



UNIVERSITÀ DI PISA

CORSO DI LAUREA IN MATEMATICA

# Il problema dei momenti e una sua applicazione statistica

TESI DI LAUREA TRIENNALE  
IN MATEMATICA

ANNO ACCADEMICO 2021/2022

RELATORE:

**Dario Trevisan**  
UNIVERSITÀ DI PISA

CANDIDATO:

**Emanuele Pardini**



# Indice

<b>Abstract</b>	<b>i</b>
<b>1 Problema dei momenti: Esistenza</b>	<b>1</b>
1.1 Problema dei momenti e funzionali . . . . .	1
1.2 Spazi di Choquet e teorema di Haviland . . . . .	3
1.3 Polinomi positivi su intervalli di $\mathbb{R}$ . . . . .	9
1.4 Soluzione al problema dei momenti su intervalli di $\mathbb{R}$ . . . . .	17
<b>2 Problema dei momenti: Unicità</b>	<b>25</b>
2.1 Teorema di Denjoy-Carleman . . . . .	25
2.2 Unicità nel problema dei momenti su $\mathbb{R}$ . . . . .	34
2.3 Teorema di Petersen per il caso multidimensionale . . . . .	41
<b>3 Un utilizzo statistico dell'unicità</b>	<b>45</b>
3.1 Presentazione del problema e metodi utilizzati . . . . .	45
3.1.1 Primo metodo . . . . .	46
3.1.2 Secondo metodo . . . . .	47
3.2 Risultati ottenuti . . . . .	49
3.3 Codici utilizzati . . . . .	56
<b>Ringraziamenti</b>	<b>65</b>
<b>Bibliografia</b>	<b>67</b>



# Abstract

Data una misura  $\mu : \mathcal{B}(\mathbb{R}^n) \rightarrow [0, +\infty]$  ed un vettore  $\alpha = (\alpha_i)_{i=1, \dots, n} \in \mathbb{N}^n$  la quantità

$$\int_{\mathbb{R}^n} x^\alpha d\mu(x)$$

in cui  $x^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ , con  $x = (x_i)_{i=1, \dots, n}$ , è chiamata **momento  $\alpha$ -esimo** di  $\mu$ .

I momenti appaiono in modo naturale in vari campi della matematica e della fisica. Ad esempio data una variabile aleatoria reale  $X$ , indicando con  $\mathbb{P}_X$  la legge di  $X$ , il valore atteso di  $X^k$  non è altro che il momento  $k$ -esimo di  $\mathbb{P}_X$ , ossia

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k d\mathbb{P}_X(x).$$

Oppure dato un corpo rigido  $\mathcal{C}$  in  $\mathbb{R}^3$  con densità di massa  $\rho(x_1, x_2, x_3)$ , il momento di inerzia rispetto al primo asse è

$$I_1 = \int_{\mathbb{R}^3} (x_2^2 + x_3^2) \rho(x_1, x_2, x_3) dx_1 dx_2 dx_3$$

che sarebbe la somma dei momenti  $(0, 2, 0)$ -esimo e  $(0, 0, 2)$ -esimo della misura di densità  $\rho$  rispetto alla misura di Lebesgue su  $\mathbb{R}^3$ .

Nella sua forma più classica il problema dei momenti è formulabile come segue:

**Problema 1 (dei momenti su  $\mathbb{R}^d$ ):** *Siano  $K$  un sottoinsieme chiuso di  $\mathbb{R}^d$  ed  $(s_\alpha)_{\alpha \in \mathbb{N}^d}$  una successione di numeri reali. Sotto quali condizioni esiste una misura di Radon  $\mu$  concentrata su  $K$  t.c.  $\forall \alpha \in \mathbb{N}^d$   $s_\alpha = \int_K x^\alpha d\mu(x)$ ?  
E, se una tale misura esiste, è unica?*

Questo problema fu ufficialmente formulato per la prima volta alla fine del diciannovesimo secolo dall'illustre matematico olandese T. J. Stieltjes (1894) che lo trattò nel caso in cui  $K = [0, +\infty)$ . Successivamente i casi con  $K = \mathbb{R}$  o  $K$  intervallo compatto furono trattati dai matematici H. Hamburger (1920) e F. Hausdorff (1920). Ulteriori non meno importanti contributi furono dati da M. Riesz, T. Carleman ed altri.

Nel primo capitolo studieremo l'esistenza di una misura con momenti assegnati. Ri-formuleremo il problema in termini di funzionali lineari agenti sullo spazio dei polinomi

a coefficienti reali, cosa che ci porterà ad un primo importante risultato sull'esistenza di una tale misura. Successivamente sarà fatto un approfondito studio del problema su  $\mathbb{R}$  in tutti i suoi casi.

Nel secondo capitolo verranno studiati criteri per stabilire quando vale l'unicità, ossia si stabiliranno delle condizioni sufficienti a concludere che una misura è caratterizzata dai propri momenti. Nella prima parte si studierà il caso di misure sulla retta reale, mentre in una seconda parte si faranno dei cenni al caso multidimensionale.

Tali cenni saranno fondamentali nel terzo capitolo, in cui si presenterà un'applicazione della teoria ad un problema statistico che può essere formulato come segue:

**Problema 2:** *Dati due vettori aleatori  $X = (X_i)_{i=1}^d$  e  $Y = (Y_i)_{i=1}^d$  ed un terzo vettore aleatorio  $Q = (Q_i)_{i=1}^d$ , vogliamo trovare una matrice simmetrica e definita positiva  $A \in \mathbb{R}^{d \times d}$  t.c.  $\mathbb{P}_{X-AY}$  sia più simile possibile a  $\mathbb{P}_Q$ .*

in cui l'espressione "più simile possibile" si tradurrà nel minimizzare determinate funzioni costo legate ai momenti delle due leggi  $\mathbb{P}_{X-AY}$  e  $\mathbb{P}_Q$  al variare di  $A$  nella varietà delle matrici simmetriche definite positive  $d \times d$ . Tali condizioni sulla matrice  $A$  sono motivate dall'applicazione di tale problema all'analisi di dati provenienti da un problema di ingegneria robotica. Nonostante questo noi ci limiteremo ad utilizzare dati simulati numericamente. La strategia e le tecniche numeriche utilizzate per approssimare la matrice  $A$  verranno esposte e verranno analizzati alcuni risultati ottenuti.

# Problema dei momenti: Esistenza

In questo primo capitolo affronteremo la parte del problema dei momenti riguardante l'esistenza di una misura soluzione.

Nella prima parte riformuleremo il problema in termini di funzionali lineari, presenteremo il fondamentale teorema di Haviland e lo useremo per fornire un primo criterio per l'esistenza di una soluzione nel caso generale di un chiuso  $K \subset \mathbb{R}^d$ .

Successivamente ci restringeremo a studiare il problema unidimensionale e dimostreremo diversi criteri per i vari casi di questo problema.

## 1.1 Problema dei momenti e funzionali

La parte del problema dei momenti riguardante l'esistenza di una misura soluzione è strettamente collegata alla rappresentazione integrale di particolari funzionali. In questa sezione si descriverà questa relazione e dimostreremo i risultati su cui si baserà gran parte dello studio successivo.

Per completezza richiamiamo la definizione di misura di Radon.

**Definizione 1.1.1:** Sia  $X$  uno spazio topologico. Indichiamo con  $\mathcal{B}(X)$  la  $\sigma$ -algebra dei boreliani di  $X$ . Una **misura di Radon** su  $X$  è una misura boreliana  $\mu : \mathcal{B}(X) \rightarrow [0, +\infty]$  t.c.

- (1)  $\forall C \subset X$  compatto  $\mu(C) < +\infty$ ;
- (2)  $\forall A \subset X$  aperto  $\mu(A) = \sup\{\mu(C) \mid C \text{ compatto, } C \subset A\}$ ;
- (3)  $\forall B \in \mathcal{B}(X)$   $\mu(B) = \inf\{\mu(A) \mid A \text{ aperto, } B \subset A\}$ .

Inoltre diremo che la misura  $\mu$  è **concentrata** su  $K \subset X$  se  $\forall B \in \mathcal{B}(X)$  vale  $\mu(B) = \mu(B \cap K)$ .

A proposito di misure di Radon ricordiamo che vale il seguente teorema.

**Teorema 1.1.2:** *Sia  $X$  uno spazio topologico localmente compatto di Hausdorff e secondo numerabile, allora ogni misura boreliana su  $X$  finita sui compatti di  $X$  è di Radon.*

*Dimostrazione.* Si veda ad esempio [3, Theorem 7.8]. □

**Osservazione 1.1.3:** Dal Teorema 1.1.2 segue, in particolare, che ogni misura boreliana su  $\mathbb{R}^d$  finita, o anche solamente finita sui compatti, è di Radon.

In quanto segue, se non specificato diversamente,  $X$  sarà uno spazio topologico localmente compatto di Hausdorff.

Consideriamo il seguente problema di rappresentazione.

**Problema 3:** *Sia  $E$  uno sottospazio vettoriale di  $C(X, \mathbb{R})$ . Dati  $K$  un sottoinsieme chiuso di  $X$  ed  $L : E \rightarrow \mathbb{R}$  un funzionale lineare. Esiste una misura di Radon  $\mu$  concentrata su  $K$  t.c.*

$$L(f) = \int_X f d\mu \quad \forall f \in E? \tag{1.1}$$

In caso di risposta affermativa alla domanda appena posta, il funzionale  $L$  sarà detto **K-rappresentabile**, o solo **rappresentabile** se  $K = X$ . Inoltre diremo che  $\mu$  **rappresenta**  $L$ .

**Definizione 1.1.4:** Siano  $E$  uno sottospazio vettoriale di  $C(X, \mathbb{R})$ ,  $\mathcal{E} = \{e_i\}_{i \in I}$ , con  $I$  insieme di indici, una sua base ed  $\mathfrak{s} = (s_i)_{i \in I}$  una famiglia di numeri reali indicizzati su  $I$ . Chiamiamo **funzionale di Riesz** associato ad  $\mathfrak{s}$  rispetto alla base  $\mathcal{E}$  il funzionale lineare  $L_{\mathfrak{s}, \mathcal{E}} : E \rightarrow \mathbb{R}$  t.c.  $L_{\mathfrak{s}, \mathcal{E}}(e_i) = s_i \quad \forall i \in I$ .

**Osservazione 1.1.5:** Ovviamente, nel contesto della definizione precedente, la domanda (1.1) ha risposta affermativa per  $L = L_{\mathfrak{s}, \mathcal{E}}$  se e solo se  $s_i = \int_X e_i(x) d\mu(x) \quad \forall i \in I$ .

Quest'ultima osservazione rende esplicito il collegamento tra il molto generale Problema 3 ed il problema dei momenti.



**Definizione 1.1.6:** Sia  $K$  un sottoinsieme chiuso di  $\mathbb{R}^d$  ed  $\mathfrak{s} = (s_\alpha)_{\alpha \in \mathbb{N}^d}$  una successione di numeri reali. Diremo che  $\mathfrak{s}$  è una **successione di K-momenti** se esiste una misura di Radon  $\mu$  su  $\mathbb{R}^d$  concentrata su  $K$  t.c.  $s_\alpha = \int_K x^\alpha d\mu(x) \quad \forall \alpha \in \mathbb{N}^d$ .

**Teorema 1.1.7:** Sia  $K$  un sottoinsieme chiuso di  $\mathbb{R}^d$  ed  $\mathfrak{s} = (s_\alpha)_{\alpha \in \mathbb{N}^d}$  una successione di numeri reali. Allora  $K$  è localmente compatto e di Hausdorff e le seguenti due affermazioni sono equivalenti:

- (1)  $\mathfrak{s}$  è una successione di  $K$ -momenti;
- (2) il funzionale di Riesz  $L_{\mathfrak{s}, \mathcal{M}} : \mathbb{R}[x_1, \dots, x_d] \rightarrow \mathbb{R}$  associato a  $\mathfrak{s}$  rispetto alla base  $\mathcal{M} = \{x^\alpha\}_{\alpha \in \mathbb{N}^d}$  è  $K$ -rappresentabile.

*Dimostrazione.* Segue da quanto detto nell'Osservazione 1.1.5. □

## 1.2 Spazi di Choquet e teorema di Haviland

Dato un sottospazio vettoriale  $E$  di  $C(X, \mathbb{R})$  chiameremo

$$E_+ = \{f \in E \mid f(x) \geq 0 \quad \forall x \in X\}.$$

**Definizione 1.2.1:** Un sottospazio vettoriale  $E$  di  $C(X, \mathbb{R})$  è detto **spazio di Choquet** se gode delle seguenti proprietà:

- (1)  $E = E_+ - E_+$ ;
- (2)  $\forall x \in X$  esiste  $f_x \in E$  t.c.  $f_x(x) > 0$ ;
- (3)  $\forall f \in E$  esiste  $g \in E_+$  t.c.  $\forall \varepsilon > 0 \quad \exists K_\varepsilon \subset X$  compatto t.c.  $|f(x)| \leq \varepsilon g(x) \quad \forall x \in X \setminus K_\varepsilon$ .

**Osservazione 1.2.2:** La condizione (3) nel caso in cui  $X = \mathbb{R}^d$  è equivalente a dire che

$$\left| \frac{f(x)}{g(x)} \right| \rightarrow 0 \quad \text{per } \|x\| \rightarrow +\infty.$$

**Lemma 1.2.3:** Sia  $E$  è un spazio di Choquet in  $C(X, \mathbb{R})$ . Per ogni  $f \in C_c(X, \mathbb{R})_+$  esiste  $g \in E_+$  t.c.  $g(x) \geq f(x) \quad \forall x \in X$ .

*Dimostrazione.* Fissiamo  $x \in X$ . Per la proprietà (2) nella Definizione 1.2.1 si ha che esiste  $g_x \in E_+$  t.c.  $g_x(x) > 0$  ed a meno di moltiplicare  $g_x$  per una qualche costante

positiva possiamo supporre che  $g_x(x) > f(x)$ . Per continuità  $\exists U_x \subset X$  intorno aperto di  $x$  t.c.  $g_x(y) > f(y) \quad \forall y \in U_x$ . La famiglia  $\{U_x\}_{x \in X}$  è un ricoprimento aperto di  $\text{supp}(f)$  che è compatto in  $X$ , dunque esistono  $n \in \mathbb{N}$  e  $\{x_1, \dots, x_n\} \subset X$  t.c.  $\text{supp}(f) \subset \bigcup_{j=1}^n U_{x_j}$ . Presa  $g := \sum_{j=1}^n g_{x_j}$  si ha che  $g(x) > f(x) \quad \forall x \in X$ .  $\square$

**Definizione 1.2.4:** Dati un sottospazio vettoriale  $E$  di  $C(X, \mathbb{R})$ ,  $L : E \rightarrow \mathbb{R}$  un funzionale e  $D \subset E$  diremo che  $L$  è **D-positivo**, o solo **positivo** nel caso  $D = E_+$ , se  $L(f) \geq 0 \quad \forall f \in D$ .

Nel seguito ci serviranno i seguenti celebri teoremi di cui si riportano gli enunciati.

**Teorema 1.2.5 (di rappresentazione di Riesz-Markov):** *Sia  $X$  uno spazio topologico localmente compatto e di Hausdorff ed  $L : C_c(X, \mathbb{R}) \rightarrow \mathbb{R}$  un funzionale lineare positivo. Esiste un'unica misura di Radon  $\mu : \mathcal{B}(X) \rightarrow \mathbb{R}$  t.c.*

$$L(f) = \int_X f d\mu \quad \forall f \in C_c(X, \mathbb{R}).$$

*Dimostrazione.* Si veda ad esempio [7, Section 21.4]  $\square$

**Teorema 1.2.6 (di Hahn-Banach):** *Siano  $V$  uno spazio vettoriale reale e  $p : V \rightarrow \mathbb{R}$  un funzionale t.c.*

- (1)  $p(\lambda f) = \lambda p(f) \quad \forall f \in V \quad \forall \lambda \in [0, +\infty)$ ;
- (2)  $p(f + g) \leq p(f) + p(g) \quad \forall f, g \in V$ .

*Se  $E$  è un sottospazio vettoriale di  $V$  ed  $L : E \rightarrow \mathbb{R}$  è un funzionale lineare t.c.  $L(f) \leq p(f) \quad \forall f \in E$ , allora esiste  $\tilde{L} : V \rightarrow \mathbb{R}$  lineare che estende  $L$  e t.c.  $\tilde{L}(f) \leq p(f) \quad \forall f \in V$ .*

*Dimostrazione.* Si veda ad esempio [6, Theorem 1.1].  $\square$

**Definizione 1.2.7:** Sia  $V$  uno spazio vettoriale reale. Un **cono convesso** di  $V$  è un  $C \subset V$  t.c.

- (1)  $f + g \in C \quad \forall f, g \in C$ ;
- (2)  $\lambda f \in C \quad \forall f \in C \quad \forall \lambda \in [0, +\infty)$ .

**Lemma 1.2.8:** *Sia  $V$  uno spazio vettoriale reale,  $E$  un suo sottospazio vettoriale e  $C$  un suo cono convesso e supponiamo che  $V = E + C$ . Allora ogni funzionale lineare*

$(C \cap E)$ -positivo  $L : E \rightarrow \mathbb{R}$  si può estendere ad un funzionale lineare  $C$ -positivo  $\tilde{L} : V \rightarrow \mathbb{R}$ .

*Dimostrazione.* Definiamo il funzionale  $q : V \rightarrow \mathbb{R}$  t.c.  $q(f) = \inf\{L(g) \mid g \in E, g - f \in C\}$ . Preso  $f \in V$ , essendo  $V = E + C$ , esistono  $g \in E$  e  $c \in C$  t.c.  $-f = -g + c$  che equivale a  $c = g - f \in C$ . Di conseguenza l'insieme  $\{L(g) \mid g \in E, g - f \in C\}$  è non-vuoto, quindi  $q$  è a valori finiti. Vediamo che  $q$  è omogenea, sublineare e che  $\forall g \in E$  vale  $q(g) = L(g)$ .

Siano  $f \in E$  e  $\lambda > 0$ , essendo  $C$  un cono convesso vale che

$$\begin{aligned} q(\lambda f) &= \inf\{L(g) \mid g \in E, g - \lambda f \in C\} \\ &= \inf\left\{L(g) \mid g \in E, \frac{1}{\lambda}g - f \in C\right\} \\ &= \inf\{L(\lambda h) \mid h \in E, h - f \in C\} = \lambda q(f) \end{aligned} \tag{1.2}$$

che prova l'omogeneità di  $q$ .

Siano  $f, g \in E$ , allora  $q(f + g) = \inf\{L(h) \mid h \in E, h - (f + g) \in C\}$ . Prendiamo  $h_1, h_2 \in E$  t.c.  $h_1 - f, h_2 - g \in C$ , essendo  $C$  cono convesso si ha  $h_1 + h_2 - (f + g) = (h_1 - f) + (h_2 - g) \in C$  e di conseguenza

$$q(f + g) \leq \inf\{L(h_1) + L(h_2) \mid h_1, h_2 \in E, h_1 - f, h_2 - g \in C\} \leq q(f) + q(g)$$

che prova la sublinearità di  $q$ .

Adesso fissiamo  $g \in E$  e prendiamo  $h \in E$  con  $c := h - g \in C$ , allora  $c \in C \cap E$  e  $L(c) \geq 0$ . Di conseguenza, essendo  $h = c + g$ , per linearità si ha

$$L(h) = L(c) + L(g) \geq L(g)$$

da cui  $q(g) \geq L(g)$  ed ovviamente si ha anche  $q(g) \leq L(g)$ , dunque  $q(g) = L(g)$  come voluto.

Sono quindi verificate tutte le ipotesi del Teorema 1.2.6 di Hahn-Banach che ci fornisce un'estensione lineare  $\tilde{L}$  di  $L$  t.c.  $\tilde{L}(f) \leq q(f) \quad \forall f \in V$ . Sia  $h \in C$  e poniamo  $g = 0$  e  $f = -h$ , allora  $g - f = h \in C$  e  $q(-h) \leq L(0) = 0$ , da cui

$$-\tilde{L}(h) = \tilde{L}(-h) \leq q(-h) \leq 0$$

che equivale a  $\tilde{L}(h) \geq 0$ , ossia  $\tilde{L}$  è  $C$ -positivo.  $\square$

**Teorema 1.2.9:** *Sia  $E$  uno spazio di Choquet in  $C(X, \mathbb{R})$ . Per ogni funzionale lineare  $L : E \rightarrow \mathbb{R}$  sono equivalenti:*

- (1)  $L$  è un funzionale  $E_+$ -positivo;  
 (2)  $L$  è un funzionale rappresentabile.

*Dimostrazione.* L'implicazione (2)  $\Rightarrow$  (1) è ovvia, infatti presa  $\mu$  misura di Radon su  $X$  t.c.  $L(f) = \int_X f d\mu \quad \forall f \in E$  allora per ogni  $f \in E_+$  si ha  $L(f) = \int_X f d\mu \geq 0$ . Dimostriamo che (1)  $\Rightarrow$  (2). Poniamo

$$\tilde{E} = \{f \in C(X, \mathbb{R}) \mid \exists g \in E_+ \text{ t.c. } |f(x)| \leq g(x) \quad \forall x \in X\}$$

e dimostriamo che  $\tilde{E} = E + (\tilde{E})_+$ . Essendo  $E$  spazio di Choquet  $E + (\tilde{E})_+ \subset \tilde{E}$ . Infatti presa  $f \in E$ , per la proprietà (3) nella Definizione 1.2.1, esiste una  $g \in E_+$  ed un compatto  $K \subset X$  t.c.  $|f| \leq g$  su  $X \setminus K$ . Inoltre  $|f| \mathbb{1}_K \in C_c(X, \mathbb{R})$  dunque per il Lemma 1.2.3 esiste  $g_K \in E_+$  t.c.  $|f| \mathbb{1}_K \leq g_K$  su  $X$ . Allora chiaramente  $|f| \leq g + g_K \in E_+$  su  $X$ . Presa invece  $h \in (\tilde{E})_+ \subset \tilde{E}$  possiamo scegliere  $g' \in E_+$  t.c.  $h \leq g'$  su  $X$ . Quindi

$$|f + h| \leq |f| + h \leq g + g_K + g' \in E_+.$$

Dimostriamo adesso che  $E + (\tilde{E})_+ \supset \tilde{E}$ . Sia  $f \in \tilde{E}$  scegliamo  $g \in E_+$  t.c.  $|f| \leq g$ . Allora si ha  $f + g \in (\tilde{E})_+$ ,  $-g \in E$  e

$$f = -g + (g + f) \in E + (\tilde{E})_+$$

dunque effettivamente  $\tilde{E} = E + (\tilde{E})_+$ .

Chiaramente  $(\tilde{E})_+$  è un cono convesso di  $\tilde{E}$  e  $E_+ = (\tilde{E})_+ \cap E$ , dunque per il Lemma 1.2.8  $L$  si estende ad un funzionale lineare  $(\tilde{E})_+$ -positivo  $\tilde{L} : \tilde{E} \rightarrow \mathbb{R}$ . Notiamo che  $C_c(X, \mathbb{R}) \subset \tilde{E}$  per il Lemma 1.2.3, quindi il Teorema 1.2.5 di rappresentazione di Riesz-Markov ci fornisce una misura di Radon  $\mu$  su  $X$  t.c.  $\tilde{L}(f) = \int_X f d\mu \quad \forall f \in C_c(X, \mathbb{R})$ . Essendo  $E$  spazio di Choquet vero è che  $E = E_+ - E_+$ , dunque per concludere la dimostrazione basta mostrare che ogni  $f \in E_+$  è  $\mu$ -integrabile e soddisfa  $L(f) = \tilde{L}(f) = \int_X f d\mu$ .

Fissiamo  $f \in E_+$  e consideriamo la famiglia

$$\mathcal{U} := \{\eta \in C_c(X, \mathbb{R}) \mid 0 \leq \eta(x) \leq 1 \quad \forall x \in X\}.$$

Ovviamente per  $\eta \in \mathcal{U}$ ,  $f\eta \in C_c(X, \mathbb{R})$  e  $\tilde{L}(f\eta) = \int_X f\eta d\mu$ . Usando questo e la  $E_+$ -positività di  $\tilde{L}$ , grazie al teorema di Beppo Levi, possiamo affermare che

$$\int_X f d\mu = \sup_{\eta \in \mathcal{U}} \int_X f\eta d\mu = \sup_{\eta \in \mathcal{U}} \tilde{L}(f\eta) \leq \tilde{L}(f) = L(f) < +\infty \quad (1.3)$$

ossia  $f$  è  $\mu$ -integrabile e  $L(f) \geq \int_X f d\mu$ .

Dimostriamo che  $L(f) \leq \int_X f d\mu$ . Per la proprietà (3) in Definizione 1.2.1 esiste una funzione  $g \in E_+$  t.c.  $\forall \varepsilon > 0 \exists K_\varepsilon \subset X$  compatto t.c.  $|f(x)| \leq \varepsilon g(x)$  per ogni  $x \in X$ . Fissiamo  $\varepsilon > 0$  e prendiamo  $K_\varepsilon$  compatto dato dalla condizione associata a  $g$  e  $\eta_\varepsilon \in \mathcal{U}$  t.c.  $\eta_\varepsilon = 1$  su  $K_\varepsilon$ . Allora  $f \leq \varepsilon g + f\eta_\varepsilon$  ed essendo  $f\eta_\varepsilon \leq f$  si ottiene

$$\begin{aligned} L(f) &= \tilde{L}(f) \leq \varepsilon \tilde{L}(g) + \tilde{L}(f\eta_\varepsilon) \\ &= \varepsilon L(g) + \int_X f\eta_\varepsilon d\mu \\ &\leq \varepsilon L(g) + \int_X f d\mu \\ &= O(\varepsilon) + \int_X f d\mu \quad \text{per } \varepsilon \rightarrow 0^+ \end{aligned}$$

dunque prendendo il limite per  $\varepsilon \rightarrow 0^+$  si ottiene la disuguaglianza che conclude la dimostrazione.  $\square$

Notiamo che se  $X$  compatto e di Hausdorff, allora  $C(X, \mathbb{R}) = C_c(X, \mathbb{R})$  e la condizione (3) della Definizione 1.2.1 è soddisfatta in modo ovvio da ogni  $E \subset C(X, \mathbb{R})$ . Infatti fissiamo  $f \in E$  e  $\varepsilon > 0$ , allora  $\frac{|f|}{\varepsilon} \in C_c(X, \mathbb{R})_+$  e per il Lemma 1.2.3 esiste  $g \in E_+$  t.c.  $\frac{|f|}{\varepsilon} \leq g$  su  $X$ . Ma allora  $|f| \leq \varepsilon g$  su  $X$  e la condizione (3) della Definizione 1.2.1 è ovviamente soddisfatta. In ogni caso, con queste più forti ipotesi topologiche, vale la seguente proposizione che non si appoggia in nessun modo alla nozione di spazio di Choquet.

**Proposizione 1.2.10:** *Sia  $X$  uno spazio topologico compatto e di Hausdorff ed  $E$  un sottospazio vettoriale di  $C(X, \mathbb{R})$  che contiene una funzione  $e \in E$  t.c.  $e(x) > 0 \forall x \in X$ . Allora per ogni funzionale lineare  $L : E \rightarrow \mathbb{R}$  sono equivalenti:*

- (1)  $L$  è un funzionale  $E_+$ -positivo;
- (2)  $L$  è un funzionale rappresentabile.

*Dimostrazione.* Anche in questo caso l'implicazione (2)  $\Rightarrow$  (1) è ovvia. Dimostriamo che (1)  $\Rightarrow$  (2). Poniamo  $V = C(X, \mathbb{R})$ ,  $C = C(X, \mathbb{R})_+$  e fissiamo  $f \in V$ . Essendo  $X$  compatto  $f$  è limitata e  $\min_X e > 0$ , quindi  $\lambda = \frac{\max_X f}{\min_X e} > 0$  ed è t.c.  $f(x) \leq \lambda e(x) \forall x \in X$ , fatto che ci permette di scrivere  $\lambda e - f \in C$ . Questo insieme al fatto che  $\lambda e \in E$  implica  $f = \lambda e - (\lambda e - f) \in E + C$ , da cui si deduce  $V \subset E + C$  e quindi  $V = E + C$ .

Applicando il Lemma 1.2.8 possiamo estendere  $L$  ad un funzionale  $\tilde{L}$  su  $V = C_c(X, \mathbb{R})$   $C$ -positivo ed invocando il Teorema 1.2.5 di rappresentazione di Riesz-Markov si conclude che  $\tilde{L}$ , e quindi anche  $L$ , è rappresentabile.  $\square$

**Osservazione 1.2.11:** Nella dimostrazione del Teorema 1.2.9 si è rivelata di fondamentale importanza la proprietà (3) della Definizione 1.2.1. Difatti questa è un'ipotesi necessaria per la validità della tesi.

Come prova di questa affermazione consideriamo lo spazio  $E = C_c(\mathbb{R}, \mathbb{R}) + \mathbb{R} \cdot \mathbf{1}_{\mathbb{R}} \subset C(\mathbb{R}, \mathbb{R})$  e prendiamo il funzionale lineare  $L : E \rightarrow \mathbb{R}$  t.c.

$$L(f + \lambda \mathbf{1}_{\mathbb{R}}) = \lambda \quad \forall f \in C_c(\mathbb{R}, \mathbb{R}) \quad \text{e} \quad \forall \lambda \in \mathbb{R}$$

Notiamo facilmente che  $E$  gode delle proprietà (1) e (2) della Definizione 1.2.1 e che  $L$  è  $E_+$ -positivo, ma  $L$  non è rappresentabile. Infatti se  $\mu$  fosse una misura di Radon t.c.  $L(f) = \int_X f d\mu \quad \forall f \in E$ , essendo  $L(f) = 0 \quad \forall f \in C_c(X, \mathbb{R})$  si avrebbe  $\mu = 0$ , ma questo darebbe un assurdo con  $L(\mathbf{1}_{\mathbb{R}}) = 1$ .

Comunque possiamo sempre considerare  $E$  come sottospazio di  $C(\overline{\mathbb{R}}, \mathbb{R})$ , in cui  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$  è la compattificazione di Alexandrov di  $\mathbb{R}$ , ponendo  $(f + \lambda \mathbf{1}_{\mathbb{R}})(\infty) = \lambda \quad \forall f \in C_c(X, \mathbb{R}) \quad \forall \lambda \in \mathbb{R}$ . Allora in questo caso

$$L(f + \lambda \mathbf{1}_{\mathbb{R}}) = \lambda = (f + \lambda \mathbf{1}_{\mathbb{R}})(\infty) = \int_{\overline{\mathbb{R}}} (f + \lambda \mathbf{1}_{\mathbb{R}}) d\delta_{\infty}$$

ossia  $L$  è rappresentabile tramite  $\delta_{\infty}$  misura di Dirac in  $\infty \in \overline{\mathbb{R}}$ .

Adesso applichiamo la teoria sviluppata fin'ora, in particolare il Teorema 1.2.9, per dare un potente criterio per l'esistenza di una soluzione al problema dei momenti su un qualsiasi sottoinsieme chiuso  $K$  di  $\mathbb{R}^d$ .

**Lemma 1.2.12:** *Sia  $K \subset \mathbb{R}^d$  un sottoinsieme chiuso. Allora  $E = \mathbb{R}[x_1, \dots, x_d]|_K$  è uno spazio di Choquet in  $C(K, \mathbb{R})$ .*

*Dimostrazione.* La proprietà (2) della Definizione 1.2.1 è ovvia. La proprietà (1) segue dalla relazione  $4p = (p+1)^2 - (p-1)^2$ . Dimostriamo la proprietà (3). Sia  $p \in E_+$  e consideriamo  $g(x) = \|x\|^2 p(x)$  per  $x \in K$ , allora

$$\frac{p(x)}{g(x)} = \frac{1}{\|x\|^2} \rightarrow 0 \quad \text{per} \quad \|x\| \rightarrow +\infty$$

che equivale alla proprietà (3) per  $p$ . Dunque (3) vale per ogni  $p \in E_+$ . Essendo  $E = E_+ - E_+$ , preso  $p \in E$  esistono  $p_1, p_2 \in E_+$  t.c.  $p = p_1 - p_2$  ed esistono  $g_1, g_2 \in E_+$  dati dalla proprietà (3) per  $p_1$  e  $p_2$ . Si ha

$$\frac{|p(x)|}{g_1(x) + g_2(x)} \leq \frac{p_1(x) + p_2(x)}{g_1(x) + g_2(x)} \leq \frac{p_1(x)}{g_1(x)} + \frac{p_2(x)}{g_2(x)} \rightarrow 0 \text{ per } \|x\| \rightarrow 0$$

che dimostra la proprietà (3) per  $p$ .  $\square$

**Definizione 1.2.13:** Sia  $K$  un sottoinsieme chiuso di  $\mathbb{R}^d$ , definiamo

$$\text{Pos}(K) = \{p \in \mathbb{R}[x_1, \dots, x_d] \mid p(x) \geq 0 \ \forall x \in K\}.$$

**Teorema 1.2.14 (di Haviland):** Sia  $K$  un sottoinsieme chiuso di  $\mathbb{R}^d$  ed  $L : \mathbb{R}[x_1, \dots, x_d] \rightarrow \mathbb{R}$  un funzionale lineare. Sono equivalenti:

(1)  $L$  è  $\text{Pos}(K)$ -positivo;

(2)  $L$  è  $K$ -rappresentabile.

*Dimostrazione.* L'implicazione (2)  $\Rightarrow$  (1) è ovvia. Dimostriamo che (1)  $\Rightarrow$  (2). Chiamiamo  $\mathbb{R}[x_1, \dots, x_d]_{|K} := \{p|_K \mid p \in \mathbb{R}[x_1, \dots, x_d]\}$ . Consideriamo il funzionale lineare  $L_K : \mathbb{R}[x_1, \dots, x_d]_{|K} \rightarrow \mathbb{R}$  t.c.  $L_k(p|_K) = L(p) \ \forall p \in \mathbb{R}[x_1, \dots, x_d]$  ed applichiamo il Teorema 1.2.9 con  $X = K$  ed  $E = \mathbb{R}[x_1, \dots, x_d]_{|K} \subset C(K, \mathbb{R})$  notando che  $L_k$  è  $E_+$ -positivo essendo  $(\mathbb{R}[x_1, \dots, x_d]_{|K})_+ = \text{Pos}(K)|_K = \{p|_K \mid p \in \text{Pos}(K)\}$ . Quindi troviamo una misura di Radon  $\mu_K : \mathcal{B}(K) \rightarrow [0, +\infty]$  su  $K$  che rappresenta  $L_k$ .

Definiamo  $\mu : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, +\infty]$  nel seguente modo

$$\mu(B) = \mu_K(B \cap K) \ \forall B \in \mathcal{B}(\mathbb{R}^d)$$

ovviamente  $\mu$  è ben definita ed è una misura concentrata su  $K$ . Essendo  $\mu_K$  di Radon possiamo concludere che anche  $\mu$  è di Radon. Infatti se  $C \subset \mathbb{R}^d$  è un compatto si ha  $\mu(C) = \mu_K(C \cap K) < +\infty$  in quanto  $C \cap K$  è compatto in  $K$ . Inoltre se  $p \in \mathbb{R}[x_1, \dots, x_d]$  si ha

$$L(p) = L_K(p|_K) = \int_K p|_K d\mu_K = \int_K p d\mu. \quad \square$$

### 1.3 Polinomi positivi su intervalli di $\mathbb{R}$

In questa sezione dimostreremo dei risultati sulla struttura degli spazi di polinomi positivi sui vari tipi di intervallo chiuso che ci serviranno per lo studio del problema dei momenti su  $\mathbb{R}$ .

Iniziamo facendo alcune semplici considerazioni sulla decomposizione di polinomi. Sia  $p \in \mathbb{R}[x]$ , essendo  $p$  a coefficienti reali, se  $\lambda \in \mathbb{C}$  è una radice di  $p$  lo è anche  $\bar{\lambda}$  e  $p$  si scomporrà nella forma

$$p(x) = a(x - \alpha_1)^{n_1} \dots (x - \alpha_r)^{n_r} (x - \lambda_1)^{j_1} (x - \bar{\lambda}_1)^{j_1} \dots (x - \lambda_k)^{j_k} (x - \bar{\lambda}_k)^{j_k} \quad (1.4)$$

in cui  $a \in \mathbb{R}$ ,  $\alpha_i \in \mathbb{R} \quad \forall i \in \{1, \dots, r\}$  e  $\lambda_i \in \mathbb{C} \setminus \mathbb{R} \quad \forall i \in \{1, \dots, k\}$  t.c.  $\alpha_i \neq \alpha_j$ ,  $\lambda_i \neq \lambda_j, \bar{\lambda}_j$  per  $i \neq j$ . Essendo per ogni  $\lambda = u + iv \in \mathbb{C}$   $(x - \lambda)(x - \bar{\lambda}) = (x - u)^2 + v^2$  la scomposizione (1.4) è riscrivibile nella forma

$$p(x) = a(x - \alpha_1)^{n_1} \dots (x - \alpha_r)^{n_r} ((x - u_1)^2 + v_1^2)^{j_1} \dots ((x - u_k)^2 + v_k^2)^{j_k} \quad (1.5)$$

in cui  $u_i = \operatorname{Re}(\lambda_i)$  e  $v_i = \operatorname{Im}(\lambda_i) \quad \forall i \in \{1, \dots, k\}$ .

L'eq. (1.5) esprime  $p$  come prodotto di una costante  $a$ , potenze di polinomi lineari con zero reale  $x - \alpha_i$  e potenze di polinomi quadratici con radici non reali  $(x - u_i)^2 + v_i^2$ .

**Proposizione 1.3.1:** *Valgono le seguenti uguaglianze:*

$$(1) \operatorname{Pos}(\mathbb{R}) = \{f^2 + g^2 \mid f, g \in \mathbb{R}[x]\} = \Sigma^2;$$

$$(2) \operatorname{Pos}([0, +\infty)) = \{f + xg \mid f, g \in \Sigma^2\} = \Sigma^2 + x\Sigma^2.$$

*Dimostrazione.* (1) Sia  $p \in \operatorname{Pos}(\mathbb{R})$ ,  $p \neq 0$  e consideriamo la sua scomposizione come nell'eq. (1.5). Essendo  $p \geq 0$  su tutto  $\mathbb{R}$  segue che  $a > 0$  ed i numeri  $k_1, \dots, k_r$  sono pari. Infatti se fosse  $a < 0$ , preso  $x_0 \in \mathbb{R}$  nel quale ogni fattore della decomposizione è positivo si avrebbe  $p(x_0) < 0$  che è un assurdo. Tale  $x_0$  esiste, infatti se  $m_\alpha = \max_{i \in \{1, \dots, r\}} \alpha_i$  basta prendere  $x_0 > m_\alpha$ . Mentre se esistesse un  $i_0 \in \{1, \dots, r\}$  con  $\alpha_{i_0}$  dispari, essendo gli  $\alpha_i$  tutti diversi,  $p$  cambierebbe di segno in un intorno di  $\alpha_{i_0}$  e questo da ovviamente un assurdo per la positività di  $p$ . Inoltre per qualsiasi  $a, b, c, d \in \mathbb{R}$  vale

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$$

ossia prodotto di due somme di quadrati è una somma di quadrati, dunque applicando tale formula ai fattori  $(x - u_i)^2 + v_i^2$  si ottiene che effettivamente  $p$  è scrivibile come somma di quadrati di elementi di  $\mathbb{R}[x]$ . Quindi  $\operatorname{Pos}(\mathbb{R}) \subset \Sigma^2$ , ma ovviamente  $\operatorname{Pos}(\mathbb{R}) \supset \Sigma^2$ , dunque  $\operatorname{Pos}(\mathbb{R}) = \Sigma^2$ .



(2) Anche in questo caso ovviamente si ha il contenimento  $\text{Pos}([0, +\infty)) \supset \Sigma^2 + x\Sigma^2$ . Dimostriamo l'inclusione opposta. Notiamo che  $\Sigma^2 + x\Sigma^2$  è chiuso per moltiplicazione, infatti se  $f_1, g_1, f_2, g_2 \in \Sigma^2$  si ha

$$(f_1 + xg_1)(f_2 + xg_2) = (f_1f_2 + x^2g_1g_2) + x(f_1g_2 + f_2g_1) \in \Sigma^2 + x\Sigma^2$$

in quanto prodotto di somme di quadrati è somma di quadrati. Sia  $p \in \text{Pos}([0, +\infty))$ ,  $p \neq 0$  e consideriamo la scomposizione di  $p$  come nell'eq. (1.5). Avendo mostrato che  $\Sigma^2 + x\Sigma^2$  è chiuso per moltiplicazione, basta provare che ogni fattore di tale scomposizione appartiene a  $\Sigma^2 + x\Sigma^2$ . Iniziamo col dire che  $(x - u)^2 + v^2 \in \Sigma^2 \subset \Sigma^2 + x\Sigma^2 \quad \forall u, v \in \mathbb{R}$ . Dunque manca solo da dimostrare che  $a$  ed i polinomi lineari  $x - \alpha_i$  appartengono a  $\Sigma^2 + x\Sigma^2$ . Essendo  $p \geq 0$  su  $[0, +\infty)$  si ha  $\lim_{x \rightarrow +\infty} p(x) = +\infty$  ed il limite di ogni fattore per  $x \rightarrow +\infty$  è uguale a  $+\infty$ , dunque per mantenere il segno positivo deve necessariamente essere  $a > 0$ . Invece  $p$  cambia segno in un intorno di ogni  $\alpha_i$ , dunque necessariamente  $\alpha_i \leq 0$ . Quindi  $a \in \Sigma^2 + x\Sigma^2$  e  $x - \alpha_i = -\alpha_i + x \in \Sigma^2 + x\Sigma^2$ .

□

**Corollario 1.3.2:** *Sia  $c \in \mathbb{R}$ , allora*

$$\text{Pos}([c, +\infty)) = \{f + (x - c)g \mid f, g \in \Sigma^2\} = \Sigma^2 + (x - c)\Sigma^2.$$

*Dimostrazione.* Ovviamente  $\text{Pos}([c, +\infty)) \supset \Sigma^2 + (x - c)\Sigma^2$ . Dimostriamo l'inclusione opposta. Sia  $p \in \text{Pos}([c, +\infty))$ , allora il polinomio  $\bar{p}(t) = p(t + c) \in \text{Pos}([0, +\infty))$ . Quindi esistono  $f, g \in \Sigma^2$  t.c.

$$\bar{p}(t) = f(t) + tg(t)$$

e ponendo  $x = t + c$  abbiamo  $t = x - c$  e

$$p(x) = \bar{p}(x - c) = f(x - c) + (x - c)g(x - c) \in \Sigma^2 + (x - c)\Sigma^2. \quad \square$$

**Definizione 1.3.3:** Dato un polinomio  $p \in \mathbb{R}[x]$  con  $\deg(p) = m$  e dato  $n \geq m$ , definiamo la **trasformata di Goursat** di grado  $n \in \mathbb{N}$  di  $p$  il polinomio

$$p_{(n)}(x) = (1 + x)^n p\left(\frac{1 - x}{1 + x}\right).$$

**Lemma 1.3.4 (di Goursat):** *Sia  $p \in \mathbb{R}[x]$ ,  $p \neq 0$  con  $\deg(p) = m$  e consideriamo la trasformata di Goursat di grado  $m$  di  $p$*

$$p_{(m)}(x) = (1 + x)^m p\left(\frac{1 - x}{1 + x}\right).$$

Valgono i seguenti fatti

- (1)  $\deg(p_{(m)}) \leq m$  e  $\deg(p_{(m)}) = m$  se e solo se  $p(-1) \neq 0$ ;
- (2)  $p \in \text{Pos}([-1, 1])$  se e solo se  $p_{(m)} \in \text{Pos}([0, +\infty))$ ;
- (3)  $p > 0$  su  $[-1, 1]$  se e solo se  $p_{(m)} > 0$  su  $[0, +\infty)$  e  $\deg(p_{(m)}) = m$ .

*Dimostrazione.* (1) Sia  $p(x) = \sum_{k=0}^m a_k x^k$ . Ovviamente la sua trasformata di Goursat

$$p_{(m)}(x) = \sum_{k=0}^m a_k (1+x)^{m-k} (1-x)^k \quad (1.6)$$

è un polinomio di grado  $\deg(p_{(m)}) \leq m$ . Per ogni  $k \in \{0, \dots, m\}$  si ha

$$a_k (1+x)^{m-k} (1-x)^k = a_k (-1)^k x^m + R_k(x)$$

con  $\deg(R_k) \leq m-1$ , dunque il coefficiente di  $x^m$  di  $p_{(m)}$  è  $\sum_{k=0}^m a_k (-1)^k = p(-1)$ . Di conseguenza  $\deg(p_{(m)}) = m$  se e solo se  $p(-1) \neq 0$ .

- (2) Per  $x \neq -1$  consideriamo il cambio di variabile  $t = \frac{1-x}{1+x}$ . Allora  $t \neq -1$  e  $x = \frac{1-t}{1+t}$ . In particolare  $[0, +\infty) \ni t \mapsto x = \frac{1-t}{1+t} \in (-1, 1]$  è una bigezione e vale

$$p_{(m)}(t) = (1+t)^m p(x) \quad \text{per } t \in [0, +\infty) \text{ e con } x = \frac{1-t}{1+t} \quad (1.7)$$

quindi  $p \geq 0$  su  $(-1, 1]$  se e solo se  $p_{(m)} \geq 0$  su  $[0, +\infty)$ . Inoltre per continuità di  $p$  si ha  $p \geq 0$  su  $[-1, 1]$  se e solo se  $p \geq 0$  su  $(-1, 1]$ . Mettendo insieme si ottiene che  $p \geq 0$  su  $[-1, 1]$  se e solo se  $p_{(m)} \geq 0$  su  $[0, +\infty)$ .

- (3) Sempre per l'eq. (1.7)  $p > 0$  su  $(-1, 1]$  se e solo se  $p_{(m)} > 0$  su  $[0, +\infty)$  e, se questo vale, per continuità si ottiene anche  $p(-1) \geq 0$ . Infine per il punto (1) sarà  $p(-1) > 0$  se e solo se  $\deg(p_{(m)}) = m$ .

□

**Osservazione 1.3.5:** Supponiamo che  $p \in \mathbb{R}[x]$  sia un polinomio con  $\deg(p) = m$  e  $p(-1) \neq 0$ , in modo che  $\deg(p_{(m)}) = m$ . Calcoliamo  $(p_{(m)})_{(m)}$

$$(p_{(m)})_{(m)}(x) = (1+x)^m \left(1 + \frac{1-x}{1+x}\right)^m p\left(\frac{1 - \frac{1-x}{1+x}}{1 + \frac{1-x}{1+x}}\right) = 2^m p(x)$$

ed in particolare  $\deg((p_{(m)})_{(m)}) = \deg(p) = m = \deg(p_{(m)})$ , dunque per il Lemma 1.3.4 (1) si ha  $p_{(m)}(-1) \neq 0$ . Quindi, chiamando  $\mathbb{P}_m = \{p \in \mathbb{R}[x] \mid p(-1) \neq 0 \text{ } \deg(p) = m\}$  l'operatore

$$\begin{aligned} \mathfrak{G}_m : \mathbb{P}_m &\rightarrow \mathbb{P}_m \\ p &\mapsto p_{(m)} \end{aligned}$$

è ben definito ed è bigettivo con inversa  $\mathfrak{G}_m^{-1} = 2^{-m}\mathfrak{G}_m$ .

In realtà anche se  $p(-1) = 0$ , cioè  $\deg(p_{(m)}) = s < m = \deg(p)$ , possiamo comunque eseguire il conto fatto in precedenza e trovare che

$$(p_{(m)})_{(m)}(x) = 2^m p(x).$$

**Proposizione 1.3.6:** *Per ogni  $a, b \in \mathbb{R}$ ,  $a < b$  vale*

$$\text{Pos}([a, b]) = \{f + (b-x)(x-a)g \mid f, g \in \Sigma^2\} =: \Sigma^2 + (b-x)(x-a)\Sigma^2.$$

*Dimostrazione.* Basta dimostrare l'enunciato per  $\text{Pos}([-1, 1])$ , in quanto preso un qualsiasi intervallo  $[a, b] \subset \mathbb{R}$  ed un polinomio  $p \in \text{Pos}([a, b])$ , possiamo considerare

$$\bar{p}(x) = p\left(\frac{(b-a)x + (b+a)}{2}\right) \in \text{Pos}([-1, 1])$$

e se esistono  $f, g \in \Sigma^2$  t.c.  $\bar{p} = f + (1-x)(1+x)g$ , ponendo  $x = \frac{2t-(b+a)}{b-a}$ , si ottiene

$$p(t) = \bar{p}\left(\frac{2t-(b+a)}{b-a}\right) = f' + (b-t)(t-a)g'$$

in cui  $f'(t) = f\left(\frac{2t-(b+a)}{b-a}\right) \in \Sigma^2$  e  $g'(t) = 4g\left(\frac{2t-(b+a)}{b-a}\right) \in \Sigma^2$ .

Dunque dimostriamo l'asserto per  $\text{Pos}([-1, 1])$ . Ovviamente  $\text{Pos}([-1, 1]) \supset \Sigma^2 + (1-x)(x+1)\Sigma^2$ . Sia  $p \in \text{Pos}([-1, 1])$  con  $\deg(p) = m$ . Consideriamo la sua trasformata di Goursat di grado  $m$ ,  $p_{(m)}$ . Essendo  $p \geq 0$  su  $[-1, 1]$ , dal Lemma 1.3.4 (2) segue che  $p_{(m)} \geq 0$  su  $[0, +\infty)$  e di conseguenza per la Proposizione 1.3.1 (2) segue che esistono  $f_1, f_2, g_1, g_2 \in \mathbb{R}[x]$ , con  $r_i := \deg(f_i) \leq \frac{m}{2}$  e  $s_i := \deg(g_i) \leq \frac{m}{2} \quad \forall i \in \{1, 2\}$ , t.c.

$$p_{(m)}(x) = \sum_{i=1}^2 f_i(x)^2 + x \sum_{i=1}^2 g_i(x)^2.$$

Effettuando un'ulteriore trasformata di Goursat di grado  $m$  si ottiene

$$\begin{aligned} 2^m p(x) &= \sum_{i=1}^2 (1+x)^{m-2r_i} (f_i)_{(r_i)}(x)^2 + \frac{1-x}{1+x} \sum_{i=1}^2 (1+x)^{m-2s_i} (g_i)_{(s_i)}(x)^2 \\ &= \sum_{i=1}^2 (1+x)^{m-2r_i} (f_i)_{(r_i)}(x)^2 + (1-x) \sum_{i=1}^2 (1+x)^{m-1-2s_i} (g_i)_{(s_i)}(x)^2. \end{aligned}$$

Se  $m$  è pari  $p \in \Sigma^2 + (1-x)(1+x)\Sigma^2$  in quanto in tal caso  $\frac{m}{2} \in \mathbb{N}$  e

$$2^m p(x) = \sum_{i=1}^2 \left( (1+x)^{\frac{m}{2}-r_i} (f_i)_{(r_i)}(x) \right)^2 + (1-x)(1+x) \sum_{i=1}^2 \left( (1+x)^{\frac{m}{2}-1-s_i} (g_i)_{(s_i)}(x) \right)^2.$$

Diversamente, se  $m$  è dispari,  $p \in (1+x)\Sigma^2 + (1-x)\Sigma^2$ , in quanto in questo caso  $\frac{m-1}{2} \in \mathbb{N}$  e

$$2^m p(x) = (1+x) \sum_{i=1}^2 \left( (1+x)^{\frac{m-1}{2}-r_i} (f_i)_{(r_i)}(x) \right)^2 + (1-x) \sum_{i=1}^2 \left( (1+x)^{\frac{m-1}{2}-s_i} (g_i)_{(s_i)}(x) \right)^2.$$

Notiamo che essendo

$$1 \pm x = \frac{(1 \pm x)^2}{2} + \frac{1}{2}(1+x)(1-x) \in \Sigma^2 + (1+x)(1-x)\Sigma^2$$

si ha l'uguaglianza insiemistica  $\Sigma^2 + (1-x)(1+x)\Sigma^2 = (1+x)\Sigma^2 + (1-x)\Sigma^2$ , da cui segue la tesi.  $\square$

**Lemma 1.3.7 (di Polya):** *Sia  $p \in \mathbb{R}[x]$  t.c.  $p > 0$  su  $[0, +\infty)$ . Allora  $\exists N \in \mathbb{N}$  t.c. il polinomio  $(1+x)^N p(x)$  ha solo coefficienti strettamente positivi, cioè*

$$(1+x)^N p(x) = \sum_{k=0}^m b_k x^k \quad \text{con } b_k > 0 \quad \forall k \in \{1, \dots, m\} \quad \text{ed } m \in \mathbb{N}.$$

*Dimostrazione.* Per  $z, t \in \mathbb{R}$  e  $j \in \mathbb{N}$  chiamiamo  $(z)_t^j := z(z-t)(z-2t)\dots(z-(j-1)t)$ . In modo tale che  $\forall k, m, j \in \mathbb{N}$ ,  $j < k$ , risulti

$$\begin{aligned} \left( \frac{k}{m} \right)_{\frac{1}{m}}^j &= \frac{k}{m} \left( \frac{k}{m} - \frac{1}{m} \right) \left( \frac{k}{m} - \frac{2}{m} \right) \dots \left( \frac{k}{m} - \frac{j-1}{m} \right) \\ &= \frac{k}{m} \frac{k-1}{m} \frac{k-2}{m} \dots \frac{k-j+1}{m} = \frac{k!}{(k-j)! m^j}. \end{aligned}$$

Supponiamo  $p(x) = \sum_{j=0}^n a_j x^j$  con  $a_n \neq 0$  e per  $x, y, t \in \mathbb{R}$  definiamo

$$\begin{aligned} P(x, y) &= \sum_{j=0}^n a_j x^j y^{n-j} \\ P_t(x, y) &= \sum_{j=0}^n a_j (x)_t^j (y)_t^{n-j}. \end{aligned}$$

Essendo  $p > 0$  su  $[0, +\infty)$ , il polinomio  $P$  è strettamente positivo su  $\Delta = \{(x, y) \in [0, +\infty) \times [0, +\infty) \mid x + y = 1\}$ , che è compatto, dunque  $P$  vi ha un minimo  $c > 0$ .

Preso  $N \in \mathbb{N}_+$ , usando l'usuale sviluppo del binomio, si ottiene

$$(x+y)^N P(x,y) = \sum_{j=0}^n \sum_{i=0}^N a_j \binom{N}{i} x^{i+j} y^{N+n-i-j}. \quad (1.8)$$

Fissiamo  $k \in \mathbb{N}$  t.c.  $k \leq N+n$  e poniamo  $m = N+n$  e  $l = m - k$ . Il coefficiente  $b_k$  di  $x^k y^l = x^k y^{N+n-k}$  nell'eq. (1.8) è

$$\begin{aligned} b_k &= \sum_{j=0}^k a_j \binom{N}{k-j} = \sum_{j=0}^k a_j \frac{N!}{(k-j)!(N-k+j)!} \\ &= \frac{N!m^n}{k!l!} \sum_{j=0}^k a_j \frac{k!}{(k-j)!m^j} \frac{l!}{(l+n-j)!m^{n-j}} \\ &= \frac{N!m^n}{k!l!} \sum_{j=0}^k a_j \left(\frac{k}{m}\right)_{\frac{1}{m}}^j \left(\frac{l}{m}\right)_{\frac{1}{m}}^{n-j} \\ &= \frac{N!m^n}{k!l!} P_{\frac{1}{m}}\left(\frac{k}{m}, \frac{l}{m}\right). \end{aligned}$$

Inoltre  $P_t \rightarrow P$  uniformemente su  $\Delta$  per  $t \rightarrow 0^+$ , infatti  $\forall (x,y) \in \Delta$  vale

$$\begin{aligned} |P_t(x,y) - P(x,y)| &\leq \sum_{j=0}^k |a_j| \left| x^j y^{k-j} - (x)_t^j (y)_t^{n-j} \right| \\ &= \sum_{j=0}^k |a_j| \left| x^j y^{k-j} - x(x-t)\dots(x-(j-1)t)y(y-t)\dots(y-(k-j-1)t) \right| \\ &= \sum_{j=0}^k |a_j| |t| |R_j(x,y,t)| \end{aligned}$$

in cui  $R_j(x,y,t)$  è un polinomio in tre variabili  $\forall j \in \{0, 1, \dots, k\}$ . In particolare  $\forall j \in \{0, 1, \dots, k\}$  la funzione  $|R_j|$  è positiva e continua su  $\Delta \times [0, 1]$ , che è compatto, dunque vi ammette un massimo  $M_j \geq 0$ . Quindi per  $t \in [0, 1]$

$$\sup_{(x,y) \in \Delta} |P_t(x,y) - P(x,y)| \leq |t| \sum_{j=0}^k |a_j| M_j \rightarrow 0 \text{ per } t \rightarrow 0^+.$$

Notiamo che  $(\frac{k}{m}, \frac{l}{m}) \in \Delta \quad \forall N \in \mathbb{N}$  e  $\forall k \in \{0, \dots, m\}$ . Quindi grazie alla convergenza uniforme, ricordando che  $P(x,y) \geq c > 0$  su  $\Delta$ , possiamo dire che per  $N$ , quindi  $m = N+n$ , sufficientemente grande  $P_{\frac{1}{m}}(\frac{k}{m}, \frac{l}{m}) > 0 \quad \forall k \in \{0, 1, \dots, N+n\}$ . Di conseguenza  $b_k > 0 \quad \forall k \in \{0, 1, \dots, N+n\}$  per  $N$  sufficientemente grande. A questo punto la tesi segue notando che  $(1+x)^N p(x) = (1+x)^N P(x,1) = \sum_{k=0}^m b_k x^k$ .  $\square$

**Teorema 1.3.8 (di Bernstein):** Sia  $p \in \mathbb{R}[x]$ ,  $p > 0$  su  $[-1, 1]$ . Allora esistono  $m \in \mathbb{N}$  ed  $\{a_{k,l}\}_{k,l \in \{0,1,\dots,m\}} \subset [0, +\infty)$  t.c.

$$p(x) = \sum_{k,l=0}^m a_{k,l} (1-x)^k (1+x)^l.$$

*Dimostrazione.* Sia  $n := \deg(p)$ . Essendo  $p > 0$  su  $[-1, 1]$ , per il Lemma 1.3.4 (3), si ha  $p_{(n)} > 0$  su  $[0, +\infty)$  e  $\deg(p_{(n)}) = n$ . Possiamo quindi applicare il Lemma 1.3.7 per trovare  $N \in \mathbb{N}$  e  $\{a'_j\}_{j \in \{0,1,\dots,N+n\}} \subset (0, +\infty)$  t.c.

$$(1+t)^N p_{(n)}(t) = \sum_{j=0}^{N+n} a'_j t^j. \quad (1.9)$$

Poniamo  $m := N + n$  e  $t = \frac{1-x}{1+x}$  per  $x \neq -1$ . Allora  $x = \frac{1-t}{1+t}$  e  $1-t = \frac{2x}{1+x}$  da cui  $\frac{1}{1+t} = \frac{1+x}{2}$  e  $p_{(n)}(t) = (1+t)^n p(x)$ . Usando questi fatti e l'eq. (1.9) otteniamo

$$\begin{aligned} p(x) &= (1+t)^{-n} p_{(n)}(t) = (1+t)^{N-m} p_{(n)}(t) \\ &= (1+t)^{-m} \sum_{j=0}^m a'_j t^j = \left(\frac{1+x}{2}\right)^m \sum_{j=0}^m a'_j \left(\frac{1-x}{1+x}\right)^j \\ &= \sum_{j=0}^m 2^{-m} a'_j (1-x)^j (1+x)^{m-j}. \end{aligned}$$

Questo conclude la dimostrazione, in quanto per ottenere un'espressione analoga a quella dell'enunciato basta porre  $a_{j,m-j} := 2^{-m} a'_j$  e  $a_{j,l} = 0 \quad \forall l \in \{0, 1, \dots, m\} \setminus \{m-j\} \quad \forall j \in \{0, \dots, m\}$ .  $\square$

**Corollario 1.3.9:** Sia  $p \in \mathbb{R}[x]$ ,  $p > 0$  su  $[a, b] \subset \mathbb{R}$ . Allora esistono  $m \in \mathbb{N}$  ed  $\{a_{k,l}\}_{k,l \in \{0,1,\dots,m\}} \subset [0, +\infty)$  t.c.

$$p(x) = \sum_{k,l=0}^m a_{k,l} (b-x)^k (a+x)^l.$$

*Dimostrazione.* Consideriamo la bigezione  $[-1, 1] \ni t \mapsto \frac{a(1-t)+b(1+t)}{2} \in [a, b]$  e definiamo il polinomio

$$\bar{p}(t) = p\left(\frac{a(1-t)+b(1+t)}{2}\right).$$

Allora  $\bar{p} > 0$  su  $[-1, 1]$  e per il Teorema 1.3.8 di Bernstein esistono  $m \in \mathbb{N}$  e  $\{a'_{k,l}\}_{k,l \in \{0,1,\dots,m\}} \subset [0, +\infty)$  t.c.

$$\bar{p}(t) = \sum_{k,l=0}^m a'_{k,l} (1-t)^k (1+t)^l. \quad (1.10)$$

Ponendo  $t = \frac{2x-(a+b)}{b-a}$  l'eq. (1.10) diventa

$$p(x) = \sum_{k,l=0}^m a'_{k,l} \left( \frac{2}{b-a} \right)^{k+l} (b-x)^k (x-a)^l. \quad \square$$

## 1.4 Soluzione al problema dei momenti su intervalli di $\mathbb{R}$

Storicamente è usata la seguente terminologia per indicare i vari casi del problema dei momenti su  $\mathbb{R}$ , i quali si distinguono per la scelta dell'intervallo chiuso  $K$  su cui si vuole concentrata la misura soluzione:

- il **problema dei momenti di Hamburger** è il caso in cui  $K = \mathbb{R}$ ;
- il **problema dei momenti di Stieltjes** è il caso in cui  $K = [0, +\infty)$ ;
- il **problema dei momenti di Stieltjes generalizzato** è il caso in cui  $K = [c, +\infty)$  con  $c \in \mathbb{R}$ ;
- il **problema dei momenti di Hausdorff** è il caso in cui  $K = [0, 1]$ ;
- il **problema dei momenti di Hausdorff generalizzato** è il caso in cui  $K = [a, b]$  con  $a, b \in \mathbb{R}$   $a < b$ .

Inoltre in tutto il resto della sezione sarà fissata  $\mathcal{M} = \{x^i\}_{i \in \mathbb{N}}$  come base di  $\mathbb{R}[x]$ . Quindi data una successione di numeri reali  $\mathfrak{s}$  indicheremo con  $L_{\mathfrak{s}}$  al posto di  $L_{\mathfrak{s}, \mathcal{M}}$  il funzionale di Riesz agente su  $\mathbb{R}[x]$  associato ad  $\mathfrak{s}$  rispetto alla base  $\mathcal{M}$ .

**Definizione 1.4.1:** Una successione di numeri reali  $(s_n)_{n \in \mathbb{N}}$  è **semidefinita positiva** se  $\forall n \in \mathbb{N}$  e  $\forall \beta_0, \beta_1, \dots, \beta_n \in \mathbb{R}$  vale

$$\sum_{k,l=0}^n s_{k+l} \beta_k \beta_l \geq 0. \quad (1.11)$$

Indicheremo con  $\mathcal{P}\mathfrak{s}(\mathbb{N})$  la famiglia delle successioni reali semidefinite positive.

**Definizione 1.4.2:** Chiameremo  $E$  l'operatore di **shift** sulle successioni di numeri reali. Ossia se  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  è una successione di numeri reali  $E\mathfrak{s} = (s_{n+1})_{n \in \mathbb{N}}$ .

**Osservazione 1.4.3:** Sia  $\mathfrak{s}$  una successione di numeri reali e sia  $L_{\mathfrak{s}}$  il funzionale di Riesz ad essa associato. Vale  $L_{E\mathfrak{s}}(p) = L_{\mathfrak{s}}(xp) \quad \forall p \in \mathbb{R}[x]$ .

**Definizione 1.4.4:** Presa  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali la matrice

$$H_n(\mathfrak{s}) = \begin{pmatrix} s_0 & s_1 & \dots & s_n \\ s_1 & s_2 & \dots & s_{n+1} \\ \dots & \dots & \dots & \dots \\ s_n & s_{n+1} & \dots & s_{2n} \end{pmatrix}$$

è detta **matrice di Hankel**.

Siamo pronti per esporre i principali risultati della sezione, i quali risolveranno la questione riguardante l'esistenza di una misura soluzione per i vari tipi di problema dei momenti su intervalli chiusi di  $\mathbb{R}$ . Si esibiranno criteri di esistenza derivanti dal fondamentale Teorema 1.2.14 di Haviland e da essi si arriverà anche a criteri potenzialmente più pratici che si baseranno sulla segnatura delle matrici di Hankel.

**Teorema 1.4.5 (Soluzione al problema dei momenti di Hamburger):** *Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali. Le seguenti affermazioni sono equivalenti:*

- (1)  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti;
- (2)  $\mathfrak{s} \in \text{Ps}(\mathbb{N})$ ;
- (3) la matrice  $H_n(\mathfrak{s})$  è semidefinita positiva per ogni  $n \in \mathbb{N}$ ;
- (4) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è Pos( $\mathbb{R}$ )-positivo;
- (5) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è t.c.  $L_{\mathfrak{s}}(p^2) \geq 0 \quad \forall p \in \mathbb{R}[x]$ .

*Dimostrazione.* Dal Teorema 1.2.14 di Haviland segue l'implicazione (4)  $\Rightarrow$  (1). E per quanto provato nella Proposizione 1.3.1 (1) si ha anche l'equivalenza tra (4) e (5). Supponiamo che valga (1) e prendiamo  $n \in \mathbb{N}$  e  $\beta_0, \dots, \beta_n \in \mathbb{R}$ , allora se  $p(x) = \sum_{j=0}^n \beta_j x^j$  si ha

$$\begin{aligned} \sum_{k,l=0}^n s_{k+l} \beta_k \beta_l &= \sum_{k,l=0}^n \int_{\mathbb{R}} x^{k+l} d\mu(x) \beta_k \beta_l \\ &= \int_{\mathbb{R}} \left( \sum_{k,l=0}^n x^{k+l} \beta_k \beta_l \right) d\mu(x) \\ &= \int_{\mathbb{R}} \left( \sum_{k,l=0}^n \beta_k x^k \beta_l x^l \right) d\mu(x) \\ &= \int_{\mathbb{R}} p(x)^2 d\mu(x) \geq 0 \end{aligned}$$



che prova l'implicazione (1)  $\Rightarrow$  (2). L'equivalenza tra (2) e (3) si ottiene facilmente notando che  $H_n(\mathfrak{s})$  è la matrice associata alla forma quadratica definita dall'eq. (1.11). In conclusione notiamo che preso  $p(x) = \sum_{j=0}^n a_j x^j \in \mathbb{R}[x]$  si ha

$$L_{\mathfrak{s}}(p^2) = \sum_{k,l=0}^n a_k a_l L_{\mathfrak{s}}(x^{k+l}) = \sum_{k,l=0}^n a_k a_l s_{k+l}$$

che prova l'equivalenza tra (2) e (5).  $\square$

**Teorema 1.4.6 (Soluzione al problema dei momenti di Stieljes generalizzato):**

Siano  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali e  $c \in \mathbb{R}$ . Le seguenti affermazioni sono equivalenti:

- (1)  $\mathfrak{s}$  è una successione di  $[c, +\infty)$ -momenti;
- (2)  $\mathfrak{s} \in \text{Ps}(\mathbb{N})$  e  $(E\mathfrak{s} - c\mathfrak{s}) \in \text{Ps}(\mathbb{N})$ ;
- (3) le matrici  $H_n(\mathfrak{s})$  e  $H_n(E\mathfrak{s} - c\mathfrak{s})$  sono semidefinite positive per ogni  $n \in \mathbb{N}$ ;
- (4) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è  $\text{Pos}([c, +\infty))$ -positivo;
- (5) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è t.c.  $L_{\mathfrak{s}}(p^2) \geq 0$  e  $L_{\mathfrak{s}}((x-c)p^2) \geq 0 \quad \forall p \in \mathbb{R}[x]$ .

*Dimostrazione.* Dal Teorema 1.2.14 di Haviland segue l'implicazione (4)  $\Rightarrow$  (1) e per quanto provato nel Corollario 1.3.2 si ha anche l'equivalenza tra (4) e (5). Supponiamo che valga (1) e prendiamo  $n \in \mathbb{N}$  e  $\beta_0, \dots, \beta_n \in \mathbb{R}$ , allora se  $p(x) = \sum_{j=0}^n \beta_j x^j$  si ha

$$\begin{aligned} \sum_{k,l=0}^n s_{k+l} \beta_k \beta_l &= \sum_{k,l=0}^n \int_c^{+\infty} x^{k+l} d\mu(x) \beta_k \beta_l \\ &= \int_c^{+\infty} \left( \sum_{k,l=0}^n x^{k+l} \beta_k \beta_l \right) d\mu(x) \\ &= \int_c^{+\infty} \left( \sum_{k,l=0}^n \beta_k x^k \beta_l x^l \right) d\mu(x) \\ &= \int_c^{+\infty} p(x)^2 d\mu(x) \geq 0 \end{aligned}$$

e

$$\begin{aligned}
\sum_{k,l=0}^n (s_{k+l+1} - cs_{k+l})\beta_k\beta_l &= \int_c^{+\infty} \left( \sum_{k,l=0}^n (x^{k+l+1} - cx^{k+l})\beta_k\beta_l \right) d\mu(x) \\
&= \int_c^{+\infty} (x-c) \left( \sum_{k,l=0}^n x^{k+l}\beta_k\beta_l \right) d\mu(x) \\
&= \int_c^{+\infty} (x-c)p(x)^2 d\mu(x) \geq 0
\end{aligned}$$

che provano l'implicazione (1)  $\Rightarrow$  (2). L'equivalenza tra (2) e (3) si ottiene facilmente notando che  $H_n(\mathfrak{s})$  è la matrice associata alla forma quadratica definita dall'eq. (1.11). In conclusione notiamo che preso  $p(x) = \sum_{j=0}^n a_j x^j \in \mathbb{R}[x]$  si ha

$$L_{\mathfrak{s}}(p^2) = \sum_{k,l=0}^n a_k a_l L_{\mathfrak{s}}(x^{k+l}) = \sum_{k,l=0}^n a_k a_l s_{k+l}$$

e

$$\begin{aligned}
L_{\mathfrak{s}}((x-c)p^2) &= \sum_{k,l=0}^n a_k a_l L_{\mathfrak{s}}((x-c)x^{k+l}) \\
&= \sum_{k,l=0}^n a_k a_l \left( L_{\mathfrak{s}}(x^{k+l+1}) + cL_{\mathfrak{s}}(x^{k+l}) \right) \\
&= \sum_{k,l=0}^n a_k a_l (s_{k+l+1} - cs_{k+l})
\end{aligned}$$

che prova l'equivalenza tra (2) e (5). □

**Osservazione 1.4.7:** Notiamo che questo da l'esistenza di una misura soluzione anche al problema dei momenti su intervalli del tipo  $(-\infty, -d]$  con  $d \in \mathbb{R}$ , infatti basta considerare la misura  $\mu$  soluzione al problema su  $[d, +\infty)$  e prendere  $\nu : \mathcal{B}((-\infty, -d]) \rightarrow [0, +\infty]$  la misura definita da  $\nu(B) = \mu(-B) \quad \forall B \in \mathcal{B}((-\infty, -d])$ . Allora ovviamente  $\nu$  risolve il problema su  $(-\infty, -d]$ .

**Teorema 1.4.8 (Soluzione al problema dei momenti di Hausdorff generalizzato):** *Siano  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali e  $a, b \in \mathbb{R}$ ,  $a < b$ . Le seguenti affermazioni sono equivalenti:*

(1)  $\mathfrak{s}$  è una successione di  $[a, b]$ -momenti;

- (2)  $\mathfrak{s} \in \text{Ps}(\mathbb{N})$  e  $((a+b)E\mathfrak{s} - E(E\mathfrak{s}) - ab\mathfrak{s}) \in \text{Ps}(\mathbb{N})$ ;
- (3) le matrici  $H_n(\mathfrak{s})$  e  $H_n((a+b)E\mathfrak{s} - E(E\mathfrak{s}) - ab\mathfrak{s})$  sono semidefinite positive per ogni  $n \in \mathbb{N}$ ;
- (4) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è  $\text{Pos}([a, b])$ -positivo;
- (5) il funzionale lineare  $L_{\mathfrak{s}}$  su  $\mathbb{R}[x]$  è t.c.  $L_{\mathfrak{s}}(p^2) \geq 0$  e  $L_{\mathfrak{s}}((b-x)(x-a)p^2) \geq 0 \quad \forall p \in \mathbb{R}[x]$ .

*Dimostrazione.* Dal teorema 1.2.14 di Haviland segue l'implicazione (4)  $\Rightarrow$  (1). E per quanto provato nella proposizione 1.3.6 si ha anche l'equivalenza tra (4) e (5). Supponiamo che valga (1) e prendiamo  $n \in \mathbb{N}$  e  $\beta_0, \dots, \beta_n \in \mathbb{R}$ , allora se  $p(x) = \sum_{j=0}^n \beta_j x^j$  si ha

$$\begin{aligned} \sum_{k,l=0}^n s_{k+l} \beta_k \beta_l &= \sum_{k,l=0}^n \int_a^b x^{k+l} d\mu(x) \beta_k \beta_l \\ &= \int_a^b \left( \sum_{k,l=0}^n x^{k+l} \beta_k \beta_l \right) d\mu(x) \\ &= \int_a^b \left( \sum_{k,l=0}^n \beta_k x^k \beta_l x^l \right) d\mu(x) = \int_a^b p(x)^2 d\mu(x) \geq 0 \end{aligned}$$

e

$$\begin{aligned} &\sum_{k,l=0}^n ((a+b)s_{k+l+1} - s_{k+l+2} - abs_{k+l}) \beta_k \beta_l \\ &= \int_a^b \left( \sum_{k,l=0}^n ((a+b)x^{k+l+1} - x^{k+l+2} - abx^{k+l}) \beta_k \beta_l \right) d\mu(x) \\ &= \int_a^b ((a+b)x - x^2 - ab) \left( \sum_{k,l=0}^n x^{k+l} \beta_k \beta_l \right) d\mu(x) \\ &= \int_a^b (x-a)(b-x)p(x)^2 d\mu(x) \geq 0 \end{aligned}$$

che provano l'implicazione (1)  $\Rightarrow$  (2). L'equivalenza tra (2) e (3) si ottiene facilmente notando che  $H_n(\mathfrak{s})$  è la matrice associata alla forma quadratica definita dall'eq. (1.11). In conclusione notiamo che preso  $p(x) = \sum_{j=0}^n a_j x^j \in \mathbb{R}[x]$  si ha

$$L_{\mathfrak{s}}(p^2) = \sum_{k,l=0}^n a_k a_l L_{\mathfrak{s}}(x^{k+l}) = \sum_{k,l=0}^n a_k a_l s_{k+l}$$

e

$$\begin{aligned}
L_{\mathfrak{s}}((b-x)(x-a)p^2) &= \sum_{k,l=0}^n a_k a_l L_{\mathfrak{s}}((b-x)(x-a)x^{k+l}) \\
&= \sum_{k,l=0}^n \left( (a+b)L_{\mathfrak{s}}(x^{k+l+1}) - L_{\mathfrak{s}}(x^{k+l+2}) - abL_{\mathfrak{s}}(x^{k+l}) \right) \\
&= \sum_{k,l=0}^n ((a+b)s_{k+l+1} - s_{k+l+2} - abs_{k+l})
\end{aligned}$$

che prova l'equivalenza tra (2) e (5).  $\square$

Dal Teorema 1.3.8 di Bernstein si può derivare un criterio per il problema dei momenti di Hausdorff generalizzato che non si basa sui quadrati di polinomi.

**Teorema 1.4.9:** *Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali. Allora  $\mathfrak{s}$  è una successione di  $[a, b]$ -momenti se e solo se*

$$L_{\mathfrak{s}}((b-x)^n(x-a)^k) \geq 0 \quad \forall n, k \in \mathbb{N}. \quad (1.12)$$

*Dimostrazione.* Se  $\mathfrak{s}$  è una successione di  $[a, b]$ -momenti per il Teorema 1.4.8  $L_{\mathfrak{s}}$  è  $\text{Pos}([a, b])$ -positivo e per ogni  $n, k \in \mathbb{N}$  il polinomio  $(b-x)^n(x-a)^k \in \text{Pos}([a, b])$ , da cui segue (1.12).

Viceversa se vale (1.12), per il Corollario 1.3.9 del Teorema 1.3.8 di Bernstein,  $L_{\mathfrak{s}}$  è positivo su tutti i polinomi strettamente positivi su  $[a, b]$ . Inoltre preso  $p \in \text{Pos}([a, b])$  e considerato un qualunque  $\varepsilon > 0$  si ha che  $p + \varepsilon \mathbf{1}_{\mathbb{R}}$  è un polinomio strettamente positivo su  $[a, b]$ . Di conseguenza

$$L_{\mathfrak{s}}(p) + \varepsilon L_{\mathfrak{s}}(\mathbf{1}_{\mathbb{R}}) = L_{\mathfrak{s}}(p + \varepsilon \mathbf{1}_{\mathbb{R}}) \geq 0 \quad \forall \varepsilon > 0. \quad (1.13)$$

Prendendo il limite per  $\varepsilon \rightarrow 0^+$  nell'eq. (1.13) si ottiene  $L_{\mathfrak{s}}(p) \geq 0$ . Quindi  $L_{\mathfrak{s}}$  è  $\text{Pos}([a, b])$ -positivo. Infine applichiamo il Teorema 1.2.14 di Haviland per concludere che  $\mathfrak{s}$  è una successione di  $[a, b]$ -momenti.  $\square$

Infine il Teorema 1.4.9 ci permette di dimostrare il seguente famoso risultato riguardante il problema dei momenti di Hausdorff.

**Definizione 1.4.10:** Una successione di numeri reali  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  è detta **completamente monotona** se

$$((I - E)^n \mathfrak{s})_k = \sum_{j=0}^n (-1)^j \binom{n}{j} s_{k+j} \geq 0 \quad \forall k, n \in \mathbb{N}$$

in cui  $I$  è l'operatore identità sulle successioni di numeri reali,  $I\mathfrak{s} = \mathfrak{s}$ .

**Teorema 1.4.11 (di Hausdorff):** *Una successione di numeri reali  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  è una successione di  $[0, 1]$ -momenti se e solo se è completamente monotona.*

*Dimostrazione.* Dal Teorema 1.4.9 si ha che  $\mathfrak{s}$  è una successione di  $[0, 1]$ -momenti se e solo se  $L_{\mathfrak{s}}((1-x)^n x^k) \geq 0 \quad \forall k, n \in \mathbb{N}$ , che equivale alla tesi. Infatti vale

$$\begin{aligned} L_{\mathfrak{s}}((1-x)^n x^k) &= \sum_{j=0}^n (-1)^j \binom{n}{j} L_{\mathfrak{s}}(x^{k+j}) \\ &= \sum_{j=0}^n (-1)^j \binom{n}{j} s_{k+j} \\ &= \sum_{j=0}^n (-1)^j \binom{n}{j} (E^j \mathfrak{s})_k \\ &= ((I - E)^n \mathfrak{s})_k. \end{aligned}$$

□



## Problema dei momenti: Unicità

Passiamo adesso a trattare la seconda parte del problema dei momenti, quella riguardante l'unicità della misura soluzione.

Nella prima sezione del capitolo dimostreremo il teorema di Denjoy-Carleman, strumento che si rivelerà fondamentale nella seconda sezione in cui si studieranno vari criteri per l'unicità nel caso unidimensionale.

Infine nella terza ed ultima parte del capitolo si presenterà il teorema di Petersen, un interessante criterio per l'unicità nel caso multidimensionale.

### 2.1 Teorema di Denjoy-Carleman

**Definizione 2.1.1:** Una successione di numeri reali positivi  $(x_n)_{n \in \mathbb{N}}$  è detta **logaritmicamente convessa**, o in breve **log-convessa**, se  $\forall n \in \mathbb{N}_+$  vale la disuguaglianza

$$x_n^2 \leq x_{n-1}x_{n+1}.$$

**Osservazione 2.1.2:** Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di  $\mathbb{R}$ -momenti t.c.  $s_{2n} \neq 0$  per ogni  $n \in \mathbb{N}$ , allora la successione  $(\sqrt{s_{2n}})_{n \in \mathbb{N}}$  è una successione log-convessa. Infatti, se  $\mu$  è una misura boreliana con successione dei momenti  $\mathfrak{s}$ , usando la disuguaglianza di Cauchy-Schwarz si ottiene

$$\begin{aligned} s_{2n} &= \int_{\mathbb{R}} x^{2n} d\mu(x) \\ &\leq \left( \int_{\mathbb{R}} x^{2(n-1)} d\mu(x) \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}} x^{2(n+1)} d\mu(x) \right)^{\frac{1}{2}} = \sqrt{s_{2(n-1)}} \sqrt{s_{2(n+1)}}. \end{aligned}$$

Il teorema della sezione è il seguente.

**Teorema 2.1.3 (di Denjoy-Carleman):** Sia  $\mathbf{m} = (M_n)_{n \in \mathbb{N}}$  una successione di numeri reali positivi log-convessa. Supponiamo che valga la condizione

$$\sum_{n \in \mathbb{N}_+} M_n^{-\frac{1}{n}} = +\infty. \quad (2.1)$$

Se  $f \in C^\infty(\mathbb{R})$  è t.c.  $|f^{(n)}(x)| \leq KM_n \quad \forall x \in \mathbb{R} \quad \forall n \in \mathbb{N}$  con  $K > 0$  costante ed esiste  $a \in \mathbb{R}$  per il quale  $f^{(n)}(a) = 0 \quad \forall n \in \mathbb{N}$  allora  $f(x) = 0 \quad \forall x \in \mathbb{R}$ .

Prima di affrontare la dimostrazione vera e propria abbiamo bisogno di costruire gli strumenti su cui si questa si baserà.

Chiamiamo  $\mathbb{R}^{\mathbb{N}}$  l'insieme delle successioni di numeri reali, questo insieme ha una naturale struttura di spazio vettoriale reale con le operazioni di somma e prodotto per scalari definite componente per componente.

Fissiamo  $(\epsilon_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  una qualsiasi successione positiva, infinitesima e strettamente decrescente.

**Definizione 2.1.4:** Per ogni  $\mathbf{u} = (u_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  definiamo

$$\|\mathbf{u}\| = \inf_{k \in \mathbb{N}} [\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\}]. \quad (2.2)$$

**Lemma 2.1.5:** Se  $\mathbf{u} = (u_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  soddisfa una condizione del tipo

$$\epsilon_{k_1} \leq \|\mathbf{u}\| \leq \epsilon_{k_2} \quad \text{in cui } k_1, k_2 \in \mathbb{N}, \quad k_2 \leq k_1$$

allora esiste un  $k \in \mathbb{N}$ ,  $k_2 \leq k \leq k_1$  t.c.

$$\|\mathbf{u}\| = \max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\}. \quad (2.3)$$

*Dimostrazione.* L'ipotesi  $\|\mathbf{u}\| \leq \epsilon_{k_2}$  implica che un  $p < k_2$  non può realizzare (2.3). Infatti se  $p < k_2$  vale

$$\|\mathbf{u}\| \leq \epsilon_{k_2} < \epsilon_p \leq \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\}.$$

Invece per  $p > k_1$  si ha

$$\epsilon_p < \epsilon_{k_1} \leq \|\mathbf{u}\| \leq \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\}$$

che implica

$$\max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\} = \max_{0 \leq n \leq p} |u_n| \quad \forall p > k_1$$



ma il membro di destra dell'equazione appena scritta è termine di una successione non decrescente in  $p$ , quindi necessariamente esiste un  $k \in \mathbb{N}$   $k_2 \leq k \leq k_1$  che realizza (2.3).  $\square$

**Osservazione 2.1.6:** Notiamo che nel caso in cui  $\mathbf{u} \in \mathbb{R}^{\mathbb{N}}$  è t.c.  $\|\mathbf{u}\| > \epsilon_0$  si ha

$$\|\mathbf{u}\| = |u_0| = \max\{\epsilon_0, |u_0|\}.$$

Infatti  $\|\mathbf{u}\| > \epsilon_0$  implica

$$\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\} > \epsilon_0 \quad \forall k \in \mathbb{N}$$

ma  $\epsilon_k \leq \epsilon_0 \quad \forall k \in \mathbb{N}$ , dunque

$$\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\} = \max_{0 \leq n \leq k} |u_n| \quad \forall k \in \mathbb{N}$$

e quindi effettivamente

$$\|\mathbf{u}\| = \inf_{k \in \mathbb{N}} [\max_{0 \leq n \leq k} |u_n|] = |u_0| = \max\{\epsilon_0, |u_0|\}.$$

**Lemma 2.1.7:** *La funzione  $\|\cdot\| : \mathbb{R}^{\mathbb{N}} \rightarrow [0, +\infty)$  gode delle seguenti proprietà:*

- (1)  $\|\mathbf{u}\| \geq 0 \quad \forall \mathbf{u} \in \mathbb{R}^{\mathbb{N}}$ ;
- (2)  $\|\mathbf{u}\| = 0$  se e solo se  $\mathbf{u} = \underline{0}$ , in cui  $\underline{0}$  è la successione identicamente nulla;
- (3)  $\|-\mathbf{u}\| = \|\mathbf{u}\| \quad \forall \mathbf{u} \in \mathbb{R}^{\mathbb{N}}$ ;
- (4)  $\|t\mathbf{u}\| \leq |t|\|\mathbf{u}\| \quad \forall \mathbf{u} \in \mathbb{R}^{\mathbb{N}}, \quad \forall t \in \mathbb{R} \quad \text{con } |t| < 1$ ;
- (5)  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ .

*Dimostrazione.* Le proprietà (1),(3) e (4) seguono direttamente dalla definizione di  $\|\cdot\|$ . Dimostriamo (2). Se  $\mathbf{u} = \underline{0}$  allora  $\|\mathbf{u}\| = 0$ . Viceversa se  $\mathbf{u} \in \mathbb{R}^{\mathbb{N}}$  t.c.  $\mathbf{u} \neq \underline{0}$ , esisterà un  $k_0 \in \mathbb{N}$  t.c.  $|u_{k_0}| > 0$ . Ma per  $k \geq k_0$

$$\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\} \geq |u_{k_0}|$$

che implica

$$\inf_{k \geq k_0} [\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\}] \geq |u_{k_0}| > 0$$

inoltre essendo  $\epsilon_k > 0$  per ogni  $k \in \mathbb{N}$  si ha anche

$$\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\} \geq \epsilon_k > 0 \quad \forall k \in \{0, \dots, k_0\}.$$

Mettendo insieme si ottiene

$$\|\mathbf{u}\| = \inf_{k \in \mathbb{N}} [\max\{\epsilon_k, \max_{0 \leq n \leq k} |u_n|\}] > 0$$

che conclude la dimostrazione di (2). Dimostriamo infine (5). Notiamo che se almeno uno tra  $\mathbf{u}$  e  $\mathbf{v}$  è  $\underline{0}$  la disuguaglianza è ovvia. Siano quindi  $\mathbf{u}, \mathbf{v} \neq \underline{0}$ . Per (2) si ha  $\|\mathbf{u}\|, \|\mathbf{v}\| > 0$ , dunque il Lemma 2.1.5 e l'Osservazione 2.1.6 garantiscono l'esistenza di  $p, q \in \mathbb{N}$  per i quali, se  $\mathbf{u} = \{u_n\}_{n \in \mathbb{N}}$  e  $\mathbf{v} = \{v_n\}_{n \in \mathbb{N}}$ , vale

$$\|\mathbf{u}\| = \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\} \quad \text{e} \quad \|\mathbf{v}\| = \max\{\epsilon_q, \max_{0 \leq n \leq q} |v_n|\}$$

e senza perdita di generalità posso supporre  $p \leq q$ . Quindi per  $n \in \mathbb{N}$ ,  $n \leq p \leq q$  vale

$$\begin{aligned} |u_n + v_n| &\leq |u_n| + |v_n| \\ &\leq \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\} + \max\{\epsilon_q, \max_{0 \leq n \leq q} |v_n|\} \\ &= \|\mathbf{u}\| + \|\mathbf{v}\| \end{aligned}$$

da cui segue

$$\max_{0 \leq n \leq p} |u_n + v_n| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Inoltre vale anche

$$\epsilon_p \leq \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n|\} = \|\mathbf{u}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Dunque abbiamo

$$\max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n + v_n|\} \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

che ci permette di scrivere

$$\|\mathbf{u} + \mathbf{v}\| \leq \max\{\epsilon_p, \max_{0 \leq n \leq p} |u_n + v_n|\} \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

ossia (5). □

Adesso dimostriamo un'importante proprietà di continuità per la funzione  $\|\cdot\|$ .

**Lemma 2.1.8:** Sia  $(\mathbf{u}^{(m)})_{m \in \mathbb{N}}$  una successione in  $\mathbb{R}^{\mathbb{N}}$ , in cui  $\mathbf{u}^{(m)} = (u_i^{(m)})_{i \in \mathbb{N}}$  per ogni  $m \in \mathbb{N}$ . Supponiamo che valga

$$u_i^{(m)} \longrightarrow u_i \in \mathbb{R} \quad \text{per } m \rightarrow +\infty \quad \forall i \in \mathbb{N}.$$

Allora presa  $\mathbf{u} = (u_i)_{i \in \mathbb{N}}$  si ha

$$\|\mathbf{u}^{(m)} - \mathbf{u}\| \longrightarrow 0 \quad \text{per } m \rightarrow +\infty.$$

*Dimostrazione.* Fissiamo  $\delta > 0$ . Allora esiste  $p \in \mathbb{N}$  t.c.  $\epsilon_p < \delta$ . Essendo  $u_i^{(m)} \longrightarrow u_i \in \mathbb{R}$  per  $m \rightarrow +\infty$ , è garantita l'esistenza di uno  $z_i \in \mathbb{N}$  t.c.

$$|u_i^{(m)} - u_i| < \epsilon_p \quad \forall m \geq z_i \quad \forall i \in \mathbb{N}.$$

Dunque preso  $q \in \mathbb{N}$ ,  $q > p$  e definito  $z'_q = \max\{z_0, z_1, \dots, z_q\}$ , si ottiene

$$\max_{0 \leq n \leq q} |u_n^{(m)} - u_n| < \epsilon_p \quad \forall m \geq z'_q$$

che implica

$$\max\{\epsilon_q, \max_{0 \leq n \leq q} |u_n^{(m)} - u_n|\} < \max\{\epsilon_q, \epsilon_p\} = \epsilon_p.$$

Come conseguenza di quanto appena detto, preso un qualsiasi  $q \in \mathbb{N}$ ,  $q > p$ , si ha

$$\|\mathbf{u}^{(m)} - \mathbf{u}\| \leq \max\{\epsilon_q, \max_{0 \leq n \leq q} |u_n^{(m)} - u_n|\} < \epsilon_p < \delta \quad \forall m \geq z'_q$$

che prova la tesi. □

**Proposizione 2.1.9:** Per ogni  $n \in \mathbb{N}$  sia  $u_n : \mathbb{R} \rightarrow \mathbb{R}$  una funzione reale continua. Allora la funzione  $g : \mathbb{R} \rightarrow \mathbb{R}$  definita da  $g(x) = \|(u_n(x))_{n \in \mathbb{N}}\|$  è continua.

*Dimostrazione.* Fissiamo  $x_0 \in \mathbb{R}$  e prendiamo  $(y_m)_{m \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  t.c.  $y_m \rightarrow x_0$  per  $m \rightarrow +\infty$ . Per continuità  $u_n(y_m) \rightarrow u_n(x_0)$  per  $m \rightarrow +\infty$  per ogni  $n \in \mathbb{N}$ , dunque usando il Lemma 2.1.8 si ottiene

$$\|(u_n(y_m))_{n \in \mathbb{N}} - (u_n(x_0))_{n \in \mathbb{N}}\| \longrightarrow 0 \quad \text{per } m \rightarrow +\infty$$

e, grazie anche alla proprietà (5) del Lemma 2.1.7, possiamo scrivere

$$\begin{aligned} |g(y_m) - g(x_0)| &= \left| \|(u_n(y_m))_{n \in \mathbb{N}}\| - \|(u_n(x_0))_{n \in \mathbb{N}}\| \right| \\ &\leq \|(u_n(y_m))_{n \in \mathbb{N}} - (u_n(x_0))_{n \in \mathbb{N}}\| \longrightarrow 0 \quad \text{per } m \rightarrow +\infty. \end{aligned}$$

L'arbitrarietà della successione  $(y_m)_{m \in \mathbb{N}}$  dimostra la continuità di  $g$  in  $x_0$  ed essendo  $x_0$  un qualunque punto di  $\mathbb{R}$  segue la tesi. □

Con questa ultima proposizione abbiamo concluso la costruzione degli strumenti che ci serviranno.

Passiamo quindi a dimostrare due ulteriori lemmi fondamentali per la nostra dimostrazione.

**Lemma 2.1.10 (Disuguaglianza di Carleman):** *Se  $(a_n)_{n \in \mathbb{N}_+}$  è una successione di numeri reali positivi vale la disuguaglianza*

$$\sum_{n \in \mathbb{N}_+} (a_1 a_2 \dots a_n)^{\frac{1}{n}} \leq e \sum_{n \in \mathbb{N}_+} a_n.$$

*Dimostrazione.* Siano  $(c_n)_{n \in \mathbb{N}}$  dei numeri reali positivi, allora per la nota disuguaglianza tra media geometrica e media aritmetica si ha

$$(a_1 \dots a_n)^{\frac{1}{n}} = (c_1 \dots, c_n)^{-\frac{1}{n}} (c_1 a_1 \dots c_n a_n)^{\frac{1}{n}} \leq (c_1 \dots, c_n)^{-\frac{1}{n}} \frac{1}{n} \sum_{m=1}^n c_m a_m.$$

Prendendo  $c_n = \frac{(n+1)^n}{n^{n-1}}$  si ha  $(c_1 \dots c_n)^{\frac{1}{n}} = n + 1$  e

$$\begin{aligned} \sum_{n \in \mathbb{N}_+} (a_1 a_2 \dots a_n)^{\frac{1}{n}} &\leq \sum_{n \in \mathbb{N}_+} \frac{1}{n(n+1)} \sum_{m=1}^n c_m a_m = \\ &= \sum_{n \in \mathbb{N}_+} \frac{1}{n(n+1)} \sum_{m=1}^n \left(1 + \frac{1}{m}\right)^m m a_m \\ &\leq e \sum_{m \in \mathbb{N}_+} a_m \end{aligned}$$

□

**Proposizione 2.1.11:** *Sia  $\mathbf{m} = (M_n)_{n \in \mathbb{N}}$  una successione di numeri reali positivi log-convessa. Allora le seguenti due condizioni sono equivalenti:*

- (1)  $\sum_{n \in \mathbb{N}_+} M_n^{-\frac{1}{n}} = +\infty;$
- (2)  $\sum_{n \in \mathbb{N}_+} \frac{M_{n-1}}{M_n} = +\infty.$

*Dimostrazione.* L'ipotesi di convessità logaritmica ci garantisce che la successione  $\left(\frac{M_{n-1}}{M_n}\right)_{n \in \mathbb{N}_+}$  sia non crescente. La non crescita appena provata ci fa ottenere

$$\left(\frac{M_{n-1}}{M_n}\right)^n \leq \frac{M_0}{M_n}$$

che implica

$$\frac{M_{n-1}}{M_n} \leq M_0^{\frac{1}{n}} M_n^{-\frac{1}{n}} \leq M_0 M_n^{-\frac{1}{n}}$$

da cui segue che (2) implica (1). Il viceversa segue facilmente dal Lemma 2.1.10 applicato con  $a_n = \frac{M_{n-1}}{M_n}$ .  $\square$

Fissiamo il contesto in cui lavoreremo d'ora in avanti. Siano  $f \in C^\infty(\mathbb{R})$  ed  $\mathbf{m} = (M_n)_{n \in \mathbb{N}}$  una successione di numeri reali positivi logicamente convessa per le quali valgano le ipotesi del Teorema 2.1.3 con  $K = 1$ .

Poniamo per ogni  $n \in \mathbb{N}$

$$u_n^{(f)}(x) = \frac{f^{(n)}(x)}{e^n M_n}$$

$$\epsilon_n = e^{-n}$$

e definiamo per ogni  $x \in \mathbb{R}$

$$\|f\|_x = \|(u_n^{(f)}(x))_{n \in \mathbb{N}}\| = \inf_{k \in \mathbb{N}} \left[ \max \left\{ e^{-k}, \max_{0 \leq n \leq k} \frac{|f^{(n)}(x)|}{e^n M_n} \right\} \right].$$

**Osservazione 2.1.12:** La funzione  $\mathbb{R} \ni x \mapsto \|f\|_x$  appena definita è una funzione continua per la Proposizione 2.1.9.

**Lemma 2.1.13:** Sia  $x \in \mathbb{R}$  e supponiamo  $\|f\|_x \geq e^{-p}$  per un qualche  $p \in \mathbb{N}$ . Per ogni  $h > 0$  vale

$$\|f\|_{x+h} \leq \|f\|_x \exp \left( e|h| \frac{M_{l_x}}{M_{l_x-1}} \right)$$

in cui  $l_x \in \mathbb{N}$  soddisfa

$$\|f\|_x = \max \left\{ e^{-l_x}, \max_{0 \leq n \leq l_x} \frac{|f^{(n)}(x)|}{e^n M_n} \right\}. \quad (2.4)$$

*Dimostrazione.* L'esistenza di un  $l_x \in \mathbb{N}$  che realizza (2.4) è garantita dal Lemma 2.1.5 o dall'Osservazione 2.1.6, in quanto  $\|f\|_x \geq e^{-p} > 0$  con  $p \in \mathbb{N}$ .

Vale la disuguaglianza

$$\|f\|_{x+h} \leq \max \left\{ e^{-l_x}, \max_{0 \leq n \leq l_x} \frac{|f^{(n)}(x+h)|}{e^n M_n} \right\}.$$

Se  $e^{-l_x} = \max \left\{ e^{-l_x}, \max_{0 \leq n \leq l_x} \frac{|f^{(n)}(x+h)|}{e^n M_n} \right\}$  allora

$$\|f\|_{x+h} \leq e^{-l_x} \leq \|f\|_x \leq \|f\|_x \exp \left( e|h| \frac{M_{l_x}}{M_{l_x-1}} \right)$$

ossia la tesi vale. In caso contrario comunque, per ogni  $n \in \mathbb{N}$ ,  $0 \leq n \leq l_x$ , si ha

$$\|f\|_{x+h} \leq \frac{f^{(n)}(x+h)}{e^n M_n}.$$

Sviluppiamo  $f^{(n)}$  in serie di Taylor nel punto  $x$  con incremento  $h$  ottenendo

$$\begin{aligned} |f^{(n)}(x+h)| &\leq \sum_{i=0}^{l_x-n-1} \frac{|h|^i}{i!} |f^{(n+i)}(x)| + \frac{|h|^{l_x-n}}{(l_x-n)!} |f^{(l_x)}(\zeta_{x,h})| \\ &\leq \|f\|_x \sum_{i=0}^{l_x-n} \frac{|h|^i}{i!} e^{n+i} M_{n+i} \end{aligned}$$

in cui si è usato che  $|f^{(n)}(x)| \leq e^n M_n \|f\|_x$  per ogni  $0 \leq n \leq l_x - 1$  e che  $|f^{(l_x)}(\zeta_{x,h})| \leq M_{l_x} \leq M_{l_x} \|f\|_x e^{l_x}$ . Quanto appena mostrato implica

$$\begin{aligned} \frac{|f^{(n)}(x+h)|}{e^n M_n} &\leq \|f\|_x \sum_{i=0}^{l_x-n} \frac{|h|^i}{i!} e^i \frac{M_{n+i}}{M_n} \\ &\leq \|f\|_x \sum_{i=0}^{l_x-n} \frac{|h|^i}{i!} e^i \left( \frac{M_{l_x}}{M_{l_x-1}} \right)^i \\ &\leq \|f\|_x \exp \left( e|h| \frac{M_{l_x}}{M_{l_x-1}} \right). \end{aligned}$$

in cui si è usata la convessità logaritmica di  $(M_n)_{n \in \mathbb{N}}$  per dire che  $\frac{M_{n+i}}{M_n} \leq \left( \frac{M_{l_x}}{M_{l_x-1}} \right)^i$  per ogni  $i \in \{0, \dots, l_x - n\}$ .  $\square$

Siamo finalmente pronti per la dimostrazione del Teorema 2.1.3.

*Dimostrazione del Teorema 2.1.3.* Supponiamo per assurdo che  $f$  soddisfi tutte le ipotesi ma che non sia identicamente nulla su  $\mathbb{R}$ . Senza perdita di generalità possiamo supporre  $K = 1$ . Per la proprietà (2) del Lemma 2.1.7 esiste un  $y \in \mathbb{R}$  t.c.

$$\|f\|_y > 0 = \|f\|_a.$$

Allora esiste il minimo  $p \in \mathbb{N}$  t.c.

$$0 < e^{-p} \leq \|f\|_y$$

sia  $p_0$  tale minimo. Per la Proposizione 2.1.9 la funzione  $g(x) = \|f\|_x$  è continua. Questo ci permette di utilizzare il teorema dei valori intermedi che ci assicura l'esistenza un  $x_0$  tra  $a$  e  $y$  t.c.

$$\|f\|_{x_0} = e^{-p_0}$$

ed usando allo stesso modo ripetutamente lo stesso teorema dei valori intermedi, posso costruire una successione di punti  $(x_j)_{j \in \mathbb{N}}$  strettamente decrescente, contenuti tra  $a$  e  $y$  t.c.

$$\|f\|_{x_j} = e^{-(p_0+j)} \quad \forall j \in \mathbb{N}.$$

Applicando il Lemma 2.1.13 con  $x = x_j$  e  $x + h = x_{j-1}$  per  $j \in \mathbb{N}_+$  otteniamo

$$e^{-(p_0+j-1)} \leq e^{-p_0-j} \exp \left\{ e|x_j - x_{j-1}| \frac{M_{l_j}}{M_{l_{j-1}}} \right\} \quad \forall j \in \mathbb{N}_+$$

che passando ai logaritmi diventa

$$\frac{M_{l_{j-1}}}{M_{l_j}} \leq e|x_j - x_{j-1}| \quad \forall j \in \mathbb{N}_+ \quad (2.5)$$

in cui per ogni  $j \in \mathbb{N}_+$  il numero  $l_j \in \mathbb{N}$  soddisfa

$$\|f\|_{x_j} = e^{-(p_0+j)} = \max \left\{ e^{-l_j}, \max_{0 \leq n \leq l_j} \frac{|f^{(n)}(x_j)|}{e^n M_n} \right\}.$$

Vale

$$e^{-(p_0+j)} = \|f\|_{x_j} \leq e^{-(p_0+j-1)}$$

dunque per il Lemma 2.1.5 si può supporre  $p_0 + j - 1 \leq l_j \leq p_0 + j$ . Di conseguenza, usando la convessità logaritmica di  $(M_n)_{n \in \mathbb{N}}$ , si ottiene

$$\frac{M_{l_{j-1}}}{M_{l_j}} \geq \frac{M_{p_0+j-1}}{M_{p_0+j}}$$

per cui l'eq. (2.5) implica

$$\frac{M_{p_0+j-1}}{M_{p_0+j}} \leq e|x_j - x_{j-1}| \quad \forall j \in \mathbb{N}_+.$$

Sommando sui  $j \in \mathbb{N}_+$  ed usando che  $\sum_{j \in \mathbb{N}_+} |x_j - x_{j-1}| \leq |a - x_0|$  si ottiene

$$\sum_{n=p_0+1}^{+\infty} \frac{M_{n-1}}{M_n} \leq e|x - a| < +\infty$$

che contraddice la divergenza della serie  $\sum_{n \in \mathbb{N}_+} \frac{M_{n-1}}{M_n}$  che, per il Lemma 2.1.11, equivale all'ipotesi  $\sum_{n \in \mathbb{N}_+} M_n^{\frac{1}{n}} = +\infty$ . Il teorema è quindi dimostrato.  $\square$

## 2.2 Unicità nel problema dei momenti su $\mathbb{R}$

**Definizione 2.2.1:** Sia  $K \subset \mathbb{R}$  un chiuso e  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali.  $\mathfrak{s}$  è una successione di  $K$ -momenti **determinata** se esiste un'unica misura di Radon  $\mu$  concentrata su  $K$  t.c.  $s_n = \int_K x^n d\mu(x) \quad \forall n \in \mathbb{N}$ .

**Esempio 2.2.2:** Non tutte le successioni di momenti sono determinate, un esempio è dato dalla successione dei momenti della distribuzione di probabilità log-normale data dalla densità

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathbb{1}_{(0,+\infty)}(x) \frac{1}{x} e^{-\frac{\log^2(x)}{2}}$$

rispetto alla misura di Lebesgue  $\mathcal{L}^1$  su  $\mathbb{R}$ , ossia  $\mu = f(x)d\mathcal{L}^1(x)$ . Infatti per  $n \in \mathbb{N}$ , ponendo  $y = \log(x)$  e poi  $t = y - n$  si ha

$$\begin{aligned} s_n &= \int_{\mathbb{R}} x^n f(x) dx = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^{n-1} e^{-\frac{\log^2(x)}{2}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ny} e^{-\frac{y^2}{2}} dy \\ &= e^{\frac{n^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(y-n)^2}{2}} dy = e^{\frac{n^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{t^2}{2}} dt = e^{\frac{n^2}{2}} \end{aligned}$$

dunque tale misura di probabilità ha successione dei momenti  $\mathfrak{s} = \left( e^{\frac{n^2}{2}} \right)_{n \in \mathbb{N}}$ .

Adesso definiamo, per  $c \in [-1, 1]$  la misura positiva  $\mu_c$  data da

$$\mu_c = [1 + c \sin(2\pi \log(x))] d\mu(x)$$

e notiamo che avendo  $\mu$  momenti finiti, anche  $\mu_c$  li ha finiti. Inoltre in modo simile a quanto fatto in precedenza si ottiene

$$\begin{aligned} \int_{\mathbb{R}} x^n \sin(2\pi \log(x)) d\mu(x) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ny} \sin(2\pi y) e^{-\frac{y^2}{2}} dy \\ &= e^{\frac{n^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(n-y)^2}{2}} \sin(2\pi y) dy = e^{\frac{n^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{t^2}{2}} \sin(2\pi(t+n)) dt = 0 \end{aligned}$$

in cui si è usato che la funzione  $\mathbb{R} \ni t \mapsto \sin(2\pi(t+n))$  è dispari. Di conseguenza la misura  $\mu_c$  ha come successione di momenti proprio  $\mathfrak{s}$  per ogni  $c \in [-1, 1]$ , dunque  $\mathfrak{s}$  non è determinata.

Iniziamo trattando il caso di misure concentrate su un compatto di  $\mathbb{R}$ .

**Lemma 2.2.3:** Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}} \subset \mathbb{R}$  una successione di  $\mathbb{R}$ -momenti e sia  $\mu$  una misura boreliana che ha  $\mathfrak{s}$  come successione dei momenti. Preso  $c > 0$  le seguenti affermazioni sono equivalenti:



- (1)  $\mu$  è concentrata su  $[-c, c]$ ;
- (2) esiste un numero  $d > 0$  t.c.  $|s_n| \leq dc^n \quad \forall n \in \mathbb{N}$ ;
- (3) esiste un numero  $d > 0$  t.c.  $s_{2n} \leq dc^{2n} \quad \forall n \in \mathbb{N}$ ;
- (4)  $\liminf_{n \rightarrow +\infty} s_{2n}^{\frac{1}{2n}} \leq c$ .

*Dimostrazione.* Le implicazioni (1)  $\Rightarrow$  (2)  $\Rightarrow$  (3)  $\Rightarrow$  (4) sono vere in modo ovvio con  $d = s_0 = \mu([-c, c]) > 0$ . Dunque basta dimostrare che (4) implica (1). Sia  $a > 0$  e definiamo  $M_a = \int_{\mathbb{R}} \mathbb{1}_{\mathbb{R} \setminus (-a, a)} d\mu$ . Allora per ogni  $n \in \mathbb{N}$

$$\begin{aligned} M_a a^{2n} &= \int_{\mathbb{R}} a^{2n} \mathbb{1}_{\mathbb{R} \setminus (-a, a)}(x) d\mu(x) \\ &\leq \int_{\mathbb{R}} x^{2n} \mathbb{1}_{\mathbb{R} \setminus (-a, a)}(x) d\mu(x) \\ &\leq \int_{\mathbb{R}} x^{2n} d\mu(x) = s_{2n} \end{aligned}$$

da cui si deduce  $M_a^{\frac{1}{2n}} a \leq s_{2n}^{\frac{1}{2n}}$  per ogni  $n \in \mathbb{N}$ . Sia  $S = \liminf_{n \rightarrow +\infty} s_{2n}^{\frac{1}{2n}}$ . Se  $M_a > 0$  allora passando al  $\liminf$  per  $n \rightarrow +\infty$  si ottiene  $a \leq S$ . Di conseguenza se  $a > S$  si ha necessariamente  $M_a = \mu(\mathbb{R} \setminus (-a, a)) = 0$ . Ma per ipotesi  $c \geq S$ , dunque per ogni  $b > c$  si ha  $\mu(\mathbb{R} \setminus (-b, b)) = 0$ . Quindi  $\mu$  risulta essere concentrata su  $[-S, S] \subset [-c, c]$ , da cui segue (1).  $\square$

**Teorema 2.2.4:** *Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  è una successione di numeri reali. Se esiste una misura boreliana  $\mu$  concentrata su un compatto di  $\mathbb{R}$  che ha  $\mathfrak{s}$  come successione di momenti allora  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti determinata.*

*Dimostrazione.* Siano  $\mu_1, \mu_2$  misure di Radon su  $\mathbb{R}$  che hanno  $\mathfrak{s}$  come successione dei momenti. Dalla dimostrazione del Lemma 2.2.3 (4)  $\Rightarrow$  (1) si evince che entrambe  $\mu_1$  e  $\mu_2$  sono concentrate su  $[-S, S]$  con  $S = \liminf_{n \rightarrow +\infty} s_{2n}^{\frac{1}{2n}}$ . Quindi per ogni  $p \in \mathbb{R}[x]$  si ha

$$\int_{-S}^S p(x) d\mu_1(x) = \int_{-S}^S p(x) d\mu_2(x).$$

Di conseguenza, essendo i polinomi uniformemente densi in  $C^0([-S, S], \mathbb{R})$  per il teorema di approssimazione di Stone-Weierstrass, si ha

$$\int_{-S}^S f(x) d\mu_1(x) = \int_{-S}^S f(x) d\mu_2(x)$$

per ogni funzione continua  $f$ . Questo in particolare ci permette di concludere che  $\mu_1 = \mu_2$ .  $\square$

Adesso passiamo a studiare più in generale i problemi dei momenti di Hamburger e di Stieltjes.

Gli strumenti che useremo saranno il Teorema 2.1.3 di Denjoy-Carleman presentato nella sezione precedente e la trasformata di Fourier di misure finite su  $\mathbb{R}$ . Ricordiamo la definizione e le proprietà per noi più importanti di quest'ultima.

**Definizione 2.2.5:** Sia  $\mu$  una misura boreliana finita su  $\mathbb{R}$ . La **trasformata di Fourier** della misura  $\mu$  è la funzione  $\hat{\mu} : \mathbb{R} \rightarrow \mathbb{C}$  definita da

$$\hat{\mu}(y) = \int_{\mathbb{R}} e^{-ixy} d\mu(x) \quad \forall y \in \mathbb{R}.$$

**Proposizione 2.2.6:** Sia  $\mu$  una misura boreliana su  $\mathbb{R}$  t.c.  $\int_{\mathbb{R}} |x|^n d\mu(x) < +\infty$  per ogni  $n \in \mathbb{N}$ . Allora  $\hat{\mu} \in C^\infty(\mathbb{R}, \mathbb{C})$  e vale

$$\hat{\mu}^{(n)}(y) = \int_{\mathbb{R}} (-ix)^n e^{-ixy} d\mu(x) \quad \forall n \in \mathbb{N}. \quad (2.6)$$

*Dimostrazione.* Dimostriamo per induzione su  $n \in \mathbb{N}$  t.c. per ogni  $n \in \mathbb{N}$  esiste  $\hat{\mu}^{(n)}$  e vale (2.6). Il caso  $n = 0$  segue dalla definizione di  $\hat{\mu}$ . Supponiamo quindi che esistano le derivate di  $\hat{\mu}$  fino all'ordine  $n$  e che per esse valga (2.6). Fissiamo  $y \in \mathbb{R}$  e consideriamo per ogni  $h \in \mathbb{R} \setminus \{0\}$  le funzioni

$$\begin{aligned} \varphi_h(x) &= \frac{e^{-ixh} - 1}{h} \\ \psi_h(x) &= \varphi_h(x)(-ix)^n e^{-ixy} \end{aligned}$$

ora  $\varphi_h(x) \rightarrow -ix$  per  $h \rightarrow 0$ , dunque  $\psi_h \rightarrow (-ix)^{n+1} e^{-ixy}$  per  $h \rightarrow 0$ . Inoltre per il teorema del valor medio vale

$$|\varphi_h(x)| = |\varphi_h(x) - \varphi_h(0)| \leq |x| \sup\{|\varphi'_h(t)| \mid t \in \mathbb{R}, |t| \leq |x|\} = |x|$$

in cui nell'ultima uguaglianza si è usato che  $|\varphi'_h(t)| = |e^{-ith}| = 1$ . Quindi si ha

$$|\psi_h(x)| = |\varphi_h(x)||x|^n \leq |x|^{n+1}$$

e per ipotesi  $|x|^{n+1} \in L^1(\mathbb{R}, \mu)$ , dunque si può applicare il teorema di convergenza dominata di Lebesgue e concludere che

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\hat{\mu}^{(n)}(y+h) - \hat{\mu}^{(n)}(y)}{h} &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} \psi_h(x) d\mu(x) \\ &= \int_{\mathbb{R}} (-ix)^{n+1} e^{-ixy} d\mu(x) \end{aligned}$$

ossia esiste  $\hat{\mu}^{(n+1)}$  e vale (2.6) anche per  $n + 1$ . □

**Teorema 2.2.7:** *Ogni misura boreliana finita  $\mu$  su  $\mathbb{R}$  è univocamente determinata dalla sua trasformata di Fourier  $\hat{\mu}$ .*

*Dimostrazione.* Si veda ad esempio [8, Theorem 6.35].  $\square$

Il seguente teorema contiene potenti condizioni per la determinazione nel problema dei momenti di Hamburger e di Stieltjes.

**Teorema 2.2.8 (di Carleman):** *Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  una successione di numeri reali.*

(1) *Se  $\mathfrak{s}$  è semidefinita positiva e soddisfa*

$$\sum_{n \in \mathbb{N}_+} s_{2n}^{-\frac{1}{2n}} = +\infty \quad (2.7)$$

*allora  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti determinata.*

(2) *Se  $\mathfrak{s}$  ed  $E\mathfrak{s} = (s_{n+1})_{n \in \mathbb{N}}$  sono semidefinite positive e vale*

$$\sum_{n \in \mathbb{N}_+} s_n^{-\frac{1}{2n}} = +\infty \quad (2.8)$$

*allora  $\mathfrak{s}$  è una successione di  $[0, +\infty)$ -momenti determinata.*

*Dimostrazione.* Per il Teorema 1.4.5 o il Teorema 1.4.6  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti o di  $[0, +\infty)$ -momenti a seconda se stiamo considerando il primo o il secondo caso rispettivamente. Dunque basta dimostrare la determinazione di  $\mathfrak{s}$  in entrambi i casi.

(1) Se esiste un  $n \in \mathbb{N}_+$  t.c.  $s_{2n} = 0$  allora necessariamente una misura boreliana  $\mu$  con  $\mathfrak{s}$  come successione dei momenti sarà concentrata in  $\{x \in \mathbb{R} \mid x^{2n} = 0\} = \{0\}$  ossia necessariamente  $\mu = s_0 \delta_0$ . Dunque supponiamo senza perdita di generalità  $s_{2n} \neq 0 \ \forall n \in \mathbb{N}$ . Siano  $\mu_1, \mu_2$  due misure di Radon su  $\mathbb{R}$  con  $\mathfrak{s}$  come successione dei momenti e poniamo  $f = \hat{\mu}_1 - \hat{\mu}_2$ . Per la Proposizione 2.2.6  $f \in C^\infty(\mathbb{R}, \mathbb{C})$ . Definiamo per ogni  $n \in \mathbb{N}$  il numero  $M_n = \sqrt{s_{2n}}$ . Per l'Osservazione 2.1.2 la successione  $(M_n)_{n \in \mathbb{N}}$  è logaritmicamente convessa e per ipotesi vale

$$\sum_{n \in \mathbb{N}_+} M_n^{-\frac{1}{n}} = \sum_{n \in \mathbb{N}_+} s_{2n}^{-\frac{1}{2n}} = +\infty. \quad (2.9)$$

Se  $f = u + iv$  vale  $u, v \in C^\infty(\mathbb{R})$  ed usando (2.6) si ottiene

$$\begin{aligned} |u^{(n)}(y)|, |v^{(n)}(y)| &\leq |f^{(n)}(y)| \\ &\leq \int_{\mathbb{R}} |x|^n d\mu_1(x) + \int_{\mathbb{R}} |x|^n d\mu_2(x) \\ &\leq 2\sqrt{s_0} M_n \quad \forall y \in \mathbb{R}, \quad \forall n \in \mathbb{N}. \end{aligned}$$

Usando sempre (2.6) si ottiene anche

$$\begin{aligned} f^{(n)}(0) &= \hat{\mu}_1^{(n)}(0) - \hat{\mu}_2^{(n)}(0) \\ &= \int_{\mathbb{R}} (-ix)^n d\mu_1(x) - \int_{\mathbb{R}} (-ix)^n d\mu_2(x) \\ &= (-i)^n s_n - (-i)^n s_n = 0 \end{aligned}$$

che in particolare implica  $u^{(n)}(0) = v^{(n)}(0) = 0$ . Quindi posso applicare il Teorema 2.1.3 di Denjoy-Carleman e dire che  $u(y) = v(y) = 0$  per ogni  $y \in \mathbb{R}$ , da cui si conclude  $f(y) = \hat{\mu}_1(y) - \hat{\mu}_2(y) = 0$  per ogni  $y \in \mathbb{R}$ . Dunque per il Teorema 2.2.7  $\mu_1 = \mu_2$ .

- (2) Siano  $\nu_1, \nu_2$  due misure di Radon su  $\mathbb{R}$  concentrate su  $[0, +\infty)$  con  $\mathfrak{s}$  come successione dei momenti. Definiamo le misure

$$\mu_j = \frac{k_{\#}\nu_j + (-k)_{\#}\nu_j}{2} \quad \text{per } j = 1, 2$$

in cui  $k(x) = \sqrt{x}$  per ogni  $x \in [0, +\infty)$  e  $k_{\#}\nu_j(A) = \nu_j(k^{-1}(A))$ ,  $(-k)_{\#}\nu_j(A) = \nu_j((-k)^{-1}(A))$  per ogni  $A \in \mathcal{B}(\mathbb{R})$  per  $i = 1, 2$ . Allora  $\mu_1, \mu_2$  sono misure di Radon con

$$\int_{\mathbb{R}} x^n d\mu_j(x) = \frac{1}{2} \int_0^{+\infty} ((\sqrt{x})^n + (-\sqrt{x})^n) d\nu_j(x) \quad \forall n \in \mathbb{N}.$$

Questo integrale è nullo per  $n$  dispari ed è uguale a  $s_k$  se  $n$  è pari con  $n = 2k$ . Dunque  $\mu_j$  sono due soluzioni al problema dei momenti di Hamburger con successione dei momenti  $\hat{\mathfrak{s}} = (\hat{s}_n)_{n \in \mathbb{N}}$  t.c.  $\hat{s}_{2n} = s_n$  e  $\hat{s}_{2n+1} = 0$  per ogni  $n \in \mathbb{N}$ . Ma, essendo  $\hat{s}_{2n} = s_n$ , dall'assunzione (2.8) segue che (2.7) vale per  $\hat{\mathfrak{s}}$ , che quindi è una successione di  $\mathbb{R}$ -momenti determinata. Quindi  $\mu_1 = \mu_2$ , da cui segue  $\nu_1 = \nu_2$ . □

**Osservazione 2.2.9:** La condizione (2.7) è famosa in letteratura col nome di **condizione di Carleman**.

**Corollario 2.2.10:** *Sia  $\mathfrak{s}$  una successione di numeri reali.*

- (1) *Se  $\mathfrak{s}$  è semidefinita positiva ed esiste una costante  $M > 0$  t.c.*

$$s_{2n} \leq M^n (2n)! \quad \forall n \in \mathbb{N}_+ \tag{2.10}$$

*allora la condizione di Carleman (2.7) è soddisfatta ed  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti determinata.*

(2) Se  $\mathfrak{s}$  ed  $E\mathfrak{s}$  sono semidefinite positive ed esiste una costante  $M > 0$  t.c.

$$s_n \leq M^n (2n)! \quad \forall n \in \mathbb{N}_+ \quad (2.11)$$

allora la condizione (2.8) è soddisfatta ed  $\mathfrak{s}$  è una successione di  $[0, +\infty)$ -momenti determinata.

*Dimostrazione.* (1) Se  $n \in \mathbb{N}_+$  allora  $(2n)! \leq (2n)^{2n}$ . Da questo si trova  $\frac{1}{2n} \leq [(2n)!]^{-\frac{1}{2n}}$ , per cui vale

$$M^{-\frac{1}{2}} \frac{1}{2n} \leq M^{-\frac{1}{2}} [(2n)!]^{-\frac{1}{2n}} \leq s_{2n}^{-\frac{1}{2n}}$$

da cui segue

$$\sum_{n \in \mathbb{N}_+} s_{2n}^{-\frac{1}{2n}} \geq \frac{M^{-\frac{1}{2}}}{2} \sum_{n \in \mathbb{N}_+} \frac{1}{n} = +\infty$$

ossia la condizione di Carleman (2.7) è verificata. Dunque, per il Teorema di Carleman 2.2.8 (1),  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti determinata.

(2) In modo analogo al punto precedente si ottiene

$$M^{-\frac{1}{2}} \frac{1}{2n} \leq M^{-\frac{1}{2}} [(2n)!]^{-\frac{1}{2n}} \leq s_n^{-\frac{1}{2n}}$$

da cui segue

$$\sum_{n \in \mathbb{N}_+} s_n^{-\frac{1}{2n}} \geq \frac{M^{-\frac{1}{2}}}{2} \sum_{n \in \mathbb{N}_+} \frac{1}{n} = +\infty$$

ossia la condizione (2.8) è verificata. Dunque, per il Teorema 2.2.8 di Carleman (2),  $\mathfrak{s}$  è una successione di  $[0, +\infty)$ -momenti determinata. □

**Corollario 2.2.11:** (1) Sia  $\mu$  una misura boreliana su  $\mathbb{R}$ . Se esiste un numero  $\varepsilon > 0$  t.c.

$$\int_{\mathbb{R}} e^{\varepsilon|x|} d\mu(x) < +\infty \quad (2.12)$$

allora  $\mu$  ammette momenti di ogni ordine ed è univocamente determinata da essi.

(2) Sia  $\mu$  una misura boreliana su  $\mathbb{R}$  concentrata su  $[0, +\infty)$ . Se esiste un numero  $\varepsilon > 0$  t.c.

$$\int_{\mathbb{R}} e^{\varepsilon\sqrt{|x|}} d\mu(x) < +\infty \quad (2.13)$$

allora  $\mu$  ammette momenti di ogni ordine ed è univocamente determinata da essi tra le misure concentrate su  $[0, +\infty)$ .

*Dimostrazione.* (1) Per ogni  $n \in \mathbb{N}$  ed  $x \in \mathbb{R}$  vale la disuguaglianza  $e^{\varepsilon|x|} \leq \frac{(\varepsilon|x|)^{2n}}{(2n)!}$ , dunque

$$x^{2n} e^{-\varepsilon|x|} \leq \varepsilon^{-2n} (2n)!$$

e quindi

$$\begin{aligned} \int_{\mathbb{R}} x^{2n} d\mu(x) &= \int_{\mathbb{R}} x^{2n} e^{-\varepsilon|x|} e^{\varepsilon|x|} d\mu(x) \\ &\leq \varepsilon^{-2n} (2n)! \int_{\mathbb{R}} e^{\varepsilon|x|} d\mu(x) < +\infty \quad \forall n \in \mathbb{N}. \end{aligned} \quad (2.14)$$

Inoltre usando la disuguaglianza di Cauchy-Schwarz si ha

$$\int_{\mathbb{R}} |x|^{2n+1} \leq \left( \int_{\mathbb{R}} x^{4n} d\mu(x) \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}} x^2 d\mu(x) \right)^{\frac{1}{2}} < +\infty \quad \forall n \in \mathbb{N}.$$

Dunque  $\mu$  ammette momenti finiti di ogni ordine. Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  la successione dei momenti di  $\mu$ , allora da (2.14) segue anche  $s_{2n} \leq M^n (2n)!$  con  $M > 0$  una costante. Quindi la condizione (2.10) è soddisfatta e dal Corollario 2.2.10 (1), segue che  $\mathfrak{s}$  è una successione di  $\mathbb{R}$ -momenti determinata e questo equivale a dire che  $\mu$  è determinata univocamente dai suoi momenti.

(2) Per ogni  $n \in \mathbb{N}$  ed  $x \in \mathbb{R}$  vale la disuguaglianza  $e^{\varepsilon\sqrt{|x|}} \leq \frac{\varepsilon^{2n}|x|^n}{(2n)!}$ , dunque

$$|x|^n e^{-\varepsilon\sqrt{|x|}} \leq \varepsilon^{-2n} (2n)!$$

e quindi

$$\begin{aligned} \int_{\mathbb{R}} |x|^n d\mu(x) &= \int_{\mathbb{R}} |x|^n e^{-\varepsilon\sqrt{|x|}} e^{\varepsilon\sqrt{|x|}} d\mu(x) \\ &\leq \varepsilon^{-2n} (2n)! \int_{\mathbb{R}} e^{\varepsilon\sqrt{|x|}} d\mu(x) < +\infty \quad \forall n \in \mathbb{N}. \end{aligned} \quad (2.15)$$

Da questo segue che  $\mu$  ammette momenti finiti di ogni ordine. Sia  $\mathfrak{s} = (s_n)_{n \in \mathbb{N}}$  la successione dei momenti di  $\mu$ , allora da (2.15) segue anche  $s_n \leq M^n (2n)!$  con  $M > 0$  una costante. Quindi la condizione (2.11) è soddisfatta e dal Corollario 2.2.10 (2), segue che  $\mathfrak{s}$  è una successione di  $[0, +\infty)$ -momenti determinata. Questo equivale a dire che  $\mu$  è determinata univocamente dai suoi momenti tra le misure concentrate su  $[0, +\infty)$ . □

**Osservazione 2.2.12:** In teoria della probabilità le condizioni (2.12) e (2.13) sono note rispettivamente come **condizione di Cramer** e **condizione di Hardy**.

## 2.3 Teorema di Petersen per il caso multidimensionale

Data una misura boreliana  $\mu : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, +\infty]$  ed una funzione boreliana  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^m$  è ben definita la misura  $\varphi_{\#}\mu : \mathcal{B}(\mathbb{R}^m) \rightarrow [0, +\infty]$  t.c.

$$\varphi_{\#}\mu(A) = \mu(\varphi^{-1}(A)) \quad \forall A \in \mathcal{B}(\mathbb{R}^m)$$

e vale la formula

$$\int_{\mathbb{R}^m} f(y) d\varphi_{\#}\mu(y) = \int_{\mathbb{R}^d} f(\varphi(x)) d\mu(x).$$

Nel caso particolare in cui la funzione boreliana considerata è la proiezione sulla  $j$ -esima componente  $\pi_j : \mathbb{R}^d \rightarrow \mathbb{R}$ , la misura  $\pi_{j\#}\mu : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty]$  è chiamata **j-esima misura marginale** di  $\mu$ .

**Definizione 2.3.1:** Sia  $V$  uno spazio vettoriale reale e sia  $B \subset V$  un suo sottoinsieme convesso. Un  $x \in B$  è detto **punto estremale** se  $\forall y, z \in B$  e  $\forall \lambda \in (0, 1)$  l'uguaglianza  $x = \lambda y + (1 - \lambda)z$  implica  $x = y = z$ .

**Definizione 2.3.2:** Data  $\mu$  una misura boreliana su  $\mathbb{R}^d$  avente momenti di ogni ordine finiti, chiameremo  $\mathfrak{M}_\mu$  la famiglia composta da tutte le misure boreliane su  $\mathbb{R}^d$  con momenti finiti ed uguali a quelli di  $\mu$ .

**Osservazione 2.3.3:** La famiglia  $\mathfrak{M}_\mu$  è un insieme convesso visto come sottoinsieme dello spazio vettoriale delle misure di Radon con segno finite. In cui una misura di Radon con segno non è altro che una misura con segno la cui decomposizione di Jordan è formata da due misure di Radon.

**Lemma 2.3.4:** Sia  $\mu$  una misura boreliana su  $\mathbb{R}^d$  con momenti di ogni ordine finiti. Una misura  $\nu \in \mathfrak{M}_\mu$  è un punto estremale di  $\mathfrak{M}_\mu$  se e solo se  $\mathbb{C}[x_1, \dots, x_d]$  è denso in  $L^1(\mathbb{R}^d, \nu)$ .

*Dimostrazione.* Supponiamo che  $\nu \in \mathfrak{M}_\mu$  non sia punto estremale. Allora esistono due misure  $\nu_1, \nu_2 \in \mathfrak{M}_\mu$ ,  $\nu_1 \neq \nu_2$ , t.c.  $\nu = \lambda_0 \nu_1 + (1 - \lambda_0) \nu_2$  per un qualche  $\lambda_0 \in (0, 1)$ . In particolare  $\nu_1$  è assolutamente continua rispetto a  $\nu$ , dunque per il teorema di Radon-Nikodym esiste  $g : \mathbb{R}^d \rightarrow [0, +\infty]$  misurabile t.c.  $\nu_1 = g.\nu$ . Dovendo essere  $\nu_1 \neq \nu_2$

necessariamente  $g \neq 1$  su un insieme di misura  $\nu$ -positiva. Inoltre

$$\begin{aligned} \int_{\mathbb{R}^d} p(1-g)d\nu &= \int_{\mathbb{R}^d} pd\nu - \int_{\mathbb{R}^d} pg\nu \\ &= \int_{\mathbb{R}^d} pd\nu - \int_{\mathbb{R}^d} pd\nu_1 = 0 \quad \forall p \in \mathbb{C}[x_1, \dots, x_d]. \end{aligned}$$

ossia  $1-g$  definisce un funzionale lineare continuo non identicamente nullo su  $L^1(\mathbb{R}^d, \nu)$  che si annulla su tutto  $\mathbb{C}[x_1, \dots, x_d]$ . Da questo segue che  $\mathbb{C}[x_1, \dots, x_d]$  non può essere denso in  $L^1(\mathbb{R}^d, \nu)$ , altrimenti il funzionale definito da  $1-g$  sarebbe per continuità identicamente nullo.

Adesso prendiamo sempre  $\nu \in \mathfrak{M}_\mu$  e supponiamo che  $\mathbb{C}[x_1, \dots, x_d]$  non sia denso in  $L^1(\mathbb{R}^d, \nu)$ . Allora esiste un funzionale lineare continuo su  $L^1(\mathbb{R}^d, \nu)$  non identicamente nullo ed unitario che si annulla su  $\mathbb{C}[x_1, \dots, x_d]$ . Inoltre  $\nu$  è finita, dunque, usando il teorema di rappresentazione di Riesz per il duale di  $L^1(\mathbb{R}^d, \nu)$ , è possibile trovare una  $g \in L^\infty(\mathbb{R}^d, \nu)$  con  $\|g\|_{L^\infty(\mathbb{R}^d, \nu)} = 1$  t.c.

$$\int_{\mathbb{R}^d} pgd\nu = 0 \quad \forall p \in \mathbb{C}[x_1, \dots, x_d]. \quad (2.16)$$

Siano  $\nu_\pm = (1 \pm g).\nu$ , essendo  $\|g\|_{L^\infty(\mathbb{R}^d, \nu)} = 1$  entrambe  $\nu_+$  e  $\nu_-$  sono misure positive. Inoltre grazie a (2.16) si ha  $\nu_+, \nu_- \in \mathfrak{M}_\mu$ . Ma  $\nu = \frac{\nu_+ + \nu_-}{2}$ , quindi  $\nu$  non è punto estremale di  $\mathfrak{M}_\mu$ .  $\square$

Nella dimostrazione del teorema principale della sezione sarà fondamentale il seguente famoso risultato di densità dovuto ad M. Riesz.

**Teorema 2.3.5 (di Riesz):** *Sia  $\mu$  una misura boreliana su  $\mathbb{R}$  con momenti di ogni ordine finiti e determinata da essi. Allora  $\mathbb{C}[x]$  è denso in  $L^2(\mathbb{R}, \mu)$ .*

*Dimostrazione.* Si veda ad esempio [2, Corollary 2.3.3].  $\square$

Siamo pronti per affrontare il rinomato teorema per la determinazione nel problema dei momenti multidimensionale di L. C. Petersen.

**Teorema 2.3.6 (di Petersen):** *Sia  $\mu$  una misura boreliana su  $\mathbb{R}^d$  con momenti di ogni ordine finiti. Se tutte le misure marginali di  $\mu$ ,  $\{\pi_{j\#}\mu\}_{j=1}^d$ , sono univocamente determinate dai propri momenti, allora anche  $\mu$  è univocamente determinata dai propri momenti.*



*Dimostrazione.* Sia  $\nu \in \mathfrak{M}_\mu$ . Siano  $\mathbb{1}_1, \dots, \mathbb{1}_d$  funzioni indicatrici di boreliani di  $\mathbb{R}$  e  $p_1, \dots, p_d$  polinomi a coefficienti complessi. Usando la disuguaglianza di Cauchy-Schwarz si ottiene

$$\begin{aligned} & \|\mathbb{1}_1(x_1) \dots \mathbb{1}_d(x_d) - p_1(x_1) \dots p_d(x_d)\|_{L^1(\mathbb{R}^d, \nu)} \\ & \leq \sum_{j=1}^d \|p_1(x_1) \dots p_{j-1}(x_{j-1}) [\mathbb{1}_j(x_j) - p_j(x_j)] p_{j+1}(x_{j+1}) \dots p_d(x_d)\|_{L^1(\mathbb{R}^d, \nu)} \\ & \leq \sum_{j=1}^d \|\mathbb{1}_j(x_j) - p_j(x_j)\|_{L^2(\mathbb{R}^d, \nu)} \|p_1(x_1) \dots p_{j-1}(x_{j-1}) \mathbb{1}_{j+1}(x_{j+1}) \dots \mathbb{1}_d(x_d)\|_{L^2(\mathbb{R}^d, \nu)}. \end{aligned} \quad (2.17)$$

Sia  $j \in \{1, \dots, d\}$ . Chiaramente avendo  $\mu$  e  $\nu$  momenti uguali anche  $\pi_{j\#}\mu$  e  $\pi_{j\#}\nu$  hanno momenti uguali e di conseguenza  $\pi_{j\#}\mu = \pi_{j\#}\nu$  per ipotesi. Dunque possiamo scrivere

$$\begin{aligned} \|\mathbb{1}_j(x_j) - p_j(x_j)\|_{L^2(\mathbb{R}^d, \nu)} &= \|\mathbb{1}_j(x_j) - p_j(x_j)\|_{L^2(\mathbb{R}, \pi_{j\#}\nu)} \\ &= \|\mathbb{1}_j(x_j) - p_j(x_j)\|_{L^2(\mathbb{R}, \pi_{j\#}\mu)}. \end{aligned} \quad (2.18)$$

Fissiamo  $\varepsilon > 0$ . Essendo  $\pi_{j\#}\mu$  determinata dai propri momenti, per il Teorema 2.3.5,  $\mathbb{C}[x_j]$  è denso in  $L^2(\mathbb{R}, \pi_{j\#}\mu)$  e grazie all'eq. (2.18) è possibile scegliere  $p_1 \in \mathbb{C}[x_1]$  t.c.

$$\|\mathbb{1}_1(x_1) - p_1(x_1)\|_{L^2(\mathbb{R}^d, \nu)} \leq \frac{\varepsilon}{d \|\mathbb{1}_2(x_2) \dots \mathbb{1}_d(x_d)\|_{L^2(\mathbb{R}^d, \nu)}}$$

poi  $p_2 \in \mathbb{C}[x_2]$  t.c.

$$\|\mathbb{1}_2(x_2) - p_2(x_2)\|_{L^2(\mathbb{R}^d, \nu)} \leq \frac{\varepsilon}{d \|p_1(x_1) \mathbb{1}_3(x_3) \dots \mathbb{1}_d(x_d)\|_{L^2(\mathbb{R}^d, \nu)}}$$

e così via ricorsivamente per ogni  $j \in \{1, \dots, d\}$ , fino quindi ad arrivare a scegliere  $p_d \in \mathbb{C}[x_d]$  t.c.

$$\|\mathbb{1}_d(x_d) - p_d(x_d)\|_{L^2(\mathbb{R}^d, \nu)} \leq \frac{\varepsilon}{d \|p_1(x_1) \dots p_{d-1}(x_{d-1})\|_{L^2(\mathbb{R}^d, \nu)}}$$

Dunque tenendo a mente l'eq. (2.17) si ottiene

$$\|\mathbb{1}_1(x_1) \dots \mathbb{1}_d(x_d) - p_1(x_1) \dots p_d(x_d)\|_{L^1(\mathbb{R}^d, \nu)} \leq \varepsilon.$$

ed essendo lo span delle funzioni del tipo  $\mathbb{1}_1(x_1) \dots \mathbb{1}_d(x_d)$  denso in  $L^1(\mathbb{R}^d, \nu)$  si ottiene che anche i polinomi  $\mathbb{C}[x_1, \dots, x_d]$  sono densi in  $L^1(\mathbb{R}^d, \nu)$ . Quindi, per il Lemma 2.3.4,  $\nu$  è un punto estremale di  $\mathfrak{M}_\mu$ . L'arbitrarietà di  $\nu$  ci suggerisce che ogni punto di  $\mathfrak{M}_\mu$  è estremale, in particolare lo è  $\frac{\mu+\nu}{2} \in \mathfrak{M}_\mu$ , quindi  $\nu = \mu$ .  $\square$



## Un utilizzo statistico dell'unicità

Nel capitolo precedente abbiamo dimostrato che sotto determinate condizioni una misura finita su  $\mathbb{R}^d$  è completamente determinata dai suoi momenti. In questo capitolo ci concentreremo nel caso in cui la misura considerata è una probabilità ed in particolare studieremo la possibilità di utilizzare questa caratterizzazione per scopi statistici, ossia studieremo quanto efficacemente possiamo ricostruire una distribuzione di probabilità basandoci sui suoi momenti. Per farlo ci baseremo su un particolare problema che andremo a descrivere nella prima parte del capitolo insieme ai due metodi utilizzati per cercare di risolverlo, entrambi basati sulla caratterizzazione attraverso i momenti descritta in precedenza. Infine nell'ultima parte commenteremo alcuni risultati ottenuti con essi.

### 3.1 Presentazione del problema e metodi utilizzati

Siano date  $X, Y, Q$  tre variabili aleatorie a valori in  $\mathbb{R}^d$ . Prendiamo la combinazione tra  $X$  ed  $Y$  data da

$$X - AY$$

con  $A$  una matrice  $d \times d$ . Il nostro scopo è quello di determinare numericamente una matrice  $A$  simmetrica e definita positiva che rende la legge di  $X - AY$  il più simile possibile alla legge della terza variabile aleatoria  $Q$ .

Il problema preliminare è quello di dare un significato al concetto di somiglianza tra due leggi. Ed è proprio questo concetto che differenzia i due metodi utilizzati.

Notiamo preliminarmente il seguente fatto.

**Proposizione 3.1.1:** *Siano  $E \subset \mathbb{R}^d$  ed  $F \subset \mathbb{R}^n$  boreliani ed  $f : E \rightarrow F$  un omeomorfismo. Due variabili aleatorie  $X$  e  $Y$  a valori in  $E$  hanno la stessa legge se e solo se  $f(X)$  e  $f(Y)$  hanno la stessa legge su  $F$ .*

*Dimostrazione.* Osserviamo che essendo anche l'inversa  $f^{-1} : F \rightarrow E$  continua basta dimostrare che se  $X$  ed  $Y$  hanno la stessa legge allora anche  $f(X)$  ed  $f(Y)$  hanno la stessa legge. Sia  $B \subset F$  boreliano, allora effettivamente vale

$$\mathbb{P}_{f(X)}(B) = \mathbb{P}_X(f^{-1}(B)) = \mathbb{P}_Y(f^{-1}(B)) = \mathbb{P}_{f(Y)}(B). \quad \square$$

### 3.1.1 Primo metodo

**Teorema 3.1.2:** *Sia  $Z$  una variabile aleatoria a valori in  $\mathbb{R}^d$  limitata. Allora  $Z$  ha tutti i momenti finiti e la sua legge è caratterizzata da essi.*

*Dimostrazione.* Sia  $M > 0$  t.c. l'immagine di  $Z$  è contenuta in  $B(0, M)$ . Sia  $Z = (Z_i)_{i=1, \dots, d}$ . Prendiamo  $j \in \{1, \dots, d\}$  e notiamo che la  $j$ -esima misura marginale di  $\mathbb{P}_Z$  è esattamente la legge di  $Z_j$  su  $\mathbb{R}$ , infatti se  $B \in \mathcal{B}(\mathbb{R})$  abbiamo

$$\pi_{j\#} \mathbb{P}_Z(B) = \mathbb{P}_Z(\pi_j^{-1}(B)) = \mathbb{P}(Z \in \pi_j^{-1}(B)) = \mathbb{P}_{Z_j}(B).$$

Notiamo che per ogni  $j \in \{1, \dots, n\}$  la misura  $\mathbb{P}_{Z_j}$  è univocamente determinata dai suoi momenti. Infatti  $\mathbb{P}_{Z_j}$  è concentrata sul compatto  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ , quindi si utilizza il Teorema 2.2.4 per ottenere la determinazione voluta. Per poter applicare il Teorema 2.3.6 di Petersen e concludere la dimostrazione, manca soltanto da dimostrare che  $\mathbb{P}_Z$  ha momenti di ogni ordine finiti. Sia  $\alpha \in \mathbb{N}^d$  ed  $M > 0$  t.c.  $B(0, M) \supset [-\frac{\pi}{2}, \frac{\pi}{2}]^d$ , vale

$$\int_{\mathbb{R}^d} |z^\alpha| d\mathbb{P}_Z(z) = \int_{B(0, M)} |z^\alpha| d\mathbb{P}_Z(z) \leq M^{|\alpha|} < +\infty$$

dunque effettivamente  $\mathbb{P}_Z$  ammette momenti di ogni ordine finiti. Come già anticipato il Teorema 2.3.6 di Petersen conclude la dimostrazione.  $\square$

Grazie a questo risultato e alla Proposizione 3.1.1 possiamo costruire un concetto di somiglianza tra le leggi di due variabili aleatorie basato sui momenti. Siano date due variabili aleatorie  $Z$  e  $Q$  e consideriamo l'omeomorfismo  $f : \mathbb{R}^d \rightarrow (-\frac{\pi}{2}, \frac{\pi}{2})^d$  che applica l'arcotangente ad ogni componente. Allora  $Z$  e  $Q$  saranno equidistribuite se e solo se lo sono  $f(Z)$  ed  $f(Q)$  e queste lo sono se e solo se hanno gli stessi momenti. Questo ci porta a dire che due variabili aleatorie hanno legge simile se i momenti delle loro immagini mediante  $f$  sono a due a due vicini.

Torniamo quindi al problema presentato inizialmente. Il metodo numerico che abbiamo utilizzato per determinare la matrice  $A$  definita positiva si basa proprio questa idea di somiglianza tra le leggi considerate. Per ovvi motivi computazionali si confrontano solo

un numero finito di momenti e si minimizza la loro distanza attraverso la minimizzazione del funzionale

$$r_1^N(A) = \sum_{k=0}^N \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha|=k}} (\mathbb{E}[(X - AY)^\alpha] - \mathbb{E}[Q^\alpha])^2 \quad (3.1)$$

con un metodo iterativo di ottimizzazione sulla varietà delle matrici simmetriche definite positive. Ovviamente i valori attesi che appaiono in (3.1) si sono approssimati campionariamente usando il dataset a disposizione.

Per attuare l'esecuzione numerica si è utilizzato un pacchetto matlab di ottimizzazione su varietà chiamato `manopt`, il metodo iterativo utilizzato è del tipo trust-region su varietà riemanniane, il suo funzionamento esula dagli scopi di questo elaborato, di conseguenza si rimanda a [5] per eventuali approfondimenti al riguardo.

### 3.1.2 Secondo metodo

**Definizione 3.1.3:** Sia  $X$  una variabile aleatoria a valori in  $\mathbb{R}^d$ . Consideriamo  $\mathcal{D}_X = \{t \in \mathbb{R}^d \mid \mathbb{E}[e^{t \cdot X}] < +\infty\}$ . Definiamo la sua **funzione generatrice dei momenti** come la funzione  $\psi_X : \mathcal{D}_X \rightarrow [0, +\infty)$  t.c.

$$\psi_X(t) = \mathbb{E}[e^{t \cdot X}] \quad \forall t \in \mathcal{D}_X.$$

**Osservazione 3.1.4:** Può avvenire che  $\mathcal{D}_X = \{0\}$ .

**Teorema 3.1.5:** Sia  $Z$  una variabile aleatoria a valori in  $\mathbb{R}^d$  t.c.  $\mathbb{E}[\|Z\|^k] < \frac{Mk!}{(dr)^k}$  in cui  $M, r > 0$  e la norma  $\|\cdot\|$  è quella della massima componente. Allora la variabile aleatoria  $Z$  ha tutti i momenti finiti, ammette funzione generatrice dei momenti definita almeno su  $(-r, r)^d$  e la sua legge è caratterizzata da essa.

*Dimostrazione.* Vale lo sviluppo in serie di Taylor, nella variabile  $t \in \mathbb{R}^d$

$$e^{(t \cdot Z)} = \sum_{k \in \mathbb{N}} \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha|=k}} \frac{(Z^\alpha)}{\alpha!} t^\alpha$$

inoltre per  $t \in (-r, r)^d$  vale

$$\begin{aligned} \sum_{k \in \mathbb{N}} \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha|=k}} \frac{\mathbb{E}[|Z^\alpha|]}{\alpha!} |t^\alpha| &\leq \sum_{k \in \mathbb{N}} \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha|=k}} \frac{\mathbb{E}[||Z||^k]}{\alpha!} |t^\alpha| \\ &= \sum_{k \in \mathbb{N}} \mathbb{E}[||Z||^k] \frac{(|t_1| + \dots + |t_d|)^k}{k!} \\ &\leq \sum_{k \in \mathbb{N}} \mathbb{E}[||Z||^k] \frac{(||t||d)^k}{k!} \leq M \sum_{k \in \mathbb{N}} \left( \frac{||t||}{r} \right)^k < +\infty \end{aligned}$$

dunque posso scambiare serie ed integrali ed ottenere

$$\psi_Z(t) = \sum_{k \in \mathbb{N}} \sum_{\substack{\alpha \in \mathbb{N}^d \\ |\alpha|=k}} \frac{\mathbb{E}[Z^\alpha]}{\alpha!} t^\alpha \quad \forall t \in (-r, r)^d.$$

Quindi  $\psi_Z$  è analitica almeno su  $(-r, r)^d$  ed univocamente determinata dai momenti di  $Z$  che a loro volta determinano la legge di  $Z$  per il Teorema 3.1.2, dunque  $\psi_Z$  determina univocamente  $\mathbb{P}_Z$ .  $\square$

Similmente a quanto fatto prima, grazie a questo risultato e alla Proposizione 3.1.1 è possibile costruire un concetto di somiglianza tra leggi di due variabili aleatorie basato stavolta sulla funzione generatrice dei momenti. Siano date due variabili aleatorie  $Z$  e  $Q$  e consideriamo l'omeomorfismo  $g : \mathbb{R}^d \rightarrow (-1, 1)^d$  che applica la funzione  $\frac{2}{\pi} \arctan$  ad ogni componente. Allora  $Z$  e  $Q$  sono equidistribuite se e solo se lo sono  $g(Z)$  ed  $g(Q)$  e queste lo sono se e solo se hanno la stessa funzione generatrice dei momenti in un intorno dell'origine, infatti essendo

$$\mathbb{E}[||g(Z)||^k] \leq 1 \leq \frac{k!}{(dr_d)^k} \quad \forall k \in \mathbb{N}$$

con  $r_d = \frac{1}{d}$ , il Teorema 3.1.5 si applica.

Tornando al problema iniziale, il metodo numerico utilizzato si basa sul confronto del valore della funzione generatrice dei momenti su vari punti di un intorno dell'origine. Si minimizza quindi la loro distanza tramite la minimizzazione sulla varietà delle matrici  $A$  definite positive del funzionale

$$r_2^N(A) = \sum_{k=1}^N \left( \mathbb{E}[e^{(t_k \cdot (X - AY))}] - \mathbb{E}[e^{(t_k \cdot Q)}] \right)^2$$

in cui  $\{t_k\}_{k=1, \dots, N}$  sono punti scelti nell'intorno dell'origine con distribuzione di probabilità uniforme. Nel nostro caso l'intorno sarà un ipercubo con centro l'origine di lato

$\frac{2}{d+1}$ . Anche in questo caso l'esecuzione numerica è stata fatta con lo stesso metodo del pacchetto matlab `manopt` utilizzato per il primo metodo.

## 3.2 Risultati ottenuti

Analizziamo a questo punto i risultati ottenuti attraverso i vari metodi considerati. Come già anticipato tutte le sperimentazioni sono state eseguite in matlab. La metodologia usata è stata la seguente: si sono scelte due distribuzioni di probabilità numericamente simulabili per le componenti delle variabili aleatorie  $X$  ed  $Y$  e si è scelta una matrice  $A_0$  simmetrica e definita positiva. Dunque si sono simulati 1425 campioni costruendo il dataset per  $X$  ed  $Y$ . Si è fatto la stessa cosa per altre due variabili aleatorie  $X_1$  e  $Y_1$  che sono state usate per comporre la terza variabile aleatoria  $Q = X_1 - A_0 Y_1$ . A questo punto si è costruita l'approssimazione campionaria della funzione  $r_i^N(A)$  ( $i = 1, 2$  a seconda di quale metodo si stava implementando) e si è minimizzata al variare di  $A$  nella varietà delle matrici simmetriche definite positive utilizzando il pacchetto `manopt`. Il metodo iterativo si interrompeva quando la norma del gradiente della funzione costo scendeva sotto la soglia  $10^{-8}$ . Lo scopo delle prove era osservare se effettivamente la matrice calcolata alla fine del metodo fosse stata sufficientemente simile ad  $A_0$ . Per ogni prova si sono fatti cinque esperimenti e si è fatta una media dei risultati ottenuti in essi.

Elenchiamo di seguito alcune prove significative tra quelle fatte. Iniziamo analizzando il primo metodo.

**Prova 1:** Per la prima sperimentazione si è presa  $X$  con componenti di legge gaussiana con media 2 e varianza 3 ed  $Y$  con componenti di legge log-normale standard. La matrice  $A_0$  usata è stata

$$A_0^{(1)} = \begin{bmatrix} 6 & -4 \\ -4 & 6 \end{bmatrix}$$

e confrontando i momenti fino all'ordine  $N = 4$ , in un tempo medio pari a  $t^{(1)} = 1.9633s$ , il primo metodo ha dato come risultato medio la matrice

$$A_c^{(1,1)} = \begin{bmatrix} 6.022 & -4.045 \\ -4.045 & 6.220 \end{bmatrix}$$

che è una buona approssimazione della matrice  $A_0$  considerata. Osserviamo che il risultato è stato buono nonostante si sia usata la distribuzione log-normale che, come è noto, non è caratterizzata dai momenti.

**Prova 2:** La distribuzione log-normale però, nonostante non sia caratterizzata da essi, ammette momenti di ogni ordine, dunque come seconda prova si è scelto di sostituire la legge log-normale con una di Cauchy che invece non ammette momento di nessun ordine, in modo da mettere il primo metodo ancor di più alla prova. In questo caso la matrice  $A_0$  utilizzata è stata

$$A_0^{(2)} = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

e confrontando i momenti fino all'ordine  $N = 4$ , in un tempo medio pari a  $t^{(2)} = 2.3669s$ , il primo metodo ha dato come risultato medio la matrice

$$A_c^{(1,2)} = \begin{bmatrix} 3.245 & -1.931 \\ -1.931 & 6.777 \end{bmatrix}$$

che continua ad essere una buona approssimazione anche in questo caso patologico, contrariamente a quanto accade se non si applica l'arcotangente ai dati, infatti senza tale accorgimento, in un tempo medio pari a  $t = 151.641643s$ , il metodo ha dato come risultato medio la matrice

$$\begin{bmatrix} 189.739 & -63.095 \\ -63.095 & 49.623 \end{bmatrix}$$

che non approssima nemmeno lontanamente la matrice obiettivo  $A_0^{(2)}$  utilizzata. Dunque effettivamente la trasformazione con la funzione arcotangente utilizzata funziona e dà un notevole vantaggio, in quanto ci garantisce che il metodo funzioni anche quando le variabili aleatorie considerate non ammettono tutti i momenti.

**Prova 3:** La prossima sperimentazione ha avuto lo scopo di testare l'efficienza del metodo all'aumentare della dimensione delle variabili aleatorie considerate. Si è aumentata la dimensione da 2 a 4 e si è presa  $X$  con componenti gaussiane standard ed  $Y$  con componenti con legge di Cauchy. La matrice  $A_0$  utilizzata è stata

$$A_0^{(3)} = \begin{bmatrix} 10 & -2 & -5 & 8 \\ -2 & 14 & 5 & -8 \\ -5 & 5 & 10 & -8 \\ 8 & -8 & -8 & 20 \end{bmatrix}$$



s e confrontando i momenti fino all'ordine  $N = 4$ , in un tempo medio pari a  $t^{(3)} = 282.0367$ , il metodo in questione ha dato come risultato medio

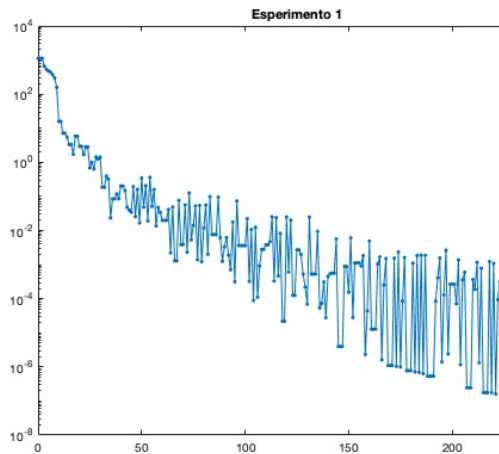
$$A_c^{(1,3,4)} = \begin{bmatrix} 9.291 & -1.642 & -5.447 & 8.519 \\ -1.642 & 13.685 & 4.953 & -8.160 \\ -5.447 & 4.953 & 10.282 & -8.057 \\ 8.519 & -8.160 & -8.057 & 16.607 \end{bmatrix}.$$

Notiamo che stavolta il tempo necessario per l'ottimizzazione è aumentato notevolmente anche solo avendo incrementato di poco la dimensione del problema. Questo era però prevedibile visto che i momenti di ordine  $n$  in dimensione  $d$  hanno un andamento nell'ordine di  $\mathcal{O}(d^n)$ , dunque la complessità computazionale del metodo cresce velocemente all'aumentare di  $d$  se si decide di confrontare momenti di ordine alto. Comunque la matrice  $A_c^{(3)}$  può essere ritenuta una buona approssimazione de  $A_0^{(3)}$ , dunque nonostante il costo computazionale maggiore possiamo constatare che il metodo funziona bene anche in dimensione più alta. Nella figura 3.1 è possibile osservare l'andamento della norma del gradiente nel corso dell'esecuzione del metodo nei cinque esperimenti di questa prova. Da quanto appena osservato è naturale domandarsi quanto l'approssimazione peggiora se l'ordine massimo dei momenti confrontati viene diminuito, in modo da rendere il metodo più leggere computazionalmente. In quest'ottica abbiamo quindi diminuito l'ordine massimo  $N$  a 2 e abbiamo ripetuto la prova ottenendo, in un tempo medio pari a  $t = 48.483153s$ , il risultato medio

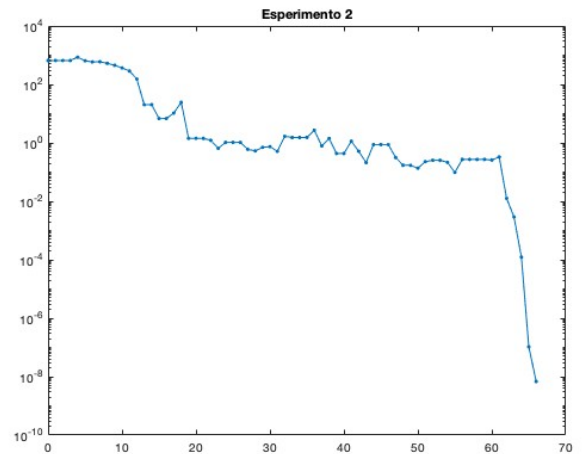
$$A_c^{(1,3,2)} = \begin{bmatrix} 10.197 & -1.233 & -4.302 & 7.613 \\ -1.233 & 11.720 & 3.722 & -6.890 \\ -4.302 & 3.722 & 7.285 & -5.485 \\ 7.613 & -6.890 & -5.485 & 12.360 \end{bmatrix}.$$

Come previsto il tempo di calcolo è diminuito sensibilmente, ma l'approssimazione ottenuta è molto meno precisa di  $A_c^{(1,3,4)}$ , calcolata confrontando i momenti fino al quarto ordine.

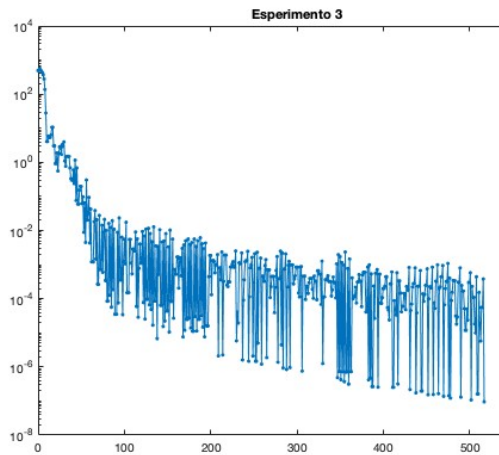
**Prova 4:** Tutte le prove fatte fino ad ora sottintendono che le componenti delle variabili aleatorie considerate siano indipendenti tra di loro ed è spontaneo chiedersi come il metodo si comporterebbe nel caso in cui invece non lo siano. L'obbiettivo della prossima prova è proprio quello di rispondere a questa domanda. Il procedimento utilizzato per questa sperimentazione è leggermente diverso da quello usato fino ad adesso ovviamente, in quanto bisogna generare il dataset diversamente. Innanzitutto ci siamo riportati per



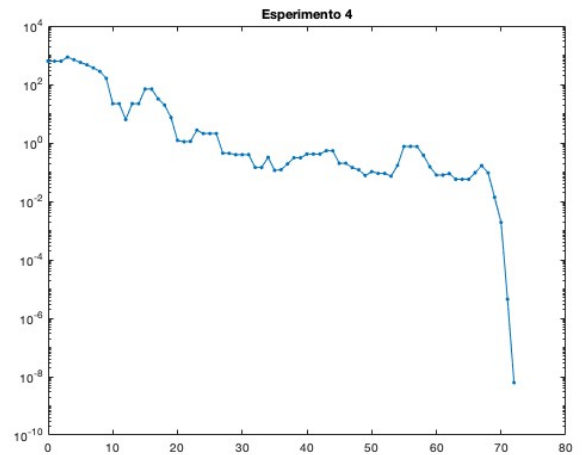
(a) Esperimento 1



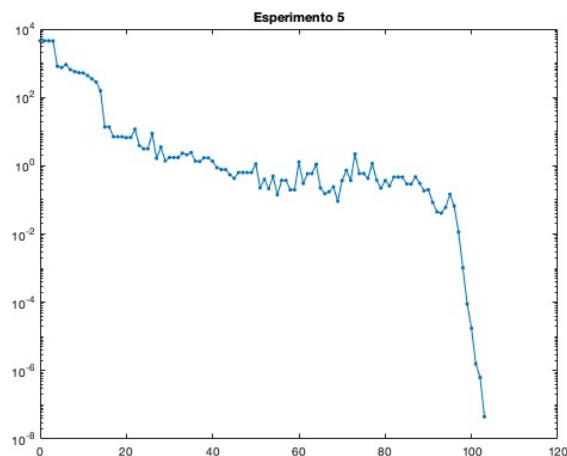
(b) Esperimento 2



(c) Esperimento 3



(d) Esperimento 4



(e) Esperimento 5

Figura 3.1: Andamento della norma del gradiente della funzione costo nella Prova 3.

comodità in dimensione 2 e per simulare un dataset con componenti dipendenti si sono generati i campioni della prima componente di  $X$  e di  $Y$  e poi si è posta come seconda componente la prima a cui si è applicata una trasformazione affine. Nel nostro caso particolare per costruire i campioni di  $X$  si è scelta la legge gaussiana standard per la sua prima componente e per la seconda abbiamo applicato la trasformazione  $x \mapsto 3x + 2$ , mentre per costruire  $Y$  abbiamo scelto la legge di Cauchy per la prima componente ed abbiamo applicato la trasformazione  $x \mapsto 2x + 1$  per la seconda. La matrice  $A_0$  utilizzata per questa prova è stata

$$A_0^{(4)} = \begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix}$$

e confrontando i momenti fino all'ordine  $N = 4$ , in un tempo medio pari a  $t^{(4)} = 4.9219s$ , il metodo ha dato come risultato medio

$$A_c^{(1,4)} = \begin{bmatrix} 3.957 & -1.989 \\ -1.989 & 5.837 \end{bmatrix}$$

che è una buona approssimazione della matrice obiettivo  $A_0^{(4)}$ . Dunque il metodo pare comportarsi bene anche in casi come questo in cui non c'è indipendenza tra le componenti dei campioni.

**Prova 5:** Un ulteriore aspetto che merita di essere indagato è quello del comportamento del metodo nel caso in cui la matrice obiettivo  $A_0$  non sia definita positiva. Ci chiediamo in particolare se il metodo converge comunque ad una matrice sul bordo topologico della varietà su cui stiamo lavorando. Si è quindi scelta come matrice obiettivo la matrice semidefinita positiva:

$$A_0^{(6)} = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}$$

e confrontando i momenti fino all'ordine  $N = 4$ , in un tempo medio pari a  $t = 10.1447s$  si è ottenuto come risultato medio la matrice

$$A_c^{(1,6)} = \begin{bmatrix} 3.958 & -1.943 \\ -1.943 & 1.102 \end{bmatrix}.$$

Dal risultato possiamo notare come il metodo tenda ad avvicinarsi alla matrice obiettivo  $A_0^{(5)}$  e quindi effettivamente il metodo può tendere ad una matrice sul bordo topologico della varietà delle matrici simmetriche definite positive.

Passiamo adesso al secondo metodo, in cui si confrontano i valori delle funzioni generatrici dei momenti. Si sono ripetute, usando stavolta il secondo metodo, alcune delle precedenti prove con i medesimi dati (matrice  $A_0$  e dataset), in modo da fare un confronto tra i due metodi.

**Prova 6:** Siamo partiti dalla seconda prova che col secondo metodo, confrontando la funzione generatrice in  $N = 50$  punti casuali in  $[-\frac{1}{3}, \frac{1}{3}]^2$ , in un tempo medio pari a  $t = 27.1548s$ , ha dato come risultato medio

$$A_c^{(2,2)} = \begin{bmatrix} 3.423 & -1.916 \\ -1.916 & 6.587 \end{bmatrix}$$

che può essere considerata una buona approssimazione di  $A_0^{(2)}$ . Questo secondo metodo è però risultato più lento rispetto al primo e nonostante questo non ha dato un'approssimazione migliore della matrice obiettivo. Nella figura 3.2 è possibile osservare l'andamento della norma del gradiente della funzione costo nei cinque esperimenti di questa prova.

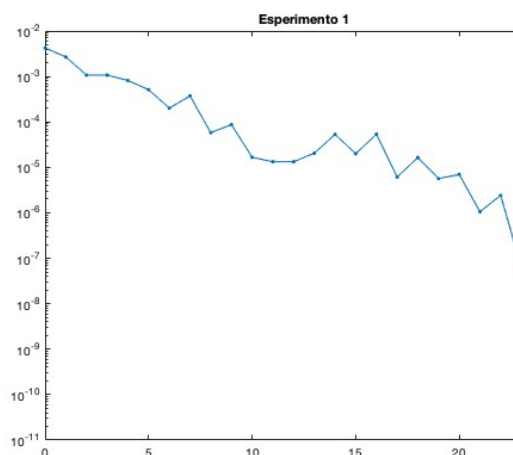
**Prova 7:** Passiamo adesso alla terza prova in dimensione  $d = 4$ . In un tempo medio pari a  $t = 208.0176s$ , confrontando  $N = 50$  punti casuali in  $[-\frac{1}{5}, \frac{1}{5}]^4$ , abbiamo ottenuto come risultato medio

$$A_c^{(2,3,50)} = \begin{bmatrix} 16.203 & -3.908 & -6.619 & 6.591 \\ -3.908 & 16.288 & 4.625 & -5.695 \\ -6.619 & 4.625 & 6.273 & -5.229 \\ 6.591 & -5.695 & -5.229 & 10.798 \end{bmatrix}$$

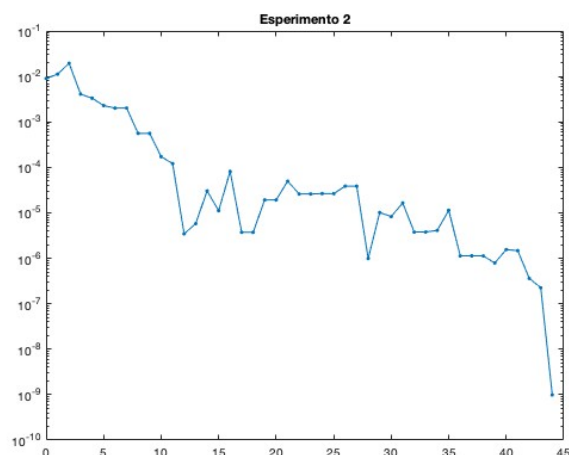
che stavolta non può essere ritenuta una buona approssimazione componente per componente della matrice  $A_0^{(3)}$  in confronto al risultato del primo metodo  $A_c^{(1,3)}$ . Comunque abbiamo deciso di proseguire la sperimentazione aumentando i punti confrontati in  $[-\frac{1}{5}, \frac{1}{5}]^4$  ad  $N = 100$  ed in un tempo medio pari a  $t = 286.4681s$ , si è ottenuto come risultato medio

$$A_c^{(2,3,100)} = \begin{bmatrix} 15.762 & -5.793 & -7.592 & 11.903 \\ -5.793 & 12.118 & 5.437 & -6.328 \\ -7.592 & 5.437 & 12.828 & -11.444 \\ 11.903 & -6.328 & -11.444 & 18.486 \end{bmatrix}$$

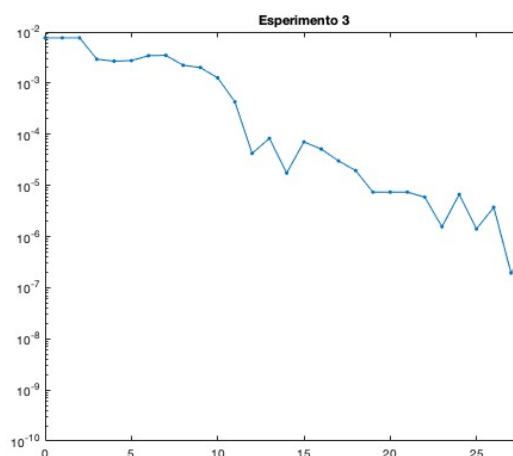
che ancora non può essere ritenuta una buona approssimazione componente per componente della matrice obiettivo in confronto alla matrice  $A_c^{(1,3)}$  ottenuta attraverso il primo metodo. Come conseguenza di quanto osservato possiamo dire che il secondo metodo diminuisce nettamente la sua efficienza all'aumentare della dimensione del problema.



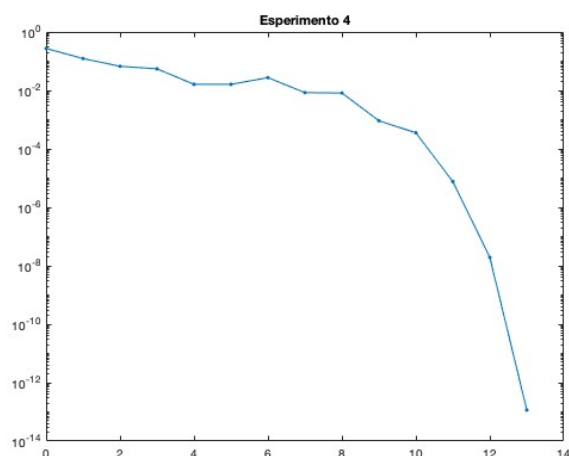
(a) Esperimento 1



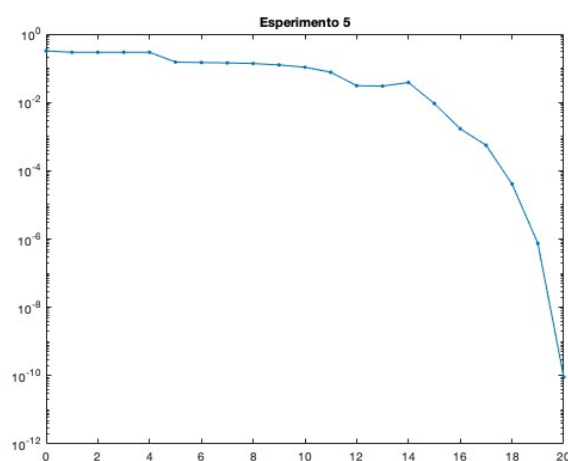
(b) Esperimento 2



(c) Esperimento 3



(d) Esperimento 4



(e) Esperimento 5

Figura 3.2: Andamento della norma del gradiente della funzione costo nella Prova 6.

La poca precisione del metodo può essere dovuta alla discretizzazione del dominio della funzione generatrice. Ossia il metodo, per funzionare in modo adeguato, richiederebbe il confronto di un numero elevato di punti, ma questo comporterebbe un aumento sostanziale della complessità numerica dell'algoritmo e questo fatto non lo rende adeguato all'utilizzo nel caso si debba lavorare in dimensioni elevate.

**Prova 8:** Nella Prova 6 si è notato che in dimensione bassa questo secondo metodo risulta essere utilizzabile, quindi continuiamo a studiare il suo comportamento in dimensione 2. In modo analogo a quanto fatto col primo metodo andiamo ad osservare il comportamento del secondo nel caso in cui le componenti non siano indipendenti tra loro. Utilizziamo gli stessi dataset, la stessa matrice obiettivo e le stesse trasformazioni della Prova 4 fatta col primo metodo. Confrontando la funzione generatrice in  $N = 50$  punti casuali in  $[-\frac{1}{3}, \frac{1}{3}]^2$ , in un tempo medio pari a  $t = 21.1022s$  il secondo metodo ha dato come risultato medio

$$A_c^{(2,4)} = \begin{bmatrix} 3.998 & -2.002 \\ -2.002 & 6.975 \end{bmatrix}$$

che è una buona approssimazione della matrice obiettivo  $A_0^{(4)}$  anche se meno precisa di  $A_c^{(1,4)}$  ottenuta col primo. Quindi anche in questo caso il secondo metodo sembra comportarsi bene, nonostante sia più lento e meno preciso rispetto al primo.

In conclusione possiamo affermare che il primo metodo è preferibile per affrontare al meglio il problema, soprattutto quando siamo costretti a lavorare in dimensioni elevate.

### 3.3 Codici utilizzati

In questa sezione conclusiva riportiamo i codici utilizzati per fare le prove descritte precedentemente.

I campioni utilizzati sono stati tutti creati utilizzando uno script matlab di simulazione di distribuzioni di probabilità chiamato `randraw` (che si può trovare al seguente indirizzo <https://it.mathworks.com/matlabcentral/fileexchange/7309-randraw?tab=discussions>). Nel seguito, per dare un esempio, si sono impostate tutte le distribuzioni come gaussiane standard e si è impostata la dimensione del problema a due.

Lo script utilizzato per il primo metodo è il seguente:

```
n = 1425; % numero di campioni
```

```

d = 2;                                % dimensione delle v.a
.

X = randraw('norm', [0,1], [d n]);
Y = randraw('norm', [0,1], [d n]);
dlX = dlarray(X);
dlY = dlarray(Y);

%-----scelta di A0-----
%A0=...;
%-----

X1 = randraw('normal', [], [d n]);
Y1 = randraw('normal', [], [d n]);
Q = X1-A0*Y1;
dlQ = dlarray(Q);
Qr = atan(dlQ);

%[mQ1,mQ2,mQ3,mQ4,mQ5] = moments(Qr,d); % momenti fino ord.
5
[mQ1,mQ2,mQ3,mQ4] = moments(Qr,d);    % momenti fino ord.
4
%[mQ1,mQ2,mQ3] = moments(Qr,d);      % momenti fino ord.
3
%-----

% immetto la varieta' sulla quale minimizzare
manifold = sympositivedefinitefactory(d);
problem.M = manifold;
% immetto la funzione costo
problem.cost = @(A) cost_atan(A,dlX,dlY,mQ1,mQ2,mQ3,mQ4,d);
% immetto il gradiente ed hessiana della funzione costo
% (differenziazione automatica)
problem = manoptAD(problem);

```

```

% immetto condizione iniziale
in = eye(d);

% numero max di iterazioni
options.maxiter = 700;
% soglia sulla norma del gradiente
options.tolgradnorm = 1e-8;
% tempo a disposizione in sec
options.maxtime = 1000;
% passo minimo che il metodo puo' compiere
options.minstepsize = 1e-10;
options.useRand = 1;

% applico l' algoritmo di minimizzazione
[Ac, costs, info, options] = trustregions(problem,in,options)
;

% stampo la matrice calcolata
Ac

in cui cost_atan è la funzione:

function c = cost_atan(A, X, Y, mQ1, mQ2, mQ3, mQ4, d)

    % se si sono scelti in 'moments.m' momenti di ordine
    % maggiore a 4 aggiungere in input
    % la variabile mQ5 (nel caso
    % si sia scelto l'ordine 5) tra mQ4 e d.

    R = X-A*Y
    Rr = atan(R);

    % costruzione dei momenti di  $Rr = \arctan(X - AY)$ .
    % A seconda dell'ordine massimo dei momenti considerato
    % in
    % 'moments.m' scegliere le linee di codice indicate (

```



```

    default: ord. 4)

% per momenti fino all'ordine 5
%[mR1,mR2,mR3,mR4,mR5] = moments(Rr,d);
% per momenti fino all'ordine 4
[mR1,mR2,mR3,mR4] = moments(Rr,d);
% per momenti fino all'ordine 3
%[mR1,mR2,mR3] = moments(Rr,d);

% calcolo della funzione costo (di default costruita
% per momenti fino all'ordine 4, se si sono scelti in
% 'moments.m' anche ordini superiori aggiungere gli
% addendi corrispondenti).

c = l2loss(mR1,mQ1,DataFormat='U',Reduction='sum',
    NormalizationFactor='none')+...
    l2loss(mR2,mQ2,DataFormat='U',Reduction='sum',
    NormalizationFactor='none')+...
    l2loss(mR3,mQ3,DataFormat='U',Reduction='sum',
    NormalizationFactor='none')+...
    l2loss(mR4,mQ4,DataFormat='U',Reduction='sum',
    NormalizationFactor='none');%+...
%l2loss(mR5,mQ5,DataFormat='U',Reduction='sum',
    NormalizationFactor='none');
% per momenti fino all'ordine 5

```

che calcola la funzione costo da minimizzare  $r_1^N$  descritta nella sezione precedente e che utilizza a sua volta le funzioni:

```

function [m1,m2,m3,m4] = moments(X,d)

% X e' una matrice dxn in cui ogni colonna e' un
% campione del v.a.
% (X1,...,Xd).
% Per aumentare l'ordine dei momenti considerati (
% default: fino all'ordine 4)

```

```
% togliere il commento alle sezioni corrispondenti ed
    aggiungere le
% variabili m4 (per momenti di ordine 4) e m5 (per
    momenti di ordine 5) in output.

%- momenti di ordine 1 -----

for i = 1 : d
    M1(i) = approx(X(i,:));
end
m1 = reshape(M1,[],1);

%- momenti di ordine 2 -----

for i = 1 : d
    for j = i : d
        aux = X(i,:).*X(j,:);
        M2(i,j) = approx(aux);
    end
end
m2 = reshape(M2,[],1);

%- momenti di ordine 3 -----

for i = 1 : d
    for j = i : d
        for k = j : d
            aux = X(i,:).*X(j,:).*X(k,:);
            M3(i,j,k) = approx(aux);
        end
    end
end
m3 = reshape(M3,[],1);

%- momenti di ordine 4 -----
```

```

for i = 1 : d
    for j = i : d
        for k = j : d
            for h = k : d
                aux = X(i,:) .* X(j,:) .* X(k,:) .* X(h,:);
                M4(i,j,k,h) = approx(aux);
            end
        end
    end
end
m4 = reshape(M4, [], 1);
%- momenti di ordine 5 -----

%   for i = 1 : d
%       for j = i : d
%           for k = j : d
%               for h = k : d
%                   for l = h : d
%                       aux = X(i,:) .* X(j,:) .* X(k,:) .* X(
h,:) .* X(l,:);
%                           M5(i,j,k,h,l) = approx(aux);
%                       end
%                   end
%               end
%           end
%       end
%   end
%   m5 = reshape(M5, [], 1);

function m = approx(Y)
    s = size(Y);
    n = s(2);
    m = (1./n) .* sum(Y);

```

Per quanto riguarda il secondo metodo invece, si è usato il seguente script:

```
n = 1425; % quanti campioni
```

```

d = 2; % dimensione delle v.a.
N = 50; % quanti punti confrontare
K = 2./(d+1); % larghezza della regione

X = randraw('norm', [0,1], [d n]);
Y = randraw('norm', [0,1], [d n]);
X = dlarray(X);
Y = dlarray(Y);

%-----scelta di A0-----
%A0=...;
%-----

X1 = randraw('norm', [0,1], [d n]);
Y1 = randraw('norm', [0,1], [d n]);
Q = X1-A0*Y1;
Q = dlarray(Q);
Qr = (2./pi).*atan(Q);

% costruisco regione discretizzata con punti casuali
D = randraw('uniform', [-(K/2), K/2],[N d]);

% calcolo exp(tQ) per tutti i campioni e per tutti i punti
ExpQr = exp(D*Qr);

% approssimo E[exp(tQ)] per ogni punto
for l = 1 : N
    MomQr(l) = approx(ExpQr(l,:));
end

% immetto la varieta' sulla quale minimizzare
manifold = sympositivedefinitefactory(d);
problem.M = manifold;
% immetto la funzione costo
problem.cost = @(A) cost_gen(A,X,Y,MomQr,D,n,N);

```

```
% immetto il gradiente ed hessiana della funzione costo
% (differenziazione automatica)
problem = manoptAD(problem);

in = eye(d);

% numero max di iterazioni
options.maxiter = 1e3;
% soglia sulla norma del gradiente
options.tolgradnorm = 1e-8;
% tempo a disposizione in sec
options.maxtime = 1000;
% passo minimo che il metodo puo' compiere
options.minstepsize = 1e-10;
options.useRand = 1;

[Ac, cost, info, options] = trustregions(problem,in,options);

% stampo la matrice calcolata
Ac

che utilizza a sua volta la funzione:

function c = cost_gen(A,X,Y,MomQr,D,N)

    R = X-A*Y;
    Rr = (2./pi).*atan(R);

    % calcolo exp(tRr) per tutti i campioni e per tutti i
    punti
    ExpRr = exp(D*Rr);

    % approssimo E[exp(tQ)] per ogni punto
    for l = 1 : N
        MomRr(l) = approx(ExpRr(l,:));
```

```
end
```

```
c = l2loss(MomRr, MomQr, DataFormat='U', Reduction='sum',  
          NormalizationFactor='none');
```

che calcola la funzione costo da minimizzare  $r_2^N$  descritta nella sezione precedente e che a sua volta utilizza la funzione `approx` già riportata precedentemente.

# Ringraziamenti

È arrivato il momento di ringraziare tutti coloro che hanno contribuito alla stesura di questo elaborato e tutti coloro che mi hanno accompagnato durante questo arduo ma soddisfacente percorso.

Ringrazio il mio relatore, il prof. Dario Trevisan, per avermi guidato nella realizzazione di questa tesi con infinita pazienza e incessante disponibilità. Ringrazio il dott. Nevio Dubbini per averci posto il problema statistico che abbiamo studiato nell'elaborato ed il prof. Fabio Durastante che ci ha consigliato la fondamentale libreria `manopt` utilizzata per le sperimentazioni.

Ringrazio tutta la mia famiglia per essere stati sempre presenti e per avermi aiutato, anche indirettamente, con il loro calore. Un ringraziamento speciale va ai miei genitori che mi hanno dato la possibilità e la libertà di studiare, che mi hanno sempre sostenuto in ogni modo possibile e a cui sarò grato per il resto della mia esistenza.

Ringrazio la mia ragazza Giada Dinucci che mi è stata sempre vicino, che ha sempre creduto in me e che sempre mi ha spronato ad andare avanti, anche nei momenti peggiori. La ringrazio per tutti i momenti di spensieratezza e di felicità che ha saputo regalarmi. Con lei ho vissuto questi anni, con lei ho imparato e con lei sono cresciuto.

Per ultimi, ma non per importanza, ringrazio tutti i miei colleghi e amici del dipartimento di matematica con cui ho condiviso gioie, delusioni e fatiche di questi primi anni trascorsi insieme. Li ringrazio per le discussioni matematiche, per i consigli che hanno saputo darmi e per tutti i momenti di divertimento passati insieme.

*Questi anni sono stati completamente fuori controllo.*





# Bibliografia

- [1] Thøger Bang. «The theory of metric spaces applied to infinitely differentiable functions». In: *Mathematica Scandinavica* (1953), pp. 137–152.
- [2] N. I. Akhiezer. *The Classical Moment Problem and Some Related Questions in Analysis*. Oliver & Boyd, 1965.
- [3] Gerald B Folland. *Real Analysis: Modern Techniques and Their Applications*. Vol. 40. John Wiley & Sons, 1999.
- [4] V. Powers e B. Reznick. «Polynomials that are positive on an interval». In: *Transactions of the American Mathematical Society* 352.10 (2000), pp. 4677–4692.
- [5] P. A. Absil, C. G. Baker e K. A. Gallivan. «Trust-region methods on Riemannian manifolds». In: *Foundations of Computational Mathematics* 7.3 (2007), pp. 303–330.
- [6] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.
- [7] H.L. Royden e P. Fitzpatrick. *Real Analysis*. Prentice Hall, 2010.
- [8] Luigi Ambrosio, Giuseppe Da Prato e Andrea Menzucchi. *Introduction to Measure Theory and Integration*. Vol. 10. Springer Science & Business Media, 2012.
- [9] Mitul Islam. «Denjoy-Carleman theorem and quasi-analyticity». In: (2014). URL: [https://www.academia.edu/19505975/Denjoy\\_Carleman\\_Theorem\\_and\\_Quasi\\_Analytic\\_Functions](https://www.academia.edu/19505975/Denjoy_Carleman_Theorem_and_Quasi_Analytic_Functions).
- [10] Maria Infusino. «Quasi-analyticity and determinacy of the full moment problem from finite to infinite dimensions». In: (2016), pp. 161–194.
- [11] K. Schmüdgen. *The Moment Problem*. Vol. 9. Springer, 2017.
- [12] K. Schmüdgen. «Ten lectures on the moment problem». In: *arXiv preprint arXiv:2008.12698* (2020).