

DIPARTIMENTO
DI MATEMATICA



SAPIENZA
UNIVERSITÀ DI ROMA

Introduzione alle equazioni cinetiche per il corso di IFM 2024-2025

© 2024 di Dario Benedetto con licenza attribuzione - non commerciale
- condividi allo stesso modo 4.0 internazionale CC BY-NC-SA 4.0



Dario Benedetto - <http://brazil.mat.uniroma1.it/dario>

Sapienza Università di Roma
Dipartimento di Matematica
Piazzale Aldo Moro n. 5, 00185 Roma
www.mat.uniroma1.it

Introduzione alle equazioni cinetiche per il corso di IFM 2024/2025

30 maggio 2025

Queste note sono “appunti” e non “dispense”: vuol dire che trovate gli argomenti e le dimostrazioni, ma le motivazioni sono piuttosto stringate.

Possibile materiale di approfondimento

- Dispense di E. Caglioti <https://sites.google.com/site/ecaglioti/didattica/MR> su
 - Gerarchia e equazione di Vlasov in “Dispense III”
 - Limite di campo medio in “Il limite di campo medio”
 - Equazione di Boltzmann e teorema H in “Dispense III”
 - Limite idrodinamico per l’equazione di Boltzmann in “Hydrodynamic limit for the Boltzmann Equation”

- C. Cercignani, R. Illner, M. Pulvirenti **The Mathematical Theory of Dilute Gases** Applied Mathematical Sciences 106 - Springer

- Sul problema del passaggio da una descrizione irreversibile a una reversibile potete guardare sul testo precedente, e anche informarvi sul “modello di Kac”, un modellino che contiene molte delle difficoltà concettuali del passaggio al limite da sistemi reversibili a sistemi irreversibili.

G.A. Gottwald, M. Oliver: *Boltzmann’s Dilemma: An Introduction to Statistical Mechanics via the Kac Ring* SIAM Review **51** 3, pp 613–635 2009

<https://epubs.siam.org/doi/pdf/10.1137/070705799>

- C. Villani: *Topics in Optimal Transportation* Graduate Studies in Mathematics vol. 58, Springer 2003

This is the first comprehensive introduction to the theory of mass transportation with its many—and sometimes unexpected—applications. In a novel approach to the subject, the book both surveys the topic and includes a chapter of problems, making it a particularly useful graduate textbook.

- F. Santambrogio: *Optimal Transport for Applied Mathematicians* PNLDE vol. 87, Birkhäuser 2015

Testo che contiene anche sviluppi più recenti, in particolare le interpretazioni di varie importanti equazioni come flussi gradiente di particolari funzionali rispetto a opportune metriche.

- H. Brezis: *Analisi funzionale* Liguori 1983

Da qui ho preso il materiale per l'appendice sulla trasformata di Legendre in spazi di Banach.

Indice

1	Dall'equazione di Liouville alla gerarchia	4
2	Equazioni di campo medio	7
2.1	Proprietà dell'equazione di Vlasov	9
2.2	Altri modelli	9
2.2.1	Il modello per il flocking di Cucker e Smale	10
2.2.2	Il modello di Kuramoto per la sincronizzazione	10
2.2.3	Un algoritmo per la determinazione dei minimi	11
2.2.4	Una descrizione unificata	11
2.3	La distanza di Wasserstein	12
2.4	Il problema di Monge e il problema di Kantorovich	18
2.4.1	La distanza delle misure empiriche dalla legge	19
2.5	Costruzione delle soluzioni	21
3	Il limite di campo medio	24
3.1	La convergenza alla Dobrushin	24
3.2	La convergenza della gerarchia	24
4	L'equazione di Boltzmann	26
4.1	Urti tra sfere dure	26
4.2	La gerarchia	28
4.3	Il limite di Boltzmann-Grad	33
4.4	Proprietà dell'equazione di Boltzmann	34
4.5	Il limite idrodinamico	37
A	Disuguaglianza triangolare per la distanza di Wasserstein	40
B	La derivazione classica della gerarchia	41
C	La trasformata di Legendre	43
C.1	Funzioni convesse	44
C.2	Funzioni semi continue inferiori	44
C.3	La trasformata di Legendre	46

La descrizione fisico-matematica dei sistemi macroscopici, come gas, fluidi, solidi (in generale “mezzi continui”, in cui la scala non permette l'analisi delle particelle che li compongono), si fa attraverso opportune equazioni, che vengono derivate in vari modi. Uno è quello di utilizzare i principi primi dei sistemi di particelle (conservazione della massa, bilancio della quantità di moto e dell'energia). In questo modo però non si ottengono equazioni chiuse e vanno definite delle relazioni tra alcune delle grandezze in gioco attraverso le “relazioni costitutive”. Un altro approccio è quello di tentare di determinare il comportamento di un sistema macroscopico a partire dal comportamento dei suoi costituenti atomici (suggerito

anche da Hilbert nel suo sesto problema). Seguendo questa idea, si possono “passare al limite” le equazioni per un sistema di N particelle, mandando N a infinito, ottenendo delle equazioni macroscopiche, in un qualche senso più “fondate” delle equazioni che si ottengono imponendo relazioni costitutive. Questo modo di procedere viene chiamato spesso “derivazione formale” delle equazioni.

La vera sfida consiste nel mostrare la **validità** delle equazioni macroscopiche, cioè di derivarle rigorosamente dai modelli microscopici: si tratta di mostrare che le soluzioni del sistema microscopico convergono in qualche senso alle soluzioni delle equazioni macroscopiche. Si noti che questo fatto non è assicurato dalla convergenza formale delle equazioni.

Per alcune equazioni cinetiche (e per alcuni sistemi idrodinamici) questo programma ha avuto parziale successo.

1 Dall'equazione di Liouville alla gerarchia

Le equazioni cinetiche descrivono sistemi di molte particelle, come un gas, in cui l'incognita è la densità di probabilità di una particella nello spazio delle fasi, che indicherò con $f(t, \mathbf{x}, \mathbf{v})$ (o anche $f_t(\mathbf{x}, \mathbf{v})$). Da questa funzione si possono ottenere le misure di qualunque osservabile. In particolare

$$\rho(t, \mathbf{x}) = \int f(t, \mathbf{x}, \mathbf{v}) d\mathbf{v}$$

è la densità di probabilità spaziale della particella. Se ci sono N particelle,

$$N \int_A \rho(t, \mathbf{x}) d\mathbf{x}$$

sarà (circa) il numero di particelle nella regione A .

Quello che però sappiamo di un sistema di N particelle è che la sua dinamica è governata dalle equazioni di Newton, che sono equivalenti all'equazione di Liouville, la cui incognita è la densità di probabilità $f^N(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{v}_1, \dots, \mathbf{v}_N)$, funzione di posizione e velocità di ogni particella. Nota f^N , otteniamo la distribuzione della probabilità della prima particella nello spazio integrando nelle variabili delle altre. Indico con $\mathbf{x}_k^N = (\mathbf{x}_k, \mathbf{x}_{k+1}, \dots, \mathbf{x}_N)$ e uso l'analogia notazione \mathbf{v}_k^N .

$$f_1(\mathbf{x}_1, \mathbf{v}_1) = \int d\mathbf{x}_2^N d\mathbf{v}_2^N f^N(\mathbf{x}_1^N, \mathbf{v}_1^N)$$

Se invece integrassimo nelle variabili relative alla particella 1 e alle particelle 3, 4, \dots , N , otterremmo la distribuzione della seconda particella, e così via.

Se vogliamo sperare di ottenere una descrizione del sistema usando solo la distribuzione di una particella, questa distribuzione non può dipendere dal nome della particella. La condizione affinché sia così è che f^N sia invariante per permutazione degli indici delle particelle. In fisica questa proprietà di f^N viene descritta come proprietà delle particelle, che in questo caso vengono dette **indistinguibili**. Per esempio se $N = 3$, sotto ipotesi di indistinguibilità

$$f_1^3(\mathbf{x}, \mathbf{v}) = \int d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{v}_2 d\mathbf{v}_3 f^3(\mathbf{x}, \mathbf{x}_2, \mathbf{x}_3, \mathbf{v}, \mathbf{v}_2, \mathbf{v}_3) = \int d\mathbf{x}_1 d\mathbf{x}_3 d\mathbf{v}_1 d\mathbf{v}_3 f^3(\mathbf{x}_1, \mathbf{x}, \mathbf{x}_3, \mathbf{v}_1, \mathbf{v}, \mathbf{v}_3)$$

etc., cioè f_1^3 è sia la distribuzione della prima particella, sia la distribuzione della seconda, sia la distribuzione della terza.

è dunque evidente che se conosco f^N conosco f_1^N . Se voglio che f_1^N rappresenti tutto il sistema devo invece poter ottenere f^N conoscendo f_1^N . In generale fissare le marginali non permette di determinare la distribuzione, ma c'è un caso facile in cui questo accade, ed è quello in cui assumiamo che le particelle siano distribuite **indipendentemente** con una certa f . In tal caso, poiché per ogni i la particella i -esima ha densità di probabilità $f(\mathbf{x}_i, \mathbf{v}_i)$, per indipendenza si ha che f^N è la distribuzione prodotto:

$$f^N(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{v}_1, \dots, \mathbf{v}_N) = \prod_{i=1}^N f(\mathbf{x}_i, \mathbf{v}_i).$$

Questa proprietà prende il nome di **caos molecolare**, ed equivale proprio a pensare che la coppia posizione - velocità delle particelle siano variabili aleatorie i.i.d..

Introduco una notazione frequente in teorie cinetiche che ci sarà utile anche in altri contesti. Sia $g_1 : \Omega_1 \rightarrow \mathbb{R}$, e $g_2 : \Omega_2 \rightarrow \mathbb{R}$, la funzione “prodotto tensore” di g_1 e g_2 è definita su $\Omega_1 \times \Omega_2$ come

$$(g_1 \otimes g_2)(\mathbf{z}_1, \mathbf{z}_2) = g_1(\mathbf{z}_1)g_2(\mathbf{z}_2).$$

La notazione si estende al prodotto tensore di più funzioni e, in particolare, data $g : \Omega \rightarrow \mathbb{R}$, la sua “tensorizzata” j volte è $g^{\otimes j} : \Omega^j \rightarrow \mathbb{R}$ definita come

$$g^{\otimes j}(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_j) = \prod_{i=1}^j g(\mathbf{z}_i).$$

In questa notazione la condizione di caos molecolare è

$$f^N = f^{\otimes N}, \quad \text{con } f = f_1^N$$

(la condizione che f sia la marginale a una particella si ottiene integrando).

Riassumendo, se il sistema è fatto di particelle identiche ed è in uno stato di caos molecolare, allora conoscere la distribuzione di una particella permette di descrivere l'intero sistema. La domanda che dobbiamo farci è se le equazioni che governano la dinamica consentono a queste due assunzioni di essere verificate. Le equazioni della meccanica permettono di descrivere l'evoluzione della funzione f^N , attraverso l'equazione di Liouville. Per comprendere se è possibile ottenere una descrizione “ridotta” del sistema, cioè attraverso la sola distribuzione di una particella, è necessario studiarla.

Considero un sistema di particelle di massa unitaria, interagenti a coppie con una forza di potenziale F di potenziale $U(\mathbf{x})$, cioè $\mathbf{F}(\mathbf{x}) = -\partial_x U(\mathbf{x})$.

$$\begin{cases} \dot{\mathbf{x}}_i = \mathbf{v}_i \\ \dot{\mathbf{v}}_i = \mathbf{F}_i = \sum_{j \neq i} \mathbf{F}(\mathbf{x}_i - \mathbf{x}_j) = -\sum_{j \neq i} \partial_x U(\mathbf{x}_i - \mathbf{x}_j) \end{cases} \quad (1.1)$$

Nei casi fisici più interessanti, U è il potenziale gravitazionale (attrattivo) o il potenziale coulombiano repulsivo, o un potenziale “empirico” di interazione molecolare come quello di Lennard-Jones. Qui per semplicità assumerò che U sia un potenziale simmetrico, repulsivo, a supporto compatto.

Poiché le variabili $(\mathbf{x}_i, \mathbf{v}_i)$ sono canoniche coniugate, il sistema conserva la misura e l'equazione di Liouville coincide con l'equazione del trasporto:

$$\partial_t f^N + \sum_{i=1}^N \mathbf{v}_i \cdot \partial_{\mathbf{x}_i} f^N + \sum_{i=1}^N \mathbf{F}_i \cdot \partial_{\mathbf{v}_i} f^N = 0$$

(si noti che anche in caso di massa diversa da 1 il sistema rimane a divergenza nulla e l'equazione di Liouville coincide con l'equazione del trasporto).

Lemma 1.1. *Si assuma che f^N al tempo $t = 0$ sia invariante per permutazioni degli indici delle particelle. Allora f^N è invariante a tutti i tempi.*

Dimostrazione per esercizio.

Scriviamo l'equazione per le distribuzioni marginali a j particelle, sotto l'ipotesi di invarianza per permutazioni. Per definizione, la marginale a k particelle è

$$f_j^N(\mathbf{x}_1^j, \mathbf{v}_1^j) = \int f^N(\mathbf{x}_1^N, \mathbf{v}_1^N) d\mathbf{x}_{j+1}^N d\mathbf{v}_{j+1}^N.$$

Si noti che per l'ipotesi di simmetria, considerare la marginale nelle prime j variabili, o nelle ultime j , o in qualsiasi sottoinsieme delle j particelle dà la stessa funzione. Supponendo che U sia regolare e che il dato iniziale sia regolare a supporto compatto, si ottiene facilmente che f^N al tempo t è regolare e a supporto compatto. Questo fatto ci permetterà di scambiare integrali e derivate senza problemi.

Integro l'equazione di Liouville in $\mathbf{x}_{j+1}^N, \mathbf{v}_{j+1}^N$. Il termine $\partial_t f^N$ diventa $\partial_t f_j^N$. In $\sum_{i=1}^n \mathbf{v}_i \cdot \partial_{\mathbf{x}_i} f^N$ i termini con $i \leq j$ e con $i > j$ danno contributi differenti: se $i > j$, scrivo $\mathbf{v}_i \cdot \partial_{\mathbf{x}_i} f^N = \text{div}_{\mathbf{x}_i}(\mathbf{v}_i f^N)$ il cui integrale nella variabile $d\mathbf{x}_i$ è nullo perché f^N è a supporto compatto. Se $i \leq j$, posso scambiare l'ordine di integrazione e derivazione. Complessivamente, il termine di trasporto dà

$$\sum_{i=1}^j \mathbf{v}_i \cdot \partial_{\mathbf{x}_i} f_j^N.$$

Resta da capire cosa accade al termine con le forze, che è

$$\sum_{i=1}^N \sum_{h \neq i} \int F(\mathbf{x}_i - \mathbf{x}_h) \cdot \partial_{\mathbf{v}_i} f^N d\mathbf{x}_1^j d\mathbf{v}_1^j.$$

Gli integrali dei termini con $\partial_{\mathbf{v}_i}$, $i > j$ sono nulli, perché f^N è a supporto compatto. Rimane dunque solo la somma per i da 1 a j . Separo la somma su h in $h \leq j$ e $h > j$. La prima somma dà l'analogo a j particelle del termine presente nell'equazione di Liouville:

$$\sum_{i=1}^j \sum_{h \neq i} F(\mathbf{x}_i - \mathbf{x}_h) \partial_{\mathbf{v}_i} f_j^N.$$

L'altro termine contiene, per $i \leq j$, la somma per $h > j$ di

$$\int F(\mathbf{x}_i - \mathbf{x}_k) \cdot \partial_{\mathbf{v}_i} f^N d\mathbf{x}_{j+1}^N d\mathbf{v}_{j+1}^N$$

Questa espressione è indipendente dall'indice k per l'invarianza per permutazioni. Fisso dunque $k = j + 1$, e la somma su h dà $N - j$ termini uguali a

$$\int F(\mathbf{x}_i - \mathbf{x}_{j+1}) \cdot \partial_{v_i} f_{j+1}^N dx_{j+1} dv_{j+1}.$$

In conclusione, l'equazione per la marginale a j particelle è

$$\begin{aligned} \partial_t f_j^N + \sum_{i=1}^j \mathbf{v}_i \cdot \partial_{x_i} f_j^N + \sum_{i=1}^j \sum_{h \neq i}^j F(\mathbf{x}_i - \mathbf{x}_h) \cdot \partial_{v_i} f_j^N + \\ + \sum_{i=1}^j (N - j) \int F(\mathbf{x}_i - \mathbf{x}_{j+1}) \partial_{v_i} f_{j+1}^N(\mathbf{x}_1^{j+1}, \mathbf{v}_1^{j+1}) d\mathbf{x}_{j+1} d\mathbf{v}_{j+1} = 0. \end{aligned}$$

In particolare l'equazione per la distribuzione a una particella è

$$\partial_t f_1^N + \mathbf{v}_1 \cdot \partial_{x_1} f_1^N + (N - 1) \int F(\mathbf{x}_1 - \mathbf{x}_2) \partial_{v_1} f_2^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) d\mathbf{x}_2 d\mathbf{v}_2 = 0.$$

Come si vede, l'equazione per j particelle coinvolge l'equazione per $j + 1$ particelle, quindi non abbiamo ottenuto equazioni chiuse, ma solo una **gerarchia** di equazioni, che prende il nome di gerarchia BBGKY (da Bogoliubov e altri quattro che l'hanno ricavata indipendentemente). L'ultima equazione, quella per $j = N$, è l'equazione di Liouville, che invece è un'equazione chiusa.

La speranza di ottenere una descrizione ridotta del sistema, cioè un'equazione chiusa in f_1 , è legata alla possibilità di soddisfare i due seguenti aspetti:

- l'ultimo termine deve essere di ordine $O(N^0)$
- $f_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)$ deve potersi scrivere in termini di f_1 .

Quest'ultima condizione si verifica nel caso di particelle indipendentemente e identicamente distribuite, perché in tal caso f_2 si fattorizza in $f_1(\mathbf{x}_1, \mathbf{v}_1)f_1(\mathbf{x}_2, \mathbf{v}_2)$. Però assumere che il dato iniziale è fattorizzato, cioè che $f^N(\mathbf{x}_1^N, \mathbf{v}_1^N, 0) = \prod_{i=1}^N f(\mathbf{x}_i, \mathbf{v}_i, 0)$ non è sufficiente: ogni interazione crea correlazione tra le particelle. Infatti si può verificare immediatamente che la gerarchia BBGKY (e dunque l'equazione di Liouville) non ammette soluzioni fattorizzate (lo si provi per esercizio).

2 Equazioni di campo medio

Nel cosiddetto limite di campo medio, si assume che l'intensità dell'interazione sia di ordine $1/N$, dunque il sistema di particelle viene modificato in

$$\begin{cases} \dot{\mathbf{x}}_i = \mathbf{v}_i \\ \dot{\mathbf{v}}_i = \mathbf{F}_i = \frac{1}{N} \sum_{j \neq i} \mathbf{F}(\mathbf{x}_i - \mathbf{x}_j) = -\frac{1}{N} \sum_{j \neq i} \partial_x U(\mathbf{x}_i - \mathbf{x}_j) \end{cases} \quad (2.1)$$

Nel limite $N \rightarrow +\infty$, la correlazione che si crea tra una particella e l'altra sarà di ordine $1/N$, dunque ci si può aspettare che la gerarchia limite supporti soluzioni fattorizzate (cioè

nel limite le particelle sono indipendenti). D'altra parte il contributo totale delle interazioni è di ordine 1, dunque l'equazione per la marginale a una particella sarà non banale. L'equazione per f_1^N è

$$\partial_t f_1^N + \mathbf{v}_1 \cdot \partial_{\mathbf{x}_1} f_1^N + (N-1) \frac{1}{N} \int F(\mathbf{x}_1 - \mathbf{x}_2) \partial_{\mathbf{v}_1} f_2^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) d\mathbf{x}_2 d\mathbf{v}_2 = 0,$$

Si vede che se N tende a infinito si ottiene un'espressione che non dipende da N , che rimane non banale. Supponiamo, formalmente, che f_j^N convergono a delle funzioni f_j . Passando al limite, le equazioni che si ottengono sono:

$$\partial_t f_j + \sum_{i=1}^j \mathbf{v}_i \cdot \partial_{\mathbf{x}_i} f_j + \sum_{i=1}^j \int F(\mathbf{x}_i - \mathbf{x}_{j+1}) \partial_{\mathbf{v}_i} f_{j+1}(\mathbf{x}_1^{j+1}, \mathbf{v}_1^{j+1}) d\mathbf{x}_{j+1} d\mathbf{v}_{j+1} = 0,$$

in cui il contributo di interazione tra le coppie di particelle è svanito, perché di ordine $1/N$, e ne rimane traccia solo nell'ultimo termine.

Il fatto nuovo è che queste equazioni ammettono una soluzione fattorizzata, come mostreremo. In particolare, scriviamo l'equazione per f_1 , ipotizzando che f_2 si fattorizzi in $f_1(\mathbf{x}_1, \mathbf{v}_1) f_1(\mathbf{x}_2, \mathbf{v}_2)$. L'equazione per f_1 è l'equazione di campo medio, che chiamerò un po' impropriamente equazione di Vlasov:

$$\partial_t f(\mathbf{x}, \mathbf{v}) + \mathbf{v} \cdot \partial_{\mathbf{x}} f(\mathbf{x}, \mathbf{v}) + \left(\int F(\mathbf{x} - \mathbf{y}) f(\mathbf{y}, \mathbf{w}) d\mathbf{y} d\mathbf{w} \right) \cdot \partial_{\mathbf{v}} f(\mathbf{x}, \mathbf{v}) = 0.$$

Lemma 2.1. *Si supponga che $f(t, \mathbf{x}, \mathbf{v})$ risolva l'equazione di Vlasov. La gerarchia limite è soddisfatta da $f_j(\mathbf{x}_1^j, \mathbf{v}_1^j) = \prod_{i=1}^j f(t, \mathbf{x}_i, \mathbf{v}_i)$.*

Spiego perché si chiama campo medio. Ma prima devo richiamare una delle varie versioni della **legge dei grandi numeri**.

Teorema 2.1. *Siano \mathbf{z}_i , $i = 1 \dots N$, variabili aleatorie i.i.d. di media nulla e varianza σ . Allora il valore atteso di $\frac{1}{N} |\sum_i \mathbf{z}_i|$ tende a 0.*

In altri termini, se $\mu(d\mathbf{z})$ è la legge comune alle variabili \mathbf{z}_i ,

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \int |\sum_i \mathbf{z}_i| \mu^{\otimes N}(d\mathbf{z}_1^N) = 0$$

La prova è facile: la variabile aleatoria $\bar{\mathbf{z}} = \frac{1}{N} \sum_i \mathbf{z}_i$ ha media nulla e, per indipendenza delle \mathbf{z}_i ha come varianza la somma delle varianze, cioè σ^2/N . Dunque

$$\int |\bar{\mathbf{z}}| \mu^{\otimes N}(d\mathbf{z}_1^N) = \int_{|\bar{\mathbf{z}}| < \varepsilon} |\bar{\mathbf{z}}| \mu^{\otimes N}(d\mathbf{z}_1^N) + \int_{|\bar{\mathbf{z}}| \geq \varepsilon} |\bar{\mathbf{z}}| \mu^{\otimes N}(d\mathbf{z}_1^N)$$

il primo termine è minore di ε , il secondo lo stimo con Chebychev: poiché $|\bar{\mathbf{z}}| \geq \varepsilon$

$$\int_{|\bar{\mathbf{z}}| \geq \varepsilon} |\bar{\mathbf{z}}| \mu^{\otimes N}(d\mathbf{z}_1^N) \leq \frac{1}{\varepsilon} \int |\bar{\mathbf{z}}|^2 \mu^{\otimes N}(d\mathbf{z}_1^N) = \frac{1}{\varepsilon} \frac{1}{N} \sigma^2$$

Scegliendo $\varepsilon = N^{-1/2}$ si ottiene la tesi.

Si può anche dimostrare, ma non lo farò, che vale la legge forte dei grandi numeri, cioè che $\bar{\mathbf{z}}$ converge a 0 quasi certamente (cioè a meno di eventi di probabilità nulla).

Supponiamo di prendere N particelle i.i.d. con densità $f(\mathbf{x}, \mathbf{v})$. La variabile aleatoria $\mathbf{F}(\mathbf{x} - \mathbf{x}_i)$ ha come media $\int \mathbf{F}(\mathbf{x} - \mathbf{y}) f(d\mathbf{y}, d\mathbf{w})$. La legge forte dei grandi numeri assicura che

$$\frac{1}{N} \sum_{i=1}^N \mathbf{F}(\mathbf{x} - \mathbf{x}_i) \rightarrow \int \mathbf{F}(\mathbf{x} - \mathbf{y}) f(\mathbf{y}, \mathbf{w}) d\mathbf{y} d\mathbf{w}.$$

2.1 Proprietà dell'equazione di Vlasov

Definisco

$$\mathcal{F}[f](\mathbf{x}) = \int F(\mathbf{x} - \mathbf{y})f(\mathbf{y}, \mathbf{w}) \, d\mathbf{y} \, d\mathbf{w}$$

che è il campo di forza generato da tutte le particelle, calcolato nel punto \mathbf{x} . In termini di questo campo, l'equazione di Vlasov è l'equazione di Liouville per il flusso a una particella

$$\begin{cases} \dot{\mathbf{X}}_t = \mathbf{V}_t \\ \dot{\mathbf{V}}_t = \mathcal{F}[f_t](\mathbf{X}_t) \end{cases} \quad (2.2)$$

dove ho messo in evidenza la dipendenza da t di f , per sottolineare che non si tratta di un sistema autonomo. Il sistema rimane dunque hamiltoniano, di hamiltoniana

$$\frac{1}{2}|\mathbf{v}|^2 + \int U(\mathbf{x} - \mathbf{y})f(t, \mathbf{y}, \mathbf{w}) \, d\mathbf{y} \, d\mathbf{w}$$

che dipende dal tempo. Si noti che l'hamiltoniana è funzione di f che è funzione del flusso generato da f .

Lemma 2.2. *La quantità*

$$\mathcal{H} = \frac{1}{2} \int |\mathbf{v}|^2 f(t, \mathbf{x}, \mathbf{v}) \, d\mathbf{x} \, d\mathbf{v} + \frac{1}{2} \int U(\mathbf{x} - \mathbf{y})f(t, \mathbf{x}, \mathbf{v})f(t, \mathbf{y}, \mathbf{w}) \, d\mathbf{x} \, d\mathbf{v} \, d\mathbf{y} \, d\mathbf{w},$$

è conservata dall'equazione.

Si noti che la variazione prima in f di \mathcal{H} è

$$\frac{\delta \mathcal{H}}{\delta f} = \frac{1}{2}|\mathbf{v}|^2 + \int U(\mathbf{x} - \mathbf{y})f(t, \mathbf{y}, \mathbf{w}) \, d\mathbf{y} \, d\mathbf{w},$$

dunque

$$\frac{d}{dt} \mathcal{H} = \int \frac{\delta \mathcal{H}}{\delta f} \partial_t f$$

Sostituendo a $\partial_t f$ la sua espressione, si ottiene la tesi integrando per parti e notando che $\partial_v \frac{\delta \mathcal{H}}{\delta f} = \mathbf{v}$ e $\partial_x \frac{\delta \mathcal{H}}{\delta f} = \mathcal{F}[f](\mathbf{x})$.

Si noti, infine, che, almeno formalmente, per $N \rightarrow +\infty$ l'energia del sistema di particelle tende a \mathcal{H} .

2.2 Altri modelli

Prima di discutere le proprietà matematiche dell'equazione di Vlasov e come si ottenga dal sistema di particelle, introduco qualche altro esempio, in ambito non fisico. Da vari decenni vengono studiati modelli per descrivere il comportamento di folle, il moto di sciame o greggi, o anche come cambiano le opinioni in un collettivo umano (“dinamica delle opinioni”), o come si comportano gli agenti sui mercati finanziari. In questi modelli le variabili riguardano gli “agenti” (posizione e velocità di uccello, opinione o disponibilità economica di un individuo), e queste variabili cambiano per interazione tra gli agenti. Spesso gli agenti sono tutti uguali, e si differenziano solo per il valore delle variabili. Gli aspetti più affascinanti di questi modelli sono le cosiddette **proprietà emergenti**, cioè i comportamenti coordinati a livello collettivo,

che non vengono imposte nel modello ma sono conseguenza delle dinamiche microscopiche (in questo senso “emergono” dalla dinamica degli agenti). Fenomeni di questo tipo sono l’allineamento degli uccelli in volo: non esiste una “regia” esterna che dice ai singoli uccelli cosa fare, ma il fatto che si allineano deve essere conseguenza delle regole dell’interazione tra essi. Un fenomeno dello stesso tipo è l’emerge del consenso su una particolare opinione (si pensi all’esplosione che sembra improvvisa di una moda nel vestire, o nel parlare, o nel gusto).

Illustro alcuni di questi modelli.

2.2.1 Il modello per il flocking di Cucker e Smale

In questo caso gli agenti sono uccelli (o più in generale animali), le variabili sono posizione e velocità $(\mathbf{x}_i, \mathbf{v}_i)$ e l’interazione è fatta in modo che ogni uccello modifica la sua velocità per uniformarsi alla velocità degli uccelli vicini.

$$\begin{aligned}\dot{\mathbf{x}}_i &= \mathbf{v}_i \\ \dot{\mathbf{v}}_i &= \lambda \sum_j p_{ij}(\mathbf{v}_j - \mathbf{v}_i)\end{aligned}\tag{2.3}$$

Il contributo dell’agente j all’accelerazione dell’agente i ha intensità $p_{ij}|\mathbf{v}_j - \mathbf{v}_i|$ e fa decrescere la differenza di velocità (si pensi alla soluzione dell’equazione $\dot{\mathbf{v}} = p(\mathbf{w} - \mathbf{v})$, con \mathbf{w} fissato: si ottiene una convergenza esponenziale di $\mathbf{v}(t)$ al valore \mathbf{w}). I coefficienti (positivi o nulli) p_{ij} formano la **matrice di comunicazione**.

Nel modello di Cucker e Smale questi coefficienti sono del tipo

$$p_{ij} = \frac{1}{1 + |\mathbf{x}_i - \mathbf{x}_j|^\alpha}$$

per qualche $\alpha > 0$. In questo modo, la velocità dell’agente i tende ad allinearsi a quella degli uccelli vicini, mentre è poco influenzata dagli uccelli lontani.

Si noti che se p_{ij} è simmetrico, allora la velocità media si conserva (lo si provi per esercizio). Si può dimostrare che se α è abbastanza piccolo, asintoticamente le velocità diventano uguali alla velocità media, e gli uccelli si muovono su traiettorie parallele, mantenendo distanze fisse tra loro.

In modelli di opinioni, si possono considerare invece sistemi di questo tipo nelle sole variabili \mathbf{v} , con opportune matrici di comunicazione. In questo caso l’allineamento delle \mathbf{v}_i corrisponde all’emergere del consenso.

2.2.2 Il modello di Kuramoto per la sincronizzazione

Il fenomeno della sincronizzazione spontanea è molto affascinante e ne esistono esempi naturali e artificiali: potete cercare in rete informazioni e video su lucciole e metronomi che si sincronizzano.

Ogni agente è descritto da una variabile periodica (il tempo per l’alternarsi acceso / spento di una lucciola, o la posizione sull’orbita chiusa dell’asta del metronomo). Assumendo che tutti gli agenti siano uguali, possiamo usare delle **fasi** cioè delle variabili angolari ϑ_i di periodicità 2π . Nel caso più semplice, il modello di Kuramoto è

$$\dot{\vartheta}_i = \omega - \frac{1}{N} \sum_j \sin(\vartheta_i - \vartheta_j)$$

Quando la differenza tra $\vartheta_i - \vartheta_j$ il contributo dell'agente j alla velocità della fase dell'agente i è nella direzione di far diminuire la differenza.

Si può provare che a meno di dati iniziali molto particolari, asintoticamente $\dot{\vartheta}_i \rightarrow \omega$, e $|\vartheta_i - \vartheta_j| \rightarrow 0$, realizzando così la sincronizzazione.

Si noti che le variabili $\gamma_i = \vartheta_i - t\omega$ verificano

$$\dot{\gamma}_i = -\frac{1}{N} \sum_j \sin(\gamma_i - \gamma_j)$$

Questo è un **sistema gradiente**, infatti

$$\begin{aligned} \dot{\gamma}_i &= -\partial_{\gamma_i} U(\boldsymbol{\gamma}) \\ U(\boldsymbol{\gamma}) &= -\frac{1}{2} \sum_{i,j} \cos(\gamma_i - \gamma_j) \end{aligned}$$

(lo si provi per esercizio). Dunque U decresce lungo il moto (ed è una funzione di Liapunov per il sistema). Si noti però che la funzione U ha infiniti minimi globali, che si realizzano per valori coincidenti delle γ_i , e anche infiniti punti stazionari che non sono minimi (quelli in cui γ_i coincidono o differiscono per π , cioè sono opposti).

2.2.3 Un algoritmo per la determinazione dei minimi

In moltissime situazioni applicative è necessario determinare il minimo globale di una funzione $U(\mathbf{x})$. In casi importanti, U può non essere regolare, \mathbf{x} può vivere in uno spazio di dimensione enorme, e soprattutto U può avere un panorama complesso di punti di minimo locale (può rientrare in questo caso l'addestramento dei parametri di una rete neurale).

Esistono molti algoritmi per risolvere questo problema, per esempio si può considerare la dinamica

$$\dot{\mathbf{x}} = -\partial_x U$$

che garantisce che U decresce. Ma a seconda del dato iniziale, il sistema finisce in un minimo locale. In genere viene aggiunta una stocasticità che permette alla traiettoria di sfuggire dai minimi troppo alti ("simulated annealing"). Un altro possibile modo per curare questo problema è avere molti agenti che esplorano lo spazio dei minimi, e che tendono a richiamarsi tra loro:

$$\dot{\mathbf{x}}_i = -\partial_x U(\mathbf{x}_i) - \sum_j p_{ij}(\mathbf{x}_i - \mathbf{x}_j).$$

In questo modo è possibile che l'agente che sta nel minimo più profondo, da cui dunque esce con maggiore difficoltà, riesca a richiamare gli altri agenti, permettendo una rapida individuazione del minimo.

Si può assumere che p_{ij} decresca con la distanza, e se è di ordine $1/N$, si può studiare questo sistema nel limite di campo medio ottenendo indicazioni sul suo funzionamento.

2.2.4 Una descrizione unificata

Tutti i modelli che ho considerato, in cui l'interazione è di ordine $1/N$ possono essere scritti nella forma

$$\dot{\mathbf{z}}_i = \mathbf{H}(\mathbf{z}_i) + \frac{1}{N} \sum_j \mathbf{G}(\mathbf{z}_i, \mathbf{z}_j)$$

(per esercizio si scrivano \mathbf{z} , \mathbf{H} e \mathbf{G} nel caso del sistema di particelle, per Cucker-Smale, per Kuramoto, per il sistema per la ricerca dei minimi).

La corrispondente equazione di Liouville è

$$\partial_t f^N(t, \mathbf{z}_1^N) + \sum_{i=1}^N \operatorname{div}_{\mathbf{z}_i} (\mathbf{H}(\mathbf{z}_i) f^N(t, \mathbf{z}_1^N)) + \sum_{i=1}^N \operatorname{div}_{\mathbf{z}_i} \left(\frac{1}{N} \sum_{h=1}^N \mathbf{G}(\mathbf{z}_i, \mathbf{z}_h) f^N(t, \mathbf{z}_1^N) \right) = 0$$

La gerarchia è

$$\partial_t f_j^N(t, \mathbf{z}_1^j) + \sum_{i=1}^j \operatorname{div}_{\mathbf{z}_i} (\mathbf{H}(\mathbf{z}_i) f_j^N(t, \mathbf{z}_1^j)) + \sum_{i=1}^j \operatorname{div}_{\mathbf{z}_i} \left(\frac{N-j}{N} \int d\mathbf{z}_{j+1} \mathbf{G}(\mathbf{z}_i, \mathbf{z}_{j+1}) f_{j+1}^N(t, \mathbf{z}_1^{j+1}) \right) = 0$$

Nel limite $N \rightarrow +\infty$, se le marginali convergono, si ottiene la gerarchia infinita:

$$\partial_t f_j(t, \mathbf{z}_1^j) + \sum_{i=1}^j \operatorname{div}_{\mathbf{z}_i} (\mathbf{H}(\mathbf{z}_i) f_j(t, \mathbf{z}_1^j)) + \sum_{i=1}^j \operatorname{div}_{\mathbf{z}_i} \left(\int d\mathbf{z}_{j+1} \mathbf{G}(\mathbf{z}_i, \mathbf{z}_{j+1}) f_{j+1}(t, \mathbf{z}_1^{j+1}) \right) = 0$$

Ipotizzando che

$$f_j^N \rightarrow f^{\otimes j}$$

tutte le equazioni della gerarchia sono soddisfatte se è soddisfatta la prima, che diventa

$$\begin{aligned} \partial_t f_t + \operatorname{div}_z (\mathbf{H}f) + \operatorname{div}_z (f_t \mathcal{K}[f_t]) &= 0 \\ \mathcal{K}[f](\mathbf{z}) := \int \mathbf{G}(\mathbf{z}, \tilde{\mathbf{z}}) f(\tilde{\mathbf{z}}) &= 0 \end{aligned} \tag{2.4}$$

che è appunto l'equazione di campo medio associata al sistema di partenza.

Per discutere dell'esistenza e unicità delle soluzioni di questa equazione e se sia affettivamente derivabile dal modello microscopico ci serve esplorare la nozione di convergenza debole delle misure.

2.3 La distanza di Wasserstein

Definizione

Date due misure di probabilità μ e ν , la loro distanza di Wasserstein-1 è

$$W_1(\mu, \nu) = \inf_{P \in \Gamma(\mu, \nu)} \int |\mathbf{x} - \mathbf{y}| P(d\mathbf{x}, d\mathbf{y})$$

dove $\Gamma(\mu, \nu)$ è l'insieme delle misure di probabilità congiunte di μ e ν , cioè delle misure di probabilità P sullo spazio prodotto che hanno μ e ν come marginali. Si noti che se μ e/o ν hanno supporto illimitato, la loro distanza può essere infinita, in base al fatto che il primo momento di una delle due, per esempio $\int |\mathbf{x}| \mu(d\mathbf{x})$, sia infinito. Questa difficoltà nella definizione non c'è se si considerano misure con supporto in un compatto.

Il seguente teorema mostra l'utilità di questa nozione.

Teorema 2.2. *Sia K un compatto di \mathbb{R}^n . Sia $\mathcal{P}(K)$ l'insieme delle misure di probabilità su K . Allora W_1 è una metrica su $\mathcal{P}(K)$, e $\mathcal{P}(K)$ è un compatto con tale metrica.*

Inoltre W_1 metrizza la convergenza debole nel senso delle misure, cioè

$$\mu_n \rightharpoonup \mu \iff W_1(\mu_n, \mu) \rightarrow 0.$$

Ricordo che $\mu_n \rightarrow \mu$ per definizione se e solo se per ogni $\phi(\mathbf{x})$ continua si ha

$$\int \phi(\mathbf{x})\mu_n(d\mathbf{x}) \rightarrow \int \phi(\mathbf{x})\mu(d\mathbf{x}).$$

Prima di dare una (parziale) dimostrazione di questo teorema, introduco la seguente caratterizzazione **duale**.

Teorema 2.3 (Dualità di Kantorovich).

$$W_1(\mu, \nu) = \sup_{\varphi, \psi \in C(K): \varphi(x) + \psi(y) \leq |\mathbf{x} - \mathbf{y}|} \int \varphi d\mu + \int \psi d\nu.$$

Presento qui una dimostrazione comprensibile, che si basa sulla teoria della trasformata di Legendre (ma alcuni dettagli sono in appendice).

Userò le seguenti notazioni “compatte”: date due misure di probabilità μ e ν , e date due funzioni continue φ e ψ :

$$\begin{aligned} d(\mu \otimes \nu) &= \mu \otimes \nu(d\mathbf{x}, d\mathbf{y}) = \mu(d\mathbf{x})\nu(d\mathbf{y}) \\ \varphi \oplus \psi &= (\varphi \oplus \psi)(\mathbf{x}, \mathbf{y}) = \varphi(x) + \psi(\mathbf{y}) \end{aligned}$$

In particolare

$$\int \varphi \oplus \psi d(\mu \otimes \nu) = \int \varphi d\mu + \int \psi d\nu.$$

Si noti che $d(\mu \otimes \nu) \in \Gamma(\mu, \nu)$, infatti è la misura di probabilità congiunta che si ottiene pensando di estrarre x con μ , e y con ν , indipendenti. Inoltre indicherò con $c = c(\mathbf{x}, \mathbf{y})$ il costo $|\mathbf{x} - \mathbf{y}|$.

L’idea per provare la dualità è quella di usare due funzioni φ e ψ per esprimere il vincolo sulle distribuzioni congiunte. Sia $\mathcal{M}(K \times K)$ l’insieme delle misure con segno, cioè il duale di $C(K \times K)$, sia \mathcal{M}^+ il sottoinsieme delle misure positive. Evidentemente

$$\Gamma(\mu, \nu) \subset \mathcal{M}^+ \subset \mathcal{M}.$$

Consideriamo la differenza

$$\int \varphi \oplus \psi d(\mu \otimes \nu) - \int \varphi \oplus \psi d\gamma$$

per una generica $\gamma \in \mathcal{M}$. Si noti che questa differenza vale

$$\begin{aligned} &\int \varphi(x)\mu(d\mathbf{x}) + \int \psi(\mathbf{y})\nu(d\mathbf{y}) - \int (\varphi(\mathbf{x}) + \psi(\mathbf{y}))\gamma(dx, d\mathbf{y}) = \\ &\int \varphi(\mathbf{x})(\mu(d\mathbf{x}) - \gamma^1(d\mathbf{x})) + \int \psi(\mathbf{y})(\nu(d\mathbf{y}) - \gamma^2(d\mathbf{y})) \end{aligned}$$

dove γ^1 e γ^2 sono le due marginali di γ . Se $\gamma \in \Gamma(\mu, \nu)$, la differenza è 0 qualunque siano φ, ψ . Al contrario, se $\gamma \notin \Gamma(\mu, \nu)$ $\mu \neq \gamma^1$ o $\nu \neq \gamma^2$ (o entrambi), e dunque esistono $\bar{\varphi}, \bar{\psi}$ su cui la differenza da un valore non nullo. Pertanto

$$\sup_{\varphi, \psi \in C(K)} \left(\int \varphi \oplus \psi d(\mu \otimes \nu) - \int \varphi \oplus \psi d\gamma \right) = \begin{cases} 0 & \text{se } \gamma \in \Gamma(\mu, \nu) \\ +\infty & \text{altrimenti} \end{cases}$$

infatti nel primo caso la differenza è nulla per ogni coppia di funzioni, nel secondo caso moltiplicando la coppia $(\bar{\varphi}, \bar{\psi})$ per una costante arbitrarie, opportuno, si ottiene un qualunque valore reale, pertanto il sup è infinito.

Nella definizione di W_1 posso considerare invece dell'inf su $P \in \Gamma$, l'inf su $\gamma \in \mathcal{M}^+$, pagando però $+\infty$ se $\gamma \notin \Gamma$:

$$W_1(\mu, \nu) = \inf_{P \in \Gamma(\mu, \nu)} \int c dP = \inf_{\gamma \in \mathcal{M}^+} \sup_{\varphi, \psi} \int c d\gamma + \int \varphi \oplus \psi d(\mu \otimes \nu) - \int \varphi \oplus \psi d\gamma. \quad (2.5)$$

(ricordo che il costo c è $c = |\mathbf{x} - \mathbf{y}|$. Per una qualunque funzione F di due variabili si prova facilmente che in generale

$$\inf_{a \in A} \sup_{b \in B} F(a, b) \geq \sup_{b \in B} \inf_{a \in A} F(a, b).$$

In qualche caso vale l'uguaglianza (si pensi a una funzione F concava in a e convessa in b , in tal caso la coppia (a, b) che realizza gli estremi è la stessa indipendentemente dall'ordine, e individua un punto di sella di F). Dunque

$$\inf_{P \in \Gamma(\mu, \nu)} \int c dP \geq \sup_{\varphi, \psi} \inf_{\gamma \in \mathcal{M}^+} \int c d\gamma + \int \varphi \oplus \psi d(\mu \otimes \nu) - \int \varphi \oplus \psi d\gamma$$

Noto che l'inf è solo sui due termini in γ :

$$\inf_{\gamma \in \mathcal{M}^+} \int (c - \varphi \oplus \psi) d\gamma$$

È evidente che se $\varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq c(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$ l'inf è raggiunto su $\gamma = 0$, e vale 0. Al contrario, se esistono valori $(\mathbf{x}_0, \mathbf{y}_0)$ tali che $\varphi(\mathbf{x}_0) + \psi(\mathbf{y}_0) > |\mathbf{x}_0 - \mathbf{y}_0|$ basta concentrare γ su questi valori, con valori divergenti di massa per ottenere che l'inf è $-\infty$. Dunque l'inf su \mathcal{M}^+ è in realtà un vincolo sulla coppia di funzioni φ, ψ . Abbiamo dunque provato che

$$W_1(\mu, \nu) \geq \sup_{\varphi, \psi \in C(K), \varphi \oplus \psi \leq c} \int \varphi \oplus \psi d(\mu \otimes \nu).$$

Rimane da provare l'uguaglianza. Ci sono varie strade, una è mostrare nei passaggi precedenti che per le funzioni che stiamo considerando si può scambiare inf sup con sup inf. Userò invece la trasformata di Legendre del seguente funzionale definito su $C(K \times K)$: data $p \in C(K \times K)$

$$H(p) = - \sup_{\varphi, \psi \in C(K), \varphi \oplus \psi \leq c-p} \int \varphi \oplus \psi d(\mu \otimes \nu).$$

Data $\gamma \in \mathcal{M}^+$, la trasformata di Legendre di H calcolata in γ è

$$H^*(\gamma) := \sup_{p \in C(K \times K)} \int p d\gamma - H(p).$$

(Si noti che H è definito sullo spazio di Banach $C(K \times K)$. H^* è definita sul suo duale. La prova consiste nei seguenti passi.

Passo 1

H è un funzionale convesso e semicontinuo inferiore.

Si noti intanto che H è finito per ogni p . Date due funzioni continue p_1 e p_2 siano φ_i, ψ_i , $i = 1, 2$, tali che

$$\varphi_i \oplus \psi_i \leq c - p_i.$$

Qualunque siano $\lambda_i \in (0, 1)$, $i = 1, 2$, con $\lambda_1 + \lambda_2 = 1$, si ha

$$\sum_{i=1,2} \lambda_i \varphi_i \oplus \psi_i \leq c - (\lambda_1 p_1 + \lambda_2 p_2),$$

inoltre

$$\sum_{i=1,2} \lambda_i (\varphi_i \oplus \psi_i) = \left(\sum_{i=1,2} \lambda_i \varphi_i \right) \oplus \left(\sum_{i=1,2} \lambda_i \psi_i \right)$$

(questa scrittura può risultare sospetta, ma si ricordi che \oplus è una somma, non una moltiplicazione). Pertanto l'estremo superiore sulle coppie φ, ψ che soddisfano il vincolo $\varphi \oplus \psi \leq c - (\lambda_1 p_1 + \lambda_2 p_2)$ è maggiore o uguale all'estremo superiore fatto sulle coppie della forma $\sum_{i=1,2} \lambda_i (\varphi_i \oplus \psi_i)$ con i vincoli assegnati per $i = 1, 2$. Pertanto

$$\begin{aligned} H(\lambda_1 p_1 + \lambda_2 p_2) &\leq - \sup_{\varphi_i, \psi_i \varphi_i \oplus \psi_i \leq c - p_i} \int \left(\sum_{j=1,2} \lambda_j (\varphi_j \oplus \psi_j) \right) d(\mu \otimes \nu) \\ &= \sum_{j=1,2} -\lambda_j \sup_{\varphi_i, \psi_i \varphi_i \oplus \psi_i \leq c - p_i} \int \varphi_j \oplus \psi_j d(\mu \otimes \nu) = \lambda_1 H(p_1) + \lambda_2 H(p_2). \end{aligned}$$

Verifichiamo la semicontinuità inferiore. Sia $p_n \rightarrow p$, nella norma dell'estremo superiore. Dato $\varepsilon > 0$, definitivamente in n si ha $|p - p_n| < \varepsilon$, in particolare $p + \varepsilon > p_n$. Definisco $\varphi \oplus \psi$ "ammissibile" per p se vale il vincolo $\varphi \oplus \psi \leq c - p$, e scrivo $\varphi \oplus \psi \in \mathcal{A}(p)$.

Si ha che $\varphi \oplus \psi \in \mathcal{A}(p_n)$ se e solo se $(\varphi - \varepsilon) \oplus \psi \in \mathcal{A}(p_n + \varepsilon)$. Poiché $p_n + \varepsilon \geq p$, ne segue che $(\varphi - \varepsilon) \oplus \psi \in \mathcal{A}(p)$. Dunque

$$-H(p) = \sup_{\varphi \oplus \psi \in \mathcal{A}(p)} \int \varphi \oplus \psi d(\mu \otimes \nu) \geq \sup_{\varphi \oplus \psi \in \mathcal{A}(p_n)} \int (\varphi - \varepsilon) \oplus \psi d(\mu \otimes \nu) = -H(p_n) - \varepsilon.$$

Abbiamo dunque provato che per ogni ε , definitivamente $H(p) \leq H(p_n) + \varepsilon$ cioè che $H(p) \leq \liminf_n H(p_n)$.

Passo 2

Calcolo di H^* .

$$H^*(\gamma) = \begin{cases} \sup_{\varphi, \psi} \int (c - \varphi \oplus \psi) d\gamma + \int \varphi \oplus \psi d(\mu \otimes \nu) & \text{se } \gamma \in M^+ \\ +\infty & \text{altrimenti} \end{cases}$$

Passo 3

Relazione tra H^* e W_1 .

La formula (2.5) dice esattamente che

$$W_1(\mu, \nu) = \inf_{\gamma \in M^+} H^*(\gamma).$$

Passo 4

Relazione tra W_1 e H^{**} .

$$H^{**}(p) = \sup_{\gamma \in M} \int p \, d\gamma - H^*(\gamma),$$

dunque

$$H^{**}(0) = \sup_{\gamma \in M} -H^*(\gamma) = - \inf_{\gamma \in M} H^*(\gamma) = - \inf_{\gamma \in M^+} H^*(\gamma),$$

infatti l'estremo inferiore è certamente raggiunto in M^+ , che contiene il dominio di H^* . Dunque $W_1(\mu, \nu) = -H^{**}(0)$.

Passo 5

Conclusione.

La funzione H è convessa e semicontinua inferiore (passo 1), dunque $H^{**} = H$ (vedi appendice), pertanto si ottiene la tesi

$$W_1(\mu, \nu) = -H(0) = \sup_{\varphi, \psi \in C(K), \varphi \otimes \psi \leq c} \int \varphi \oplus \psi \, d(\mu \otimes \nu).$$

La dualità di Kantorovich vale in generale per funzioni costo $c(\mathbf{x}, \mathbf{y})$, di cui $|\mathbf{x} - \mathbf{y}|$ è un caso particolare. Invece il prossimo teorema è una proprietà che vale solo per la W_1 .

Teorema 2.4 (Distanza 1-lipschitziana).

$$W_1(\mu, \nu) = \sup_{\varphi, \psi \in C(K), \varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|} \int \varphi \, d\mu + \int \psi \, d\nu = \sup_{\varphi \in \text{Lip}_1} \int \varphi \, (d\mu - d\nu),$$

dove Lip_1 è l'insieme delle funzioni lipschitziane di costante di Lipschitz minore o uguale a 1.

La prima affermazione è il contenuto del teorema precedente. Dimostro la seconda. La prima osservazione è che se φ è 1-lipschitziana (cioè è in Lip_1) la coppia di funzioni $\varphi(\mathbf{x})$, $\psi(\mathbf{y}) = -\varphi(\mathbf{y})$ verifica il vincolo richiesto, dunque

$$\sup_{\varphi, \psi: \varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|} \int \varphi \, d\mu + \int \psi \, d\nu \geq \sup_{\varphi \in \text{Lip}_1} \int \varphi \, (d\mu - d\nu).$$

Dimostro la disuguaglianza opposta. Date φ e ψ con $\varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|$, osservo che

$$\varphi(\mathbf{x}) \leq \psi^*(\mathbf{x}) := \inf_{\mathbf{y}} |\mathbf{x} - \mathbf{y}| - \psi(\mathbf{y}),$$

infatti per ogni \mathbf{y} , $\varphi(\mathbf{x}) \leq |\mathbf{x} - \mathbf{y}| - \psi(\mathbf{y})$, e dunque $\varphi(\mathbf{x})$ è minore dell'estremo inferiore in \mathbf{y} . Si ha dunque

$$\int \varphi \, d\mu \leq \int \psi^* \, d\mu$$

D'altra parte, poiché

$$\psi^*(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|$$

si ha che

$$\psi(\mathbf{y}) \leq \psi^{**}(\mathbf{y}) := \inf_{\mathbf{x}} |\mathbf{x} - \mathbf{y}| - \psi^*(\mathbf{x}).$$

Pertanto

$$\sup_{\varphi, \psi: \varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|} \int \varphi d\mu + \int \psi d\nu = \sup_{\psi} \left(\int \psi^* d\mu + \int \psi^{**} d\nu \right)$$

Si prova facilmente che ψ^* è una funzione 1-lipschitziana, dunque

$$\sup_{\varphi, \psi: \varphi(\mathbf{x}) + \psi(\mathbf{y}) \leq |\mathbf{x} - \mathbf{y}|} \int \varphi d\mu + \int \psi d\nu \leq \sup_{\varphi \in \text{Lip}_1} \int \varphi d\mu + \int \varphi^* d\nu.$$

Infine, è facile provare che essendo φ una funzione 1-lipschitziana,

$$\varphi^*(\mathbf{y}) = -\varphi(\mathbf{y}),$$

e questo fatto conclude la prova.

Prova del teorema 2.2. Considero dimostrato dai corsi di analisi che l'insieme delle misure di probabilità su un compatto è compatto nella topologia debole.

Provo che W_1 è una metrica La simmetria è ovvia. Inoltre, se $W_1(\mu, \nu) = 0$, allora per qualunque funzione lipschitziana

$$\int \phi d\mu = \int \phi d\nu,$$

procedendo per densità si ottiene che la stessa affermazione vale per ϕ continua. Questo basta a dichiarare uguali le due misure. Viceversa, scegliendo $P(d\mathbf{x}, d\mathbf{y}) = \delta(\mathbf{x} - \mathbf{y})\mu(d\mathbf{x}) d\mathbf{y}$, si ottiene $W_1(\mu, \mu) = 0$.

La disuguaglianza triangolare è di facile dimostrazione usando l'equivalenza con la distanza 1-lipschitziana:

$$W_1(\mu, \nu) = \sup_{\varphi \in \text{Lip}_1} \int \varphi(d\mu - d\nu) = \sup_{\varphi \in \text{Lip}_1} \int \varphi(d\mu - d\tilde{\mu}) + \int \varphi(d\tilde{\mu} - d\nu) \leq W_1(\mu, \tilde{\mu}) + W_1(\tilde{\mu}, \nu).$$

Una traccia di una prova che usa direttamente la definizione è riportata in appendice.

Rimane infine da provare che la convergenza in W_1 equivale alla convergenza debole. Osservo che per ogni funzione φ

$$\int \varphi(d\mu_n - d\mu) = \int (\varphi(\mathbf{x}) - \varphi(\mathbf{y}))P(d\mathbf{x}, d\mathbf{y})$$

per ogni $P \in \Gamma(\mu, \nu)$. Se φ funzione test lipschitziana, passando ai moduli e all'inf su P si ottiene

$$\left| \int \varphi(d\mu_n - d\mu) \right| \leq LW_1(\mu, \nu)$$

dove L è la costante di Lipschitz di φ . Con il consueto argomento di densità per le osservabili si prova che se la distanza W_1 tende a 0, allora si ha la convergenza debole.

Viceversa, sia data μ_n tale che per ogni φ continua

$$\int \varphi d\mu_n \rightarrow \int \varphi d\mu.$$

In particolare questo è vero le funzioni 1-lipschitziane. Sia ora φ_n una funzione 1-lipschitziana tale che

$$W_1(\mu_n, \mu) \leq \int \varphi_n d\mu_n - \int \varphi_n d\mu + \frac{1}{n}$$

(che esiste per la definizione di sup). Noto che nella definizione di distanza 1-lipschitziana, posso aggiungere a φ qualunque costante, senza modificare nulla. Questo vuol dire che posso considerare il sup sulle funzioni a media nulla. È facile provare che le funzioni a media nulla e 1-lipschitziane sono limitate dal diametro del dominio. Posso utilizzare il teorema di Ascoli-Arzelà per concludere che a meno di sottosequenze φ_n converge uniformemente a $\varphi \in \text{Lip}_1$. Indico con n la sottosequenza.

$$\int \varphi_n d\mu_n - \int \varphi_n d\mu = \int (\varphi_n - \varphi) d\mu_n - \int (\varphi_n - \varphi) d\mu + \int \varphi (d\mu_n - d\mu)$$

I primi integrali sono stimati da $\|\varphi_n - \varphi\|_\infty$ che tende a 0. L'ultimo integrale va a 0 perché $\mu_n \rightarrow \mu$. In conclusione si ottiene che $W_1(\mu_n, \mu) \rightarrow 0$.

□

2.4 Il problema di Monge e il problema di Kantorovich

La distanza di Wasserstein W_1 è anche detta distanza di Kantorovich-Rubinstein, e, insieme alle altre W_p è legata ai problemi di **trasporto di massa**. Li illustro facendo esempi.

Considerando il caso di due distribuzioni singolari:

$$\begin{cases} \mu(d\mathbf{x}) = \frac{1}{n} \sum_1^n \delta(\mathbf{x} - \mathbf{x}_i) d\mathbf{x} \\ \nu(d\mathbf{x}) = \frac{1}{n} \sum_1^n \delta(\mathbf{x} - \mathbf{y}_i) d\mathbf{x} \end{cases}$$

dove i punti \mathbf{x}_i sono distinti tra loro, e i punti \mathbf{y}_i sono distinti tra loro. Qualunque distribuzione congiunta P di μ e di ν , ha necessariamente supporto sugli n^2 punti $\{(\mathbf{x}_i, \mathbf{y}_j)\}_{i,j}$ (lo si provi per esercizio). Dunque

$$P(d\mathbf{x}, d\mathbf{y}) = \frac{1}{n} \sum_{i,j} a_{ij} \delta(\mathbf{x} - \mathbf{x}_i) \delta(\mathbf{y} - \mathbf{y}_i) d\mathbf{x} d\mathbf{y}$$

con $a_{ij} \geq 0$ e

$$\sum_i a_{ij} = 1 = \sum_j a_{ij}$$

condizione che si ottiene imponendo che P sia una distribuzione congiunta. In pratica, la delta in \mathbf{x}_1 , viene "frantumata" in n porzioni, di masse a_{1j} , e la massa a_{1j} viene associata

alla posizione \mathbf{y}_j . Si può pesare che la scelta degli a_{ij} sia un modo per **trasportare** μ in ν . Il corrispondente valore del funzionale è

$$\frac{1}{n} \sum_{ij} a_{ij} |\mathbf{x}_i - \mathbf{y}_j|$$

e viene interpretato come **costo** associato al trasporto descritto da P attraverso la matrice a . Quindi la distanza W_1 tra μ e ν è il minimo costo che bisogna pagare per trasportare μ in ν . Si dimostra abbastanza facilmente che il modo migliore di trasportare n masse puntiformi uguali in n masse puntiformi uguali si ottiene associando una e una sola massa iniziale a una e una sola massa finale, e quindi

$$\inf_{a_{ij}} \sum_{ij} a_{ij} |\mathbf{x}_i - \mathbf{y}_j| = \min_{\pi} \sum_i |\mathbf{x}_i - \mathbf{y}_{\pi_i}|$$

dove π è una permutazione degli indici.

Questo problema discreto si incontra in molti campi della matematica, per esempio in ottimizzazione combinatoria, e a volte ha il nome di problema del matching bipartito, perché si vuole trovare il migliore accoppiamento di due sequenze di n punti, rispetto alla distanza. È risolvibile numericamente con il cosiddetto *algoritmo ungherese*, con complessità polinomiale. Nel contesto del trasporto di massa, questo problema prende il nome di **problema di Monge**. Si può formulare anche al continuo, restringendo la ricerca della distribuzione congiunta P che garantisce il minimo costo, a quelle che si ottengono associando le misure attraverso un push-forward nello spazio (in pratica un cambio di variabili). Il problema di trovare l'estremo inferiore sulle generiche distribuzioni congiunte è invece il problema del trasporto di massa nella formulazione di Kantorovich, che è più generale, ma si riduce al problema di Monge tranne che nei casi di misure con differenti tipi di singolarità.

2.4.1 La distanza delle misure empiriche dalla legge

Il problema di Monge si trova all'intersezione di molte discipline, tra cui la meccanica statistica, l'ottimizzazione combinatoria, la teoria della probabilità, l'analisi.

Qui mi serve solo richiamare un risultato importante, di cui do una dimostrazione relativamente facile e con qualche buco. Stime precise sono molto più difficili e oggetto di lavori molto recenti.

Intanto do una definizione: dati $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ chiamo **misura empirica** associata la misura

$$\pi(\mathbf{z}_1^N, d\mathbf{z}) = \frac{1}{N} \sum_1^N \delta(\mathbf{z} - \mathbf{z}_i) d\mathbf{z}$$

Si può provare che ogni misura μ può essere ben approssimata da una misura empirica, cioè in modo che $W_1(\pi, \mu)$ tende a 0 per $N \rightarrow +\infty$.

Qui però voglio discutere di un aspetto probabilistico.

Teorema 2.5. *Sia μ una misura di probabilità a supporto compatto, per ogni N estraggo \mathbf{z}_i in modo indipendente, secondo la misura di probabilità μ . Allora*

$$\int W_1(\pi(\mathbf{z}_1^N), \mu) \mu^{\otimes N}(d\mathbf{z}_1^N) \rightarrow 0$$

cioè il valore atteso della distanza di Wasserstein tra la misura empirica e la legge con cui estraggo i punti va a 0.

In lavori recenti vengono anche suggeriti o dimostrati precisi andamenti asintotici in N , in generale per le distanze di ordine p . Qui mi accontento di provare che nel limite si ha 0. Per la prova, divido il supporto in $L = c\varepsilon^{-d}$ regioni di diametro di ordine $c\varepsilon$, dove d è la dimensione dello spazio. Chiamo $p_k = \int_{Q_k} \mu$, notando che la somma di questi numeri è 1. Se qualche p_k è nullo, escludo la regione (sarei fuori dal supporto di μ). Estraggo ora i punti \mathbf{z}_i , e indico con N_k il numero dei punti che cadono in Q_k . Suppongo che non siano 0 (poi darò indicazioni su come rimuovere questa condizione).

Costruisco una nuova misura $\tilde{\pi}$, sempre concentrata nei punti \mathbf{z}_i in questo modo: se $\mathbf{z}_i \in Q_h$, gli assegno massa p_h/N_h . In questo modo, μ e $\tilde{\pi}$ hanno la stessa massa sui Q_h , dunque posso scegliere una misura congiunta che ha supporto in $\sum_k Q_k \times Q_k$ (cioè accoppio le masse dentro in Q_k , posso perché sono uguali). Ottengo

$$W_1(\mu, \tilde{\pi}) \leq cL^{-d}.$$

Resta da stimare la distanza tra π e $\tilde{\pi}$. Uso la rappresentazione duale e considero per ogni funzione φ con $Lip(\varphi) \leq 1$:

$$\sum_h \left(\int_{Q_h} \pi \varphi - \int_{Q_h} \tilde{\pi} \varphi \right) = \sum_h \left(\frac{1}{N} - \frac{p_h}{N_h} \right) \sum_{\mathbf{z}_i \in Q_h} \varphi(i)$$

Osservo che se cambio φ di una costante questo valore non cambia dunque posso assumere che sia 0 in almeno un punto. Ne segue che $|\varphi| \leq D$, dove D è il diametro del supporto della misura μ . Quindi stimo tutto con

$$D \sum_h \left| \frac{1}{N} - \frac{p_h}{N_h} \right| N_j$$

e quindi, passando al sup su φ

$$W_1(\pi, \tilde{\pi}) \leq D \sum_h \left| \frac{N_h}{N} - p_h \right|$$

Mi aspetto che questo termine sia piccolo perché il valore atteso di N_h/N è proprio p_h . Per la precisione, N_h è una variabile binomiale, con probabilità di successo p_h , varianza $p_h(1-p_h)N$, dunque, usando Cauchy-Schwartz, il valore atteso di $|N_h/N - p_h|$ si stima con $\sqrt{p_h(1-p_h)/N}$. In definitiva

$$W_1(\mu, \pi) \leq \frac{1}{L^d} + c \frac{L}{\sqrt{N}}$$

Ora osservo che a L fissato, la probabilità che almeno un valore N_h si nullo è stimata dalla somma delle probabilità che $N_i = 0$, $i = 1, \dots, L$, più la probabilità che $N_{i_1} = N_{i_2} = 0$, per $i_1 < i_2$, etc. e queste somme sono pari a

$$\sum_{k=1}^{L-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq L} (1 - (p_{i_1} + \dots + p_{i_k}))^N$$

che è minore di $C(L)c^N$ con $c < 1$, dunque va a 0 in N . Nel calcolo del valore atteso escludo questa regione dello spazio di probabilità, che ha misura che va a 0 in N (l'integrando è stimato da D). Dunque, per ogni L

$$\limsup_{N \rightarrow +\infty} \int W_1(\pi, \mu) \mu^{\otimes N} \leq \varepsilon$$

Mandando $\varepsilon \rightarrow 0$ (cioè $L \rightarrow +\infty$) ottengo la tesi.

2.5 Costruzione delle soluzioni

A questo punto abbiamo gli strumenti per costruire le soluzioni dell'equazione di campo medio nella sua forma generale (2.4).

Osservo intanto che conoscendo $\mathcal{K}[f_t]$, l'equazione di campo medio è l'equazione di Liouville per il flusso (non autonomo), generato dal campo $\mathcal{K}[f_t]$. In particolare, detto $\mathbf{Z}_t^f(\mathbf{z})$ il flusso generato da f_t , ho che la soluzione è il push-forward di f_0 mediante il flusso:

$$f_t = \mathbf{Z}_t^f \# f_0.$$

Questa espressione non è la soluzione, perché non conosco il flusso se non conosco la soluzione stessa, però è utile perché permette di notare che per l'esistenza di f_t non ho bisogno di regolarità di f_0 ma solo del flusso. D'altra parte il campo \mathcal{K} è generato da un integrale, quindi per la sua regolarità ho solo bisogno della regolarità della funzione \mathbf{G} , e non di f_t .

Queste considerazioni si concretizzeranno in un teorema di costruzione delle soluzioni per dati iniziali misure di probabilità. Come ipotesi supporrò che f_0 sia a supporto limitato, e che la funzione \mathbf{G} sia localmente limitata e lipschitziana. L'idea è che, data g_t con $g_0 = f_0$, costruisco un flusso \mathbf{Z}_t^g , e con questo flusso costruisco $\mathbf{Z}_t^g \# f_0$, ottenendo una nuova traiettoria nello spazio delle distribuzioni di una particella. Mostrando che questa mappa è contrattiva otterrò l'esistenza e unicità delle soluzioni.

Lo faremo in parte attraverso due proposizioni; la prima che assicura che data f_t possiamo costruire un flusso con buone proprietà, la seconda che assicura che dato un flusso possiamo trasportare f_0 in f_t in modo da soddisfare le ipotesi che permettono di costruire il flusso. Infine, dovremo mettere insieme questi due aspetti.

Lemma 2.3 (Regolarità del campo). *Sia \mathbf{G} una funzione lipschitziana e limitata.*

Sia f_t una famiglia di misure di probabilità, con $t \in [0, T]$, con i supporti tutti contenuti in un chiuso limitato B , e sia inoltre debolmente continua in t , cioè $W_1(f_t, f_s) \rightarrow 0$ se $s \rightarrow t$.

Allora $\mathcal{K}[f_t](\mathbf{z})$ è un campo limitato e lipschitziano, inoltre è continuo in $t \in [0, T]$.

Dimostrazione per esercizio. Questo lemma implica che il flusso \mathbf{Z}_t^f è ben definito ovunque sui compatti.

Osservazione. Il caso fisicamente più interessante è quello dei sistemi conservativi con interazione di coppia, da cui siamo partiti, in tal caso \mathbf{G} è lipschitziana, ma non è limitata (una delle sue componenti è \mathbf{v}). Non è difficile superare questa difficoltà, perché come si vedrà serve solo la limitatezza sul supporto di f_t . Lascio i dettagli al lettore. Invece il caso reale è quello di interazione con potenziale gravitazionale o elettrostatico. In tal caso è tutto più difficile e interessante.

Proposizione 2.1 (Dipendenza del flusso dalla distribuzione). *Siano f_t e g_t due famiglie di misure di probabilità, con $t \in [0, T]$, con i supporti tutti contenuti in un chiuso limitato B .*

Siano inoltre debolmente continue in t , cioè $W_1(f_t, f_s), W_1(g_t, g_s) \rightarrow 0$ se $s \rightarrow t$.

Allora $\mathcal{K}[f_t]$ e $\mathcal{K}[g_t]$ sono funzioni continue in t , limitate e lipschitziane uniformemente in $t \in [0, T]$, pertanto i corrispondenti flussi sono ben definiti, e vale

$$|\mathbf{Z}_t^f(\mathbf{z}_1) - \mathbf{Z}_t^g(\mathbf{z}_2)| \leq e^{Lt} \left(|\mathbf{z}_1 - \mathbf{z}_2| + L \int_0^t ds W_1(f_s, g_s) \right). \quad (2.6)$$

per un'opportuna costante L .

Dimostro la (2.6). Scrivo i due flussi in forma integrale:

$$\begin{aligned}\mathbf{Z}_t^f(\mathbf{z}_1) &= \mathbf{z}_1 + \int_0^t ds \mathcal{K}[f_t](\mathbf{Z}_s^f(\mathbf{z}_1)) \\ \mathbf{Z}_t^g(\mathbf{z}_2) &= \mathbf{z}_2 + \int_0^t ds \mathcal{K}[g_t](\mathbf{Z}_s^g(\mathbf{z}_2))\end{aligned}$$

Sottraendo

$$|\mathbf{Z}_t^f(\mathbf{z}_1) - \mathbf{Z}_t^g(\mathbf{z}_2)| \leq |\mathbf{z}_1 - \mathbf{z}_2| + \int_0^t ds |\mathcal{K}[f_s](\mathbf{Z}_s^f(\mathbf{z}_1)) - \mathcal{K}[g_s](\mathbf{Z}_s^g(\mathbf{z}_2))|$$

Per stimare l'ultimo termine, sommo e sottraggo il campo generato da f_t ma calcolato nel flusso generato da g , ottenendo i due termini

$$\begin{aligned}(1) &= |\mathcal{K}[f_s](\mathbf{Z}_s^f(\mathbf{z}_1)) - \mathcal{K}[f_s](\mathbf{Z}_s^g(\mathbf{z}_2))| \\ (2) &= |\mathcal{K}[f_s](\mathbf{Z}_s^g(\mathbf{z}_2)) - \mathcal{K}[g_s](\mathbf{Z}_s^g(\mathbf{z}_2))|\end{aligned}$$

Il primo termine si stima semplicemente usando la lipschitzianità del campo \mathcal{K} , con la costante di Lipschitz L che è quella di \mathbf{G} :

$$|\mathcal{K}[f_s](\mathbf{Z}_s^f(\mathbf{z}_1)) - \mathcal{K}[f_s](\mathbf{Z}_s^g(\mathbf{z}_2))| \leq L |(\mathbf{Z}_s^f(\mathbf{z}_1)) - (\mathbf{Z}_s^g(\mathbf{z}_2))|$$

Nel secondo caso si tratta di stimare in $\mathbf{z} = \mathbf{Z}_s^g(\mathbf{z}_2)$ la differenza

$$|\mathcal{K}[f_s](\mathbf{z}) - \mathcal{K}[g_s](\mathbf{z})|.$$

Questa stima è meno immediata, e qui interviene la distanza di Wasserstein.

$$\begin{aligned}\mathcal{K}[f_s](\mathbf{z}) - \mathcal{K}[g_s](\mathbf{z}) &= \int \mathbf{G}(\mathbf{z} - \mathbf{z}_1) f_s(d\mathbf{z}_1) - \int \mathbf{G}(\mathbf{z} - \mathbf{z}_2) g_s(d\mathbf{z}_2) \\ &= \int (\mathbf{G}(\mathbf{z} - \mathbf{z}_1) - \mathbf{G}(\mathbf{z} - \mathbf{z}_2)) P(d\mathbf{z}_1, d\mathbf{z}_2)\end{aligned}$$

per una qualunque distribuzione congiunta di f_s e g_s (si verifichi questo fatto spezzando l'integrale e ricordando la definizione di distribuzione congiunta).

A questo punto siamo in grado fare la stima passando ai moduli e usando di nuovo la lipschitzianità di \mathbf{G} :

$$(2) \leq L \int |\mathbf{z}_1 - \mathbf{z}_2| P(d\mathbf{z}_1, d\mathbf{z}_2).$$

Passando all'estremo inferiore in P si ottiene

$$(2) \leq W_1(f_s, g_s)$$

Mettendo insieme le due stime si ha

$$|\mathbf{Z}_t^f(\mathbf{z}_1) - \mathbf{Z}_t^g(\mathbf{z}_2)| \leq |\mathbf{z}_1 - \mathbf{z}_2| + L \int_0^t ds W_1(f_s, g_s) + L \int_0^t ds |\mathbf{Z}_s^f(\mathbf{z}_1) - \mathbf{Z}_s^g(\mathbf{z}_2)|$$

Usando il lemma di Gronwall “congelando” la dipendenza temporale del termine in W_1 , si ottiene la tesi (dettagli al lettore).

Proposizione 2.2 (Dipendenza della distribuzione dal flusso). *Sia f_0 una distribuzione di probabilità con supporto in un compatto B_0 . Siano f_t^i , $i = 1, 2$ le soluzioni per $t \in [0, T]$ dell'equazione di Liouville di flussi \mathbf{Z}_t^i , generati dai campi di distribuzione g_t^i , di dati iniziali f_0^i . Le misure f_t^i hanno supporto in un compatto B che dipende solo da T , e sono debolmente continue in t . Inoltre esistono $C > 0$ e $\gamma \in (0, 1)$ tali che per T sufficientemente piccolo*

$$\sup_{t \in [0, T]} W_1(f_t^1, f_t^2) \leq CW_1(f_0^1, f_0^2) + \gamma \sup_{t \in [0, T]} W_1(g_t^1, g_t^2).$$

Poiché \mathbf{G} è limitato, chiamando M il suo massimo è immediato concludere che $|\mathbf{Z}^t(\mathbf{z}) - \mathbf{z}| \leq Mt$. Queste disuguaglianze provano che f_t^i hanno supporto in un opportuno compatto B . Sia $P_t(d\mathbf{z}_1, d\mathbf{z}_2)$ una distribuzione congiunta di f_t^1 e f_t^2 . Allora la scrittura

$$\int \phi(\mathbf{Z}_t^1(\mathbf{z}_1), \mathbf{Z}_t^2(\mathbf{z}_2)) P_0(d\mathbf{z}_1, d\mathbf{z}_2) = \int \phi(\mathbf{z}_1, \mathbf{z}_2) P_t(d\mathbf{z}_1, d\mathbf{z}_2)$$

definisce una distribuzione congiunta di f_0^1 e f_0^2 . Analogamente, ogni distribuzione congiunta di f_0^1 e f_0^2 viene mappata dai due flussi in una distribuzione congiunta di f_t^1 , f_t^2 . Dunque, usando il lemma precedente

$$\begin{aligned} \int |\mathbf{z}_1 - \mathbf{z}_2| P_t(d\mathbf{z}_1, d\mathbf{z}_2) &= \int |(\mathbf{Z}_t^1(\mathbf{z}_1) - \mathbf{Z}_t^2(\mathbf{z}_2))| P_0(d\mathbf{z}_1, d\mathbf{z}_2) \\ &\leq e^{Lt} \int |\mathbf{z}_1 - \mathbf{z}_2| P_0(d\mathbf{z}_1, d\mathbf{z}_2) + e^{Lt} \int_0^t W_1(g_t^1, g_t^2) \end{aligned}$$

Passando all'estremo inferiore in P_0 si conclude che

$$W_1(f_t^1, f_t^2) \leq e^{Lt} W_1(f_0^1, f_0^2) + e^{Lt} \int_0^t W_1(g_t^1, g_t^2)$$

da cui, passando all'estremo superiore in $[0, T]$ per T sufficientemente piccolo si ottiene la tesi.

A questo punto è semplice costruire la soluzione. Dirò che f_t è una soluzione debole dell'equazione di campo medio di dato iniziale f_0 se e solo se $f_t = \mathbf{Z}_t^f \# f_0$.

Teorema 2.6. *Sia data f_0 misura di probabilità al tempo 0, con supporto in un compatto B_0 . L'equazione di campo medio ammette un'unica soluzione debole di dato f_0 per $t \in [0, +\infty)$. La soluzione è debolmente continua nel dato iniziale.*

Sia T il tempo determinato nell'ultimo lemma, e sia g_t una famiglia di misure di probabilità a supporto in B e debolmente continua in t , con $g_0 = f_0$. Considera la mappa \mathcal{M}_t che associata alla famiglia di misure di probabilità g_t di dato iniziale f_0 , la famiglia di misure di probabilità $\mathcal{M}_t(g)$ soluzione al tempo t di dato iniziale f_0 ottenuta trasportando f_0 con il flusso generato dal campo determinato da g_t . Per il lemma precedente, \mathcal{M}_t è una contrazione, nel senso

$$\sup_{t \in [0, T]} W_1(\mathcal{M}_t(g^1), \mathcal{M}_t(g^2)) \leq \gamma \sup_{t \in [0, T]} W_1(g_t^1, g_t^2),$$

con $\gamma < 1$. Quindi in un intorno di f_0 esiste $\{f_t\}_{t \in [0, T]}$ tale che

$$\mathcal{M}_t(f) = f_t,$$

che dunque risolve l'equazione di Vlasov. è facile mostrare l'unicità e la continuità nel dato iniziale (esercizio).

3 Il limite di campo medio

3.1 La convergenza alla Dobrushin

Teorema 3.1 (Limite di campo medio). *Sia $\mathbf{z}_1^N(0)$ nel compatto B_0 dato iniziale per il sistema di particelle*

$$\frac{d}{dt}\mathbf{z}_i(t) = \frac{1}{N} \sum_{j=1}^N \mathbf{G}(\mathbf{z}_i(t), \mathbf{z}_j(t)). \quad (3.1)$$

e sia $\mathbf{z}_1^N(t)$ la corrispondente soluzione. Sia

$$\pi(\mathbf{z}_1^N(t), d\mathbf{z}) = \frac{1}{N} \sum_1^N \delta(\mathbf{z} - \mathbf{z}_i(t)) d\mathbf{z}$$

la **misura empirica** associata. Sia f_t una soluzione regolare dell'equazione di Vlasov. Se $\mu_0^N \rightharpoonup f_0$ allora $\mu_t^N \rightharpoonup f_t$.

La dimostrazione si basa sulla continuità nel dato iniziale delle soluzioni deboli dell'equazione di Vlasov, dopo aver notato che la misura empirica $\pi(\mathbf{z}_1^N(t), d\mathbf{z})$ è una soluzione debole dell'equazione di Vlasov.

Il punto chiave è il seguente: considero il campo generato dalla misura empirica π

$$\mathcal{K}[\pi](\mathbf{z}) = \int \mathbf{G}(\mathbf{z}, \tilde{\mathbf{z}}) \pi(\mathbf{z}_1^N(t), d\tilde{\mathbf{z}}) = \frac{1}{N} \sum_j \mathbf{G}(\mathbf{z}, \mathbf{z}_j(t))$$

Questo campo nel punto $\mathbf{z}_i(t)$ è proprio il campo che muove $\mathbf{z}_i(t)$, dunque il flusso \mathbf{Z}_t^π , che è definito in tutto lo spazio, muove le particelle esattamente come il sistema (3.1), cioè

$$\mathbf{Z}_t^N \# \pi(\mathbf{z}_1^N(0), d\mathbf{z}) = \pi(\mathbf{z}_1^N(t), d\mathbf{z}).$$

A questo punto, per continuità delle soluzioni deboli rispetto alla convergenza debole dei dati iniziali, si ottiene la tesi.

3.2 La convergenza della gerarchia

Abbiamo provato un risultato molto forte di validità dell'equazione di campo medio: se prendo dati iniziali che approssimano (debolmente) una distribuzione iniziale di probabilità, questa approssimazione vale anche al tempo t .

Si tratta però di un punto di vista completamente diverso da quello da cui siamo partiti, in cui la domanda era sulla convergenza della 1-marginale a una soluzione dell'equazione di campo medio. Più in generale ci si può chiedere se, ipotizzando che $f_0^N = f_0^{\otimes N}$,

$$\lim_{N \rightarrow +\infty} f_j^N(t) = f_t^{\otimes N}$$

dove f_t risolve l'equazione di campo medio di dato iniziale f_0 .

La risposta a questa domanda è positiva ed è nota da tempo, ma come sottoprodotto dell'argomento precedente. In tempi recenti è diventato possibile dimostrarla indipendentemente, con una tecnica sviluppata per analizzare il limite di campo medio per sistemi quantistici, in cui l'argomento di Dobrushin non si applica.

Ricordo che

$$f_t = \mathbf{Z}_t^f \# f_0, \quad f_t^N = \mathbf{Z}_1^N \# f_0^{\otimes N}.$$

Tensorizzando \mathbf{Z}_t^f ho che

$$f_t^{\otimes M} = (\mathbf{Z}_t^f)^{\otimes M} \# f_0^{\otimes M}$$

Voglio stimare $W_1(f_t, f_1^N(t))$. Per farlo costruisco una particolare distribuzione congiunta. Sia

$$P_0(d\mathbf{z}_1^N, d\mathbf{w}_1^N) = \delta(\mathbf{z}_1^N - d\mathbf{w}_1^N) f_0^{\otimes N}$$

Questa distribuzione congiunta è quella che realizza la distanza di Wasserstein nulla tra $f_0^{\otimes N}$ e $f_0^{\otimes N}$. Sia ora

$$P_t = (\mathbf{Z}_t^f)^{\otimes N} \otimes \mathbf{Z}_t^N \# P_0$$

Questa è una distribuzione congiunta di $f_t^{\otimes N}$ e f_t^N , infatti, se φ è una funzione test

$$\int \varphi(\mathbf{z}_1^N, \mathbf{w}_1^N) P_t(d\mathbf{z}_1^N, d\mathbf{w}_1^N) := \int \varphi((\mathbf{Z}_t^f)^{\otimes N}(\mathbf{z}_1^N), \mathbf{Z}_t^N(\mathbf{z}_1^N)) f_0^{\otimes N}(d\mathbf{z}_1^N)$$

Se φ non dipende dalle \mathbf{w}_i si ottiene φ integrato contro $f_t^{\otimes N}$, se non dipende dalle \mathbf{z}_i si ottiene φ integrato contro f_t^N .

Posso ora tentare la stima:

$$W_1(f_t, f_1^N(t)) \leq \int |\mathbf{z}_1 - \mathbf{w}_1| P_t(d\mathbf{z}_1^N, d\mathbf{w}_1^N)$$

infatti, la marginale di P_t nelle variabili \mathbf{z}_1 e \mathbf{w}_1 è una distribuzione congiunta di f_t e $f_1^N(t)$. Si noti che la stessa espressione vale per ogni i :

$$W_1(f_t, f_1^N(t)) \leq \int |\mathbf{z}_i - \mathbf{w}_i| P_t(d\mathbf{z}_1^N, d\mathbf{w}_1^N).$$

Per definizione di P_t :

$$W_1(f_t, f_1^N(t)) \leq \int |\mathbf{Z}_t^f(\mathbf{z}_i) - (\mathbf{Z}_t^N(\mathbf{z}_1^N))_i| f_0^{\otimes N}(d\mathbf{z}_1^N) \quad (3.2)$$

dove con $(\mathbf{Z}_t^N)_i$ intendo la componente del flusso relativa alla i -esima particella. Si noti che questo valore dipende da tutte le particelle, al contrario del termine in \mathbf{Z}_t^f che dipende solo da \mathbf{z}_i .

Stimo ora la differenza

$$|\mathbf{Z}_t^f(\mathbf{z}_i) - (\mathbf{Z}_t^N(\mathbf{z}_1^N))_i| \leq \int_0^t ds |\mathcal{K}[f_t](\mathbf{Z}_s^f(\mathbf{z}_i)) - \mathcal{K}[\pi(\mathbf{Z}_t^N(\mathbf{z}_1^N))](\mathbf{Z}_s^N(\mathbf{z}_1^N))_i|$$

Operando come nella proposizione 2.1, sommando e sottraendo $\mathcal{K}[f_t](\mathbf{Z}_t^N(\mathbf{z}_1^N))_i$ e notando che i dati iniziali sono gli stessi, si ottiene

$$|\mathbf{Z}_t^f(\mathbf{z}_i) - (\mathbf{Z}_t^N(\mathbf{z}_1^N))_i| \leq Le^{Lt} \int_0^t W_1(f_s, \pi(\mathbf{Z}_s^N(\mathbf{z}_1^N))) ds.$$

A questo punto uso la disuguaglianza triangolare, sommando e sottraendo la misura empirica calcolata nei punti evoluti con \mathbf{Z}_t^f :

$$W_1(f_s, \pi(\mathbf{Z}_s^N(\mathbf{z}_1^N))) \leq W_1(f_s, \pi(\mathbf{Z}_s^f \otimes \mathbf{Z}_s^N(\mathbf{z}_1^N))) + W_1(\pi(\mathbf{Z}_s^f \otimes \mathbf{Z}_s^N(\mathbf{z}_1^N)), \pi(\mathbf{Z}_s^N(\mathbf{z}_1^N)))$$

Per stimare il secondo termine uso l'accoppiamento particella $\mathbf{Z}_s^f(\mathbf{z}_i)$ con particella $(\mathbf{Z}_s^N(\mathbf{z}_1^N))_i$ per ogni i , ottenendo

$$W_1(\pi(\mathbf{Z}_s^{f^{\otimes N}}(\mathbf{z}_1^N)), \pi(\mathbf{Z}_s^N(\mathbf{z}_1^N))) \leq \frac{1}{N} \sum_j |\mathbf{Z}_s^f(z_j) - (\mathbf{Z}_s^N(\mathbf{z}_1^N))_j|$$

Inserendo questa espressione nella stima precedente si ottiene, per ogni i ,

$$|\mathbf{Z}_t^f(\mathbf{z}_i) - (\mathbf{Z}_t^N(\mathbf{z}_1^N))_i| \leq \frac{1}{N} \sum_j L e^{Lt} \int_0^t |\mathbf{Z}_s^f(z_j) - (\mathbf{Z}_s^N(\mathbf{z}_1^N))_j| + L e^{Lt} \int_0^t W_1(f_s, \pi(\mathbf{Z}_s^{f^{\otimes N}}(\mathbf{z}_1^N))) ds.$$

Sommando su i e dividendo per N , usando Gronwall si ha

$$\frac{1}{N} \sum_i |\mathbf{Z}_t^f(\mathbf{z}_i) - (\mathbf{Z}_t^N(\mathbf{z}_1^N))_i| \leq C(T) \int_0^t W_1(f_s, \pi(\mathbf{Z}_s^{f^{\otimes N}}(\mathbf{z}_1^N))) ds.$$

Ora sommo su i e divido per N la stima (3.2), ottenendo

$$W_1(f_t, f_1^N(t)) \leq C(T) \int_0^t ds W_1(f_s, \pi(\mathbf{Z}_s^{f^{\otimes N}}(\mathbf{z}_1^N))) f_0^{\otimes N}(d\mathbf{z}_1^N) = C(T) \int_0^t ds W_1(f_s, \pi(\mathbf{z}_1^N)) f_s^{\otimes N}(d\mathbf{z}_1^N)$$

che tende a 0 per $N \rightarrow +\infty$ per la legge dei grandi numeri.

4 L'equazione di Boltzmann

C'è un altro caso in cui si riesce a fare una descrizione ridotta di un sistema di particelle mediante un'equazione cinetica, ed è quello dell'equazione di Boltzmann che descrive i gas rarefatti. La ragione fisico-matematica che lo permette è completamente diversa da quella delle equazioni di campo medio.

Per illustrarla considero il caso più classico, considerando un modello di interazione di particelle molto semplificato, quello delle sfere dure: invece di considerare particelle microscopiche e un potenziale di interazione tra loro, si considerano particelle sferiche, che interagiscono solo tramite urti elastici. L'equazione di Boltzmann si può anche ottenere per potenziali di interazione a corto range, in tal caso ci sarà solo una differente espressione del cosiddetto nucleo di collisione (che nel seguito verrà indicato con B) (4.2)).

Per prima cosa bisogna descrivere con precisione la dinamica.

4.1 Urti tra sfere dure

Supponiamo che le particelle abbiano diametro $r > 0$, e che a un certo istante di tempo vengano in contatto (urto). In tal caso le posizioni sono $\mathbf{x}_2 = \mathbf{x}_1 + r\mathbf{n}$ dove \mathbf{n} è il versore di $\mathbf{x}_2 - \mathbf{x}_1$. Indico con \mathbf{v}_1 e \mathbf{v}_2 le velocità "entrantanti", cioè prima dell'urto. Affinché l'urto possa accadere, le particelle si devono avvicinare. Supponiamo che l'urto avvenga al tempo $t = 0$. Prima dell'urto $t < 0$ e le posizioni delle particelle sono $\mathbf{x}_i(t) = \mathbf{x}_i + t\mathbf{v}_i$ dunque

$$|\mathbf{x}_2(t) - \mathbf{x}_1(t)|^2 = |\mathbf{x}_2 - \mathbf{x}_1 + t(\mathbf{v}_2 - \mathbf{v}_1)|^2 = r^2 + 2tr\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) + t^2|\mathbf{v}_1 - \mathbf{v}_2|^2$$

che in $t = 0$ è decrescente solo se vale la condizione

$$\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0.$$

(la stessa conclusione si poteva trarre da considerazioni geometriche). Dopo l'urto, la posizione delle particelle è invariata, ma cambiano le velocità. Indico con \mathbf{v}'_i le velocità "uscenti" e provo a determinarne l'espressione in termini delle velocità entranti.

Anche se l'urto è una descrizione semplificata di un'interazione, devono valere i principi della fisica. In particolare, deve conservarsi la quantità di moto (perché non ci sono forze esterne), e, nell'ipotesi di urto elastico, deve conservarsi l'energia. Per imporre queste due condizioni è utile scrivere le velocità in termini della velocità del baricentro e della semi-differenza:

$$\bar{\mathbf{v}} = \frac{\mathbf{v}_1 + \mathbf{v}_2}{2}, \quad \mathbf{w} = \frac{\mathbf{v}_1 - \mathbf{v}_2}{2},$$

così che

$$\begin{aligned} \mathbf{v}_1 &= \bar{\mathbf{v}} + \mathbf{w} & \mathbf{v}'_1 &= \bar{\mathbf{v}} + \mathbf{w}' \\ \mathbf{v}_2 &= \bar{\mathbf{v}} - \mathbf{w} & \mathbf{v}'_2 &= \bar{\mathbf{v}} - \mathbf{w}'. \end{aligned} \quad \text{e}$$

Nell'ultima espressione ho usato che la velocità del baricentro si conserva, dunque

$$\frac{\mathbf{v}'_1 + \mathbf{v}'_2}{2} = \bar{\mathbf{v}}$$

e ho indicato con \mathbf{w}' la semi-somma delle velocità uscenti. In questo modo mi sono assicurato la conservazione della quantità di moto nell'urto, e ho ridotto il problema alla determinazione di \mathbf{w}' .

Impongo la conservazione dell'energia cinetica

$$E = \frac{1}{2}(|\mathbf{v}_1|^2 + |\mathbf{v}_2|^2) = |\bar{\mathbf{v}}|^2 + |\mathbf{w}|^2 = \frac{1}{2}(|\mathbf{v}'_1|^2 + |\mathbf{v}'_2|^2) = |\bar{\mathbf{v}}|^2 + |\mathbf{w}'|^2$$

che si riduce alla condizione

$$|\mathbf{w}'|^2 = |\mathbf{w}|^2.$$

Si noti che in un sistema unidimensionale ci sono solo due possibilità: $w' = w$ o $w' = -w$. Nel primo caso le velocità non cambiano nell'urto, nel secondo le particelle si scambiano le velocità (che è quello che accade colpendo in linea una biglia ferma con un'altra). Nel limite $r \rightarrow 0$ queste due possibilità danno lo stesso sistema: infatti nel secondo caso stiamo semplicemente scambiando il nome delle particelle dopo l'urto, senza nessuna variazione rispetto al moto libero.

In dimensione maggiore, la condizione di conservazione dell'energia non è evidentemente sufficiente a determinare il valore di \mathbf{w}' . Nel modello di sfere dure si aggiunge però una richiesta ragionevole: che la variazione di impulso avvenga nella direzione di \mathbf{n} , cioè in modo perpendicolare alla superficie delle sfere urtanti. Si pensi per esempio a una sfera che urta contro una parete e rimbalza elasticamente: la componente normale della velocità cambia segno, le componenti tangenti restano costanti.

Indico con $\mathbf{a} \otimes \mathbf{b}$ la matrice di componenti $a_i b_j$ che dunque agisce sui vettori in questo modo

$$\mathbf{a} \otimes \mathbf{b} \mathbf{v} = \mathbf{a} (\mathbf{b} \cdot \mathbf{v})$$

(la chiamerò un po' impropriamente "prodotto tensore"). Noto che $\mathbf{n} \otimes \mathbf{n}$ è l'operatore di proiezione lungo \mathbf{n} . Riscrivo la condizione di conservazione dell'energia imponendo che le componenti di \mathbf{w} ortogonali a \mathbf{n} non cambino, cioè che

$$(\mathbb{I} - \mathbf{n} \otimes \mathbf{n})\mathbf{w}' = (\mathbb{I} - \mathbf{n} \otimes \mathbf{n})\mathbf{w}.$$

Poiché $\mathbb{I} = (\mathbb{I} - \mathbf{n} \otimes \mathbf{n}) + \mathbf{n} \otimes \mathbf{n}$ ottengo

$$(\mathbf{n} \cdot \mathbf{w}')^2 = (\mathbf{n} \cdot \mathbf{w})^2$$

che mi da due scelte, una che mi ridà i valori delle velocità entranti, e l'altra che corrisponde a

$$\mathbf{n} \cdot \mathbf{w}' = -\mathbf{n} \cdot \mathbf{w};$$

in questo caso la componente della velocità relativa lungo \mathbf{n} viene invertita dall'urto. Ora posso scrivere le velocità uscenti in funzione delle entranti:

$$\begin{aligned}\mathbf{v}'_1 &= \mathbf{v}_1 - ((\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{n}) \mathbf{n} \\ \mathbf{v}'_2 &= \mathbf{v}_2 + ((\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{n}) \mathbf{n}\end{aligned}$$

Osservazione. Si osservi con attenzione l'espressione delle velocità uscenti per le due particelle: è esattamente la stessa. Inoltre, è immediato verificare che è un'espressione invariante per scambio tra velocità entranti e uscenti:

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{v}'_1 - ((\mathbf{v}'_1 - \mathbf{v}'_2) \cdot \mathbf{n}) \mathbf{n} \\ \mathbf{v}_2 &= \mathbf{v}'_2 + ((\mathbf{v}'_1 - \mathbf{v}'_2) \cdot \mathbf{n}) \mathbf{n}\end{aligned}$$

come è giusto: si tratta di un sistema fisico conservativo, dunque non c'è modo di distinguere velocità prima dell'urto da velocità dopo l'urto senza usare la posizione delle particelle.

La condizione di velocità entranti è $(\mathbf{v}_2 - \mathbf{v}_1) \cdot \mathbf{n} < 0$ mentre dopo l'urto vale $(\mathbf{v}'_2 - \mathbf{v}'_1) \cdot \mathbf{n} > 0$ (che è la condizione di "velocità uscenti").

Osservazione. L'urto conserva la misura nello spazio delle velocità:

$$d\mathbf{v}_1 d\mathbf{v}_2 = 2^d d\bar{\mathbf{v}} d\mathbf{w} = 2^d d\bar{\mathbf{v}} d\mathbf{w}' = d\mathbf{v}'_1 d\mathbf{v}'_2$$

4.2 La gerarchia

La scrittura della gerarchia per la dinamica delle sfere dure presenta qualche difficoltà concettuale. Intanto qualche riflessione sul flusso dato dalla dinamica delle sfere dure. È definito su $\Omega_r = \{(\mathbf{x}_1^N, \mathbf{v}_1^N) \mid \forall i \neq j, |\mathbf{x}_i - \mathbf{x}_j| > r\}$. Si noti che se $|\mathbf{x}_i - \mathbf{x}_j| = r$ la dinamica è ancora ben definita: se le velocità sono uscenti il moto allontana $\mathbf{x}_i, \mathbf{x}_j$, se sono entranti prima avviene l'urto, e poi il moto allontana le particelle.

In Ω_r il moto è libero, ogni particella si muove di moto rettilineo a velocità costante fino a che non urta. Dunque $\Phi^t(\mathbf{z}_1^N)$ è ben definito in Ω_r , incluso il bordo, e fuori dal bordo coincide con il flusso libero. Dunque $\mathbf{x}_1^N(t)$ è una funzione continua, $\mathbf{v}_1^N(t)$ è una funzione costante a tratti.

C'è un'ulteriore difficoltà: bisogna escludere il caso di urti contemporanei di più particelle, in cui la dinamica non è definita (e non si può definire in modo da conservare la continuità nelle variabili spaziali: si pensi a due particelle di velocità opposte che collidono frontalmente, e una terza, che urta con entrambe da una direzione ortogonale: perturbando di poco il tempo di urto delle prime due, la terza collide o con una o con l'altra, andando in direzioni differenti). Inoltre, bisogna assicurarsi che non si verificano infiniti urti in tempo finito (come invece può accadere con urti anelastici, che dissipano energia). Entrambe queste patologie si possono curare mostrando che l'insieme dei dati iniziali per cui la dinamica comporta urti multipli o infiniti urti in tempo finito è di misura nulla. Indico ancora con Ω_r questo insieme.

L'equazione di Liouville in Ω_r è particolarmente semplice perché il moto è un flusso libero tra un urto e l'altro:

$$\partial_t f^N + \sum_{i=1}^N \mathbf{v}_i \cdot \partial_{x_i} f^N = 0$$

Questa equazione coincide formalmente con l'equazione del flusso libero, ma va considerato che vale solo nella regione Ω_r e che al bordo devono valere delle condizioni di compatibilità, in accordo con la definizione del flusso. Marginalizzandola e con un bel po' di lavoro si ottiene la gerarchia.

Consideriamo il caso semplice di sole due particelle. Assumeremo anche che $f_0^2(\mathbf{z}_1, \mathbf{z}_2)$ abbia supporto, compatto, nella regione dello spazio delle fasi definita da $|\mathbf{x}_1 - \mathbf{x}_2| > r$, e sia continua. Prima dell'urto

$$\begin{aligned} \mathbf{x}_1(t) &= \mathbf{x}_1 + t\mathbf{v}_1 \\ \mathbf{x}_2(t) &= \mathbf{x}_2 + t\mathbf{v}_2 \\ \mathbf{v}_1(t) &= \mathbf{v}_1 \\ \mathbf{v}_2(t) &= \mathbf{v}_2 \end{aligned}$$

Dopo l'urto

$$\begin{aligned} \mathbf{x}_1(t) &= (\mathbf{x}_1 + \tau\mathbf{v}_1) + (t - \tau)\mathbf{v}'_1 \\ \mathbf{x}_2(t) &= \mathbf{x}_2 + \tau\mathbf{v}_2 + (t - \tau)\mathbf{v}'_2 \\ \mathbf{v}_1(t) &= \mathbf{v}'_1 \\ \mathbf{v}_2(t) &= \mathbf{v}'_2 \end{aligned}$$

dove τ è il primo tempo positivo tale che

$$|\mathbf{x}_2 - \mathbf{x}_1 + \tau(\mathbf{v}_2 - \mathbf{v}_1)| = r.$$

Chiediamoci cosa accade a f^2 quando il suo supporto raggiunge la regione di urto. Consideriamo un dato iniziale che urta al tempo τ nella configurazione $\bar{\mathbf{x}}_1 = \mathbf{x}_1 + \tau\mathbf{v}_1$ $\bar{\mathbf{x}}_2 = \mathbf{x}_2 + \tau\mathbf{v}_2$. Poiché f deve essere costante lungo il flusso, deve accadere che

$$\lim_{t \rightarrow \tau^-} f_t^2(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{v}_1, \mathbf{v}_2) = f_0(\mathbf{z}) = \lim_{t \rightarrow \tau^+} f_t^2(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{v}'_1, \mathbf{v}'_2)$$

Si noti che l'espressione a sinistra è calcolata per $\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0$, mentre quella a destra per $\mathbf{n} \cdot (\mathbf{v}'_2 - \mathbf{v}'_1) > 0$; se queste condizioni sono violate, i valori di f devono essere nulli. Stiamo imponendo la condizione di "riflessione" al bordo, data da

$$f_t^2(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{v}_1, \mathbf{v}_2) = f_t^2(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{v}'_1, \mathbf{v}'_2), \quad \text{per } \mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0.$$

Scrivendo questa condizione al bordo per il caso di N particelle e integrando l'equazione di Liouville si può ottenere la gerarchia, con un bel po' di lavoro. C'è un altro modo di procedere, che è forse più breve e ha anche il vantaggio di mettere in evidenza alcuni aspetti decisamente sottili della derivazione, anche solo formale, dell'equazione di Boltzmann.

In pratica, trasformerò la condizione al contorno in un termine singolare da aggiungere all'equazione di Liouville. In questo modo sarà più semplice marginalizzare.

Semplifico le notazioni:

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2), \quad \mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2), \quad \mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2) = (\mathbf{x}_1, \mathbf{v}_1, \mathbf{x}_2, \mathbf{v}_2);$$

indico con $f(\mathbf{z})$ la distribuzione delle due particelle, e con $\mathbf{Z}_t(\mathbf{z})$ il flusso di dato iniziale \mathbf{z} . In particolare con $\mathbf{v}_i(t)$ e $\mathbf{x}_i(t)$ indico velocità e posizione della particella i .

Sia $\Delta_t = \mathbf{x}_2 - \mathbf{x}_1 + t(\mathbf{v}_2 - \mathbf{v}_1)$; si noti che questa espressione coincide con $\mathbf{x}_2(t) - \mathbf{x}_1(t)$ solo prima dell'urto. Per semplicità, considero T tale che sul supporto di f_0 si ha $|\Delta_t| = r$ al più per un solo valore di $t \in [0, T]$. Questa condizione tecnica serve a poter identificare l'istante di urto con la condizione $|\Delta_t| = r$.

Sia $\alpha(\mathbf{z})$ una un funzione test. Sempre sotto le ipotesi che ho fatto per f_0 e per $t \leq T$,

$$\begin{aligned} \int \alpha(\mathbf{z}) f_t(\mathbf{z}) d\mathbf{z} &= \int \alpha(\mathbf{Z}_t(\mathbf{z})) f_0(\mathbf{z}) d\mathbf{z} \\ &= \int \alpha(\mathbf{Z}_t(\mathbf{z})) f_0(\mathbf{z}) \mathcal{X}\{|\Delta_t| > r\} + \int \alpha(\mathbf{Z}_t(\mathbf{z})) f_0(\mathbf{z}) \mathcal{X}\{|\Delta_t| < r\} \end{aligned}$$

Voglio derivare in t questa espressione per ottenere la versione distribuzionale dell'equazione di Liouville. Nella regione per $|\Delta_t| > r$, il flusso è quello libero, nell'altra \mathbf{Z}_t assume i valori dopo l'urto. In entrambi i casi,

$$\dot{\mathbf{x}}_i(t) = \mathbf{v}_i(t), \quad \dot{\mathbf{v}}_i(t) = 0$$

e dunque

$$\frac{d}{dt} \alpha(\mathbf{Z}_t(\mathbf{z})) = \sum_i \mathbf{v}_i(t) \cdot \partial_{x_i} \alpha(\mathbf{Z}_t(\mathbf{z}))$$

Vanno però considerate anche le due funzioni indicatrici, che hanno derivata in t opposta perché la loro somma è 1. Questo contributo è

$$\int d\mathbf{z} f_0(\mathbf{z}) \frac{d}{dt} \mathcal{X}\{|\Delta_t| > r\} (\alpha(\mathbf{Z}_t^-(\mathbf{z})) - \alpha(\mathbf{Z}_t^+(\mathbf{z})))$$

dove con \mathbf{Z}_t^\pm indico il valore del flusso al tempo t dopo e prima dell'urto (si noti che f_t è continua nell'urto per costruzione, ma questo fatto non vale necessariamente per la funzione test α).

Vale

$$\frac{d}{dt} \mathcal{X}\{|\Delta_t| > r\} = \delta(|\Delta_t| - r) \frac{d}{dt} |\Delta_t|$$

Inoltre

$$\frac{d}{dt} |\Delta_t| = \frac{1}{|\Delta_t|^2} \Delta_t \cdot (\mathbf{v}_2 - \mathbf{v}_1)$$

che va calcolato esattamente quando $|\Delta_t| = r$; in tal caso $|\Delta_t| = \mathbf{x}_2(t) - \mathbf{x}_1(t)$, con $|\mathbf{x}_2(t) - \mathbf{x}_1(t)| = r$. Risulta quindi

$$\frac{d}{dt} \mathcal{X}\{|\Delta_t| < r\} = \delta(|\mathbf{x}_2(t) - \mathbf{x}_1(t)| - r) \mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1).$$

Si noti le variabili di integrazione $\mathbf{v}_1, \mathbf{v}_2$ sono le velocità iniziali, che non cambiano fino all'urto, dunque posso scrivere $\mathbf{v}_2 - \mathbf{v}_1 = \mathbf{v}_2(t)^- - \mathbf{v}_1(t)^-$. Il primo termine dell'integrale è dunque

$$\int f_0(\mathbf{z}) \alpha(\mathbf{Z}_t^-) \delta(|\mathbf{x}_2(t) - \mathbf{x}_1(t)| - r) [\mathbf{n} \cdot (\mathbf{v}_2(t)^- - \mathbf{v}_1(t)^-)]^-$$

dove con $[x]^-$ si intende la parte negativa di x , cioè x se x è negativo, 0 altrimenti. Si noti che la funzione caratteristica sul fatto che le velocità siano entranti è implicita nella condizione $|\Delta_t| > r$. Tornando indietro con il flusso, l'integrale diventa

$$\int d\mathbf{z} \alpha(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) f_t(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) \delta(|\mathbf{x}_2 - \mathbf{x}_1| - r) [\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)]^-$$

dove $r\mathbf{n} = \mathbf{x}_2 - \mathbf{x}_1$. Notando che

$$\delta(|\mathbf{x}_1 - \mathbf{x}_2| - r) d\mathbf{x}_2 = r^2 d\mathbf{n} \delta_{\mathbf{x}_1 + r\mathbf{n}}(d\mathbf{x}_2)$$

posso riscrivere il termine più esplicitamente come

$$-r^2 \int d\mathbf{x}_1 d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| d\mathbf{n} f_t(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \alpha(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2)$$

L'altro termine è

$$\begin{aligned} & - \int f_0(\mathbf{z}) \alpha(\mathbf{Z}_t^+) \delta(|\mathbf{x}_2(t) - \mathbf{x}_1(t)| - r) [\mathbf{n} \cdot (\mathbf{v}_2^-(t) - \mathbf{v}_1^-(t))]^- \\ & = - \int f_0(\mathbf{z}) \alpha(\mathbf{x}_1(t), \mathbf{x}_2(t), \mathbf{v}_1^-(t)', \mathbf{v}_2^-(t)') \delta(|\mathbf{x}_2(t) - \mathbf{x}_1(t)| - r) [\mathbf{n} \cdot (\mathbf{v}_2^-(t) - \mathbf{v}_1^-(t))]^- . \end{aligned}$$

Tornando indietro con il flusso ottengo che l'osservabile α è calcolato nelle velocità uscenti:

$$- \int f_t(\mathbf{z}) \alpha(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1', \mathbf{v}_2') \delta(|\mathbf{x}_2 - \mathbf{x}_1| - r) [\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)]^- .$$

Procedendo esattamente come sopra ottengo

$$r^2 \int d\mathbf{x}_1 d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| d\mathbf{n} f_t(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \alpha(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1', \mathbf{v}_2')$$

Cambio variabili di integrazione, da $\mathbf{v}_1, \mathbf{v}_2$ a $\mathbf{v}_1', \mathbf{v}_2'$, e poi cambio nome alle variabili stesse, infine cambio variabile di integrazione anche da \mathbf{n} a $-\mathbf{n}$ (il modulo del determinante di questo cambiamento di coordinate è 1). Ottengo dunque

$$r^2 \int d\mathbf{x}_1 d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| d\mathbf{n} f_t(\mathbf{x}_1, \mathbf{x}_1 - r\mathbf{n}, \mathbf{v}_1', \mathbf{v}_2') \alpha(\mathbf{x}_1, \mathbf{x}_1 - r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2)$$

Sono quindi pronto per scrivere la versione distribuzionale dell'equazione di Liouville nel caso di due particelle:

$$\begin{aligned} \partial_t f_t(\mathbf{z}) + \sum_{i=1}^2 \mathbf{v}_i \cdot \nabla f_t(\mathbf{z}) &= \mathcal{T}_{12} f_t := \\ r^2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| & (\delta_{\mathbf{x}_1 - r\mathbf{n}}(d\mathbf{x}_2) f_t(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1', \mathbf{v}_2') - \delta_{\mathbf{x}_1 + r\mathbf{n}}(d\mathbf{x}_2) f_t(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2)) \end{aligned}$$

Si noti che il secondo membro è una distribuzione, che ha supporto solo sul bordo. Abbiamo ottenuto un'espressione relativamente comprensibile: nella regione al bordo data da $|\mathbf{x}_1 - \mathbf{x}_2| = r$, dove avviene l'urto, f varia perché compaiono i valori calcolati nelle velocità uscenti

(il primo termine è detto termine di “gain”, guadagno) e si azzerano i termini nelle velocità entranti (il secondo termine è detto termine di “loss”, perdita).

È ora semplice riscrivere l'equazione di Liouville nel caso generale:

$$\begin{aligned} \partial_t f_t^N(\mathbf{z}) + \sum_{i=1}^N \mathbf{v}_i \cdot \nabla f_t(\mathbf{z}) &= \sum_{i < j} \mathcal{T}_{ij} f^N \\ \mathcal{T}_{ij} f^N(\mathbf{z}) &:= r^2 \int_{\mathbf{n} \cdot (\mathbf{v}_j - \mathbf{v}_i) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_j - \mathbf{v}_i)| \left(\delta_{\mathbf{x}_i - r\mathbf{n}}(d\mathbf{x}_j) f^N(\tilde{\mathbf{z}}_{ij}) - \delta_{\mathbf{x}_i + r\mathbf{n}}(d\mathbf{x}_j) f^N(\mathbf{z}) \right) \end{aligned}$$

dove $\tilde{\mathbf{z}}_{ij}$ si ottiene da \mathbf{z} considerando le velocità uscenti per le particelle i e j .

Lemma 4.1.

$$\int d\mathbf{x}_i d\mathbf{x}_j d\mathbf{v}_i d\mathbf{v}_j \mathcal{T}_{ij} f^N = 0$$

Questa proprietà non deve sorprendere, perché la probabilità totale si conserva. Consideriamo il caso \mathcal{T}_{12} : l'integrale diventa

$$r^2 \int d\mathbf{x} d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| \left(f(\mathbf{x}, \mathbf{x} - r\mathbf{n}, \mathbf{v}'_1, \mathbf{v}'_2) - f^2(\mathbf{x}, \mathbf{x} + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \right)$$

Nel termine di gain, a \mathbf{x} fissato passo all'integrale nelle velocità uscenti, e poi cambio nome alle velocità:

$$\begin{aligned} &\int d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| f(\mathbf{x}, \mathbf{x} - r\mathbf{n}, \mathbf{v}'_1, \mathbf{v}'_2) \\ &= \int d\mathbf{v}'_1 d\mathbf{v}'_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| f(\mathbf{x}, \mathbf{x} - r\mathbf{n}, \mathbf{v}'_1, \mathbf{v}'_2) \\ &= \int d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}'_2 - \mathbf{v}'_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}'_2 - \mathbf{v}'_1)| f(\mathbf{x}, \mathbf{x} - r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \\ &= \int d\mathbf{v}_1 d\mathbf{v}_2 \int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) > 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| f(\mathbf{x}, \mathbf{x} - r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \end{aligned}$$

infine, cambio segno a \mathbf{n} , e osservo che ho ottenuto un termine che cancella esattamente quello di loss.

Considero ora l'equazione di Liouville, e integro in $d\mathbf{z}_2^N$. Il termine di trasporto dà $\mathbf{v}_1 \cdot \nabla_{x_1} f_1^N$, l'integrale dei termini a destra dà zero nel caso $2 \leq i < j$. Sopravvivono gli $N - 1$ termini dati da $1 = i < k$. Si vede facilmente che sono tutti uguali, dunque si ottiene la seguente equazione per la 1-marginale:

$$\begin{aligned} \partial_t f_1^N + \mathbf{v}_1 \cdot \nabla_{x_1} f_1^N &= (N - 1) r^2 \int d\mathbf{v}_2 \\ &\int_{\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| \left(f_2^N(\mathbf{x}_1, \mathbf{x}_2 - r\mathbf{n}, \mathbf{v}'_1, \mathbf{v}'_2) - f_2^N(\mathbf{x}_1, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2) \right) \end{aligned} \quad (4.1)$$

Non è difficile scrivere le successive equazioni della gerarchia. Le trovate scritte nell'appendice B, dedicata alla derivazione tradizionale.

4.3 Il limite di Boltzmann-Grad

Abbiamo ottenuto

$$\partial_t f_1^N + \mathbf{v}_1 \cdot \partial_{\mathbf{x}_1} f_1^N = G - L$$

dove i termini di gain e di loss sono dati da

$$G := (N-1)r^2 \int d\mathbf{v}_2 \int_{S^-} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| f_2^N(\mathbf{x}_1, \mathbf{x}_2 - r\mathbf{n}, \mathbf{v}'_1, \mathbf{v}'_2)$$

$$L := (N-1)r^2 \int d\mathbf{v}_2 \int_{S^-} d\mathbf{n} |\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1)| f_2^N(\mathbf{x}_1, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2)$$

dove S^- è la semi-superficie sferica individuata da $\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_1) < 0$.

Rileggiamo questa equazione in termini più fisici. Una particella con velocità \mathbf{v}_1 , di raggio r in un tempo τ occupa nello spazio delle configurazioni una regione di volume di ordine $\tau|\mathbf{v}_1|r^2$. Nella direzione \mathbf{n} , il volume spannato è $\tau|\mathbf{v}_1 \cdot \mathbf{n}|r^2$. Se $g(\mathbf{v}_2)$ è la densità di probabilità delle particelle con velocità \mathbf{v}_2 , e N è il numero di particelle, il numero di urti con particelle di velocità \mathbf{v}_2 nella direzione \mathbf{n} è

$$\tau|(\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{n}|r^2 N g(\mathbf{v}_2) d\mathbf{v}_2$$

Il termine $r^2|(\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{n}|$ prende il nome di **sezione d'urto**, ed è la densità di probabilità di un urto nella direzione \mathbf{n} tra particelle con velocità $\mathbf{v}_1, \mathbf{v}_2$ nell'unità di tempo. Integrando in $d\mathbf{n}$ con $\mathbf{n} \in S^-$ si ottiene la sezione d'urto totale. Il **cammino libero medio** è la distanza che una particella percorre prima di urtare. Dall'espressione precedente si ottiene che la particella fa un solo urto se percorre uno spazio $\tau|\mathbf{v}_1|$ di ordine $1/(r^2N)$,

Passando al limite $N \rightarrow +\infty$, per ottenere un termine di ordine 1, è necessario che $r^2N = \lambda > 0$ sia fissato. Questo è il limite di **bassa densità**, detto anche limite di Boltzmann-Grad. In questo caso il cammino libero medio di una particella è di ordine $1/\lambda$, e quindi anche il tempo tra un urto e l'altro è di ordine $1/\lambda$ (si osservi che nel limite di campo medio il cammino libero medio è invece 0). In questo senso la densità è bassa: le particelle si muovono di moto libero, facendo un urto in un tempo di ordine 1. Il limite di bassa densità non è dunque adatto alla descrizione di un gas atmosferico in condizioni normali, ma lo è per esempio per i rarefatti gas della mesosfera.

Nel limite $N \rightarrow +\infty, r \rightarrow 0, Nr^2 \rightarrow \lambda$, ipotizzando che $f_2^N \rightarrow f_2$ e $f_1^N \rightarrow f_1$ si ottiene

$$\partial_t f_1 + \mathbf{v}_1 \cdot \partial_{\mathbf{x}_1} f_1 = \lambda \int d\mathbf{v}_2 \int_{S^-} d\mathbf{n} |(\mathbf{v}_1 - \mathbf{v}_2) \cdot \mathbf{n}| (f_2(\mathbf{x}_1, \mathbf{x}_1, \mathbf{v}'_1, \mathbf{v}'_2) - f_2(\mathbf{x}_1, \mathbf{x}_1, \mathbf{v}_1, \mathbf{v}_2))$$

e equazioni analoghe per f_j con $j \geq 2$. Questa gerarchia di equazioni ammette soluzioni fattorizzate se $f_1 = f$ risolve l'equazione di Boltzmann, che si ottiene supponendo che

$$f_2(\mathbf{x}_1, \mathbf{x}_2, \mathbf{v}_1, \mathbf{v}_2) = f(\mathbf{x}_1, \mathbf{v}_2) f(\mathbf{x}_2, \mathbf{v}_2),$$

ed è

$$\partial_t f + \mathbf{v} \cdot \partial_{\mathbf{x}} f = \lambda \int d\mathbf{v}_* \int_{S^-} d\mathbf{n} |(\mathbf{v} - \mathbf{v}_*) \cdot \mathbf{n}| (f' f'_* - f f_*)$$

dove

$$f = f(\mathbf{x}, \mathbf{v}), \quad f_* = f(\mathbf{x}, \mathbf{v}_*)$$

è f calcolata nelle velocità precollisionali, e

$$f' = f(\mathbf{x}, \mathbf{v}'), \quad f'_* = f(\mathbf{x}, \mathbf{v}_*)$$

è f calcolata nelle velocità precollisionali con velocità uscenti \mathbf{v}, \mathbf{v}_* . Infine, poiché non c'è più dipendenza da \mathbf{n} nelle variabili spaziali, si può integrare su tutto S dividendo per 2. Scalando λ di un coefficiente 2 si ottiene

$$\begin{aligned} \partial_t f + \mathbf{v} \cdot \partial_x f &= \lambda Q(f, f) \\ Q(f, f) &= \int_S d\mathbf{n} \int d\mathbf{v}_* |(\mathbf{v} - \mathbf{v}_*) \cdot \mathbf{n}| (f' f'_* - f f_*). \end{aligned} \quad (4.2)$$

Nel seguito porrò $\lambda = 1$.

Un'ultima complicata osservazione: per scrivere l'equazione abbiamo ipotizzato nel termine di loss che f_2 sia fattorizzata nelle variabili precollisionali, e lo stesso abbiamo fatto nel termine di gain. Se f_2 nel termine di gain si fattorizzasse nelle variabili uscenti i due termini si sarebbero compensati dando 0.

La prova rigorosa della validità dell'equazione di Boltzmann è concettualmente sottile, ed è stata ottenuta da Lanford nel 1973, per tempi dell'ordine del tempo di collisione. La difficoltà non dovrebbe sorprendere, perché l'equazione di Boltzmann è un'equazione irreversibile, come vedremo nel prossimo paragrafo.

Lo scaling di Boltzmann-Grad

C'è un modo più formale per giustificare il nome di limite di bassa densità. Fissiamo r , scegliamo un parametro ε , e definiamo le funzioni $f^{N,\varepsilon}$ nel seguente modo:

$$f_j^N(t, \mathbf{x}_1^j, \mathbf{v}_1^j) = \varepsilon^{3j} f_j^{N,\varepsilon}(\varepsilon t, \varepsilon \mathbf{x}_1^j, \mathbf{v}_1^j).$$

Si noti che le velocità non scalano perché spazio e tempo sono scalate nello stesso modo. Le nuove funzioni sono normalizzate a 1, cioè sono ancora densità di probabilità. Ipotizziamo che $f_j^{N,\varepsilon}$ varia su scala 1 in t e \mathbf{x} ; in tal caso stiamo assumendo che f_j^N varia su scala ε . In questo senso, f^N è scritta in variabili "microscopiche", mentre $f^{N,\varepsilon}$ è scritta in variabili "macroscopiche". Sostituendo nell'equazione per la prima marginale, si ottiene che il termine di collisione ha un coefficiente davanti dato da

$$\frac{1}{\varepsilon} \varepsilon^{-3} \varepsilon^6 (N-1) r^2 = (N-1) (\varepsilon r)^2,$$

dove ε^{-3} viene dalla normalizzazione di $f_1^{N,\varepsilon}$, ε^6 da quella di $f_2^{N,\varepsilon}$, mentre $1/\varepsilon$ viene dal rapporto tra le derivate in t e \mathbf{x} microscopiche e le corrispondenti derivate macroscopiche.

Il limite di Boltzmann-Grad si ottiene mandando N a $+\infty$ e $\varepsilon \rightarrow 0$, in modo che $N \varepsilon^2 r^2 = \lambda$ fissato non nullo. Fissando questo valore λ , si ha che il numero di particelle in un cubo di lato $1/\varepsilon$ in variabili microscopiche non va come il volume $1/\varepsilon^3$, ma come $N \approx \lambda / (r^2 \varepsilon^2)$. Dunque si tratta di un limite di bassa densità.

4.4 Proprietà dell'equazione di Boltzmann

Indico con $Q(f, f)$ il termine di collisione

$$Q(f, f) = \int d\mathbf{v}_* \int_S d\mathbf{n} B (f' f'_* - f f_*)$$

dove $B = B(|\mathbf{v} - \mathbf{v}_*|, \mathbf{n} \cdot (\mathbf{v} - \mathbf{v}_*))$ è il nucleo di collisione, che dipende in generale da $|\mathbf{v} - \mathbf{v}_*|$ e dall'angolo tra \mathbf{n} e $\mathbf{v} - \mathbf{v}_*$. Nel caso delle sfere dure è $|\mathbf{n} \cdot (\mathbf{v} - \mathbf{v}_*)|$. Sia $\varphi(\mathbf{v})$ un qualunque osservabile regolare nelle velocità.

Lemma 4.2.

$$\begin{aligned} \int \varphi(\mathbf{v}) Q(f, f)(\mathbf{v}) d\mathbf{v} &= \frac{1}{4} \int d\mathbf{v} d\mathbf{v}_* \int_S d\mathbf{n} B(f' f'_* - f f_*) (\varphi + \varphi_* - \varphi' \varphi'_*) = \\ &= \int d\mathbf{v} d\mathbf{v}_* \int_S d\mathbf{n} B f f_* (\varphi' + \varphi'_* - \varphi - \varphi_*) \end{aligned}$$

La prova si basa sul fatto che $B(f' f'_* - f f_*)$ è invariante nello scambio $\mathbf{v} \leftrightarrow \mathbf{v}_*$, e cambia di segno nello scambio $(\mathbf{v}, \mathbf{v}_*) \leftrightarrow (\mathbf{v}', \mathbf{v}'_*)$ mentre $d\mathbf{v} d\mathbf{v}_* = d\mathbf{v}' d\mathbf{v}'_*$. Dettagli al lettore.

Per economia espositiva, discuto per ora dell'equazione di Boltzmann omogenea, cioè assumo $f(\mathbf{x}, \mathbf{v}) = f(\mathbf{v})$, pensando che \mathbf{x} viva in un dominio limitato senza bordo, in pratica in un toro. Si noti che l'integrale nella variabile \mathbf{v} di f non deve essere 1, ma l'inverso della misura del toro.

Teorema 4.1 (Quantità conservate). *Se f è una soluzione regolare dell'equazione di Boltzmann omogenea*

$$\partial_t f = Q(f, f)$$

allora si conservano

- *la densità spaziale di probabilità (che a volte chiamerò densità di massa)*

$$\int f(t, \mathbf{v}) d\mathbf{v} = \rho$$

- *la densità di quantità di moto*

$$\int \mathbf{v} f(t, \mathbf{v}) d\mathbf{v} = \rho \mathbf{u},$$

dove \mathbf{u} è la velocità media

- *la densità di energia*

$$\frac{1}{2} \int |\mathbf{v}|^2 f(t, \mathbf{v}) d\mathbf{v} = \frac{1}{2} \rho |\mathbf{u}|^2 + \rho e$$

dove

$$e = \frac{1}{2\rho} \int |\mathbf{v} - \mathbf{u}|^2 f(t, \mathbf{v}) d\mathbf{v}$$

è l'energia interna specifica (cioè per unità di massa), anch'essa costante.

La prova è immediata, faccio solo notare che la decomposizione dell'energia si ottiene in questo modo:

$$\frac{1}{2} \int |\mathbf{v}|^2 f = \frac{1}{2} \int |\mathbf{v} - \mathbf{u} + \mathbf{u}|^2 f = \frac{1}{2} \int |\mathbf{v} - \mathbf{u}|^2 f + \int (\mathbf{v} - \mathbf{u}) \cdot \mathbf{u} f + \frac{1}{2} \int |\mathbf{u}|^2 f$$

Il primo termine è ρe , il secondo è nullo perché \mathbf{u} esce dall'integrale e l'integrale dà 0 perché \mathbf{u} è la media di \mathbf{v} , il terzo è l'energia cinetica "del baricentro", cioè quella che si avrebbe se tutta la massa fosse concentrata nel baricentro, e vale $\rho |\mathbf{u}|^2 / 2$.

Teorema 4.2 (Il teorema H). *Sia f una soluzione regolare dell'equazione di Boltzmann omogenea. L'entropia $\mathcal{H}(f) = \int f \log f$ è una funzione crescente del tempo, inoltre f converge per $t \rightarrow +\infty$ alla distribuzione di Maxwell-Boltzmann*

$$M_{\rho, \mathbf{u}, T} = \frac{\rho}{(2\pi k_B T)^{3/2}} e^{-\frac{|\mathbf{v}-\mathbf{u}|^2}{2k_B T}}$$

dove T è la temperatura, con $e = \frac{3}{2}k_B T$ (k_B è la costante di Boltzmann), e i parametri massa ρ , $\rho \mathbf{u}$, pe sono quelli del dato iniziale.

Dimostrazione.

$$\frac{d}{dt} \mathcal{H}(f) = \int (\log f + 1) \partial_t f = \int \log f Q(f, f)$$

dove nell'ultima uguaglianza ho usato che $\int \partial_t f = \partial_t \int f = 0$. Usando il lemma con $\varphi = \log f$ si ottiene

$$\frac{d}{dt} \mathcal{H}(f) = \frac{1}{4} \int B \log \frac{f f_*}{f' f'_*} (f' f'_* - f f_*)$$

che è negativo perché $(x - y) \log(y/x) \leq 0$, ed è zero solo se $x = y$.

Determino possibili equilibri del sistema. Deve valere $f' f'_* = f f_*$, ovvero $\log f' + \log f'_* = \log f + \log f_*$, cioè $\log f$ è un **invariante di collisione**. Si può mostrare che gli unici invarianti di collisione sono le combinazioni lineari di 1, delle componenti di \mathbf{v} e di $|\mathbf{v}|^2$. Imponendo tale condizione si ottiene $\log f(\mathbf{v}) = a + \mathbf{b} \cdot \mathbf{v} + c|\mathbf{v}|^2$, dove a, \mathbf{b}, c sono cinque costanti. Quindi f è l'esponenziale del membro di destra. Imponendo che l'equilibrio abbia gli stessi parametri conservati del dato iniziale, si ottiene l'espressione della maxwelliana data sopra (dettagli al lettore, si noti solo che la maxwelliana ha 'matrice di covarianza' $\rho k_B T \mathbb{I}$, dunque il valore atteso di $|\mathbf{v} - \mathbf{u}|^2$ è $3\rho k_B T$).

Fissati gli integrali primi, l'equilibrio è unico, e si può mostrare, con una sorta di teorema di Liapunov, che effettivamente $f(t) \rightarrow M$. □

Proposizione 4.1 (Estremi dell'entropia). *Sia $\mathcal{M}_{\rho, \mathbf{u}, T}$ l'insieme delle misure di probabilità nelle \mathbf{v} con ρ, \mathbf{u}, T fissati. L'entropia \mathcal{H} è un funzionale convesso su $\mathcal{M}_{\rho, \mathbf{u}, T}$ e ha un minimo stretto in $M_{\rho, \mathbf{u}, T}$.*

Il minimo di \mathcal{H} è raggiunto nella maxwelliana $M_{\rho, \mathbf{u}, T}$.

Dimostrazione. Si noti che affinché \mathcal{H} sia definita è necessario che f sia a.c. rispetto a Lebesgue. \mathcal{H} viene definita $+\infty$ se questa condizione non è soddisfatta. Si ottiene facilmente che $M_{\rho, \mathbf{u}, T}$ è l'unico punto stazionario di \mathcal{H} in \mathcal{M} minimizzando \mathcal{H} con i moltiplicatori di Lagrange per i parametri idrodinamici ρ, \mathbf{u}, T . □

Nel seguito, per semplicità di notazioni, assorbirò la costante k_B dentro T , in pratica scriverò le equazioni con $k_B = 1$. Considero ora il caso in cui f non è spazialmente omogenea. Poiché il termine $\mathbf{v} \cdot \partial_x f$ coincide con il termine di divergenza $\text{div}_x(\mathbf{v}f)$, sia ha che per ogni osservabile della sola variabile \mathbf{v}

$$\int d\mathbf{x} \int d\mathbf{v} \varphi(\mathbf{v}) \mathbf{v} \cdot \partial_x f = \int d\mathbf{x} \text{div}_x \left(\int d\mathbf{v} \varphi(\mathbf{v}) \mathbf{v} f \right) = 0$$

per il teorema della divergenza (assumendo che f decada abbastanza rapidamente a infinito, per esempio che sia supporto compatto in \mathbf{x} , o che il dominio sia un toro). Da questa osservazione discende il seguente teorema

Teorema 4.3 (Quantità conservate - caso non omogeneo). *Se f è una soluzione regolare dell'equazione di Boltzmann*

$$\partial_t f + \mathbf{v} \cdot \partial_x f = Q(f, f)$$

allora si conservano

- *la probabilità totale*

$$\int f(t, \mathbf{v}) \, d\mathbf{v} \, d\mathbf{x} = 1$$

- *la quantità di moto totale*

$$\int \mathbf{v} f(t, \mathbf{v}) \, d\mathbf{v} \, d\mathbf{x}$$

- *l'energia totale*

$$\frac{1}{2} \int |\mathbf{v}|^2 f(t, \mathbf{v}) \, d\mathbf{v} \, d\mathbf{x}$$

Inoltre l'entropia $\int \mathcal{H}(f) \, d\mathbf{x}$ decresce.

Si noti che tutti i teoremi precedenti sono solo formali, perché non sono inseriti in una teoria che inizia con risultati di esistenza e unicità (che comunque è disponibile per il caso omogeneo).

4.5 Il limite idrodinamico

L'esistenza degli invarianti di collisione suggerisce la possibilità di ottenere equazioni per le quantità macroscopiche legate agli invarianti, integrando nella sola variabile \mathbf{v} . Per esempio, integrando in \mathbf{v} l'equazione, e usando che $\int Q(f, f) \, d\mathbf{v} = 0$, si ottiene la prima equazione dell'idrodinamica, l'**equazione di continuità**:

$$\partial_t \int f \, d\mathbf{v} + \operatorname{div}_x \int \mathbf{v} f \, d\mathbf{v} = 0$$

Definendo

$$\rho(\mathbf{x}, t) = \int f(t, \mathbf{x}, \mathbf{v}, t) \, d\mathbf{v}, \quad \rho(\mathbf{x}, t) \mathbf{u}(\mathbf{x}, t) = \int f(t, \mathbf{x}, \mathbf{v}, t) \mathbf{v} \, d\mathbf{v}$$

l'equazione precedente diventa

$$\partial_t \rho + \operatorname{div}_x(\rho \mathbf{u}) = 0$$

che esprime la **legge di conservazione** della massa. Infatti afferma che la **corrente** della densità di massa è proprio l'impulso $\rho \mathbf{u}$ (si veda l'introduzione all'equazione di Liouville per chiarire questa affermazione).

Moltiplicando l'equazione di Boltzmann per \mathbf{v} e integrando in \mathbf{v} si ottiene

$$\partial_t \int \mathbf{v} f \, d\mathbf{v} + \operatorname{div}_x \int \mathbf{v} \otimes \mathbf{v} f \, d\mathbf{v} = 0$$

dove, se A è una matrice $n \times n$, con $\operatorname{div}_x A$ si intende il vettore composto dalle divergenze delle righe di A :

$$(\operatorname{div}_x A)_i = \sum_j \partial_{x_j} A_{ij}.$$

Il primo termine dell'equazione è semplicemente $\partial_t(\rho\mathbf{u})$. Il secondo richiede qualche manipolazione per essere compreso meglio:

$$\begin{aligned}\int \mathbf{v} \otimes \mathbf{v} f \, d\mathbf{v} &= \int (\mathbf{v} - \mathbf{u}) \otimes \mathbf{v} f \, d\mathbf{v} + \mathbf{u} \otimes \int \mathbf{v} f \, d\mathbf{v} \\ &= \int (\mathbf{v} - \mathbf{u}) \otimes (\mathbf{v} - \mathbf{u}) f \, d\mathbf{v} + \left(\int (\mathbf{v} - \mathbf{u}) f \, d\mathbf{v} \right) \otimes \mathbf{u} + \mathbf{u} \otimes \int \mathbf{v} f \, d\mathbf{v}\end{aligned}$$

Il terzo termine è $\rho\mathbf{u} \otimes \mathbf{u}$, il secondo è nullo, perché la media di \mathbf{v} è proprio \mathbf{u} , il primo termine è una matrice (un tensore) che indico con N .

Dunque l'equazione del **bilancio della quantità di moto** è

$$\partial_t(\rho\mathbf{u}) + \operatorname{div}_x(\rho\mathbf{u} \otimes \mathbf{u}) = -\operatorname{div}_x N$$

I due termini hanno due interpretazioni differenti: il primo è un termine di corrente, e infatti dipende solo da ρ e \mathbf{u} . Il termine a destra è un termine dovuto a cambiamenti in \mathbf{v} della distribuzione f . In idrodinamica questo termine è legato alle "forze interne" al fluido.

Definisco la densità specifica di energia interna come

$$\rho e := \int \frac{1}{2}(\mathbf{v} - \mathbf{u})^2 f \, d\mathbf{v}.$$

Moltiplicando l'equazione di Boltzmann per $|\mathbf{v}|^2$ e integrando si ottiene l'equazione del **bilancio dell'energia**:

$$\partial_t \frac{1}{2} \int |\mathbf{v}|^2 f \, d\mathbf{v} + \operatorname{div}_x \frac{1}{2} \int |\mathbf{v}|^2 \mathbf{v} f \, d\mathbf{v} = 0$$

Sommando e sottraendo \mathbf{u} , il primo termine dà facilmente

$$\partial_t \left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right)$$

Per il secondo termine

$$\begin{aligned}\int |\mathbf{v}|^2 \mathbf{v} f &= \mathbf{u}(\rho |\mathbf{u}|^2 + 2\rho e) + \int |\mathbf{v}|^2 (\mathbf{v} - \mathbf{u}) f \\ &= \mathbf{u}(\rho |\mathbf{u}|^2 + 2\rho e) + \int |\mathbf{v} - \mathbf{u}|^2 (\mathbf{v} - \mathbf{u}) f + 2 \left(\int (\mathbf{v} - \mathbf{u}) \otimes (\mathbf{v} - \mathbf{u}) f \right) \mathbf{u} + |\mathbf{u}|^2 \int (\mathbf{v} - \mathbf{u}) f\end{aligned}$$

Nell'ultima riga, il quarto termine è nullo, il terzo è $2N\mathbf{u}$, il secondo lo denoto con $2\mathbf{q}$. L'equazione del bilancio dell'energia è dunque

$$\partial_t \left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) + \operatorname{div}_x \left(\left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) \mathbf{u} \right) = -\operatorname{div}_x(N\mathbf{u}) - \operatorname{div}_x \mathbf{q}.$$

Il termine di sinistra è un termine di corrente, a destra ci sono i termini dovuti al modificarsi della distribuzione in \mathbf{v} .

Le tre equazioni che abbiamo ottenuto hanno la stessa struttura delle equazioni che si ottengono in idrodinamica invocando le equazioni di Newton (con l'aggiunta del teorema di Cauchy sul tensore degli sforzi):

$$\begin{aligned}\partial_t \rho + \operatorname{div}(\rho\mathbf{u}) &= 0 \\ \partial_t(\rho\mathbf{u}) + \operatorname{div}(\rho\mathbf{u} \otimes \mathbf{u}) &= \operatorname{div} S \\ \partial_t \left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) + \operatorname{div} \left(\left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) \mathbf{u} \right) &= \operatorname{div}(S\mathbf{u}) - \operatorname{div} \mathbf{q}\end{aligned}$$

dove S è il tensore degli sforzi, \mathbf{q} il flusso di calore, che nel nostro caso sono espressi in termini di osservabili in \mathbf{v} .

Questo fatto non sorprende perché queste equazioni esprimono esattamente la conservazione “microscopica” di massa, quantità di moto ed energia, che sono le conseguenze dei principi della dinamica. In ogni caso si tratta di un sistema di equazioni che **non è chiuso**. Infatti, nei termini di destra compaiono quantità di un ordine superiore in \mathbf{v} rispetto a quelli di sinistra. In idrodinamica questo problema, viene affrontato modellizzando più o meno rigorosamente il tensore degli sforzi S e il flusso di calore \mathbf{q} , in funzione dei campi idrodinamici (ρ, \mathbf{u}, e) .

A partire dall'equazione di Boltzmann, c'è un modo rigoroso di derivare le equazioni dell'idrodinamica, attraverso il così detto **limite idrodinamico**, riuscendo a esprimere N e \mathbf{q} in termini dei campi idrodinamici. Si procede con uno *scaling* idrodinamico, cioè si osserva il sistema su scale temporali e spaziali grandi, di ordine $1/\varepsilon$, dove ε sarà un parametro piccolo:

$$f^\varepsilon(t, \mathbf{x}, \mathbf{v}) = f(t/\varepsilon, \mathbf{x}/\varepsilon, \mathbf{v})$$

Si noti che f^ε è valutata in variabili spazio-temporali di ordine 1 se f è valutata in variabili di ordine $1/\varepsilon$. Inoltre, non si mantiene la normalizzazione (che implicherebbe f^ε localmente divergente). La velocità non viene scalata, perché spazio e tempo scalano nello stesso modo. L'equazione di Boltzmann diventa

$$\partial_t f^\varepsilon + \mathbf{v} \cdot \partial_x f^\varepsilon = \frac{1}{\varepsilon} Q(f^\varepsilon, f^\varepsilon)$$

Mandare ε a zero vuol dire operare un cosiddetto “limite singolare”, infatti, portando ε all'altro membro, stiamo mandando a zero i termini con le derivate. Questo implica che $\lim_{\varepsilon \rightarrow 0} Q(f^\varepsilon, f^\varepsilon)$ deve essere 0. Questo può accadere se f^ε converge a una maxwelliana, per ogni valore di \mathbf{x} . I parametri della maxwelliana però possono dipendere da \mathbf{x} , ma in tal caso devono essere governati dalle equazioni dell'idrodinamica.

Lo spiego meglio: supponiamo che

$$f^\varepsilon(\mathbf{x}, \mathbf{v}, t) = M_{\rho, \mathbf{u}, T} + O(\varepsilon)$$

con ρ, \mathbf{u}, T dipendenti da \mathbf{x} e t e da determinare. All'ordine ε^{-1} l'equazione di Boltzmann è risolta, perché Q è nullo sulle maxwelliane. Nelle equazioni dell'idrodinamica, i membri di destra sono ora esplicitamente calcolabili, a meno di $O(\varepsilon)$, infatti, per esempio,

$$N = \int (\mathbf{v} - \mathbf{u}) \otimes (\mathbf{v} - \mathbf{u}) f^\varepsilon = \int (\mathbf{v} - \mathbf{u}) \otimes (\mathbf{v} - \mathbf{u}) M_{\rho, \mathbf{u}, T} + O(\varepsilon)$$

L'integrale è facilmente computabile, ricordando che

$$\frac{1}{(2\pi)^{3/2}} \int d\xi \xi \otimes \xi e^{-|\xi|^2/2} = \frac{1}{3} \mathbb{I}.$$

si ottiene

$$N = \rho T \mathbb{I} + O(\varepsilon).$$

Il calcolo di \mathbf{q} è più semplice, perché le maxwelliane sono pari nella variabile $\mathbf{v} - \mathbf{u}$, quindi \mathbf{q} è 0 al primo ordine in ε . Dunque nel limite $\varepsilon \rightarrow 0$ ci si aspetta di ottenere

$$\begin{aligned} \partial_t \rho + \operatorname{div}_x(\rho \mathbf{u}) &= 0 \\ \partial_t(\rho \mathbf{u}) + \operatorname{div}_x(\rho \mathbf{u} \otimes \mathbf{u}) &= -\partial_x(\rho T) \\ \partial_t \left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) + \operatorname{div}_x \left(\left(\frac{1}{2} \rho |\mathbf{u}|^2 + \rho e \right) \mathbf{u} \right) &= -\operatorname{div}_x(\rho T \mathbf{u}) \end{aligned}$$

dove $e = \frac{3}{2}T$. Queste sono le equazioni dell'idrodinamica di un **gas perfetto**, infatti nel bilancio della quantità di moto compare $-\partial_x(\rho T)$ che si identifica con il gradiente della pressione e la pressione è ρT , data dalla legge di stato dei gas perfetti (a meno di coefficienti). Nell'equazione del bilancio dell'energia si è annullato il termine di diffusione del calore. Attraverso qualche manipolazione che non faccio, si ottiene che la terza equazione equivale ad affermare che l'entropia specifica è trasportata dal campo \mathbf{u} . Dunque stiamo descrivendo un gas isoentropico, in cui, appunto, non c'è scambio di calore tra le parti. Questa derivazione può essere resa rigorosa, attraverso la cosiddetta **espansione di Hilbert** che permette di trovare una soluzione f^ε dell'equazione di Boltzmann intorno alla maxwelliana con i campi che obbediscono alle equazioni dell'idrodinamica dei gas perfetti.

A Disuguaglianza triangolare per la distanza di Wasserstein

Do una prova alternativa della disuguaglianza triangolare per la distanza di Wasserstein, che richiede un po' di teoria della misura. L'idea è la seguente: sia $\gamma_{12}(dx_1, dx_2)$ una distribuzione congiunta di μ_1 e μ_2 , e sia $\gamma_{23}(dx_2, dx_3)$ una distribuzione congiunta di μ_2 e μ_3 . Supponiamo ora di saper costruire $\sigma(dx_1, dx_2, dx_3)$ tale che la marginale ottenuta integrando in dx_1 sia γ_{23} e la marginale ottenuta integrando in dx_3 sia γ_{12} . Indico con $\sigma_{13}(dx_1, dx_3)$ la marginale che si ottiene integrando in dx_2 . Evidentemente σ_{13} è una distribuzione congiunta di μ_1 e μ_3 . Infatti, integrando in dx_3 e commutando l'ordine di integrazione con dx_2 si ottiene l'integrale in dx_2 di γ_{23} , che è appunto μ_1 . Analogamente si prova che integrando in dx_1 si ottiene μ_3 . Ne segue che

$$W_1(\mu_1, \mu_3) \leq \int |x_1 - x_3| \sigma(dx_1, dx_2, dx_3).$$

Sommando e sottraendo x_2 in $|x_1 - x_3|$ si ottiene

$$\begin{aligned} W_1(\mu_1, \mu_3) &\leq \int |x_1 - x_2| \sigma(dx_1, dx_2, dx_3) + \int |x_2 - x_3| \sigma(dx_1, dx_2, dx_3) \\ &= \int |x_1 - x_2| \gamma_{12}(dx_1, dx_2) + \int |x_2 - x_3| \gamma_{23}(dx_2, dx_3) \end{aligned}$$

Per l'arbitrarietà di γ_{12} e γ_{23} si ottiene la tesi.

Resta da costruire σ . Per capire come fare, suppongo che γ_{ij} ammettano densità regolare rispetto a Lebesgue, che continuo a indicare con γ_{ij} e assumo che anche μ_2 abbia densità regolare g_2 . Allora $\gamma_{12}(x_1, x_2)/g_2(x_2)$ è la densità di probabilità della variabile x_1 , condizionata a x_2 . Si noti che questa espressione ha senso se viene posta a 0 dove $g_2(x_2)$ è nulla, e che se $g_2(x_2)$ è zero, allora anche $\gamma_{12}(x_1, x_2) = 0$ per qualunque x_1 . Analogamente $\gamma_{23}(x_2, x_3)/g_2(x_2)$ è la densità di probabilità della variabile x_3 , condizionata a x_2 . La densità di probabilità di σ si ottiene come

$$g_2(x_2) \frac{\gamma_{12}(x_1, x_2)}{g_2(x_2)} \frac{\gamma_{23}(x_2, x_3)}{g_2(x_2)}.$$

Si verifichi per esercizio che questa densità ha le proprietà richieste.

Nel caso di misure non assolutamente continue, si procede nello stesso modo. L'idea è che $\gamma_{12}(dx_1, dx_2)$ si possa scrivere come

$$\gamma_{12}(dx_1, dx_2) = \mu_2(dx_2) \Gamma_{12}(dx_1 | x_2)$$

dove $\Gamma_{12}(dx_1|x_2)$ è la misura di probabilità di x_1 condizionata a x_2 . L'espressione precedente è la "disintegrazione" della misura congiunta, e l'esistenza μ_2 -q.o. della misura condizionata $\Gamma_{12}(dx_1|x_2)$ tale che valga la formula di disintegrazione è appunto una conseguenza del teorema di disintegrazione delle misure di probabilità.

La prova si conclude scegliendo

$$\sigma(dx_1, dx_2, dx_3) = \mu_2(dx_2)\Gamma_{12}(dx_1|x_2)\Gamma_{23}(dx_3|x_2).$$

B La derivazione classica della gerarchia

Assumeremo f^N continua, dunque quando $\mathbf{x}_j = \mathbf{x}_i + r\mathbf{n}$:

$$f_t^N(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{v}'_i, \dots, \mathbf{v}'_j, \dots) = f_t^N(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots)$$

dove $\mathbf{v}_i, \mathbf{v}_j$ sono le velocità entranti (dunque $(\mathbf{v}_i - \mathbf{v}_j) \cdot \mathbf{n} > 0$), e $\mathbf{v}'_i, \mathbf{v}'_j$ sono le velocità uscenti.

Per ottenere la gerarchia BBGKY bisogna integrare tenendo conto dei "buchi" nello spazio delle configurazioni. Per fare questi calcoli premetto un lemma che dà un'idea di come si fanno questi integrali.

Lemma B.1. *Sia g una funzione di una o due variabili, regolare e a supporto compatto. Allora*

$$\int_{|\mathbf{x}-\mathbf{z}|>r} \partial_z g(\mathbf{z}) d\mathbf{z} = - \int \mathbf{n} g(\mathbf{z}) \delta(|\mathbf{x} - \mathbf{z}| = r) d\mathbf{z} = -r^2 \int_S \mathbf{n} g(\mathbf{x} + r\mathbf{n}) d\mathbf{n} \quad (\text{B.1})$$

$$\begin{aligned} \int_{|\mathbf{x}-\mathbf{z}|>r} \partial_x g(\mathbf{x}, \mathbf{z}) d\mathbf{z} &= \partial_x \int_{|\mathbf{x}-\mathbf{z}|>r} g(\mathbf{x}, \mathbf{z}) d\mathbf{z} + \int \mathbf{n} g(\mathbf{x}, \mathbf{z}) \delta(|\mathbf{x} - \mathbf{z}| = r) d\mathbf{z} \\ &= \partial_x \int_{|\mathbf{x}-\mathbf{z}|>r} g(\mathbf{x}, \mathbf{z}) d\mathbf{z} + r^2 \int_S \mathbf{n} g(\mathbf{x}, \mathbf{x} + r\mathbf{n}) d\mathbf{n}. \end{aligned} \quad (\text{B.2})$$

dove $\mathbf{n} = (\mathbf{z} - \mathbf{x})/|\mathbf{z} - \mathbf{x}|$.

Dimostrazione. Per la prima equazione:

$$\int_{|\mathbf{x}-\mathbf{z}|>r} \partial_z g(\mathbf{z}) d\mathbf{z} = \int_{|\mathbf{y}|>r} \partial_y g(\mathbf{x} + \mathbf{y}) d\mathbf{y} = -r^2 \int_S \mathbf{n} g(\mathbf{x} + r\mathbf{n}) d\mathbf{n}$$

dove nel secondo passaggio ho usato che la normale esterna a S è opposta alla normale esterna al dominio di interazione, e che $r^2 d\mathbf{n}$ è la misura su $|\mathbf{x} - \mathbf{z}| = r$.

Per la seconda equazione valuto

$$\partial_x \int_{|\mathbf{x}-\mathbf{z}|>r} g(\mathbf{x}, \mathbf{z}) d\mathbf{z}$$

cambio variabile di integrazione: $\mathbf{z} = \mathbf{x} + \mathbf{y}$ e ottenendo

$$\partial_x \int_{|\mathbf{y}|>r} g(\mathbf{x}, \mathbf{x} + \mathbf{y}) d\mathbf{y} = \int_{|\mathbf{y}|>r} (\partial_x g(\mathbf{x}, \mathbf{x} + \mathbf{y}) + \partial_y g(\mathbf{x}, \mathbf{x} + \mathbf{y})) d\mathbf{y}.$$

Il primo termine è

$$\int_{|\mathbf{y}|>r} \partial_x g(\mathbf{x}, \mathbf{z}) d\mathbf{z},$$

il secondo è

$$\int_{|\mathbf{z}-\mathbf{x}|>r} \partial_z g(\mathbf{x}, \mathbf{z}) d\mathbf{z} = - \int_{|\mathbf{z}-\mathbf{x}|=r} \mathbf{n} g(\mathbf{x}, \mathbf{z}) \sigma(dz)$$

dove \mathbf{n} è la normale esterna in \mathbf{z} sul bordo della palla e dove ho usato il teorema della divergenza. Poiché la palla ha raggio r , posso scrivere l'ultimo integrale come

$$-r^2 \int_S \mathbf{n} g(\mathbf{x}, \mathbf{x} + r\mathbf{n}) d\mathbf{n}$$

dove S è la superficie della palla unitaria. □

Torno alla gerarchia. Per semplicità ricavo solo l'equazione per la f_1^N , integrando in $d\mathbf{z}_2^N$. La parte in ∂_t dà semplicemente $\partial_t f_1^N$. Per l'altro termine va distinto $i = 1$ da $i > 1$.

Caso $i = 1$

In questo caso compare la derivata ∂_{x_1} , con \mathbf{x}_1 che non è una variabile di integrazione:

$$\mathbf{v}_1 \cdot \int \partial_{x_1} f^N d\mathbf{x}_2 \dots d\mathbf{x}_N d\mathbf{v}_2 \dots d\mathbf{v}_N.$$

Da questo integrale deve venire il termine di flusso libero a una particella $\mathbf{v}_1 \cdot \partial_{x_1} f_1^N$, però non si può semplicemente permutare l'integrale con la derivata, perché il dominio di integrazione dipende da \mathbf{x}_1 . Procedo come per la prova di (B.2). Sia

$$\tilde{\Omega} = \{\mathbf{y}_2^N : \forall h, k \geq 2, k \neq h, |\mathbf{y}_h| > r, |\mathbf{y}_h - \mathbf{y}_k| > r\}.$$

Cambiando variabili di integrazione si ha:

$$\begin{aligned} \int \mathbf{v}_1 \cdot \partial_{x_1} f^N d\mathbf{x}_2^N d\mathbf{v}_2^N &= \int_{\tilde{\Omega}} \mathbf{v}_1 \cdot \partial_{x_1} f^N(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{y}_2, \dots, \mathbf{x}_1 + \mathbf{y}_n, \mathbf{v}_1^N) d\mathbf{y}_2^N d\mathbf{v}_2^N \\ &= \mathbf{v}_1 \cdot \partial_{x_1} f_1^N - \int_{\tilde{\Omega}} \mathbf{v}_1 \cdot \partial_{x_1} [f^N(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{y}_2, \dots, \mathbf{x}_1 + \mathbf{y}_n, \mathbf{v}_1^N)] d\mathbf{y}_2^N d\mathbf{v}_2^N \\ &= \mathbf{v}_1 \cdot \partial_{x_1} f_1^N - \sum_{j=2}^N \int_{\tilde{\Omega}} \mathbf{v}_1 \cdot \partial_{y_j} f^N(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{y}_2, \dots, \mathbf{x}_1 + \mathbf{y}_n, \mathbf{v}_1^N) d\mathbf{y}_2^N d\mathbf{v}_2^N. \end{aligned}$$

Integro la derivata in ∂_{y_j} con il teorema della divergenza, ricordando che fissate le altre \mathbf{y}_h , $h \neq i$, il bordo è composto da $|\mathbf{y}_j| = r$ e $|\mathbf{y}_j - \mathbf{y}_h| = r$. Inoltre noto che l'integrale nelle $d\mathbf{z}_k$, con $k \neq 1, j$ dà f_2^N :

$$\begin{aligned} \int_{\tilde{\Omega}} \mathbf{v}_1 \cdot \partial_{y_j} f^N(\mathbf{x}_1, \mathbf{x}_1 + \mathbf{y}_2, \dots, \mathbf{x}_1 + \mathbf{y}_n, \mathbf{v}_1^N) d\mathbf{y}_2^N d\mathbf{v}_2^N \\ = -r^2 \int d\mathbf{v}_j \int_S d\mathbf{n} \mathbf{v}_1 \cdot \mathbf{n} f_2^N(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_j) - \sum_{h \neq 1, j} \int \mathbf{v}_1 \cdot \mathbf{n}_{hj} \delta(|\mathbf{x}_j - \mathbf{x}_h| = r) f^N(\mathbf{z}_1^N) d\mathbf{z}_k^N \end{aligned}$$

Sommando anche in $j \geq 2$, la somma al secondo termine dà zero per simmetria, perché $\mathbf{n}_{hj} = -\mathbf{n}_{jh}$. Notando che la variabile di integrazione \mathbf{v}_j è una variabile muta, la somma in j del primo termine dà

$$(N-1)r^2 \int d\mathbf{v}_2 \int_S d\mathbf{n} \mathbf{v}_1 \cdot \mathbf{n} f_2^N(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2).$$

Caso $i \geq 2$

Per questi termini posso usare direttamente il teorema della divergenza nella variabile \mathbf{x}_i

$$\int \mathbf{v}_i \cdot \partial_{x_i} f^N d\mathbf{z}_2^N = - \sum_{j \neq i} \int \mathbf{v}_1 \cdot \mathbf{n}_{ij} f^N(\mathbf{z}_1^N) \delta(|\mathbf{x}_i - \mathbf{x}_j| = r) d\mathbf{z}_2^N$$

Considero il contributo per $j = 1$: integrando nelle \mathbf{z}_h con $h \neq i$ si ottiene un'espressione che dipende dalla distribuzione congiunta delle particelle 1 e i , che, per invarianza per permutazioni, è esattamente la $f_2^N(\mathbf{z}_1, \mathbf{z}_i)$. Cambiando il nome delle variabili di integrazione e sommando su i si ottiene

$$-r^2(N-1) \int_S d\mathbf{v}_2 \int \mathbf{v}_2 \cdot \mathbf{n} f_2^N(\mathbf{x}_1, \mathbf{x}_1 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2).$$

Ci sono ancora da considerare i contributi dati da $|\mathbf{x}_i - \mathbf{x}_k| = r$, con $k \neq 1, i$. Procedendo esattamente come sopra, integrando sulle variabili libere e rinominando $\mathbf{z}_i = \mathbf{z}_2$ e $\mathbf{z}_k = \mathbf{z}_3$ si ha

$$I = -r^2 \int_S d\mathbf{n} \int \mathbf{v}_2 \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) d\mathbf{x}_2 d\mathbf{v}_2 d\mathbf{v}_3$$

Cambiando variabile di integrazione $\mathbf{x}_2 \rightarrow \mathbf{x}_2 + r\mathbf{n}$, scambiando in f_3^N le variabili 2 e 3, e cambiando $\mathbf{n} \rightarrow -\mathbf{n}$ si ottiene lo stesso integrale con \mathbf{v}_3 e il segno cambiato, dunque $I = -\frac{1}{2}r^2 \int_S d\mathbf{n} J$ dove

$$J = \int (\mathbf{v}_2 - \mathbf{v}_3) \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) d\mathbf{x}_2 d\mathbf{v}_2 d\mathbf{v}_3$$

Mostro che J è nullo:

$$\begin{aligned} J &= \int (\mathbf{v}_2 - \mathbf{v}_3) \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}'_2, \mathbf{v}'_3) d\mathbf{x}_2 d\mathbf{v}_2 d\mathbf{v}_3 \\ &= - \int (\mathbf{v}'_2 - \mathbf{v}'_3) \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}'_2, \mathbf{v}'_3) d\mathbf{x}_2 d\mathbf{v}_2 d\mathbf{v}_3 \\ &= - \int (\mathbf{v}'_2 - \mathbf{v}'_3) \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}'_2, \mathbf{v}'_3) d\mathbf{x}_2 d\mathbf{v}'_2 d\mathbf{v}'_3 \\ &= - \int (\mathbf{v}_2 - \mathbf{v}_3) \cdot \mathbf{n} f_3^N(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2 + r\mathbf{n}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) d\mathbf{x}_2 d\mathbf{v}_2 d\mathbf{v}_3 = -J \end{aligned}$$

dove nel primo passaggio ho usato che f_3^N è continua nell'urto, nel secondo che $\mathbf{n} \cdot (\mathbf{v}_2 - \mathbf{v}_3)$ cambia segno nell'urto, nel terzo che $d\mathbf{v}_2 d\mathbf{v}_3 = d\mathbf{v}'_2 d\mathbf{v}'_3$, e infine nel quarto ho semplicemente rinominato le variabili. Poiché $J = -J$ l'integrale è nullo.

C La trasformata di Legendre

Anche se non è indispensabile, per mostrare il legame tra formalismo hamiltoniano e formalismo lagrangiano ho introdotto la trasformata di Legendre, perché è uno strumento che ha una notevole rilevanza in vari argomenti della fisica-matematica (in particolare in meccanica statistica, e più in generale nei metodi probabilistici). In questo corso mi servirà ancora per alcuni risultati delle teorie cinetiche, per questo gli dedico un'appendice dettagliata, in cui mostro anche la forte connessione che ha con la teoria delle funzioni convesse.

C.1 Funzioni convesse

Sia f una funzione da \mathbb{R}^d a valori in $\overline{\mathbb{R}} := \mathbb{R} \cup +\infty$; si chiama **propria** se è finita in almeno un punto, e si chiama **dominio** di f l'insieme

$$\text{dom}(f) := \{\mathbf{x} \mid f(\mathbf{x}) < +\infty\}.$$

Ricordo che una funzione è **convessa** se $\forall \lambda \in [0, 1]$ e $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$:

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y}).$$

È preferibile pensare le funzioni convesse definite sempre su tutto lo spazio a valori in $\overline{\mathbb{R}}$, dichiarandole infinite fuori dal loro dominio di definizione,

È di immediata verifica la validità della seguente proposizione.

Proposizione C.1. *Se f è convessa, allora $\text{dom}(f)$ è un convesso.*

Prova per esercizio.

Teorema C.1 (Funzioni convesse in \mathbb{R}^d). *Sia $f : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ convessa e sia A un aperto di $\text{dom}(f)$. Allora f è localmente lipschitziana in A .*

Per una guida alla prova vedi il foglio di esercizi es02-bis.

Tratterò funzioni convesse e trasformata di Legendre in un contesto più astratto. Sia \mathcal{B} uno spazio di Banach, ricordo che \mathcal{B}^* (il suo duale topologico) è lo spazio dei funzionali lineari e continui (equivalentemente funzionali limitati) su \mathcal{B} . Userò una notazione astratta, introdotta da Dirac per spazi di Hilbert in relazione alla meccanica quantistica, e che in analisi viene chiamata “pairing”. Sia $\varphi \in \mathcal{B}$ e sia $\mu \in \mathcal{B}^*$. L'azione di μ su φ si indica come $\langle \mu, \varphi \rangle$. Questa notazione “bra-ket” serve a tenere distinti con chiarezza gli elementi di \mathcal{B} e \mathcal{B}^* anche nei casi in cui sono indistinguibili, come accade per spazi euclidei o spazi di Hilbert; per esempio in \mathbb{R}^d si può pensare a φ come un vettore colonna, e a $\mu \in (\mathbb{R}^d)^*$ come a un vettore riga. Ricordo anche che l'operazione $\mu \rightarrow \langle \mu, \varphi \rangle$ definisce un funzionale lineare continuo su \mathcal{B}^* , e dunque \mathcal{B} si identifica con un sottospazio di \mathcal{B}^{**} .

Sia $F : \mathcal{B} \rightarrow \overline{\mathbb{R}}$, una funzione propria; si chiama **epigrafico** di F il sottoinsieme di $\mathcal{B} \times \mathbb{R}$ dato da

$$\text{epi}(F) = \{(\varphi, a) \mid a \geq F(\varphi)\},$$

(si noti che se $(\varphi, a) \in \text{epi}(F)$ allora $\varphi \in \text{dom}(F)$). Per ispezione diretta si mostra facilmente che vale la seguente proposizione.

Proposizione C.2 (Funzioni convesse e epigrafico).

F è convessa se e solo se $\text{epi}(F)$ è un convesso

Dimostrazione da inserire, ma fatta a lezione.

C.2 Funzioni semi continue inferiori

In analisi, il “teorema ponte” in \mathbb{R}^d assicura l'equivalenza tra continuità e continuità per successioni. Questa equivalenza non è in generale vera.

Do qualche definizione a questo proposito. Un sottoinsieme C di uno spazio topologico (\mathcal{X}, τ) (\mathcal{X} è un insieme, τ è la topologia) è **sequenzialmente chiuso** se e solo ogni successione

convergente di elementi di C converge a un elemento di C . È facile mostrare che i chiusi sono chiusi per successioni, infatti se C è chiuso, $x_n \in C$, e $x_n \rightarrow x$, se $x \notin C$, x_n non può convergere a x perché il complementare C^c di C è un aperto che contiene x e $x_n \notin C^c$ per ogni n .

In generale non è vero il viceversa. Uno spazio topologico è detto **sequenziale** se i suoi chiusi sono esattamente i sottoinsiemi chiusi per successione.

Sia $f : \mathcal{X} \rightarrow \mathcal{Y}$; f è detta **sequenzialmente continua** se e solo se per ogni successione $x_n \in \mathcal{X}$ convergente si ha $\lim_n f(x_n) = f(\lim_n x_n)$.

Proposizione C.3. *Se \mathcal{X} è sequenziale, f è continua se e solo se è sequenzialmente continua.*

Dimostrazione. È un fatto generale che ogni funzione continua è sequenzialmente continua (prova per esercizio). Mostro il viceversa. Considero la controimmagine $f^{-1}(K)$ di un chiuso K di \mathcal{Y} . Sia x_n una sequenza, convergente a un qualche x , di elementi di $f^{-1}(K)$. Allora $f(x_n) \in K$, e per ipotesi $f(x_n) \rightarrow f(x)$. Ma K è un chiuso dunque è sequenzialmente chiuso, e quindi $x \in f^{-1}(K)$. Ne segue che $f^{-1}(K)$ è sequenzialmente chiuso, e dunque, per l'ipotesi di sequenzialità di \mathcal{X} , è chiuso. Ho così dimostrato che la controimmagine di un chiuso è un chiuso, cioè che f è continua. \square

Proposizione C.4. *Gli spazi metrici sono sequenziali.*

Dimostrazione. Sia C un sottoinsieme sequenzialmente chiuso. Suppongo per assurdo che non sia chiuso, cioè che C^c non sia aperto. Ne segue che esiste $x \in C^c$ tale che ogni aperto che contiene x ha intersezione non vuota con C . Scegliendo come aperti le palle $B_{1/n}(x)$ si trova una successione $x_n \in B_{1/n}(x) \cap C$, che per come è definita converge a $x \notin C$, contro l'ipotesi che C sia sequenzialmente chiuso. Si osservi che quello che realmente è servito è l'esistenza di una base numerabile di intorni; questa dimostrazione si estende infatti agli spazi che soddisfano il primo assioma di numerabilità. \square

Ricordo che la topologia della semicontinuità inferiore su \mathbb{R} è quella che ha come aperti le sole semirette $(b, +\infty)$, con $b \in \mathbb{R}$. Una funzione da uno spazio topologico \mathcal{X} a valori in \mathbb{R} è **semicontinua inferiore** (scrittura che abbrevierò in "s.c.i.") se la controimmagine di $(b, +\infty)$ è aperta per ogni $b \in \mathbb{R}$, cioè se è continua rispetto alla topologia della semicontinuità inferiore. Se lo spazio \mathcal{X} è sequenziale, f è s.c.i. se e solo se è sequenzialmente s.c.i.. Provare per esercizio che la s.c.i. sequenziale equivale al fatto che per ogni successione convergente $x_n \rightarrow x$ vale $f(x) \leq \underline{\lim}_n f(x_n)$.

Proposizione C.5 (Funzioni s.c.i. e epigrafico).

$F : \mathcal{B} \rightarrow \overline{\mathbb{R}}$ è s.c.i. se e solo se $\text{epi}(F)$ è chiuso.

Dimostrazione. \mathcal{B} è uno spazio metrico, dunque è sufficiente mostrare che F è sequenzialmente s.c.i. se e solo se $\text{epi}(F)$ è sequenzialmente chiuso.

Sia $\text{epi}(F)$ chiuso, e sia $\varphi_n \rightarrow \varphi$, con $F(\varphi_n) < +\infty$. Passando a un'opportuna sottosequenza, posso assumere che $F(\varphi_n)$ converga al valore $\underline{\lim} \varphi_n$. Poiché $(\varphi_n, F(\varphi_n))$ è nell'epigrafico che è chiuso, $(\varphi, \underline{\lim} F(\varphi_n))$ è nell'epigrafico e dunque $F(\varphi) \leq \underline{\lim} F(\varphi_n)$.

Viceversa sia $\text{epi}(F) \ni (\varphi_n, a_n) \rightarrow (\varphi, a)$. Passando al limite la relazione $F(\varphi_n) \leq a_n$, per semicontinuità inferiore si ottiene $F(\varphi) \leq a$ e dunque $(\varphi, a) \in \text{epi}(F)$. \square

Osservo che il vero contenuto di questa proposizione è proprio il fatto che la s.c.i. sequenziale è equivalente alla chiusura sequenziale dell'epigrafico; tornerò su questa osservazione a fine del paragrafo, e questo fatto è vero qualunque sia la topologia che consideriamo su \mathcal{B} .

Convessità e s.c.i. si conservano nel passaggio al sup, come è facile dimostrare.

Proposizione C.6. *Sia I un insieme di indici (di cardinalità qualunque), e siano date delle funzioni $F_i : \mathcal{B} \rightarrow \overline{\mathbb{R}}$. Sia*

$$F(\varphi) = \sup_{i \in I} F_i(\varphi).$$

Allora

$$\text{epi}(F) = \bigcap_{i \in I} \text{epi}(F_i).$$

Come conseguenza, se le F_i sono s.c.i. allora F è s.c.i. (perché l'intersezione di chiusi è un chiuso), e se le F_i sono convesse allora F è convessa (perché l'intersezione di convessi è un convesso).

Dimostrazione per esercizio.

C.3 La trasformata di Legendre

Definisco ora la trasformata di Legendre. Data F propria,

$$F^*(\mu) = \sup_{\varphi \in \mathcal{B}} (\langle \mu, \varphi \rangle - F(\varphi)),$$

che è un funzionale su \mathcal{B}^* . Per la proposizione precedente, $F^*(\mu)$ è convesso e s.c.i.. Analogamente, data $F^*(\mu)$ in \mathcal{B}^* propria, si definisce $F^{**} : \mathcal{B}^* \rightarrow \overline{\mathbb{R}}$ come

$$F^{**}(\varphi) = \sup_{\mu \in \mathcal{B}^*} (\langle \mu, \varphi \rangle - F^*(\mu))$$

(Si noti che questa definizione non è esattamente come quella di prima, perché F^{**} viene definita su \mathcal{B} e non su \mathcal{B}^{**}). Dalla definizione segue la seguente importante disuguaglianza.

Teorema C.2 (Disuguaglianza di Young). *Per ogni $\varphi \in \mathcal{B}$ e $\mu \in \mathcal{B}^*$ si ha*

$$\begin{aligned} \langle \mu, \varphi \rangle &\leq F(\varphi) + F^*(\mu) \\ \langle \mu, \varphi \rangle &\leq F^{**}(\varphi) + F^*(\mu) \end{aligned}$$

Indago ora sulla relazione tra F^{**} e F . Osservo immediatamente che dalla disuguaglianza di Young si ottiene che $F^{**} \leq F$.

Mi serviranno due lemmi, la cui prova lascio per esercizio.

Lemma C.1.

Se $F \leq G$ allora $F^ \geq G^*$.*

Se $F^ \leq G^*$ allora $F^{**} \geq G^{**}$.*

Lemma C.2.

Sia $\mu_0 \in \mathcal{B}^$ e $b_0 \in \mathbb{R}$, e sia $L_{\mu_0, b_0}(\varphi) = \langle \mu_0, \varphi \rangle + b_0$. Allora*

$$L_{\mu_0, b_0}^*(\mu) = \begin{cases} -b_0 & \text{se } \mu = \mu_0 \\ +\infty & \text{altrimenti} \end{cases}$$

Inoltre

$$L_{\mu_0, b_0}^{**} = L_{\mu_0, b_0}.$$

Ricordo ora la definizione di inviluppo convesso. Sia $F : \mathcal{B} \rightarrow \overline{\mathbb{R}}$ propria. Il suo inviluppo convesso è

$$F^c(\varphi) = \sup_{(\mu, b) \in A_F} \langle \mu, \varphi \rangle + b,$$

dove

$$A_F = \{(\mu, b) \in \mathcal{B}^* \times \mathbb{R} : \forall \psi \langle \mu, \psi \rangle + b \leq F(\psi)\}.$$

Teorema C.3. $F^c = F^{**}$

Dimostrazione. Suppongo che A_F è non vuoto, se no F^c non è definita (si noti che se A_F è vuoto allora F^* coincide con $+\infty$, dunque neanche F^{**} è definita, oppure F^{**} si può considerare $-\infty$, come anche F^c). Sia $(\mu, b) \in A_F$. Per ogni φ

$$\langle \mu, \psi \rangle + b \leq F(\psi).$$

Passando alla doppia trasformata entrambi i membri, usando i due lemmi, ottengo

$$\langle \mu, \psi \rangle + b \leq F^{**}(\psi),$$

dunque, passando al sup su A_F , ottengo che $F^c \leq F^{**}$.

Per la disuguaglianza di Young, per ogni $\mu \in \mathcal{B}^*$, $(\mu, -F^*(\mu)) \in A_F$, dunque

$$F^c(\varphi) \geq \sup_{\mu \in \mathcal{B}^*} (\langle \mu, \varphi \rangle - F^*(\mu)) = F^{**}(\varphi).$$

□

Teorema C.4 (Fenchel-Moreau). F è convessa e s.c.i. se e solo se $F = F^{**}$

Dimostrazione. Abbiamo già provato che F^{**} è convessa e s.c.i.. Proviamo il viceversa. Sia F convessa e s.c.i.; il suo epigrafico è dunque un convesso chiuso. Supponiamo che esista ψ tale che $F^{**}(\psi) < F(\psi)$. Per la seconda forma geometrica del teorema di Hahn-Banach esiste un funzionale lineare continuo su $\mathcal{B} \times \mathbb{R}$ che separa il punto $(\psi, F^{**}(\psi))$ dall'epigrafico, cioè esistono $\mu \in \mathcal{B}^*$, $k, c \in \mathbb{R}$, $\varepsilon > 0$ tale che

$$\begin{aligned} -\langle \mu, \psi \rangle + kF^{**}(\psi) &\leq c - \varepsilon \\ -\langle \mu, \varphi \rangle + ka &\geq c + \varepsilon, \quad \forall (\varphi, a) \in \text{epi}(F). \end{aligned}$$

Si noti che deve essere $k \geq 0$, se no la seconda relazione porterebbe ad un assurdo, potendo prendere a arbitrariamente grande. Ora si devono distinguere vari casi.

Caso $k > 0$.

Scegliendo $a = F(\varphi)$, con $\varphi \in \text{dom}(F)$, nella seconda relazione, si ha che per ogni $\varphi \in \text{dom}(F)$ vale $-\langle \mu, \varphi \rangle + kF(\varphi) > c + \varepsilon$. Poiché $k > 0$, si ottiene

$$\langle \mu/k, \varphi \rangle + (c + \varepsilon)/k < F(\varphi)$$

Questa relazione è vera anche se φ non è nel dominio, dunque posso passare alla doppia trasformata e ottenere

$$\langle \mu/k, \varphi \rangle + (c + \varepsilon)/k \leq F^{**}(\varphi)$$

per ogni φ . Ne segue

$$-\langle \mu, \varphi \rangle + kF^{**}(\varphi) \geq c + \varepsilon$$

che per $\varphi = \psi$ è falso.

Caso $k = 0$ e $F \geq 0$. Si ha $\langle \mu, \psi \rangle \leq c - \varepsilon$ e $\langle \mu, \phi \rangle \geq c + \varepsilon$ per ogni $\varphi \in \text{dom}(F)$. Purtroppo non è noto se $\psi \in \text{dom}(F)$ dunque per ottenere una contraddizione si deve lavorare un po'. Poichè $F^{**}(\psi)$ è finito, esiste δ sufficientemente piccolo tale che $-\langle \mu, \psi \rangle + \delta F^{**}(\psi) \leq c$. Inoltre, per l'ipotesi di positività di F , per ogni $\varphi \in \text{dom}(F)$ si ha $-\langle \mu, \varphi \rangle + \delta F(\varphi) \geq c + \varepsilon$. Ci siamo così ricondotti al caso $k > 0$ già trattato.

Caso generale. I due casi precedenti concludono il teorema nel caso di F positiva. Sia ora F generale. Per ipotesi F è finita in almeno un punto φ_0 , dunque $(\varphi_0, \lambda) \notin \text{epi}(F)$ per $\lambda < F(\varphi_0)$. Sempre per Hahn-Banach, esiste $(\bar{\mu}, k) \in \mathcal{B}^* \times \mathbb{R}$ e $c \in \mathbb{R}, \varepsilon > 0$ tali che

$$\begin{aligned} -\langle \bar{\mu}, \varphi_0 \rangle + k\lambda &\leq c - \varepsilon \\ -\langle \bar{\mu}, \varphi \rangle + kF(\varphi) &\geq c + \varepsilon, \quad \forall \varphi \in \text{dom}(F) \end{aligned}$$

Poichè $\varphi_0 \in \text{dom}(F)$, $kF(\varphi_0) > k\lambda$, dunque $k > 0$, perché $\lambda < F(\varphi_0)$. Abbiamo così trovato $\mu_0 = \bar{\mu}/k$ e $b_0 = (c + \varepsilon)/k$, tale che

$$L_{\mu_0, b_0}(\varphi) \leq F(\varphi).$$

La funzione $F(\varphi) - L_{\mu_0, b_0}(\varphi)$, che è convessa e s.c.i., è positiva e dunque

$$F - L_{\mu_0, b_0} = (F - L_{\mu_0, b_0})^{**} = (F - L_{\mu_0, b_0})^c.$$

Troviamo l'involuppo convesso:

$$(F - L_{\mu_0, b_0})^c(\varphi) = \sup_{(\mu, b) \in A_{F - L_{\mu_0, b_0}}} \langle \mu, \varphi \rangle + b = -L_{\mu_0, b_0}(\varphi) + \sup_{(\mu, b) \in A_{F - L_{\mu_0, b_0}}} \langle (\mu + \mu_0), \varphi \rangle + b + b_0$$

È immediato verificare che $(\mu, b) \in A_{F - L_{\mu_0, b_0}}$ se e solo se $(\mu + \mu_0, b + b_0) \in A_F$; quindi

$$(F - L_{\mu_0, b_0})^c(\varphi) = -L_{\mu_0, b_0}(\varphi) + F^c = -L_{\mu_0, b_0}(\varphi) + F^{**}.$$

□

Sarà utile conoscere delle proprietà della trasformata di Legendre rispetto alla convergenza debole. Ricordo che su \mathcal{B} si definisce la topologia debole come la minima topologia che rende continua l'azione degli elementi di \mathcal{B}^* su \mathcal{B} . Questa topologia contiene come aperti tutte le controimmagini di intervalli aperti di \mathbb{R} dell'applicazione lineare $\varphi \rightarrow \langle \mu, \varphi \rangle \in \mathbb{R}$, al variare di $\mu \in \mathcal{B}^*$. Dunque per ogni a reale e per ogni μ , i semispazi "aperti" $\{\varphi | \langle \mu, \varphi \rangle < a\}$ e $\{\varphi | \langle \mu, \varphi \rangle > a\}$ sono aperti della topologia debole. Una base di aperti della topologia debole è data da intersezioni finite di questi semispazi. Si noti che $\{\varphi | \langle \mu, \varphi \rangle \leq a\}$ e $\{\varphi | \langle \mu, \varphi \rangle \geq a\}$ sono invece insiemi chiusi della topologia debole. Una sequenza $\varphi_n \in \mathcal{B}$ converge a $\varphi \in \mathcal{B}$ in questa topologia se e solo se per ogni $\mu \in \mathcal{B}^*$ si ha

$$\langle \mu, \varphi_n \rangle \rightarrow \langle \mu, \varphi \rangle.$$

La convergenza debole si indica con il simbolo $\varphi_n \rightharpoonup \varphi$. Ricordo che la convergenza nella norma di \mathcal{B} implica la convergenza debole, ma non vale il viceversa.

Proposizione C.7. *Sia C un convesso di un Banach. Allora C è chiuso per la topologia forte se e solo se è sequenzialmente chiuso per la topologia debole.*

Dimostrazione. Se C non è chiuso per la topologia forte, esiste una successione in C che converge fortemente (e dunque debolmente) a un elemento fuori di C , che dunque non è neanche sequenzialmente debolmente chiuso. Questo prova, in generale, che se C è sequenzialmente chiuso nella topologia debole, è chiuso nella topologia forte. Per i convessi vale anche il viceversa. Sia C è un convesso chiuso e supponiamo che non sia sequenzialmente debolmente chiuso. Dunque esiste $C \ni \psi_n \rightharpoonup \psi \notin C$. Per Hahn-Banach, esiste un $\mu \in \mathcal{B}^*$ e un $c \in \mathbb{R}$ tale che $\langle \mu, \psi \rangle < c$, mentre per tutti gli elementi $\varphi \in C$ si ha che $\langle \mu, \varphi \rangle > c$. In particolare, $\langle \mu, \psi_n \rangle > c$. Ma allora è impossibile che $\langle \mu, \psi_n \rangle$ converga a $\langle \mu, \psi \rangle$, in contrasto con l'ipotesi di convergenza debole sequenziale. \square

Come conseguenza immediata si ottiene il seguente risultato.

Teorema C.5. F è convessa e sequenzialmente debolmente s.c.i. se e solo se $F = F^{**}$

Dimostrazione. F è convessa e sequenzialmente debolmente s.c.i. se e solo se $\text{epi}(F)$ è convesso e sequenzialmente debolmente chiuso, come abbiamo provato precedentemente. Per la proposizione precedente, questo fatto è vero se e solo se $\text{epi}(F)$ è convesso e chiuso per la topologia indotta dalla metrica, e quindi se e solo se F è convessa e s.c.i. Ma quest'ultima asserzione è vera se e solo se $F = F^{**}$. \square

Osservazione finale: per provare questo teorema serve quello precedente per lo spazio $\mathcal{B} \times \mathbb{R}$. Va verificato che $(\varphi_n, a_n) \rightharpoonup (\varphi, a)$ se e solo se $\varphi_n \rightharpoonup \varphi$ e $a_n \rightarrow a$. Lascio ai lettori i restanti dettagli.