

ALMA MATER STUDIORUM · UNIVERSITÁ DI BOLOGNA

DIPARTIMENTO DI MATEMATICA
Corso di Laurea (Triennale) in Matematica

Distanza di Wasserstein
e
propagazione del Chaos

Tesi di Laurea in Probabilità

Relatore:
Chiar.mo Prof.
PASCUCCI ANDREA

Presentata da:
BRUNI GIOVANNI

Anno Accademico 2024-2025

A tutto ciò che é Famiglia.

Introduzione

Negli ultimi decenni, l'interesse per i problemi di ottimizzazione del trasporto ha avuto una crescita significativa, grazie alle sue applicazioni in diversi campi. In particolare, ha contribuito alla nascita di oggetti matematici moderni come la distanza di Wasserstein, che ha assunto un ruolo centrale nello studio di fenomeni di tipo probabilistico e nell'analisi della convergenza di misure.

Un'applicazione rilevante di questi strumenti é lo studio della *propagazione del Chaos*, concetto introdotto da Kac e poi sviluppato da McKean negli anni '50,'60 del Novecento, per poter analizzare ed esprimere il comportamento di un grande numero di particelle interagenti tra loro. Il loro studio ha permesso di osservare che, al crescere del numero N di particelle (ovvero quando N tende a infinito), il sistema tenda a comportarsi in modo "indipendente". La distanza di Wasserstein é uno dei concetti che ha sicuramente giocato un ruolo fondamentale nello sviluppo della loro teoria.

Lo scopo di questa tesi é quello di presentare in modo rigoroso i principali concetti relativi al problema di ottimizzazione del trasporto e alla distanza di Wasserstein, per poi applicarli allo studio della propagazione del Chaos in grandi sistemi di particelle interagenti.

Nel primo capitolo verrà introdotto il problema di ottimizzazione del trasporto, la sua prima formulazione sviluppata da Kantorovich e la sua espressione duale. Il secondo capitolo é dedicato alla definizione e allo studio della distanza di Wasserstein, concentrandosi in particolare sulle varie formulazioni che acquisisce sotto determinate ipotesi, sulle peculiari proprietà dello spa-

zio $(\mathcal{P}_p(\mathcal{X}), W_p)$ e sulle qualità che le permettono alla distanza W_p di essere molto utilizzata in vari ambiti della matematica applicata. Infine, il terzo capitolo tratta i processi di diffusione non lineare di McKean e il loro contributo allo sviluppo della teoria della propagazione del Chaos, mostrando come quest'ultima viene tutt'oggi utilizzata in molti ambiti della matematica e non solo.

Indice

Introduzione	i
1 Problema di ottimizzazione del trasporto	1
1.1 Introduzione al problema di Kantorovich	1
1.2 Dualità di Kantorovich	3
2 Distanza di Wasserstein	7
2.1 Definizione e primi risultati	7
2.2 Distanza 1-Wasserstein tra distribuzioni reali	14
2.3 Completezza e separabilità di $(\mathcal{P}_p(\mathcal{X}), W_p)$	19
2.4 Applicazioni	27
3 Propagazione del Chaos	29
3.1 Processi di diffusione non lineare di McKean	30
3.2 Convergenza del sistema di diffusione	36
3.3 Applicazioni	39
Bibliografia	43

Capitolo 1

Problema di ottimizzazione del trasporto

Nel seguente capitolo tratteremo del problema di ottimizzazione del trasporto di Kantorovich, che porrà le basi su cui si fonderà tutto il resto della trattazione.

1.1 Introduzione al problema di Kantorovich

Il problema di ottimizzazione del trasporto introdotto da Kantorovich negli anni '40 del Novecento é intuitivamente descrivibile attraverso il seguente semplice modello.

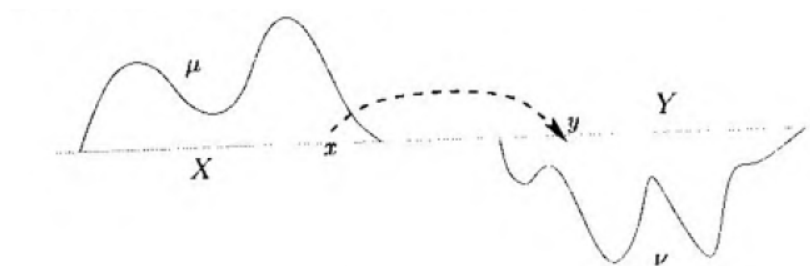


Figura 1.1: *Il problema di trasporto.* Crediti: Figura 1 [Vil03]

Assumiamo di dover riempire completamente una cavità con una pila di

sabbia dello stesso volume, considerando per semplicità che abbiano massa unitaria. Modelliamo ora la pila e la cavità introducendo due misure di probabilità μ, ν definite rispettivamente su degli spazi misurabili $\mathcal{X}$ e $\mathcal{Y}$. Per ogni A, B sottoinsiemi misurabili rispettivamente di $\mathcal{X}, \mathcal{Y}$, $\mu(A)$ rappresenta la misura di quanta sabbia si trova in A , mentre $\nu(B)$ rappresenta quanta sabbia può essere inserita in B . Muovere la sabbia necessita ovviamente di uno sforzo che può essere considerato come una funzione di costo c , misurabile e definita su $\mathcal{X} \times \mathcal{Y}$; intuitivamente $c(x, y)$ ci descrive quanto costa trasportare un'unità di massa da x a y .

Il problema è quindi: *Come si realizza il trasporto a costo minimo?*

Per cominciare si associa un piano di trasporto ad una misura di probabilità π sullo spazio $\mathcal{X} \times \mathcal{Y}$. Affinché un piano di trasporto $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ sia ammissibile, è necessario che tutta la massa presa da un punto x coincida con $d\mu(x)$ e che tutta la massa trasferita ad un punto y coincida $d\nu(y)$, più rigorosamente che

$$\pi(A \times \mathcal{Y}) = \mu(A), \quad \pi(\mathcal{X} \times B) = \nu(B), \quad (1.1)$$

per ogni sottoinsieme misurabile A di $\mathcal{X}$ e B di $\mathcal{Y}$.

Definizione 1.1.1 ($\Pi(\mu, \nu)$). Diciamo che le misure di probabilità π che soddisfano (1.1) hanno *marginali* μ e ν e sono piani di trasporto ammissibili. Definiamo l'insieme di tali misure di probabilità come

$$\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \mid (1.1) \text{ vale per ogni } A \subseteq \mathcal{X}, B \subseteq \mathcal{Y} \text{ misurabili}\}.$$

Ora abbiamo una chiara definizione matematica del problema di Kantorovich, ovvero minimizzare

$$I[\pi] := \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) \quad \text{con } \pi \in \Pi(\mu, \nu).$$

1.2 Dualità di Kantorovich

Nel corso della trattazione ci focalizzeremo su spazi descritti come nella seguente definizione.

Definizione 1.2.1 (Spazio Polacco, σ -algebra di Borel e $\mathcal{P}(\mathcal{X})$). $\mathcal{X}$ si dice spazio Polacco se $\mathcal{X}$ é uno spazio metrico completo e separabile su cui definiamo la σ -algebra di Borel come la σ -algebra generata dagli aperti. Definiamo, invece, $\mathcal{P}(\mathcal{X})$ come l'insieme delle misure di probabilità Boreliane su $\mathcal{X}$.

In questo contesto é possibile dare una formulazione *duale* del problema di minimizzazione introdotto da Kantorovich nel 1942 e si ottiene cosí il seguente risultato.

Teorema 1.2.1 (Dualità di Kantorovich). Siano $\mathcal{X}, \mathcal{Y}$ spazi Polacchi, sia $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ e $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ una funzione continua. Per ogni $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ e $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, definiamo

$$I[\pi] := \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y), \quad J(\varphi, \psi) := \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu.$$

Definito, invece:

$\Phi_c := \{(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu) \mid \varphi(x) + \psi(y) \leq c(x, y) \text{ } \mu\text{-}q.o.\text{ in } \mathcal{X} \text{ e } \nu\text{-}q.o.\text{ in } \mathcal{Y}\}$, allora

$$\inf_{\Pi(\mu, \nu)} I[\pi] = \sup_{\Phi_c} J(\varphi, \psi). \quad (1.2)$$

Diamo prima la dimostrazione di un risultato piú semplice riguardante una parte della dualità di Kantorovich.

Proposizione 1.2.1. Sotto le ipotesi del teorema precedente (1.2.1) valgono le seguenti disuguaglianze:

$$\sup_{\Phi_c \cap C_b} J(\varphi, \psi) \leq \sup_{\Phi_c} J(\varphi, \psi) \leq \inf_{\Pi(\mu, \nu)} I[\pi].$$

Dimostrazione. La disuguaglianza a sinistra é banale e segue dal fatto che $C_b(\mathcal{X}) \times C_b(\mathcal{Y}) \subseteq L^1(d\mu) \times L^1(d\nu)$.

Dimostriamo invece la disuguaglianza di destra.

Siano $(\varphi, \psi) \in \Phi_c$ e sia $\pi \in \Pi(\mu, \nu)$, allora per definizione di Π si ha che:

$$J(\varphi, \psi) = \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu = \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y).$$

Siano ora N_x, N_y insiemi misurabili rispettivamente di $\mathcal{X}$ e $\mathcal{Y}$ tali che $\mu[N_x] = 0, \nu[N_y] = 0$ e $\varphi(x) + \psi(y) \leq c(x, y) \quad \forall (x, y) \in N_x^c \times N_y^c$. Ora, per definizione di π notiamo che $\Pi[\mathcal{X} \times N_y] = \nu[N_y] = 0$ e $\Pi[N_x \times \mathcal{Y}] = \mu[N_x] = 0$ per cui $\pi[(N_x^c \times N_y^c)^c] = 0$. Allora:

$$\begin{aligned} & \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \\ &= \int_{(N_x^c \times N_y^c)} [\varphi(x) + \psi(y)] d\pi(x, y) + \int_{(N_x^c \times N_y^c)^c} [\varphi(x) + \psi(y)] d\pi(x, y) \\ &\leq \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) = I[\pi]. \end{aligned}$$

Si ottiene quindi la tesi passando al sup a sinistra e all'inf a destra. $\square$

Osservazione 1.2.1. Grazie alla proposizione appena dimostrata di può osservare che

$$\sup_{\Phi_c \cup C_b} J = \inf_{\Pi} I \quad \Rightarrow \quad \sup_{\Phi_c} J = \inf_{\Pi} I.$$

Diamo ora una dimostrazione formale della formula di dualità (1.2).

Dimostrazione. Notiamo che vale la seguente uguaglianza:

$$\inf_{\pi \in \Pi(\mu, \nu)} I[\pi] = \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \left(I[\pi] + \begin{cases} 0 & \text{se } \pi \in \Pi(\mu, \nu) \\ +\infty & \text{altrimenti} \end{cases} \right),$$

dove $M_+(\mathcal{X} \times \mathcal{Y})$ é lo spazio delle misure di Borel non negative su $\mathcal{X} \times \mathcal{Y}$.

Notiamo ora che:

$$\begin{cases} 0 & \text{se } \pi \in \Pi(\mu, \nu) \\ +\infty & \text{altrimenti} \end{cases} = \sup_{(\varphi, \psi)} \left[\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \right],$$

dove $(\varphi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})$. Infatti se $\pi \in \Pi(\mu, \nu)$, per definizione la quantità di destra si annulla, mentre se esistesse $\bar{\pi} \notin \Pi(\mu, \nu)$ tale per cui la quantità di destra si annullasse, allora avremmo che

$$\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\bar{\pi}(x, y) = 0$$

per ogni $(\varphi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})$. Scegliendo però $\varphi = \mathbb{1}_A$ e $\psi \equiv 0$ con $A \subseteq \mathcal{X}$ otterremmo che $\mu(A) = \bar{\pi}(A \times \mathcal{Y})$, e per arbitrarietà di A avremmo l'uguaglianza per ogni $A \subseteq \mathcal{X}$. Analogamente avremmo anche che $\nu(B) = \bar{\pi}(\mathcal{X} \times B)$, $\forall B \subseteq \mathcal{Y}$ scegliendo $\varphi \equiv 0$ e $\psi = \mathbb{1}_B$, da cui si può dedurre che $\bar{\pi} \in \Pi(\mu, \nu)$.

Perciò:

$$\begin{aligned} & \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \sup_{(\varphi, \psi)} \left[\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) + \int_{\mathcal{X}} \varphi d\mu \right. \\ & \quad \left. + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \right] \\ &= \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \left\{ I[\pi] + \sup_{(\varphi, \psi)} \left[\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \right] \right\} \\ &= \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \left(I[\pi] + \begin{cases} 0 & \text{se } \pi \in \Pi(\mu, \nu) \\ +\infty & \text{altrimenti} \end{cases} \right) = \inf_{\pi \in \Pi(\mu, \nu)} I[\pi], \end{aligned}$$

ma si nota anche che:

$$\begin{aligned} & \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \sup_{(\varphi, \psi)} \left[\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) + \int_{\mathcal{X}} \varphi d\mu \right. \\ & \quad \left. + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \right] \\ &= \sup_{(\varphi, \psi)} \inf_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \left[\int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y) + \int_{\mathcal{X}} \varphi d\mu \right. \\ & \quad \left. + \int_{\mathcal{Y}} \psi d\nu - \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)] d\pi(x, y) \right] \\ &= \sup_{(\varphi, \psi)} \left\{ \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \sup_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y) - c(x, y)] d\pi(x, y) \right\}. \end{aligned}$$

Concentriamoci ora sul sup interno:

Definita $\xi(x, y) = \varphi(x) + \psi(y) - c(x, y)$ avremo che se $\exists(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ tale che $\xi(x_0, y_0) > 0$, allora scegliendo $\pi = \lambda \delta_{(x_0, y_0)}$ si nota che mandando $\lambda \rightarrow +\infty$ si avrà che il sup sarà sicuramente $+\infty$.

Se invece ξ é non positiva il sup é chiaramente 0.

Da questa osservazione possiamo concludere che:

$$\sup_{\pi \in M_+(\mathcal{X} \times \mathcal{Y})} \int_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y) - c(x, y)] d\pi(x, y) = \begin{cases} 0 & \text{se } (\varphi, \psi) \in \Phi_c \\ +\infty & \text{altrimenti} \end{cases}$$

perció otteniamo che:

$$\sup_{(\varphi, \psi)} \left[\int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{Y}} \psi d\nu - \begin{cases} 0 & \text{se } (\varphi, \psi) \in \Phi_c \\ +\infty & \text{altrimenti} \end{cases} \right] = \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi)$$

Possiamo quindi concludere che

$$\inf_{\Pi(\mu, \nu)} I[\pi] = \sup_{\Phi_c} J(\varphi, \psi).$$

□

Il problema di ottimizzazione del trasporto di Kantorovich e la sua formulazione duale sono i concetti alla base della distanza di Wasserstein come vedremo meglio nel capitolo seguente.

Capitolo 2

Distanza di Wasserstein

Il seguente capitolo tratta della distanza di Wasserstein, delle sue applicazioni e proprietà. Vedremo come si esprime nella sua formulazione duale e come, sotto determinate ipotesi, acquisisce espressioni semplici da calcolare soprattutto tra distribuzioni reali 1-dimensionali. Infine daremo un risultato di convergenza e uno di completezza e separabilità.

2.1 Definizione e primi risultati

Per prima cosa, dato $\mathcal{X}$ spazio Polacco e $p \geq 1$, definiamo il sottospazio di $\mathcal{P}(\mathcal{X})$ delle misure di probabilità di ordine p , su cui poi definiremo con generalità la distanza p -Wasserstein W_p .

Definizione 2.1.1 ($\mathcal{P}_p(\mathcal{X})$). Sia $\mathcal{X}$ uno spazio Polacco. Per ogni $p \geq 1$ indichiamo con $\mathcal{P}_p(\mathcal{X})$ il sottospazio di $\mathcal{P}(\mathcal{X})$ delle misure di probabilità di ordine p e lo definiamo nel seguente modo: fissato $x_0 \in \mathcal{X}$ qualunque, allora

$$\mathcal{P}_p(\mathcal{X}) = \left\{ \mu \in \mathcal{P}(\mathcal{X}) \mid \int_{\mathcal{X}} (x, x_0)^p d\mu(x) < +\infty \right\}. \quad (2.1)$$

Mostriamo ora che la precedente definizione é ben posta, ovvero che non dipende da $x_0 \in \mathcal{X}$.

Osservazione 2.1.1. Notiamo che per ogni $a, b \geq 0$ e per ogni $p \geq 1$ vale

$$|a + b|^p \leq 2^p(|a|^p + |b|^p),$$

infatti, dato $c = \max\{|a|, |b|\}$ si ha

$$|a + b| \leq |a| + |b| \leq 2|c| \Rightarrow |a + b|^p \leq 2^p |c|^p \leq 2^p(|a|^p + |b|^p).$$

Osservazione 2.1.2. Grazie alla precedente osservazione notiamo che la definizione (2.1) é ben posta, infatti dato $x_1 \in \mathcal{X}$, $x_1 \neq x_0$ si ha che

$$\begin{aligned} \int_{\mathcal{X}} d(x, x_1)^p d\mu(x) &\leq \int_{\mathcal{X}} (d(x, x_0) + d(x_0, x_1))^p d\mu(x) \\ &\leq \int_{\mathcal{X}} [2^p d(x, x_0)^p + 2^p d(x_0, x_1)^p] d\mu(x) \\ &= 2^p \int_{\mathcal{X}} d(x, x_0)^p d\mu(x) + 2^p d(x_0, x_1)^p < +\infty. \end{aligned}$$

Siamo ora pronti a definire la distanza p-Wasserstein.

Definizione 2.1.2 ($W_p(\mu, \nu)$). Dato $(\mathcal{X}, d)$ uno spazio Polacco, per ogni $p \geq 1$ e $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ la distanza p-Wasserstein $W_p(\mu, \nu)$ é definita nel seguente modo:

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left[\int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) \right]^{1/p}.$$

Come anticipato nel precedente capitolo la forma della distanza di Wasserstein é simile a quella del problema di ottimizzazione di trasporto introdotto da Kantorovich, per cui possiamo darne una formulazione duale come segue nella prossima proposizione.

Proposizione 2.1.1. Se $(\mathcal{X}, d)$ é uno spazio Polacco, $p \geq 1$ e $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$, allora:

$$W_p(\mu, \nu)^p = \sup_{\varphi(x) + \psi(y) \leq d(x, y)^p} \left[\int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{X}} \psi(y) d\nu(y) \right] \quad (2.2)$$

con $\varphi, \psi \in C_b(\mathcal{X})$.

Prima della dimostrazione della proposizione diamo la definizione di mappa di trasporto e distribuzione *push-forward*.

Definizione 2.1.3 (*push-forward*). Dato $\mathcal{X}$ uno spazio Polacco, $\mu \in \mathcal{P}(\mathcal{X})$ e $\psi: \mathcal{X} \rightarrow \mathcal{X}$ una funzione misurabile, allora diciamo che $\mu \circ \psi^{-1}$ é la distribuzione *push-forward* di μ e la indichiamo con $\psi_{\#}\mu$.

Definizione 2.1.4 (Mappa di trasporto). Dato $\mathcal{X}$ uno spazio Polacco, $\mu, \nu \in \mathcal{P}(\mathcal{X})$ diciamo che $\psi: \mathcal{X} \rightarrow \mathcal{X}$ misurabile é una mappa di trasporto da μ a ν se la *push-forward* $\mu \circ \psi^{-1}$ corrisponde con ν .

Siamo ora pronti ad affrontare la dimostrazione della proposizione.

Dimostrazione. La dimostrazione si sviluppa in quattro step.

Primo Step.

Definiamo la quantità a destra della (2.2) con $\tilde{W}_p(\mu, \nu)$ e proviamo che $\tilde{W}_p(\mu, \nu)$ soddisfa la disuguaglianza triangolare, cioè che date μ, ν e $\theta \in \mathcal{P}_p(\mathcal{X})$ si ha:

$$\tilde{W}_p(\mu, \nu) \leq \tilde{W}_p(\mu, \theta) + \tilde{W}_p(\theta, \nu)$$

Sappiamo che $\forall a, b \geq 0$ e $p \geq 1$

$$(a + b)^p = \inf_{s \in (0,1)} (s^{p-1}a^p + (1-s)^{p-1}b^p) \quad (2.3)$$

e l'uguaglianza si ottiene per $s = \frac{a}{a+b}$. Siano $\varphi, \psi \in C_b(\mathcal{X})$ che soddisfano $\varphi(x) + \psi(y) \leq d(x, y)^p$ per ogni $x, y \in X$. Notiamo che la (2.3) implica che, per ogni $s \in (0, 1)$ e $x, y, z \in \mathcal{X}$, abbiamo che:

$$\varphi(x) + \psi(y) \leq d(x, y)^p \leq (d(x, z) + d(z, y))^p \leq s^{p-1}d(x, z)^p + (1-s)^{p-1}d(z, y)^p. \quad (2.4)$$

Definiamo ora la seguente funzione:

$$\xi(z) = \inf_{x \in \mathcal{X}} [s^{p-1}d(x, z)^p - \varphi(x)], \quad z \in \mathcal{X}.$$

Per costruzione:

$$\varphi(x) + \xi(z) \leq s^{p-1}d(x, z)^p, \quad x, z \in \mathcal{X}. \quad (2.5)$$

Inoltre:

$$\begin{aligned} \psi(y) - \xi(z) &= \varphi(x) + \psi(y) - \xi(z) - \varphi(x) \\ &\leq s^{p-1}d(x, z)^p + (1-s)^{p-1}d(y, z)^p - \xi(z) - \varphi(x), \end{aligned}$$

dove abbiamo usato le assunzioni su φ e ψ e la disuguaglianza (2.4). Dunque deduciamo che:

$$\begin{aligned}\psi(y) - \xi(z) &\leq \inf_{x \in \mathcal{X}} [s^{p-1} d(x, z)^p - \varphi(x)] + (1-s)^{p-1} d(y, z)^p - \xi(z) \\ &= (1-s)^{p-1} d(y, z)^p.\end{aligned}\tag{2.6}$$

Unendo la (2.5) e la (2.6) otteniamo:

$$\begin{aligned}&\int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{X}} \psi(y) d\nu(y) \\ &\leq \left[\int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{X}} \xi(z) d\theta(z) \right] + \left[\int_{\mathcal{X}} \psi(x) d\nu(x) - \int_{\mathcal{X}} \xi(z) d\theta(z) \right] \\ &\leq s^{p-1} \tilde{W}_p(\mu, \theta)^p + (1-s)^{p-1} \tilde{W}_p(\theta, \nu)^p,\end{aligned}$$

dove abbiamo usato la definizione di $\tilde{W}_p$. Dato che nella parte sinistra della disuguaglianza non compaiono φ e ψ , allora la disuguaglianza continua a valere anche prendendo il supremo nella parte sinistra sulle funzioni φ e ψ che soddisfano $\varphi(x) + \psi(y) \leq d(x, y)^p$. Otteniamo quindi:

$$\tilde{W}_p(\mu, \nu)^p \leq s^{p-1} \tilde{W}_p(\mu, \theta)^p + (1-s)^{p-1} \tilde{W}_p(\theta, \nu)^p.$$

Prendendo ora l'inf su $s \in (0, 1)$ nella parte destra dell'ultima disuguaglianza otteniamo che:

$$\tilde{W}_p(\mu, \nu)^p \leq [\tilde{W}_p(\mu, \theta) + \tilde{W}_p(\theta, \nu)]^p.$$

Secondo Step.

Proviamo ora che la tesi vale quando lo spazio $\mathcal{X}$ é finito.

Sia $\mathcal{X} = \{e_1, \dots, e_n\}$, in questo caso useremo la seguente notazione: $\mu(i) = \mu(\{e_i\})$ e $\nu(i) = \nu(\{e_i\})$ per $i = 1, \dots, n$. Per definizione avremo che:

$$\begin{aligned}W_p(\mu, \nu)^p &= \inf_{\substack{\pi(i, j) \geq 0, \sum_{1 \leq i \leq n} \pi(i, j) = \nu(j), \\ \sum_{1 \leq j \leq n} \pi(i, j) = \mu(i)}} \sum_{1 \leq i, j \leq n} d(e_i, e_j)^p \pi(i, j).\end{aligned}$$

Il precedente problema é riconducibile al seguente: sia $b = (b_{(i, j)} = d(e_i, e_j)^p)_{1 \leq i, j \leq n}$, $A = (A_{l, (i, j)})_{1 \leq l \leq 2n, 1 \leq (i, j) \leq n}$ con $A_{l, (i, j)} = \mathbf{1}_{i=l}$ se $l \leq n$, $A_{l, (i, j)} = \mathbf{1}_{j=l-n}$ se

$l > n$ e $c = (c(l))_{1 \leq l \leq 2n}$ con $c(l) = \mu(l)$ se $l \leq n$ e $c(l) = \nu(l - n)$ se $l > n$.

Allora avremo che

$$\inf_{\substack{\pi(i, j) \geq 0, \sum_{1 \leq i \leq n} \pi(i, j) = \nu(j), \\ \sum_{1 \leq j \leq n} \pi(i, j) = \mu(i)}} \sum_{1 \leq i, j \leq n} d(e_i, e_j)^p \pi(i, j) = \inf_{\pi(i, j) \geq 0, A\pi = c} b \cdot \pi,$$

e grazie alla teoria della dualit  per programmazione lineare finita dimensionale sappiamo che:

$$\sup_{A^T x \leq b} c \cdot x = \inf_{y \geq 0, Ay = c} b \cdot y,$$

applicando tale risultato al problema attuale, otteniamo che:

$$\begin{aligned} W_p(\mu, \nu)^p &= \inf_{\substack{\pi(i, j) \geq 0, \sum_{1 \leq i \leq n} \pi(i, j) = \nu(j), \\ \sum_{1 \leq j \leq n} \pi(i, j) = \mu(i)}} \sum_{1 \leq i, j \leq n} d(e_i, e_j)^p \pi(i, j) \\ &= \inf_{\pi(i, j) \geq 0, A\pi = c} b \cdot \pi = \sup_{A^T x \leq b} c \cdot x \\ &= \sup_{\varphi(i) + \psi(j) \leq d(i, j)^p} \left[\sum_{1 \leq i \leq n} \varphi(i) \mu(i) + \sum_{1 \leq j \leq n} \psi(j) \nu(j) \right], \end{aligned}$$

se denotiamo con $\varphi(1), \dots, \varphi(n), \psi(1), \dots, \psi(n)$ le componenti del vettore x . Ma notiamo che l'ultima espressione corrisponde alla dualit  di Kantorovich (2.2) nel caso di $\mathcal{X}$ finito.

Terzo Step.

Ora proveremo la disuguaglianza $W_p(\mu, \nu)^p \geq \tilde{W}_p(\mu, \nu)^p$ in generale. Ci  segue dal fatto che se $\varphi, \psi \in C_b(\mathcal{X})$ soddisfano $\varphi(x) + \psi(y) \leq d(x, y)^p$, allora per ogni $\pi \in \Pi(\mu, \nu)$, abbiamo che:

$$\begin{aligned} \int_{\mathcal{X}} \varphi(x) d\mu(x) + \int_{\mathcal{X}} \psi(y) d\nu(y) &= \int_{\mathcal{X} \times \mathcal{X}} \varphi(x) d\pi(x, y) + \int_{\mathcal{X} \times \mathcal{X}} \psi(y) d\pi(x, y) \\ &\leq \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y). \end{aligned}$$

Dato che la parte sinistra   indipendente da π , possiamo considerare l'inf a destra su ogni $\pi \in \Pi(\mu, \nu)$ e ottenere cos  $W_p(\mu, \nu)^p$. Possiamo ora considerare

il sup a sinistra su tutte le (φ, ψ) e ottenere cosí la disuguaglianza desiderata.

Quarto Step.

Proviamo, infine, la disuguaglianza inversa per ottenere la tesi. Sia $x_0 \in \mathcal{X}$ fissato. Dato $\epsilon > 0$, $\exists K_\epsilon \subset \mathcal{X}$ compatto tale che:

$$\int_{K_\epsilon} d(x, x_0)^p [d\mu(x) + d\nu(y)] < \epsilon^p.$$

Sia ora una partizione finita $(\mathcal{X}_\epsilon^j)_{1 \leq j \leq n}$ di K_ϵ di Boreliani di diametro al massimo ϵ , e per ogni $j \in \{1, \dots, n\}$ scegliamo un punto $x_j \in \mathcal{X}_\epsilon^j$. Infine,

$$\text{definiamo } f: \mathcal{X} \rightarrow \mathcal{X}, f(x) = \begin{cases} x_j & \text{se } x \in \mathcal{X}_\epsilon^j \\ x_0 & \text{se } x \in K_\epsilon^c \end{cases}.$$

Chiaramente f é una mappa di trasporto tra μ e $\tilde{\mu} = \mu \circ f^{-1}$ come lo é anche tra ν e $\tilde{\nu} = \nu \circ f^{-1}$. Notiamo che, per definizione di f , vale la seguente stima:

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) &= \int_{\mathcal{X}} d(x, f(x))^p d\mu(x) \\ &= \int_{K_\epsilon^c} d(x, f(x))^p d\mu(x) + \int_{K_\epsilon} d(x, f(x))^p d\mu(x) \\ &= \int_{K_\epsilon^c} d(x, x_0)^p d\mu(x) + \int_{K_\epsilon} d(x, f(x))^p d\mu(x) \\ &\leq \epsilon^p + \epsilon^p \mu(K_\epsilon) \leq 2\epsilon^p. \end{aligned} \tag{2.7}$$

Per definizione di $W_p(\mu, \tilde{\mu})$ e per la stima (2.7) avremo quindi che $W_p(\mu, \tilde{\mu}) \leq 2^{1/p}\epsilon$. I ragionamenti precedenti valgono anche per $W_p(\nu, \tilde{\nu})$, perciò segue che $W_p(\nu, \tilde{\nu}) \leq 2^{1/p}\epsilon$, e, usando la disuguaglianza triangolare per la distanza W_p otteniamo che:

$$W_p(\mu, \nu) \leq W_p(\mu, \tilde{\mu}) + W_p(\tilde{\mu}, \tilde{\nu}) + W_p(\mu, \tilde{\nu}) \leq W_p(\tilde{\mu}, \tilde{\nu}) + 2^{1+1/p}\epsilon. \tag{2.8}$$

Usando il risultato provato nello *Secondo Step*, notiamo che $W_p(\tilde{\mu}, \tilde{\nu}) = \tilde{W}_p(\tilde{\mu}, \tilde{\nu})$ e usando la disuguaglianza triangolare dimostrata nel *Primo Step*, otteniamo:

$$\begin{aligned} \tilde{W}_p(\tilde{\mu}, \tilde{\nu}) &\leq \tilde{W}_p(\tilde{\mu}, \mu) + \tilde{W}_p(\mu, \nu) + \tilde{W}_p(\nu, \tilde{\nu}) \\ &\leq W_p(\tilde{\mu}, \mu) + \tilde{W}_p(\mu, \nu) + W_p(\nu, \tilde{\nu}) \\ &\leq \tilde{W}_p(\mu, \nu) + 2^{1+1/p}\epsilon, \end{aligned} \tag{2.9}$$

dove abbiamo usato il fatto provato nel *Terzo Step*. Usando la (2.8) e la (2.9) otteniamo:

$$W_p(\mu, \nu) = \tilde{W}_p(\mu, \nu) + 2^{1+1/p}\epsilon.$$

e mandando $\epsilon \rightarrow 0^+$ concludiamo la dimostrazione. $\square$

Una conseguenza immediata della proposizione é la seguente caratterizzazione della distanza 1-Wasserstein W_1 .

Diamo prima la definizione di funzione 1-Lipschitziana.

Definizione 2.1.5 (f 1-Lipschitziana). Sia $(\mathcal{X}, d)$ uno spazio Polacco e $f: \mathcal{X} \rightarrow \mathcal{X}$, allora diciamo che f é 1-Lipschitziana se

$$d(f(x), f(y)) \leq d(x, y) \quad \forall x, y \in \mathcal{X}.$$

Possiamo ora enunciare e dimostrare il seguente corollario.

Corollario 2.1.1. Se $(\mathcal{X}, d)$ é uno spazio Polacco, e $\mu, \nu \in \mathcal{P}_1(\mathcal{X})$, allora:

$$W_1(\mu, \nu) = \sup_{\varphi \text{ 1-Lipschitziana}} \int_{\mathcal{X}} \varphi(x) (d\mu - d\nu)(x).$$

Dimostrazione. La formula di dualit  di Kantorovich ci dice che per $p = 1$:

$$W_1(\mu, \nu) = \sup_{(\varphi, \psi)} \left\{ \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{X}} \psi d\nu \right\},$$

dove il sup é preso sulle $\varphi, \psi \in C_b(\mathcal{X})$ tali che

$$\varphi(x) + \psi(y) \leq d(x, y), \quad \forall x, y \in \mathcal{X}.$$

Data ψ , definiamo per ogni $x \in \mathcal{X}$

$$\varphi(x) := \inf_{y \in \mathcal{X}} \{d(x, y) - \psi(y)\}.$$

Questa nuova funzione φ soddisfa per definizione $\varphi(x) + \psi(y) \leq d(x, y)$ e risulta anche 1-Lipschitz, poich :

$$\varphi(x) - \varphi(z) \leq \sup_{y \in \mathcal{X}} [d(x, y) - d(z, y)] \leq d(x, z).$$

Inoltre, scambiando i ruoli di x e y , si ottiene $\psi = -\varphi$. Ne segue che

$$W_1(\mu, \nu) = \sup_{\varphi \text{ 1-Lipschitziana}} \left\{ \int_{\mathcal{X}} \varphi d\mu + \int_{\mathcal{X}} \varphi d\nu \right\} = \sup_{\varphi \text{ 1-Lipschitziana}} \int_{\mathcal{X}} \varphi d(\mu - \nu).$$

$\square$

Interpretazione probabilistica in termini di variabili aleatorie

Dal momento che stiamo trattando misure di probabilità, possiamo dare un'interpretazione probabilistica alla distanza di Wasserstein.

Definizione 2.1.6. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e dato $\mathcal{X}$ uno spazio Polacco, sia X una variabile aleatoria a valori in $\mathcal{X}$. X si dice di ordine p , per $p > 1$, se $\mathbb{E}[d(x_0, X)^p] < +\infty$ per un $x_0 \in \mathcal{X}$.

Definizione 2.1.7. Dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, dato $\mathcal{X}$ uno spazio Polacco, $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ con $p \geq 1$, possiamo definire $W_p(\mu, \nu)^p$ nel seguente modo:

$$W_p(\mu, \nu)^p = \inf \{ \mathbb{E}[d(X, Z)^p] \mid X \sim \mu, Z \sim \nu \}.$$

2.2 Distanza 1-Wasserstein tra distribuzioni reali

Nel caso di due distribuzioni 1-dimensionali reali la distanza 1-Wasserstein assume una forma particolare in cui compaiono le rispettive funzioni di ripartizione. Diamo quindi le seguenti definizioni.

Definizione 2.2.1 (CDF). Sia $\mu \in \mathcal{P}(\mathbb{R})$, allora la sua funzione di ripartizione o CDF é definita da

$$F_\mu(x) := \mu([-\infty, x]), \quad x \in \mathbb{R}.$$

Definizione 2.2.2 (Dominanza stocastica). Date X, Y due variabili aleatorie reali con rispettive CDF F_X, F_Y . Diciamo che X domina Y e lo indichiamo nel seguente modo $X \succ Y$, se

$$F_X^{-1}(u) \geq F_Y^{-1}(u) \quad u \in [0, 1].$$

Vale quindi il seguente teorema.

Teorema 2.2.1. Siano $\mu, \nu \in \mathcal{P}(\mathbb{R})$ e rispettivamente F_μ, F_ν le loro CDF, allora

$$W_1(\mu, \nu) = \int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)| du.$$

Si veda Teorema 2.1 [AG21].

Vale, inoltre, anche il seguente corollario.

Corollario 2.2.1. Siano X, Y due variabili aleatorie reali e μ_X, μ_Y le loro rispettive distribuzioni. Se $X \succ Y$, allora

$$W_1(\mu_X, \mu_Y) = \mathbb{E}[X] - \mathbb{E}[Y].$$

Dimostrazione. Per il teorema (2.2.1) e per l'ipotesi di dominanza sappiamo che

$$\begin{aligned} W_1(\mu_X, \mu_Y) &= \int_0^1 |F_X^{-1}(u) - F_Y^{-1}(u)| du \\ &= \int_0^1 F_X^{-1}(u) - F_Y^{-1}(u) du \\ &= \int_0^1 F_X^{-1}(u) du - \int_0^1 F_Y^{-1}(u) du \\ &= \int_{\mathbb{R}} x d\mu_X(x) - \int_{\mathbb{R}} y d\mu_Y(y) \\ &= \mathbb{E}[X] - \mathbb{E}[Y]. \end{aligned}$$

□

Vedremo in seguito come i risultati precedenti rendano molto più facile il calcolo della distanza 1-Wasserstein tra distribuzioni reali.

Distanza 1-Wasserstein tra delta di Dirac

Sia $(\mathcal{X}, d) = (\mathbb{R}, d_\epsilon)$, $\mu = \delta_{x_1}$, $\nu = \delta_{x_2}$ con d_ϵ distanza euclidea, $x_1, x_2 \in \mathbb{R}$, $x_1 \neq x_2$ e $\delta_{x_1}, \delta_{x_2}$ misure di probabilità discrete definite nel seguente modo:

$$\delta_{x_i}(A) = \begin{cases} 1 & \text{se } x_i \in A \\ 0 & \text{altrimenti} \end{cases} \quad i \in \{1, 2\}.$$

Osserviamo che per ogni $i \in \{1, 2\}$, $\delta_{x_i} \in \mathcal{P}_1(\mathbb{R})$, infatti fissato, per comodità, $x_0 = 0$ si ha che:

$$\int_{\mathbb{R}} d(0, x) d\delta_{x_i}(x) = \int_{\mathbb{R}} |x| d\delta_{x_i}(x) = |x_i| < +\infty, \quad i \in \{1, 2\}.$$

Osservazione 2.2.1. Notiamo che dati $a, b \in \mathbb{R}$, $a \neq b$ si ha $\Pi(\delta_a, \delta_b) = \{\delta_{(a,b)}\}$. Infatti, sia $\mu \in \Pi(\delta_a, \delta_b)$ e siano $C := \mathbb{R}^2 \setminus \{(a, b)\}$, $C_x := \mathbb{R} \setminus \{a\}$ e $C_y := \mathbb{R} \setminus \{b\}$.

Sappiamo che $C = (C_x \times \mathbb{R}) \cup (\mathbb{R} \times C_y)$ e per subadditività della misura si ha che

$$\mu(C) \leq \mu(C_x \times \mathbb{R}) + \mu(\mathbb{R} \times C_y),$$

ma dato che $\mu \in \Pi(\delta_a, \delta_b)$

$$\mu(C_x \times \mathbb{R}) + \mu(\mathbb{R} \times C_y) = \delta_a(C_x) + \delta_b(C_y) = 0.$$

Perciò $\mu(C) = 0 \Rightarrow \mu(C^c) = 1$, da cui deduciamo che $\mu = \delta_{(a,b)}$.

Dato che $\Pi(\delta_{x_1}, \delta_{x_2})$ ha solo un elemento, ovvero $\delta_{(x_1, x_2)}$ delta di Dirac centrata nel punto (x_1, x_2) , per definizione di $W_1(\delta_{x_1}, \delta_{x_2})$ avremo che:

$$\inf_{\pi \in \Pi(\delta_{x_1}, \delta_{x_2})} \int_{\mathbb{R}^2} d(x, y) d\pi(x, y) = \int_{\mathbb{R}^2} |x - y| d\delta_{(x_1, x_2)}(x, y) = |x_1 - x_2|.$$

Sfruttando invece, il corollario (2.1.1) notiamo che:

$$\begin{aligned} W_1(\delta_{x_1}, \delta_{x_2}) &= \sup_{\varphi \text{ 1-Lipschitziana}} \int_{\mathbb{R}} \varphi(x) (d\delta_{x_1} - d\delta_{x_2})(x) \\ &= \sup_{\varphi \text{ 1-Lipschitziana}} \left[\int_{\mathbb{R}} \varphi(x) d\delta_{x_1} - \int_{\mathbb{R}} \varphi(x) d\delta_{x_2} \right] \\ &= \sup_{\varphi \text{ 1-Lipschitziana}} (\varphi(x_1) - \varphi(x_2)), \end{aligned}$$

da cui possiamo concludere che:

$$\sup_{\varphi \text{ 1-Lipschitziana}} (\varphi(x_1) - \varphi(x_2)) = W_1(\delta_{x_1}, \delta_{x_2}) = |x_1 - x_2|.$$

Distanza 1-Wasserstein tra distribuzioni esponenziali reali

Sia $(X, d) = (\mathbb{R}, d_\epsilon)$, $\mu = \text{Exp}_{\lambda_1}$, $\nu = \text{Exp}_{\lambda_2}$ con d_ϵ distanza euclidea, $\lambda_1, \lambda_2 \in \mathbb{R}_{>0}$ e $\text{Exp}_{\lambda_1}, \text{Exp}_{\lambda_2}$ misure di probabilità assolutamente continue con CDF:

$$F_i(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-\lambda_i x} & \text{se } x \geq 0 \end{cases}, \quad i \in \{1, 2\}.$$

Utilizzando la formula (2.2.1) avremo che

$$W_1(\text{Exp}_{\lambda_1}, \text{Exp}_{\lambda_2}) = \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)| du.$$

Dato che le CDF valgono entrambe 0 se $x < 0$ e l'integrale va da 0 a 1, calcoliamo le inverse delle CDF ristrette all'intervallo $[0, +\infty)$. Abbiamo infatti che $F_i|_{[0, +\infty)}(x) = 1 - e^{-\lambda_i x}$, perciò avremo che

$$F_i^{-1}|_{[0, +\infty)}(x) = -\frac{1}{\lambda_i} \log(1 - x).$$

Perciò, svolgendo dei semplici calcoli, avremo che

$$\begin{aligned} \int_0^1 |F_1^{-1}(u) - F_2^{-1}(u)| du &= \int_0^1 \left| \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right| |\log(1 - u)| du \\ &= \left| \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right| \int_0^1 |\log(1 - u)| du \\ &= \left| \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right| [u - u \log(u)]_0^1 \\ &= \left| \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right|. \end{aligned}$$

Possiamo quindi concludere che

$$W_1(\text{Exp}_{\lambda_1}, \text{Exp}_{\lambda_2}) = \left| \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right|.$$

Notiamo che se avessimo avuto in piú l'ipotesi che $0 < \lambda_1 < \lambda_2$, allora $F_1^{-1}(u) = -\frac{1}{\lambda_1} \log(1-u) > -\frac{1}{\lambda_2} \log(1-u) = F_2^{-1}(u)$ per ogni $u \in [0, 1]$ perché $-\log(1-u)|_{[0,1]}$ é una funzione crescente, continua e positiva.

Quindi per il corollario (2.2.1) avremo che

$$W_1(\text{Exp}_{\lambda_1}, \text{Exp}_{\lambda_2}) = \int_{\mathbb{R}} x d\text{Exp}_{\lambda_1}(x) - \int_{\mathbb{R}} y d\text{Exp}_{\lambda_2}(y) = \frac{1}{\lambda_1} - \frac{1}{\lambda_2},$$

soluzione coerente con il risultato ottenuto in precedenza sfruttando il teorema (2.2.1).

2.3 Completezza e separabilità di $(\mathcal{P}_p(\mathcal{X}), W_p)$

Nella seguente sezione introdurremo un nuovo concetto di convergenza in $\mathcal{P}_p$ e vedremo come quest'ultima sia equivalente alla convergenza in W_p . Riusciremo, inoltre, a dimostrare un risultato fondamentale, ovvero che dato $p \geq 1$ e $\mathcal{X}$ uno spazio Polacco, $(\mathcal{P}_p(\mathcal{X}), W_p)$ rimane uno spazio Polacco. Diamo per iniziare alcune definizioni che saranno utili nel corso della trattazione.

Definizione 2.3.1 (Π_p^{opt}). Sia $(\mathcal{X}, d)$ uno spazio Polacco, $p \geq 1$ e $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$. Definiamo $\Pi_p^{opt}(\mu, \nu)$ come l'insieme delle misure di probabilità congiunte per le quali l'inf é raggiunto, cioè:

$$\Pi_p^{opt}(\mu, \nu) := \left\{ \pi \in \Pi(\mu, \nu) \mid (W_p(\mu, \nu))^p = \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi(x, y) \right\}.$$

Definizione 2.3.2 (Convergenza debole). Dati $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ spazio di probabilità, $\mu_n \in \mathcal{P}(\mathcal{X})$ per ogni $n \in \mathbb{N}$ e $\mu \in \mathcal{P}(\mathcal{X})$ diciamo che μ_n converge debolmente a μ e lo indicheremo con $\mu_n \xrightarrow{d} \mu$, se

$$\int_{\mathcal{X}} f(x) d\mu_n(x) \xrightarrow{n \rightarrow +\infty} \int_{\mathcal{X}} f(x) d\mu(x)$$

per ogni funzione $f \in C_b(\mathcal{X})$.

Osservazione 2.3.1. Questo tipo di convergenza risulta piú debole rispetto alla distanza W_p , infatti W_p risulta essere inferiormente semicontinua rispetto alla topologia indotta dalla convergenza debole. Cioé, se $\mu_n \xrightarrow{d} \mu$ e $\nu_n \xrightarrow{d} \nu$, allora

$$W_p(\mu, \nu) \leq \liminf_{n \rightarrow +\infty} W_p(\mu_n, \nu_n),$$

si veda il Capitolo 4 di [Vil09]. Perciò rafforziamo la definizione di convergenza debole in $\mathcal{P}_p(\mathcal{X})$ affinché quest'ultima e la nozione di convergenza in W_p siano equivalenti.

Definizione 2.3.3 (Convergenza in $\mathcal{P}_p(\mathcal{X})$). Sia $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}_p(\mathcal{X})$ una successione e $\mu \in \mathcal{P}_p(\mathcal{X})$. Diciamo che $(\mu_n)_{n \in \mathbb{N}}$ converge a μ in $\mathcal{P}_p(\mathcal{X})$, e scriveremo $\mu_n \xrightarrow{\mathcal{P}_p} \mu$, se esiste $x_0 \in \mathcal{X}$ tale che

$$\mu_n \xrightarrow{d} \mu \quad e \quad \int_{\mathcal{X}} d(x, x_0)^p d\mu_n \xrightarrow{n \rightarrow +\infty} \int_{\mathcal{X}} d(x, x_0)^p d\mu.$$

Vale quindi il seguente risultato.

Teorema 2.3.1. Sia $\mathcal{X}$ uno spazio Polacco compatto, per ogni $p \geq 1$, se $(\mu_n)_{n \in \mathbb{N}}$ successione in $\mathcal{P}_p(\mathcal{X})$ e $\mu \in \mathcal{P}_p(\mathcal{X})$ allora, $\lim_{n \rightarrow +\infty} W_p(\mu_n, \mu) = 0$ se e solo se $\mu_n \xrightarrow{\mathcal{P}_p} \mu$.

Prima di dare una prova formale del precedente risultato, enunciamo dei teoremi necessari al fine della dimostrazione.

Teorema 2.3.2 (Disuguaglianza di Markov). Sia X una variabile aleatoria a valori in $\mathbb{R}^n$, $\delta > 0$ e $p \geq 0$, allora vale la seguente disuguaglianza

$$P(|X| \geq \delta) \leq \frac{\mathbb{E}[|X|^p]}{\delta^p}.$$

Si veda Teorema 3.1.4 [Pas24].

Teorema 2.3.3 (Teorema di Skorohod). Siano $\mu_1, \mu_2, \dots$ elementi di uno spazio metrico separabile $(\mathcal{X}, d)$ tali che $\mu_n \xrightarrow{d} \mu$. Allora su un adatto spazio di probabilità esistono $\eta \stackrel{d}{=} \mu$ e $\eta_n \stackrel{d}{=} \mu_n$, $n \in \mathbb{N}$ tali che $\eta_n \rightarrow \eta$ quasi certamente.

Si veda Teorema 3.30 [Kal21].

Teorema 2.3.4 (Lemma di Fatou). Dato $(\Omega, \mathcal{F}, \mu)$ uno spazio di misura e $\mathcal{X} \in \mathcal{F}$, sia $(f_n)_{n \geq 1}$ una successione di funzioni misurabili $f_n: \mathcal{X} \rightarrow [0, +\infty]$ per ogni $n \geq 1$. Definita $f(x) = \liminf_{n \rightarrow +\infty} f_n(x)$ per ogni $x \in \mathcal{X}$, allora f é misurabile e

$$\int_{\mathcal{X}} f d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\mathcal{X}} f_n d\mu.$$

Si veda Lemma 2.2.8 [Pas24].

Osservazione 2.3.2. Grazie al teorema di Skorohod (2.3.3) e al lemma di Fatou (2.3.4) osserviamo che:

$$\begin{aligned} \int_{\mathcal{X}} d(x_0, x)^p d\mu(x) &= \mathbb{E}[d(x_0, X)^p] \\ &\leq \liminf_{n \rightarrow +\infty} \mathbb{E}[d(x_0, X_n)^p] \\ &\leq \liminf_{n \rightarrow +\infty} \int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x). \end{aligned} \quad (2.10)$$

Vedremo in seguito come questo risultato sarà fondamentale all'interno della dimostrazione.

Possiamo ora dimostrare il teorema precedente (2.3.1).

Dimostrazione. Primo Step.

Assumiamo per ipotesi che $\lim_{n \rightarrow +\infty} W(\mu_n, \mu) = 0$. Sia $\pi_n \in \Pi_p^{opt}(\mu_n, \mu)$, per ogni $n \geq 1$, allora per ogni $f \in C_b(\mathcal{X})$ abbiamo

$$\int_{\mathcal{X}} f(x) d\mu_n(x) - \int_{\mathcal{X}} f(y) d\mu(y) = \int_{\mathcal{X} \times \mathcal{X}} (f(x) - f(y)) d\pi_n(x, y).$$

Dato che f é continua su $\mathcal{X}$ compatto, allora é uniformemente continua, cioè

$$\forall \epsilon > 0, \exists \delta > 0 \mid \forall x, y \in \mathcal{X} \quad d(x, y) < \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

Dividiamo quindi l'integrale nel seguente modo:

$$\begin{aligned} & \int_{\mathcal{X} \times \mathcal{X}} (f(x) - f(y)) d\pi_n(x, y) \\ &= \int_{d(x, y) \leq \delta} (f(x) - f(y)) d\pi_n(x, y) + \int_{d(x, y) > \delta} (f(x) - f(y)) d\pi_n(x, y). \end{aligned}$$

Notiamo ora che valgono le seguenti stime.

$$\left| \int_{d(x, y) \leq \delta} (f(x) - f(y)) d\pi_n(x, y) \right| \leq \epsilon,$$

inoltre

$$\left| \int_{d(x, y) > \delta} (f(x) - f(y)) d\pi_n(x, y) \right| \leq 2\|f\|_{\infty} \pi_n(d(x, y) > \delta)$$

e per la disuguaglianza (2.3.2) avremo che

$$\pi_n(d(x, y) > \delta) \leq \frac{1}{\delta^p} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p d\pi_n(x, y) = \frac{1}{\delta^p} W_p(\mu_n, \mu)^p.$$

Ora, grazie alle osservazioni appena fatte, abbiamo che:

$$\left| \int_{\mathcal{X}} f(x) d\mu_n(x) - \int_{\mathcal{X}} f(y) d\mu(y) \right| \leq \epsilon + \frac{2\|f\|_{\infty}}{\delta^p} W_p(\mu_n, \mu)^p,$$

e mandando ϵ a 0 otterremo che

$$\left| \int_{\mathcal{X}} f(x) d\mu_n(x) - \int_{\mathcal{X}} f(y) d\mu(y) \right| \leq \frac{2\|f\|_{\infty}}{\delta^p} W_p(\mu_n, \mu)^p \xrightarrow{n \rightarrow +\infty} 0.$$

Concludiamo quindi che $(\mu_n)_{n \geq 1}$ converge debolmente a μ .

Sia ora $\epsilon > 0$ fissato, allora esiste $c_{\epsilon} > 0$ tale che per ogni $a, b \geq 0$ vale

$(a + b)^p \leq (1 + \epsilon)a^p + c_{\epsilon}b^p$. Quindi, per $x, y \in \mathcal{X}$, valgono le seguenti disuguaglianze:

$$d(x_0, x)^p \leq (d(x_0, y) + d(y, x))^p \leq (1 + \epsilon)d(x_0, y)^p + c_{\epsilon}d(y, x)^p,$$

e integrando da entrambi i lati rispetto a $\pi_n \in \Pi_p^{opt}(\mu_n, \mu)$ otteniamo:

$$\int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x) \leq (1 + \epsilon) \int_{\mathcal{X}} d(x_0, y)^p d\mu(y) + c_{\epsilon} \int_{\mathcal{X} \times \mathcal{X}} d(y, x)^p d\pi_n(x, y).$$

Per definizione l'integrale piú a destra corrisponde a $W_p(\mu_n, \mu)^p$ che tende a 0 per $n \rightarrow +\infty$. Perciò:

$$\limsup_{n \rightarrow \infty} \int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x) \leq (1 + \epsilon) \int_{\mathcal{X}} d(x_0, y)^p d\mu(y),$$

e mandando $\epsilon \rightarrow 0$ avremo che:

$$\limsup_{n \rightarrow \infty} \int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x) \leq \int_{\mathcal{X}} d(x_0, y)^p d\mu(y).$$

Per la (2.10) sappiamo quindi che:

$$\limsup_{n \rightarrow \infty} \int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x) \leq \int_{\mathcal{X}} d(x_0, y)^p d\mu(y) \leq \liminf_{n \rightarrow \infty} \int_{\mathcal{X}} d(x_0, x)^p d\mu_n(x),$$

quindi $\mu_n \xrightarrow{\mathcal{P}_p} \mu$.

Secondo Step.

Viceversa, supponiamo ora che $\mu_n \xrightarrow{\mathcal{P}_p} \mu$. Per il teorema di Skorohod esiste una successione di variabili aleatorie $(X_n)_{n \geq 1}$ e una variabile aleatoria X definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ tali che $(X_n)_{n \geq 1}$ converge quasi certamente a X e $X \sim \mu$, $X_n \sim \mu_n$ per ogni $n \geq 1$. Per la (2.10) osserviamo che $X \in L^p$ e quindi che $X_n \xrightarrow{P} X$, inoltre:

$$W_p(\mu_n, \mu)^p \leq \mathbb{E}[d(X_n, X)^p] \xrightarrow{n \rightarrow +\infty} 0,$$

da cui possiamo concludere la dimostrazione. $\square$

Enunciamo ora il seguente teorema di completezza e separabilità di $(\mathcal{P}_p(\mathcal{X}), W_p)$ che sarà fondamentale nel terzo capitolo.

Teorema 2.3.5 (Completezza e separabilità di $(\mathcal{P}_p(\mathcal{X}), W_p)$). Sia $\mathcal{X}$ uno spazio Polacco e $p \geq 1$, allora $(\mathcal{P}_p(\mathcal{X}), W_p)$ é ancora uno spazio Polacco.

Dobbiamo prima però introdurre alcuni strumenti che saranno fondamentali al fine della dimostrazione.

Definizione 2.3.4 (*tightness*). Sia $\mathcal{X}$ uno spazio Polacco, allora un insieme di misure di probabilità $\mathcal{M}$ su $\mathcal{X}$ si dice *tight* se

$$\forall \epsilon > 0, \exists K_\epsilon \subset \mathcal{X} \text{ compatto} \mid \forall \mu \in \mathcal{M} \text{ si ha che } \mu(\mathcal{X} \setminus K_\epsilon) < \epsilon.$$

Teorema 2.3.6 (Prohorov). Sia $\mathcal{X}$ uno spazio Polacco, allora un insieme di misure di probabilità $\mathcal{M}$ su $\mathcal{X}$ é *tight* se e solo se $\mathcal{M}$ é precompatto, ovvero la sua chiusura é compatta, nella topologia della convergenza debole.

Si veda Teorema 13.29 [Kle14].

Al fine della trattazione a noi interessa il seguente corollario.

Corollario 2.3.1. Sotto le stesse ipotesi del teorema (2.3.6), se $\mathcal{M}$ é finito, allora é *tight*.

Dimostrazione. Se $\mathcal{M}$ é finito, allora é chiuso, perciò la sua chiusura é se stesso. Inoltre $\mathcal{M}$ é anche compatto, infatti supponiamo che $\mathcal{M} := \{\mu_1, \dots, \mu_n\}$, $n \in \mathbb{N}$ e consideriamo $\{U_i\}_i$ un suo ricoprimento aperto. Allora esisteranno $U_1, \dots, U_n$ tali che $\mu_i \in U_i$, $1 \leq i \leq n$, allora $\{U_i\}_{1 \leq i \leq n}$ sarà un sottoricoprimento finito di $\mathcal{M}$. Perciò per il teorema (2.3.6) $\mathcal{M}$ é *tight*. $\square$

Teorema 2.3.7. Siano μ, ν due misure di probabilità su uno spazio Polacco $(\mathcal{X}, d)$. Sia $p \geq 1$ e $x_0 \in \mathcal{X}$, allora

$$W_p(\mu, \nu) \leq 2^{\frac{1}{p'}} \left(\int_{\mathcal{X}} d(x_0, x)^p d|\mu - \nu|(x) \right)^{\frac{1}{p}}, \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

Si veda Teorema 6.15 [Vil09].

Lemma 2.3.1. Sia $\mathcal{X}$ spazio Polacco e $p \geq 1$, se $(\mu_n)_{n \in \mathbb{N}}$ é una successione di Cauchy in $(\mathcal{P}_p(\mathcal{X}), W_p)$, allora é *tight*.

Si veda Lemma 6.14 [Vil09].

Osservazione 2.3.3. Si noti che grazie al Teorema di Helly, si veda Teorema 3.3.7 [Pas24], sotto le ipotesi del lemma precedente esisteranno $(\mu_{k_n})_{n \in \mathbb{N}}$ sottosuccessione e $\mu \in \mathcal{P}(\mathcal{X})$ tali che $\mu_{k_n} \xrightarrow{d} \mu$.

Siamo ora pronti a dimostrare il teorema (2.3.5).

Dimostrazione. Proviamo prima la separabilit  e infine la completezza.

Separabilit .

Sia D un sottoinsieme denso e numerabile di $\mathcal{X}$ e sia $\mathcal{M}$ l'insieme delle misure di probabilit  del tipo

$$\sum_j a_j \delta_{x_j},$$

dove gli a_j sono razionali e gli x_j sono un numero finito di elementi di D . Mostriamo che $\mathcal{M}$ é denso in $\mathcal{P}_p(\mathcal{X})$.

Siano $\epsilon > 0$ e $x_0 \in D$ arbitrario. Sia $\mu \in \mathcal{P}_p(\mathcal{X})$, allora poniamo

$$r := \int_{\mathcal{X}} d(x, x_0)^p d\mu < +\infty.$$

Definiamo ora la seguente funzione a valori sui boreliani

$$\gamma(H) := \frac{1}{r} \int_H d(x, x_0)^p d\mu$$

e notiamo che é una misura di probabilit , quindi per il teorema (2.3.6) $\{\gamma\}$ é *tight*.

Perci  esiste $K \subset \mathcal{X}$ compatto tale per cui

$$\gamma(\mathcal{X} \setminus K) \leq \frac{\epsilon^p}{r},$$

per cui possiamo concludere che

$$\int_{\mathcal{X} \setminus K} d(x, x_0)^p d\mu \leq \epsilon^p.$$

Ora, per compattezza di K e densità di D , esiste un ricoprimento finito di K costituito da palle metriche del tipo: $B(x_k, \epsilon/2)$, con $1 \leq k \leq N$, $x_k \in D$, $N \in \mathbb{N}$. Definiamo, quindi, i seguenti insiemi:

$$\begin{aligned} T_1 &:= B(x_1, \epsilon), \\ T_k &:= B(x_k, \epsilon) \setminus \bigcup_{j < k} B(x_j, \epsilon) \quad 2 \leq k \leq N \end{aligned}$$

e notiamo che i T_k , $1 \leq k \leq N$ ricoprono ancora K e sono disgiunti. Definiamo ora $f: \mathcal{X} \rightarrow D$ nel seguente modo

$$\begin{cases} f(x) = x_k & \text{se } x \in T_k \cap \mathcal{X} \\ f(x) = x_0 & \text{se } x \in \mathcal{X} \setminus K \end{cases}.$$

Allora, per ogni $x \in K$, $d(x, f(x)) \leq \epsilon$, quindi:

$$\int_{\mathcal{X}} d(x, f(x))^p d\mu \leq \epsilon^p \int_K d\mu + \int_{\mathcal{X} \setminus K} d(x, x_0)^p d\mu \leq 2\epsilon^p.$$

Considerata quindi $f_{\#}\mu$ la *push-forward* di μ , per la disuguaglianza precedente, avremo che $W_p(\mu, f_{\#}\mu) \leq 2\epsilon^p$ con $f_{\#}\mu$ che per costruzione può essere scritta come una combinazione di delta di Dirac del tipo $\sum_{j=1}^N a_j \delta_{x_j}$, $x_j \in D$. Per concludere basta mostrare che i coefficienti a_j possono essere sostituiti da coefficienti razionali b_j a meno di un errore trascurabile in distanza W_p . Infatti grazie al teorema (2.3.7), abbiamo che

$$W_p \left(\sum_{j \leq N} a_j \delta_{x_j}, \sum_{j \leq N} b_j \delta_{x_j} \right) \leq 2^{\frac{1}{p'}} \left[\max_{k,l} d(x_k, x_l) \right] \sum_{j \leq N} |a_j - b_j|^{\frac{1}{p}},$$

e la quantità di destra può essere resa il più piccola possibile grazie a una buona scelta dei coefficienti b_j .

Perciò possiamo concludere che $\mathcal{M}$ è denso in $\mathcal{P}_p(\mathcal{X})$ e numerabile per costruzione.

Completezza.

Sia $(\mu_n)_{n \in \mathbb{N}}$ una successione di Cauchy in $(\mathcal{P}_p(\mathcal{X}), W_p)$. Per il lemma (2.3.1), la successione è *tight* e, per l'osservazione (2.3.3), ammette dunque una sottosuccessione $(\mu_{k_n})_{n \in \mathbb{N}}$ che converge debolmente ad una $\mu \in \mathcal{P}(\mathcal{X})$. Ma, siccome

W_p é inferiormente semicontinua rispetto alla convergenza debole

$$\begin{aligned} \int_{\mathcal{X}} d(x, x_0)^p d\mu = W_p(\delta_{x_0}, \mu)^p &\leq \liminf_{n \rightarrow +\infty} W_p(\delta_{x_0}, \mu_{k_n})^p \\ &= \liminf_{n \rightarrow +\infty} \int_{\mathcal{X}} d(x, x_0)^p d\mu_{k_n} < +\infty, \end{aligned}$$

quindi $\mu \in \mathcal{P}_p(\mathcal{X})$. Inoltre, sempre per inferiore semicontinuitá di W_p , si ha che

$$W_p(\mu, \mu_{k_n}) \leq \liminf_{m \rightarrow +\infty} W_p(\mu_{k_m}, \mu_{k_n}) \quad \forall n \in \mathbb{N},$$

da cui possiamo concludere che

$$\begin{aligned} \limsup_{n \rightarrow +\infty} W_p(\mu, \mu_{k_n}) &\leq \limsup_{n \rightarrow +\infty} \liminf_{m \rightarrow +\infty} W_p(\mu_{k_m}, \mu_{k_n}) \\ &\leq \limsup_{n, m \rightarrow +\infty} W_p(\mu_{k_m}, \mu_{k_n}) = 0. \end{aligned}$$

Perció $(\mu_{k_n})_{n \in \mathbb{N}}$ converge a μ attraverso W_p , e dato che $(\mu_{k_n})_{n \in \mathbb{N}}$ é di Cauchy per ipotesi e ammette una sottosuccessione convergente, allora si ha che $(\mu_n)_{n \in \mathbb{N}}$ converge a μ attraverso W_p .

Possiamo quindi concludere che $(\mathcal{P}_p(\mathcal{X}), W_p)$ é completo. $\square$

La struttura di spazio Polacco di $(\mathcal{P}_p(\mathcal{X}), W_p)$, ma soprattutto la sua completezza, é alla base della teoria dei processi di diffusione non lineare di McKean, servirá infatti a mostrarne l'esistenza e l'unicitá come vedremo nel capitolo successivo.

2.4 Applicazioni

La distanza di Wasserstein ha delle ottime caratteristiche e viene sfruttata in statistica e machine learning, in seguito tratteremo di alcune proprietà che rendono la distanza di Wasserstein così versatile.

Baricentro di Wasserstein

Quando vogliamo fare la media di diversi oggetti, come distribuzioni o immagini, vorremmo ottenere come risultato un oggetto simile a quelli di partenza.

Supponiamo di avere $\mu_1, \dots, \mu_N$ distribuzioni e definiamo il *baricentro di Wasserstein* come la distribuzione μ che minimizza

$$\sum_{i=1}^N W(\mu, \mu_i).$$

Scegliendo la media Euclidea delle distribuzioni $\bar{\mu} := \frac{1}{N} \sum_{i=1}^N \mu_i$ non si ottiene nulla di simile alle $\mu_i, i = 1, \dots, N$ di partenza, mentre utilizzando il baricentro di Wasserstein si ottengono ottimi risultati.

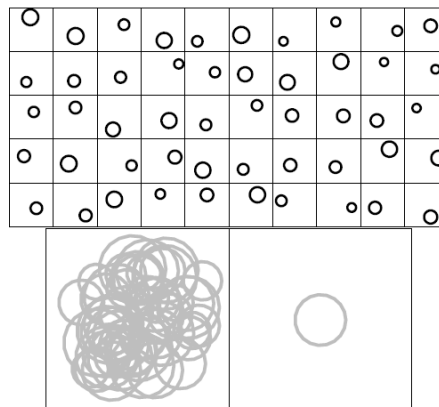


Figura 2.1: *In alto: alcune distribuzioni. In basso a sinistra: media Euclidea. In basso a destra: baricentro di Wasserstein.*

Un esempio sono le immagini precedenti che mostrano alcune distribuzioni, ciascuna concentrata su delle circonferenze. La figura in basso a sinistra mostra la media Euclidea delle distribuzioni come una massa grigia, mentre la figura in basso a destra mostra il baricentro di Wasserstein che offre un risultato migliore.

Geodetica di Wasserstein

Ricordiamo che dato $\mathcal{X}$ uno spazio Polacco, $p \geq 1$, $\mu \in \mathcal{P}_p(\mathcal{X})$ e $T: \mathcal{X} \rightarrow \mathcal{X}$ una funzione misurabile, allora diciamo che $\mu \circ T^{-1}$ é la distribuzione *push-forward* di μ e la indichiamo con $T_{\#}\mu$.

Siano ora $\mu_0, \mu_1 \in \mathcal{P}_p(\mathcal{X})$ due distribuzioni. Consideriamo una funzione $c: [0, 1] \rightarrow \mathcal{P}_p(\mathcal{X})$ tale che, $c(0) = \mu_0$ e $c(1) = \mu_1$, allora $(\mu_t : 0 \leq t \leq 1)$ é un cammino che connette μ_0 e μ_1 , dove $c(t) = \mu_t$, $\forall t \in [0, 1]$. Definiamo la lunghezza di c nel seguente modo

$$L(c) := \sup \sum_{i=1}^m W_p(c(t_{i-1}), c(t_i)),$$

dove il sup é preso su tutti gli m e $0 = t_1 < \dots < t_m = 1$.

Si può mostrare che esiste un cammino c tale che $L(c) = W_p(\mu_0, \mu_1)$, in altre parole $(\mu_t : 0 \leq t \leq 1)$ é la geodetica che connette μ_0 e μ_1 . Inoltre vale anche che $\mu_t = F_{t\#}\pi$, dove $\pi \in \Pi_p^{opt}(\mu_0, \mu_1)$ e $F_t(x, y) := (1 - t)x + ty$.

Un esempio é mostrato nell'immagine seguente.



Figura 2.2: Trasformazione di un'immagine in un'altra attraverso la geodetica di Wasserstein.

Capitolo 3

Propagazione del Chaos

Quando Boltzmann pubblicó il suo articolo piú famoso [Bol66], lo studio di grandi sistemi di particelle interagenti era motivato dalla modellizzazione microscopica dei sistemi termodinamici. Sebbene fosse tutt'altro che un'idea accettata all'epoca, Boltzmann postuló che, poiché un volume macroscopico di gas contiene una miriade di particelle elementari, é inutile tenere traccia di ogni singola particella e si dovrebbe piuttosto cercarne una descrizione statistica. Derivó cosí l'equazione che ora porta il suo nome in base a un presupposto cruciale chiamato *chaos molecolare* che, informalmente, si traduce nell'idea che, nonostante la moltitudine di interazioni, due particelle casualmente scelte dovrebbero essere statisticamente indipendenti quando il numero totale di particelle cresce a $+\infty$.

Kac fu il primo a dare una rigorosa definizione matematica di chaos [Kac56], introducendo l'idea che in sistemi che si evolvono nel tempo, anche il chaos dovesse propagarsi, una proprietà che venne poi chiamata *propagazione del Chaos*. Poco tempo dopo Kac, McKean introdusse una classe di modelli di diffusione che soddisfacevano la proprietà di propagazione del chaos precedentemente introdotta [McK69].

In questo capitolo, studieremo i processi di diffusione non lineare, introdotti da McKean negli anni '60 e riusciremo a concludere che in un sistema di N particelle interagenti, prese casualmente n particelle, allora quest'ultime si

comporteranno come se fossero indipendenti quando $N \rightarrow +\infty$.

3.1 Processi di diffusione non lineare di McKean

Un processo di diffusione non lineare di McKean é descritto da un'equazione differenziale stocastica in cui la legge del processo di diffusione é essa stessa un fattore incognito nell'equazione. Per un processo n -dimensionale, la forma dell'equazione é

$$\begin{cases} X_t = X_0 + \int_0^t \bar{b}^{\mu_s}(X_s) ds + \int_0^t \bar{\sigma}^{\mu_s}(X_s) dW_s \\ X_t \sim \mu_t \end{cases}, \quad t \geq 0 \quad (3.1)$$

dove utilizziamo la seguente notazione:

$$\bar{\varphi}^\nu(x) := \int_{\mathbb{R}^n} \varphi(x, y) d\nu(y)$$

con $\nu \in \mathcal{P}(\mathbb{R}^n)$ e φ una funzione misurabile definita su $\mathbb{R}^n \times \mathbb{R}^n$.

W é un moto Browniano, la condizione iniziale X_0 é una variabile aleatoria indipendente da W , e le due funzioni $b: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ e $\sigma: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ sono funzioni misurabili dette rispettivamente *drift* e *coefficiente di diffusione*, dove indichiamo con $\mathbb{R}^{n \times n}$ lo spazio delle matrici $n \times n$ a coefficienti reali.

In sostanza l'equazione (3.1) puó essere vista, in determinate condizioni, come limite di N equazioni differenziali stocastiche:

$$\begin{aligned} X_t^{(i,N)} = X_0^{(i,N)} + \frac{1}{N} \sum_{j=1}^N \int_0^t b(X_s^{(i,N)}, X_s^{(j,N)}) ds \\ + \frac{1}{N} \sum_{j=1}^N \int_0^t \sigma(X_s^{(i,N)}, X_s^{(j,N)}) dW_s^i, \end{aligned} \quad (3.2)$$

con $t \geq 0$, $1 \leq i \leq N$. Qui ogni equazione differenziale stocastica é relativa ad un moto Browniano W^i indipendente dalle altre.

Osservazione 3.1.1. Il sistema (3.2) é interpretabile come una descrizione su scala microscopica di particelle fisiche che interagiscono nello stesso ambiente, mentre il modello iniziale (3.1) fornisce la prospettiva macroscopica di una singola particella rappresentativa. Le diverse prospettive sono molto utili, soprattutto nelle varie applicazioni, per comprendere come la media microscopica operi su scala macroscopica.

Si può mostrare che, sotto particolari ipotesi su drift e coefficiente di diffusione, esiste ed é unica la soluzione al processo di diffusione non lineare (3.1). Prima di fare ciò diamo però la seguente definizione.

Definizione 3.1.1 (Funzione globalmente Lipschitziana). Dati $(\mathcal{X}, d_x), (\mathcal{Y}, d_y)$ due spazi metrici e data $f: \mathcal{X} \rightarrow \mathcal{Y}$, f si dice globalmente Lipschitziana se $\exists L > 0$ tale che

$$d_y(f(x) - f(y)) \leq L d_x(x - y) \quad \forall x, y \in \mathcal{X}.$$

Enunciamo ora due risultati che serviranno per la dimostrazione del successivo teorema (3.1.2).

Lemma 3.1.1 (Lemma di Grönwall). Sia $u \in L^1([0, T])$ che verifica la seguente stima integrale

$$u(t) \leq a + b \int_0^t u(s) ds, \quad t \in [0, T],$$

per certi $a, b \in \mathbb{R}_{\geq 0}$. Allora si ha

$$u(t) \leq a e^{bt}, \quad t \in [0, T].$$

Si veda Lemma 18.4 [Kal21].

Teorema 3.1.1 (Convergenza dominata). Sia $(X_n)_{n \in \mathbb{N}}$ successione di variabili aleatorie su $(\Omega, \mathcal{F}, P)$ tale che $X_n \rightarrow X$ P -q.c. e $|X_n| \leq Y \in L^1(\Omega, P)$ per ogni n , allora

$$\lim_{n \rightarrow +\infty} \int_{\Omega} X_n dP = \int_{\Omega} X dP.$$

Si veda Teorema 2.2.12 [Pas24].

Possiamo ora enunciare e dimostrare il seguente teorema.

Teorema 3.1.2 (Esistenza e Unicit ). Supponiamo che $b : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ sia limitata e globalmente Lipschitziana e $\sigma(x, y) = \text{Id}$. Allora esiste ed   unica la soluzione all'equazione della diffusione non lineare

$$\begin{cases} X_t = X_0 + \int_0^t \bar{b}^{\mu_s}(X_s) ds + \int_0^t \bar{\sigma}^{\mu_s}(X_s) dW_s \\ X_t \sim \mu_t \end{cases}, \quad t \geq 0.$$

Procediamo ora con la dimostrazione del teorema.

Dimostrazione. La dimostrazione si baser  su un argomento di punto fisso. Perci  introduciamo la distanza di Wasserstein $W_T(\cdot, \cdot)$, definita sull'insieme $\mathcal{P}(C_T)$ delle misure di probabilit  sulle traiettorie continue $C_T := C([0, T])$:

$$W_T(\nu_1, \nu_2) := \inf_{\nu \in \Pi(\nu_1, \nu_2)} \left[\int_{C_T \times C_T} \left(\sup_{s \leq T} |\omega_{1,s} - \omega_{2,s}| \wedge 1 \right) d\nu(\omega_1, \omega_2) \right],$$

con la notazione $(\sup_{s \leq T} |\omega_{1,s} - \omega_{2,s}| \wedge 1) := \min\{\sup_{s \leq T} |\omega_{1,s} - \omega_{2,s}|, 1\}$. Grazie al teorema (2.3.5), questa metrica rende completo $\mathcal{P}(C_T)$ rispetto alla topologia della convergenza in legge.

Esistenza.

Definiamo ora $\Phi : \mathcal{P}(C_T) \rightarrow \mathcal{P}(C_T)$, tale che, data $\nu \in \mathcal{P}(C_T)$, $\Phi(\nu)$ sia la legge della soluzione di

$$X_t = X_0 + \int_0^t \bar{b}^{\nu_s}(X_s) ds + W_t.$$

Ora, date $\nu_1, \nu_2 \in \mathcal{P}(C_T)$, compariamo le due equazioni differenziali stocastiche X_1 e X_2 :

$$\begin{aligned} & \sup_{s \leq T} |X_{1,s} - X_{2,s}| \\ & \leq \int_0^T \left| \int_{C_T} b(X_{1,s}, \omega_{1,s}) d\nu_1(\omega_1) - \int_{C_T} b(X_{2,s}, \omega_{2,s}) d\nu_2(\omega_2) \right| ds \\ & \leq \int_0^T \left[L_b |X_{1,s} - X_{2,s}| + \int_{C_T \times C_T} (L_b |\omega_{1,s} - \omega_{2,s}| \wedge 2 \|b\|_\infty) d\nu(\omega_1, \omega_2) \right] ds, \end{aligned}$$

dove L_b é la costante di Lipschitzianità di b .

Posto $K := \max(L_b, 2\|b\|_\infty)$, prendendo l'inf su tutte le congiunte ν e grazie all'ausilio del lemma (3.1.1), si ha

$$\sup_{s \leq t} |X_{1,s} - X_{2,s}| \leq K \int_0^T [|X_{1,s} - X_{2,s}| + W_s(\nu_1, \nu_2)] ds \leq Ke^{KT} \int_0^T W_s(\nu_1, \nu_2) ds,$$

da cui deduciamo un controllo su $W_t(\Phi(\nu_1), \Phi(\nu_2))$, infatti

$$W_T(\Phi(\nu_1), \Phi(\nu_2)) \leq Ke^{KT} \int_0^t W_s(\nu_1, \nu_2) ds.$$

Sia ora $\nu \in \mathcal{P}(C_T)$, costruiamo la seguente successione ricorsiva

$$\begin{cases} \nu^{(0)} = \nu \\ \nu^{(n)} = \Phi(\nu^{(n-1)}) \end{cases}$$

e notiamo che

$$W_T(\nu^{(1)}, \nu^{(2)}) = W_T(\Phi(\nu), \Phi(\nu^{(1)})) \leq Ke^{KT} \int_0^t W_s(\nu, \nu^{(1)}) ds.$$

Per la seconda iterata otterremo

$$\begin{aligned} W_T(\nu^{(2)}, \nu^{(3)}) &\leq Ke^{KT} \int_0^T W_s(\nu^{(1)}, \nu^{(2)}) ds \\ &\leq Ke^{KT} \int_0^T Ke^{KT} \int_0^s W_r(\nu, \nu^{(1)}) dr ds \\ &\leq (Ke^{KT})^2 W_T(\nu, \nu^{(1)}) \int_0^T \int_0^s dr ds \\ &= \frac{(Ke^{KT})^2}{2} W_T(\nu, \nu^{(1)}), \end{aligned}$$

quindi, per ricorsione, deduciamo che

$$W_T(\nu^{(n)}, \nu^{(n+1)}) \leq \frac{(Ke^{KT})^n}{n!} W_T(\nu, \nu^{(1)}).$$

Notiamo ora che

$$\sum_{n=0}^{\infty} W_T(\nu^{(n)}, \nu^{(n+1)}) \leq W_T(\nu, \nu^{(1)}) \sum_{n=0}^{\infty} \frac{(Ke^{KT})^n}{n!} = W_T(\nu, \nu^{(1)}) e^{Ke^{KT}},$$

quindi la serie converge e di conseguenza $\{(\nu^n)\}_{n \in \mathbb{N}}$ é una successione di Cauchy. Dato che $(\mathcal{P}(C_T), W_T(C_T))$ é completo, la successione $\{(\nu^n)\}_{n \in \mathbb{N}}$ converge a $\nu^* \in \mathcal{P}(C_T)$.

Mostriamo ora che Φ é continua. Sia $(\nu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(C_T)$ successione e sia $\nu \in \mathcal{P}(C_T)$ tale che $\nu_n \xrightarrow{d} \nu$. Confrontiamo ora le due equazioni stocastiche cosí definite

$$\begin{aligned} X_t^{\nu_n} &:= X_0 + \int_0^t \bar{b}^{(\nu_n)s}(X_s^{\nu_n}) ds + W_t, \\ X_t^\nu &:= X_0 + \int_0^t \bar{b}^{\nu s}(X_s^\nu) ds + W_t. \end{aligned}$$

Sia $\delta_t^n := |X_t^{\nu_n} - X_t^\nu|$, allora

$$\begin{aligned} \delta_t^n &\leq \int_0^t |\bar{b}^{(\nu_n)s}(X_s^{\nu_n}) - \bar{b}^{\nu s}(X_s^\nu)| ds \\ &\leq \int_0^t |\bar{b}^{(\nu_n)s}(X_s^{\nu_n}) - \bar{b}^{(\nu_n)s}(X_s^\nu)| ds + \int_0^t |\bar{b}^{(\nu_n)s}(X_s^\nu) - \bar{b}^{\nu s}(X_s^\nu)| ds. \end{aligned}$$

Notiamo che

$$\begin{aligned} |\bar{b}^{(\nu_n)s}(X_s^{\nu_n}) - \bar{b}^{(\nu_n)s}(X_s^\nu)| &\leq \int_{\mathbb{R}^n} |b(X_s^{\nu_n}, y) - b(X_s^\nu, y)| d\nu_n(y) \\ &\leq L |X_s^{\nu_n} - X_s^\nu|, \end{aligned}$$

perció

$$\delta_t^n \leq L \int_0^t \delta_s^n ds.$$

Inoltre abbiamo che

$$\begin{aligned} |\bar{b}^{(\nu_n)s}(X_s^\nu) - \bar{b}^{\nu s}(X_s^\nu)| &\leq \int_{\mathbb{R}^n} |b(X_s^\nu, y)| |d\nu_n - d\nu|(y) \leq \|b\|_\infty \quad \text{e} \\ \int_{\mathbb{R}^n} |b(X_s^\nu, y)| |d\nu_n - d\nu|(y) &\xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

quindi, per il teorema della convergenza dominata

$$\epsilon_n := \int_0^t |\bar{b}^{(\nu_n)s}(X_s^\nu) - \bar{b}^{\nu s}(X_s^\nu)| ds \xrightarrow{n \rightarrow \infty} 0.$$

Riassumendo si ha che

$$\delta_t^n \leq L \int_0^t \delta_s^n ds + \epsilon_n,$$

quindi per il lemma (3.1.1) $\delta_t^n \leq \epsilon_n e^{Lt} \xrightarrow{n \rightarrow \infty} 0$. Da cui possiamo concludere che $\forall \epsilon > 0 \quad P(\delta_t^n > \epsilon) \leq P(\epsilon_n e^{Lt} > \epsilon) \xrightarrow{n \rightarrow \infty} 0$, perché $\epsilon_n e^{Lt} \xrightarrow{n \rightarrow \infty} 0$ in modo deterministico. Perciò $\Phi(\nu_n) \xrightarrow{P} \Phi(\nu) \Rightarrow \Phi(\nu_n) \xrightarrow{d} \Phi(\nu)$, quindi Φ é continua.

Ora, per continuità di Φ , avremo che

$$\begin{aligned} \Phi(\nu^{(n)}) &\xrightarrow{n \rightarrow \infty} \Phi(\nu^*), \\ \Phi(\nu^{(n)}) = \nu^{(n+1)} &\xrightarrow{n \rightarrow \infty} \nu^*. \end{aligned}$$

Perciò esiste $\nu^* \in \mathcal{P}(C_T)$ tale che $\Phi(\nu^*) = \nu^*$, ora mostriamo la sua unicità.

Unicità.

Supponiamo per assurdo che esistano $\nu^*, \mu^* \in \mathcal{P}(C_T)$, $\nu^* \neq \mu^*$ punti fissi, cioè tali che $\Phi(\nu^*) = \nu^*$ e $\Phi(\mu^*) = \mu^*$. Allora avremmo che

$$W_t(\nu^*, \mu^*) = W_t(\Phi(\nu^*), \Phi(\mu^*)) \leq K e^{KT} \int_0^t W_s(\nu^*, \mu^*) ds \quad \forall t \in [0, T]$$

e per il lemma (3.1.1)

$$W_t(\nu^*, \mu^*) = 0 \quad \forall t \in [0, T],$$

in particolare $W_T(\nu^*, \mu^*) = 0$, quindi $\nu^* = \mu^*$. □

3.2 Convergenza del sistema di diffusione

Sotto le stesse ipotesi considerate precedentemente, mostreremo come ogni equazione differenziale stocastica $X^{(i,N)}$ del sistema (3.2) converga all'equazione differenziale stocastica $X^{(i)}$ (3.1) spiegando la transizione tra scala macroscopica e scala microscopica menzionata a inizio capitolo. Mostriamo quindi il seguente teorema.

Teorema 3.2.1. Consideriamo le ipotesi del teorema precedente (3.1.2) e denotiamo con μ_t la legge comune di $(X_t^{(i)})_i$ soluzione di

$$X_t^{(i)} = X_0^{(i)} + \int_0^t \bar{b}^{\mu_s}(X_s^{(i)}) ds + W_t^i, \quad t \geq 0, \quad 1 \leq i \leq N,$$

dove $(X_0^{(i)})_i$ sono variabili aleatorie indipendenti di legge μ_0 data e $(W^i)_i$ sono moti Browniani indipendenti.

Per $t \geq 0$ e $1 \leq i \leq N$, poniamo

$$X_t^{(i,N)} = X_0^{(i)} + \frac{1}{N} \sum_{j=1}^N \int_0^t b(X_s^{(i,N)}, X_s^{(j,N)}) ds + W_t^i.$$

Allora abbiamo

$$\sup_{1 \leq i \leq N} \mathbb{E} \left(\sup_{t \leq T} |X_t^{(i,N)} - X_t^{(i)}| \right) = O \left(\frac{1}{\sqrt{N}} \right) \quad \text{per } N \rightarrow +\infty.$$

Dimostrazione. Per la disuguaglianza triangolare abbiamo che

$$\begin{aligned} |X^{(i,N)} - X^{(i)}|_T &:= \sup_{t \leq T} |X_t^{(i,N)} - X_t^{(i)}| \\ &\leq \int_0^T \frac{1}{N} \sum_{j=1}^N |b(X_s^{(i,N)}, X_s^{(j,N)}) - b(X_s^{(i)}, X_s^{(j,N)})| ds \\ &\quad + \int_0^T \frac{1}{N} \sum_{j=1}^N |b(X_s^{(i)}, X_s^{(j,N)}) - b(X_s^{(i)}, X_s^{(j)})| ds \\ &\quad + \int_0^T \left| \frac{1}{N} \sum_{j=1}^N b(X_s^{(i)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(i)}) \right| ds \end{aligned}$$

$$\begin{aligned} &\leq L_b \int_0^T |X_s^{(i,N)}, X_s^{(i)}| ds + L_b \int_0^T \frac{1}{N} \sum_{j=1}^N |X_s^{(j,N)} - X_s^{(j)}| ds \\ &\quad + \int_0^T \left| \frac{1}{N} \sum_{j=1}^N b(X_s^{(i)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(i)}) \right| ds \end{aligned}$$

Per simmetria della legge, $\mathbb{E}[|X^{(1,N)} - X^{(1)}|_T]$ non dipende piú da i , e quindi, sfruttando il lemma (3.1.1), deduciamo che

$$\begin{aligned} &\mathbb{E}[|X^{(1,N)} - X^{(1)}|_T] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[|X^{(i,N)} - X^{(i)}|_T] \\ &\leq 2L_b \int_0^T \frac{1}{N} \sum_{j=1}^N \mathbb{E}[|X_s^{(j,N)} - X_s^{(j)}|_T] ds + \int_0^T \mathbb{E} \left[\left| \frac{1}{N} \sum_{j=1}^N b(X_s^{(1)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(1)}) \right| \right] ds \\ &\leq e^{2L_b T} \int_0^T \mathbb{E} \left[\left| \frac{1}{N} \sum_{j=1}^N b(X_s^{(1)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(1)}) \right| \right] ds. \end{aligned}$$

Ma, per la disuguaglianza di Cauchy-Schwarz,

$$\begin{aligned} &\left\{ \mathbb{E} \left[\left| \frac{1}{N} \sum_{j=1}^N b(X_s^{(1)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(1)}) \right| \right] \right\}^2 \\ &\leq \frac{1}{N^2} \mathbb{E} \left[\sum_{j,k=1}^N (b(X_s^{(1)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(1)}))^\perp (b(X_s^{(1)}, X_s^{(k)}) - \bar{b}^{\mu_s}(X_s^{(1)})) \right]. \end{aligned}$$

Possiamo notare che i termini per $j \neq k$ e $j \neq 1$ sono nulli, perché da una parte la legge comune di $X_s^{(j)}$ é μ_s e $\bar{b}^{\mu_s}$ é la media di b rispetto alla seconda variabile (cioé $\mathbb{E}[b(x, X_s^{(j)})] = \bar{b}^{\mu_s}(x)$), e d'altra parte $X^{(1)}$ e $X^{(j)}$ sono indipendenti. I termini per $j \neq k$ e $k \neq 1$ si annullano per la stessa ragione. Quindi, sfruttando il fatto che b é limitata, deduciamo che

$$\frac{1}{N^2} \mathbb{E} \left[\sum_{j,k=1}^N (b(X_s^{(1)}, X_s^{(j)}) - \bar{b}^{\mu_s}(X_s^{(1)}))^\perp (b(X_s^{(1)}, X_s^{(k)}) - \bar{b}^{\mu_s}(X_s^{(1)})) \right] \leq \frac{4\|b\|_\infty^2}{N},$$

e in conclusione

$$\mathbb{E}[|X^{(i,N)} - X^{(i)}|_T] \leq T e^{2L_b T} \frac{2\|b\|_\infty}{\sqrt{N}}.$$

□

Possiamo quindi enunciare il seguente risultato.

Teorema 3.2.2. Sotto le ipotesi del teorema precedente (3.2.1), la legge di $(X^{(i_1,N)}, \dots, X^{(i_n,N)})$ (per tutte le n-uple $1 \leq i_1 < \dots < i_n \leq N$) si comporta asintoticamente (cioé quando $N \rightarrow +\infty$) come un n-upla di particelle indipendenti.

Ne troviamo varie dimostrazioni in [CD22].

Osservazione 3.2.1. Ciò ci permette di concludere che nella realtà, al di là del comportamento di una particella specifica, possiamo descrivere la dipendenza limite di un gruppo di n particelle quando il numero totale N di particelle nell'ambiente tende a infinito. Pur interagendo, alla fine si comportano in modo indipendente.

3.3 Applicazioni

I sistemi di particelle interagenti sono ormai presenti in molte applicazioni e, recentemente, è in crescita il numero di modelli in matematica applicata in cui la nozione di chaos gioca un ruolo centrale. Nella seguente sezione descriveremo alcune applicazioni concrete in cui entrano in gioco i processi di diffusione non lineare di McKean.

Fenomeni di aggregazione nelle popolazioni

I processi di diffusione non lineare vengono utilizzati anche per descrivere il comportamento aggregativo della formica schiava *Polyergus rufescens*, una specie parassita sociale che dipende da altre formiche per svolgere compiti vitali. Le colonie sono composte solo da soldati incapaci di occuparsi della prole o del nido, perciò rapiscono larve o pupe di altre specie per allevarle come schiave.

Le incursioni per ottenere nuovi schiavi iniziano con un'esplorazione da parte di alcune formiche, seguita da un rapido raduno di centinaia di soldati richiamati da segnali chimici. L'esercito si muove in formazione compatta e ordinata fino al nido bersaglio, dove attacca, uccide le difensori e rapisce larve e pupe. Il ritorno, invece, avviene in modo disorganizzato, con formiche che si muovono singolarmente. Il modello che è stato proposto per spiegare l'aggregazione trasversale dell'esercito, trascurando l'influenza della scia chimica nella direzione del movimento, è il seguente.

La variazione della posizione delle formiche è data da un sistema di equazioni del seguente tipo

$$dX = f(X, t) dt + g(X, t) dW_t,$$

dove $f(X, t)$ e $g(X, t)$ sono funzioni deterministiche e $\{W_t\}$ è una famiglia di moti Browniani. In particolare si considera $g(X, t) = \sigma(N)$, affinché g dipenda solo dal numero N di "particelle" e non dalla loro distribuzione. Ciò può essere interpretato dicendo che si considera solo il movimento casuale di ogni

formica in modo tale che i cammini liberi possibili si riducono all'aumentare di N .

Da notare come i processi di diffusione non lineare (3.1) siano alla base del modello precedente e siano fondamentali per la sua descrizione. Si veda piú approfonditamente [BCM00].

Stabilità del sistema interbancario e rischio sistemico

Si consideri un sistema finanziario con N banche; le tesorerie delle banche dipendono l'una dall'altra perché, in caso di necessità di finanziamento a breve termine, possono prestarsi denaro a vicenda. Sia $X^{(i,N)}$ il logaritmo della riserva monetaria della banca i . Semplicemente modellando il sistema di prestito interbancario, come si può vedere in [FL13], otteniamo la seguente dinamica:

$$X_t^{(i,N)} = X_0^{(i,N)} + \int_0^t \frac{\alpha}{N} \sum_{j=1}^N (X_s^{(j,N)} - X_s^{(i,N)}) ds + \sigma W_t^i, \quad t \geq 0, \quad 1 \leq i \leq N,$$

Il fattore α viene interpretato come un *tasso di prestito interbancario autorizzato*. Questo parametro può essere modificato dalle autorità di regolamentazione per garantire una maggiore stabilità del sistema. Si noti che l'equazione precedente è effettivamente della forma (3.2).

Equazione di Burgers

In matematica l'equazione di Burgers [Bur48] è un'equazione differenziale alle derivate parziali che si presenta nella seguente forma.

Per una data funzione $f(x, y)$ di due variabili, la forma generale dell'equazione di Burgers è

$$\frac{\partial f}{\partial y} + f \frac{\partial f}{\partial x} - a \frac{\partial^2 f}{\partial x^2} = 0 \quad a \in \mathbb{R}$$

e prende il nome di forma *avvettiva*. Quando $a = 0$, l'equazione diventa *inviscida*

$$\frac{\partial f}{\partial y} + f \frac{\partial f}{\partial x} = 0$$

ed é un prototipo per equazioni per le quali la soluzione puó sviluppare discontinuitá come una funzione costante a tratti.

In generale l'equazione é fondamentale per la meccanica dei fluidi e utile anche in molte aree della matematica applicata, quali la modellazione della gasdinamica e del flusso del traffico.

L'equazione che interessa la trattazione é della seguente forma:

$$\frac{\partial v}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 v}{\partial x^2} - v \frac{\partial v}{\partial x}, \quad (3.3)$$

dove v rappresenta una velocitá e $\nu = \frac{1}{2}\sigma^2$ é il coefficiente di viscositá. Allora v si interpreta come la funzione di ripartizione della misura μ associata alla diffusione non lineare con $n = 1$, $\sigma(x, y) = \sigma$ e $b(x, y) = \mathbb{1}_{x>y}$ (la funzione di Heaviside).

Infatti, ponendo

$$u(t, x) = \int_{-\infty}^x d\mu_t(y),$$

si puó verificare formalmente che u risolve la (3.3).

Bibliografia

- [AG21] Marco De Angelis and Ander Gray. Why the 1-wasserstein distance is the area between the two marginal cdfs, 2021.
- [BCM00] Silvia Boi, Vincenzo Capasso, and Daniela Morale. Modeling the aggregative behavior of ants of the species *polyergus rufescens*. volume 1, pages 163–176. 2000. Spatial heterogeneity in ecological models (Alcalá de Henares, 1998).
- [Bol66] L. Boltzmann. Further studies on the thermal equilibrium of gas molecules. In S. G. Brush, editor, *Kinetic Theory, Volume 2*, pages 88–174. Pergamon Press, Oxford, 1966. Translation of: *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen*, Sitzungsberichte der Akademie der Wissenschaften, 66 (1872), 275–370.
- [Bur48] J. M. Burgers. A mathematical model illustrating the theory of turbulence. In Richard von Mises and Theodore von Kármán, editors, *Advances in Applied Mechanics*, pages 171–199. Academic Press, New York, 1948.
- [CD18] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications. I*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018. Mean field FBSDEs, control, and games.

- [CD22] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *Kinet. Relat. Models*, 15(6):895–1015, 2022.
- [FL13] Jean-Pierre Fouque and Joseph A. Langsam, editors. *Handbook on systemic risk*. Cambridge University Press, Cambridge, 2013.
- [Gob16] Emmanuel Gobet. *Monte-Carlo methods and stochastic processes*. CRC Press, Boca Raton, FL, 2016. From linear to non-linear.
- [HM14] Maxime Hauray and Stéphane Mischler. On Kac’s chaos and related problems. *J. Funct. Anal.*, 266(10):6055–6157, 2014.
- [Kac56] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. Univ. California Press, Berkeley-Los Angeles, Calif., 1956.
- [Kal21] Olav Kallenberg. *Foundations of modern probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer, Cham, third edition, [2021] ©2021.
- [Kle14] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.
- [McK69] H. P. McKean. Propagation of chaos for a class of non-linear parabolic equations. In A. K. Aziz, editor, *Lecture Series in Differential Equations, Volume 2*, volume 19 of *Van Nostrand Mathematical Studies*, pages 177–194. Van Nostrand Reinhold Company, New York, 1969.
- [Pas24] Andrea Pascucci. *Probability theory. Vol. I. Random variables and distributions*, volume 165 of *Unitext*. Springer, Cham, [2024] ©2024. Translated from the 2020 Italian, *La Matematica per il 3+2*.

-
- [Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.