caotico per le equazioni differenziali autonome bisogna porsi in almeno $d = 3$. Corrispondentemente per le mappe invertibili il numero di dimensioni minime risulta essere $d = 2$. Basti a tal proposito pensare ad una mappa di Poincaré associata ad un'equazione differenziale autonoma in tre dimensioni. Tuttavia le mappe non invertibili possono presentare un comportamento caotico anche in $d = 1$ (e.g. la mappa logistica).

Consideriamo adesso un esempio di mappa caotica unidimensionale

■ **Esempio 5.4 — Bernoulli shift.** La sua definizione ricorsiva è la seguente:

$$x(t+1) = \mathcal{G}(x(t)) = 2x(t) \pmod{1} = \begin{cases} 2x(t) & 0 \leq x(t) < 1/2 \\ 2x(t) - 1 & 1/2 \leq x(t) < 1 \end{cases} \quad (5.6)$$

La sua caoticità risiede nella seguente proprietà: un errore iniziale anche piccolo si propaga amplificandosi in modo drammatico dopo un numero relativamente limitato di iterazioni.

Per analizzare la caoticità del sistema consideriamo la rappresentazione binaria di $x_0 \in [0, 1]$:

$$x_0 = \sum_{n=1}^{\infty} \frac{a_n}{2^n} = \frac{a_1}{2} + \frac{a_2}{4} + \frac{a_3}{8} + \frac{a_4}{16} + \dots = 0.a_1a_2a_3\dots$$

dove gli $a_n = 0, 1$. Secondo la regola si avrà quindi

$$x_1 = 2x_0 \pmod{1} = \left[a_1 + \underbrace{\frac{a_2}{2} + \frac{a_3}{4} + \dots}_{<1} \right] \pmod{1}$$

Se $a_1 = 1$, poiché $\sum_{n=2}^{\infty} \frac{a_n}{2^n} < 1$ il suo $\text{mod } 1$ è pari alla somma stessa. Invece se $a_1 = 0$ esso è già non presente. In definitiva $x_1 = 0.a_2a_3a_4\dots$, cioè lo stesso numero con la prima cifra tagliata. Consideriamo dunque $x_0 \simeq x'_0$ entro un $\mathcal{O}(2^{-M})$ ², cioè $a_n = a'_n$ per $n = 1, \dots, M$, con ad esempio $M = 20$. L'errore iniziale è dell'ordine di $2^{-20} \simeq 10^{-6}$. Tuttavia esso si propaga secondo:

$$\delta x_t \sim 2^t \delta x_0 = \delta x_0 e^{t \ln 2} \quad (5.7)$$

cioè cresce esponenzialmente con esponente $\lambda = \ln 2$. Dopo M passi l'errore diventa $\mathcal{O}(1)$, e quindi i due numeri sono completamente diversi. Per concludere le osservazioni riguardo questa mappa evidenziamo come se $x_0 \in \mathbb{Q}$, allora la mappa è periodica, oppure converge a 0 dopo un numero finito di iterazioni. ■

■ **Esempio 5.5 — Equazioni di Hamilton.** Un sistema dinamico è ad esempio costituito dalle equazioni differenziali, il cui caso più importante, nonché particolarmente difficile da studiare, è quello delle **Equazioni di Hamilton**, dove lo spazio delle fasi è quello ad energia costante e la misura invariante è la **misura microcanonica**: $\mu = \rho_{micro}$.

Iniziamo col caratterizzare i sistemi Hamiltoniani dal punto di vista dei sistemi dinamici: un sistema Hamiltoniano è univocamente descritto dal vettore

$$\underline{x} \equiv (\underline{q}_1, \dots, \underline{q}_N; \underline{p}_1, \dots, \underline{p}_N)$$

Qui lo spazio delle fasi è una superficie iso-energetica di energia E e il semi-gruppo di evoluzione viene dalla soluzione delle equazioni di Hamilton:

$$\begin{aligned} \Omega &= \{\underline{x} | \mathcal{H}(\underline{x}) = E\} \\ \dot{\underline{x}} &= \underline{F}(\underline{x}) \rightarrow \underline{x}(t) = S^t \underline{x}(0) \end{aligned}$$

Cerchiamo di capire come è fatta la ρ . Se definiamo l'energia E entro un certo intervallo Δ , questa si la possiamo scrivere come:

$$\rho_m = \begin{cases} 0 & \text{se } \mathcal{H}(\underline{x}) \in [E, E + \Delta] \\ \frac{1}{P_{\Delta}(E)} & \text{se } \mathcal{H}(\underline{x}) \notin [E, E + \Delta] \end{cases}$$

con $P_{\Delta}(E) = \int_{E < \mathcal{H} < E + \Delta} d\underline{x}$. Si può dimostrare che se $\Delta \rightarrow 0$ allora

$$\rho_m(\underline{x}) d\underline{x} = \frac{d\Sigma(\underline{x})}{|\nabla \mathcal{H}|}$$

dove il $d\Sigma$ è l'elemento di superficie a energia costante che, essendo $\underline{x} \in \mathbb{R}^{6N}$, sarà una superficie $(6N - 1)$ -dimensionale. ■

■ **Esempio 5.6 — Oscillatore armonico unidimensionale.** La superficie ad energia costante è un'ellisse unidimensionale che vive in \mathbb{R}^2 poiché vale la conservazione dell'energia meccanica:

$$E = \frac{p^2}{2m} + \frac{m\omega^2}{2} x^2$$

²Per chiarezza, le due quantità x_0 e x'_0 sono uguali fino all' M -esima cifra. Attuando il Bernoulli shift la prima cifra di entrambe le quantità verrà eliminata e dunque saranno uguali fino all' $(M-1)$ -esima cifra.

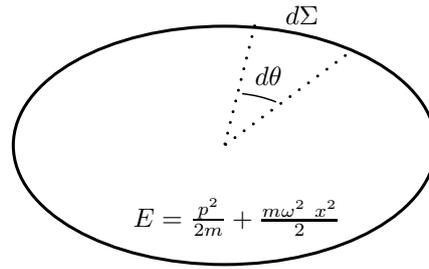


Figura 5.2: Superficie iso-energetica dell'oscillatore armonico

In questo caso si ha che $\rho_m(\mathbf{x})d\mathbf{x} \rightarrow f(\theta)d\theta$, dove $f(\theta)$ include anche il fattore di normalizzazione. ■

5.1 Ergodicità

Torniamo adesso alla questione dell'approccio probabilistico, chiedendoci come vada ad inserirsi in questo contesto. Nonostante tutte le nostre considerazioni sui sistemi dinamici sono di natura deterministica, è evidente che nella realtà ad ogni punto \underline{x} si associa una certa incertezza ϵ , dunque una densità di probabilità. Da un punto di vista fisico risulta dunque più interessante chiedersi come evolve la densità di probabilità associata ad un certo dato iniziale $\rho_0(\underline{x}) \rightarrow \rho_1(\underline{x}) \rightarrow \dots$, ponendosi per essa le stesse domande fatte nel caso dei sistemi stocastici, cioè se ad esempio esiste una densità di probabilità asintotica $\lim_{t \rightarrow \infty} \rho_t(\underline{x}) = \rho_\infty(\underline{x})$.

Si immagini di voler fare una misura su un sistema fisico di un *osservabile* $A(\mathbf{x})$. Tipicamente, a causa della risoluzione temporale finita dello strumento, ciò che realmente si misura è una *media* di A sul tempo T in cui stiamo effettuando la misura:

$$\bar{A} = \frac{1}{T} \int_t^{t+T} ds A(\mathbf{x}(s))$$

A questo punto però, il calcolo di \bar{A} da un punto di vista teorico presenta una serie di problemi:

- conoscere la condizione iniziale $\mathbf{x}(0)$;
- studiare l'evoluzione $\mathbf{x}(0) \rightarrow \mathbf{x}(t) = S^t \mathbf{x}(0)$;
- calcolare l'integrale.

Quindi abbiamo in primis il problema tecnologico della misura delle condizioni iniziali, poi quello matematico del risolvere le equazioni del moto e infine, ma non meno banale, calcolare l'integrale. Quando studiamo un sistema macroscopico conoscere le condizioni iniziali di un numero di particelle come $N \sim 10^{20}$ e studiarne l'evoluzione è nella pratica impossibile. Storicamente fu Boltzmann a risolvere il problema assumendo che il sistema sia **ergodico**, ossia:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} ds A(\underline{x}(s)) = \int \rho_m(\underline{x}) d\underline{x} A(\underline{x}) \quad (5.8)$$

L'idea di mandare $T \rightarrow \infty$ è fisicamente sensata, poiché la misura avviene su intervalli di tempi molto lunghi rispetto a quelli tipici atomici o molecolari. Se l'ipotesi è valida, allora non è più necessario conoscere la dinamica del sistema e il tutto si riconduce alla conoscenza di ρ_m . Il problema, formulato originariamente in ambito di Meccanica Statistica, può essere riscritto anche in senso più astratto attraverso la seguente definizione:

Definizione 5.2 — Ergodicità. Dato un sistema dinamico (Ω, S^t, μ) questo si dice ergodico se vale:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} ds A(\underline{x}(s)) = \int d\mu(\underline{x}) A(\underline{x}) \quad (5.9)$$

Se poi possiamo definire opportunamente una *derivata*, scriviamo $d\mu = \rho^{inv}(\mathbf{x})d\mathbf{x}$.

Matematicamente si ha quindi una certa legge di evoluzione temporale che deve essere messa in relazione con una misura. Per cui "l'ergodicità risulta un problema di trasformazione su spazi di misura". Quello di cui ci interesseremo noi però è il punto di vista fisico del problema.

5.2 Mixing

Un altro problema che ci si può porre è quello del *Mixing*, che fu introdotto invece da Gibbs.

Abbiamo introdotto la definizione di "ergodicità" perché questa individua i sistemi interessanti dal punto di vista della Meccanica Statistica. Esistono però dei casi in cui il sistema, seppur ergodico, risulta essere banale da studiare: è il caso dei *sistemi ergodici mescolanti (mixing)*. Si considerino $A, B \subseteq \Omega$ e la preimmagine di B : intersechiamola con A e facciamone la misura: $\mu(A \cap S^{-t}B)$.

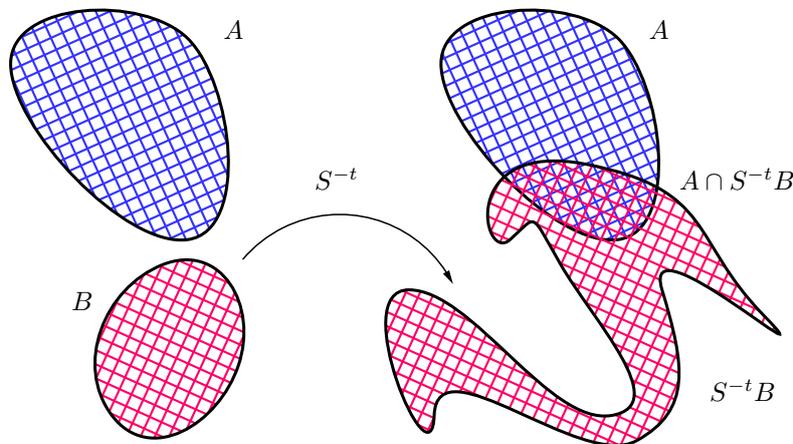


Figura 5.3: Rappresentazione grafica dell'evoluto $S^t B$ e della sua intersezione con A

Definizione 5.3 — Mixing. Il sistema è detto in mixing se vale che:

$$\lim_{t \rightarrow \infty} \mu(A \cap S^{-t}B) = \mu(A)\mu(B) \quad \forall A, B \subseteq \Omega \quad (5.10)$$

Scritta così la definizione è di difficile interpretazione fisica. Assumiamo che valgano "ipotesi di invertibilità", ossia di poter sostituire il $-t$ con t (non sempre è vero). Il mixing diventa:

$$\mu(A \cap S^t B) \rightarrow \mu(A)\mu(B)$$

Scritto così, riusciamo anche a capire cosa fisicamente è un sistema mixing. Si immagina di avere un bicchiere d'acqua, di volume Ω , di cui considero un sottoinsieme A e uno B . Suppongo che in B ci siano particelle di inchiostro. Se mescolo il contenuto del bicchiere, mi aspetto che le particelle in B si diffondano in tutto il volume, come in [fig. 5.4](#).

Dire che il sistema è mixing vuol dire allora che l'inchiostro si diffonde *uniformemente* nel bicchiere, per cui la frazione di inchiostro che parte da B e arriva in A (che è $\mu(A \cap S^t B)/\mu(B)$) coincide con la frazione di volume occupato da A (che è $\mu(A)$).

Il problema può essere posto in maniera equivalente in termini di densità di probabilità:

1. Parto da una certa distribuzione: $\rho(\mathbf{x}, 0)$
2. La faccio evolvere fino al tempo t $\rho(\mathbf{x}, t)$
3. Mi chiedo: converge a una $\rho_{inv}(\mathbf{x})$?

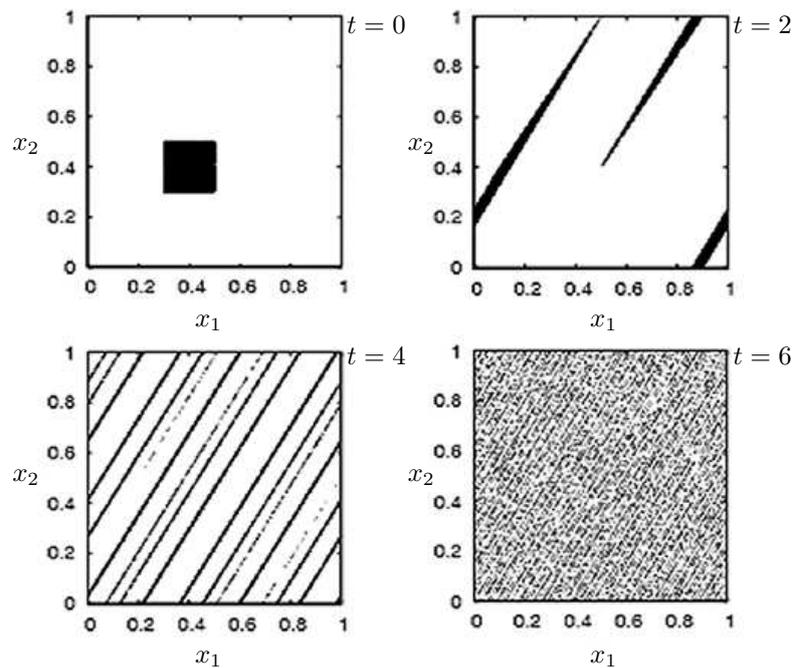


Figura 5.4: Le “gocce di inchiostro” sono confinate in quadrato a $t = 0$. All’evolvere di t , si diffondono uniformemente in tutto il bicchiere

E’ praticamente lo stesso problema posto dal Boltzmann nel *Teorema H*³, solo che stavolta ragioniamo nello *spazio delle fasi* Ω e non nello spazio μ . Si può far vedere che un sistema mixing è sempre ergodico, ma non vale il contrario.

$$\begin{aligned} \text{MIXING} &\implies \text{ERGODICO} \\ \text{ERGODICO} &\not\implies \text{MIXING} \end{aligned}$$

5.2.1 Calcolo della misura invariante

Data una mappa di un sistema dinamico, assegniamo una distribuzione iniziale $\rho_0(\mathbf{x})$ e vediamo come si comporta la generica ρ_t . Nel caso delle *catene di Markov*, studiavamo un problema di evoluzione della probabilità a tempi discreti:

$$\underline{P}(t+1) = \widehat{W}\underline{P}(t)$$

qui la probabilità invariante risolveva l’equazione: $\underline{P}^{inv} = \widehat{W}\underline{P}^{inv}$. Cerchiamo di scrivere delle regole ricorsive analoghe per la $\rho(x)$.

■ **Esempio 5.7 — Bernoulli shift.** Mettiamoci in una dimensione nel caso semplice in cui la mappa è:

$$x_{t+1} = 2x_t \pmod{1}$$

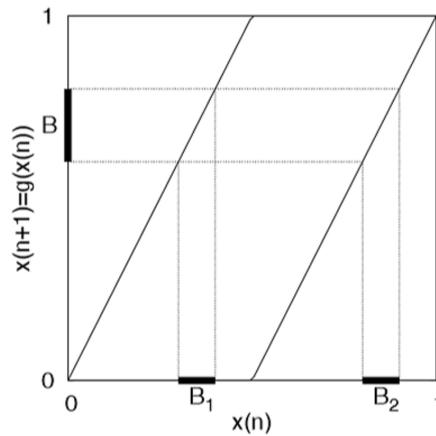
La probabilità di trovare x_{t+1} in x , entro un intorno Δx piccolo, è data dalla probabilità di avere $x = 2x_t \pmod{1}$. Per cui:

³Nel caso di Boltzmann la ρ_{inv} era quella di Maxwell-Boltzmann.

$$\begin{aligned}
\rho_{t+1}(x)\Delta x &= P(x < x_{t+1} < x + \Delta x) = \\
&= P\left(\frac{x}{2} < x_t < \frac{x}{2} + \frac{\Delta x}{2}\right) + P\left(\frac{x}{2} + \frac{1}{2} < x_t < \frac{x}{2} + \frac{1}{2} + \frac{\Delta x}{2}\right) = \\
&= \rho_t\left(\frac{x}{2}\right)\frac{\Delta x}{2} + \rho_t\left(\frac{x}{2} + \frac{1}{2}\right)\frac{\Delta x}{2}
\end{aligned}$$

Per cui ricavo l'equazione di evoluzione:

$$\rho_{t+1}(x) = \frac{1}{2}\rho_t\left(\frac{x}{2}\right) + \frac{1}{2}\rho_t\left(\frac{x}{2} + \frac{1}{2}\right)$$



A questo punto è chiaro che nel caso generale noi dobbiamo guardare le **pre-immagini** della mappa, per definire fuori una regola di evoluzione. Prendiamo una mappa $x_{t+1} = f(x)$ e definiamo i punti $x^{(k)}$ tali che $x = f(x^{(k)})$. Allora a un x definito entro un Δx corrispondono preimmagini $x^{(k)}$ definite entro $\Delta x^{(k)} = \Delta x / |f'(x^{(k)})|$, come si vede graficamente.

La regola di evoluzione generale della ρ_t allora è:

$$\rho_{t+1}(x) = \sum_{x^{(k)}} \frac{1}{|f'(x^{(k)})|} \rho_t(x^{(k)}) \quad (5.11)$$

dove $f(x^{(k)}) = x$. Il numero delle preimmagini, in generale, dipende dall' x considerato.

Otteniamo quindi una (complicata) regola ricorsiva lineare. Questa è comunemente scritta come:

$$\rho_{t+1} = \mathcal{L}_{PF} \rho_t \quad (5.12)$$

dove \mathcal{L}_{PF} è definito come **Operatore di Perron-Frobenius**. Questa è analoga delle catene di Markov, però a dimensione infinita (da cui la complicazione).

Per definizione ora la ρ_{inv} è l'auto-funzione dell'operatore:

$$\rho_{inv} = \mathcal{L}_{PF} \rho_{inv}$$

Tornando all'esempio del Bernoulli shift, troviamo:

$$\rho_{inv}(x) = \frac{1}{2}\rho_{inv}\left(\frac{x}{2}\right) + \frac{1}{2}\rho_{inv}\left(\frac{x}{2} + \frac{1}{2}\right)$$

che è risolta da $\rho_{inv} = \text{cost.}$ ■

5.2.2 Sistemi ergodici non in mixing

Come abbiamo già sottolineato, un sistema ergodico non è necessariamente in mixing. Per dimostrarlo consideriamo un controesempio:

■ **Esempio 5.8 — La rotazione sul cerchio.** Come esempio di sistema *ergodico*, ma non *mixing*, consideriamo in particolare il sistema la regola ricorsiva:

$$x_{t+1} = (x_t + \omega) \bmod 1$$

Lo spazio delle fasi è il *cerchio* $\Omega = [0, 1]$, definito con condizioni periodiche al bordo:

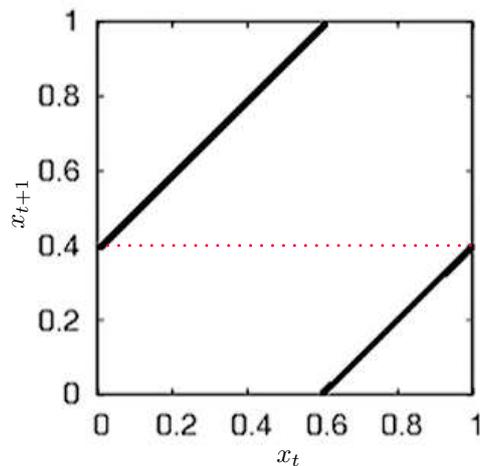


Figura 5.5: Rappresentazione grafica della mappa $x_{t+1} = x_t + \omega(\bmod 1)$

Calcolo della misura invariante

Nel caso della *rotazione sul cerchio*, in analogia con quanto ottenuto nell'eq. (5.11), si ottiene l'equazione:

$$\rho_{t+1}(x) = \rho_t(x - \omega)$$

dove tutto inteso (mod 1). Ciò che accade è che, ad ogni istante di tempo, la distribuzione di probabilità viene tralata rigidamente di ω .

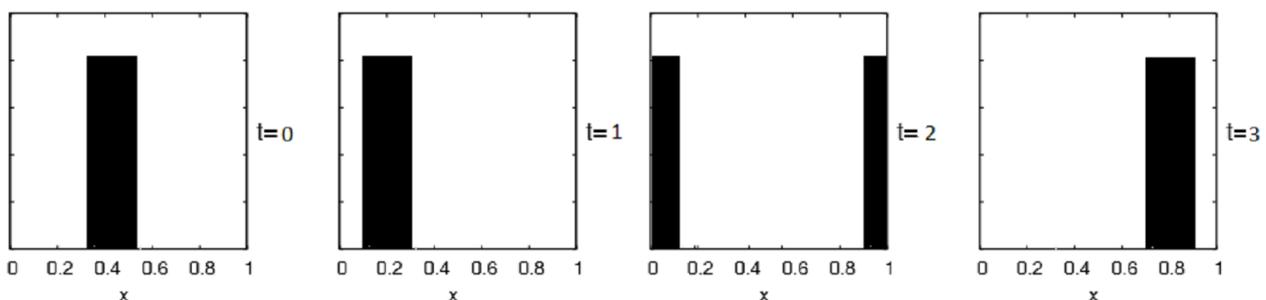


Figura 5.6: Evoluzione di una probabilità a gradino da $t = 0$. Poiché la legge di evoluzione è una traslazione rigida, la forma a gradino si mantiene ad ogni istante successivo.

Il che vuol dire che se si parte da una $\rho_0(x)$ che non è uniforme, non lo diventerà mai ad istanti successivi. Per cui il sistema *non è mixing*. Notiamo comunque una ρ_{inv} , come ad esempio quella

$\rho_{inv}(x) = 1$, a cui convergo però solo se $\rho_0 = \rho_{inv}$.

Facciamo vedere allora che il sistema è *ergodico* rispetto alla misura invariante $\rho_{inv}(x) = 1$, se $\omega \in \mathbb{R} - \mathbb{Q}$.

Ergodicità del sistema

Abbiamo capito che il sistema ammette $\rho_{inv} = 1$ e che non è mixing. Vogliamo dimostrare che è ergodico per ω irrazionale.

Consideriamo un osservabile analitica $A \in C^\infty$. Per studiare l'ergodicità, dobbiamo calcolare la media sulla misura, quella temporale e confrontarle.

Espandiamo allora A in serie di Fourier:

$$A(x) = A_0 + \sum_{n \neq 0} A_n e^{2\pi i n x}$$

Il valor medio di A sulla densità $\rho_{inv} = 1$ si scrive come:

$$\int_0^1 \rho_{inv}(x) dx A(x) = A_0$$

Per calcolare la media temporale, notiamo che:

$$x_{t+1} = x_t + \omega \implies x_t = (x_0 + \omega t) \pmod{1}$$

Quindi ottengo:

$$\frac{1}{T} \sum_{t=1}^T A(x_t) = \frac{1}{T} \sum_{t=1}^T [A_0 + \sum_{n \neq 0} A_n e^{2\pi i n x_t}] = A_0 + \sum_{n \neq 0} A_n e^{2\pi i n x_0} \left[\frac{1}{T} \sum_{t=1}^T e^{2\pi i n \omega t} \right]$$

Pongo $\alpha = e^{2\pi i n \omega}$. Devo valutare $\frac{1}{T} \sum_{t=1}^T \alpha^t$. Se ω è irrazionale ho, per ogni n , $|\alpha| < 1$ (l'esponente non è mai multiplo di 2π).

Per cui si ha

$$\frac{1}{T} \sum_{t=1}^T \alpha^t = \frac{1}{T} \frac{\alpha - \alpha^T}{1 - \alpha} \rightarrow 0, \quad \text{per } T \rightarrow \infty$$

Ma allora, per $T \rightarrow \infty$, vale

$$\frac{1}{T} \sum_{t=1}^T A(x_t) \rightarrow A_0$$

ed è verificata l'ergodicità.

Vediamo che succede invece per ω razionale. Ci sarà qualche \bar{n} per cui $e^{2\pi i \bar{n} \omega} = 1$. Per questi \bar{n} la somma sui tempi mi dà 1. Allora avrò ad ogni T :

$$\frac{1}{T} \sum_{t=1}^T A(x_t) \longrightarrow A_0 + \sum_{n \neq 0}^* A_n e^{2\pi i n x_0}$$

dove \sum^* indica la somma sugli \bar{n} discussi prima. Anche qui la media temporale converge sempre, ma a una funzione che dipende da x_0 (come predetto dal primo teorema di Birkhoff), ossia perdiamo l'ergodicità. La spiegazione fisica di questo fenomeno è semplice: se ω è irrazionale, ogni volta che il sistema inizia un nuovo giro, riempie punti nuovi. Per cui alla fine riempie densamente tutto il cerchio. Se invece è razionale, il moto è periodico e passa per un numero finito di punti. ■

5.3 Teoremi Ergodici

Sono teoremi formulati da Birkhoff e Von Neumann che danno informazioni di stampo matematico sull'ergodicità.

Teorema 5.3.1 Dato un sistema dinamico, consideriamo l'evoluzione del suo stato meccanico a partire da una condizione iniziale assegnata:

$$\mathbf{x}(0) \rightarrow \mathbf{x}(t)$$

Preso un osservabile $A(\mathbf{x})$ sufficientemente regolare, la sua media temporale converge *quasi ovunque* a una funzione di $\mathbf{x}(0)$, che chiamo \bar{A} :

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt A(\mathbf{x}(t)) = \bar{A}(\mathbf{x}(0))$$

Essendo la convergenza quasi ovunque, le funzioni al primo e secondo membro possono differire su sottoinsiemi dello spazio delle fasi a misura nulla.

Quindi l'unica informazione che abbiamo dal teorema è la convergenza della media temporale⁴.

Alla luce di questo teorema, l'eventuale proprietà di ergodicità del sistema sta nell'indipendenza di \bar{A} da $\mathbf{x}(0)$.

Teorema 5.3.2 — Teorema della transitività metrica. Si consideri ancora un sistema dinamico: il sistema è **ergodico** $\iff \exists A \subseteq \Omega$ tale che

$$\mu(A) \neq \begin{cases} 0 \\ 1 \end{cases} \quad \text{e } A \text{ è invariante (cioè } S^t A = A)$$

Ossia: non posso spezzare Ω in due suoi complementi non-banali A e A' , tale che se parto da A rimango in A e se rimango in A' rimango in A' (vedi fig. 5.7).

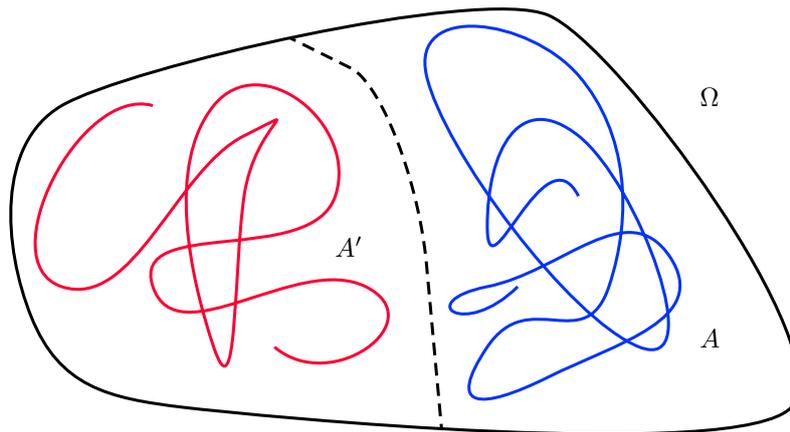


Figura 5.7: Divido Ω in due sottoinsiemi A e A' , tali che le traiettorie che partono da uno dei due sottoinsiemi vi restano dentro.

Dimostrazione. Dimostriamo il teorema solo da sinistra verso destra.

Assumiamo che esista un A che rispetti le ipotesi. Se trovo una funzione per cui la media temporale

⁴Fisicamente questa ce l'aspettiamo sempre, per cui il teorema risulta poco interessante in questo senso.

non coincide con quella sullo spazio delle fasi, ho negato l'ipotesi di ergodicità e la tesi è verificata. Considero la *funzione caratteristica di A*:

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

So che se parto da A rimango in A ad ogni istante successivo, e idem se parto fuori da A ne rimango fuori. Ma allora vale che:

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

Viceversa, se considero la media sullo spazio delle fasi ho che:

$$\int d\mu(\mathbf{x}) f(\mathbf{x}) = \mu(A)$$

I due risultati sono diversi se $\mu(A) \neq 0$ e $\mu(A) \neq 1$. Per cui violo l'ipotesi di ergodicità.

Da questo teorema si può anche far vedere che ogni sistema *mixing* è anche *ergodico*. Si consideri infatti un sistema mixing e sia $A \subseteq \Omega$ un qualsiasi insieme invariante: $A = S^{-t}A$.

Sia $B = A$.

Per definizione di mixing vale: $\mu(A \cap S^{-t}A) \rightarrow \mu(A)\mu(A)$

Ma allora, essendo A invariante, vale:

$$\mu(A) = \mu(A)\mu(A) \implies \mu(A) = 0 \text{ oppure } \mu(A) = 1$$

Ma allora è verificata l'ipotesi destra del teorema e il sistema risulta essere ergodico. ■

5.4 Ergodicità e frequenza

Notiamo che l'ergodicità ha anche a che fare con la *frequenza* con cui il sistema passa in un sottoinsieme dello spazio delle fasi. Per vederlo, si consideri un sottoinsieme B nello spazio delle fasi e la sua funzione caratteristica

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

Come già visto abbiamo che, per definizione

$$\int \rho_m(x) dx A(x) = \mu(B)$$

L'integrale temporale di $A(x(t))$ è per definizione il tempo che $x(t)$ passa in B . Per cui

$$\frac{1}{T} \int_0^T dt A(x(t)) = f_T(B)$$

è la frazione di tempo che il sistema passa in B su un tempo totale T . Se il sistema è ergodico le due medie coincidono. Per cui un altro modo per enunciare l'ergodicità, per un sistema del genere, è il seguente:

"Se il sistema è ergodico, su tempi lunghi la probabilità di stare in un volume B è uguale alla frequenza con cui la traiettoria passa per il volume." // In questo senso l'ergodicità, che non è strettamente vera per tutti i sistemi hamiltoniani, è irrinunciabile poiché ci dice che se si osserva il sistema abbastanza a lungo lo si può studiare in termini probabilistici. Con questo enunciato l'ergodicità può essere intesa come una generalizzazione della *legge dei grandi numeri*: mentre la seconda è una semplice generalizzazione

della disuguaglianza di Chebichev per variabili indipendenti, questa vale anche per variabili dipendenti (prese in situazioni deterministiche)! Inoltre, un sistema Hamiltoniano *unidimensionale* è sempre periodico (ha una traiettoria unidimensionale che si muove in uno spazio delle fasi unidimensionale). A partire da questo è possibile dimostrare che è sempre ergodico. In due dimensioni questo risultato non è più vero. Con questo approccio è comunque possibile studiare le proprietà di ergodicità di un sistema a partire da principi primi (ossia le *equazioni di Hamilton*), anche se è scomodo nella maggior parte dei casi.

Riferimenti bibliografici

- [1] M. Cencini, F. Cecconi e A.Vulpiani. *Chaos: From Simple Models to Complex Systems*. World Scientific, 2009.

6. Termodinamica di non equilibrio

Tra gli interessi del giovane Einstein ci fu la ricerca di un'evidenza sperimentale convincente dell'esistenza degli atomi: ci riuscì con la scoperta del moto Browniano (1905), ma già in precedenza produsse alcuni articoli in cui si concentrò sul problema delle **fluttuazioni**. Per molto tempo queste ultime erano di fatto passate inosservate (anche agli occhi di Boltzmann e Gibbs) perché considerate "trascurabili".

■ **Esempio 6.1** Ricordiamo la relazione (valida nell'ensemble canonico) che lega le fluttuazioni dell'energia ed il **calore specifico** c_v :

$$\langle (E - \langle E \rangle)^2 \rangle = k_B T^2 c_v, \quad (6.1)$$

Se il sistema è composto da N particelle, introducendo l'energia per particella $e = \frac{E}{N}$ possiamo riscrivere l'eq. (6.1) come

$$N^2 \langle (e - \langle e \rangle)^2 \rangle = N k_B T^2 c_v^{(1)}, \quad (6.2)$$

con $c_v^{(1)}$ il calore specifico di singola particella. La fluttuazione di energia per particella è perciò

$$\langle (e - \langle e \rangle)^2 \rangle = \frac{k_B T^2 c_v^{(1)}}{N} \quad (6.3)$$

che riproduce il ben noto risultato secondo cui $\sigma(e) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$. Potremmo concludere (ingenuamente) che le fluttuazioni non contino nulla per grandi N e che l'importante sia il valor medio $\langle e \rangle$.

Einstein manifestò invece grande interesse per questa relazione, che ammette una diversa interpretazione: se riuscissi a misurare le fluttuazioni, noto $c_v^{(1)}$ potrei risalire alla costante fondamentale k_B . ■

Evidenziamo l'analogia concettuale con il problema del moto browniano, in cui il coefficiente di diffusione definito come:

$$D = \lim_{t \rightarrow \infty} \frac{\langle x^2(t) \rangle}{2t} = \frac{k_B T}{6\pi\eta a} \quad (6.4)$$

rappresenta a sua volta una misura delle fluttuazioni.

Tali fluttuazioni sono osservabili, benché indirettamente. Ricordando che il calore specifico è dato da

$$c_v^{(1)} = \frac{d\langle e \rangle}{dT} \quad (6.5)$$

notiamo come tale quantità sia misurabile, in Termodinamica, attraverso l'osservazione di quantità macroscopiche: abbiamo così collegato fluttuazioni su scala microscopica con la variazione di grandezze macroscopiche. Perciò, se è vero che le fluttuazioni sono piccole, il coefficiente davanti a $\frac{1}{N}$ è tuttavia una quantità di interesse. Non si tratta chiaramente di una peculiarità dell'energia: quello visto è soltanto un esempio appartenente alla classe delle relazioni di **fluttuazione e dissipazione**.

L'eq. (6.5) ci sta dicendo come si modifica l'energia (il cui valor medio ipotizziamo valere $\langle e \rangle$ a una data temperatura T) in risposta a una variazione di temperatura

$$\begin{aligned} T &\rightarrow T + \delta T \\ \langle e \rangle &\rightarrow \langle e \rangle + \langle \delta e \rangle \end{aligned} \quad (6.6)$$

e questa modifica dell'energia (che ha a che fare con la fluttuazione) è legata a quella della temperatura T tramite

$$\langle \delta e \rangle = c_v^{(1)} \delta T . \quad (6.7)$$

Siamo di fronte a una delle idee centrali della Meccanica Statistica del non equilibrio: *per capire come fluttua un sistema, osservo come quest'ultimo risponde a una perturbazione* e, viceversa, *per sapere come il sistema risponde a una perturbazione, basta osservarne le fluttuazioni*. La potenza di questo concetto sta nella possibilità di conoscere la risposta del sistema anche senza eseguire davvero la perturbazione.

Einstein prese le mosse da una semplice osservazione:

$$S = k_B \ln W \quad \Rightarrow \quad W = e^{\frac{S}{k_B}} \quad (6.8)$$

(*principio di Einstein-Boltzmann*). In senso matematico le due forme sono equivalenti, ma fisicamente abbiamo completamente ribaltato il punto di vista. L'entropia S è una quantità termodinamica, mentre W è inerente alla probabilità; le due versioni corrispondono quindi a due diversi modi di procedere:

1. Date le configurazioni microscopiche, posso calcolare l'entropia.
2. L'entropia come oggetto macroscopico *definisce* le probabilità: nota l'entropia, risalgo quindi alle probabilità.

Ci proponiamo perciò di scrivere una relazione del tipo

$$P(\dots) \simeq e^{\frac{S(\dots)}{k_B}} \quad (6.9)$$

6.1 Teoria delle fluttuazioni di Einstein

Consideriamo un sistema a N gradi di libertà descritto da $\underline{x} = (\underline{q}, \underline{p}) \in \mathbb{R}^{6N}$ e supponiamo di avere accesso alle osservabili macroscopiche $\alpha_1 \dots \alpha_N$, ciascuna funzione delle variabili microscopiche attraverso $\alpha_n = g_n(\underline{x})$ (ad esempio l'energia, la densità, il numero di particelle in una data regione, la magnetizzazione...). Poniamoci nel canonico, dove possiamo definire la distribuzione

$$P(\underline{x}, \mathcal{P}) = \frac{e^{-\beta H(\underline{x})}}{z(\mathcal{P})} = e^{-\beta H(\underline{x})} e^{\beta F(\mathcal{P})} \quad (6.10)$$

che dipende, oltre che dalle variabili \underline{x} , da un set¹ di parametri \mathcal{P} . Nella eq. (6.10) si è fatto uso dell'energia libera F , che soddisfa all'equilibrio la relazione

$$z(\mathcal{P}) = e^{-\beta F(\mathcal{P})} \quad (6.11)$$

¹Nel Canonico avremo $\mathcal{P} = (T, N, V)$, mentre ad esempio nel Gran Canonico sarà $\mathcal{P} = (T, \mu, V)$.

dove $z(\mathcal{P})$ è la funzione di partizione.

Ricaviamo ora la distribuzione di probabilità delle variabili macroscopiche $\underline{\alpha}$. Trattandosi di funzioni delle \underline{x} , con passaggi elementari si ottiene²

$$\begin{aligned} P(\underline{\alpha}) &= P(\alpha_1 \dots \alpha_n) \\ &= \int P(\underline{x}, \mathcal{P}) \prod_{i=1}^n \delta(\alpha_i - g_i(\underline{x})) d\underline{x} \\ &= e^{\beta F(\mathcal{P})} \int e^{-\beta H(\underline{x})} \prod_{i=1}^n \delta(\alpha_i - g_i(\underline{x})) d\underline{x} . \end{aligned} \quad (6.12)$$

Riconosciamo nell'integrale che compare nell'ultimo passaggio la funzione di partizione vincolata alle $\delta(\alpha_i - g_i(\underline{x}))$; usando la eq. (6.11), possiamo scrivere

$$\begin{aligned} P(\underline{\alpha}) &= e^{-\beta[F(\underline{\alpha}|\mathcal{P}) - F(\mathcal{P})]} \\ &= e^{\frac{\delta S(\underline{\alpha})}{k_B}} \end{aligned} \quad (6.13)$$

dove si è chiamata $F(\underline{\alpha}|\mathcal{P})$ l'energia libera in cui si sono fissati i valori delle α_i . Si è inoltre interpretata la quantità

$$-\beta[F(\underline{\alpha}|\mathcal{P}) - F(\mathcal{P})] := \frac{\delta S(\underline{\alpha})}{k_B} \quad (6.14)$$

come una *differenza di entropia*³ tra la situazione ad $\underline{\alpha}$ fissate e quella di equilibrio:

$$\begin{aligned} \delta S(\underline{\alpha}) &= S(\underline{\alpha}|\mathcal{P}) - S(\mathcal{P}) \\ &= -\frac{1}{T}[F(\underline{\alpha}|\mathcal{P}) - F(\mathcal{P})] . \end{aligned} \quad (6.15)$$

Fissare le $\underline{\alpha}$ significa forzare uno spostamento dalla situazione di equilibrio: perciò la distribuzione eq. (6.13) che abbiamo ricavato descrive la fluttuazione di queste variabili. Nell'ipotesi in cui tali fluttuazioni siano piccole (ragionevole, dal momento che si tratta di quantità macroscopiche normalizzate su un grande numero di particelle e si è visto che la loro variazione è $\mathcal{O}\left(\frac{1}{N}\right)$), vogliamo ora descrivere gli spostamenti dalla condizione di equilibrio

$$\langle \alpha_i \rangle = \alpha_i^{eq} \quad (6.16)$$

tramite uno sviluppo parabolico attorno al valore massimo di S : sappiamo dalla Termodinamica che questo è assunto, appunto, in corrispondenza della situazione di equilibrio.

Notiamo *en passant* che abbiamo ritrovato, come auspicavamo nella eq. (6.9), la forma

$$P(\alpha_1 \dots \alpha_n) = \mathcal{N} e^{\frac{\delta S(\underline{\alpha})}{k_B}} , \quad (6.17)$$

²Si ricordi la regola per il cambio di variabili $y = f(x)$,

$$\rho_y(y) = \int \rho_x(x) \delta(y - f(x)) dx$$

che corrisponde a marginalizzare $\rho_x(x)$ con il vincolo $y = f(x)$.

³Si tratta davvero di un'entropia solamente nel microcanonico, ma a parte il piccolo abuso di linguaggio la si può comunque identificare con un altro potenziale termodinamico. Nel canonico, ad esempio, si ha $F = U - TS$ e la eq. (6.14) rappresenta la variazione dell'energia libera nel caso in cui la U è supposta costante (situazione ottenibile, ad esempio, scegliendo la U come una delle α_i).

dove \mathcal{N} è un'opportuna costante di normalizzazione. Calcoliamo allora

$$\begin{aligned} S(\underline{\alpha}) &\simeq S(\underline{\alpha}^{eq}) + \frac{1}{2} \sum_{i,j} \left. \frac{\partial^2 S}{\partial \alpha_i \partial \alpha_j} \right|_{\underline{\alpha}^{eq}} \delta \alpha_i \delta \alpha_j \\ &= S(\underline{\alpha}^{eq}) - \frac{1}{2} \sum_{i,j} A_{ij} \delta \alpha_i \delta \alpha_j \end{aligned} \quad (6.18)$$

dove abbiamo definito la matrice simmetrica

$$A_{ij} := - \left. \frac{\partial^2 S}{\partial \alpha_i \partial \alpha_j} \right|_{\underline{\alpha}^{eq}} \quad (6.19)$$

che sappiamo di per certo (ancora in virtù del secondo principio) trattarsi di una forma quadratica definita positiva. Otteniamo così la gaussiana multivariata:

$$P(\alpha_1 \dots \alpha_n) = \mathcal{N} e^{-\frac{1}{2k_B} \sum_{i,j} A_{ij} \delta \alpha_i \delta \alpha_j} . \quad (6.20)$$

Si può verificare che è soddisfatta la **relazione di fluttuazione**:

$$\langle \delta \alpha_i \delta \alpha_j \rangle = k_B (\hat{A}^{-1})_{ij} \quad (6.21)$$

e ricordiamo dalla definizione eq. (6.19) che \hat{A} è calcolata nei valori medi: si tratta cioè di un oggetto termodinamico e, come tale, accessibile. Notiamo ancora l'analogia con

$$\langle \delta E^2 \rangle = k_B T^2 c_v$$

in cui il membro a destra rappresenta una quantità termodinamica misurabile.

Abbiamo così sfatato il mito per cui la Termodinamica sarebbe *fenomenologica*, perché ci dà soltanto i valori medi. In realtà dice anche cose che apparentemente non dovrebbe dire, a patto che vi si aggiunga opportunamente la Meccanica Statistica: nel nostro caso essa compare nel principio eq. (6.8) di Einstein-Boltzmann che abbiamo posto alla base del ragionamento. Dalla Termodinamica si possono capire le fluttuazioni e da queste ultime la Termodinamica.

6.1.1 Dimostrazione della relazione per le fluttuazioni

Per la dimostrazione si può sfruttare il teorema di equipartizione⁴:

Teorema 6.1.1 — Teorema di equipartizione. Dato un ensemble canonico, il teorema afferma che:

$$\left\langle \alpha_i \frac{\partial S}{\partial \alpha_j} \right\rangle = -k_B \delta_{ij} \quad (6.22)$$

Dimostrazione. Si consideri la distribuzione di probabilità data dalla formula:

$$P(\underline{\alpha}) = N e^{\frac{S(\underline{\alpha})}{k_B}}$$

La definizione stessa di valor medio afferma che:

$$\left\langle \alpha_i \frac{\partial S}{\partial \alpha_j} \right\rangle = N \int \alpha_i \frac{\partial S}{\partial \alpha_j} e^{\frac{S(\underline{\alpha})}{k_B}} d\underline{\alpha} = N \int \alpha_i k_B \frac{\partial}{\partial \alpha_j} e^{\frac{S(\underline{\alpha})}{k_B}} d\underline{\alpha}$$

⁴Abbiamo studiato questo teorema nel microcanonico come: $\langle x_i \frac{\partial H}{\partial x_j} \rangle = k_B T \delta_{ij}$

Con un'integrazione per parti si ottiene:

$$\left\langle \alpha_i \frac{\partial S}{\partial \alpha_j} \right\rangle = N \alpha_i k_B e^{\frac{S(\underline{\alpha})}{k_B}} \Big|_{\text{bordi}} - N k_B \int \delta_{ij} e^{\frac{S(\underline{\alpha})}{k_B}} d\underline{\alpha}$$

dove il primo pezzo, essendo una probabilità sui bordi, è nullo mentre il secondo dà un contributo $-k_B \delta_{ij}$ poiché $N \int e^{\frac{S(\underline{\alpha})}{k_B}} d\underline{\alpha} = 1$ essendo l'insieme delle probabilità. Il risultato è vero per ogni S . ■

Procediamo quindi a dimostrare la relazione di fluttuazione:

$$\langle \alpha_i \alpha_j \rangle = k_B [A^{-1}]_{ij}$$

Dimostrazione. Per semplicità consideriamo $\alpha_n^{eq} = \langle \alpha_n \rangle = 0$ e consideriamo la relazione:

$$S(\underline{\alpha}) = -\frac{1}{2} \sum_{i,j} A_{ij} \alpha_i \alpha_j$$

possiamo calcolarne la derivata rispetto ad α_j :

$$\frac{\partial S}{\partial \alpha_j} = -\sum_n A_{nj} \alpha_n$$

a questo punto la sostituiamo nell'eq. (6.22) ottenendo:

$$\left\langle \alpha_i \frac{\partial S}{\partial \alpha_j} \right\rangle = -\sum_n A_{nj} \langle \alpha_i \alpha_n \rangle = -k_B \delta_{ij}$$

che, semplificando i meno, possiamo riscrivere come:

$$\sum_n A_{nj} \langle \alpha_i \alpha_n \rangle = k_B \delta_{ij}$$

L'unico modo per cui questa relazione risulta essere soddisfatta è che sia vera l'equazione:

$$\langle \alpha_i \alpha_n \rangle = [A^{-1}]_{in} k_B \quad (6.23)$$

■

Si può pervenire allo stesso risultato diagonalizzando le matrici. In questo caso abbiamo:

$$\sum_{i,j} A_{ij} \alpha_i \alpha_j = (\underline{\alpha}, \hat{A} \underline{\alpha}) \quad (6.24)$$

e possiamo scrivere α in un modo più opportuno secondo la relazione:

$$\underline{\alpha} = \hat{C} \underline{z}$$

A questo punto sostituiamo la relazione di α nell'eq. (6.24) ottenendo:

$$\sum_{i,j} A_{ij} \alpha_i \alpha_j = (\hat{C} \underline{z}, \hat{A} \hat{C} \underline{z}) = (\underline{z}, (\hat{C}^{-1} \hat{A} \hat{C}) \underline{z}) \quad (6.25)$$

Qui il trucco sta nel trovare la giusta matrice \hat{C} che renda diagonale \hat{A} . A partire da questa matrice diagonale con autovalori a_1, \dots, a_n , avremo:

$$\langle z_n z_{n'} \rangle = \frac{\delta_{nn'}}{a_n} k_B$$

per cui possiamo calcolare $\langle \alpha_i \alpha_j \rangle$ sapendo che $\alpha_i = \sum_n C_{in} z_n$, da cui:

$$\langle \alpha_i \alpha_j \rangle = \left\langle \sum_{nn'} C_{in} z_n C_{jn'} z_{n'} \right\rangle = k_B \sum_n \frac{C_{in} C_{jn}}{a_n} \quad (6.26)$$

L'eq. (6.26) potrebbe sembrare il prodotto di due matrici, ma facendo attenzione alla forma della matrice diagonale notiamo che:

$$A_D^{-1} = \begin{bmatrix} \frac{1}{a_1} & 0 & \dots & 0 \\ 0 & \frac{1}{a_2} & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & \frac{1}{a_n} \end{bmatrix}$$

per cui l'eq. (6.26) è il prodotto tra tre matrici:

$$\langle \alpha_i \alpha_j \rangle = k_B \sum_{n,k} C_{in} A_{nk}^{-1} C_{kj}^{-1} = k_B [C A^{-1} C^{-1}]_{ij} = k_B [A^{-1}]_{ij}$$

6.2 Teoria delle perturbazioni lineari

L'importanza dell'eq. (6.23) risiede nella connessione tra le fluttuazioni e le proprietà termodinamiche del sistema: le prime saranno calcolabili a partire dalle seconde che sono effettivamente osservabili.

In un'ottica un po' più generale consideriamo un sistema di N particelle descritto dall'hamiltoniana H_0 a cui applichiamo una perturbazione $A(x)$ legata al sistema nella seguente maniera:

$$H_0 \longrightarrow H_0 + \lambda A(x)$$

con $\lambda \ll 1$. Un esempio potrebbe essere l'applicazione di un campo magnetico h che cambia leggermente la magnetizzazione del sistema:

$$H_0 - h \sum_i m_i$$

Siamo interessati all'osservabile B i cui valori medi, per un sistema imperturbato e uno perturbato, saranno:

$$\langle B \rangle_0 \longrightarrow \langle B \rangle_\lambda$$

per cui avremo una variazione:

$$\langle \Delta B \rangle = \langle B \rangle_\lambda - \langle B \rangle_0 \quad (6.27)$$

Proseguendo nell'esempio del campo magnetico avremo $\langle m \rangle_0 = 0$, ma per calcolare $\langle m \rangle_\lambda$ necessitiamo della densità di probabilità. Per l'hamiltoniana imperturbata la densità la conosciamo:

$$\rho_0(x) = \frac{e^{-\beta H_0}}{\int e^{-\beta H_0} dx}$$

per quella perturbata avremo invece:

$$\rho_\lambda(x) = \frac{e^{-\beta H_0} e^{-\beta \lambda A}}{\int e^{-\beta H_0} e^{-\beta \lambda A} dx}$$

Poiché $\lambda \ll 1$ possiamo espandere gli esponenziali della relazione precedente ottenendo:

$$\rho_\lambda(x) = \frac{(1 - \beta\lambda A)e^{-\beta H_0}}{\int (1 - \beta\lambda A)e^{-\beta H_0} dx} \quad (6.28)$$

Il denominatore può essere riscritto come:

$$\begin{aligned} \int (1 - \beta\lambda A)e^{-\beta H_0} dx &= \int e^{-\beta H_0} dx - \beta\lambda \int A e^{-\beta H_0} dx = \\ &= \int e^{-\beta H_0} dx \left[1 - \beta\lambda \frac{\int A e^{-\beta H_0} dx}{\int e^{-\beta H_0} dx} \right] = \\ &= \int e^{-\beta H_0} dx [1 - \beta\lambda \langle A \rangle_0] \end{aligned}$$

Sostituendolo nell'eq. (6.28) otteniamo:

$$\begin{aligned} \rho_\lambda(x) &= \frac{(1 - \beta\lambda A)e^{-\beta H_0}}{\int e^{-\beta H_0} dx [1 - \beta\lambda \langle A \rangle_0]} = \\ &= \frac{1 - \beta\lambda A}{1 - \beta\lambda \langle A \rangle_0} \rho_0(x) \end{aligned}$$

Ricordando che $\frac{1}{1-\varepsilon} \simeq 1 + \varepsilon$ avremo:

$$\begin{aligned} \rho_\lambda(x) &= \frac{1 - \beta\lambda A}{1 - \beta\lambda \langle A \rangle_0} \rho_0(x) = \\ &= (1 - \beta\lambda A)(1 + \beta\lambda \langle A \rangle_0) \rho_0(x) = \\ &= [1 - \beta\lambda(A - \langle A \rangle_0) + O(\lambda^2)] \rho_0(x) \end{aligned}$$

Possiamo ora calcolare il valore dell'eq. (6.27):

$$\begin{aligned} \langle \Delta B \rangle &= \int B(\rho_\lambda - \rho_0) dx = \\ &= \int B \left[-\beta\lambda \rho_0 (A(x) - \langle A(x) \rangle_0) \right] dx = \\ &= -\beta\lambda \left[\int \rho_0(x) B(x) A(x) dx - \int \rho_0(x) B(x) \langle A \rangle_0 dx \right] \end{aligned}$$

ottenendo l'equazione che esplica la risposta del sistema ad una perturbazione esterna:

$$\langle \Delta B \rangle = -\beta\lambda \left[\langle AB \rangle_0 - \langle B \rangle_0 \langle A \rangle_0 \right] \quad (6.29)$$

L'importanza di questa relazione sta nella possibilità di capire la risposta del sistema ad una perturbazione senza tuttavia applicarla realmente. Per capire meglio l'analogia: ci basta guardare una persona in faccia per capire come reagirà a una serie di schiaffi, non è necessario prendercela realmente. Tutto ciò ha validità finché la perturbazione è piccola e "sfruguglia" poco il sistema.

■ **Esempio 6.2 — Campo magnetico.** Analizziamo il caso del campo magnetico considerando $A = B = m$ e $\lambda = -h$, per cui:

$$H_0 \longrightarrow H_0 - hm$$

Applicando la formula dell'eq. (6.29) otteniamo:

$$\langle m \rangle_h = -\beta(-h) \langle \delta m^2 \rangle_0 = \beta h \langle m^2 \rangle_0 \quad (6.30)$$

Dividendo $\langle m \rangle_h$ per $h \ll 1$ otteniamo:

$$k_B T \frac{\partial \langle m \rangle}{\partial h} = \langle m^2 \rangle_0 \quad (6.31)$$

per cui le fluttuazioni della magnetizzazione sono proporzionali alla *suscettività* identificata dalla derivata rispetto al campo magnetico della magnetizzazione calcolata nel sistema perturbato. ■

■ **Esempio 6.3 — Campo magnetico.** La relazione dell'eq. (6.31) è analiticamente molto simile alla formula per la variazione di energia in un sistema termodinamico:

$$\langle \delta E^2 \rangle = k_B T^2 \frac{\partial \langle E \rangle}{\partial T} \quad (6.32)$$

questo perché, in realtà, sono la stessa formula, sono due casi particolari dell'eq. (6.29). Per il caso dell'energia, ponendo sempre $A = B = E$, avremo:

$$\beta H_0 \longrightarrow (\beta + \delta\beta) H_0$$

Scriviamo ora $\langle \Delta B \rangle$ come la variazione di energia $\langle \delta E \rangle$ e sostituiamo nell'eq. (6.29):

$$\langle \delta E \rangle = -\beta \delta\beta \langle \delta E^2 \rangle \quad (6.33)$$

Come nel caso de campo magnetico, per $\delta\beta \ll 1$ avremo:

$$\frac{\partial \langle E \rangle}{\partial \beta} = -\beta \langle \delta E^2 \rangle \quad (6.34)$$

Si può notare che la derivata può essere riscritta come:

$$\frac{\partial}{\partial \beta} = \frac{\partial}{\partial T} \frac{\partial T}{\partial \beta} = \frac{\partial}{\partial T} \frac{\partial}{\partial \beta} \left(\frac{1}{k_B \beta} \right) = -\frac{\partial}{\partial T} \frac{1}{k_B \beta^2} = -\frac{\partial}{\partial T} k_B T^2 \quad (6.35)$$

In entrambi i casi analizzati siamo partiti da una situazione di equilibrio e siamo finiti di nuovo all'equilibrio. Nel caso in cui invece considerassimo un campo elettrico, questo implicherebbe una corrente e dunque non avremmo più l'equilibrio. Un'altra cosa di cui bisogna tenere conto è che in queste relazioni non abbiamo il tempo: aspettiamo semplicemente che il sistema ritorni all'equilibrio. Tutte queste relazioni culminano nel moto browniano con cui riusciamo a calcolare k_B . Intorno al 1930 il fisico norvegese Onsager introdusse il tempo in questi processi. ■

6.3 Relazioni di Onsager

Ripartiamo dalla formula di Einstein per la probabilità di trovare un set di variabili macroscopiche $\underline{\alpha}$ a un certo valore con $\underline{\alpha}_{eq} = 0$, senza perdere di generalità.

$$P(\underline{\alpha}) = \mathcal{N} e^{\frac{S(\underline{\alpha})}{k_B}} \simeq \mathcal{N} e^{-\frac{1}{2k_B} \sum_{ij} g_{ij} \alpha_i \alpha_j} \quad (6.36)$$

Dove $g_{ij} = -\frac{\partial^2 S(\underline{\alpha})}{\partial \alpha_i \partial \alpha_j} |_{eq}$ costituiscono gli elementi della matrice \hat{g} .

Definiamo quella che i chimici chiamano l'*affinità* di (mentre i fisici *campo coniugato* a) α_n come $X_n = \frac{\partial S(\underline{\alpha})}{\partial \alpha_n}$ così da poter riscrivere

$$dS = \sum_n X_n d\alpha_n \quad (6.37)$$

capiamo dunque chi sono le X_n dal confronto della eq. (6.37) con il differenziale dell'entropia (ad es.) nell'ensemble microcanonico dove $S(E, V, \{n_k\})$ (con n_k numero di molecole nella k -esima specie molecolare che compone il sistema) e cioè $\alpha_1 = E$, $\alpha_2 = V$, $\alpha_k = n_k$ ($k > 2$)

$$dS = \frac{1}{T}dE + \frac{P}{T}dV + \sum_i \frac{\mu_i}{T}dn_i \quad (6.38)$$

e allora risulta $X_1 = \frac{1}{T}$, $X_2 = \frac{P}{T}$, $X_k = \frac{\mu_k}{T}$ ($k > 2$).

L'obiettivo è ora quello di scrivere delle equazioni differenziali stocastiche che descrivano l'evoluzione delle $\underline{\alpha}$ del tipo

$$\frac{d\underline{\alpha}}{dt} = \underline{F}(\underline{\alpha}) + \hat{b}\underline{\eta} \quad (6.39)$$

Il sogno sarebbe quello di ricavare delle equazioni del tipo eq. (6.39) a partire dai principi primi (Hamiltoniana del sistema). Non riuscendovi partiamo direttamente scrivendo delle equazioni della forma eq. (6.39), utilizzando la fisica del sistema per porre dei vincoli su $\underline{F}(\underline{\alpha})$ e \hat{b} . Ci mettiamo sotto la stessa ipotesi di Einstein di piccole fluttuazioni e quindi ci aspettiamo che la soluzione stazionaria della Fokker-Planck associata alle eq. (6.39) sia proprio eq. (6.36), cioè una Gaussiana multivariata: soluzione, in generale, di una Fokker-Planck lineare. Dunque, essendo le $\underline{\alpha}$ piccole (e nulle all'equilibrio), possiamo approssimare

$$\underline{F}(\underline{\alpha}) \simeq -\hat{M}\underline{\alpha} \quad (6.40)$$

e le eq. (6.39) saranno del tipo $\frac{d\underline{\alpha}_n}{dt} = -\sum_j M_{nj}\alpha_j + \sum_j b_{nj}\eta_j$ con Fokker-Planck associata

$$\frac{\partial P(\underline{\alpha})}{\partial t} = \sum_{nj} M_{nj} \frac{\partial}{\partial \alpha_n} (\alpha_j P(\underline{\alpha})) + \frac{1}{2} \sum_{nj} \left(\sum_k b_{nk} b_{kj} \right) \frac{\partial^2 P(\underline{\alpha})}{\partial \alpha_j \partial \alpha_n} \quad (6.41)$$

che ha come soluzione una Gaussiana multivariata. Un primo vincolo sulla \hat{M} e sulla \hat{b} lo pone la richiesta che la Gaussiana sia proprio eq. (6.36). Ma questa condizione sulla probabilità di equilibrio non basta a definire univocamente la dinamica stocastica - time dependent - delle $\underline{\alpha}$, cioè le equazioni eq. (6.39). Tuttavia possiamo porre qualche vincolo ulteriore, che ha a che fare con le proprietà dinamiche del sistema.

Usiamo il fatto che $\underline{X} = -\hat{g}\underline{\alpha}$ poi, essendo \hat{g} simmetrica (\rightarrow invertibile) $\underline{\alpha} = -\hat{g}^{-1}\underline{X}$. Allora definiamo la matrice di Onsager $\hat{L} = \hat{M}\hat{g}^{-1}$. Con questa nuova notazione scriviamo

$$\frac{d\underline{\alpha}}{dt} = \hat{L}\underline{X} + \hat{b}\underline{\eta} \quad (6.42)$$

Quindi ci siamo riscritti i flussi $\underline{J} = \hat{L}\underline{X}$ come combinazione lineare delle affinità. Cerchiamo ora i vincoli che la fisica impone sulla matrice \hat{L} .

Il problema originale (deterministico) è **time-reversal**. Nell'ambito stocastico quanto più si avvicina a questa condizione è il bilancio dettagliato, ma, nel caso di un problema gaussiano come quello in analisi, basta richiedere che valga

$$\langle \alpha_i(t)\alpha_j(t + \Delta t) \rangle = \epsilon_1\epsilon_2 \langle \alpha_j(t)\alpha_i(t + \Delta t) \rangle \quad (6.43)$$

che, tramite iterazione, può essere estesa a un generico intervallo temporale τ . $\epsilon_1\epsilon_2 = \pm 1$ a seconda che nel sistema siano o meno coinvolti campi magnetici. Infatti, quando è presente un campo magnetico,

la time reversibility vale a patto di invertire - oltre alle velocità - il segno del campo. La dinamica stocastica gaussiana è completamente determinata dai valori medi delle $\underline{\alpha}$ (che già sappiamo essere nulli) e dalle correlazioni a coppie tra di esse. Quindi richiedere la eq. (6.43) implicherà che valga il bilancio dettagliato.

Teorema 6.3.1 L'eq. (6.43) è soddisfatta se $L_{ij} = \epsilon_i \epsilon_j L_{ji}$.

Dimostrazione. Partiamo da

$$\left\langle \alpha_i \frac{\partial S}{\partial \alpha_j} \right\rangle = \langle \alpha_i X_j \rangle = -k_B \delta_{ij} \quad (6.44)$$

L'idea è quella di sviluppare, partendo dalla eq. (6.42), $\alpha_i(t + \Delta t) = \alpha_i(t) + \sum_k L_{ik} X_k(t) \Delta t + \sum_k b_{ik} \eta_k(t) \Delta t + O(\Delta t^2)$ e poi di tenere conto che $\langle \eta_k \rangle = 0 \forall k$. Per questa ragione eviteremo di considerare la parte di rumore nei calcoli che seguono.

$$\begin{aligned} \langle \alpha_i(t) \alpha_j(t + \Delta t) \rangle &\simeq \left\langle \alpha_i(t) \left[\alpha_j(t) + \sum_k L_{jk} X_k(t) \Delta t \right] \right\rangle = \\ &= \langle \alpha_i(t) \alpha_j(t) \rangle + \Delta t \sum_k L_{jk} \langle \alpha_i(t) X_k(t) \rangle = \\ &= \langle \alpha_i(t) \alpha_j(t) \rangle + \Delta t \sum_k L_{jk} (-k_B \delta_{ik}) = \\ &= \langle \alpha_i(t) \alpha_j(t) \rangle - k_B \Delta t L_{ji} \end{aligned} \quad (6.45)$$

Analogamente si trova, sviluppando $\alpha_i(t + \Delta t)$, $\langle \alpha_i(t + \Delta t) \alpha_j(t) \rangle \simeq \langle \alpha_i(t) \alpha_j(t) \rangle - k_B \Delta t L_{ij}$.

Quindi, uguagliando le due espressioni appena trovate e semplificando il termine comune, si ha che la eq. (6.43) risulta verificata solo se $L_{ij} = \epsilon_i \epsilon_j L_{ji}$, che è proprio la nostra tesi. ■

Questo risultato è importate perché, sebbene non ci dica come siano fatte le eq. (6.42) (e quindi le eq. (6.39)), quantomeno pone delle condizioni ulteriori sui modelli che possiamo inventare: se non è soddisfatto che $L_{ij} = L_{ji}$ (\hat{L} simmetrica) sicuramente il modello è sbagliato. Capiamo infine anche il perché della notazione introdotta con le affinità \underline{X} : queste permettono di riscrivere comodamente ⁵ $\hat{M}\underline{\alpha} = -\hat{L}\underline{X}$ poiché la time-reversibility impone condizioni di simmetria su \hat{L} e non su \hat{M} .

6.3.1 Connessione tra Onsager e Bilancio Dettagliato

Abbiamo fino ad ora studiato la connessione tra le relazioni di Onsager e l'inversione temporale; sappiamo, inoltre, che esiste una relazione tra quest'ultima ed il bilancio dettagliato: siamo interessati a studiare, ora, la connessione tra la relazione di Onsager ed il bilancio dettagliato.

Partiamo da una relazione dinamica tra le due quantità discrete n_1 ed n_2 : per avere un'idea più 'concreta' di quello che stiamo facendo, possiamo visualizzare tali valori come il numero di particelle presenti in due regioni dello spazio o, equivalentemente, come due specie chimiche presenti all'interno dello stesso volume.

$$n_1, n_2 \rightarrow (n_1 \pm 1), (n_2 \pm 1) \quad (6.46)$$

formalmente, si tratta di processi descritti da equazioni differenziali di tipo Master Equation:

$$\frac{dP_n}{dt} = \sum_{n' \neq n} T_{n' \rightarrow n} P_{n'} - c_n P_n + \text{rumore} \quad (6.47)$$

⁵Matematicamente la trattazione si potrebbe fare anche senza introdurre le affinità ma per ragioni storiche e di compattezza si fa questa scelta

Possiamo leggere questa equazione come una equazione stocastica per le variabili n . La relazione di bilancio dettagliato impone che:

$$P_n^s T_{n \rightarrow n'} = P_{n'}^s T_{n' \rightarrow n} \quad (6.48)$$

Vogliamo mostrare ora che la relazione di Onsager che descrive questo sistema è proprio il Bilancio Dettagliato. Partiamo con lo scrivere l'entropia del sistema: suppongo di avere m specie di particelle

$$n_1, \dots, n_m \quad | \quad \sum_{j=1}^m n_j = N \quad (6.49)$$

definisco, dunque, la probabilità che la particella del tipo i -esimo si trovi in uno stato dato P_1, \dots, P_n ed assumo che il numero di particelle del sistema appartenenti ad una data specie sia distribuito secondo una probabilità multinomiale. La probabilità del sistema $P(n_1, \dots, n_m)$ definisce, dunque, l'entropia del sistema stesso.

$$P(n_1, \dots, n_m) = \frac{N!}{\prod_j n_j!} \prod_{j=1}^m P_j^{n_j} = e^{\frac{S(n_1, \dots, n_m)}{k_B}} \quad (6.50)$$

Procediamo, ora, calcolando il valore di S intorno al suo massimo, in corrispondenza di n_j^{eq} . Al fine di valutare questa quantità introduciamo una variabile α definita in modo tale che

$$\alpha_j = n_j - n_j^{eq}$$

Sviluppamo i conti necessari per ricavare l'espressione dell'entropia:

$$\begin{aligned} \ln P(n_1, \dots, n_m) &= \frac{S(n_1, \dots, n_m)}{k_B} = \\ &= \ln N! + \sum_j n_j \ln P_j - \sum_j \ln n_j! = \\ &= N \ln N - N + \sum_j n_j \ln P_j - \sum_j n_j \ln n_j - \sum_j n_j = \\ &= - \sum_j n_j \ln \frac{n_j}{NP_j} \end{aligned} \quad (6.51)$$

Dove il logaritmo naturale del fattoriale è stato approssimato dalla formula di Stirling. Abbiamo dunque ottenuto che l'entropia del sistema è definita come:

$$\frac{S(n_1, \dots, n_m)}{k_B} = - \sum_j n_j \ln \frac{n_j}{NP_j}$$

Ricordando, ora, che il valore di n_j all'equilibrio è pari a n_j^{eq} , e che il valor atteso della multinomiale è proprio NP_j , si ottiene:

$$\frac{S(n_1, \dots, n_m)}{k_B} = - \sum_j n_j \ln \frac{n_j}{n_j^{eq}} \quad (6.52)$$

Possiamo fare un primo check di consistenza verificando che il valore massimo dell'entropia, $S = 0$, si ottiene per $n_j = n_j^{eq}$. Voglio capire ora come si comporta l'entropia del sistema per uno spostamento di n_j dal suo valore all'equilibrio: per fare ciò introduco le variabili α in modo tale che:

$$n_j = n_j^{eq} + \alpha_j$$

Sostituendo nella eq. (6.52) ottengo

$$\begin{aligned}
\frac{S(n_1, \dots, n_m)}{k_B} &= - \sum_j (n_j^{eq} + \alpha_j) \ln \left(\frac{n_j^{eq}}{n_j^{eq}} + \frac{\alpha_j}{n_j^{eq}} \right) = \\
&= - \sum_j (n_j^{eq} + \alpha_j) \ln \left(1 + \frac{\alpha_j}{n_j^{eq}} \right) = \\
&= - \sum_j (n_j^{eq} + \alpha_j) \left[\frac{\alpha_j}{n_j^{eq}} - \frac{\alpha_j^2}{2 \cdot (n_j^{eq})^2} + \mathcal{O}(\alpha_j^3) \right] = \\
&= - \sum_j \left(\alpha_j + \frac{1}{2} \frac{\alpha_j^2}{n_j^{eq}} + \mathcal{O}(\alpha_j^3) \right)
\end{aligned}$$

Essendo $\sum_j \alpha_j = 0$ ed essendo $S_{eq} = 0$ si ottiene:

$$S = \delta S = -k_B \sum_j \frac{\alpha_j^2}{2 \cdot n_j^{eq}}$$

Ricordando che il campo X_j coniugato ad α_j è definito come:

$$X_j = \frac{\partial S}{\partial \alpha_j} = -k_B \frac{\alpha_j}{n_j^{eq}} \quad (6.53)$$

Proviamo ora a scriverci l'equazione per α

$$\frac{d\alpha_j}{dt} = \sum T_{n \rightarrow j} \alpha_n + \text{rumore} \quad (6.54)$$

Posso riscrivere questa relazione nella forma di Onsager, dove non tengo conto del termine dovuto al rumore essendo interessati unicamente ai termini non diagonali. Nella forma di Onsager l'eq. (6.54) diventa

$$\frac{d\alpha_j}{dt} = \sum L_{jn} X_n$$

Ricordando, però, la relazione eq. (6.53)

$$\frac{d\alpha_j}{dt} = - \sum T_{n \rightarrow j} n_n^{eq} \frac{X_n}{k_B}$$

Identifico i coefficienti di Onsager con i valori di equilibrio: ricordando ora che $n_j^{eq} = NP_j^{eq}$ e sostituendo otteniamo:

$$L_{jn} = - \frac{1}{k_B} T_{n \rightarrow j} n_n^{eq} = - \frac{N}{k_B} T_{n \rightarrow j} P_n^{eq}$$

Il bilancio dettagliato in questo modello di reazione è equivalente all'inversione temporale. Abbiamo dunque che la relazione di Onsager implica la relazione di bilancio dettagliato.

6.3.2 Teoria di Onsager-Machlup

Definiamo il vettore α descrivente l'evoluzione temporale dello stato, continuo, di un sistema a tempi discreti che, per semplicità, supporremo equidistanti:

$$\dot{\alpha} = f(\alpha) + \hat{b}\eta$$

Supponiamo di partire dalla posizione α_0 considerare il set di tempi discreti t_1, \dots, t_N in modo tale che la posizione del corpo al tempo i -esimo sia compresa tra $(\alpha_i; \alpha_i + \delta\alpha)$. Voglio trovare l'espressione della densità di probabilità di una data traiettoria. Riconosciamo una evidente analogia con l'analisi di un processo markoviano in tempo discreto in cui la probabilità di avere una successione di stati S_0, S_1, \dots, S_N è data da

$$\begin{aligned} P(S_0, \dots, S_N) &= P(S_0)P(S_0 \rightarrow S_1) \cdots P(S_{N-1} \rightarrow S_N) = \\ &= P(S_0) \cdot \prod_{t=1}^{N-1} P(S_t \rightarrow S_{t+1}) \end{aligned} \quad (6.55)$$

Procedendo in modo completamente analogo possiamo ricavare l'espressione della probabilità della traiettoria $\alpha_0, \dots, \alpha_N$ nel problema in esame:

$$\begin{aligned} P(\alpha_0, \dots, \alpha_N) &= P(\alpha_0)P(\alpha_0 \rightarrow \alpha_1) \cdots P(\alpha_{N-1} \rightarrow \alpha_N) = \\ &= P(\alpha_0) \cdot \prod_{t=1}^{N-1} P(\alpha_t \rightarrow \alpha_{t+1}) \end{aligned} \quad (6.56)$$

Caso unidimensionale

Partiamo dalla analisi del problema nel caso unidimensionale: in questo caso il sistema è descritto dalla relazione

$$\dot{\alpha} = -\gamma\alpha + \sqrt{2C}\eta$$

sappiamo, dunque, che la variabile α al tempo t_1 è un processo stocastico che dipende in modo lineare da α_0 sommato ad un termine di rumore:

$$\alpha(t_1) = \alpha_1 = \alpha_0 e^{-\gamma\tau} + \sigma_1 w_1$$

Dove $\alpha_0 e^{-\gamma\tau}$ è proprio il valor medio di α_1 dove $\tau = t_1 - t_0$, σ_1 ne rappresenta, invece, la varianza moltiplicata per w_1 che è un numero gaussiano con media nulla e varianza unitaria. Sappiamo, ora che

$$\sigma_1^2 = \langle \alpha^2 \rangle (1 - e^{-2\gamma\tau})$$

da cui ricaviamo

$$\sigma_1^2 \sim \langle \alpha^2 \rangle (1 - 1 + 2\gamma\tau) = 2 \langle \alpha^2 \rangle \gamma\tau \quad (6.57)$$

Supponiamo ora per semplicità che τ sia piccolo così da poter espandere l'esponenziale al primo ordine ottenendo:

$$\alpha_1 \sim \alpha_0 (1 - \tau\gamma) + \sqrt{2 \langle \alpha^2 \rangle \gamma\tau} w$$

Essendo la densità di probabilità di passare da uno stato α_0 ad uno stato α_1 un processo gaussiano, risulta:

$$P(\alpha_0 \rightarrow \alpha_1) \propto e^{-w^2/2} \propto e^{-\frac{(\alpha_1 - \alpha_0 + \alpha_0\gamma\tau)^2}{4 \langle \alpha^2 \rangle \gamma\tau}} \quad (6.58)$$

Estendendo tali relazioni ad ogni coppia di tempi successivi in modo ricorsivo possiamo ricavare la densità di probabilità di una traiettoria:

$$P(\alpha_0, \dots, \alpha_N) \propto P(\alpha_0) e^{-\frac{\sum_j (\alpha_j - \alpha_{j-1} + \alpha_{j-1}\gamma\tau)^2}{4 \langle \alpha^2 \rangle \gamma\tau}}$$

Osserviamo ora che:

$$\alpha_j - \alpha_{j-1} + \alpha_{j-1}\gamma\tau = \left(\frac{\alpha_j - \alpha_{j-1}}{\tau} + \alpha_{j-1}\gamma \right) \tau$$

Ricordando inoltre che $\langle \alpha^2 \rangle = C/\gamma$ la densità di probabilità di una traiettoria diventa:

$$P(\alpha_0, \dots, \alpha_N) \propto P(\alpha_0) e^{-\frac{\sum_j \tau (\frac{\alpha_j - \alpha_{j-1}}{\tau} + \alpha_{j-1}\gamma)^2}{4C}}$$

Osserviamo ora che

$$P(\alpha_0, \dots, \alpha_N) \rightarrow P(\alpha_0) e^{-\frac{1}{4C} \int_{t_0}^{t_f} (\dot{\alpha}(t) + \alpha(t)\gamma)^2 dt}$$

ma, ricordando che $\dot{\eta} \propto \dot{\alpha} + \gamma\alpha$, abbiamo ottenuto che la densità di probabilità di avere un dato percorso dipende da η^2 . Facciamo ora due osservazioni: in primo luogo ciò che abbiamo ricavato è l'espressione di una densità di probabilità, per ottenere la probabilità della traiettoria è necessario integrare sulle $\delta\alpha$; in secondo luogo osserviamo che la relazione trovata è puramente formale non essendoci noi interessati delle divergenze dei coefficienti. Mandando τ a zero e definendo opportunamente una funzione Lagrangiana L possiamo riassumere i risultati trovati nella relazione:

$$P(\{\alpha(t)\}, t_0 < t < t_f) \propto e^{-\frac{1}{4C} \int_{t_0}^{t_f} L(\dot{\alpha}, \alpha) dt} \quad (6.59)$$

formalmente simile alla probabilità di avere un dato stato nota l'hamiltoniana del sistema. Ad un risultato molto simile giunge Feynman pochi anni prima nell'ambito della teoria dei campi attraverso la definizione di 'integrali su i cammini'.

Riferimenti bibliografici

- [1] M. Falcioni e A. Vulpiani. *Meccanica Statistica Elementare*. Springer-Verlag, 2014.
- [2] H.B.G. Casimir. "On Onsager's Principle of Microscopic Reversibility". In: *Rev. Mod. Phys.* 17.343 (1945).
- [3] S.R. de Groot e P. Mazur. *Non-equilibrium Thermodynamics*. Dover, 1984.

7. Teoremi di fluttuazione

L'obiettivo della teoria della risposta lineare è quello di capire alcune proprietà di non equilibrio partendo da informazioni sulla situazione di equilibrio. L'esempio classico è quello di un circuito elettrico: si vuole sapere quanta corrente passa attraverso una resistenza ai cui capi è applicata una certa differenza di potenziale senza realmente applicarla. L'ipotesi fondamentale è la conoscenza completa di quello che riguarda la situazione imperturbata (anche in funzione del tempo), ad esempio le funzioni di correlazione. Ci chiediamo quindi quali informazioni possiamo trarre sulla situazione perturbata; un problema del genere è molto importante, in quanto ci permetterebbe, nel caso dell'esempio del circuito elettrico, di ricavare la conducibilità elettrica dalle proprietà microscopiche del sistema.

7.1 Sistemi lineari

Partiamo dall'equazione di Langevin lineare unidimensionale:

$$\dot{v} = -\frac{v}{\tau} + \sqrt{2c\eta} \quad (7.1)$$

di cui conosciamo $\langle v(t) \rangle_0 = 0$ e $\langle v(t)v(0) \rangle_0 = \langle v^2 \rangle_0 e^{-\frac{t}{\tau}}$. Si è posto l'accento con il pedice $\langle \cdot \rangle_0$ sul fatto che i valori medi sono fatti sul sistema imperturbato, cioè $\langle (\cdot) \rangle_0 = \int (\cdot) P_S(v) dv$, con $P_S(v)$ distribuzione di probabilità stazionaria del sistema imperturbato. Il passo successivo è perturbarlo; l'equazione da studiare diventa quindi:

$$\dot{v} = -\frac{v}{\tau} + \sqrt{2c\eta} + \epsilon F(t) \quad (7.2)$$

con ϵ parametro piccolo (la richiesta che sia piccolo nei casi lineari come questo è irrilevante) e in cui si è posta la massa $m = 1$, altrimenti sarebbe stato $\frac{F(t)}{m}$.

Il campo $F(t)$ può avere una dipendenza più o meno complicata nel tempo, generalmente assumeremo che sia una $\theta(t)$, cioè che sia acceso al tempo $t = 0$ e che poi sia una costante F per $t > 0$. Ci aspettiamo quindi che una volta acceso il campo ci sia un passaggio da $\langle v(t) \rangle_0 = 0$ a $\langle v(t) \rangle_\epsilon$ incognito, dove il pedice ϵ sta ad indicare che la media è fatta sul sistema perturbato.

Per calcolare il valore medio sul sistema perturbato si può introdurre una funzione di risposta $R(t, t')$ tale che:

$$\langle v(t) \rangle_\epsilon = \epsilon \int_{-\infty}^t F(t') R(t, t') dt' \quad (7.3)$$

che si basa sul fatto che la risposta al tempo t di una osservabile perturbata (in questo caso la velocità) dipende dalla perturbazione $F(t)$ a ogni tempo $t' < t$. La funzione di risposta $R(t, t')$ deve essere invariante sotto traslazioni temporali e quindi può dipendere solo da $t - t'$ e inoltre per preservare la causalità deve essere 0 per $t' > t$, cioè il valor medio perturbato non può dipendere dalla perturbazione a tempi maggiori del tempo che sto considerando. Per conoscere la funzione di risposta bisognerebbe porre $F(t) = \delta(t)$ cosicché nella eq. (7.3) si farebbe l'integrale con la δ e resterebbe soltanto la funzione di risposta calcolata a $t = 0$. Se fossimo in grado di risolvere il problema con una perturbazione impulsiva quindi avremmo la funzione di risposta.

Nel nostro caso per ricavare la funzione di risposta possiamo argomentare che, essendo la perturbazione istantanea al tempo $t' = 0$ ha soltanto l'effetto di cambiare la condizione iniziale, si passa cioè da $v(0)$ a $v(0) + \epsilon$ (o meglio ϵ per qualcosa che ha le dimensioni di una velocità). La soluzione della eq. (7.2) è quindi una semplice modifica della soluzione della eq. (7.1) che avevamo ricavato tempo fa:

$$v(t) = (v(0) + \epsilon)e^{-\frac{t}{\tau}} + \sqrt{2c} \int_0^t e^{-\frac{t-t'}{\tau}} \eta(t') dt' \quad (7.4)$$

da cui mediando

$$\langle v(t) \rangle_\epsilon = \langle v(0) \rangle_0 e^{-\frac{t}{\tau}} + \epsilon e^{-\frac{t}{\tau}} + \sqrt{2c} \int_0^t e^{-\frac{t-t'}{\tau}} \langle \eta(t') \rangle_0 dt' \quad (7.5)$$

in cui il primo e il terzo termine sono nulli perché $v(0) = 0$ e $\langle \eta(t) \rangle_0 = 0$. Riconosciamo quindi che la funzione di risposta altro non è che $R(t - t') = e^{-\frac{t-t'}{\tau}}$.

Nel caso di una perturbazione $F(t)$ generica possiamo scrivere quindi:

$$\langle v(t) \rangle_\epsilon = \epsilon \int_{-\infty}^t F(t') e^{-\frac{t-t'}{\tau}} dt' \quad (7.6)$$

■ **Esempio 7.1** Utilizzando la precedente nel semplice caso in cui $F(t) = F\theta(t)$ (cosicché l'integrazione parte da $t' = 0$) abbiamo che per t che va all'infinito

$$\langle v(\infty) \rangle_\epsilon = \epsilon F \lim_{t \rightarrow \infty} \int_0^t e^{-\frac{t-t'}{\tau}} dt' = \epsilon F \tau \lim_{t \rightarrow \infty} [1 - e^{-t/\tau}] = \epsilon F \tau \quad (7.7)$$

Nel caso lineare osserviamo che la funzione di risposta $R(t - t') = e^{-\frac{t-t'}{\tau}}$ è un esponenziale, la stessa forma della funzione di correlazione $\langle v(t)v(0) \rangle_0 = \langle v^2 \rangle_0 e^{-\frac{t}{\tau}}$. Per sottolineare questo punto possiamo scrivere $\frac{\langle v(t) \rangle_\epsilon}{\epsilon} = e^{-\frac{t}{\tau}}$ e in maniera ancor più simbolica

$$\frac{\langle \delta v(t) \rangle_\epsilon}{\delta v(0)} = e^{-\frac{t}{\tau}} = \frac{\langle v(t)v(0) \rangle_0}{\langle v^2 \rangle_0} \quad (7.8)$$

il primo termine è come risponde il sistema ad una perturbazione ϵ rispetto alla perturbazione stessa. Abbiamo quindi che la risposta media normalizzata con la perturbazione iniziale è l'esponenziale e in questo caso l'esponenziale è anche la funzione di correlazione. Con una verifica diretta in questa situazione si è trovato che il sistema risponde ad una perturbazione nello stesso modo in cui fluttua (**relazioni di fluttuazione-dissipazione**). In parole povere noi diamo una "botta" al sistema e il sistema in media riassorbe la botta come la funzione di correlazione. Il sistema rilassa come fluttua (come le variabili imperturbate fluttuano), per dirla alla Onsager.

Osserviamo che potevamo giungere all'espressione a tempi grandi $\langle v(\infty) \rangle_\epsilon$ semplicemente mediando l'eq. (7.8) (per semplicità poniamo $\epsilon = 1$)

$$\langle \dot{v} \rangle = -\frac{\langle v \rangle}{\tau} + \sqrt{2c} \langle \eta \rangle + \langle F \rangle \quad (7.9)$$

Dato che abbiamo a che fare con un rumore bianco $\langle \eta \rangle = 0$, che per tempi lunghi $\langle \dot{v} \rangle = 0$ e che $\langle F \rangle = F$ l'equazione si riduce a

$$-\frac{\langle v(\infty) \rangle_\epsilon}{\tau} + F = 0 \quad (7.10)$$

da cui $\langle v(\infty) \rangle_\epsilon = \frac{F}{m}\tau$ (dove si è reintrodotta la massa per comodità per il passaggio successivo). Inoltre sappiamo che il coefficiente di diffusione $D = \langle v^2 \rangle_0 \tau = \frac{k_B T}{m} \tau$ da cui $\tau = \frac{Dm}{k_B T}$. In definitiva quindi

$$\langle v(\infty) \rangle_\epsilon = \frac{D}{k_B T} F = \mu F \quad (7.11)$$

con μ la mobilità, che è l'analogo della conducibilità. Quest'ultima relazione quindi ci dice che la corrente è proporzionale al campo, e il coefficiente di proporzionalità μ è una proprietà di non equilibrio. Il punto è che questa proprietà di non equilibrio in questo caso si può ricavare da caratteristiche di equilibrio (D, T) , per ottenere il coefficiente di diffusione infatti non è necessario perturbare il sistema.

7.1.1 Caso in più dimensioni

Generalizziamo il procedimento al caso dell'equazione di Langevin lineare in più dimensioni. L'equazione imperturbata da studiare è

$$\dot{\underline{x}} = \hat{A}\underline{x} + \hat{B}\eta \quad (7.12)$$

La distribuzione stazionaria di questo processo è una gaussiana multivariata, in cui assumiamo tutte le medie nulle. In analogia con il caso unidimensionale calcoliamo la funzione di correlazione sul sistema imperturbato

$$C_{ij}(t) = \langle x_i(t)x_j(0) \rangle_0 \quad (7.13)$$

Risolviamo innanzitutto l'equazione omogenea associata $\dot{\underline{x}} = \hat{A}\underline{x}$ che ha come soluzione

$$\underline{x}(t) = e^{\hat{A}t}\underline{x}(0)$$

da cui introducendo l'elemento di matrice $G_{ik}(t) = (e^{\hat{A}t})_{ik}$ abbiamo $x_i(t) = \sum_k G_{ik}(t)x_k(0)$.

Possiamo scrivere formalmente la soluzione della eq. (7.12) quindi come $\underline{x}(t) = \hat{G}(t)\underline{x}(0) + \int_0^t \hat{G}(t-t')\hat{B}\eta(t')dt'$. La funzione di correlazione è quindi

$$C_{ij}(t) = \langle (\sum_k G_{ik}(t)x_k(0) + \int_0^t \sum_k (\hat{G}(t-t')\hat{B})_{ik}\eta_k(t')dt')x_j(0) \rangle_0 \quad (7.14)$$

La media opera soltanto sulle variabili x in quanto G_{ik} sono dei numeri e dato che $\langle \eta(t')x_j(0) \rangle = 0$ si ottiene

$$C_{ij}(t) = \sum_k G_{ik}(t)\langle x_k(0)x_j(0) \rangle_0 = \sum_k G_{ik}(t)C_{kj}(0) \quad (7.15)$$

oppure nel formalismo matriciale $\hat{C}(t) = \hat{G}(t)\hat{C}(0)$. Le funzioni di correlazioni dipendono quindi dalla matrice \hat{G} , che a sua volta dipende soltanto dalla parte deterministica \hat{A} , e dalle funzioni di correlazioni $\langle x_k(0)x_j(0) \rangle_0$ che invece dipendono anche dal rumore.

Lo step successivo è ora quello di aggiungere la perturbazione ϵ . La nostra ipotesi, che essendo il problema lineare si può facilmente generalizzare, è che la perturbazione (assumendo che sia una $\delta(t)$) agisce soltanto sulla componente j -esima del vettore \underline{x} . Si può quindi scrivere

$$\dot{x}_i = \sum_k A_{ik}x_k + \sum_k B_{ik}\eta_k + \epsilon\delta_{ij}\delta(t) \quad (7.16)$$

Inoltre, come nel caso unidimensionale, abbiamo che la perturbazione impulsata va a modificare soltanto la condizione iniziale, quindi $\delta\underline{x}(0) = (0, 0, \dots, \epsilon, 0, \dots, 0)$ con ϵ sulla componente j -esima. Ora dobbiamo

vedere come evolve $\delta\mathbf{x}(t) = \mathbf{x}(t)_\epsilon - \mathbf{x}(t)_0$, che altro non è l'evoluzione temporale della risposta del sistema alla perturbazione. In teoria potrebbero esserci due situazioni distinte:

1. Utilizziamo lo stesso rumore sia per il sistema imperturbato che per quello perturbato, in altre parole abbiamo un certo rumore per il sistema imperturbato, diamo una botta e il rumore resta lo stesso, quindi quando sottraiamo le due situazioni (che è quello che il vettore $\delta\mathbf{x}(t)$ rappresenta) il rumore stesso si cancella: $\delta\dot{\mathbf{x}} = \hat{A}\delta\mathbf{x}$.
2. Più realisticamente i rumori sono diversi e scorrelati e quindi si andranno a sommare quadraticamente: $\delta\dot{\mathbf{x}} = \hat{A}\delta\mathbf{x} + \sqrt{2}\hat{B}\eta$

Entrambe le possibilità però per i nostri scopi sono equivalenti in quanto siamo interessati alle quantità mediate, ed essendo il rumore bianco la sua media è nulla. Abbiamo quindi

$$\langle \delta\mathbf{x}(t) \rangle_\epsilon = \hat{G}(t) \langle \delta\mathbf{x}(0) \rangle_\epsilon \quad (7.17)$$

e per la singola componente

$$\langle \delta x_i(t) \rangle_\epsilon = G_{ij}(t) \delta x_j(0) \quad (7.18)$$

con $\delta x_j(0) = \epsilon \delta_{jl}$, l la componente che abbiamo supposto perturbata. Questa relazione è analoga alla eq. (7.8), ci dice come risponde la componente i -esima del sistema ad una perturbazione ϵ che modifica solo la componente j -esima. In maniera più simbolica, essendo $\hat{G}(t) = \hat{C}(t)\hat{C}^{-1}(0)$ e ricordando l'espressione per $\hat{C}(t)$, abbiamo

$$\frac{\langle \delta x_i(t) \rangle_\epsilon}{\delta x_j(0)} = G_{ij}(t) = \sum_k \langle x_i(t) x_k(0) \rangle_0 (\hat{C}^{-1}(0))_{kj} \quad (7.19)$$

Quest'ultima relazione ci dice che il sistema risponde alla perturbazione con una combinazione di funzioni di correlazione sul sistema imperturbato, similmente al caso unidimensionale.

Per concludere si vuole sottolineare come le relazioni di fluttuazione-dissipazione non valgono soltanto se si perturba un sistema all'equilibrio (anche se questa è la situazione tipica). Possiamo anche perturbare un sistema fuori dall'equilibrio (che ad esempio era stato precedentemente perturbato) e le relazioni di fluttuazione-dissipazione continuano a valere, a patto che la perturbazione sia piccola.

7.2 Sistemi hamiltoniani

Nei problemi lineari siamo in grado di risolvere le equazioni: da queste soluzioni poi, facendo le medie, abbiamo ottenuto le relazioni di fluttuazione-dissipazione. Ora cerchiamo di affrontare il problema in maniera più generale.

Abbiamo una Hamiltoniana imperturbata indipendente dal tempo $\mathcal{H}_0(\underline{X})$ con $\underline{X} = (Q, P)$ a cui aggiungiamo una perturbazione "piccola" dipendente dal tempo $F(t)A(\underline{X})$. Notare che per semplicità è il parametro "piccolo" ϵ è inglobato in A , mentre verrà comunque utilizzato come pedice per le medie sul sistema perturbato. Come al solito supponiamo che la perturbazione F sia accesa al tempo $t = 0$ e sia una $\theta(t)$ o più in generale che si assesti su un valore costante per tempi grandi. L'Hamiltoniana del sistema in esame diventa quindi

$$\mathcal{H}(\underline{X}, t) = \mathcal{H}_0(\underline{X}) - F(t)A(\underline{X}) \quad (7.20)$$

Il nostro obiettivo non è quello di ricavare la distribuzione di probabilità del sistema perturbato (che in generale sarà una funzione complicata di x , t e della perturbazione), ma è quello di ricavare informazioni su un generico osservabile B che ha un certo valore medio sul sistema imperturbato $\langle B \rangle_0$, ossia siamo interessati a:

$$\langle \delta B(t) \rangle_\epsilon = \langle B(t) \rangle_\epsilon - \langle B \rangle_0$$

Se la perturbazione è piccola ci aspettiamo una risposta lineare e trascurando i termini di ordine superiore possiamo scrivere, introducendo come in precedenza una funzione di risposta $R(t - t')$:

$$\langle \delta B(t) \rangle_\epsilon = \int_{-\infty}^t R(t - t') F(t') dt' \quad (7.21)$$

La funzione di risposta $R(t - t')$ in generale dipende sia da $A(\underline{X})$ che da B , e se guardiamo il sistema per $t \rightarrow \infty$ abbiamo che (dato che la perturbazione è costante per tempi grandi) l'unico termine che conta è $\int_{-\infty}^{+\infty} R(t - t') dt'$. Questo approccio va sotto il nome di **teoria della risposta lineare**.

■ **Esempio 7.2** Per comprendere l'utilità di questo approccio possiamo considerare il caso di un campo elettrico lungo la direzione x . Abbiamo quindi una perturbazione del tipo $qE(t) \sum_{n=1}^N x_n$ con N numero delle particelle. Supponiamo che a campo spento $\langle v \rangle_0 = 0$. Con $E \neq 0$ avremo in generale

$$\langle v(t) \rangle_E = q \int_{-\infty}^t E(t') R(t - t') dt' \quad (7.22)$$

A tempi asintotici, supponendo $E(t) = E\theta(t)$ abbiamo $\langle v \rangle_E = qE \int_0^\infty R(t) dt$. Questa, a meno di fattori, altro non è che la legge di Ohm e vediamo come $\int_0^\infty R(t) dt$ sia strettamente legato alla conducibilità. ■

Il problema ora è come calcolare la funzione di risposta R . Il risultato che diamo, la cui dimostrazione viene omessa¹, è la **formula di Kubo**:

$$R(t) = \beta \left\langle B(t) \frac{dA}{dt} \Big|_{t=0} \right\rangle_0 \quad (7.23)$$

in cui l'osservabile B è calcolata nelle coordinate al tempo t e la media è fatta sul sistema imperturbato. Ovviamente R dipende da come perturbiamo il sistema ($\frac{dA}{dt} \Big|_{t=0}$) e da cosa guardiamo ($B(t)$).

Abbiamo quindi trovato che anche in questo caso la risposta è una funzione di correlazione del sistema imperturbato.

■ **Esempio 7.3** Tornando all'esempio del campo elettrico e supponendo per semplicità che il campo agisca solo su una particella la perturbazione è $qE(t)x(t)$. Abbiamo quindi $A = x$ e prendiamo come osservabile la velocità della particella stessa $B = v$. Utilizzando la relazione di Kubo eq. (7.23) si può scrivere (essendo $\langle v \rangle_0 = 0$ da cui $\langle \delta v(t) \rangle_E = \langle v(t) \rangle_E$) $R(t) = \langle v(t) \dot{x}(0) \rangle_0 = \langle v(t)v(0) \rangle_0$ e quindi

$$\langle v(t) \rangle_E = q\beta \int_{-\infty}^t E(t') \langle v(t')v(0) \rangle_0 dt' \quad (7.24)$$

Ora tutta la difficoltà sta nel trovare $\langle v(t')v(0) \rangle_0$ che in generale è un oggetto complicato. Se però siamo interessati a $t \rightarrow \infty$ e supponendo al solito $E(t) = E\theta(t)$ abbiamo

$$\langle v \rangle_E = q\beta E \int_0^\infty \langle v(t)v(0) \rangle_0 dt \quad (7.25)$$

Il termine $\int_0^\infty \langle v(t)v(0) \rangle_0 dt$ è una costante legata alla conducibilità. Il risultato più importante è proprio questo: in linea di principio (se conoscessi le funzioni di correlazione del sistema imperturbato) potrei calcolare la conducibilità senza accendere la perturbazione, senza applicare una differenza di potenziale. Si possono fare delle ipotesi sulla funzione di correlazione, ad esempio possiamo dire che $\langle v(t)v(0) \rangle_0 \sim \langle v^2 \rangle_0 e^{-\frac{t}{\tau}}$ con τ un certo tempo caratteristico. Abbiamo quindi che a tempi asintotici la eq. (7.25) diventa (assumendo l'equilibrio termodinamico)

$$\langle v \rangle_E \sim q\beta E \langle v^2 \rangle_0 \int_0^\infty e^{-\frac{t}{\tau}} dt \sim q\beta E \langle v^2 \rangle_0 \tau \sim \frac{k_B T}{m} \tau q\beta E \quad (7.26)$$

¹Si tratta di una dimostrazione un po' noiosa che inoltre fa uso di ipotesi che sono soggette ad obiezioni. Vedremo nel dettaglio una formulazione più generale.

A tempi lunghi quindi dobbiamo risolvere integrali di funzioni di correlazione. Si è trovata un'espressione esplicita per la conducibilità che dipende da τ . Il tempo caratteristico è determinato dal modello microscopico che si utilizza. ■

7.2.1 Obiezione di Van Kampen

Van Kampen formulò un'obiezione alla teoria della risposta lineare basata sul fatto che nella dimostrazione si manipolano oggetti che si suppongono essere sempre piccoli (ad esempio $\delta x_i(t)$, differenze tra variabili perturbate e variabili imperturbate).

In particolare egli sosteneva la sua invalidità in presenza di sistemi caotici, dato che l'assunzione che lo spostamento dalla "traiettoria" imperturbata $\delta x(t)$ è piccolo ad ogni t è evidentemente sbagliata. Infatti, in un sistema caotico, nonostante le traiettorie distino $\epsilon \ll 1$ a $t = 0$ si allontanano esponenzialmente nel tempo. Se a $t = 0$ perturbo il sistema con una perturbazione piccola, avrò qualcosa del tipo $\delta x_i(0) \sim \epsilon$, ma nel caso di un sistema caotico può crescere in maniera esponenziale $\delta x_i(t) \sim \epsilon e^{\lambda t}$. Per cui l'ipotesi che i valori medi di queste variabili siano dell'ordine di una quantità piccola Δ è valida solo per tempi $t < t^*$, con t^* che si può ricavare approssimativamente da $\langle \delta x_i(t) \rangle_\epsilon \sim \langle \epsilon e^{\lambda t} \rangle_\epsilon \sim O(\Delta)$ da cui

$$t^* \sim \frac{1}{\lambda} \ln \frac{\Delta}{\epsilon}$$

■ **Esempio 7.4** Per un campo elettrico possiamo stimare la perturbazione iniziale massima affinché la risposta lineare sia valida per $t < 1s$. Facendo il conto si ottiene che tale campo elettrico massimo è dell'ordine di $10^{-20} \frac{V}{m}$, praticamente zero. Secondo Van Kampen, data questa condizione così stringente sulla perturbazione iniziale, la teoria della risposta lineare non può essere fondata. ■

Il ragionamento mosso da Van Kampen trascura però la presenza delle fluttuazioni stocastiche, le quali portano $\delta x(t)$ ad essere piccolo ad ogni t . La giustificazione di Kubo è che l'instabilità delle traiettorie favorisce la stabilità delle funzioni di distribuzione a causa del processo di mescolamento. Fisicamente ci sono delle oscillazioni e delle cancellazioni che mantengono piccoli anche i valori medi su $\delta x_i(t)$ (il ragionamento di Van Kampen infatti sarebbe stato valido se avessimo fatto $\langle |\delta x_i(t)| \rangle$).

Questa argomentazione così posta è però debole, ma non è neppure necessaria dato che si trovò una formulazione della relazione di Kubo dipendente dalla sola $\delta x(0)$ e non dalla sua evoluzione.

7.3 Teorema di fluttuazione-dissipazione generalizzato

La forma più generale del teorema di fluttuazione-dissipazione è la seguente:

Teorema 7.3.1 — Fluttuazione-dissipazione generalizzato. Consideriamo una variabile markoviana \underline{x} a n componenti, che ammette una probabilità stazionaria $P_s(\underline{x}) \neq 0$ per ogni \underline{x} . Supponiamo questo sistema sia soggetto ad una perturbazione istantanea a $t = 0$, ossia

$$x_j \rightarrow x_j + \delta x_j(0)$$

allora, detto $\overline{\delta B(x)}^t = \langle B \rangle_\epsilon(t) - \langle B \rangle_0(t)$, si ha che

$$\overline{\delta B(x)}^t = - \sum_j \left\langle B(\underline{x}(t)) \frac{\partial \ln P_s(\underline{y})}{\partial x_j} \Big|_{\underline{x}(0)} \delta x_j(0) \right\rangle_0 \quad (7.27)$$

Come verificato nei casi studiati in precedenza, la risposta del sistema è proporzionale al valor medio sul sistema imperturbato dell'osservabile per una funzione che dipende dalla perturbazione e dallo stato iniziale

$$\frac{\langle \delta x_i(t) \rangle_\epsilon}{\delta x_j(0)} = -\langle x_i(t) F_j(\underline{x}(0)) \rangle_0 \quad (7.28)$$

L'unico punto delicato sta nell'ipotesi per cui $P_S(\underline{x}) \neq 0$, ma possiamo dire che nei casi fisicamente rilevanti questa ipotesi è sempre verificata.

Dimostrazione. Ci poniamo nel caso in cui la perturbazione modifica per semplicità la sola componente j -esima. A seguito di questa perturbazione siamo interessati allo studio delle variazioni di un generico osservabile $B(\underline{x})$ al variare del tempo. Abbiamo dunque per definizione

$$\overline{\delta B(\underline{x})}^t = \int d\underline{x} B(\underline{x}) P_\epsilon(\underline{x}, t) - \int d\underline{x} B(\underline{x}) P_0(\underline{x}, t), \quad (7.29)$$

dove la distribuzione perturbata è P_ϵ e $P_0(\underline{x}, t)$ quella imperturbata, mentre \overline{A}^t indica la media al tempo t .² Sfruttando la markovianità del processo in esame è possibile scrivere la probabilità perturbata e imperturbata al tempo t propagando opportunamente quelle al tempo $t = 0$

$$\begin{aligned} P_\epsilon(\underline{x}, t) &= \int d\underline{y} P_\epsilon(\underline{y}, 0) W(\underline{y} \rightarrow \underline{x}, t) \\ P_0(\underline{x}, t) &= \int d\underline{y} P_0(\underline{y}, 0) W(\underline{y} \rightarrow \underline{x}, t), \end{aligned}$$

che sostituite in eq. (7.29) portano a

$$\overline{\delta B(\underline{x})}^t = \int d\underline{x} d\underline{y} B(\underline{x}) [P_\epsilon(\underline{y}, 0) - P_0(\underline{y}, 0)] W(\underline{y} \rightarrow \underline{x}, t). \quad (7.30)$$

Così scritta l'equazione è poco maneggevole, ma è possibile semplificarla facendo alcune considerazioni:

1. il sistema imperturbato è all'equilibrio, da cui $P_0(\underline{x}, 0) = P_s(\underline{x})$;
2. il sistema è perturbato da una forza impulsiva, il che porta il sistema ad essere traslato rigidamente. Nello specifico abbiamo $x_\epsilon = x + \epsilon \rightarrow \langle x_\epsilon \rangle = \epsilon$, dato che $\langle x \rangle = 0$. Questo si riflette su una semplice traslazione della densità di probabilità

$$P_\epsilon(\underline{x}, 0) = P_s(\underline{x} - \delta \underline{x}(0)).$$

Dunque la eq. (7.30) diventa

$$\overline{\delta B(\underline{x})}^t = \int d\underline{x} d\underline{y} B(\underline{x}) [P_s(\underline{y} - \delta \underline{x}(0)) - P_s(\underline{y})] W(\underline{y} \rightarrow \underline{x}, t), \quad (7.31)$$

la quale non è ancora della forma desiderata. Quanto vogliamo è infatti correlare la risposta del sistema con una qualche funzione di correlazione, le quali sono della forma, restringendoci al caso unidimensionale

$$\langle x(t)x(0) \rangle = \int dx dy xy P(y, 0, x, t),$$

dove $P(y, 0, x, t)$ è la probabilità che il sistema sia in y al tempo $t = 0$ e in x al tempo t . Sfruttando la markovianità si ottiene

$$\langle x(t)x(0) \rangle = \int dx dy xy P_s(y) W(y \rightarrow x, t).$$

Dunque, vogliamo ottenere una formula analoga in eq. (7.31), il che può essere conseguito banalmente moltiplicando e dividendo per $P_s(\underline{y})$, ottenendo

²Si sta compiendo per il momento una trattazione generale, in cui il sistema perturbato non è detto che sia all'equilibrio.

$$\begin{aligned}
\overline{\delta B(x)}^t &= \int d\underline{x} d\underline{y} B(\underline{x}) \underbrace{\frac{P_s(\underline{y} - \delta\underline{x}(0)) - P_s(\underline{y})}{P_s(\underline{y})}}_{F(\underline{y}; \delta\underline{x}(0))} P_s(\underline{y}) W(\underline{y} \rightarrow \underline{x}, t) \\
&= \left\langle B(\underline{x}(t)) F(\underline{y}; \delta\underline{x}(0)) \right\rangle_0
\end{aligned} \tag{7.32}$$

dove in $F(\underline{y}; \delta\underline{x}(0))$ l'uso del “;” è per indicare che il primo termine è una variabile mentre il secondo un parametro fisso. La formula ottenuta è esatta e indipendente dal fatto che $\delta\underline{x}(0)$ sia piccolo. Ora, possiamo muovere ulteriori interessanti considerazioni sfruttando ciò. Se $\delta\underline{x}(0)$, di componenti $\delta x_i = \delta_{ij} \delta x_j$ (cioè con la sola componente j non nulla), è piccolo allora

$$P_s(\underline{y} - \delta\underline{x}(0)) = P_s(\underline{y}) - \left. \frac{\partial P_s(\underline{y})}{\partial x_j} \right|_{\delta x_j=0} \delta x_j(0) + \mathcal{O}(\delta x_j^2),$$

che sostituita in eq. (7.32)

$$\begin{aligned}
\overline{\delta B(x)}^t &= \int d\underline{x} d\underline{y} B(\underline{x}) \frac{-\left. \frac{\partial P_s(\underline{y})}{\partial x_j} \right|_{\delta x_j=0} \delta x_j(0)}{P_s(\underline{y})} P_s(\underline{y}) W(\underline{y} \rightarrow \underline{x}, t) \\
&= - \int d\underline{x} d\underline{y} B(\underline{x}) \left. \frac{\partial \ln P_s(\underline{y})}{\partial x_j} \right|_{\underline{x}(0)} P_s(\underline{y}) \delta x_j(0) W(\underline{y} \rightarrow \underline{x}, t) \\
&= - \left\langle B(\underline{x}(t)) \left. \frac{\partial \ln P_s(\underline{y})}{\partial x_j} \right|_{\underline{x}(0)} \delta x_j(0) \right\rangle_0
\end{aligned} \tag{7.33}$$

■

Si osservi l'assenza di $\delta\underline{x}(t)$, eliminando il problema sollevato da Van Kampen.³ La formula qui ottenuta è estremamente generale, descrivendo la dinamica di tutti quei fenomeni per cui $P_s(\underline{x}) \neq 0$. Questo significa la sua inefficienza in sistemi caotici dissipativi, dato che in tale circostanza $P_s(\underline{x}) = 0$ per qualche \underline{x} . Basti pensare ad una pallina in una buca con attrito, dove è evidente che $P_s(x) = \delta(x - x_{\min})$ in cui x_{\min} è il punto di minimo della buca. In tali circostanze si può tentare di salvare la situazione includendo del rumore, che può spingere il sistema a esplorare ad ogni t con probabilità non nulla qualsiasi stato.⁴

7.4 Applicazioni della teoria della risposta lineare

7.4.1 Sistema Hamiltoniano

Consideriamo un sistema di Hamiltoniana $H = H_0 - A(Q)F(t)$, in cui H_0 è la Hamiltoniana imperturbata e $A(Q) = \mathcal{O}(\epsilon)$ e $F(t) = \delta(t)$ (perturbazione impulsiva come nel caso precedente). A seguito dell'accensione della perturbazione quanto si modificano sono i soli impulsi, dato che

$$\begin{aligned}
\dot{Q}_n &= \frac{\partial H}{\partial P_n} = \frac{\partial H_0}{\partial P_n} \\
\dot{P}_n &= -\frac{\partial H}{\partial Q_n} = -\frac{\partial H_0}{\partial Q_n} + \frac{\partial A(Q)}{\partial Q_n} \delta(t),
\end{aligned}$$

³Si noti che $\left. \frac{\partial P_s(\underline{y})}{\partial x_j} \right|_{\delta x_j=0}$ è equivalente a $\left. \frac{\partial P_s(\underline{y})}{\partial x_j} \right|_{\underline{x}(0)}$.

⁴Il sistema esposto non è caotico, ma è per illustrare l'idea.

da cui $\delta P_n(0) = \partial A(Q)/\partial Q_n$. Essendo in un sistema Hamiltoniano all'equilibrio la probabilità stazionaria del sistema imperturbato è quella canonica

$$P_s(Q, P) = \mathcal{N} e^{-\beta H_0(Q, P)} \rightarrow \ln P_s(Q, P) = -\beta H_0 + \ln \mathcal{N} \rightarrow \frac{\partial \ln P_s(Q, P)}{\partial P_n} = -\beta \frac{\partial H_0}{\partial P_n} = -\beta \dot{Q}_n,$$

in cui si è calcolata la derivata rispetto alla componente P_n dell'impulso essendo la quantità che si perturba (si veda eq. (7.33)). Dunque abbiamo

$$\begin{aligned} \overline{\delta B(x)^t} &= -\left\langle B(\underline{x}(t)) \sum_n \frac{\partial \ln P_s(y)}{\partial P_n} \Big|_{\underline{P}(0)} \delta P_n(0) \right\rangle_0 \\ &= \beta \left\langle B(\underline{x}(t)) \sum_n \dot{Q}_n \frac{\partial A(Q)}{\partial Q_n} \Big|_{t=0} \right\rangle_0 \\ &= \beta \left\langle B(t) \frac{dA}{dt} \Big|_{t=0} \right\rangle_0, \end{aligned}$$

dove si è riconosciuta la definizione di derivata temporale di $A(Q)$. Abbiamo in questo modo ottenuto lo stesso risultato di Kubo.

7.4.2 Sistema caotico

L'obiezione mossa da Van Kampem non è però del tutto insensata. Consideriamo un sistema caotico, in cui vogliamo studiare la variazione di $B(x) = \underline{x}$

$$\frac{\overline{\delta x_i(t)}}{\delta x_i(0)} = -\left\langle x_i(t) \frac{\partial \ln P_s}{\partial x_i} \Big|_{\underline{x}(0)} \right\rangle_0,$$

dove si noti ci curiamo della singola componente i -esima. Per poter procedere a questo calcolo possiamo svolgere analiticamente il conto, oppure compiere una media campionaria. Infatti, invocando l'ergodicità del sistema, i due risultati devono essere equivalenti. La procedura che seguiamo per compiere questa media è la seguente:

1. al tempo $t = 0$ perturbiamo impulsivamente il sistema

$$x_i(0) \rightarrow x_i^{(k)} = x_i(0) + \delta x_i(0),$$

dove $k = 1, \dots, N$ indica quante volte il sistema è stato perturbato;

2. lo lasciamo evolvere sino ad un tempo $t = \tau$, in corrispondenza del quale lo "proiettiamo" sulla traiettoria imperturbata $x_i(\tau)$;

3. poniamo $\tau = 0$ e ripetiamo il procedimento un numero di volte N grande, aumentando k di 1.

Si faccia attenzione al fatto che la perturbazione iniziale è indipendente da k , cioè è sempre la stessa. Indicando con $\delta x_i^{(k)}(t) = x_i^{(k)}(t) - x_i(t)$, dove si ricordi che $x_i(t)$ è la componente imperturbata, si ottiene la seguente sequenza

$$\delta x_i^{(1)}(t), \delta x_i^{(2)}(t), \dots, \delta x_i^{(N)}(t),$$

da cui possiamo stimare la variazione di $x_i(t)$ come media sulle variazioni

$$\frac{\overline{\delta x_i(t)}}{\delta x_i(0)} = \frac{1}{\delta x_i(0)} \sum_{k=1}^N \delta x_i^{(k)}(t).$$

Nonostante $|\delta x_i(t)| \sim \epsilon e^{\lambda t}$ i contributi molto grandi si cancellano rendendo la quantità piccola come atteso. Ciononostante, è vero che la varianza cresce esponenzialmente $\sigma^2 \sim \epsilon^2 e^{2\lambda t}$.

7.4.3 Sistemi gradiente

Consideriamo l'equazione di Langevin unidimensionale, in cui il termine deterministico è sempre esprimibile in termini di derivata di una funzione $V(x)$ (potenziale), tale che

$$\dot{x} = -\frac{dV}{dx} + \sqrt{2c}\eta(t).$$

Scrivendo la Fokker-Planck di questo sistema è evidente che la densità di probabilità stazionaria sia $P_s(x) = \mathcal{N}e^{-V(x)/c} \neq 0$ per ogni x . Dato che sono soddisfatte le ipotesi alla base della teoria della risposta lineare, si ha

$$\frac{\overline{\delta x(t)}}{\overline{\delta x(0)}} = \left\langle x(t) \frac{1}{c} \frac{dV}{dx} \Big|_{x(0)} \right\rangle_0.$$

7.5 Conclusioni

Dopo aver analizzato alcuni esempi di applicazione della teoria della risposta lineare, sembrerebbe che lo studio delle risposte di un sistema arbitrario sia possibile. Ciononostante, la situazione non è così semplice. Infatti, si noti che il calcolo della risposta di un sistema è vincolata alla conoscenza della probabilità stazionaria $P_s(x)$. Quest'ultima può essere in linea di principio ricavata dalla risoluzione dell'equazione di Fokker-Planck, ma è nella pratica spesso impossibile. La eq. (7.33) consente però di affermare che la risposta di un sistema è pari ad una qualche correlazione, seppur complicata, consentendo di dire che un sistema risponde come fluttua.

Riferimenti bibliografici

- [1] U. Marini Bettolo Marconi, A. Puglisi, L. Rondoni e A. Vulpiani. "Fluctuation- dissipation: Response theory in statistical physics". In: *Phys. Rep.* 461.111 (2008).

8. Grandi deviazioni

8.1 Approssimazione gaussiana e secondo principio

Introduciamo il percorso che ci porterà alla trattazione del problema delle grandi deviazioni, partendo dalla distribuzione dell'energia in un ensemble canonico. Consideriamo a tal proposito un sistema composto da N particelle, di hamiltoniana H e a temperatura β . Nota la distribuzione di probabilità canonica $P_{\underline{X}}(\underline{X})$, è possibile ricavare quella dell'energia attraverso la relazione:

$$P_E(E) = \int d\underline{X} P_{\underline{X}}(\underline{X}) \delta(H - E) = \quad (8.1)$$

$$= \frac{\int e^{-\beta H(\underline{X})} \delta(H(\underline{X}) - E) d\underline{X}}{Z} = \quad (8.2)$$

$$= \frac{\omega(E) e^{-\beta E}}{\int dE \omega(E) e^{-\beta E}} \quad (8.3)$$

dove la funzione $\omega(E)$ è definita come

$$\omega(E) = \partial \Sigma(E) / \partial E = \frac{\partial}{\partial E} \int d\underline{X} \theta(E - H(\underline{x})) = \int d\underline{X} \delta(H(\underline{X}) - E)$$

Equivalentemente essa può essere definita a partire dalla relazione $\omega(E) = e^{S(E)/k_B}$, dove $S(E)$ è l'entropia in funzione dell'energia, che in virtù della sua natura estensiva può essere scritta in funzione dell'energia per particella $S(E) = NS(E/N) = NS(\epsilon)$. Sostituendo questa espressione nella eq. (8.3), si ottiene $P_E(E) = e^{-\beta(E-TS(E))}/Z$. Da quest'ultima è immediato ricavare la distribuzione dell'energia per particella:

$$P(\epsilon) = \frac{e^{-\beta N(\epsilon-TS(\epsilon))}}{\int d\epsilon e^{-\beta N(\epsilon-TS(\epsilon))}} \quad (8.4)$$

L'integrale al denominatore può essere sviluppato per $N \gg 1$ con il metodo di Laplace ottenendo

$$\int d\epsilon e^{-\beta N(\epsilon-TS(\epsilon))} \simeq e^{-\beta N \min_{\epsilon}(\epsilon-TS(\epsilon))} = e^{-\beta N f(T)}$$

con $f(T)$ energia libera per particella. Sostituendo tutto si ottiene $P(\epsilon) = e^{-\beta N[\epsilon - TS(\epsilon) - f(T)]}$. Indichiamo con ϵ^* il valore dell'energia per particella che minimizza la quantità $\epsilon - TS(\epsilon)$. L'approssimazione gaussiana consiste nello sviluppare l'esponente della distribuzione intorno a tale punto.

$$\epsilon - TS(\epsilon) - f(T) = -\frac{T}{2} \left. \frac{\partial^2 S(\epsilon)}{\partial \epsilon^2} \right|_{\epsilon^*} (\epsilon - \epsilon^*)^2 + \mathcal{O}(\epsilon - \epsilon^*)^3 \quad (8.5)$$

Da questo segue $P(\epsilon) \simeq e^{-\beta N \left(-\frac{1}{2} \left. \frac{\partial^2 S(\epsilon)}{\partial \epsilon^2} \right|_{\epsilon^*} (\epsilon - \epsilon^*)^2 \right)}$. Perché questa sia effettivamente una gaussiana è necessario che sia abbia $\left. \frac{\partial^2 S(\epsilon)}{\partial \epsilon^2} \right|_{\epsilon^*} < 0$. Una possibilità è quella di ricercare l'approssimazione gaussiana, e una volta trovata imporre la condizione per cui essa funzioni. Tuttavia esistono fortunatamente motivi più profondi per cui la condizione precedente debba essere rispettata.

La concavità della funzione entropia infatti, discende direttamente dal secondo principio della termodinamica. Per mostrare ciò consideriamo due sistemi inizialmente separati composti da N_1 e N_2 rispettivamente. Ipotizziamo inoltre che il sistema 2 si trovi ad una temperatura maggiore rispetto a 1 ($\beta_2 < \beta_1$). L'entropia nella situazione iniziale sarà data da $S_I = N_1 S(\epsilon_1) + N_2 S(\epsilon_2) = N [P_1 S(\epsilon_1) + P_2 S(\epsilon_2)] = N [P_1 S(\epsilon_1) + (1 - P_1) S(\epsilon_2)] = N \langle S(\epsilon) \rangle$, dove $N = N_1 + N_2$, $P_1 = N_1/N$ e $P_2 = N_2/N$.

Immaginiamo adesso di mescolare i due sistemi e chiediamoci quando valga l'entropia in questa nuova configurazione. L'energia finale sarà semplicemente la somma delle energie, cioè $E_f = E_1 + E_2 = N_1 \epsilon_1 + N_2 \epsilon_2 = N(P_1 \epsilon_1 + P_2 \epsilon_2) = N \langle \epsilon \rangle$. L'entropia finale sarà quindi $S_f = NS(\epsilon_f) = NS(\langle \epsilon \rangle)$. Il secondo principio richiede che valga la disuguaglianza $S_f \geq S_I$. Questa è automaticamente soddisfatta se S è una funzione concava:

Teorema 8.1.1 — Disuguaglianza di Jensen. Sia $f(x)$ una funzione concava ($f'' < 0$) nel suo dominio. Allora vale la disuguaglianza

$$\langle f(x) \rangle \leq f(\langle x \rangle) \quad (8.6)$$

Dimostrazione. Assumendo per semplicità f derivabile, allora $\forall x_0$ nel dominio di f vale la disuguaglianza:

$$f(x) \leq f(x_0) + f'(x_0)(x - x_0) \quad (8.7)$$

Prendendo il valor medio ambo i membri della precedente disuguaglianza si ottiene $\langle f(x) \rangle \leq f(x_0) + f'(x_0)(\langle x \rangle - x_0)$. La dimostrazione si conclude ponendo $x_0 = \langle x \rangle$, in virtù della sua arbitrarietà.

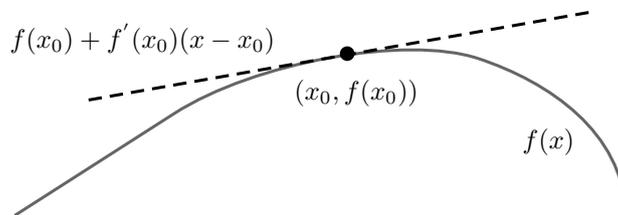


Figura 8.1: Funzione concava e la retta tangente in un punto. ■

Inoltre da questa trattazione emerge come essendo ϵ_f pari alla media tra l'energia per particella dei sistemi 1 e 2, si ha $\epsilon_1 < \epsilon_f < \epsilon_2$, con un corrispondente flusso di energia dal sistema a temperatura maggiore a quello a temperatura minore.

8.2 La funzione di Cramér

Concludiamo dicendo che l'approssimazione gaussiana corrisponde a richiedere che l'entropia sia una funzione parabolica di ϵ . Tuttavia qualsiasi entropia che abbia una forma funzionale concava risulterebbe essere adeguata. Introduciamo dunque il problema delle grandi deviazioni, che si pone come strumento attraverso il quale superare l'approssimazione gaussiana. Considerando la distribuzione dell'energia per particella questo può essere ottenuto cercando una funzione $C(\epsilon)$ tale per cui:

$$P(\epsilon) \propto e^{-N[\beta(\epsilon - TS(\epsilon) - f(T))]} = e^{-NC(\epsilon)} \quad (8.8)$$

La C , detta **funzione di Cramer**, deve inoltre rispettare le seguenti proprietà:

$$\begin{cases} C(\epsilon^*) = 0, & ; & \epsilon^* = \langle \epsilon \rangle \\ C(\epsilon) > 0, & & \epsilon \neq \epsilon^* \\ C(\epsilon) \simeq \frac{1}{2\sigma^2} (\epsilon - \epsilon^*)^2, & & \epsilon \simeq \epsilon^* \end{cases} \quad (8.9)$$

Richiedendo infine che la terza proprietà sia valida ad ogni temperatura se ne ottiene una ulteriore, cioè che $\frac{\partial^2 C}{\partial \epsilon^2} > 0$.

In quest'ottica l'approssimazione gaussiana emerge come sviluppo della funzione di Cramér intorno al suo massimo. È importante sottolineare come le suddette proprietà di C discendono tutte da considerazioni di carattere puramente matematico, e possono in quanto tali essere ricavate indipendentemente da qualsivoglia legame con la fisica.

Il problema delle grandi deviazioni fu affrontato per la prima dal matematico svedese H. Cramér, in relazione al suo lavoro presso le agenzie di assicurazione. Infatti sebbene il teorema del limite centrale descriva bene la distribuzione di probabilità di una somma di variabili indipendenti intorno al loro valor medio, lo stesso non si può dire nel momento in cui vengano considerati eventi rari (si pensi nell'ottica assicurativa al paragone tra il rischio associato ad un impiegato e quello all'amministratore delegato di un grande impianto industriale).

Serve dunque uno strumento matematico che permette di andare oltre il TLC (che poi altro non è se non l'approssimazione gaussiana). Prima di fare ciò introduciamo però il suddetto teorema, analizzandone proprietà e limitazioni.

Teorema 8.2.1 — Limite Centrale (TLC). Siano x_1, \dots, x_N N variabili indipendenti identicamente distribuite, con valor medio $\langle x \rangle$ e varianza finita $\sigma^2 < \infty$. Definita y_N come $y_N = \frac{1}{N} \sum_{n=1}^N x_n$, allora vale il seguente teorema del limite centrale:

Per $N \gg 1$ la distribuzione di probabilità della y_N risulta essere asintoticamente pari a:

$$P(y_N) \sim \sqrt{\frac{N}{2\pi\sigma^2}} e^{-N \frac{(y_N - \langle x \rangle)^2}{2\sigma^2}} \quad (8.10)$$

Cominciamo a vedere una limitazione sull'uso di questo teorema. Consideriamo a tal proposito un set di variabili tali che $x_n = 0, 1$, da cui segue che $0 < y_N < 1$. Conseguentemente la $P(y_N)$ dovrà essere nulla al di fuori di questo intervallo, proprietà che non si ritrova nella distribuzione di probabilità fornita dal TLC, che essendo gaussiana è definita su tutto l'asse reale. Sebbene la differenza tra le due distribuzioni al di fuori $[0,1]$ risulti in genere molto piccola questa risulta tuttavia rilevante (ad esempio nel calcolo dei momenti di grado elevato, vedi oltre).

Analogamente a quanto fatto nella sezione precedente superiamo l'approssimazione gaussiana introducendo una funzione di Cramér $C(y)$ tale che $P(y_N) \sim e^{-NC(y_N)}$ per $N \gg 1$. Questa deve rispettare le condizioni eq. (8.9). Dunque $C(\langle y \rangle) = 0$ e $C(y) > 0$ per $y \neq \langle y \rangle$, in modo che la distribuzione sia massima in corrispondenza del valor medio e che le altre realizzazioni siano invece depresse esponenzialmente. Dalla condizione di raccordo con il TLC segue poi $C(y) \simeq \frac{(y - \langle y \rangle)^2}{2\sigma^2}$. Non è invece ancora

matematicamente evidente perché debba essere $\frac{d^2C}{dy^2} > 0$. In figura fig. 8.2 è riportato il confronto tra il TCL e le grandi deviazioni per il caso binomiale per la quale la funzione di Cramer assume espressione $C(y) = \frac{1+y}{2} \ln \frac{1+y}{2p} + \frac{1-y}{2} \ln \frac{1-y}{2(1-p)}$.

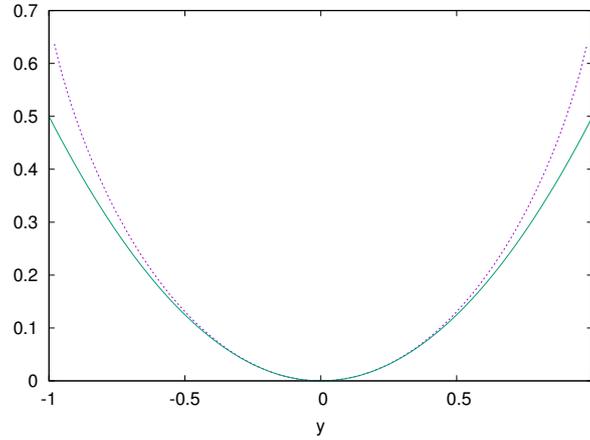


Figura 8.2: Confronto tra la funzione di Cramer $C(x)$ (linea tratteggiata) e l'approssimazione parabolica (linea continua), per la distribuzione binomiale con $p=0.5$

8.2.1 I limiti del TLC: la distribuzione log-normale

Per approfondire le limitazioni del TLC consideriamo adesso la distribuzione lognormale (il perché di questo nome sarà chiaro a breve). Lo studio di questa distribuzione è legata al problema ingegneristico della determinazione del numero ideale di "botte" da dare ad un blocco di carbone perché i frammenti derivanti abbiano le dimensioni desiderate (ad esempio sufficienti a farli passare per una tubatura). Tale problema è analogo allo studio della distribuzione delle dimensioni dei sassi in montagna (o dei granelli al mare) immaginando che siano tutti originati da un unico blocco iniziale di massa m_0 . Attraverso un ipotetico meccanismo, come ad esempio le gelate per la montagna, a partire da questo sasso se ne generano un numero generico, aventi ognuno una massa frazionaria di quella iniziale pari a $m_1 = x_1 m_0$, con $0 < x_1 < 1$. Il meccanismo può essere iterato un numero arbitrario di volte, ottenendo in generale, $m_N = x_N m_{N-1} = \prod_{i=1}^N x_i m_0$. Prendendo il logaritmo di questa relazione si ottiene $\ln m_N = \sum_{i=1}^N \ln x_i + \ln m_0$ da cui a sua volta è possibile definire la variabile:

$$\ln \frac{m_N}{m_0} = \sum_{i=1}^N \ln x_i = z_N \quad (8.11)$$

la quale nel limite $N \gg 1$, per il TLC avrà distribuzione asintotica data da

$$P(z_N) \sim \frac{1}{\sqrt{2N\pi\sigma_z^2}} e^{-\frac{1}{2N\sigma_z^2}(z_N - N\langle z \rangle)^2}$$

con $\langle z \rangle = \langle \ln x \rangle$ e $\sigma_z^2 = \langle \ln x^2 \rangle - \langle \ln x \rangle^2$. A questo punto occorre passare alla variabile $y_N = e^{z_N}$, per fare ciò applichiamo la formula del cambio di variabile per una distribuzione:

$$P_{Y_N}(y_N) = \frac{P_{Z_N}(z_N(y_N))}{\left| \frac{dy_N}{dz_N} \right|} = \frac{1}{\sqrt{2N\pi\sigma_z^2}} \frac{1}{y_N} e^{-\frac{1}{2N\sigma_z^2}(\ln y_N - N\langle \ln x \rangle)^2} \quad (8.12)$$

che non è altro che una distribuzione gaussiana nella variabile $\ln y_N$. È interessante notare che la forma funzionale di tale distribuzione è la stessa a prescindere dal luogo in cui i sassi risiedono a meno di fattori di normalizzazione quale la massa iniziale m_0 .

Consideriamo a questo punto il valor medio della potenza q -esima di y_N e confrontiamo quanto si ottiene con la distribuzione originaria

$$\langle y_N^q \rangle_V = \left\langle \left(\prod_{n=1}^N x_n \right)^q \right\rangle_V = (\langle x^q \rangle_V)^N = e^{N \ln \langle x^q \rangle_V} \quad (8.13)$$

con quanto si ottiene dalla lognormale; sviluppiamo il conto di $\langle y_N^q \rangle_{LG}$:

$$\begin{aligned} \langle y_N^q \rangle_{LG} &= \frac{1}{\sqrt{2N\pi\sigma_z^2}} \int dy y^{q-1} e^{-\frac{1}{2N\sigma_z^2}(\ln y - N\langle \ln x \rangle)^2} = \frac{1}{\sqrt{2N\pi\sigma_z^2}} \int dz e^{zq} e^{-\frac{1}{2N\sigma_z^2}(z - N\langle \ln x \rangle)^2} = \\ &= \frac{1}{\sqrt{2N\pi\sigma_z^2}} e^{N \ln \langle x \rangle q} \int d\xi e^{\xi q} e^{-\frac{\xi^2}{2N\sigma_z^2}} = e^{N(\langle \ln x \rangle q + \frac{1}{2}\sigma_z^2 q^2)} \end{aligned}$$

dove si sono fatti in successione i cambi di variabile: $y = e^z$, $z - N \ln \langle x \rangle = \xi$. Abbiamo quindi:

$$\langle y_N^q \rangle_{LG} = e^{N(\langle \ln x \rangle q + \frac{1}{2}\sigma_z^2 q^2)} \quad (8.14)$$

Le due espressioni ottenute sono in accordo per $q \ll 1$. Infatti in questo limite è possibile sviluppare il valor medio sulla distribuzione vera tenendo conto che $\langle x^q \rangle = \langle e^{q \ln x} \rangle = \langle 1 + q \ln x + \frac{q^2}{2} \ln^2 x + \mathcal{O}(q^3) \rangle = 1 + q \langle \ln x \rangle + \frac{q^2}{2} \langle \ln^2 x \rangle + \mathcal{O}(q^3)$. Da questa segue:

$$\begin{aligned} \langle y_N^q \rangle_V &= e^{N \ln \langle x^q \rangle_V} \simeq e^{N \ln \left(1 + q \langle \ln x \rangle + \frac{q^2}{2} \langle \ln^2 x \rangle \right)} \simeq \\ &\simeq e^{N \left[q \langle \ln x \rangle + \frac{q^2}{2} \langle \ln^2 x \rangle - \frac{q^2}{2} \langle \ln x \rangle^2 \right]} \simeq e^{N \left[\langle \ln x \rangle q + \frac{1}{2} \sigma_z^2 q^2 \right]} \end{aligned} \quad (8.15)$$

Se si ha che la distribuzione vera è tale da essere definita tra un x minimo e uno massimo, $x_m < x < x_M$ segue che $\langle x_m^q \rangle < \langle x^q \rangle < \langle x_M^q \rangle$ e prendendo il logaritmo si ottiene infine $q \ln x_m < \ln \langle x^q \rangle < q \ln x_M$, cioè il logaritmo del valor medio della potenza q -esima deve essere compreso tra due rette. È evidente che utilizzando l'approssimazione gaussiana questa condizione non può essere rispettata per ogni q , e prima o poi si uscirà dalla zona in cui dovrebbe ricadere il valor medio associato alla distribuzione vera.

Questa limitazione del TCL è evidente anche dalla seguente osservazione. Quando si considerano momenti relativi ad alte potenze q , il contributo al valor medio è dato principalmente da zone relative ad eventi rari o che addirittura non sono permesse (come nel caso di dominio limitato).

8.2.2 Ancora sulla distribuzione dell'energia

Per chi avesse ancora dubbi sulla forma della distribuzione dell'energia proponiamo questo calcolo alternativo valido per ΔE piccolo.

$$\begin{aligned} P(E)\Delta E &= \int_{E < H(\underline{x}) < E + \Delta E} P(\underline{X}) d\underline{X} = \int_{E < H(\underline{x}) < E + \Delta E} e^{-\beta H(\underline{X})} d\underline{X} \simeq \\ &\simeq e^{-\beta E} \int_{E < H(\underline{x}) < E + \Delta E} d\underline{X} = e^{-\beta E} \Gamma(E) = e^{-\beta E} \omega(E) \Delta E \end{aligned}$$

Da cui segue il risultato finale ¹, infatti per $\Delta E \rightarrow 0$:

$$\Gamma(E) = \Sigma(E + \Delta E) - \Sigma(E) = \frac{\Sigma(E + \Delta E) - \Sigma(E)}{\Delta E} \Delta E = \frac{\partial \Sigma}{\partial E} \Delta E = \omega(E) \Delta E$$

¹Non si sono considerati i fattori costanti derivanti dalle funzioni di partizione

8.3 Il vantaggio del metodo di Cramér

Dal Teorema del Limite Centrale, date N variabili i.i.d con distribuzione $P_x(x)$ allora la grandezza $y_N = \frac{1}{N} \sum_n x_n$ è distribuita, nel limite $N \gg 1$, come

$$P_y(y_N) \propto \exp -\frac{N(y_N - \langle x \rangle)^2}{2\sigma^2}$$

tuttavia le uniche caratteristiche della P_x originale contenute nell'approssimazione del TLC sono la media $\langle x \rangle$ e la varianza σ^2 . Questo fatto comporta un problema sulla stima delle code, poiché distribuzioni con stessa media e varianza, ma momenti successivi diversi, tendono alla stessa distribuzione di y_N . Nelle grandi deviazioni invece:

$$P_y(y_N) \propto \exp -NC(y_N)$$

dove la funzione di Cramér $C(y_N)$ dipende dai dettagli della distribuzione $P_x(x)$. Questo vantaggio può risultare più chiaro con l'esempio di un problema affrontato da Boltzmann.

■ **Esempio 8.1 — Problema di Boltzmann.** Consideriamo il seguente problema: vogliamo calcolare la distribuzione della y_N , date N variabili x_n

$$x_n = \begin{cases} 1 & p \\ -1 & (1-p) \end{cases} \quad (8.16)$$

Detto k il numero di volte che si verifica l'evento $x_n = 1$, la frequenza dell'evento sarà quindi data da $f = \frac{k}{N}$, ottenendo per la y_N :

$$y_N = \frac{1}{N} \sum_{n=1}^N x_n = \frac{1}{N} (k - (N - k)) = 2f - 1 \quad (8.17)$$

La probabilità di avere $y_N = 2f - 1$ equivale a dire che si verificano k su N possibili casi, perciò y_N sarà distribuita esattamente come una binomiale:

$$P(y_N = 2f - 1) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \quad (8.18)$$

con valor medio $\langle y_N \rangle = (+1)p + (-1)(1-p) = 2p - 1$ e varianza $\sigma_y^2 = \frac{4}{N}p(1-p)$.

Siamo interessati al limite $k \rightarrow \infty$ e $N \rightarrow \infty$, per cui trattiamo l'eq. (8.18) con l'approssimazione di Stirling:

$$\begin{aligned} P(y_N) &= \exp \{ \ln N! - \ln k! - \ln(N-k)! + k \ln p + (N-k) \ln(1-p) \} = \\ &\simeq \exp \{ N \ln N - k \ln k - (N-k) \ln(N-k) + k \ln p + (N-k) \ln(1-p) \} = \\ &= \exp \left\{ N \ln N - k \ln \left(\frac{k}{p} \right) - (N-k) \ln \left(\frac{N-k}{1-p} \right) \right\} = \\ &= \exp \left\{ (k + (N-k)) \ln N - k \ln \left(\frac{k}{p} \right) - (N-k) \ln \left(\frac{N-k}{1-p} \right) \right\} = \\ &= \exp \left\{ -k \ln \left(\frac{k}{Np} \right) - (N-k) \ln \left(\frac{N-k}{N(1-p)} \right) \right\} = \\ &= \exp -N \left\{ f \ln \left(\frac{f}{p} \right) + (1-f) \ln \left(\frac{1-f}{1-p} \right) \right\} \end{aligned}$$

e ricordando che $f = \frac{k}{N} = \frac{y_N + 1}{2}$ otteniamo infine la seguente relazione:

$$P(y_N) \simeq \exp -N \left\{ \frac{1 + y_N}{2} \ln \left(\frac{1 + y_N}{2p} \right) + \frac{1 - y_N}{2} \ln \left(\frac{1 - y_N}{2(1-p)} \right) \right\} \quad (8.19)$$

ossia otteniamo una forma del tipo $P(y) \propto \exp -NC(y)$. De Moivre nel suo calcolo arrivò allo stesso risultato ma approssimò ulteriormente espandendo parabolicamente l'esponente:

$$\left. \frac{dC}{dy} \right|_{y^*} = 0 \Leftrightarrow y^* = \langle y \rangle \quad \text{e} \quad \left. \frac{d^2C}{dy^2} \right|_{y^*} = \frac{1}{\sigma_y^2}$$

che fornisce lo stesso risultato del limite centrale. Tuttavia si tratta soltanto di una prima approssimazione del risultato della legge dei grandi numeri: in effetti, avendo ottenuto un'espressione esplicita, non ha senso approssimare ulteriormente. ■

8.3.1 Calcolo della funzione di Cramér

In generale, esiste un metodo formale per calcolare la funzione di Cramér $C(y)$ a partire dalle proprietà della distribuzione iniziale. Introduciamo la funzione generatrice dei momenti di y , $g(q) = \langle e^{qNy} \rangle$ e, ricordando che le x_n sono fra loro indipendenti, osserviamo che

$$\langle e^{qNy} \rangle = \langle e^{q \sum_n x_n} \rangle = \langle e^{qx} \rangle^N = e^{NL(q)} \quad (8.20)$$

dove $L(q) = \ln \langle e^{qx} \rangle$ è la funzione generatrice dei cumulanti di x . Se invece calcoliamo esplicitamente $g(q)$ e ci poniamo nel caso $N \gg 1$, per cui $P_y(y) \simeq e^{-NC(y)}$, segue che:

$$g(q) = \int dy P_y(y) e^{qNy} \simeq \int dy e^{-N(C(y)-qy)} \quad (8.21)$$

In generale questo integrale non è risolvibile se non nel caso di $C(y)$ parabolica, a cui non siamo interessati. Volendo una forma analoga all'eq. (8.20), ci serviamo dell'approssimazione di Laplace per grandi N :

$$\int dy e^{N(qy-C(y))} \simeq e^{NL(q)}$$

dove abbiamo infine ricavato un'espressione per la $L(q)$:

$$L(q) \simeq \sup_y \{qy - C(y)\} \quad (8.22)$$

ossia la generatrice dei cumulanti è circa la *trasformata di Legendre* della funzione di Cramér. Questa espressione serve a ben poco se non è invertibile, infatti $L(q)$ è nota, mentre noi siamo interessati a $C(y)$. Nell'ipotesi in cui $C(y)$ è convessa ($C''(y) > 0 \forall y$) la relazione è invertibile:

$$C(y) \simeq \sup_y \{qy - L(q)\} \quad (8.23)$$

Se invece $C(y)$ non è convessa, la relazione non è invertibile e la quantità $\sup \{qy - L(q)\}$ rappresenta l'*inviluppo convesso* di $C(y)$. Ad esempio, la costruzione di Maxwell sulla curva isoterma di Van Der Waals, corrisponde a fare l'inviluppo convesso dell'energia libera, ossia è la minima modifica che rende l'entropia una funzione convessa.

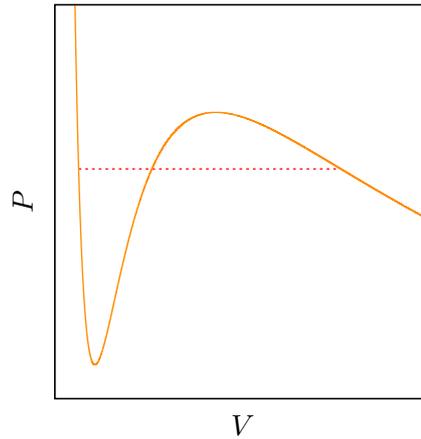


Figura 8.3: Legge di Van Der Waals e costruzione di Maxwell. La costruzione di Maxwell consiste nel trovare il segmento che interseca la curva di Van Der Waals delineando due superfici di ugual misura.

Siccome la condizione di avere una funzione convessa è restrittiva, vorremmo che si trattasse di una proprietà fisica del sistema e non di un'assunzione matematica.

8.3.2 Simmetria di Gallavotti-Cohen

Un sistema all'equilibrio è al massimo dell'entropia, perciò non produce entropia. Un sistema fuori dall'equilibrio invece, è in grado di produrre entropia al tempo t e a questa sarà associata una certa variabile aleatoria (secondo un qualche modello) che indichiamo con $\sigma(t)$. Volendo misurare l'entropia, effettuiamo una media campionaria:

$$\frac{1}{T} \int_0^T dt \sigma(t) = \frac{W(T)}{T} \equiv \widetilde{W}(T)$$

assumendo che nel limite $T \gg 1$ il sistema sia sufficientemente scorrelato da far valere le grandi deviazioni, vorremmo per questa quantità una distribuzione del tipo

$$P_T(\widetilde{W}) \simeq \exp -TC(\widetilde{W})$$

Detta $\text{Tr}(t)$ la traiettoria diretta del sistema al tempo t , a questa sarà associata la quantità \widetilde{W} . Alla traiettoria inversa sarà invece associata $-\widetilde{W}$:

$$\text{Tr}(t) \longleftrightarrow \widetilde{W} \quad \text{inv}(\text{Tr})(t) \longleftrightarrow -\widetilde{W}$$

avremo quindi che il rapporto delle probabilità associate alle traiettorie nel limite $T \gg 1$ è:

$$\frac{P(\text{Tr}(t))}{P(\text{inv}(\text{Tr})(t))} \sim \frac{\exp -TC(\widetilde{W})}{\exp -TC(-\widetilde{W})} = \exp -T [C(\widetilde{W}) - C(-\widetilde{W})]$$

Gallavotti e Cohen dimostrano che questa quantità è uguale a $e^{-T\widetilde{W}}$, ossia vale la **relazione di Gallavotti-Cohen**:

$$C(\widetilde{W}) - C(-\widetilde{W}) = \widetilde{W} \quad (8.24)$$

che rappresenta la frontiera dell'attuale meccanica statistica del non-equilibrio e il maggiore passo in avanti compiuto in questa disciplina dai tempi delle relazioni di Onsager. In qualche modo riconosce

un'ulteriore vincolo dei sistemi, ossia che note le fluttuazioni positive, allora le fluttuazioni negative sono completamente determinate. Per dimostrare questa relazione abbiamo bisogno di un'entropia che tenga in considerazione anche il tempo: introduciamo quindi i lavori di Shannon sulla teoria dell'informazione e l'entropia di Shannon.

8.4 Teoria dell'informazione

Nel 1948, Claude Shannon ingegnere alle industrie Bell, si pose il problema della misura del "disordine" (o dell'imprevedibilità) associato ad un sistema, elaborando la **teoria dell'informazione** poi formalizzata grazie all'intervento matematico di Kolmogorov e Khinchin.

■ **Esempio 8.2** Osservando 2 eventi A e B a cui sono associate delle probabilità in diverse configurazioni:

A	B
0.3	0.7
0.4	0.6
0.5	0.5

risulta intuitivo pensare come configurazione "più disordinata" quella equiprobabile. Per esempio, avendo 2 eventi con probabilità p e $1-p$, possiamo pensare ad una misura del disordine come $-|p-0.5|$, massima nel caso equiprobabile. Non è affatto intuitivo invece capire nel caso di tre o più eventi se una configurazione sia più disordinata di un'altra. ■

Shannon riuscì a trovare una funzione delle probabilità che soddisfacesse le condizioni intuitive cercate, l'**entropia di Shannon**, definita come:

$$H = -\lambda \sum_{n=1}^N P_n \ln P_n \quad (8.25)$$

Dimostrò inoltre che tale funzione è unica a meno di una costante moltiplicativa λ , che tuttavia non ne altera le proprietà². Curiosamente tale funzione già era in circolazione nel mondo della fisica (anche se in forma continua), infatti già Gibbs esprimeva l'entropia nel canonico come:

$$S = -k_B \int ds \rho(s) \ln \rho(s)$$

Teorema 8.4.1 — Teorema di $\exists!$ di Shannon. Dati N eventi a cui è associata una probabilità P_n esiste ed è unica (a meno di una costante moltiplicativa) la funzione H che soddisfa le seguenti proprietà:

1. H è massima quando $P_n = \frac{1}{N}$;
2. H è minima quando $P_n = \delta_{nk}$;
3. H non cambia se considero l'evento $(N+1)$ con $P_{N+1} = 0$;
4. $H(A, B) = H(A) + H(B)$ se A e B sono eventi indipendenti;
5. $H(A, B) = H(A) + H(B|A)$ se A e B sono dipendenti;
6. $H(B|A) \leq H(B)$.

Tale funzione è l'entropia di Shannon.

Dimostrazione. Dimostriamo solamente l'esistenza. Siccome la costante moltiplicativa è irrilevante, può essere omessa.

²L'arbitrarietà della costante è il motivo per cui la base del logaritmo non è rilevante (per esempio in informatica si usa \log_2), in fisica \ln e in matematica \log .

1. Nel nostro caso $f(x) = \ln x$ è concava poiché $f''(x) = -\frac{1}{x^2} < 0$, perciò

$$\langle \ln x \rangle = \sum_n P_n (\ln x_n) \leq \ln(\langle x \rangle) = \ln\left(\sum_n P_n x_n\right)$$

avendo sfruttato il precedente risultato per cui se f è concava, allora . Scegliendo infine $x_n = P_n^{-1}$ la relazione diventa

$$-\sum_n P_n \ln P_n \leq \ln N$$

ossia

$$H(P_1, \dots, P_N) \leq H\left(\frac{1}{N}, \dots, \frac{1}{N}\right)$$

2. Nel caso $P_n = \delta_{nk}$, abbiamo che $P_{n \neq k} = 0$, per cui, interpretando la quantità $0 \ln 0$ come il limite

$$\lim_{\epsilon \rightarrow 0^+} \epsilon \ln \epsilon = 0$$

si ha $H = 0$ che è proprio il valore minimo.

3. Se $P_{N+1} = 0$, analogamente al caso precedente, abbiamo che il termine $P_{N+1} \ln P_{N+1} = 0 \ln 0 = 0$.
 4. Siano gli eventi indipendenti $A = \{A_1, \dots, A_N\}$ con probabilità associate P_n , $B = \{B_1, \dots, B_M\}$ con probabilità associate P_m e calcoliamo:

$$H(A, B) = -\sum_{n,m} P(A_n, B_m) \ln P(A_n, B_m) = \dots$$

ma essendo A e B indipendenti, segue che $P(A_n, B_m) = P(A_n)P(B_m)$, perciò:

$$\begin{aligned} \dots &= -\sum_{n,m} P(A_n)P(B_m) \ln P(A_n) - \sum_{n,m} P(A_n)P(B_m) \ln P(B_m) = \\ &= -\sum_n P(A_n) \ln P(A_n) - \sum_m P(B_m) \ln P(B_m) = \\ &= H(A) + H(B) \end{aligned}$$

avendo sfruttato il fatto che $\sum_n P(A_n) = \sum_m P(B_m) = 1$.

5. Se invece A e B non sono indipendenti $P(A_n, B_m) = P(A_n)P(B_m|A_n)$, per cui:

$$\begin{aligned} H(A, B) &= -\sum_{n,m} P(A_n)P(B_m|A_n) \ln P(A_n) - \sum_{n,m} P(A_n)P(B_m|A_n) \ln P(B_m|A_n) = \\ &= -\sum_n P(A_n) \ln P(A_n) - \sum_n P(A_n) \left[\sum_m P(B_m|A_n) \ln P(B_m|A_n) \right] = \\ &= H(A) - \sum_n P(A_n) H(B|A_n) = H(A) + H(B|A) \end{aligned}$$

avendo sfruttato il fatto che $\sum_m P(B_m|A_n) = 1$.

6. Dalla convessità della funzione $g(x) = -x \ln x$ si ha che

$$\langle -x \ln x \rangle = -\sum_n P_n x_n \ln x_n \leq -\langle x \rangle \ln \langle x \rangle = -\left(\sum_n P_n x_n\right) \ln \left(\sum_n P_n x_n\right)$$

per cui scegliendo $x_n = P(B_m|A_n)$ e $P_n = P(A_n)$, otteniamo:

$$-\sum_n P(A_n)P(B_m|A_n) \ln P(B_m|A_n) \leq -\left(\sum_n P(A_n)P(B_m|A_n)\right) \ln \left(\sum_n P(A_n)P(B_m|A_n)\right)$$

ma per il teorema della probabilità totale $\sum_n P(A_n)P(B_m|A_n) = P(B_m)$, per cui, sommando su tutte le m , si ha infine:

$$-\sum_{n,m} P(A_n)P(B_m|A_n) \ln P(B_m|A_n) \leq -\sum_m P(B_m) \ln P(B_m)$$

ossia

$$H(B|A) \leq H(B)$$

■

Si noti che la proprietà $H(B|A) \leq H(B)$ è la traduzione dell'idea intuitiva che la misura del disordine di un oggetto B di cui posseggiamo delle informazioni A , è minore della misura senza informazioni.

8.4.1 L'entropia di Shannon

Vogliamo introdurre ora in maniera formale l'entropia di Shannon e mostrare come questa possieda le proprietà necessarie che giustificano la sua identificazione con un'entropia.

Immaginiamo di avere una macchina che produce una sequenza di caratteri $\mathcal{S} \doteq \{s_1, s_2, \dots, s_T\}$ con $T \gg 1$ in cui i singoli caratteri s_t siano presi da un alfabeto fissato di dimensione M , cioè $s_t \in \{0, 1, \dots, M-1\}$, in cui abbiamo per semplicità identificato i possibili caratteri dell'alfabeto con i numeri da 0 ad $M-1$. L'obiettivo è dunque quello di costruire una funzione capace di fornire una misura del grado di ordine o disordine o di informazione contenuto all'interno di questa sequenza. Possiamo infatti immaginare che vi sia dietro questa successione una qualche forma di regolarità: una legge deterministica, un processo markoviano; oppure che sia completamente randomica, la nostra funzione dovrà dunque essere in grado di distinguere queste due eventualità.

Per fissare le idee immaginiamo un caso semplice. Consideriamo un processo markoviano a due stati come illustrato in Figura fig. 8.4.

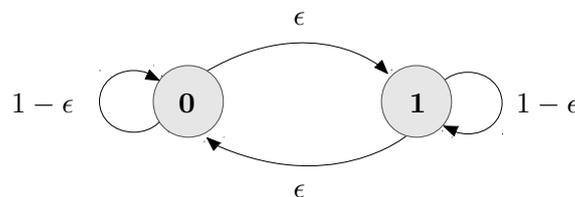


Figura 8.4: Schema di una Catena di Markov a due stati.

Lo stato del sistema definisce il carattere della sequenza, il nostro alfabeto sarà dunque $s_t \in \{0, 1\}$. Analizziamo dunque a seconda del valore del parametro ϵ quali siano i diversi tipi di sequenze che il nostro processo è in grado di generare.

$$\begin{aligned}
\varepsilon \simeq 10^{-3} & : \mathcal{S} \doteq \{0, 0, 0, \dots, 1, 1, 1, \dots, 0, 0, 0, \dots\} \\
\varepsilon = \frac{1}{2} & : \mathcal{S} \doteq \{1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, \dots\} \\
\varepsilon \simeq 1 - 10^{-3} & : \mathcal{S} \doteq \{0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \dots\}
\end{aligned}$$

Osserviamo dunque come esiste una differenza qualitativa fra i tre diversi tipi di sequenze che il nostro sistema propone al variare di ε : la prima e la terza sequenza possiedono un'evidente regolarità, sono infatti caratterizzate dall'alternanza di blocchi di zeri o di uni la cui lunghezza è in media eguale e legata al valore di ε . La seconda sequenza invece è invece completamente diversa: nessuna regolarità appare evidente, al contrario infatti la sequenza risulta completamente randomica e priva di qualunque struttura.

Sottolineiamo come le probabilità invarianti di questo sistema siano indipendenti da ε e pari ad $\mathcal{P}^{(\text{inv})} = 1/2$.

Tentiamo dunque di formalizzare ora la funzione entropia di Shannon. Proseguendo per semplicità nell'ipotesi di alfabeto binario (la generalizzazione ad un alfabeto arbitrario è immediata), consideriamo l'insieme delle sequenze di lunghezza N che indichiamo con W_N . Per fissare le idee si avrà dunque:

$$W_1 = \{0, 1\} \quad ; \quad W_2 = \{00, 01, 10, 11\} \quad ; \quad W_3 = \{000, 001, 010, 011, 100, 101, 110, 111\} \quad \dots$$

In generale l'insieme W_N conterrà 2^N elementi. L'idea è dunque quella di associare a ciascun elemento di un insieme W_N una sua probabilità, cioè $W_N \mapsto \mathcal{P}(W_N)$, e costruire poi l'entropia del blocco W_N come:

$$H_N = - \sum_{\{W_N\}} \mathcal{P}(W_N) \ln \mathcal{P}(W_N) \quad (8.26)$$

Analizziamo questo oggetto nel caso semplice in cui consideriamo i caratteri della sequenza $\mathcal{S} = \{s_t\}_{t=1}^N$ come indipendenti. In questo caso la probabilità di un elemento dell'insieme W_N , dunque di una parola, si scomporrà banalmente in:

$$\mathcal{P}(W_N) = \mathcal{P}(s_1) \mathcal{P}(s_2) \cdots \mathcal{P}(s_N) \quad (8.27)$$

Sostituendo questo oggetto nell'espressione eq. (8.26) otteniamo:

$$\begin{aligned}
H_N &= - \sum_{\{s_1, \dots, s_N\}} \mathcal{P}(s_1) \mathcal{P}(s_2) \cdots \mathcal{P}(s_N) \left[\ln \mathcal{P}(s_1) + \ln \mathcal{P}(s_2) + \cdots + \ln \mathcal{P}(s_N) \right] = \\
&= - \sum_{\{s_2, \dots, s_N\}} \mathcal{P}(s_2) \cdots \mathcal{P}(s_N) \left(\sum_{s_1} \mathcal{P}(s_1) \ln \mathcal{P}(s_1) \right) - \cdots = \\
&= - \sum_{s_1} \mathcal{P}(s_1) \ln \mathcal{P}(s_1) - \sum_{s_2} \mathcal{P}(s_2) \ln \mathcal{P}(s_2) - \cdots - \sum_{s_N} \mathcal{P}(s_N) \ln \mathcal{P}(s_N) = NH_1
\end{aligned}$$

Dove si è indicato con H_1 l'entropia associata a ciascun carattere:

$$H_1 = - \sum_{s_i} \mathcal{P}(s_i) \ln \mathcal{P}(s_i) \quad (8.28)$$

Vorremmo ora giustificare l'uso del termine entropia per questa grandezza evidenziando i parallelismi esistenti fra la funzione H_N di Shannon e l'entropia fisica di Boltzmann definita come $S = k_B \ln W$,

dove qui W è il numero di configurazioni del sistema. Una prima proprietà interessante che vorremmo verificare è *l'estensività*, cioè la dipendenza dell'entropia dal numero di particelle N del sistema: $S(N, \cdot) \simeq Ns(\cdot)$ per $N \gg 1$, dove con $s(\cdot)$ si è indicata l'entropia per particella. Ugualmente vorremo verificare che:

$$\frac{H_N}{N} \xrightarrow{N \rightarrow \infty} h_s \quad ; \quad H_N \simeq Nh_s \quad \text{per } N \rightarrow \infty \quad (8.29)$$

È possibile verificare che questo limite esiste e che risulta inoltre $0 \leq h_s \leq \ln M$. Delineiamo dunque la strategia di dimostrazione generale del teorema prima della sua introduzione formale. Si parte dalla costruzione della relazione $H_N = H_{N-1} + h_N$, si mostra poi che la grandezza $h_N = H_N - H_{N-1}$ è un oggetto decrescente all'aumentare di N e che verifica la catena di disuguaglianze: $0 \leq h_N \leq \ln M$. A questo punto invocando il teorema dei carabinieri (funzione monotona e limitata) si dimostra l'esistenza del limite per $N \rightarrow \infty$.

Teorema 8.4.2 — Esistenza limite Entropia. Si consideri una sequenza $\mathcal{S} = \{s_t\}_{t=1}^N$ con $\mathcal{S} \in W_N$ e l'entropia di Shannon H_N associata al gruppo W_N . È possibile allora costruire la decomposizione:

$$H_N = H_{N-1} + h_N \quad (8.30)$$

Inoltre esiste il limite $\exists \lim_{N \rightarrow \infty} h_N = h_s$ e si può scrivere:

$$H_N \simeq c + Nh_s \quad (8.31)$$

Dove c è un'opportuna costante.

Dimostrazione. Consideriamo innanzitutto l'entropia di Shannon per il gruppo W_N :

$$H_N = - \sum_{\{W_N\}} \mathcal{P}(W_N) \ln \mathcal{P}(W_N)$$

Osserviamo come una parola lunga N possa essere sempre decomposta come una parola di lunghezza $N-1$ cui aggiungo l'ultimo carattere s_N , si avrà cioè $W_N = (W_{N-1}, s_N)$. Potrò allora considerare la probabilità della parola W_N come la probabilità di avere la parola W_{N-1} per la probabilità di trovare nella posizione N il carattere s_N , dato la parola W_{N-1} nei precedenti spazi. In formule:

$$\mathcal{P}(W_N) = \mathcal{P}(W_{N-1}) \mathcal{P}(s_N | W_{N-1})$$

Sostituendo questa scrittura otteniamo:

$$\begin{aligned} H_N = & - \sum_{W_{N-1}, s_N} \mathcal{P}(W_{N-1}) \mathcal{P}(s_N | W_{N-1}) \ln \mathcal{P}(W_{N-1}) + \\ & - \sum_{W_{N-1}, s_N} \mathcal{P}(W_{N-1}) \mathcal{P}(s_N | W_{N-1}) \ln \mathcal{P}(s_N | W_{N-1}) \end{aligned}$$

Osserviamo i due termini che figurano nell'espressione precedente. Nel primo termine posso eseguire la sommatoria su s_N sfruttando la normalizzazione di quest'ultimo, otterremo:

$$\sum_{s_N} \mathcal{P}(W_{N-1}) \mathcal{P}(s_N | W_{N-1}) \ln \mathcal{P}(W_{N-1}) = \mathcal{P}(W_{N-1}) \ln \mathcal{P}(W_{N-1})$$

Riscriviamo invece il secondo termine evidenziando l'entropia condizionata che figura nel seguente modo:

$$- \sum_{W_{N-1}} \left(\sum_{s_N} \mathcal{P}(s_N | W_{N-1}) \ln \mathcal{P}(s_N | W_{N-1}) \right) \mathcal{P}(W_{N-1}) = \sum_{W_{N-1}} h(s_N | W_{N-1}) \mathcal{P}(W_{N-1})$$

Sostituendo nella relazione originaria i due contributi così riscritti otteniamo:

$$\begin{aligned} H_N &= - \sum_{W_{N-1}} \mathcal{P}(W_{N-1}) \ln \mathcal{P}(W_{N-1}) + \sum_{W_{N-1}} h(s_N | W_{N-1}) \mathcal{P}(W_{N-1}) \\ &= H_{N-1} + H(s_N | W_{N-1}) \end{aligned}$$

Identificando ora $h_N \equiv H(s_N | W_{N-1})$ abbiamo mostrato l'espressione eq. (8.30). Per dedurre l'andamento del termine h_N in funzione dell'argomento N possiamo vederlo sostanzialmente come il complemento della probabilità di indovinare il carattere s_N , l'ultimo, conoscendo una parola lunga N caratteri. Osserviamo che aumentando la lunghezza N il carattere s_N rimarrà sempre il singolo carattere alla fine della parola, mentre aumenterà la lunghezza della parola nota, aumenterà dunque l'informazione a nostra disposizione per indovinare il carattere s_N . Concludiamo intuitivamente che l'andamento della probabilità è crescente nell'argomento N , di conseguenza l'entropia h_N , una misura del disordine, del grado di non-informazione, avrà l'andamento opposto: sarà una funzione decrescente in N . Questo aspetto è formalmente sintetizzato nella relazione che un'entropia deve soddisfare: $H(B | A) \leq H(B)$.

Per concludere la dimostrazione è sufficiente osservare ora che la grandezza $h_N \geq 0$ è inferiormente limitata da zero, ovvio trattandosi di somma di un'entropia condizionata, e superiormente limitata da $h_N \leq \ln M$, corrispondente al caso di tutti i caratteri s_N indipendenti ed equiprobabili in un alfabeto di dimensioni M . Possiamo a questo punto invocare il teorema dei carabinieri per concludere che esiste il seguente limite:

$$\exists \lim_{N \rightarrow \infty} h_N = h_s$$

Per ottenere infine l'espressione eq. (8.31) possiamo osservare che per N sufficientemente grande possiamo sostituire ad h_N il suo valore limite h_s ed iterare la sostituzione dell'equazione eq. (8.30) per H_{N-1} e ripetere il procedimento. Si otterrà dunque che per grandi valori di N l'andamento di H_N sarà esprimibile come:

$$H_N \simeq c + Nh_s$$

Dove la costante c è stata introdotta per tener conto delle correzioni introdotte quando si considera la ricorsione fino ad N piccoli. Il teorema risulta così dimostrato. ■

Presentiamo ora un breve esempio di un calcolo esplicito della grandezza $H(s_N | W_{N-1})$ nel caso del processo di Markov descritto dalla Figura fig. 8.4. Partiamo dalla definizione:

$$H(s_N | W_{N-1}) = - \sum_{W_{N-1}} \left[\sum_{s_N} \mathcal{P}(s_N | W_{N-1}) \ln \mathcal{P}(s_N | W_{N-1}) \right] \mathcal{P}(W_{N-1}) \quad (8.32)$$

Osserviamo ora che per un processo di Markov vi è una notevole semplificazione: la definizione stessa di Processo markoviano è formalizzata come nell'assenza di memoria del sistema, lo stato successivo è determinato solo dallo stato precedente e non dalla storia. La conoscenza dell'intera parola W_{N-1} risulta dunque del tutto superflua per la determinazione di s_N , è sufficiente infatti conoscere il carattere immediatamente precedente s_{N-1} . Avremo dunque che, per un processo markoviano, sarà valida la semplificazione:

$$\mathcal{P}(s_N | W_{N-1}) = \mathcal{P}(s_N | s_{N-1}) \quad (8.33)$$

Avremo dunque che nell'espressione eq. (8.32) tutte le somme sui caratteri $\{s_t\}$ con $t \neq \{N, N-1\}$ saranno banalmente uno per la normalizzazione di $\mathcal{P}(W_{N-1})$. Rimarrà dunque:

$$\begin{aligned} H(s_N | W_{N-1}) &= - \sum_{s_N, s_{N-1}} \left(\mathcal{P}_{s_{N-1} \rightarrow s_N} \ln \mathcal{P}_{s_{N-1} \rightarrow s_N} \right) \mathcal{P}_{s_{N-1}} = \\ &= - \sum_{i,j} \mathcal{P}_i^{(\text{inv})} \mathcal{P}_{i \rightarrow j} \ln \mathcal{P}_{i \rightarrow j} \end{aligned}$$

Dove nel secondo passaggio si è semplicemente cambiato nome alle variabili per rendere più esplicito il legame con il formalismo e linguaggio tipico di un processo markoviano. Ricordando che nel sistema markoviano di Figura fig. 8.4 si ha $\mathcal{P}_1^{(\text{inv})} = \mathcal{P}_0^{(\text{inv})} = 1/2$, possiamo sostituire i valori ottenendo la relazione finale:

$$H(s_N | W_{N-1}) = - \left[\varepsilon \ln \varepsilon + (1 - \varepsilon) \ln (1 - \varepsilon) \right] \quad (8.34)$$

È immediato verificare che questa funzione ha un massimo per $\varepsilon = 1/2$ e si annulla per i valori estremali $\varepsilon = 0, \varepsilon = 1$. Sottolineiamo come questa distinzione sia l'analisi quantitativa offerta dell'entropia di Shannon per il sistema di cui prima si era potuta offrire solo uno studio qualitativo.

Vogliamo ora approfondire il legame fra l'entropia di Shannon h_s e l'entropia di Boltzmann $S = k_b \ln W$, in particolare vogliamo mostrare che la celebre relazione $W \simeq \exp\left(N \frac{S}{k_B}\right)$ per il calcolo del numero di configurazioni (sottolineiamo qui il ruolo dell'entropia come contatore logaritmico) ammette un diretto corrispettivo per l'entropia di Shannon. Ciò sarà dunque l'ulteriore e definitiva prova che l'entropia di Shannon è di fatto un'entropia. Questo risultato è formalizzato nel Teorema di Shannon-McMillan.

Teorema 8.4.3 — Shannon-McMillan. Si consideri l'insieme delle parole di lunghezza N indicato con $\{W_N\}$. Per $N \gg 1$ si distinguono due classi $\Omega_N^{(1)}$ ed $\Omega_N^{(2)}$:

$$W_N \in \begin{cases} \Omega_N^{(1)} \\ \Omega_N^{(2)} \end{cases} \quad \text{dove} \quad \begin{cases} \sum_{W_N \in \Omega_N^{(1)}} \mathcal{P}(W_N) \xrightarrow{N \rightarrow \infty} 1 \\ \sum_{W_N \in \Omega_N^{(2)}} \mathcal{P}(W_N) \xrightarrow{N \rightarrow \infty} 0 \end{cases} \quad (8.35)$$

Si avrà allora che se $W_N \in \Omega_N^{(1)}$ avremo:

$$\mathcal{P}(W_N) \simeq e^{-h_s N} \quad (8.36)$$

Inoltre, indicando con $\mathcal{N}(N) =$ numero di parole in $\Omega_N^{(1)}$, si avrà:

$$\mathcal{N}(N) \simeq e^{N h_s} \quad (8.37)$$

Dimostrazione. Dimostriamo questo teorema nei due casi in cui abbiamo una sorgente di caratteri $\{s_t\}_{t=1}^N$ indipendenti ed identicamente distribuiti (i.i.d) ed il caso di sorgente markoviana.

Caso variabili i.i.d.

Consideriamo una successione di caratteri i.i.d $\{s_t\}_{t=1}^N$ che compone la parola di lunghezza N :

$$W_N = (s_1, s_2, \dots, s_N) \quad ; \quad s_n \in \{0, 1, \dots, M-1\}$$

Sfruttando l'indipendenza dei singoli caratteri possiamo facilmente scrivere la probabilità dell'intera parola $\mathcal{P}(W_N)$ come semplice prodotto delle probabilità dei singoli caratteri:

$$\mathcal{P}(W_N) = \prod_{n=1}^N \mathcal{P}(s_n) = \exp \left[\sum_{n=1}^N \ln \mathcal{P}(s_n) \right]$$

Introduciamo ed indichiamo con il simbolo n_j il numero di volte che il carattere $s_n = j$ nella parola W_N . Sfruttando questa nuova variabile possiamo riscrivere l'espressione precedente per la probabilità della parola.

$$\mathcal{P}(W_N) = \exp \left[\sum_{n=1}^N \ln \mathcal{P}(s_n) \right] = \exp \left[\sum_j n_j(W_N) \ln \mathcal{P}(s_n = j) \right] = \exp \left[\sum_j n_j(W_N) \ln \mathcal{P}_j \right]$$

Possiamo a questo punto moltiplicare e dividere l'esponente che compare nell'espressione precedente per N , otteniamo dunque le frequenze dell'occorrenza del simbolo $s_n = j$:

$$\frac{n_j(W_N)}{N} \equiv f_j(W_N)$$

Abbiamo dunque di fatto ridotto l'espressione della probabilità ad una forma che richiama la struttura dell'entropia di Shannon dove tuttavia figurano le frequenze e non le probabilità, ci domandiamo dunque se nel limite per $N \gg 1$ risulti verificato:

$$\exists? \lim_{N \rightarrow \infty} f_j(W_N) = P_j$$

Osserviamo come questa sia esattamente la tesi della legge dei grandi numeri, invocando i risultati possiamo dunque affermare che:

$$\mathcal{P}(W_N) = \exp \left[-N \left(- \sum_j f_j(W_N) \ln \mathcal{P}_j \right) \right] \xrightarrow{N \rightarrow \infty} \exp \left[-N \left(- \sum_j P_j \ln P_j \right) \right] = e^{-Nh_s}$$

Non solo, il teorema del limite centrale (e le grandi deviazioni) garantisce inoltre che la probabilità di deviazione di $f_j(W_N)$ dal suo valore asintotico P_j è soppressa esponenzialmente all'aumentare di N . Possiamo dunque identificare a questo punto due classi $\Omega_N^{(1)}$ ed $\Omega_N^{(2)}$ dove:

$$\begin{aligned} \Omega_N^{(1)} &:= \left\{ W_N \in \{W_N\} : \lim_{N \rightarrow \infty} f_j(W_N) = P_j \right\} \\ \Omega_N^{(2)} &:= \left\{ W_N \in \{W_N\} : \lim_{N \rightarrow \infty} f_j(W_N) \neq P_j \right\} \end{aligned}$$

Per quanto enunciato precedentemente concludiamo dunque che:

$$\begin{aligned} \mathcal{P}(W_N \in \Omega_N^{(1)}) &\xrightarrow{N \rightarrow \infty} 1 \\ \mathcal{P}(W_N \in \Omega_N^{(2)}) &\xrightarrow{N \rightarrow \infty} 0 \end{aligned}$$

La classe $\Omega_N^{(1)}$ diviene dunque totalmente dominante nel limite $N \gg 1$ e risultano dimostrata le tesi eq. (8.35) ed eq. (8.37) in quanto:

$$\sum_{W_N \in \Omega_N^{(1)}} \mathcal{P}(W_N) \simeq \sum_{W_N \in \Omega_N^{(1)}} e^{-Nh_s} = \mathcal{N}(N) e^{-Nh_s} = 1 \quad ; \quad N \rightarrow \infty \quad (8.38)$$

$$\mathcal{N}(N) e^{-Nh_s} = 1 \quad \Rightarrow \quad \mathcal{N}(N) = e^{Nh_s} \quad (8.39)$$

Caso Processo Markoviano

Consideriamo ora il caso di una sequenza generata da un processo di Markov. I singoli caratteri $\{s_t\}_{t=1}^N$ non saranno più indipendenti, tuttavia sfruttando la proprietà di memoria esclusiva dello stato precedente possiamo sempre scomporre la probabilità della parola $\mathcal{P}(W_N)$ come:

$$\mathcal{P}(W_N) = P(s_1) \mathcal{P}_{s_1 \rightarrow s_2} \mathcal{P}_{s_2 \rightarrow s_3} \cdots \mathcal{P}_{s_{N-1} \rightarrow s_N}$$

A questo punto lo schema concettuale della dimostrazione è assolutamente analogo. Introduciamo le due quantità:

$$f_j(W_N) \equiv \frac{n_j}{N} = \frac{1}{N} (\text{numero di volte } s_n = j)$$

$$F_{j \rightarrow i} \equiv \frac{(\text{numero di salti } j \rightarrow i, s_n = j, s_{n+1} = i)}{(\text{numero di volte } s_n = j)}$$

Possiamo a questo punto costruire la catena di uguaglianze:

$$\mathcal{P}(W_N) = \mathcal{P}(s_1) \exp \left[\sum_{n=1}^{N-1} \ln \mathcal{P}_{s_n \rightarrow s_{n+1}} \right] = \mathcal{P}(s_1) \exp \left[N \sum_{i,j} f_j F_{j \rightarrow i} \ln \mathcal{P}_{j \rightarrow i} \right]$$

A questo punto invochiamo l'ergodicità del Processo di Markov per affermare che le frequenze f_j ed $F_{j \rightarrow i}$ definite (la prima la frequenza invariante del sito $s_n = j$, la seconda la frequenza di salto da j ad i) per $N \gg 1$ tendono alle probabilità $f_j \xrightarrow{N \rightarrow \infty} P_j^{(\text{inv})}$ ed $F_{j \rightarrow i} \xrightarrow{N \rightarrow \infty} P_{j \rightarrow i}$. Non solo: la probabilità che ciò non avvenga è soppressa esponenzialmente come per il caso di variabili i.i.d, dunque si definiscono anche in questo caso le due classi $\Omega_N^{(1)}$ ed $\Omega_N^{(2)}$. Scriveremo infine:

$$\mathcal{P}(W_N) \simeq \mathcal{P}(s_1) \exp \left[-N \left(- \sum_{i,j} P_j^{(\text{inv})} P_{j \rightarrow i} \ln P_{j \rightarrow i} \right) \right] \simeq e^{-Nh_s} \quad ; \quad N \rightarrow \infty$$

Essendo la derivazione delle eq. (8.35) e delle eq. (8.37) assolutamente analoga al caso precedente, il teorema risulta dimostrato. ■

Vogliamo ora sottolineare un aspetto molto importante, spesso frainteso, di questo teorema e cioè che la classe $\Omega_N^{(1)}$, dominante nel limite per $N \gg 1$ **non** contiene la parola W_N più probabile.

Lo schema della ripartizione delle regioni di probabilità nelle classi $\Omega_N^{(1)}$ ed $\Omega_N^{(2)}$ è illustrata in figura fig. 8.5. Per comprendere il perché di questa struttura è sufficiente considerare il lancio di una moneta truccata. Immaginiamo di avere:

$$T := \begin{cases} 0 & p = 0.9 \\ 1 & p = 0.1 \end{cases} \quad (8.40)$$

Nonostante la sequenza più probabile sia $\mathbf{S} = \{0, 0, 0, \dots\}$ in realtà non ci aspettiamo di trovare questa successione quando il numero di elementi della successione è tale che $N \gg 1$, questo perché, nonostante questa risulti effettivamente la sequenza più probabile, è una sola, mentre la sproporzionata maggioranza delle sequenze ricade nella classe $\Omega_N^{(1)}$ (notare il taglio degli assi) dove si avrà in media $\bar{n}_0 = 90\%$ ed $\bar{n}_1 = 10\%$. La crescita del numero di sequenze con un numero proporzionato alla probabilità di zeri ed uno cresce infatti con un fattore combinatorio, dunque esponenziale.

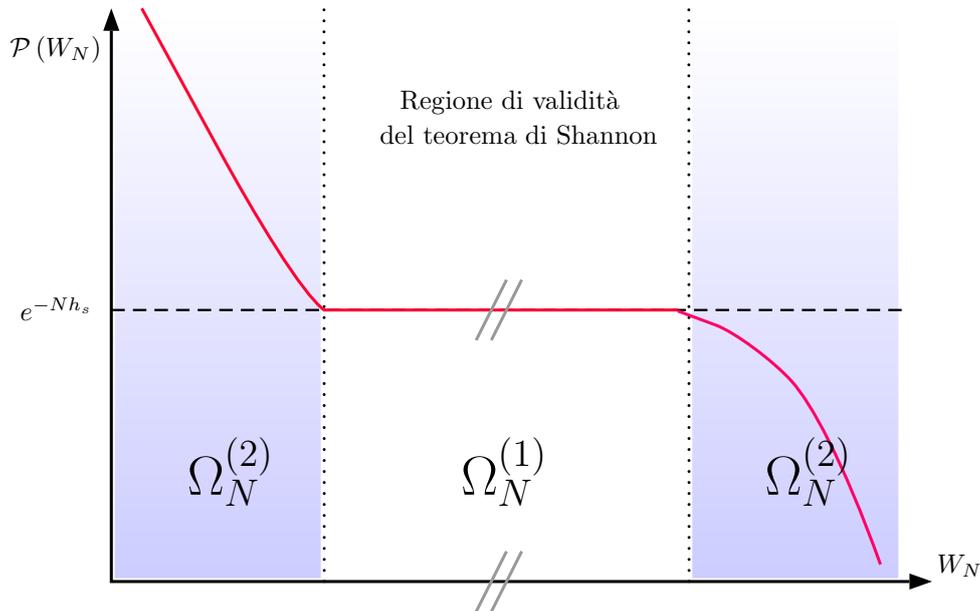


Figura 8.5: Grafico andamento delle probabilità della parola e identificazione delle regioni $\Omega_N^{(1)}$ ed $\Omega_N^{(2)}$

8.4.2 Comprimità dell'informazione

Preso un alfabeto di M simboli (o più in generale M stati) $(0, \dots, M-1)$ consideriamo una sequenza di questi lunga T (S_1, S_2, \dots, S_T) , generata tramite un processo stocastico definito sui suddetti stati. Al processo sarà associata una certa entropia di Shannon $h_S \in [0, \ln M]$, cosicché le sequenze di T simboli che nel $\lim_{T \rightarrow \infty}$ hanno probabilità $\rightarrow 1$ di essere generate (ciò quelle appartenenti alla 'classe buona', $\Omega_N^{(1)}$ secondo la notazione adottata nella scorsa lezione) saranno $\mathcal{N}_{eff}(T) \sim e^{Th_S}$.

Vogliamo comprimere la sequenza (S_1, S_2, \dots, S_T) in una di L caratteri $(\sigma_1, \sigma_2, \dots, \sigma_L)$, con $L < T$. A priori vorremmo che il rapporto $\frac{T}{L}$ sia il più grande possibile, invece si può dimostrare - e ne daremo un breve argomento in un caso particolare $M = 2$ - che

$$\frac{T}{L} \leq \frac{\ln M}{h_S} \quad (8.41)$$

cioè esiste un limite teorico nella comprimibilità di una sequenza di caratteri, dato dall'entropia di Shannon del processo con cui essi vengono estratti. Questo vuol dire che nella sequenza iniziale di T simboli c'è "informazione" - quantificata da h_S - che si deve conservare nella compressione (se vogliamo poi essere in grado di decomprimere $(\sigma_1, \sigma_2, \dots, \sigma_L)$ e farla tornare la sequenza originale). Come anche intuitivo che sia, l'eq. (8.41) ci dice che se $h_S = \ln M$ (cioè se tutti i simboli che compongono la sequenza sono equiprobabili per ogni sito) allora la comprimibilità che possiamo ottenere è nulla: la sequenza è completamente random e non ci sarà alcuna regolarità che possiamo utilizzare a nostro vantaggio nella costruzione di una codifica alternativa dell'informazione; se invece $h_S = 0$ avremo la più totale regolarità della sequenza (un solo simbolo che si ripete T volte) e dunque l'informazione in essa contenuta risulterà completamente comprimibile nel simbolo che la costituisce e nella lunghezza della sequenza stessa.

Per chiarire meglio l'idea consideriamo un **processo Markoviano a due stati**. Si noti come il caso di un alfabeto binario sia di particolare interesse nell'ambito informatico e delle telecomunicazioni. La probabilità invariante per ciascuno stato è $P_{inv} = \frac{1}{2}$ e da essa abbiamo calcolato nella scorsa lezione

l'entropia di Shannon del processo $h_S = -[p \ln(p) + (1-p) \ln(1-p)]$. Nel limite di $p \sim 0$ o di $p \sim 1$ - in cui la sequenza estratta sarà molto regolare - si può approssimare $h_S \sim p \ln\left(\frac{1}{p}\right) \ll 1$. Questo ci dice che le sequenze che il processo genererà saranno molto comprimibili, vediamo in che modo.

Nel limite in cui ci siamo messi una tipica sequenza estratta - che equivale a dire una sequenza della classe 'buona' - sarà del tipo $\underbrace{00\dots0}_{n_1} | \underbrace{11\dots1}_{n_2} | \dots | \dots | \underbrace{\dots}_{n_N}$. Questa sequenza è comprimibile in $0, n_1, n_2, \dots, n_N$,

cioè nel simbolo con cui inizia la sequenza seguito dal numero di volte che, alternandosi, ciascun simbolo si ripete. Tale codifica è perfetta e non ci fa perdere alcuna informazione sulla stringa di caratteri. Ma di quanto siamo riusciti a comprimere la sequenza inizialmente lunga, diciamo, T ?

Sappiamo che $\sum_{j=1}^k n_j = T$ e che nel $\lim_{T \rightarrow \infty}$, per la legge dei grandi numeri, $T \simeq k \langle n \rangle = kN$. Dunque il nuovo messaggio che dovremo inviare sarà costituito da un primo simbolo più k numeri $n_i \sim O(N)$, in binario ci serviranno quindi $O(k \log_2 N)$ bit di informazione. Se consideriamo che $N \sim \frac{1}{p}$ otteniamo che il costo del nuovo messaggio sarà circa

$$k \log_2 N = \frac{T}{N} \log_2(N) = Tp \log_2\left(\frac{1}{p}\right) \quad (8.42)$$

Si noti che questo risultato è in accordo con eq. (8.41), che nel caso specifico in analisi ci dice $\frac{T}{L} \leq \frac{\ln 2}{p \ln\left(\frac{1}{p}\right)}$ e - tramite la formula del cambiamento di base per i logaritmi - restituisce come limite teorico proprio il risultato da noi ottenuto. Diamo ora una intuizione sul perché **questa è proprio la codifica migliore**.

Sappiamo che $\mathcal{N}_{eff}(N) \sim e^{N h_S}$; se prendiamo una sequenza di T simboli (S_1, S_2, \dots, S_T) possiamo porre un limite inferiore sul costo del messaggio ottimamente compresso. Dividiamo la parola di T simboli in sotto-sequenze, ciascuna di N simboli. Otteniamo così $k = \frac{T}{N}$ parole $W_N^{(1)}, W_N^{(2)}, \dots, W_N^{(k)}$. Il caso migliore ai fini della compressione della sequenza è che tutte le $W_N^{(i)}$ siano massimamente stereotipate e quindi appartenenti alla classe $\Omega_N^{(1)}$ delle $\sim e^{N h_S}$ parole equamente distribuite. Codifichiamo queste parole associandole ai numeri che vanno da 1 a $e^{N h_S}$. In questo modo il costo per trasmettere una parola appartenente alla classe $\Omega_N^{(1)}$ sarà - poiché siamo nel caso di un alfabeto binario - dell'ordine $\log_2 e^{N h_S} = \frac{N h_S}{\ln 2}$. Ne segue che il costo della parola lunga T compressa è

$$L = k * costo(W_N^{(i)}) = \frac{T}{N} \frac{N h_S}{\ln 2} = \frac{T h_S}{\ln 2} \quad (8.43)$$

Ricordando che ci siamo limitati al caso ottimale, in cui tutto le parole lunghe N appartengano alla classe 'buona', è chiaro come questo risultato costituisca un limite inferiore per il costo di una parola lunga T codificata. Questo risultato è in accordo con quello più generale dell'eq. (8.41).

8.4.3 Entropia topologica

Ora anziché domandarci quale sia il numero 'effettivo' di parole lunghe N - che per grandi N sia più probabile ottenere - vogliamo sapere il numero totale $\mathcal{N}_{tot}(N)$ di sequenze di stati ottenibili come frutto di un certo processo stocastico. In pratica in $\mathcal{N}_{tot}(N)$ non ci deve essere alcuna informazione legata alla probabilità di ciascuna parola, deve essere solo un contatore di tutte quelle possibili. Sicuramente sappiamo che esse saranno $\mathcal{N}_{tot}(N) \leq M^N$, numero di parole possibili qualora gli stati venissero estratti indipendentemente da una distribuzione di probabilità definita su di essi e non nulla su ciascuno. Ma se vogliamo fare di meglio possiamo definire h_T l'*entropia topologica* del processo. Vogliamo che essa sia tale che

$$\mathcal{N}_{tot}(N) = e^{N h_T}$$

Già sappiamo che di certo $h_S \leq h_T$ poiché $\mathcal{N}_{eff}(N) \leq \mathcal{N}_{tot}(N)$ dove l'uguaglianza vale solo se le parole sono tutte equiprobabili (cioè per $h_S = h_T = \ln M$).

■ **Esempio 8.3** Consideriamo una catena di Markov a 2 stati come quella in [fig. 8.6](#). In questo tipo di processo il simbolo 1 può essere solo preceduto e seguito dallo 0, questo riduce il numero di sequenze possibili rispetto al caso in cui entrambi gli stati siano sempre disponibili; ma di quanto?

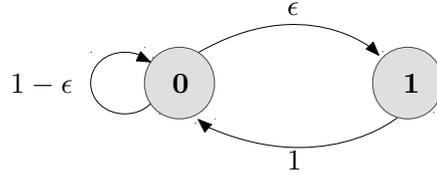


Figura 8.6: Catena di Markov a 2 stati, in cui lo stato 1 può andare solo in 0

Vedremo che, mentre il $\mathcal{N}_{eff}(N)$ per un processo Markoviano - e dunque h_S - dipende dalle probabilità invarianti dello stesso e dalle probabilità di transizione tra stati, invece $\mathcal{N}_{tot}(N)$ - quindi h_T - non vi dipenderà, come del resto ci aspettiamo. Definiamo $\mathcal{N}_i(n)$ il numero di parole che il processo può generare lunghe n e che finiscano con l' i -esimo simbolo dell'alfabeto, appare evidente che nel nostro caso $\mathcal{N}_{tot}(N) = \mathcal{N}_1(N) + \mathcal{N}_0(N)$. Del resto una parola che finisca in 1 avrà come penultimo stato per forza lo 0 $\rightarrow \mathcal{N}_1(N) = \mathcal{N}_0(N-1)$, mentre un se ho come stato finale lo 0 esso può essere preceduto sia da 0 che da 1 $\rightarrow \mathcal{N}_0(N) = \mathcal{N}_0(N-1) + \mathcal{N}_1(N-1) = \mathcal{N}_{tot}(N-1)$. Quindi abbiamo

$$\mathcal{N}_{tot}(N) = \mathcal{N}_{tot}(N-1) + \mathcal{N}_0(N-1) = \mathcal{N}_{tot}(N-1) + \mathcal{N}_{tot}(N-2) \quad (8.44)$$

dove nell'ultimo passaggio è stato iterato il ragionamento fatto per $\mathcal{N}_0(N)$. Riconosciamo nella [eq. \(8.44\)](#) la regola di ricorsione di Fibonacci; la dividiamo per $\mathcal{N}_{tot}(N-1)$ ottenendo l'equazione

$$\frac{\mathcal{N}_{tot}(N)}{\mathcal{N}_{tot}(N-1)} = 1 + \frac{\mathcal{N}_{tot}(N-2)}{\mathcal{N}_{tot}(N-1)} \quad (8.45)$$

Definendo $\alpha = \frac{\mathcal{N}_{tot}(N)}{\mathcal{N}_{tot}(N-1)}$ nel limite di $N \rightarrow \infty$ riusciamo a chiudere la [eq. \(8.45\)](#) asintoticamente come

$$\alpha = 1 + \frac{1}{\alpha} \quad (8.46)$$

ovvero $\alpha^2 - \alpha - 1 = 0$ di cui l'unica soluzione fisicamente sensata è quella positiva $\alpha = \frac{1+\sqrt{5}}{2} \equiv \varphi$, ossia la **sezione aurea** (non ci stupisce, essendo connesso alla successione di Fibonacci). Perciò asintoticamente $\mathcal{N}_{tot}(N) = \phi^N = e^{N \ln \phi}$ e $h_T = \ln \phi$. ■

Esiste tuttavia un risultato più generale per l'**entropia topologica per un processo markoviano**:

■ **Esempio 8.4** Immaginiamo di avere un processo come [fig. 8.7](#) e facciamo un ragionamento simile a quello precedente

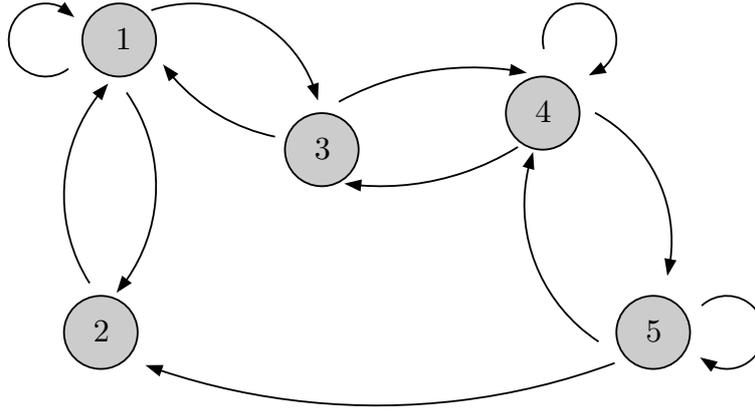


Figura 8.7: Catena di Markov a 5 stati, sono indicati solo gli stati connessi, senza le relative probabilità di transizione, dal momento che h_T non dipenderà da queste.

Avremo sicuramente che

$$\mathcal{N}_{tot}(N) = \sum_{i=1}^5 \mathcal{N}_i(N) \quad (8.47)$$

con

$$\begin{cases} \mathcal{N}_1(N) = \mathcal{N}_1(N-1) + \mathcal{N}_2(N-1) + \mathcal{N}_3(N-1) \\ \mathcal{N}_2(N) = \mathcal{N}_1(N-1) + \mathcal{N}_5(N-1) \\ \mathcal{N}_3(N) = \mathcal{N}_1(N-1) + \mathcal{N}_4(N-1) \\ \mathcal{N}_4(N) = \mathcal{N}_3(N-1) + \mathcal{N}_4(N-1) + \mathcal{N}_5(N-1) \\ \mathcal{N}_5(N) = \mathcal{N}_4(N-1) + \mathcal{N}_5(N-1) \end{cases} \quad (8.48)$$

Costruendo la matrice \hat{A} t.c.

$$A_{ij} = \begin{cases} 1 & \text{se } P_{ij} \neq 0 \\ 0 & \text{se } P_{ij} = 0 \end{cases} \quad (8.49)$$

il sistema eq. (8.48) si può riscrivere in forma vettoriale come

$$\begin{bmatrix} \mathcal{N}_1(N) \\ \vdots \\ \mathcal{N}_5(N) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{\hat{A}} \begin{bmatrix} \mathcal{N}_1(N-1) \\ \vdots \\ \mathcal{N}_5(N-1) \end{bmatrix} \quad (8.50)$$

che può essere iterata fino a risalire a

$$\underline{\mathcal{N}}(N) = \hat{A}^N \underline{\mathcal{N}}(1) \quad (8.51)$$

L'eq. (8.51) è vera in generale; chiamando λ l'autovalore massimo in modulo della matrice \hat{A} , avremo che, nel $\lim N \rightarrow \infty$

$$\mathcal{N}_{tot}(N) = \sum_i \mathcal{N}_i(N) \sim \lambda^N = e^{N \ln \lambda} \quad (8.52)$$

poiché ogni componente di $\underline{\mathcal{N}}(N)$ scala asintoticamente come λ^N , per cui anche una loro combinazione lineare scalerà ugualmente. Dall'eq. (8.52) risulta chiaro che $h_T = \ln \lambda$.

Si può verificare facilmente che l'autovalore massimo della matrice \hat{A} associata a quel processo è φ . In conclusione sottolineiamo nuovamente come h_T sia una grandezza poco informativa, dal momento che pesa equamente tutte le parole possibili, a prescindere dalla loro probabilità di essere generate. Essa è molto sensibile alla presenza di probabilità di transizione estremamente piccole: se infatti per h_S fa poca differenza che un certo $P_{ij} = 10^{-9}$ piuttosto che sia nullo, per h_T la differenza che questo induce è significativa.

8.5 Collegamento tra teoria dell'informazione e non equilibrio

Immaginiamo di avere un processo stocastico su stati discreti a tempi discreti (Markoviano), il ragionamento che faremo è estendibile al caso continuo. Allo scorrere del tempo il sistema compierà delle traiettorie sugli stati, che possono essere immaginate come parole lunghe t (secondo la terminologia che abbiamo adottato in teoria dell'informazione).

Chiamiamo $T(t) = (s_1, s_2, \dots, s_t)$ la generica traiettoria compiuta dal sistema nel tempo t e $TI(t) = (s_t, s_{t-1}, \dots, s_1)$ la sua inversa. Come noto per i processi Markoviani possiamo calcolare la probabilità delle traiettorie come:

$$P[T(t)] = P^{(s)}(s_1)P_{s_1 \rightarrow s_2}P_{s_2 \rightarrow s_3} \dots P_{s_{t-1} \rightarrow s_t}$$

e analogamente $P[TI(t)] = P^{(s)}(s_t)P_{s_t \rightarrow s_{t-1}}P_{s_{t-1} \rightarrow s_{t-2}} \dots P_{s_2 \rightarrow s_1}$, dove l'apice (s) sta a indicare la probabilità stazionaria. Chiaramente se vale il bilancio dettagliato - che ricordiamo essere una proprietà estendibile ad ogni tempo - avremo che $P[T(t)] = P[TI(t)]$. La probabilità delle traiettorie (o parole) appartenenti alla classe $\Omega_N^{(1)}$ è, per $t \gg 1$, $P[T(t)] \sim e^{-th_S}$; dunque se vale il bilancio dettagliato anche $P[TI(t)] \sim e^{-th_S}$.

Definiamo la quantità

$$\tilde{W} = \frac{1}{t} \ln \frac{P[T(t)]}{P[TI(t)]} \quad (8.53)$$

Chiaramente si ha che

$$\tilde{W} = \begin{cases} = 0 & \text{se vale il Bilancio Dettagliato} \\ \neq 0 & \text{altrimenti} \end{cases}$$

che indica, a meno di costanti moltiplicative, quanta entropia sta producendo la traiettoria $T(t)$ per unità di tempo. Riprendiamo la relazione di Gallavotti-Cohen

$$C(\tilde{W}) - C(-\tilde{W}) = -\tilde{W} \quad (8.54)$$

che unita alla relazione generale della teoria delle grandi deviazioni, per $t \gg 1$

$$P_t(\tilde{W}) \sim e^{-tC(\tilde{W})} \quad (8.55)$$

abbiamo visto come conduca a una riformulazione probabilistica del secondo principio (le traiettorie che producono un'entropia pari a \tilde{W} sono esponenzialmente più probabili di quelle che la diminuiscono di $-\tilde{W}$)

$$\frac{P_t(\tilde{W})}{P_t(-\tilde{W})} \sim e^{t\tilde{W}} \quad (8.56)$$

Per vederlo è sufficiente la sostituzione:

$$\frac{P_t(\tilde{W})}{P_t(-\tilde{W})} \simeq \frac{\exp[-tC(\tilde{W})]}{\exp[-tC(-\tilde{W})]} = \exp\{-t[C(\tilde{W}) - C(-\tilde{W})]\} = e^{t\tilde{W}} \quad (8.57)$$

Un altro risultato ottenibile a partire da Gallavotti-Cohen è che, nel caso limite di un processo Gaussiano ($C(\tilde{W})$ quadratica) si ha che il valor medio dell'entropia prodotta per unità di tempo è la metà della sua varianza:

$$2\langle\tilde{W}\rangle = \sigma_{\tilde{W}}^2$$

Questa uguaglianza vale tra grandezze adimensionali. Infatti usando la eq. (8.54) si ha che

$$\begin{aligned} -\tilde{W} &= \frac{(\tilde{W} - \langle\tilde{W}\rangle)^2}{2\sigma^2} - \frac{(-\tilde{W} - \langle\tilde{W}\rangle)^2}{2\sigma^2} = \\ &= \frac{\cancel{\tilde{W}^2} + \cancel{\langle\tilde{W}\rangle^2} - 2\tilde{W}\langle\tilde{W}\rangle - \cancel{\tilde{W}^2} - \cancel{\langle\tilde{W}\rangle^2} - 2\tilde{W}\langle\tilde{W}\rangle}{2\sigma^2} \end{aligned}$$

perciò:

$$2\langle\tilde{W}\rangle = \sigma_{\tilde{W}}^2$$

Questo risultato è ottenibile anche a partire da un'equazione di Langevin con **bias** F (che rompe il bilancio dettagliato) del tipo

$$\dot{v} = -\frac{v}{\tau} + \frac{F}{m} + \sqrt{2c\eta} \quad (8.58)$$

svolgendo esplicitamente il calcolo di $\langle\tilde{W}\rangle$ e di $\sigma_{\tilde{W}}^2$. Se avessimo $F = 0$ troveremmo $\langle x(t)\rangle_0 = 0$, $\langle v(t)\rangle_0 = 0$ e per $t \gg 1$

$$\langle x^2(t)\rangle_0 = 2Dt = 2\left(\frac{k_B T}{m}\tau\right)t$$

dal momento che $D = \langle v^2\rangle_0 \tau = \frac{k_B T}{m}\tau$. Quando invece $F \neq 0$ abbiamo $\langle v\rangle_F = \frac{F}{m}\tau$.

Se ora osserviamo il sistema evolvere su un Δt quanto vale $W(\Delta t)$? Sarà dato dal rapporto tra il lavoro compiuto sul sistema in quel Δt da F e la temperatura T a cui esso si trova: $W(\Delta t) \equiv \frac{F\Delta x}{T}$. Nel continuo quindi avremo $W(t) = \frac{Fx(t)}{T}$, con valor medio

$$\langle W(t)\rangle = \frac{F}{T}\langle x(t)\rangle_F = \frac{F}{T}\langle v\rangle_F t = \frac{F^2}{Tm}\tau t$$

Da questo otteniamo

$$\langle\tilde{W}(t)\rangle \equiv \frac{W(t)}{t} = \frac{F^2}{Tm}\tau \quad (8.59)$$

Invece per $\sigma_{\tilde{W}}$, ponendo $k_B \equiv 1$ in D cosicché $\langle W\rangle$ e σ^2 abbiano la stessa dimensione, abbiamo:

$$\frac{\langle W(t)^2\rangle}{t} = \frac{F^2}{T^2}\frac{\langle x^2\rangle_F}{t} \simeq \frac{F^2}{T^2}\frac{\langle x^2\rangle_0}{t} = \frac{F^2}{T^2}\frac{2T}{m}\tau = 2\frac{F^2}{Tm}\tau \quad (8.60)$$

dove abbiamo approssimato, per F piccolo, $\langle x^2\rangle_F \simeq \langle x^2\rangle_0$. Abbiamo ritrovato $2\langle\tilde{W}\rangle = \sigma_{\tilde{W}}^2$.

8.6 Relazione di Gallavotti-Cohen

Sebbene ne abbiamo già fatto uso più volte, ancora non abbiamo giustificato come venga fuori l'eq. (8.54). Il calcolo originale di Gallavotti è abbastanza tortuoso, per questo ne presentiamo una versione più didatticamente utile (Marini, Puglisi, Rondoni, Vulpiani, 2008).

Riprendiamo la quantità definita nella eq. (8.53) - stavolta non per unità di tempo - e esplicitiamola nel caso di un processo Markoviano. Essa sarà

$$W(t) = \ln \frac{P(T(t))}{P(TI(t))} \simeq \sum_{j=1}^t \ln \frac{P_{s_j \rightarrow s_{j+1}}}{P_{s_{j+1} \rightarrow s_j}} \quad (8.61)$$

con l'approssimazione dovuta al fatto che ci sarà un contributo $O(1)$ che dipende dalla probabilità stazionaria del primo stato delle due traiettorie

Definiamo $P_i(t, W) = \text{prob}(\text{essere nello stato } i \text{ al tempo } t \text{ avendo prodotto un'entropia pari a } W)$. Allora avremo che

$$P_i(t+1, W) = \sum_j P_{j \rightarrow i} P_j(t, W - \Delta W_{ji}) \quad (8.62)$$

dove $\Delta W_{ji} = \ln \frac{P_{j \rightarrow i}}{P_{i \rightarrow j}}$ come si deduce dalla eq. (8.61).

Svolgiamo la trasformata di Laplace di $P_i(t+1, W)$, ricordando che $\mathcal{L}[P_i(t, W)] \equiv \tilde{P}_i(t, \lambda) = \int dW e^{-\lambda W} P_i(t, W)$.

$$\tilde{P}_i(t+1, \lambda) = \int dW \sum_j P_{j \rightarrow i} P_j(t, W - \Delta W_{ji}) \underbrace{e^{-\lambda(W - \Delta W_{ji})} e^{-\lambda \Delta W_{ji}}}_{e^{-\lambda W}} \quad (8.63)$$

Ma, ancora dalla eq. (8.61), $e^{-\lambda \Delta W_{ji}} = (P_{j \rightarrow i})^{-\lambda} (P_{i \rightarrow j})^\lambda$, quindi

$$\tilde{P}_i(t+1, \lambda) = \sum_j \int dW e^{-\lambda(W - \Delta W_{ji})} \underbrace{(P_{j \rightarrow i})^{1-\lambda} (P_{i \rightarrow j})^\lambda}_{A_{ij}(\lambda)} P_j(t, W - \Delta W_{ji}) \quad (8.64)$$

che può essere riscritta in forma compatta tramite la matrice $\hat{A}(\lambda)$ definita nella eq. (8.64)

$$\tilde{P}(t+1, \lambda) = \hat{A}(\lambda) \tilde{P}(t, \lambda) \quad (8.65)$$

Questa formula ricorsiva ha una struttura analoga alla eq. (8.48), perciò seguiamo un ragionamento identico a quello adottato per il calcolo dell' h_T di un generico processo Markoviano. Definito $\zeta(\lambda)$ come il logaritmo del più grande autovalore di $\hat{A}(\lambda)$, avremo che per $t \gg 1$

$$\tilde{P}(t, \lambda) \sim e^{t\zeta(\lambda)} \quad (8.66)$$

Dove abbiamo anche sommato sullo stato i di arrivo, operazione che - come detto per la eq. (8.52) - non discosta l'ordine di grandezza del risultato da quello della singola componente $\tilde{P}_i(t, \lambda)$. Ora bisogna svolgere l'antitrasformata di Laplace

$$P(t, W) = \int_{-i\infty}^{+i\infty} \tilde{P}(t, \lambda) e^{\lambda W} d\lambda \sim \int e^{t\zeta(\lambda) + \lambda W} d\lambda = \int e^{t[\zeta(\lambda) + \lambda \tilde{W}]} d\lambda \quad (8.67)$$

L'integrale finale della eq. (8.67), in generale, non si sa risolvere. Ma dato che siamo nel limite di $t \gg 1$ possiamo svilupparlo asintoticamente (metodo di Laplace) e limitarci a cercare l'estremo superiore dell'esponente, cosicché

$$P(t, W) \sim e^{t \sup_\lambda (\lambda \tilde{W} + \zeta(\lambda))} \quad (8.68)$$

Se ne evince che, affinché la eq. (8.66) sia in accordo con la eq. (8.55), $\sup_{\lambda}(\lambda\tilde{W} + \zeta(\lambda)) \equiv -C(\tilde{W})$. Vedremo che la eq. (8.54) si ottiene come vincolo sulla $C(\tilde{W})$, imposto dalla proprietà di cui $\hat{A}(\lambda)$ gode $\hat{A}^T(\lambda) = \hat{A}(1 - \lambda)$. Questa si verifica immediatamente ricordando che $A_{ij}(\lambda) = (P_{j \rightarrow i})^{1-\lambda}(P_{i \rightarrow j})^{\lambda}$ e quindi $A_{ij}^T(\lambda)(P_{i \rightarrow j})^{1-\lambda}(P_{j \rightarrow i})^{\lambda}$. Poiché la $C(\tilde{W})$ è connessa alla $\hat{A}(\lambda)$ tramite la dipendenza da $\zeta(\lambda)$, anche essa dovrà risentire di un vincolo, che è dato proprio dalla eq. (8.54).

Abbiamo ottenuto che per $t \gg 1$ vale per la relazione:

$$P(W, t) \sim \exp t \sup_{\lambda} [\lambda\tilde{W} + \zeta(\lambda)] \quad (8.69)$$

Calcolando la derivata rispetto a λ della funzione $\lambda\tilde{W} + \zeta(\lambda)$, il massimo si ha per un valore λ_* che la annulla, ovvero quando si realizza la condizione

$$\left. \frac{d\zeta(\lambda)}{d\lambda} \right|_{\lambda_*} = -\tilde{W} \quad (8.70)$$

da cui, ricordando che $\sup_{\lambda}[\lambda\tilde{W} + \zeta(\lambda)] = -C(\tilde{W})$, si ottiene

$$C(\tilde{W}) = -\lambda_*\tilde{W} - \zeta(\lambda_*) \quad (8.71)$$

Procedendo analogamente per $-\tilde{W}$, si ottiene che la derivata della funzione $-\lambda\tilde{W} + \zeta(\lambda)$ si annulla quando $\lambda = \lambda_1$, ovvero

$$\left. \frac{d\zeta(\lambda)}{d\lambda} \right|_{\lambda_1} = \tilde{W} \quad (8.72)$$

da cui

$$C(-\tilde{W}) = \lambda_1\tilde{W} - \zeta(\lambda_1) \quad (8.73)$$

Ricordando che $\zeta(\lambda)$ è il logaritmo del più grande autovalore della matrice \hat{A} , dalla relazione $\hat{A}^T(\lambda) = \hat{A}(1 - \lambda)$ segue che $\zeta(\lambda) = \zeta(1 - \lambda)$ e quindi che

$$\lambda_1 = 1 - \lambda_* \quad (8.74)$$

ottenendo così dalla eq. (8.73)

$$C(-\tilde{W}) = \tilde{W} - \lambda_*\tilde{W} - \zeta(\lambda_*) \quad (8.75)$$

Sottraendo la eq. (8.75) alla eq. (8.71) si ricava la relazione di Gallavotti-Cohen:

$$C(\tilde{W}) - C(-\tilde{W}) = -\tilde{W} \quad (8.76)$$

8.6.1 Considerazioni sulla relazione di Gallavotti

Dalla relazione di Gallavotti si possono muovere alcune considerazioni sull'entropia e il Secondo Principio della Termodinamica. Abbiamo già osservato come la relazione di Gallavotti-Cohen sia assolutamente analoga a:

$$\frac{P_t(\tilde{W})}{P_t(-\tilde{W})} = e^{t\tilde{W}} \quad (8.77)$$

Se ora immaginiamo di fissare $\tilde{W} = -a < 0$ e di identificare con \tilde{W} lo scambio di entropia del sistema, che potrà dunque essere di produzione o di assorbimento (dunque di diminuzione globale), otteniamo

che è effettivamente possibile osservare diminuzioni d'entropia, tuttavia la probabilità di quest'ultime risulterà esponenzialmente soppressa nel tempo:

$$\mathcal{P}_t(-a) = \mathcal{P}_t(a) e^{-at} \quad (8.78)$$

Per un sistema sufficientemente grande e per tempi lunghi dunque la probabilità di osservare tale diminuzione tenderà a zero (in realtà essendo il taglio esponenziale, maggiore la diminuzione d'entropia $\tilde{W} = -a$ che si desidera osservare, minore sarà il tempo sufficiente affinché la probabilità di quest'ultima risulti trascurabile).

Risulta quindi un fondamento statistico per il Secondo Principio della Termodinamica: per sistemi macroscopici l'entropia è una funzione non decrescente nel tempo. Osserviamo infine come la Gallavotti-Cohen sia nel limite gaussiano correlata con la legge di Einstein per il coefficiente di Diffusione nel moto browniano.

Riferimenti bibliografici

- [1] U. Marini Bettolo Marconi, A. Puglisi, L. Rondoni e A. Vulpiani. "Fluctuation- dissipation: Response theory in statistical physics". In: *Phys. Rep.* 461.111 (2008).
- [2] L. Onsager e S. Machlup. "Fluctuations and Irreversible Processes". In: *Phys. Rev.* 91.1505 (1953).
- [3] G. Boffetta e A. Vulpiani. *Probabilità in Fisica*. Springer- Verlag, 2012.

Bibliografia

- [1] G. Boffetta e A. Vulpiani. *Probabilità in Fisica*. Springer- Verlag, 2012.
- [2] P. Castiglione, M. Falcioni, A. Lesne e A. Vulpiani. *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge University Press, 2008.
- [3] M. Cencini, F. Cecconi e A. Vulpiani. *Chaos: From Simple Models to Complex Systems*. World Scientific, 2009.
- [4] S.R. de Groot e P. Mazur. *Non-equilibrium Thermodynamics*. Dover, 1984.
- [5] M. Falcioni e A. Vulpiani. *Meccanica Statistica Elementare*. Springer- Verlag, 2014.
- [6] M. Kardar. *Statistical Physics of Particles*. Cambridge University Press, 2007.
- [7] U. Marini Bettolo Marconi, A. Puglisi, L. Rondoni e A. Vulpiani. “Fluctuation- dissipation: Response theory in statistical physics”. In: *Phys. Rep.* 461.111 (2008).
- [8] H.B.G. Casimir. “On Onsager’s Principle of Microscopic Reversibility”. In: *Rev. Mod. Phys.* 17.343 (1945).
- [9] L. Onsager e S. Machlup. “Fluctuations and Irreversible Processes”. In: *Phys. Rev.* 91.1505 (1953).