

Università di Pisa



Facoltà di Scienze
Matematiche, Fisiche e Naturali

Corso di Laurea in Matematica
Anno Accademico 2002/2003

Tesi di Laurea
3 luglio 2003

Misure ottime per costi di trasporto e funzionali locali

Candidato
Filippo Santambrogio

Relatore
Prof. Giuseppe Buttazzo

Alla Nonna Ran

“Te voret studiaa matematica? e poi se te fet, te vivet d’espicient?”... “Me sun propri ricreduda: u vedu tuc chi scienziaa li che fan pu sé de palanca e u pensaa: anca el mee nevud el faraa i stess.” Beh, magari non è proprio così... ma di sicuro, ora più che mai, puoi vedere quanto io sia contento di quello che studio ed essere contenta con me.

Indice

Introduzione	5
Notazioni	10
1 Il problema del trasporto ottimo	11
1.1 Formulazione	11
1.1.1 Il problema di Monge	11
1.1.2 Il problema di Kantorovich	12
1.2 Piani di trasporto ottimi nel problema di Kantorovich	13
1.2.1 Esistenza	13
1.2.2 Condizioni di ottimalità	16
1.3 La formula di dualità	19
1.4 Gli spazi di Wasserstein	21
1.5 Trasporti classici ottimi	25
2 Funzionali locali e semicontinui su spazi di misure	30
2.1 Definizioni e preliminari	30
2.2 Un risultato di semicontinuità non convessa	35
2.3 Caratterizzazione dei funzionali locali semicontinui	43
3 Problemi di concentrazioni ottimali	45
3.1 Presentazione dei funzionali	45
3.2 Esistenza tramite il metodo diretto	47
3.3 Un risultato di esistenza senza compattezza	49
4 Condizioni necessarie di ottimalità	55
4.1 Ottimalità a ν fissata	55
4.1.1 Prime proprietà dei minimi	56
4.1.2 Relazione fra densità e potenziale	59
4.1.3 Regolarità e PDE	64

4.1.4	Relazione densità-potenziale per $p = 1$	67
4.2	Ottimalità a μ fissata	68
4.2.1	Cardinalità del supporto	68
4.2.2	Posizione e valore delle masse	70
4.3	Considerazioni sui minimi globali	72
5	Esistenza in \mathbb{R}^n	75
5.1	I problemi parziali	75
5.2	Esistenza per il problema globale	78
A	c-concavità e analisi convessa	86
A.1	c -concavità	86
A.1.1	Nozioni sulle funzioni c -concave	86
A.1.2	c -monotonia e funzioni c -concave	88
A.2	Convessità	89
A.2.1	Convesse coniugate e sottodifferenziale	89
A.2.2	Funzione di recessione e inf-convoluzione	92
A.2.3	Regolarità di funzioni convesse in \mathbb{R}^n	94
A.3	Multifunzioni	97
A.3.1	Convergenza di Hausdorff	97
A.3.2	Multifunzioni misurabili	101
A.3.3	Multifunzioni convesse	104
B	Γ-convergenza	107
B.1	Premesse	107
B.2	Definizione ed esempi	108
B.3	Convergenza dei minimi	111
B.4	Convergenza dei punti di minimo	114
	Bibliografia	116

Introduzione

Fra i temi più attuali della ricerca nell'ambito del calcolo delle variazioni vi è di sicuro quello delle applicazioni della teoria riguardante i problemi di trasporto ottimo a campi quali, per esempio, la pianificazione urbana.

Tipicamente, note le distribuzioni di abitazioni e uffici in una città, si cerca, per esempio, la linea di metropolitana che minimizza il costo totale del trasporto di ogni singolo abitante dalla propria casa al proprio posto di lavoro (in termini di tempo o sforzo), rispettando alcuni vincoli sul costo economico per la realizzazione di tale linea (in genere dipendente dalla sua lunghezza) o inserendo quantità relative a questo costo dentro il funzionale da minimizzare. Della teoria del trasporto ottimo si usa pertanto soprattutto la distanza che viene indotta fra le misure (densità di abitazioni e uffici) tramite il minimo costo di trasporto fra di esse (cosa che ha dato luogo all'introduzione degli spazi metrici di Wasserstein) più che l'effettiva esistenza e caratterizzazione di un'eventuale mappa di trasporto ottima.

Appare allora naturale domandarsi, fra le varie possibili città (con le loro distribuzioni di abitazioni e uffici), quale sia quella che, a prescindere dall'esistenza linee di trasporti pubblici, si presta di più allo spostamento casa-lavoro dei propri abitanti. Le incognite di questo problema saranno pertanto due misure di probabilità, rappresentanti come al solito la prima la distribuzione delle abitazioni e la seconda quella degli uffici, o, più in generale, dei servizi. Naturalmente si otterrebbe un costo di trasporto minimo se si scegliesse una coppia di misure coincidenti (cosa che corrisponde a distribuire in una qualsiasi maniera gli abitanti nella città, e fare lavorare ciascuno a casa propria), ma non è questa la scelta che vogliamo indurre, perchè non sarebbe realmente efficiente, in quanto, ragionevolmente, le due misure devono soddisfare requisiti di tipo opposto: sulle abitazioni ci si aspetta che siano il più sparse possibili, in modo da evitare sovraffollamento dei quartieri residenziali, mentre sui servizi ci si aspetta che siano rappresentati da misure molto concentrate. Bisognerà pertanto tenere conto, nel funzionale da minimizzare, di penalizzazioni da assegnare qualora le due misure non rispettino queste richieste.

Una possibilità è imporre che la misura μ (abitazioni) sia assolutamente continua con

densità μ^a rispetto alla misura di Lebesgue sul sottoinsieme di \mathbb{R}^2 che rappresenta la città e che la misura ν (servizi) sia invece puramente atomica, cioè della forma $\nu = \sum_{i=0}^{\infty} a_i \delta_{x_i}$, con $\sum_{i=0}^{\infty} a_i = 1$. Su queste misure consideriamo il funzionale

$$(\mu, \nu) \longmapsto T(\mu, \nu) + \int f(\mu^a(x)) dx + \sum_{i=0}^{\infty} g(a_i), \quad (1)$$

dove T rappresenta il minimo costo di trasporto fra le due misure, f è una funzione convessa e superlineare tale che $f(0) = f'(0) = 0$, mentre g è una funzione subadditiva tale che $g(0) = 0$ e $g'(0) = +\infty$.

In questo caso la quantità $(\mu^a(x))^{-1} f(\mu^a(x))$ rappresenta l'effetto negativo che un abitante del punto x è costretto a subire a causa di una densità di popolazione $\mu^a(x)$, valore che, per la superlinearità di f , cresce con il crescere della densità, andando dal valore 0 nel caso di assenza di abitazioni in una certa zona, fino a valori arbitrariamente grandi nel caso di sempre maggiore concentrazione. Questa quantità viene poi integrata rispetto alla misura μ stessa per ottenere lo scontento complessivo della popolazione causato dalle eventuali eccessive concentrazioni di abitazioni. Ciascun addendo della sommatoria che compare nell'ultimo termine del funzionale invece rappresenta i costi di costruzione o gestione di un polo di servizi posizionato nel punto x_i di grandezza a_i .

Questo modello riesce quindi a descrivere abbastanza bene i criteri per la scelta ottimo di queste due misure, se ciò che si vuole minimizzare relativamente alla concentrazione dei servizi è qualcosa rappresentato da un costo economico. Se invece si preferisse tenere maggior conto dell'efficienza di tali servizi (che si immagina essere maggiore se essi sono dislocati in posizioni non lontane fra di loro, per sfruttare al meglio il lavoro d'equipe) rispetto al loro costo sarebbe meglio scegliere in maniera diversa il funzionale da applicare alla sola ν (terzo addendo di quello complessivo), per favorire misure il cui supporto sia piccolo come diametro invece che come cardinalità (o, eventualmente, che abbiano supporto molto esteso, ma che diano poca massa alle parti più distanziate di tale supporto). Forniremo un esempio di possibile funzionale per svolgere questo ruolo (esempio 3.2), ma ci occuperemo in prevalenza del problema di minimo relativo al funzionale descritto dalla (1).

Il nostro obiettivo sarà quello di trovare risultati di esistenza per i minimi di questo funzionale, nonché di esibire condizioni necessarie affinché una coppia (μ, ν) risulti ottima, col duplice scopo di identificarla il più possibile (che è la cosa più utile ai fini pratici) e di studiarne alcune proprietà di regolarità.

Questo lavoro di tesi si aprirà con un capitolo in cui si riassumono brevemente alcuni fra i concetti di base della teoria del trasporto ottimo. Fra questi di sicuro vi sono gli spazi di Wasserstein e la formula di dualità. Una particolarità, infatti, del modo in cui si ricaveranno le condizioni di ottimalità nello studio del problema di minimo relativo

alla (1), sarà data dal fatto che spesso il termine $T(\mu, \nu)$ verrà espresso grazie all'uso della formula di dualità, invece che come estremo inferiore al variare delle mappe t . Il capitolo segue molto l'impostazione data all'argomento in un corso tenuto nel 2002 presso la Scuola Normale dal Prof. L.Ambrosio.

Nel secondo capitolo si affronta la teoria dei funzionali semicontinui e locali definiti su spazi di misure, sviluppata fra la fine degli anni '80 e l'inizio dei '90 da G.Bouchitté e G.Buttazzo, che presenta una vasta classe di funzionali, cui appartengono quelli che utilizziamo come penalizzazioni sulle misure μ e ν . Viene riproposta in dettaglio la dimostrazione presente in un articolo del 1990 ([5]) del fatto che funzionali, non necessariamente convessi, soddisfacenti certe condizioni siano semicontinui e poi richiamato un successivo risultato che serve a dare una caratterizzazione dei funzionali locali e semicontinui. Ricordiamo solo che un funzionale definito sullo spazio $\mathcal{M}(\Omega)$ delle misure finite vettoriali su Ω si dice locale se rispetta questa condizione di additività:

$$F(\lambda_1 + \lambda_2) = F(\lambda_1) + F(\lambda_2) \quad \lambda_1, \lambda_2 \in \mathcal{M}(\Omega) : \lambda_1 \perp \lambda_2.$$

È essenziale la possibilità di utilizzare funzionali di questo tipo che non siano convessi, se si vuole costruire, restando nell'ambito dei funzionali locali e godendo dei vantaggi dal punto di vista variazionale della semicontinuità, un funzionale che favorisca le misure concentrate.

Per la teoria esposta in questi due capitoli risultano molto utili alcuni strumenti tipici dell'analisi convessa, quali la trasformata di Fenchel, il sottodifferenziale, l'inf-convoluzione. Anche nozioni sulle multifunzioni a valori chiusi, in particolare di quelle anche a valori convessi, argomento strettamente connesso con l'analisi convessa, sono risultate talvolta necessarie per i risultati del capitolo 2. Questi argomenti, insieme a quello della c -concavità, che estende il concetto di convessità, sottodifferenziale e trasformata di Fenchel ad un contesto più astratto e che è essenziale nello studio del problema del trasporto ottimo, sono stati inseriti in un apposito capitolo, denominato appendice A.

I capitoli dal 3 al 5 invece sono dedicati allo studio variazionale dei funzionali che modellizzano il nostro problema. Nel capitolo 3 vengono presentati tali funzionali ed esposti i risultati di esistenza: si studia il caso di misure ambientate in un generico spazio metrico Ω compatto e quello in cui Ω è un convesso limitato di \mathbb{R}^n , non necessariamente chiuso. Il capitolo 4 è invece dedicato a condizioni necessarie di ottimalità, che permettono di intuire l'aspetto che avrà una configurazione (μ, ν) ottima, ma anche di scoprire alcune condizioni di regolarità. Per esempio, nel caso in cui la funzione f sia data da $f(y) = y^2/2$ ed il costo di trasporto sia calcolato usando il costo puntuale $c(x, y) = |x - y|$ si può scoprire che μ^a è una funzione 1-lipschitziana. Alcuni risultati sono validi anche per le misure che minimizzano il funzionale (1), avendo fissato l'altra

misura, e non solo per le coppie (μ, ν) ottime per l'intero funzionale. Nel capitolo 5 invece si prova un risultato di esistenza in \mathbb{R}^n (quindi senza compattezza né limitatezza dello spazio ambiente), dimostrando che si può produrre una successione minimizzante costituita da misure concentrate su un singolo insieme limitato. In questo caso è anche abbastanza chiaro come siano fatte le coppie (μ, ν) ottimali, che si dimostra essere costituite da una ν con un numero finito di atomi x_i ciascuno circondato da una palla $B(x_i, R_i)$ su cui la μ ha densità

$$\mu^a(x) = k(R_i^p - |x - x_i|^p),$$

dove k è la funzione inversa della funzione f' . La determinazione dei raggi R_i avviene tramite la minimizzazione di una funzione definita su un sottoinsieme di \mathbb{R}^N , per un opportuno N di cui si dà una stima.

Nell'appendice B sono riassunte le nozioni di Γ -convergenza che vengono usate qua e là, essenzialmente nel capitolo 4 e semplicemente come strumento per ottenere, nel caso di funzionali complicati ma approssimabili con funzionali più semplici, alcune proprietà sui rispettivi minimi.

Si può pensare che questo stesso modello variazionale si applichi in altri campi, diversi da quello della pianificazione urbana. In effetti, questo modello sembra difficilmente applicabile, in pratica, a qualcosa dell'ordine di grandezza di una città, perchè esso non dà indicazioni, come nei problemi relativi alla rete di trasporti pubblici, su come inserire nuovi elementi in un ambito urbano (per esempio la metropolitana), ma sui criteri nell'assegnazione della destinazione delle varie aree urbane, cosa che ha un suo significato solo prima della costruzione e dello sviluppo della città stessa. È più facile che questo modello possa quindi essere applicato in scala ridotta, per esempio sostituendo la città con un villaggio turistico e identificando i servizi con i bar o i centri di ritrovo, oppure nella pianificazione di un nuovo quartiere, da inserire in una città già esistente. D'altra parte questo tipo di modello si presta ad affrontare una generica situazione in cui si desidera scegliere nel modo ottimo una coppia di misure, seguendo criteri opposti di concentrazione (una sparsa, l'altra concentrata), e richiedendo che esse non siano troppo lontane se si misura tale lontananza tramite il costo del trasporto.

Una situazione di questo genere si potrebbe verificare allorquando μ e ν rappresentano due stadi successivi di una distribuzione di materiale, laddove non sia fissato a priori lo stadio finale della distribuzione (nel nostro caso μ), non vi sia cioè qualcuno che abbia ordinato tale materiale e lo desideri in un posti precisi, ma si voglia che tale stadio sia il più sparso possibile. A μ si arriva tramite uno stadio intermedio di distribuzione (rappresentato da ν), per il quale sono previsti dei prezzi che naturalmente crescono con l'aumentare del numero di consegne da effettuare (tipicamente il prezzo di ogni consegna è dato da una funzione subadditiva della quantità consegnata). Il termine di trasporto

gioca il ruolo del costo di un successivo spargimento del materiale affidato ad agenti locali, mentre il termine che costringe μ ad essere sparsa quello dell'insoddisfazione nel caso che tale misura non lo sia abbastanza.

Come esempio si può pensare allo stato maggiore di un esercito che desidera infiltrare dei soldati in tutto il territorio nemico. Per prima cosa vengono allestiti dei voli per paracadutarli, distribuendoli quindi provvisoriamente in una maniera rappresentata da una misura ν , con un costo economico dato dal terzo addendo del funzionale, che prevede una spesa per ogni singolo volo, secondo costi che crescono in maniera subadditiva con la quantità di militari trasportati (perchè se in un volo i soldati sono tanti è necessario un aereo più capiente che, per esempio, consuma di più). Successivamente essi hanno l'ordine di spostarsi a piedi in territorio nemico in modo da essere distribuiti secondo una nuova misura μ , della quale si desidera che non presenti concentrazioni di soldati troppo alte, per minimizzare il rischio che essi vengano scoperti; il secondo addendo del funzionale è un termine di rischio che misura la probabilità che ciò avvenga e i relativi effetti negativi. Il termine di trasporto rappresenta il costo, in tempo e fatica, per eseguire lo spostamento.

Notazioni

$\mathcal{P}(\Omega)$	insieme delle misure di probabilità su Ω
$\mathcal{M}(\Omega)$	spazio delle misure vettoriali a variazione finita su Ω
$t_{\sharp}\mu$	misura immagine di μ secondo t : $t_{\sharp}\mu(A) = \mu(t^{-1}(A))$
$spt(\mu)$	supporto della misura μ (minimo chiuso su cui μ è concentrata)
δ_x	misura di Dirac concentrata nel punto x
\sharp	misura che conta i punti: $\sharp A = \text{cardinalità dell'insieme } A$
$m_p(x_0, \mu)$	momento p -esimo della misura μ rispetto al punto x_0 : $(\int d(x, x_0)^p \mu(dx))^{1/p}$
\perp	ortogonalità (o singolarità) di misure su Ω : $\lambda_1 \perp \lambda_2 \Leftrightarrow \exists A, B \subset \Omega : A \cap B = \emptyset, A \cup B = \Omega, \lambda_1 (A) = \lambda_2 (B) = 0$
$C_b(\Omega)$	spazio delle funzioni continue e limitate su Ω
$C_c(\Omega)$	spazio delle funzioni continue a supporto compatto su Ω
$C_0(\Omega)$	spazio delle funzioni continue che tendono a 0 all'infinito e su $\partial\Omega$
I_A	funzione caratteristica dell'insieme A : $I_A(x) = 1$ se $x \in A$, 0 se $x \notin A$
$\delta(\cdot A)$	funzione indicatrice dell'insieme A : $\delta(x A)(x) = 0$ se $x \in A$, $+\infty$ se $x \notin A$
$a \cdot b$	prodotto scalare euclideo fra i vettori a e b
$ a $	norma euclidea del vettore a
$ A $	misura di Lebesgue dell'insieme $A \subset \mathbb{R}^n$: $\mathcal{L}^n(A)$
ω_n	misura della palla unitaria in \mathbb{R}^n : $\mathcal{L}^n(B(0, 1))$
\subset	inclusione, non necessariamente stretta, fra insiemi
\mathbb{R}^+	insieme dei numeri reali non negativi: $[0, +\infty[$

Capitolo 1

Il problema del trasporto ottimo

1.1 Formulazione

1.1.1 Il problema di Monge

Il problema del trasporto ottimo viene proposto per la prima volta da Gaspard Monge nel 1781. Traduciamo in termini moderni il suo enunciato, ma si tenga conto del fatto che, nel formularlo, Monge non aveva a disposizione la teoria della misura:

Problema 1.1 (M). *Date due funzioni misurabili non negative sullo spazio euclideo \mathbb{R}^n h_0 e h_1 con $\int_{\mathbb{R}^n} h_0(x)dx = \int_{\mathbb{R}^n} h_1(x)dx$, trovare fra tutte le mappe di trasporto t ($t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ funzione misurabile, in cui il punto $t(x)$ rappresenta la posizione in cui andrà a trovarsi dopo il trasporto la massa inizialmente presente in x) che trasformino h_0 in h_1 (tali cioè che $t_{\#}h_0 \cdot \mathcal{L}^n = h_1 \cdot \mathcal{L}^n$) quella che minimizzi il costo del trasporto, valutato in $\int_{\mathbb{R}^n} |x - t(x)|h_0(x) dx$.*

Il problema non è di facile soluzione per via della struttura classe delle funzioni t ammissibili, in cui il vincolo è molto diverso da quelli tipici dei problemi di calcolo delle variazioni. In questo caso se il trasporto t è iniettivo e lipschitziano si può vedere che esso è ammissibile se e solo se

$$h_0(x) = h_1(t(x))|det \nabla t(x)| \quad \mathcal{L}^n - q.o. x \quad (1.1)$$

dove \mathcal{L}^n indica la misura di Lebesgue nello spazio considerato. L'equazione (1.1) è evidentemente non lineare in t e pertanto non si può sperare che un limite debole di funzioni ammissibili sia ancora ammissibile, cosa che sarebbe stata molto comoda per affrontare classicamente il problema.

Naturalmente il problema di Monge può essere esteso facilmente a casi più generali. Si possono sostituire le funzioni h_0 e h_1 con due misure μ e ν non più necessariamente

assolutamente continue rispetto a \mathcal{L}^n , sostituire lo spazio euclideo con uno spazio diverso (tipicamente un generico spazio metrico localmente compatto e separabile), sostituire la distanza euclidea $|x - t(x)|$ con una generica funzione costo $c(x, t(x))$, su cui si potranno fare di volta in volta delle richieste (continuità...), e permettere addirittura che gli spazi di partenza e di arrivo di t siano diversi. Il vincolo resta che la misura immagine di μ tramite t sia ν . Pertanto, in questo ambito più generale, si cercherà

$$t_0 : (X, \mathcal{B}(X)) \rightarrow (Y, \mathcal{B}(Y)),$$

dove X e Y sono spazi metrici e $\mathcal{B}(X), \mathcal{B}(Y)$ le loro tribù borelliane (eventualmente completate rispetto alle misure μ e ν), che minimizzi

$$\int_X c(x, t(x)) \mu(dx)$$

fra tutte le t tali che $t_{\#}\mu = \nu$, date una funzione $c : X \times Y \rightarrow \overline{\mathbb{R}}$, che rappresenta il costo puntuale di trasporto, e due misure di probabilità μ, ν su X e Y rispettivamente.

Questo problema in generale non ha soluzione non solo perché potrebbe non esistere un minimo ma perché potrebbero non esistere t ammissibili!

Esempio 1.1. Se $\mu = \delta_p$ e $\nu = \frac{1}{2}\delta_p + \frac{1}{2}\delta_q$ con $p \neq q$ non esiste nessuna mappa t tale che $t_{\#}\mu = \nu$.

L'immagine tramite un qualunque trasporto di una misura concentrata in un solo punto è infatti sempre una misura dello stesso tipo.

1.1.2 Il problema di Kantorovich

La variante proposta da Leonid Kantorovich nel 1942 del problema di Monge è questa:

Problema 1.2 (K). *Date μ, ν misure di probabilità su X, Y rispettivamente, trovare γ misura su $X \times Y$ che abbia μ e ν come marginali (cioè $\gamma \in \mathcal{P}(X \times Y) : (\pi_X)_{\#}\gamma = \mu, (\pi_Y)_{\#}\gamma = \nu$, dove π_X e π_Y sono le usuali proiezioni di $X \times Y$ su X e Y) e che minimizzi la quantità $I(\gamma) = \int_{X \times Y} c d\gamma$.*

Le misure γ , chiamate piani di trasporto, rappresentano un modo per trasportare la massa dallo spazio X allo spazio Y , senza il vincolo di assegnare la stessa destinazione a ciò che in partenza si trova nello stesso punto, vincolo che si ha quando si lavora con le mappe t . In particolare $\gamma(A \times B)$ rappresenta la quantità di massa che si trovava in A e viene spostata in B . Naturalmente un trasporto alla Monge induce un piano di trasporto: precisamente al trasporto t è associato il piano $\gamma = (id \times t)_{\#}\mu$. Non solo: il costo associato da Monge ad un trasporto t è lo stesso associato in questo contesto al piano da lui indotto in quanto

$$\int_{X \times Y} c(x, y) ((id \times t)_{\#}\mu)(dx, dy) = \int_X c(x, t(x)) \mu(dx)$$

I vantaggi di questa variante sono numerosi: innanzitutto esiste sempre almeno un piano di trasporto ammissibile (per esempio $\gamma = \mu \otimes \nu$), inoltre il vincolo sulla misura γ è lineare e l'insieme dei piani ammissibile è chiuso rispetto alla convergenza debole nella dualità con le funzioni continue e limitate su $X \times Y$. Nella prossima sezione si dimostrerà, con un'ipotesi molto leggera sulla funzione costo c (la semicontinuità inferiore), l'esistenza del minimo per questo problema. È possibile poi risolvere il problema di Monge scoprendo che in certi casi la γ che realizza il minimo nel problema di Kantorovich è forzata ad essere indotta da un trasporto, che pertanto risulterà ottimo per il problema di Monge.

La variante di Kantorovich rende anche più chiara un'interpretazione probabilistica (si veda, ad esempio, [13]). Si tratta cioè di cercare due variabili aleatorie X e Y , di cui sono assegnate le leggi μ, ν in modo da minimizzare la quantità

$$E [c(X, Y)],$$

che naturalmente dipende solo dalla loro legge congiunta γ . Il problema di Monge si riduce pertanto a quello, molto meno naturale, in cui si ammettono a competere solo le coppie $(X, t(X))$ al variare della mappa t .

1.2 Piani di trasporto ottimi nel problema di Kantorovich

1.2.1 Esistenza

Lemma 1.1. *Sia Ω uno spazio metrico; se la funzione $c : \Omega \rightarrow \mathbb{R} \cup \{+\infty\}$ è semicontinua inferiormente e $\inf c > -\infty$ esiste una successione $(c_k)_k \subset C_b(\Omega)$ tale che $c_k \nearrow c$.*

Dimostrazione. Definiamo $c'_k(x) = \inf_{y \in \Omega} c(y) + kd(x, y)$. È evidente che a x fissato i valori di c'_k sono crescenti e che ogni funzione c'_k è k -lipschitziana, in quanto estremo inferiore di funzioni k -lipschitziane. Inoltre $\inf c \leq c'_k \leq c$. Si conclude pertanto se si mostra che per ogni x vale $\sup_k c'_k(x) = c(x)$ e si sceglie $c_k = c'_k \wedge k$, ottenendo in questo modo funzioni limitate anche dall'alto.

Sia ora x fissato e $L = \sup_k c'_k(x)$. Ci basta dimostrare $L \geq c(x)$, pertanto si può supporre $L < +\infty$. Sappiamo che per ogni k esiste un punto $y_k \in \Omega$ tale che

$$c(y_k) + kd(x, y_k) < c'_k(x) + \frac{1}{k} \tag{1.2}$$

e quindi

$$d(x, y_k) < \frac{c(x) + k^{-1} - \inf c}{k} \rightarrow 0$$

Pertanto $y_k \rightarrow x$ e, per semicontinuità

$$c(x) \leq \liminf_k c(y_k)$$

ma, usando la (1.2), si ha in particolare $\liminf_k c(y_k) \leq \liminf_k c'_k(x) = L$. \square

Definizione 1.1. Una successione $(\mu_h)_h \subset \mathcal{P}(\Omega)$ è detta *tesa* se per ogni $\varepsilon > 0$ esiste un compatto $K_\varepsilon \subset \Omega$ tale che per ogni h vale $\mu_h(\Omega \setminus K_\varepsilon) < \varepsilon$.

Lemma 1.2 (Prokhorov). *Data su uno spazio metrico Ω una successione di misure $(\mu_h)_h \subset \mathcal{P}(\Omega)$, se tale successione è tesa, da essa si può estrarre una sottosuccessione $(\mu_{h_k})_k$ convergente debolmente nella dualità con le funzioni C_b ad una certa $\mu \in \mathcal{P}(\Omega)$.*

Dimostrazione. Applicando la definizione 1.1 con $\varepsilon = 1/n$ si ottiene una successione, che si può supporre crescente, di compatti $(K_n)_n$ tali che per ogni h e n valga

$$\mu_h(\Omega \setminus K_n) < \frac{1}{n} \quad (1.3)$$

Definendo $\mu_{h,n} = I_{K_n} \mu_h$, ad n fissato si ha una successione relativamente compatta nella topologia che stiamo considerando. Questo viene dall'identificazione dello spazio delle misure col duale dello spazio delle funzioni continue a supporto compatto, che si riduce allo spazio di tutte le C o di tutte le C_b , quando si lavora su un compatto. Il teorema di Alaoglu permette, pertanto, di estrarre da ciascuna successione $(\mu_{h,n})_h$ una sottosuccessione debolmente convergente e, con un procedimento diagonale, si può ottenere un'unica sottosuccessione e delle misure σ_n concentrate ciascuna su K_n tali che

$$\begin{aligned} & \mu_{h_k,n} \xrightarrow{*} \sigma_n \text{ per } k \rightarrow +\infty \\ \text{cioè } \lim_k \int_{K_n} \varphi d\mu_{h_k,n} &= \int_{K_n} \varphi d\sigma_n \quad \forall \varphi \in C_b(K_n), \end{aligned}$$

condizione che si riscrive così, tenendo conto che le $\mu_{h_k,n}$ e le σ_n sono concentrate su K_n :

$$\int_{\Omega} \varphi d\mu_{h_k,n} \xrightarrow{k} \int_{\Omega} \varphi d\sigma_n \quad \forall \varphi \in C_b(\Omega)$$

La condizione (1.3) e la definizione di $\mu_{h,n}$ danno

$$|\mu_{h_k,n} - \mu_{h_k,m}|(\Omega) < \frac{1}{n} \quad \forall n < m \quad (1.4)$$

dove $|\cdot|$ indica la misura variazione totale di una misura con segno e $|\cdot|(\Omega)$ è una norma sullo spazio $\mathcal{M}(\Omega)$, semicontinua inferiormente rispetto alla convergenza debole contro C_b . Pertanto, passando al limite in k nella (1.4),

$$|\sigma_n - \sigma_m|(\Omega) \leq \frac{1}{n} \quad \forall n < m$$

La successione (σ_n) è quindi di Cauchy ed esiste allora σ tale che $\sigma_n \rightarrow \sigma$. Ora mostriamo che $\mu_{h_k} \rightarrow \sigma$. Sia $\varphi \in C_b(\Omega)$: vale allora

$$\begin{aligned} \left| \int_{\Omega} \varphi d(\mu_{h_k} - \sigma) \right| &\leq \left| \int_{\Omega} \varphi d(\mu_{h_k} - \mu_{h_k,n}) \right| + \left| \int_{\Omega} \varphi d(\mu_{h_k,n} - \sigma_n) \right| + \left| \int_{\Omega} \varphi d(\sigma_n - \sigma) \right| \\ &\leq \frac{\|\varphi\|_{C_b}}{n} + \left| \int_{\Omega} \varphi d(\mu_{h_k,n} - \sigma_n) \right| + \|\varphi\|_{C_b} |\sigma_n - \sigma|(\Omega), \end{aligned}$$

relazione ottenuta tenendo conto che μ_{h_k} e $\mu_{h_k,n}$ coincidono su K_n . Passando al limsup in k si ottiene

$$\limsup_k \left| \int_{\Omega} \varphi d(\mu_{h_k} - \sigma) \right| \leq \|\varphi\|_{C_b} \left(\frac{1}{n} + |\sigma_n - \sigma|(\Omega) \right)$$

Dall'arbitrarietà di n segue la tesi (il fatto che σ sia una misura di probabilità segue dal fatto che la convergenza debole contro C_b fa passare al limite la massa totale di una misura, testando contro la funzione costante 1. \square)

Teorema 1.3. *Siano X e Y spazi metrici completi e separabili e $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ una funzione semicontinua inferiormente e limitata dal basso. Allora il problema (K) ammette almeno una soluzione.*

Dimostrazione. Useremo il metodo diretto del calcolo delle variazioni: sceglieremo cioè una topologia sull'insieme dei γ ammissibili e dimostreremo che il funzionale che vogliamo minimizzare è sequenzialmente semicontinuo inferiormente e che lo spazio in cui cerchiamo il minimo è sequenzialmente compatto. In virtù dei lemmi 1.1 e 1.2 la topologia indotta dalla convergenza debole nella dualità con C_b sarà una buona scelta.

Detti $I_k(\gamma) = \int_{\Omega} c_k d\gamma$, il teorema di convergenza monotona ci dice che da $c_k \nearrow c$ segue $I_k \nearrow I$. Pertanto I risulta semicontinuo inferiormente in quanto estremo superiore degli I_k , che sono continui. Per concludere manca la compattezza dell'insieme dei γ ammissibili, che si ottiene notando che ogni successione di piani di trasporto è tesa. Infatti per ogni $\varepsilon > 0$ esistono dei compatti $K^X \subset X$, $K^Y \subset Y$ tali che

$$\mu(X \setminus K^X) < \frac{\varepsilon}{2}, \quad \nu(Y \setminus K^Y) < \frac{\varepsilon}{2}$$

Si scelga allora $K_\varepsilon = K^X \times K^Y$. In questo modo si ottiene, data una successione $(\gamma_n)_n$ di piani, l'esistenza di un limite γ . Che γ sia ancora un piano ammissibile segue dal fatto che la convergenza debole che stiamo considerando fa passare al limite anche i marginali. \square

1.2.2 Condizioni di ottimalità

Definizione 1.2. Data una funzione $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, un sottoinsieme $\Gamma \subset X \times Y$ è detto c -ciclicamente monotono (o c -monotono) se per ogni n , ogni n -upla di punti $((x_i, y_i))_{i=1..n} \subset \Gamma$ e ogni permutazione $\sigma \in \mathfrak{S}^n$ vale

$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^n c(x_i, y_i)$$

Osservazione 1.1. In realtà la condizione è simmetrica nei due fattori X e Y .

Osservazione 1.2. Sfruttando il fatto che ogni permutazione si decompone nel prodotto di cicli disgiunti la definizione potrebbe essere data limitandosi alle permutazioni cicliche.

Teorema 1.4. *Sia c continua. Allora ogni piano di trasporto γ ottimo per (K) ha supporto c -monotono.*

Dimostrazione. Supponiamo per assurdo che $\text{spt}(\gamma)$ non sia c -monotono: allora esso contiene una n -upla $(x_1, y_1), \dots, (x_n, y_n)$ di punti tali che:

$$\sum_{i=1}^n (c(x_i, y_i) - c(x_i, y_{\sigma(i)})) > 0$$

per una certa $\sigma \in \mathfrak{S}^n$.

Siano $U_i \times V_i$ intorno di (x_i, y_i) tali che la (1.2.2) valga anche se si sostituisce alla n -upla $(x_1, y_1) \dots (x_n, y_n)$ una qualunque n -upla $(x'_1, y'_1), \dots, (x'_n, y'_n)$ tale che per ogni indice i valga $x'_i \in U_i, y'_i \in V_i$. Per l'esistenza di tali intorno si usa la continuità di c . Siano ora

$$\begin{aligned} \gamma_i &= I_{\{U_i \times V_i\}} \gamma \quad \alpha_i = \gamma_i(X \times Y) = \gamma(U_i \times V_i) \\ \alpha &= \min \alpha_i > 0 \quad (\text{disuguaglianza che segue da } (x_i, y_i) \in \text{spt}(\gamma)), \\ Z &= \prod_{i=1}^n (U_i \times V_i), \quad \lambda = \bigotimes_{i=1}^n \frac{\gamma_i}{\alpha_i}, \quad u_i, v_i \text{ le proiezioni canoniche,} \\ \tilde{\gamma} &= \gamma + \frac{\alpha}{n} \sum_{i=1}^n [(u_i \times v_{\sigma(i)})_{\#} \lambda - (u_i \times v_i)_{\#} \lambda]. \end{aligned}$$

Verifichiamo che $\tilde{\gamma}$ è un piano di trasporto ammissibile con $I(\tilde{\gamma}) < I(\gamma)$ e avremo un assurdo. Intanto verifichiamo $\tilde{\gamma} \geq 0$: si noti che $(u_i \times v_i)_{\#} \lambda = \gamma_i / \alpha_i$. La positività di $\tilde{\gamma}$ segue da

$$\frac{\alpha}{n} \sum_{i=1}^n \frac{\gamma_i}{\alpha_i} \leq \frac{1}{n} \sum_{i=1}^n \gamma_i \leq \gamma.$$

Le misure marginali di $\tilde{\gamma}$ sono date da

$$\begin{aligned}(\pi_X)_\# \tilde{\gamma} &= (\pi_X)_\# \gamma + \frac{\alpha}{n} \sum_{i=1}^n ((u_i)_\# \lambda - (u_i)_\# \lambda) = (\pi_X)_\# \gamma = \mu, \\(\pi_Y)_\# \tilde{\gamma} &= (\pi_Y)_\# \gamma + \frac{\alpha}{n} \left(\sum_{i=1}^n (v_{\sigma_i})_\# \lambda - (v_i)_\# \lambda \right) = (\pi_Y)_\# \gamma = \nu.\end{aligned}$$

Infine

$$I(\tilde{\gamma}) - I(\gamma) = \frac{\alpha}{n} \int_Z \sum_{i=1}^n n(c(u_i, v_{\sigma_i}) - c(u_i, v_i)) d\lambda,$$

quantità che è negativa, perché, su tutto Z , $\sum_{i=1}^n (c(u_i, v_{\sigma_i}) - c(u_i, v_i)) < 0$ in quanto le applicazioni u_i, v_i sono a valori in U_i e V_i rispettivamente. \square

Teorema 1.5. *Se c è non negativa e s.c.i. ogni γ ottimo è concentrato su un insieme c -monotono e, se vale anche che c è a valori reali e $\int_{X \times Y} c d\mu \otimes \nu < +\infty$, è vero anche il viceversa, cioè ogni γ concentrato su un insieme c -monotono è ottimo per (K) .*

Dimostrazione. In questa sezione dimostriamo solo la seconda parte dell'enunciato; la prima verrà come conseguenza della formula di dualità e sarà dimostrata nella prossima sezione. Consideriamo allora un piano di trasporto γ concentrato su un insieme Γ che sia c -monotono e la funzione ψ associata a Γ dal teorema di Ruschendorf (teorema A.3). Possiamo supporre che Γ sia unione numerabile di compatti di $X \times Y$ su cui la funzione c è continua. Infatti, grazie al teorema di Lusin, per ogni numero naturale h , esiste un sottoinsieme $A_h \subset \Gamma$ tale che c coincida su A_h con una funzione continua e $\gamma(\Gamma \setminus A_h) < 1/h$. Per la regolarità della misura γ si può sostituire A_h con un suo sottoinsieme compatto A'_h tale che $\gamma(\Gamma \setminus A'_h) < 2/h$. La funzione c risulta pertanto continua se ristretta ad un qualunque insieme A'_h , e pertanto, siccome per verificare la continuità su un'unione finita di chiusi è sufficiente verificarla su ciascuno di essi, anche sugli insiemi $\Gamma_i = \bigcup_{h=1}^i A'_h$, che sono ancora compatti. Poiché si ha $\gamma(\Gamma \setminus \Gamma_i) < 1/i$, è chiaro che la misura γ è concentrata anche su $\bigcup_{i=1}^{\infty} \Gamma_i$, che è ancora un insieme c -monotono, perché sottoinsieme di Γ . Varrà:

$$\begin{aligned}\psi(x) + \psi^c(y) &\leq c(x, y) \text{ su } X \times Y; \\ \psi(x) + \psi^c(y) &= c(x, y) \text{ su } \Gamma\end{aligned}$$

(si veda A.1). Preso allora un piano γ' qualsiasi varrà:

$$\begin{aligned}
\int_{X \times Y} c d\gamma' &\geq \int_{X \times Y} (\psi(x) + \psi^c(y)) \gamma'(dx, dy) \\
&= \int_X \psi d\mu + \int_Y \psi^c d\nu \\
&= \int_{X \times Y} (\psi(x) + \psi^c(y)) \gamma(dx, dy) \\
&= \int_{X \times Y} c d\gamma.
\end{aligned} \tag{1.5}$$

Questo naturalmente mostrerebbe l'ottimalità di γ , ma, perché tutti i passaggi della (1.5) abbiano senso bisogna verificare che ψ, ψ^c siano misurabili e semi-integrabili, per poter spezzare l'integrale della loro somma nella somma degli integrali.

Misurabilità di ψ : Sia $\Gamma = \bigcup_i \Gamma_i$, con Γ_i compatto e $c|_{\Gamma_i}$ continua per ogni i . Allora si ha:

$$\psi(x) = \inf_{n,i} \inf_{(x_j, y_j) \in \Gamma_i} c(x, y_n) - c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0). \tag{1.6}$$

poiché le funzioni $f_k : X \times (\Gamma_i)^n \rightarrow \mathbb{R}$, date da

$$(x, (x_1, y_1) \dots (x_n, y_n)) \mapsto c_k(x, y_n) - c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0)$$

sono semicontinue inferiormente e convergono crescendo alla funzione di cui si fa l'inf nella (1.6), si ottiene

$$\psi(x) = \inf_{n,i} \sup_k \inf_{z \in (\Gamma_i)^n} f_k(x, z) \tag{1.7}$$

Siccome le $f_k(x, z)$ sono k -lipschitziane in x a z fissato, facendo il sup al variare di k si ottiene una funzione s.c.i.; la (1.7) mostra pertanto che ψ si scrive come inf numerabile di funzioni s.c.i., che indicheremo con $\psi_{n,i}$.

Misurabilità di ψ^c : si ha

$$\begin{aligned}
\{y \in Y \mid \psi^c(y) \geq t\} &= \{y \in Y \mid \forall x \in X \ c(x, y) - \psi(x) \geq t\} \\
&= \bigcap_p \bigcup_{n,i} \left\{ y \mid \forall x \in X \ c(x, y) - \psi_{n,i}(x) \geq t - \frac{1}{p} \right\} \\
&= \bigcap_p \bigcup_{n,i} \bigcap_q \bigcup_k \left\{ y \mid \forall x \in X \ c_k(x, y) - \psi_{n,i}(x) \geq t - \frac{1}{p} - \frac{1}{q} \right\}.
\end{aligned}$$

Le funzioni $c_k(x, y) - \psi_{n,i}$ sono s.c.s., pertanto hanno sopralivelli chiusi: che contengano tutti gli $x \in X$ oppure solo quelli di un denso numerabile $D = \{x_j \mid j \in \mathbb{N}\}$ è quindi la stessa cosa. Si può allora arrivare a

$$\{y \mid \psi^c(y) \geq t\} = \bigcap_p \bigcup_{n,i} \bigcap_q \bigcup_k \bigcap_j \left\{ y \mid c_k(x_j, y) - \psi_{n,i}(x_j) \geq t - \frac{1}{p} - \frac{1}{q} \right\},$$

che mostra che i sopralivelli di ψ^c sono borelliani.

Semi-integrabilità di ψ^c (l'osservazione A.2 crea una vera simmetria fra ψ e ψ^c , in questo caso dove c è a valori reali e $\psi < +\infty$): si ha che per μ -q.o. $x \in X$

$$\psi(x) > -\infty \text{ e } \int_Y c(x, y) \nu(dy) < +\infty \quad (1.8)$$

Si scelga x per cui valgano le disuguaglianze della (1.8), allora, siccome per ogni coppia $(x, y) \in X \times Y$ vale $\psi^c(y) \leq c(x, y) - \psi(x)$, si ha

$$\int_Y (\psi^c(y))^+ \nu(dy) \leq \int_Y c(x, y) \nu(dy) + (\psi(x))^- < +\infty$$

□

1.3 La formula di dualità

Teorema 1.6. *Ipotizziamo che la funzione costo c sia s.c.i., reale e limitata dal basso e che $\int_{X \times Y} c d\mu \otimes \nu < +\infty$. Allora*

$$\min(K) = \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \mid \varphi \in L^1(\mu), \psi \in L^1(\nu), \varphi(x) + \psi(y) \leq c(x, y) \right\}.$$

Dimostrazione. Naturalmente vale una disuguaglianza. Infatti per ogni γ piano di trasporto e per ogni φ, ψ scelte fra le funzioni su cui si fa il sup, si ha

$$\int_{X \times Y} c d\gamma \geq \int_{X \times Y} \varphi(x) + \psi(y) \gamma(dx, dy) = \int_X \varphi d\mu + \int_Y \psi d\nu. \quad (1.9)$$

Mostriamo ora l'altra disuguaglianza. Supponiamo dapprima c continua e limitata e consideriamo un piano di trasporto γ ottimo in (K). Per i teoremi 1.4 e A.3 e l'osservazione A.1 esiste una funzione c -concava che chiameremo $\hat{\psi}$ tale che

$$\hat{\psi}(x) + \hat{\psi}^c(y) = c(x, y) \quad \forall (x, y) \in \text{spt}(\gamma).$$

Allora scegliendo $\varphi = \hat{\psi}$, $\psi = \hat{\psi}^c$ si ottiene l'uguaglianza nella (1.9). Il fatto che con questa scelta di φ, ψ si ottengano funzioni misurabili e integrabili si vede con lo stesso ragionamento della parte finale del teorema 1.5 (ma, in realtà, nel caso che c sia continua si potrebbe semplificare notevolmente, perché in questo caso le funzioni c -concave sono s.c.s.), e da questo segue la tesi nel caso continuo. Si vede addirittura che il sup è un massimo realizzato da una coppia di funzioni data da una c -concava e dalla sua c -coniugata.

Supponiamo ora c solo s.c.i., e siano c_k le funzioni continue date dal lemma 1.1. Vale

$$\begin{aligned} \min(K) &= \min_{\gamma} I(\gamma) = \sup_k \min_{\gamma} I_k(\gamma) \\ &= \sup_k \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \mid \varphi \in L^1(\mu), \psi \in L^1(\nu), \varphi(x) + \psi(y) \leq c_k(x, y) \right\} \\ &\leq \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \mid \varphi \in L^1(\mu), \psi \in L^1(\nu), \varphi(x) + \psi(y) \leq c(x, y) \right\}, \end{aligned}$$

il che dimostra l'unica disuguaglianza mancante. \square

Osservazione 1.3. Assegnate una funzione ψ c -concava e un piano di trasporto ottimo γ risultano equivalenti:

1. la coppia (ψ, ψ^c) massimizza nella formula di dualità;
2. γ è concentrata su $G\partial_c\psi$.

poiché, se la funzione c è continua, abbiamo visto che il sup nella formula di dualità è un massimo realizzato da una coppia (ψ, ψ^c) e poiché per queste coppie la condizione $\psi(x) + \psi^c(y) \leq c(x, y)$ è automaticamente verificata, spesso considereremo il sup della formula di dualità come un sup fatto al variare di una sola funzione ψ fra le c -concave e diremo che una certa ψ realizza il massimo se lo fa la coppia (ψ, ψ^c) .

Esempio 1.2. Se $X = Y$, $c(x, y) = d(x, y)$ e μ, ν hanno primo momento finito, si ha

$$\min(K) = \max_{u \in Lip_1} \int_X u d(\mu - \nu).$$

Questo si ottiene tenendo conto dell'esempio (A.1), di quanto evidenziato nella dimostrazione precedente nel caso di c continua e del fatto che con le ipotesi fatte su μ, ν le funzioni lipschitziane risultano essere integrabili.

Richiamiamo qui l'enunciato della parte del teorema 1.5 non ancora dimostrata: con l'uso della formula di dualità appena mostrata saremo in grado di provare anche tale parte.

Teorema *Sia c s.c.i. e non negativa. Allora ogni piano di trasporto ottimo è concentrato su un insieme c -monotono.*

Dimostrazione. Sia γ ottimo e si consideri una successione $((\varphi_h, \psi_h))_h$ massimizzante nel sup dato dalla formula di dualità. Allora

$$\begin{aligned} c(x, y) - \varphi_h(x) - \psi_h(y) &\geq 0; \\ \lim_{h \rightarrow \infty} \int_{X \times Y} c(x, y) - \varphi_h(x) - \psi_h(y) \gamma(dx, dy) &= 0. \end{aligned}$$

Estraendo eventualmente una sottosuccessione si ha anche convergenza puntuale a 0 su un insieme Γ su cui γ è concentrata. Pertanto per $(x, y) \in \Gamma$ vale

$$c(x, y) = \lim_h \varphi_h(x) + \psi_h(y).$$

Questa condizione ci permette di dimostrare che Γ è c -monotono. Infatti, data una n -upla $((x_i, y_i))_{i=1..n}$ di punti in Γ e $\sigma \in \mathfrak{S}^n$, si ha:

$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^n \varphi_h(x_i) + \sum_{i=1}^n \psi_h(y_{\sigma(i)}) = \sum_{i=1}^n \varphi_h(x_i) + \sum_{i=1}^n \psi_h(y_i),$$

per l'arbitrarietà di h , usando il fatto che

$$\sum_{i=1}^n \varphi_h(x_i) + \sum_{i=1}^n \psi_h(y_i) \rightarrow \sum_{i=1}^n c(x_i, y_i),$$

si ottiene la c -monotonia di Γ , su cui è concentrata γ . □

1.4 Gli spazi di Wasserstein

Dato uno spazio metrico localmente compatto e separabile X e un numero reale $p \geq 1$ vogliamo definire una distanza fra misure che tenga conto del costo di trasporto fra di esse, calcolato utilizzando come costo puntuale $c(x, y) = d(x, y)^p$.

Definizione 1.3. Chiamiamo *spazio di Wasserstein di esponente p* lo spazio delle misure di probabilità su X di momento p -esimo finito

$$\mathcal{W}_p = \left\{ \mu \in \mathcal{P}(X) \mid \exists x_0 \in X : \int_X d(x, x_0)^p \mu(dx) < +\infty \right\},$$

munito della distanza W_p data da

$$W_p(\mu, \nu) = \inf_{\gamma} \left(\int_{X \times X} d(x, y)^p \gamma(dx, dy) \right)^{1/p},$$

dove l'inf è fatto al variare dei γ piani di trasporto ammissibili.

Si noti che la condizione di momento finito è verificata per ogni $x_0 \in X$ se ne esiste uno per cui lo è, che nel caso di X limitato tutti gli spazi \mathcal{W}_p coincidono insiemisticamente con lo spazio delle misure di probabilità su X e che l'inf nella definizione di W_p è in realtà un minimo.

Resta però da verificare che la W_p poc'anzi definita sia davvero una distanza. Simmetria e non negatività sono ovvie; usando il fatto che l'inf è un minimo è facile vedere anche che $W_p(\mu, \nu) = 0$ se e solo se $\mu = \nu$. Infatti se $W_p(\mu, \nu) = 0$ esiste un piano di trasporto fra le due misure concentrato sulla diagonale, il che forza i suoi marginali (μ, ν) , appunto) a coincidere. Per la disuguaglianza triangolare si può sfruttare il seguente lemma.

Lemma 1.7 (Dudley). *Siano dati tre spazi metrici separabili X, Y, Z e due misure $\gamma_1 \in \mathcal{P}(X \times Y)$, $\gamma_2 \in \mathcal{P}(Y \times Z)$ con $(\pi_y)_\# \gamma_1 = (\pi_y)_\# \gamma_2 = \nu$. Allora esiste*

$$\hat{\gamma} \in \mathcal{P}(X \times Y \times Z) : (\pi_{x,y})_\# \hat{\gamma} = \gamma_1; (\pi_{y,z})_\# \hat{\gamma} = \gamma_2.$$

Dimostrazione. Ci servirà utilizzare una disintegrazione della misura γ_1 rispetto al fattore Y , cioè una famiglia $(\gamma_{1,y})_{y \in Y}$ di misure su $X \times Y$, tali che

$$\forall A \in \mathcal{B}(X \times Y) \quad y \mapsto \gamma_{1,y}(A) \text{ è borelliana e } \int_Y \gamma_{1,y}(A) \nu(dy) = \gamma_1(A)$$

e tali che, per ν -q.o. $y \in Y$, la misura $\gamma_{1,y}$ sia concentrata su $X \times \{y\}$.

L'esistenza di una disintegrazione equivale all'esistenza di una legge condizionale per la variabile aleatoria identità di $X \times Y$ in sé rispetto alla variabile aleatoria proiezione sul secondo fattore. Le ipotesi sugli spazi ci garantiscono l'esistenza della legge condizionale ed anche la sua (essenziale) unicità (si veda ad esempio [11], pag.78-80, dove si trova esattamente la dimostrazione dell'esistenza di una disintegrazione di una misura di probabilità, sotto ipotesi che qui sono verificate).

Considerata poi anche un'analoga disintegrazione di γ_2 sempre rispetto al fattore Y , siccome le misure $\gamma_{1,y}, \gamma_{2,y}$ possono essere pensate come misure su X e Z rispettivamente, si può utilizzare la famiglia di misure $\gamma_{1,y} \otimes \gamma_{2,y}$, pensarle come misure su $X \times Y \times Z$ concentrate ciascuna su $X \times \{y\} \times Z$ e costruire la misura

$$\hat{\gamma} = \int_Y \gamma_{1,y} \otimes \gamma_{2,y} \nu(dy).$$

È facile adesso verificare che $\hat{\gamma}$ ha le proprietà richieste. □

Teorema 1.8. *La funzione W_p della definizione 1.3 soddisfa la disuguaglianza triangolare ed è pertanto una distanza.*

Dimostrazione. Per verificare la disuguaglianza triangolare si consideri una terna di misure μ, ν, σ e γ_1, γ_2 piani di trasporto ottimo fra μ e ν e fra ν e σ rispettivamente. Grazie al lemma di Dudley procuriamoci una $\hat{\gamma}$ su $X \times X \times X$ e sia γ_3 la sua proiezione sul prodotto del primo e del terzo fattore. La misura γ_3 è un piano di trasporto fra μ e σ , infatti

$$\begin{aligned} (\pi_{x_1})_\# \gamma_3 &= (\pi_{x_1})_\# (\pi_{x_1, x_3})_\# \hat{\gamma} = (\pi_{x_1})_\# (\pi_{x_1, x_2})_\# \hat{\gamma} = (\pi_{x_1})_\# \gamma_1 = \mu; \\ (\pi_{x_3})_\# \gamma_3 &= (\pi_{x_3})_\# (\pi_{x_1, x_3})_\# \hat{\gamma} = (\pi_{x_3})_\# (\pi_{x_2, x_3})_\# \hat{\gamma} = (\pi_{x_3})_\# \gamma_2 = \sigma. \end{aligned}$$

Inoltre valgono le disuguaglianze

$$\begin{aligned}
W_p(\mu, \sigma) &\leq \left(\int_{X \times X} d(x, z)^p \gamma_3(d(x, z)) \right)^{1/p} \\
&= \left(\int_{X^3} d(x, z)^p \hat{\gamma}(d(x, y, z)) \right)^{1/p} \\
&\leq \left(\int_{X^3} (d(x, y) + d(y, z))^p \hat{\gamma}(d(x, y, z)) \right)^{1/p} \\
&\leq \left(\int_{X^3} d(x, y)^p \hat{\gamma}(d(x, y, z)) \right)^{1/p} + \left(\int_{X^3} d(y, z)^p \hat{\gamma}(d(x, y, z)) \right)^{1/p} \\
&= \left(\int_{X \times X} d(x, y)^p \gamma_1(dx, dy) \right)^{1/p} + \left(\int_{X \times X} d(y, z)^p \gamma_2(d(y, z)) \right)^{1/p} \\
&= W_p(\mu, \nu) + W_p(\nu, \sigma),
\end{aligned}$$

che concludono la dimostrazione. \square

Osservazione 1.4. In generale $p \leq q$ implica $W_p \leq W_q$, ma, se lo spazio X è limitato, vale anche una disuguaglianza di tipo opposto e tutte le distanze W_p risultano addirittura equivalenti. Infatti si verifica facilmente che

$$p \geq q \Rightarrow W_p \leq (W_q)^{1-q/p} (\text{diam}(X))^{q/p}.$$

Cerchiamo ora di caratterizzare la convergenza di misure di probabilità nello spazio \mathcal{W}_p .

Teorema 1.9. *Sia X uno spazio metrico compatto e $p \geq 1$. Date $\mu, (\mu_h)_h \subset \mathcal{W}_p$, sono equivalenti:*

1. $W_p(\mu_h, \mu) \rightarrow 0$;
2. $\mu_h \rightarrow \mu$.

Dimostrazione. (1 \Rightarrow 2) Poiché $W_p(\mu_h, \mu) \rightarrow 0$ vale anche $\delta_1(\mu_h, \mu) \rightarrow 0$. Grazie alla formula di dualità $\delta_1(\mu_h, \mu) = \sup_{u \in Lip_1} \int_X u d(\mu_h - \mu)$. Pertanto da 1 segue che per ogni $u \in Lip_1$ si ha $\lim_h \int_X u d\mu_h = \int_X u d\mu$. Questo per linearità continua a valere anche per tutte le u lipschitziane e, per densità, siccome le masse delle μ_h sono equimitate (tutte uguali a 1), anche per tutte le u continue (per dimostrare che in uno spazio metrico compatto ogni funzione continua si approssima uniformemente con funzioni lipschitziane basta usare il lemma 1.1 per ottenere delle funzioni lipschitziane che convergono puntualmente crescendo alla funzione data e tenere conto che la convergenza monotona ad una funzione continua su un compatto implica quella uniforme).

(2 \Rightarrow 1) Ci basterà dimostrare che si ha convergenza in \mathcal{W}_1 in virtù dell'osservazione

1.4. Sia $K = \{u \in Lip_1 | u(x_0) = 0\}$: per il teorema di Ascoli-Arzelà K è un compatto di $C(X)$. Se $L_h : K \rightarrow \mathbb{R}$ è il funzionale $u \mapsto \int_X u d\mu_h$ (e $L(u) = \int_X u d\mu$), per l'ipotesi di convergenza debole si ha convergenza puntuale di L_h a L su tutto $C(X)$, ma i funzionali L_h sono anche equicontinui (sono tutti lineari con norma 1). Come conseguenza di Ascoli-Arzelà la convergenza sul compatto K è anche uniforme, pertanto

$$\lim_{h \rightarrow +\infty} \sup_{u \in K} |L_h(u) - L(u)| = 0.$$

Questa è proprio la condizione per cui $\delta_1(\mu_h, \mu) \rightarrow 0$. □

Enunciamo anche la caratterizzazione delle successioni convergenti in $\mathcal{W}_p(X)$ con X spazio metrico proprio non necessariamente compatto.

Definizione 1.4. Uno spazio metrico X si dice *proprio* se ogni palla chiusa e limitata di X è compatta.

Osservazione 1.5. Uno spazio metrico proprio è anche completo, localmente compatto e σ -compatto.

Teorema 1.10. Sia X uno spazio metrico proprio, $p \geq 1$ e $\mu, (\mu_h)_h \subset \mathcal{W}_p$. Allora sono equivalenti:

1. $W_p(\mu_h, \mu) \rightarrow 0$;
2. $\mu_h \xrightarrow{*} \mu$ e $m_p(x_0, \mu_h) \rightarrow m_p(x_0, \mu) \forall x_0 \in X$;
3. $\mu_h \xrightarrow{*} \mu$ e $\lim_{R \rightarrow +\infty} \sup_h \int_{X \setminus B(x_0, R)} d(x, x_0)^p \mu_h(dx) = 0 \forall x_0 \in X$;
4. $\int_X u d\mu_h \rightarrow \int_X u d\mu \forall u \in C(X)$ di crescita di ordine p (cioè $\exists C > 0, x_0 \in X : |u(x)| \leq C(1 + d(x, x_0)^p)$).

Se ci limitiamo tuttavia al caso di X compatto, il teorema 1.9 ci rende evidente che lo spazio $\mathcal{W}_p(X)$ risulta anch'esso uno spazio metrico compatto. Infatti da una qualunque successione di misure di probabilità è possibile estrarre una sottosuccessione che converga debolmente ad una misura di probabilità (come dal lemma 1.2, applicato ad una successione di misure di probabilità su un compatto, che è pertanto banalmente tesa). Questo esempio mostra invece come, in generale, non sia abbia nemmeno locale compattezza di \mathcal{W}_p .

Esempio 1.3. Nello spazio $\mathcal{W}_p(\mathbb{R})$ la misura δ_0 non ha interni compatti.

Sia $\varepsilon > 0$: esiste allora una successione $(\mu_h)_h$ tale che $W_p(\mu_h, \delta_0) \leq \varepsilon$, ma priva di sottosuccessioni convergenti in \mathcal{W}_p (il che impedisce l'esistenza di un intorno compatto

di δ_0). Per costruirla si scelga una successione di punti x_h tali che $\frac{1}{h}|x_h|^p = \varepsilon^p$. Si scelga poi

$$\mu_h = \left(1 - \frac{1}{h}\right) \delta_0 + \frac{1}{h} \delta_{x_h}.$$

La condizione sugli x_h garantisce $W_p(\mu_h, \delta_0) \leq \varepsilon$. poiché $\mu_h \xrightarrow{*} \delta_0$ ogni eventuale sottosuccessione convergente di $(\mu_h)_h$ dovrebbe convergere a δ_0 per il teorema 1.10. Tuttavia $m_p(0, \mu_h) = \frac{1}{h}|x_h|^p = \varepsilon^p$ e questo impedisce la convergenza a δ_0 per la condizione 2 dello stesso teorema.

Presento ora un risultato, la cui dimostrazione è tanto laboriosa quanto l'enunciato è intuitivo, relativo alle geodetiche dello spazio metrico $\mathcal{W}_p(\mathbb{R}^n)$.

Definizione 1.5. Uno spazio metrico X si dice *geodetico* se per ogni coppia di punti $x, y \in X$ esiste un'isometria $j : [0, d(x, y)] \rightarrow X$ con $j(0) = x$, $j(dx, dy) = y$, cioè una curva continua che connette x a y , lunga esattamente tanto quanto la distanza fra di essi.

È chiaro che, se una siffatta j esiste, essa sarà una geodetica fra i suoi estremi, in quanto non si possono realizzare curve continue che li connettano di lunghezza minore della loro distanza.

Teorema 1.11. *Lo spazio $\mathcal{W}_p(\mathbb{R}^n)$ è geodetico e, dati una coppia di suoi elementi μ, ν e un piano di trasporto ottimo γ fra di essi, una geodetica è data da:*

$$j(s) = (\pi_s)_\# \gamma,$$

dove l'applicazione $\pi_s : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ è data da $\pi_s(x, y) = x + s(y - x)$.

Nel caso in cui γ sia dato da un trasporto classico $t : \mathbb{R}^n \rightarrow \mathbb{R}^n$, cioè sia della forma $\gamma = (id \times t)_\#(\mu)$, la geodetica trovata può anche essere espressa come

$$j(s) = (t_s)_\# \mu \quad \text{dove } t_s(x) = x + s(t(x) - x),$$

il che esprime chiaramente l'idea di spostare un po' per volta le masse costituenti μ in direzione della posizione che dovranno occupare per dare luogo alla misura ν .

1.5 Trasporti classici ottimi

Sappiamo da quanto visto nella parte dedicata alle condizioni necessarie di ottimalità dei piani di trasporto ottimi che, almeno nel caso di $c : X \times Y \rightarrow \mathbb{R}$ continuo, un γ ottimo deve necessariamente avere supporto c -monotono e, dal teorema A.3, che allora esiste una funzione c -concava ψ tale che $spt(\gamma) \subset G\partial_c \psi$. Se con opportune ipotesi si riuscisse

a dimostrare che queste condizioni implicano che per μ -q.o. $x \in X$ $\text{spt}(\gamma) \cap (\{x\} \times Y)$ è un singoletto, otterremmo che γ sarebbe indotto da un trasporto classico (sarebbe cioè della forma $\gamma = (id \times t)_\# \mu$, con t mappa di trasporto) grazie al seguente teorema.

Teorema 1.12. *Un piano di trasporto γ è indotto da un trasporto classico se il suo supporto Γ è un grafico (cioè se, posto $\Gamma_x = \{y \in Y \mid (x, y) \in \Gamma\}$, si ha $\#\Gamma_x \leq 1$ per ogni $x \in X$).*

Dimostrazione. Consideriamo una disintegrazione $(\gamma_x)_{x \in X}$ di γ rispetto al fattore X e pensiamola come una famiglia di misure su Y (come nel lemma 1.7). Allora

$$1 = \gamma(\Gamma) = \int_X \gamma_x(\Gamma_x) \mu(dx);$$

pertanto per μ -q.o. $x \in X$ la misura γ_x è concentrata su Γ_x , che però è costituito al più da un punto. Allora per tali x , l'insieme Γ_x è non vuoto e costituito da un punto $t(x)$ tale che $\gamma_x = \delta_{t(x)}$. Bisogna ora solo mostrare che t è misurabile e che induce proprio il piano γ . Per la misurabilità ci si metta fuori da un insieme trascurabile in cui Γ_x potrebbe essere vuoto, e lì si consideri, fissato un $B \in \mathcal{B}(Y)$, la funzione

$$X \ni x \longmapsto \int_Y I_B(y) \gamma_x(dy) = I_B(t(x)) = I_{t^{-1}(B)}(x).$$

Essa è misurabile per definizione di disintegrazione e questo corrisponde alla misurabilità di t . Fissati ora $A \in \mathcal{B}(X)$ e $B \in \mathcal{B}(Y)$, si ha

$$\begin{aligned} \gamma(A \times B) &= \int_A \gamma_x(B) \mu(dx) = \int_A I_B(t(x)) \mu(dx) = \int_A I_{t^{-1}(B)}(x) \mu(dx) \\ &= \mu(\{x \in X \mid x \in A, t(x) \in B\}) = (id \times t)_\# \mu(A \times B), \end{aligned}$$

il che mostra come γ sia proprio il piano di trasporto indotto da t . □

Si noti che, se con questo schema dimostrativo, si ottiene che il piano γ ottimo è per forza indotto da un trasporto, si ottiene anche l'unicità di tale piano. Infatti se γ_1, γ_2 fossero entrambi ottimi allora potremmo concludere che anche $\frac{1}{2}(\gamma_1 + \gamma_2)$ lo sarebbe, ma non può essere indotto da un trasporto se lo sono γ_1 e γ_2 , a meno che essi non coincidano (la sua disintegrazione sarebbe costituita in x da una somma di due misure di Dirac con peso $1/2$, concentrate rispettivamente in $t_1(x)$ e $t_2(x)$, se t_1, t_2 sono i trasporti che inducono γ_1, γ_2 , che si riduce ad una sola massa se e solo se $t_1 = t_2$).

Il prossimo lemma lascia intuire quale possa essere il modo con cui si può mostrare che Γ_x è costretto a non contenere più di un punto. Lavoreremo con $X = Y = \mathbb{R}^n$ e $c(x, y) = h(x - y)$, dove $h : \mathbb{R}^n \rightarrow \mathbb{R}$ è una funzione strettamente convessa.

La stretta convessità ci garantisce l'invertibilità della multifunzione ∂h , cioè ci assicura che, per ogni vettore $v \in \mathbb{R}^n$, esiste un unico $x \in \mathbb{R}^n$ tale che $v \in \partial h(x)$. Indicheremo questo vettore con $(\partial h)^{-1}(v)$.

Lemma 1.13. *Sia $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ c -concava e differenziabile in un punto x_0 . Allora se $y \in \partial_c \psi(x_0)$ vale $y = x_0 - (\partial h)^{-1}(\nabla \psi(x_0))$*

Dimostrazione. La funzione $c(x, y) - \psi(x)$ assume il suo minimo per $x = x_0$, pertanto, utilizzando (A.5),

$$\nabla \psi(x_0) \in \partial h(x_0 - y),$$

da cui si ricava la tesi. \square

Osservazione 1.6. Se si usasse il costo $c(x, y) = h(x - y) = |x - y|$, dal fatto $y \in \partial_c \psi(x_0)$ si potrebbe solo ricavare la direzione del vettore $y - x_0$ e non anche il suo modulo.

In virtù di questo lemma ci sarà sufficiente trovare condizioni su μ e su h per cui ogni funzione c -concava (con $c(x, y) = h(x - y)$) sia differenziabile μ -quasi ovunque. Un primo risultato, di rapida dimostrazione, è questo.

Teorema 1.14. *Siano μ, ν a supporto compatto in \mathbb{R}^n e $\mu \ll \mathcal{L}^n$: allora esiste ed è unico un piano di trasporto ottimo γ fra μ e ν , ed è indotto da un trasporto t , che si esprime come*

$$t(x) = x - (\partial h)^{-1}(\nabla \psi(x)) \quad \mu - q.o.x, \quad (1.10)$$

per una qualsiasi funzione c -concava ψ tale che γ sia concentrata su $G\partial_c \psi$.

Dimostrazione. Per l'ipotesi di supporto compatto il problema del trasporto può anche essere ambientato in un aperto limitato $\Omega \subset \mathbb{R}^n$, e la funzione h risulterà, in quanto convessa, lipschitziana in $\Omega - \Omega$. Una generica funzione c -concava in Ω (restringendo il costo c a $\Omega \times \Omega$) si esprime tramite

$$\psi(x) = \inf_{y \in \Omega} h(x - y) - \chi(y),$$

ed è pertanto inf di funzioni lipschitziane con la stessa costante di Lipschitz (quella di h in $\Omega - \Omega$, che resta un limitato). Pertanto ψ è lipschitziana e, per il teorema di Rademacher, differenziabile quasi ovunque secondo Lebesgue, quindi anche secondo μ . I ragionamenti fatti finora ci permettono di concludere l'esistenza e unicità di una soluzione per il problema (K), che risulta essere un piano indotto da un trasporto classico. \square

Esistono naturalmente molte altre condizioni sotto cui si può ottenere lo stesso risultato. Per esempio questo è uno dei più generali.

Teorema 1.15. *Sia $h \in C_{loc}^{1,1}(\mathbb{R}^n)$, μ tale che $\mu(A) = 0$ per ogni insieme A tale che $\mathcal{H}^{n-1}(A) < +\infty$; allora esiste ed è unico un piano di trasporto ottimo γ fra μ e ν , ed è indotto da un trasporto t . Inoltre, se $\text{spt}(\nu)$ è limitato o se $h \in C^{1,1}(\mathbb{R}^n)$, vale la relazione (1.10) fra t e una qualsiasi funzione c -concava ψ tale che γ sia concentrata su $G\partial_c \psi$.*

La dimostrazione di questo teorema si ottiene mostrando una regolarità migliore per le funzioni c -concave: precisamente si scopre che sono anche semiconcave (ovvero funzioni concave cui sia stato aggiunto un termine polinomiale di grado due), cosa che segue essenzialmente dal fatto che c stesso lo è (come conseguenza dell'ipotesi su h di essere $C^{1,1}$). Questo fa sì che esse godano delle stesse proprietà di differenziabilità delle funzioni concave (o convesse), delle quali si può dimostrare che sono differenziabili in tutti i punti al di fuori di un'unione numerabile di ipersuperfici lipschitziane (il che rende naturale l'ipotesi su μ), grazie al corollario A.14.

Se si vuole evitare l'ipotesi $h \in C_{loc}^{1,1}$, che escluderebbe, per esempio, le funzioni del tipo $h(z) = |z|^p$, $1 < p < 2$, si può vedere un altro risultato, che generalizza il teorema 1.14, ma che non abbiamo presentato prima solo perché richiede una variante del lemma 1.13.

Definizione 1.6. Diciamo che una funzione $\psi : \Omega \rightarrow \mathbb{R}$ è *approssimativamente differenziabile* in un punto x_0 se esiste una funzione $\tilde{\psi}$ differenziabile in x_0 , con $\psi(x_0) = \tilde{\psi}(x_0)$ e tale che $\{ \psi = \tilde{\psi} \}$ abbia densità 1 in x_0 . In tal caso poniamo $\nabla_{ap}\psi(x_0) = \nabla\tilde{\psi}(x_0)$.

Naturalmente si può vedere che questa è una buona definizione, nel senso che se esistono due diverse $\tilde{\psi}_1, \tilde{\psi}_2$ con le stesse proprietà i loro gradienti coincidono. In questo modo possiamo estendere il lemma 1.13

Lemma 1.16. *Sia $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ una funzione c -concava e approssimativamente differenziabile in un punto x_0 . Allora se $y \in \partial_c\psi(x_0)$ vale $y = x_0 - (\partial h)^{-1}(\nabla_{ap}\psi(x_0))$*

Dimostrazione. La dimostrazione è identica a quella con la differenziabilità, l'unica differenza consiste nel mostrare per altra via che dal fatto che $h(x-y) - \psi(x)$ ha un minimo in x_0 segue $\nabla_{ap}\psi(x_0) \in \partial h(x_0 - y)$: vogliamo dimostrare che per ogni v vale

$$h(x_0 + v - y) - h(x_0 - y) \geq \nabla_{ap}\psi(x_0) \cdot v.$$

Per farlo consideriamo l'insieme dei v per cui vale la disuguaglianza opposta e notiamo che, grazie alla continuità delle funzioni in gioco (convesse o lineari), ci basta mostrare che è trascurabile (seguirà poi che sarà vuoto, essendo aperto). Tale insieme si esprime come unione numerabile

$$\bigcup_{\varepsilon > 0} \{v | h(x_0 + v - y) - h(x_0 - y) < \nabla_{ap}\psi(x_0) \cdot v - \varepsilon|v|\},$$

pertanto ci basta mostrare che ciascuno di questi insiemi dipendenti da ε è trascurabile. Siccome però, per la convessità di h , se v appartiene ad uno di essi anche λv , con $0 < \lambda < 1$, vi appartiene, ci basta mostrare che hanno densità nulla in 0. Per mostrarlo si può usare la relazione $h(x_0 + v - y) - h(x_0 - y) \geq \psi(x_0 + v) - \psi(x_0)$, sostituire ψ con $\tilde{\psi}$ fuori da un'insieme di vettori v di densità nulla in 0 e usare la differenziabilità di $\tilde{\psi}$. \square

Teorema 1.17. *Sia h strettamente convessa e $\mu \ll \mathcal{L}^n$: allora per una qualsiasi ν esiste ed è unico un piano di trasporto ottimo γ fra μ e ν , ed esso è indotto da un trasporto t , che si esprime come*

$$t(x) = x - (\partial h)^{-1}(\nabla_{ap}\psi(x)) \quad \mu - q.o.x, \quad (1.11)$$

per una qualsiasi funzione c -concava ψ tale che $spt(\gamma) \subset G\partial_c\psi$.

Dimostrazione. Ci resta solo da dimostrare che ogni funzione c -concava è approssimativamente differenziabile μ -q.ovunque. Vale:

$$\psi(x) = \inf_y h(x - y) - \chi(y)$$

e sappiamo che l'inf è un minimo se $x \in \pi_X(\Gamma)$, dove $\Gamma = spt(\gamma)$, con γ piano di trasporto ottimo. L'insieme $\pi_X(\Gamma)$ è di misura piena (secondo μ). Poniamo $\psi_R = \inf_{|y| \leq R} h(x - y) - \chi(y)$ e $A_R = \{\psi = \psi_R\}$. Vale allora $\mu(\mathbb{R}^n \setminus \bigcup_R A_R) = 0$. Inoltre le ψ_R sono localmente lipschitziane grazie al fatto che h lo è, e pertanto differenziabili quasi ovunque (sia secondo Lebesgue che secondo μ). Nei punti di A_R di densità 1 per A_R stesso e di differenziabilità di ψ_R la ψ è approssimativamente differenziabile e tali punti sono tutti fuorché un insieme μ -trascurabile. \square

Con discorsi non banali di approssimazione tramite costi del tipo $|x - y|^{1+\varepsilon}$, tenendo conto del problema variazionale secondario che nasce da questa approssimazione e decomponendo \mathbb{R}^n in raggi di trasporto unidimensionali, avendo a disposizione risultati sul trasporto per una vasta classe di costi nello spazio \mathbb{R} , si riesce a dimostrare il seguente (vedi ad esempio [3]), che dà una soluzione al problema di Monge.

Teorema 1.18. *Se $\mu \ll \mathcal{L}^n$ e $c(x, y) = |x - y|$ allora, per ogni $\nu \in \mathcal{P}(\mathbb{R}^n)$, fra i piani di trasporto ottimi fra μ e ν ve ne è almeno uno della forma $(id \times t)_{\#}\mu$.*

Capitolo 2

Funzionali locali e semicontinui su spazi di misure

In questo capitolo esporremo alcuni risultati provenienti da [5] e [6], relativi alla caratterizzazione di funzionali, definiti sullo spazio delle misure vettoriali finite su uno spazio metrico Ω , che siano s.c.i. rispetto alla convergenza debole nella dualità con C_c (o C_0) e che godano di una certa proprietà di additività. In tali lavori si studia una più vasta classe di funzionali rispetto a quelli precedentemente presenti in letteratura, che erano essenzialmente della forma

$$\lambda \longmapsto \int_{\Omega} f\left(x, \frac{d\lambda}{d\mathcal{L}}(x)\right) \mathcal{L}(dx) + \int_{\Omega} f^{\infty}\left(x, \frac{d\lambda^s}{d|\lambda^s|}(x)\right) |\lambda^s|(dx),$$

per una certa misura positiva \mathcal{L} e una certa funzione $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ tale che per ogni x $f(x, \cdot)$ fosse convessa, risultando essi stessi funzionali convessi, mentre la teoria esposta in questo capitolo tratterà anche funzionali non convessi.

2.1 Definizioni e preliminari

In tutto il capitolo Ω sarà uno spazio metrico localmente compatto e separabile, su cui sia fissata una misura di Radon positiva \mathcal{L} priva di atomi. Letture vettoriali che considereremo saranno sempre a valori in qualche \mathbb{R}^n , con n fissato.

Definizione 2.1. Diremo che un funzionale $F : \mathcal{M}(\Omega) \rightarrow [0, +\infty]$ è *locale* se rispetta questa condizione di additività: per ogni coppia di misure $\lambda_1, \lambda_2 \in \mathcal{M}(\Omega)$ tali che $\lambda_1 \perp \lambda_2$ vale

$$F(\lambda_1 + \lambda_2) = F(\lambda_1) + F(\lambda_2).$$

Quando parleremo di semicontinuità ci riferiremo invece sempre alla convergenza debole* indotta su $\mathcal{M}(\Omega)$ dalla dualità con $C_0(\Omega)$.

Incontreremo spesso funzioni $g : \mathbb{R}^n \rightarrow [0, +\infty]$ subadditive:

Definizione 2.2. Una funzione $g : \mathbb{R}^n \rightarrow [0, +\infty]$ si dice *subadditiva* se, per ogni $v, w \in \mathbb{R}^n$, vale $g(v + w) \leq g(v) + g(w)$.

Richiederemo anche che g sia semicontinua inferiormente e che valga 0 in 0. Data una g con queste proprietà, definiamo la funzione $g_0 : \mathbb{R}^n \rightarrow [0, +\infty]$ tramite

$$g_0(v) = \limsup_{t \rightarrow 0^+} \frac{g(tv)}{t}.$$

Teorema 2.1. *Se g è s.c.i., subadditiva e $g(0) = 0$ allora g_0 è convessa, s.c.i., positivamente 1-omogenea e vale*

$$g_0(v) = \sup_{t > 0} \frac{g(tv)}{t} = \lim_{t \rightarrow 0^+} \frac{g(tv)}{t}. \quad (2.1)$$

Dimostrazione. L'omogeneità è un'immediata conseguenza della definizione; se dimostriamo la prima uguaglianza in (2.1), dalla semicontinuità e subaddittività di g segue quella di g_0 , espressa come estremo superiore. La convessità segue dalla subaddittività e dalla omogeneità, grazie al successivo lemma 2.2. Ci restano pertanto da mostrare solo le uguaglianze in (2.1).

Definendo $\alpha(r) = \sup_{0 < t \leq r} \frac{g(tv)}{t}$, si ha chiaramente una funzione crescente di r , il cui limite per $r \rightarrow 0^+$ è proprio $g_0(v)$. Tuttavia, grazie alla subaddittività di g ,

$$\alpha(2r) = \sup_{0 < t \leq r} \frac{g(2tv)}{2t} \leq \sup_{0 < t \leq r} \frac{g(tv)}{t} = \alpha(r),$$

che implica che α è costante e dimostra la prima uguaglianza.

Per la seconda supponiamo per assurdo che esista una successione di punti $(t_h v)_h$, con $t_h \rightarrow 0^+$ e un $\varepsilon > 0$ con

$$\frac{g(t_h v)}{t_h} \leq \sup_{0 < t \leq r} \frac{g(tv)}{t} - \varepsilon.$$

La subaddittività di g mostra che se questa relazione vale per $t = t_h$ allora vale anche per $t = kt_h$, per ogni k e h . Tuttavia l'insieme $\{kt_h | k, h \in \mathbb{N}\}$ è denso in \mathbb{R} (perché comprende successioni aritmetiche di ragione arbitrariamente piccola, e pertanto interseca ogni intervallo) e la semicontinuità inferiore di g mostra che se la disuguaglianza $g(tv) \leq t \left(\sup_{0 < t \leq r} \frac{g(tv)}{t} - \varepsilon \right)$ vale per t in un denso allora vale per ogni t , contraddicendo la definizione di sup. \square

Si noti che la semicontinuità è essenziale per l'ultima parte della dimostrazione, come mostra questo esempio.

Esempio 2.1. Sia $g : \mathbb{R} \rightarrow [0, +\infty]$ la funzione data da

$$g(t) = \begin{cases} 0 & \text{se } x \in \mathbb{Q} \\ 1 & \text{se } x \notin \mathbb{Q} \end{cases}.$$

Questa funzione è subadditiva ma non semicontinua e naturalmente non esiste $\lim_{t \rightarrow 0^+} \frac{g(t)}{t}$.

In particolare si noti come la subadditività non implichi in alcun modo proprietà tipiche della concavità, come la decrescenza dei rapporti incrementali.

Ora, come promesso, mostriamo le relazioni fra convessità, subadditività e omogeneità.

Lemma 2.2. *Sia data una funzione $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ con $f(0) = 0$. Si considerino questi tre concetti:*

1. f è convessa;
2. f è positivamente 1-omogenea;
3. f è subadditiva.

Due qualunque di essi implicano il terzo.

Dimostrazione. (1 + 2 \Rightarrow 3) Vale

$$f(v + w) = 2f\left(\frac{v + w}{2}\right) \leq 2\left[\frac{1}{2}f(v) + \frac{1}{2}f(w)\right] = f(v) + f(w),$$

dove la prima uguaglianza segue dall'omogeneità e la disuguaglianza dalla convessità.

(2 + 3 \Rightarrow 1) Dalle ipotesi di omogeneità e subadditività di f segue che

$$f(tv + (1 - t)w) \leq f(tv) + f((1 - t)w) = tf(v) + (1 - t)f(w),$$

(3 + 1 \Rightarrow 2) Si scelga v e si consideri la restrizione di f alla semiretta per v uscente da 0: ci basta dimostrare l'omogeneità su questa arbitraria semiretta. Notiamo innanzitutto che per la subadditività di f , unita alla convessità, o $f(tv) = +\infty$ per ogni $t > 0$, oppure f assume solo valori finiti. Nel primo caso l'omogeneità è dimostrata, nel secondo la f è continua (si veda e si adatti il corollario A.11) e quindi anche s.c.i. e, dalla prima uguaglianza in (2.1), si ha

$$\lim_{t \rightarrow 0^+} \frac{f(t)}{t} = \sup_{t > 0} \frac{f(t)}{t} \geq \frac{f(t_0)}{t_0} \geq \lim_{t \rightarrow 0^+} \frac{f(t)}{t},$$

dove, per l'ultima disuguaglianza, si è tenuto conto della convessità di f , che implica la crescita dei rapporti incrementali, e del fatto $f(0) = 0$. Da questo segue la costanza dei rapporti incrementali in 0, quindi l'omogeneità □

I funzionali che studieremo saranno della forma

$$F(\lambda) = \int_{\Omega} f\left(x, \frac{d\lambda}{d\mathcal{L}}(x)\right) \mathcal{L}(dx) + \int_{\Omega \setminus A_{\lambda}} \varphi\left(x, \frac{d\lambda^s}{d|\lambda^s|}(x)\right) |\lambda^s|(dx) + \int_{A_{\lambda}} g(x, \lambda(x)) \#(dx)$$

per opportune f, φ, g tali che per ogni punto fissato $x \in \Omega$, le funzioni $f(x, \cdot)$ e $g(x, \cdot)$ siano rispettivamente convessa e subadditiva e si annullino in 0 (il significato dell'insieme A_{λ} sarà esplicitato più avanti). Sotto ipotesi di semicontinuità e ipotesi che leghino i comportamenti di f, φ e g mostreremo un risultato di semicontinuità per funzionali di questa forma, che, come si può facilmente vedere, sono sempre locali. D'ora in avanti, per funzioni $g : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$, useremo il termine subadditive per intendere che lo sono a $x \in \Omega$ fissato.

Come si vede giocheranno un ruolo importante le misure non prive di atomi, perché proprio su di loro verrà a cadere la convessità di questi funzionali. Introduciamo allora queste notazioni:

- $\mathcal{M}^0(\Omega)$ spazio delle misure in $\mathcal{M}(\Omega)$ prive di atomi
- $\mathcal{M}^{\#}(\Omega)$ spazio delle misure in $\mathcal{M}(\Omega)$ puramente atomiche
- $\mathcal{M}_k^{\#}(\Omega)$ insieme delle misure in $\mathcal{M}^{\#}(\Omega)$ costituite da al più k atomi

Le misure puramente atomiche sono quelle della forma $\sum_{i=0}^{\infty} a_i \delta_{x_i}$, con $a_i \in \mathbb{R}^n$; quando esprimeremo una misura puramente atomica in questo modo intenderemo sempre implicitamente che i punti x_i siano distinti, cosicché $\lambda(\{x_i\}) = a_i$, salvo esplicita diversa indicazione.

È chiaro che vale

$$\mathcal{M}(\Omega) = \mathcal{M}^0(\Omega) \oplus \mathcal{M}^{\#}(\Omega).$$

Con A_{λ} indichiamo l'insieme degli atomi di λ , cioè l'insieme $\{x \in \Omega | \lambda(\{x\}) \neq 0\}$ e per comodità scriveremo spesso $\lambda(x)$ in luogo di $\lambda(\{x\})$.

Lemma 2.3. $\mathcal{M}_k^{\#}(\Omega)$ è sequenzialmente debolmente* chiuso. Inoltre, per ogni funzione $g : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$ s.c.i. e subadditiva, il funzionale

$$G(\lambda) = \int_{\Omega} g(x, \lambda(x)) \#(dx)$$

è sequenzialmente debolmente* s.c.i. sullo spazio $\mathcal{M}_k^{\#}(\Omega)$.

Dimostrazione. Dimostriamo innanzitutto che $\mathcal{M}_k^{\#}(\Omega)$ è chiuso, prendendo una successione $(\lambda_h)_h \subset \mathcal{M}_k^{\#}(\Omega)$, con $\lambda_h \xrightarrow{*} \lambda$. Allora si può scrivere

$$\lambda_h = \sum_{i=1}^k a_i^h \delta_{x_i^h}.$$

Poiché $|\lambda_h|(\Omega) = \sum_{i=1}^k |a_i^h|$, e poiché questa quantità è limitata in h per via della convergenza debole, tutte le successioni $(a_i^h)_h$ sono limitate. A meno di sottosuccessioni possiamo pertanto assumere che, per ogni indice i , esista un vettore a_i tale che $\lim_h a_i^h = a_i$ e che inoltre esista un insieme di indici $I \subset \{1, \dots, k\}$ tale che, per ogni $i \in I$, esista un punto x_i che verifichi $\lim_h x_i^h = x_i$, mentre per gli indici $i \notin I$ la successione $(x_i^h)_h$ esca da ogni compatto di Ω , il che però significa

$$\sum_{i \notin I} a_i^h \delta_{x_i^h} \xrightarrow{*} 0.$$

A questo punto è chiaro che

$$\lambda_h \xrightarrow{*} \sum_{i \in I} a_i \delta_{x_i},$$

da cui segue $\lambda = \sum_{i \in I} a_i \delta_{x_i} \in \mathcal{M}_k^\#(\Omega)$ (in quest'ultimo caso la scrittura di λ , pur mostrando che ha al più un numero di atomi pari alla cardinalità di I , che è minore di k , non è stata costruita tramite atomi necessariamente distinti).

Per la semicontinuità del funzionale G , utilizzando le stesse notazioni di prima e restringendosi alla stessa sottosuccessione di prima (il che è sempre possibile se si cerca di dimostrare la semicontinuità per assurdo, trovando una successione su cui si ha un limite strettamente minore del valore voluto), si ha

$$\liminf_h G(\lambda_h) \geq \liminf_h \sum_{i \in I} g(x_i^h, a_i^h) \geq \sum_{i \in I} g(x_i, a_i) \geq \sum_{x \in A_\lambda} g(x, \lambda(x)) = G(\lambda),$$

dove si è usata prima la semicontinuità di g e poi la sua subadditività (l'ultima disuguaglianza non è in generale un'uguaglianza proprio per il fatto che i punti x_i non sono necessariamente distinti e i vettori a_i non coincidono necessariamente con $\lambda(x_i)$). \square

Le ipotesi, che nella prossima sezione esplicheremo, serviranno a garantire la semicontinuità del funzionale F e, come abbiamo già detto, legheranno fra loro i comportamenti di f, φ e g , e lo faranno coinvolgendo le funzioni g_0 , a cui abbiamo già dedicato spazio, f^∞ (vedi in particolare la definizione A.9 e l'esempio A.3 nell'appendice dedicato ad alcuni dei numerosi strumenti di analisi convessa necessari per lo studio di questi argomenti) e φ_f , che introdurremo ora. Naturalmente quando parleremo di funzione di recessione o di convessa coniugata di una funzione come queste, dipendente da due variabili, di cui una in Ω e una nello spazio euclideo, intenderemo di congelare la prima e considerare come variabile solo la seconda, per esempio

$$f^\infty(x, v) = \lim_{t \rightarrow +\infty} \frac{f(x, tv)}{t}.$$

Definiamo allora, assegnata una funzione $f : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$, la funzione φ_f :

$$\varphi_f(x, v) = \sup \left\{ u(x)v : u \in C_0(\Omega), \int_\Omega f^*(x, u(x)) \mathcal{L}(dx) < +\infty \right\}.$$

È chiaro dalla definizione che φ_f è s.c.i. in (x, v) e convessa e positivamente 1-omogenea in v a x fissato. La funzione φ_f è strettamente legata alla f^∞ da queste relazioni (che si trovano in [7]):

$$\varphi_f(x, \cdot) \leq f^\infty(x, \cdot) \text{ per } \mathcal{L} - \text{q.o. } x \in \Omega; \quad (2.2)$$

$$\varphi_f(x, \cdot) \geq f^\infty(x, \cdot) \text{ per ogni } x \in \Omega \text{ se } x \mapsto \text{epi}(f^*(x, \cdot)) \text{ è s.c.i. in } \Omega. \quad (2.3)$$

2.2 Un risultato di semicontinuità non convessa

Teorema 2.4. *Siano assegnate tre funzioni f, φ, g di Borel su $\Omega \times \mathbb{R}^n$ a valori in $[0, +\infty]$, tali che $f(x, 0) = \varphi(x, 0) = g(x, 0) = 0$ per ogni $x \in \Omega$. Supponiamo esistano un insieme \mathcal{L} -trascurabile $M \subset \Omega$ e uno numerabile $N \subset \Omega$ tali che valgano le condizioni:*

H1) $f(x, \cdot)$ è convessa s.c.i. per ogni $x \in \Omega \setminus M$;

H2) esiste una funzione $u_0 \in L^1(\Omega, \mathcal{L}; \mathbb{R}^n)$ tale che l'applicazione $x \mapsto f(x, u_0(x))$ è una funzione L^1 ;

H3) su $(\Omega \setminus M) \times \mathbb{R}^n$ le funzioni f^∞ e φ_f coincidono;

H4) g è s.c.i. su $\Omega \times \mathbb{R}^n$;

H5) per ogni $x \in \Omega$ $g(x, \cdot)$ è subadditiva;

H6) $g \leq \varphi_f$ in $\Omega \times \mathbb{R}^n$;

H7) $g_0 = \varphi = \varphi_f$ in $(\Omega \setminus N) \times \mathbb{R}^n$.

Il funzionale

$$F(\lambda) = \int_{\Omega} f\left(x, \frac{d\lambda}{d\mathcal{L}}(x)\right) \mathcal{L}(dx) + \int_{\Omega \setminus A_\lambda} \varphi\left(x, \frac{d\lambda^s}{d|\lambda^s|}(x)\right) |\lambda^s|(dx) + \int_{A_\lambda} g(x, \lambda(x)) \#(dx)$$

è allora sequenzialmente debolmente* s.c.i. in $\mathcal{M}(\Omega)$.

Osservazione 2.1. Nel caso in cui le funzioni f, φ, g non dipendano esplicitamente dalla variabile x , il teorema 2.4 assicura la semicontinuità del funzionale

$$F(\lambda) = \int_{\Omega} f\left(\frac{d\lambda}{d\mathcal{L}}(x)\right) \mathcal{L}(dx) + \int_{\Omega \setminus A_\lambda} f^\infty\left(\frac{d\lambda^s}{d|\lambda^s|}(x)\right) |\lambda^s|(dx) + \int_{A_\lambda} g(\lambda(x)) \#(dx).$$

Si è tenuto conto, per verificare che questo funzionale rispetti le ipotesi del teorema 2.4, delle condizioni (2.2).

Un primo lemma che ci servirà per la dimostrazione è il seguente.

Lemma 2.5. *Sia F un funzionale sul duale X^* di uno spazio di Banach X e definiamo per ogni $\varepsilon > 0$ il funzionale F_ε dato da*

$$F_\varepsilon(y) = F(y) + \varepsilon \|y\|_{X^*} .$$

Allora F è sequenzialmente debolmente s.c.i. se e solo se gli F_ε lo sono per ogni $\varepsilon > 0$.*

Dimostrazione. La parte del solo se è banale. Per il viceversa supponiamo che $(y_h)_h$ sia una successione tale che $y_h \xrightarrow{*} y$. Allora per un certo $K \in \mathbb{R}^+$ varrà $\|y_h\|_{X^*} \leq K$ per ogni h . Pertanto, per ogni $\varepsilon > 0$, usando la semicontinuità di F_ε , si avrà

$$F(y) \leq F_\varepsilon(y) \leq \liminf_h F_\varepsilon(y_h) \leq \liminf_h F(y_h) + \varepsilon K,$$

che dà la tesi grazie all'arbitrarietà di ε . □

Questo ci permetterà di supporre che F sia minorato da un termine proporzionale a $\|\lambda\| = |\lambda|(\Omega)$, e pertanto che esista una costante $c > 0$ tale che $f(x, v) \geq c|v|$, $\varphi(x, v) \geq c|v|$, $g(x, v) \geq c|v|$.

Come successivo passo introduciamo le multifunzioni C_ε :

$$C_\varepsilon(x) = \{w \mid v \cdot w \leq g(x, v) \forall v : |v| \leq \varepsilon\},$$

poniamo inoltre $g_\varepsilon(x, v) = \delta^*(v|C_\varepsilon(x))$.

Lemma 2.6. *Le multifunzioni C_ε sono s.c.i..*

Dimostrazione. Fissiamo $x \in \Omega$, una successione $(x_h)_h$ tale che $x_h \rightarrow x$, un punto $w \in C_\varepsilon(x)$ e un numero $\delta > 0$: ci basterà mostrare che per h grande abbastanza $(1 - \delta)w \in C_\varepsilon(x_h)$ (in questo modo qualunque aperto intersechi C_ε in un punto w dovrà intersecare definitivamente anche $C_\varepsilon(x_h)$ in punti del tipo $(1 - \delta)w$). Se supponiamo che questo non sia vero otteniamo una successione $(v_h)_h$ di vettori tali che $v_h \cdot (1 - \delta)w > g(x_h, v_h)$. Grazie alla subadditività di g , nella condizione di appartenenza a C_ε ci si sarebbe potuti limitare ai vettori v di modulo compreso fra $\frac{\varepsilon}{2}$ e ε (perché $g(x, \frac{v}{2}) \geq \frac{1}{2}g(x, v)$) e pertanto possiamo scegliere i v_h in modo che $\frac{\varepsilon}{2} \leq |v_h| \leq \varepsilon$. A meno di sottosuccessioni si può supporre che, per un certo vettore $v \neq 0$, valga $v_h \rightarrow v$. Per la semicontinuità di g si ottiene

$$g(x, v) \leq (1 - \delta)v \cdot w \leq (1 - \delta)g(x, v),$$

dove la seconda disuguaglianza segue da $w \in C_\varepsilon(x)$ e $|v| \leq \varepsilon$. poiché però $v \neq 0$ e, per l'assunzione fatta su g a seguito del lemma 2.5, $g(x, v) > 0$, questo è assurdo. □

Come conseguenza abbiamo il seguente

Corollario 2.7. *Le funzioni g_ε sono s.c.i. su $\Omega \times \mathbb{R}^n$.*

Dimostrazione. Tenendo conto della definizione di g_ε , basta applicare il teorema A.37. \square

Ora ci servirà dimostrare che le funzioni g_ε approssimano puntualmente la funzione g_0 .

Lemma 2.8. *Le funzioni g_ε convergono puntualmente crescendo a g_0 quando ε decresce a 0.*

Dimostrazione. È chiaro dalla definizione che, per ε che decresce a 0, le g_ε crescono, e crescono alla funzione di supporto dell'insieme $C(x) = \bigcup_\varepsilon C_\varepsilon(x)$ (tutto il ragionamento si può fare ad x fissato, ignorando la dipendenza da x di g e degli insiemi C_ε, C). Naturalmente tale funzione di supporto è uguale a quella di \overline{C} , mentre il candidato limite g_0 , essendo positivamente 1-omogeneo, si esprime come funzione di supporto dell'insieme $G = \{w|v \cdot w \leq g_0(v) \forall v\}$. Si verifica facilmente, esprimendo g_0 come limite, grazie al teorema 2.1, che, per ogni $\varepsilon > 0$, vale l'inclusione $C_\varepsilon \subset G$, pertanto $C \subset G$ e, poiché G è chiuso, anche $\overline{C} \subset G$. Se dimostriamo anche l'inclusione opposta avremo l'uguaglianza dei due insiemi, che implica

$$g_0 = \delta^*(\cdot|G) = \delta^*(\cdot|C) = \sup_\varepsilon g_\varepsilon.$$

G è convesso e, grazie all'ipotesi di minorazione di $g(x, v)$ con $c|v|$, contiene la palla $B(0, c)$, quindi è anche a parte interna non vuota e pertanto vale allora l'uguaglianza fra G e la chiusura della sua parte interna. Ci basterà allora mostrare che la parte interna di G è contenuta in C . Se supponiamo che questo non sia vero troviamo un vettore w , un numero $\delta > 0$ e una successione di vettori $(v_h)_h$ tali che

$$\begin{aligned} \forall z : |z| < \delta \quad w + z \in G; \\ \forall h \quad \frac{1}{2} \leq |v_h| \leq 1; \\ 2^{-h}v_h \cdot w > g(2^{-h}v_h). \end{aligned}$$

Se fissiamo un numero naturale k , per la subadditività di g varrà, per $h \geq k$

$$\frac{g(2^{-k}v_h)}{2^{-k}} \leq \frac{g(2^{-h}v_h)}{2^{-h}} \leq v_h \cdot w.$$

Si può supporre che $v_h \rightarrow v$, per un certo vettore v e passare al liminf, ottenendo

$$\frac{g(2^{-k}v)}{2^{-k}} \leq v \cdot w;$$

d'altra parte, utilizzando il fatto che $w + z$ appartiene a G per $|z| < \delta$, si ha anche $v \cdot w \leq g_0(v) - v \cdot z$ e, passando all'inf in z , si ottiene $v \cdot w \leq g_0(v) - \delta|v|$, quindi

$$\frac{g(2^{-k}v)}{2^{-k}} \leq v \cdot w \leq g_0(v) - \frac{1}{2}\delta.$$

Quest'ultima relazione è assurda per l'arbitrarietà di k , visto che $g_0(v) = \lim_k \frac{g(2^{-k}v)}{2^{-k}}$. \square

L'ultimo fatto di cui avremo bisogno è il seguente (con il simbolo \square indichiamo l'inf-convoluzione, data dalla definizione A.10).

Lemma 2.9. *Sia $f : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$ una funzione di Borel tale che $f(x, 0) = 0$ per ogni $x \in \Omega$ e che soddisfi le ipotesi H1 e H2 e $C : \Omega \rightarrow \mathbb{R}^n$ una multifunzione s.c.i. a valori convessi chiusi tale che, per ogni x , si abbia $C(x) \subset \text{dom}(\varphi_f^*(x, \cdot))$ e $B(0, c) \subset C(x)$. Allora, per ogni $\lambda \in \mathcal{M}(\Omega)$, vale*

$$\begin{aligned} & \sup \left\{ \int_{\Omega} u \, d\lambda - \int_{\Omega} f^*(x, u(x)) \mathcal{L}(dx) : u \in C_0(\Omega), u(x) \in C(x) \forall x \in \Omega \right\} \\ & = \int_{\Omega} [f(x, \cdot) \square \delta^*(\cdot | C(x))] \left(\frac{d\lambda}{d\mathcal{L}}(x) \right) \mathcal{L}(dx) + \int_{\Omega} \delta^* \left(\frac{d\lambda^s}{d|\lambda^s|}(x) | C(x) \right) |\lambda^s|(dx) \end{aligned}$$

Dimostrazione. Consideriamo la misura $m = \mathcal{L} + |\lambda^s|$ e poniamo

$$\begin{aligned} j(x, v) &= f^*(x, v) \frac{d\mathcal{L}}{dm}(x) \quad (x, v) \in \Omega \times \mathbb{R}^n; \\ J(u) &= \int_{\Omega} j(x, u(x)) m(dx) = \int_{\Omega} f^*(x, u(x)) \mathcal{L}(dx) \quad u \in L^\infty(\Omega, m; \mathbb{R}^n); \\ H &= \{u \in C_0(\Omega) : u(x) \in C(x) \forall x \in \Omega, J(u) < +\infty\}; \\ \Gamma(x) &= \overline{\{u(x) | u \in H\}} \quad x \in \Omega. \end{aligned}$$

Si noti intanto che J è a valori in $\mathbb{R} \cup \{+\infty\}$ perché si ha, grazie all'ipotesi H2,

$$J(u) = \int_{\Omega} f^*(x, u(x)) \mathcal{L}(dx) \geq \int_{\Omega} u(x) \cdot u_0(x) - f(x, u(x)) \mathcal{L}(dx) \in \mathbb{R}.$$

Inoltre, siccome $f^*(x, 0) = -\sup_v f(x, v) \leq 0$, vale $J(0) \leq 0$ e pertanto $J(0) \in \mathbb{R}$. Nel lemma successivo dimostreremo $\Gamma(x) = C(x)$ per ogni $x \in \Omega$; se per ora lo assumiamo vero possiamo applicare il teorema 2 di [7], di cui daremo l'enunciato, con $X = L^\infty(\Omega, m; \mathbb{R}^n)$, $Y = L^1(\Omega, m; \mathbb{R}^n)$, con il sottoinsieme $H \subset X$ e il funzionale J definiti prima e $\bar{u} = 0$. Ne deduciamo

$$(J + \delta(\cdot | H))^*(\omega) = \int_{\Omega} k(x, \omega(x)) m(dx) \quad \text{per ogni } \omega \in L^1(\Omega, m; \mathbb{R}^n),$$

dove $k(x, \cdot) = [j^*(x, \cdot) \square \delta^*(\cdot | C(x))]^{**}$. Tenendo conto del fatto che $\delta^*(\cdot | C(x))$ è coerciva per ogni x , perché $B(0, c) \subset C(x)$ e quindi $\delta^*(v | C(x)) \geq c|v|$, e del teorema A.9, la doppia coniugazione nella definizione di k risulta superflua; pertanto si verifica facilmente che

$$k(x, \cdot) = j^*(x, \cdot) \square \delta^*(\cdot | C(x)) = \begin{cases} f(x, \cdot) \square \delta^*(\cdot | C(x)) & \text{per } \mathcal{L} - \text{q.o. } x \in \Omega \\ \delta^*(\cdot | C(x)) & \text{per } |\lambda^s| - \text{q.o. } x \in \Omega. \end{cases}$$

Allora

$$\begin{aligned} & (J + \delta(\cdot|H))^* \left(\frac{d\lambda}{dm} \right) \\ &= \int_{\Omega} [f(x, \cdot) \square \delta^*(\cdot|C(x))] \left(\frac{d\lambda}{d\mathcal{L}} \right) \mathcal{L}(dx) + \int_{\Omega} \delta^* \left(\frac{d\lambda^s}{d|\lambda^s|}(x)|C(x) \right) |\lambda^s|(dx), \end{aligned}$$

che è equivalente alla tesi. \square

Teorema 2.10 (Theorem 2 in [7]). *Siano X, Y spazi vettoriali di funzioni misurabili su uno spazio metrico Ω munito della propria tribù borelliana e, su di essa, di una misura di Radon positiva m . Data $j : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ misurabile, convessa e s.c.i. a $x \in \Omega$ fissato, sia J il funzionale definito sullo spazio delle funzioni misurabili su Ω a valori in \mathbb{R}^n dato da*

$$J(u) = \int_{\Omega} j(x, u(x)) m(dx),$$

(posto uguale a $+\infty$ ogni qualvolta l'integrale della parte positiva dà risultato infinito). Sia $H \subset X$ un sottoinsieme stabile per combinazioni convesse tramite partizioni dell'unità (se $u_0, \dots, u_k \in H$ e $\alpha_0, \dots, \alpha_k$ è una partizione dell'unità con $\alpha_i \in C_c(\Omega)$, $i = 1..k$ allora $\sum_{i=0}^k \alpha_i u_i \in H$). Supponiamo che inoltre esista $\bar{u} \in H$ tale che $J(\bar{u}) \in \mathbb{R}$. Posto

$$\Gamma(x) = \overline{\{u(x) | u \in H \cap \text{dom}(J)\}},$$

nella dualità fra X e Y si ha

$$[J + \delta(\cdot|H)]^*(\omega) = \int_{\Omega} k(x, \omega(x)) m(dx),$$

dove $k(x, \cdot) = [j^*(x, \cdot) \square \delta^*(\cdot|\Gamma(x))]^{**}$.

Di questo risultato non diamo la dimostrazione perché non fa parte della teoria dei funzionali locali non convessi su spazi di misure. Diamo invece la dimostrazione del fatto $\Gamma(x) = C(x)$.

Lemma 2.11. *Nelle ipotesi e notazioni del lemma 2.9 vale $\Gamma(x) = C(x)$.*

Dimostrazione. Naturalmente vale l'inclusione $\Gamma(x) \subset C(x)$ perché C è chiuso, per cui dobbiamo mostrare solo quella opposta e, siccome C è un convesso chiuso comprendente la palla $B(0, c)$, quindi a parte interna non vuota, ci basterà dimostrare $\overset{\circ}{C}(x) \subset \Gamma(x)$. Consideriamo $v \in \overset{\circ}{C}(x)$ e troviamo una palla $B(v, r) \subset C(x)$. Grazie al fatto che la multifunzione C è s.c.i. e al lemma A.38 abbiamo un intorno N di x tale che, per ogni

$y \in N$, vale l'inclusione $B(v, \frac{r}{2}) \subset C(y)$. Utilizzando l'ipotesi $C(x) \subset \text{dom}(\varphi_f^*(x, \cdot))$ abbiamo, posto $\Theta(x) = \{u(x) | u \in C_0(\Omega), J(u) < +\infty\}$,

$$\begin{aligned} v \in \text{dom}(\overset{\circ}{\varphi}_f^*(x, \cdot)) &= \{w | w \cdot z \leq \overset{\circ}{\varphi}_f(x, z) \forall z\} \\ &= \{w | w \cdot z \leq \overset{\circ}{\delta}^*(z | \Theta(x)) \forall z\} \\ &= \overset{\circ}{\Theta(x)} \\ &= \Theta(x), \end{aligned}$$

dove si è usato che la condizione $w \in \bar{A}$ è equivalente a che per ogni $z \in \mathbb{R}^n$ valga $w \cdot z \leq \delta(z | A)$, fatto valido per A convesso (si può dimostrare artigianalmente usando Hahn-Banach, oppure vedere su [12]), nonché il fatto, sempre valido per A convesso e sempre verificabile grazie ad un appropriato uso di Hahn-Banach, $\overset{\circ}{\bar{A}} = \overset{\circ}{A}$.

Si trova in questo modo una funzione $u \in C_0(\Omega)$ per la quale $J(u) < +\infty$ e $u(x) = v$. Pertanto in un opportuno intorno N' di x , con $N' \subset N$, varrà $u(y) \in B(v, \frac{r}{2}) \subset C(y)$. Si può pertanto costruire una funzione $\tilde{u}(x) = t(x)u(x)$, con t funzione continua a valori in $[0, 1]$, a supporto compatto in N' , che valga 1 in un intorno $N'' \subset N'$. Varrà

$$\begin{aligned} J(\tilde{u}) &= \int_{\Omega} j(x, \tilde{u}(x))m(dx) \\ &\leq \int_{\Omega} [t(x)j(x, u(x)) + (1-t(x))j(x, 0)]m(dx) \\ &\leq J(u) + J(0) \in \mathbb{R}, \end{aligned}$$

inoltre per costruzione $\tilde{u}(x) \in C(x)$ per ogni $x \in \Omega$, perché per $x \notin N'$ si ha $\tilde{u}(x) = 0$, mentre per gli altri $\tilde{u}(x)$ è una combinazione convessa fra $u(x)$, che sapevamo appartenere a $C(x)$, e 0, che vi appartiene per ipotesi. In questo modo abbiamo $\tilde{u} \in H$ e quindi $v \in \Gamma(x)$. \square

Finalmente siamo in grado di dare una dimostrazione al teorema 2.4

Dimostrazione. (Teorema 2.4) Come abbiamo già fatto notare possiamo supporre che valgano le stime $f(x, v) \geq c|v|$, $\varphi(x, v) \geq c|v|$, $g(x, v) \geq c|v|$. Consideriamo una successione $(\lambda_h)_h$ con $\lambda_h \xrightarrow{*} \lambda$ e, per una certa costante M , avremo $\|\lambda_h\| \leq M$ per ogni h . Fissiamo $\varepsilon > 0$ e poniamo

$$A_h^\varepsilon = \{x \in \Omega \mid |\lambda_h(x)| \geq \varepsilon\}, \quad \lambda_h^{1, \varepsilon} = I_{(A_h^\varepsilon)^c} \cdot \lambda_h, \quad \lambda_h^{2, \varepsilon} = I_{A_h^\varepsilon} \cdot \lambda_h.$$

Per la proprietà di attività data dal fatto di essere locale si ha

$$F(\lambda_h) = F(\lambda_h^{1, \varepsilon}) + F(\lambda_h^{2, \varepsilon}).$$

Le misure $\lambda_h^{2,\varepsilon}$ appartengono a $\mathcal{M}_k^\sharp(\Omega)$ con $k = M/\varepsilon$. Naturalmente anche $\|\lambda_h^{1,\varepsilon}\|$ e $\|\lambda_h^{2,\varepsilon}\|$ si possono stimare con M , per cui si può supporre, a meno di passare a sottosuccessioni,

$$\lambda_h^{i,\varepsilon} \xrightarrow{*} \lambda^{i,\varepsilon}, \text{ per } i = 1, 2.$$

Naturalmente $\lambda^{1,\varepsilon} + \lambda^{2,\varepsilon} = \lambda$. Il lemma 2.3 ci dice che $\lambda^{2,\varepsilon} \in \mathcal{M}_k^\sharp(\Omega)$ e che

$$\liminf_h F(\lambda_h^{2,\varepsilon}) = \liminf_h G(\lambda_h^{2,\varepsilon}) \geq G(\lambda^{2,\varepsilon}) = F(\lambda^{2,\varepsilon}).$$

Consideriamo ora un'arbitraria funzione $u \in C_0(\Omega)$ tale che $u(x) \in C_\varepsilon(x)$ per ogni $x \in \Omega$. Con la notazione del lemma 2.8 abbiamo

$$C_\varepsilon(x) \subset G = \text{dom}(g_0(x, \cdot)) \subset \text{dom}(\varphi_f^*(x, \cdot)),$$

grazie anche all'ipotesi H6, che, dando una disuguaglianza fra g e φ_f la dà anche fra i limiti dei loro rapporti incrementali in 0, cioè fra g_0 e φ_f stessa, perché φ_f è omogenea. Sappiamo che valgono le condizioni

$$\begin{aligned} v \cdot u(x) &\leq \varphi_f(x, v) \text{ per ogni } (x, v); \\ v \cdot u(x) &\leq g(x, v) \text{ per ogni } (x, v) : |v| \leq \varepsilon. \end{aligned}$$

(la prima relazione può essere dedotta dalla definizione di φ_f se ci si limita alle u tali che $J(u) = \int_\Omega f^*(x, u(x))\mathcal{L}(dx) < +\infty$, cosa più che lecita per i nostri futuri scopi, altrimenti dal fatto $u(x) \in C_\varepsilon(x) \subset \text{dom}(\varphi_f^*(x, \cdot))$ e dalla caratterizzazione dei domini delle coniugate di funzioni omogenee; la seconda viene dalla definizione di C_ε). Tenendo conto che $|\lambda_h^{1,\varepsilon}(x)| \leq \varepsilon$ per ogni $x \in \Omega$, si ha

$$\begin{aligned} F(\lambda_h^{1,\varepsilon}) &= \int_\Omega f\left(x, \frac{d\lambda_h^{1,\varepsilon}}{d\mathcal{L}}(x)\right) \mathcal{L}(dx) \\ &\quad + \int_{\Omega \setminus A_{\lambda_h^{1,\varepsilon}}} \varphi\left(x, \frac{d(\lambda_h^{1,\varepsilon})^s}{d|(\lambda_h^{1,\varepsilon})^s|}(x)\right) |(\lambda_h^{1,\varepsilon})^s|(dx) + \int_\Omega g(x, \lambda_h^{1,\varepsilon}(x))\sharp(dx) \geq \\ &\int_\Omega \left[u(x) \cdot \frac{d\lambda_h^{1,\varepsilon}}{d\mathcal{L}}(x) - f^*(x, u(x)) \right] \mathcal{L}(dx) + \int_\Omega u d(\lambda_h^{1,\varepsilon})^s + \int_\Omega u(x) \cdot \lambda_h^{1,\varepsilon}(x)\sharp(dx) \\ &= \int_\Omega u d\lambda_h^{1,\varepsilon} - \int_\Omega f^*(x, u(x))\mathcal{L}(dx), \end{aligned}$$

(abbiamo usato l'ipotesi H7) e quindi

$$\liminf_h F(\lambda_h^{1,\varepsilon}) \geq \int_\Omega u d\lambda^{1,\varepsilon} - \int_\Omega f^*(x, u(x))\mathcal{L}(dx),$$

da cui, passando al sup in u e utilizzando il lemma 2.9, si ottiene

$$\liminf_h F(\lambda_h^{1,\varepsilon}) \geq \int_{\Omega} [f(x, \cdot) \square g_{\varepsilon}(x, \cdot)] \left(\frac{d\lambda^{1,\varepsilon}}{d\mathcal{L}}(x) \right) \mathcal{L}(dx) + \int_{\Omega} g_{\varepsilon} \left(x, \frac{d(\lambda^{1,\varepsilon})^s}{d|(\lambda^{1,\varepsilon})^s|}(x) \right) |(\lambda^{1,\varepsilon})^s|(dx).$$

Pertanto

$$\begin{aligned} \liminf_h F(\lambda_h) &\geq \int_{\Omega} [f(x, \cdot) \square g_{\varepsilon}(x, \cdot)] \left(\frac{d\lambda^{1,\varepsilon}}{d\mathcal{L}}(x) \right) \mathcal{L}(dx) \\ &\quad + \int_{\Omega} g_{\varepsilon} \left(x, \frac{d(\lambda^{1,\varepsilon})^s}{d|(\lambda^{1,\varepsilon})^s|}(x) \right) |(\lambda^{1,\varepsilon})^s|(dx) + \int_{\Omega} g(x, \lambda^{2,\varepsilon}(x)) \#(dx) \end{aligned}$$

Vogliamo passare al sup in ε : notiamo che la parte non atomica di $\lambda^{1,\varepsilon}$ è determinata univocamente (non dipende cioè da ε), perché coincide con quella di λ , pertanto, tenendo conto che g_{ε} converge crescendo a g_0 (lemma 2.8 e del teorema A.9, per convergenza monotona il primo addendo cresce a $\int_{\Omega} [f(x, \cdot) \square g_0(x, \cdot)] \left(\frac{d\lambda}{d\mathcal{L}}(x) \right) \mathcal{L}(dx)$. Tenendo conto delle ipotesi H3 e H7 si ha $g_0(x, \cdot) = f^{\infty}(x, \cdot)$ per \mathcal{L} -q.o. $x \in \Omega$, per cui si può sostituire g_0 con f^{∞} e tenere conto dell'esempio A.3, per avere $\int_{\Omega} f(x, \left(\frac{d\lambda}{d\mathcal{L}}(x) \right)) \mathcal{L}(dx)$.

Il secondo addendo invece si separa ulteriormente in due addendi, uno restringendo l'integrale agli atomi di $(\lambda^{1,\varepsilon})^s$, uno fuori da tali atomi. Quest'ultimo è dato da

$$\int_{\Omega \setminus A_{\lambda}} g_{\varepsilon} \left(x, \frac{d\lambda^s}{d|\lambda^s|}(x) \right) |\lambda^s|(dx) \nearrow \int_{\Omega \setminus A_{\lambda}} g_0 \left(x, \frac{d\lambda^s}{d|\lambda^s|}(x) \right) |\lambda^s|(dx).$$

Rimangono pertanto due addendi, che sono

$$\int_{\Omega} [g_{\varepsilon}(x, \lambda^{1,\varepsilon}(x)) + g(x, \lambda^{2,\varepsilon}(x))] \#(dx). \quad (2.4)$$

Detto semplicemente, per ogni $\varepsilon > 0$, gli atomi di λ sono suddivisi in due insiemi di atomi, eventualmente scindendone qualcuno, e ai primi si applica g_{ε} e ai secondi g . Possiamo naturalmente supporre, siccome $|\lambda^{i,\varepsilon}(x)| \leq M$ per ogni x e $i = 1, 2$, a meno del solito passaggio a sottosuccessioni, scegliendo una sottosuccessione con un tipico procedimento diagonale, che per ogni atomo x_j , $j \in \mathbb{N}$, si abbia

$$\lambda^{i,\varepsilon}(x_j) \rightarrow a_j^i, \quad j \in \mathbb{N}, i = 1, 2; \quad a_j^1 + a_j^2 = \lambda(x_j)$$

L'espressione in (2.4) si può stimare dal basso, per qualsiasi numero naturale k e $\varepsilon_0 > 0$ con

$$\sum_{j=0}^k g_{\varepsilon_0}(x_j, \lambda^{1,\varepsilon}(x_j)) + g(x_j, \lambda^{2,\varepsilon}(x_j)), \quad \text{se } \varepsilon < \varepsilon_0$$

grazie al fatto che, al diminuire di ε , le funzioni g_ε crescono. Passando ora al liminf in ε e usando la semicontinuità di g_{ε_0} , g si arriva a

$$\sum_{j=0}^k g_{\varepsilon_0}(x_j, a_j^1) + g(x, a_j^2),$$

poi, passando al sup in ε_0 , si ottiene

$$\sum_{j=0}^k g_0(x_j, a_j^1) + g(x, a_j^2) \geq \sum_{j=0}^k g(x_j, a_j^1) + g(x, a_j^2) \geq \sum_{j=0}^k g(x_j, \lambda(x_j)),$$

dove si è usato ancora $g_0 \geq g$ e la subadditività di g . Infine, col sup in k si ottiene $\sum_{j=0}^{\infty} g(x_j, \lambda(x_j))$. Abbiamo pertanto ottenuto

$$\begin{aligned} \liminf_h F(\lambda_h) &\geq \\ &\int_{\Omega} f\left(x, \left(\frac{d\lambda^{1,\varepsilon}}{d\mathcal{L}}(x)\right)\right) \mathcal{L}(dx) + \int_{\Omega \setminus A_\lambda} g_0\left(x, \frac{d\lambda^s}{d|\lambda^s|}(x)\right) |\lambda^s|(dx) + \sum_{j=0}^{\infty} g(x_j, \lambda(x_j)) \\ &= F(\lambda). \end{aligned}$$

L'ultima uguaglianza ha tenuto conto del fatto che, per l'ipotesi H7, si poteva sostituire g_0 con φ . Le ipotesi H4 e H5 sono state usate più volte, non solo in quest'ultima dimostrazione ma anche nei lemmi precedenti, mentre le ipotesi H1 e H2 servivano per poter applicare il lemma 2.9. □

2.3 Caratterizzazione dei funzionali locali semicontinui

Per completezza enunciamo anche il risultato contenuto in [6], che insieme a quello appena dimostrato costituisce una sorta di caratterizzazione dei funzionali definiti su $\mathcal{M}(\Omega)$ che siano al tempo stesso locali e debolmente* semicontinui inferiormente.

Teorema 2.12. *Sia $F : \mathcal{M}(\Omega) \rightarrow [0, +\infty]$ un funzionale locale e s.c.i.. Allora esiste una misura di Radon positiva \mathcal{L} non atomica su Ω e tre funzioni di Borel $f, \varphi, g : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$ tali che*

K1) $f(x, \cdot)$ è convessa s.c.i. e $f(x, 0) = 0$ per \mathcal{L} -q.o. $x \in \Omega$;

K2) le funzioni f^∞ e φ_f coincidono per \mathcal{L} -q.o. $x \in \Omega$;

K3) g e g^∞ sono s.c.i. su $\Omega \times \mathbb{R}^n$ e $g(x, 0) = 0$ per ogni $x \in \Omega$;

K4) in $\Omega \times \mathbb{R}^n$ valgono le disuguaglianze $g^\infty \leq \varphi_f$ e $g^\infty \leq \hat{g}$;

K5) esiste un sottoinsieme numerabile $N \subset \Omega$ tale che $g_0 = \varphi = \varphi_f$ in $(\Omega \setminus N) \times \mathbb{R}^n$;

e tali che valga questa rappresentazione integrale

$$F(\lambda) = \int_{\Omega} f \left(x, \frac{d\lambda}{d\mathcal{L}}(x) \right) \mathcal{L}(dx) + \int_{\Omega \setminus A_\lambda} \varphi \left(x, \frac{d\lambda^s}{d|\lambda^s|}(x) \right) |\lambda^s|(dx) + \int_{A_\lambda} g(x, \lambda(x)) \#(dx).$$

Questo enunciato necessita di alcuni chiarimenti. Data una funzione $g : \Omega \times \mathbb{R}^n \rightarrow [0, +\infty]$ abbiamo indicato con \hat{g} la funzione

$$\hat{g}(x, v) = \liminf_{\substack{(y, w) \rightarrow (x, v) \\ y \neq x}} g(y, w).$$

Inoltre abbiamo usato la funzione g^∞ . La definizione data nell'appendice è quella più comoda per funzioni convesse, ma non è quella utilizzata nel caso generale. Data una qualunque funzione $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, di solito si definisce

$$g^\infty(v) = \sup_{w \in \text{dom}(g)} [g(v+w) - g(w)].$$

L'equivalenza nel caso convesso con l'altra definizione può essere trovata su [12], ottenendo così anche una dimostrazione dell'indipendenza dal punto $x_0 \in \text{dom}(g)$ della definizione nell'appendice. Si verifica facilmente che g è subadditiva se e solo se $g^\infty \leq g$ e si può notare che la g la cui esistenza è invocata in questo teorema verifica solo $g^\infty \leq \hat{g}$, ma $\hat{g} \geq g$. Infatti non è detto che tale g risulti subadditiva a x fissato per ogni x , ma si dimostra che l'insieme D degli x su cui tale subadditività non è verificata è al più numerabile. Inoltre si può notare che, pur essendo molto simili alle ipotesi H1, ..., H7, le proprietà del teorema 2.12 non sono sufficienti per la semicontinuità. È però dimostrato nello stesso lavoro che se si aggiunge l'ipotesi che D sia privo di punti di accumulazione allora il funzionale F definito nel solito modo diventa debolmente* s.c.i..

Capitolo 3

Problemi di concentrazioni ottimali

3.1 Presentazione dei funzionali

Come già anticipato nell'introduzione, dove sono anche state esposte alcune possibili interpretazioni ed applicazioni dei concetti che ora affronteremo, lo scopo principale di questo lavoro sarà lo studio di problemi di minimo che permettano di scegliere in modo ottimo coppie di misure, delle quali si desidera che la prima sia molto "sparsa", che la seconda molto "concentrata" e che la loro distanza misurata in termini di costi di trasporto dall'una all'altra sia piccola.

Avremo pertanto assegnati uno spazio metrico localmente compatto e separabile (Ω, d) , una misura di Radon positiva \mathcal{L} su $(\Omega, \mathcal{B}(\Omega))$ e un funzionale definito sullo spazio $\mathcal{P}(\Omega)^2$, della forma

$$\mathfrak{F}(\mu, \nu) = T_c(\mu, \nu) + F(\mu) + G(\nu) \in [0, +\infty].$$

- $T_c(\mu, \nu)$ è il costo di un trasporto ottimo fra le due misure μ, ν relativamente ad una assegnata funzione s.c.i. $c : \Omega \times \Omega \rightarrow [0, +\infty]$, cioè

$$T_c(\mu, \nu) = \inf \left\{ \int c d\gamma \mid \gamma \text{ piano di trasporto fra } \mu \text{ e } \nu \right\}.$$

Sceghieremo generalmente $c(x, y)$ uguale ad una funzione crescente di $d(x, y)$. Qualora sia $c(x, y) = d(x, y)^p$ scriveremo T_p in luogo di T_c ; si noti che, se $\mu, \nu \in W_p$ vale $T_p(\mu, \nu) = \delta_p(\mu, \nu)^p$.

- F sarà un funzionale che penalizza nella minimizzazione le misure troppo concentrate rispetto alla misura di riferimento \mathcal{L} e lo sceghieremo della forma

$$F(\mu) = \int_{\Omega} f(\mu^a) d\mathcal{L} + K\mu^s(\Omega), \quad (3.1)$$

dove μ^a indica la densità della parte assolutamente continua di μ rispetto ad \mathcal{L} e μ^s è invece la parte singolare. perché F sia semicontinuo inferiormente rispetto alla convergenza debole delle misure su Ω nella dualità con le funzioni continue a supporto compatto, richiederemo

$$f : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \text{ convessa , } \quad \lim_{x \rightarrow +\infty} \frac{f(x)}{x} = K \in \mathbb{R} \cup \{+\infty\}.$$

Chiederemo inoltre $f(0) = f'(0) = 0$ (queste due condizioni non fanno perdere di generalità perché, in effetti, data una f convessa qualunque, sottraendole un termine polinomiale di grado uno, ci si riconduce a questo caso senza cambiare alcunché nella minimizzazione del funzionale, tranne nel caso $f(0) > 0, \mathcal{L}(\Omega) = +\infty$, in cui però $F \equiv +\infty$); in questo modo F risulta essere la restrizione alle misure di probabilità di un funzionale locale definito sulle misure reali su Ω (il termine in cui compare K è quello di recessione), per il quale vale la teoria esposta nel capitolo precedente, con particolare riguardo alla semicontinuità. Noi sceglieremo perlopiù di restringerci al caso in cui la funzione f è superlineare ($K = +\infty$), rendendo pertanto finito il valore del funzionale F solo sulle misure di probabilità assolutamente continue. Si noti che in tutti questi casi, se $\mathcal{L}(\Omega) < +\infty$, il minimo del funzionale F è realizzato (come si può facilmente vedere grazie alla disuguaglianza di Jensen) dalla misura $\mu = \frac{1}{\mathcal{L}(\Omega)} \cdot \mathcal{L}$, che per noi sarà la più sparsa in assoluto.

- La scelta del funzionale G , che deve costringere la misura ν a concentrarsi, è più delicata e dipende dal significato che si dà a questo concetto. Noi ci limiteremo ad una particolare classe di funzionali, che hanno l'effetto di favorire il fatto che ν concentri massa in singoli punti di Ω e tenere conto di quanta massa vi è in ciascuno di essi, ma non delle distanze reciproche. Costruiremo tali funzionali a partire da una funzione $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ subadditiva e s.c.i., con $g(0) = 0$:

$$G(\nu) = \sum_{x \in A_\nu} g(\nu(x)) + L\nu(\Omega \setminus A_\nu) \quad (3.2)$$

dove $\nu(x)$ è un abuso di notazione per $\nu(\{x\})$. Richiederemo

$$\lim_{x \rightarrow 0^+} \frac{g(x)}{x} = L \in \mathbb{R} \cup \{+\infty\}.$$

Anche in questo caso analizzeremo in particolare il caso $L = +\infty$. Si noti che qui il minimo del funzionale G è realizzato da una qualunque misura di Dirac in un qualunque punto di Ω e queste misure saranno per noi le più concentrate in assoluto.

Osservazione 3.1. Il funzionale G definito da (3.2) è locale e semicontinuo.

Per vederlo basta considerare la g , funzione subadditiva, in coppia con una funzione convessa s.c.i. ad essa compatibile (nel senso che le si possa applicare il corollario 2.1) e tale che il funzionale associato ad essa e alla g dalla teoria esposta nel capitolo precedente sia proprio G . Questo si ottiene scegliendo come funzione convessa $x \mapsto Lx$ (se $L + \infty$ si sceglie $x \mapsto \delta(x | \{0\})$). Si noti che queste funzioni sono le uniche funzioni convesse nulle in 0 a valori in $[0, +\infty]$ che siano anche concave. Questa scelta, se si vogliono utilizzare funzionali locali per imporre concentrazione, è obbligata, perché da un lato la convessità serve per garantire la semicontinuit , dall'altro essa   da evitare (anzi, il suo opposto, la concavit ,   meglio) se si vogliono penalizzare le misure sparse: infatti la convessit  favorisce nella minimizzazione le misture fra misure differenti, il che porta a dispersione della massa ed   pertanto contrario al nostro obiettivo.

Osservazione 3.2. Il funzionale $(\mu, \nu) \mapsto F(\mu) + G(\nu)$ pu  essere visto come un unico funzionale locale definito su misure a valori in \mathbb{R}^2 , con una parte convessa che dipende solo dalla prima componente e una subadditiva solo dalla seconda.

Esempio 3.1. Per ogni $q > 1$ e $r \in (0, 1)$ si pu  considerare il funzionale

$$(\mu, \nu) \mapsto \begin{cases} T_p(\mu, \nu) + \int (\mu^a)^q d\mathcal{L} + \sum_{x \in A_\nu} (\nu(x))^r & \text{se } \mu \in L^q(\mathcal{L}), \nu(\Omega \setminus A_\nu) = 0 \\ +\infty & \text{altrimenti,} \end{cases}$$

dove con $\mu \in L^q(\mathcal{L})$ si intende $\mu \ll \mathcal{L}, \mu^a \in L^q(\mathcal{L})$.

3.2 Esistenza tramite il metodo diretto

Per poter esporre i primi risultati di esistenza, premettiamo alcune facili considerazioni sulla convergenza debole di misure di probabilit .

Proposizione 3.1. *Se $\mu, \mu_h \in \mathcal{P}(\Omega)$ allora sono equivalenti*

1. $\mu_h \rightharpoonup \mu$ per $h \rightarrow +\infty$ (nella dualit  con le funzioni $C_b(\Omega)$);
2. $\mu_h \xrightarrow{*} \mu$ per $h \rightarrow +\infty$ (nella dualit  con le funzioni $C_c(\Omega)$).

Dimostrazione. (1 \Rightarrow 2)   banale perch  $C_c \hookrightarrow C_b$.

(2 \Rightarrow 1) Sia χ una funzione continua e limitata dalle costanti $-L, L$. Allora le funzioni $\chi + L$ e $L - \chi$ sono continue non negative e possono quindi essere espresse come estremo superiore di funzioni C_c . Questo si vede moltiplicandole per opportune funzioni di Urysohn costruite a partire da una successione di compatti che invada Ω . Pertanto le valutazioni degli integrali di $\chi + L$ e $L - \chi$ sono applicazioni s.c.i. rispetto alla convergenza

debole nella dualità con C_c (estremo superiore di continue). Allora, se $\mu_h \xrightarrow{*} \mu$, vale

$$\begin{aligned} \int_{\Omega} (\chi + L) d\mu &\leq \liminf_h \int_{\Omega} (\chi + L) d\mu_h, \\ \int_{\Omega} (L - \chi) d\mu &\leq \liminf_h \int_{\Omega} (L - \chi) d\mu_h, \end{aligned}$$

da cui, tenendo conto che μ_h, μ sono tutte misure con la stessa massa, segue la tesi. Pertanto per misure di probabilità non vi sarà ambiguità nell'espressione "convergenza debole". \square

Proposizione 3.2. *Se $\mu_h, \mu \in \mathcal{P}(\Omega)$ e $\mu_h \rightarrow \mu$, allora la successione $(\mu_h)_h$ è tesa.*

Dimostrazione. Supponiamo per assurdo che non lo sia: allora esisterebbe un numero $\varepsilon > 0$ tale che, per ogni compatto $K \subset \Omega$, esista una sottosuccessione $(\mu_{h_k})_k$ tale che $\mu_{h_k}(\Omega \setminus K) \geq \varepsilon$. Prendendo due arbitrari compatti K_1, K_2 , con K_1 contenuto nella parte interna di K_2 , e costruendo una funzione continua χ , nulla su K_1 e costantemente 1 fuori di K_2 , si avrebbe allora, scegliendo la sottosuccessione $(\mu_{h_k})_k$ in relazione al compatto K_2 ,

$$\varepsilon \leq \int_{\Omega} \chi d\mu_{h_k} \rightarrow \int_{\Omega} \chi d\mu \leq \mu(\Omega \setminus K_1),$$

ma questo è assurdo perché al variare di K_1 $\mu(\Omega \setminus K_1)$ può essere resa arbitrariamente piccola. \square

Proposizione 3.3. *Se $\mu_h \xrightarrow{*} \mu$ e le misure μ_h sono misure di probabilità allora sono equivalenti:*

1. μ è una misura di probabilità;
2. $\mu_h \rightarrow \mu$ (anche nella dualità con C_b).

Dimostrazione. (1 \Rightarrow 2) è già stato mostrato con la parte 2 \Rightarrow 1 dell'osservazione 3.1. (2 \Rightarrow 1) per vedere che μ è positiva basta testare con qualsiasi funzione continua positiva, eventualmente anche a supporto compatto (qui non serve la convergenza contro C_b), per vedere che la massa di μ è unitaria basta testare contro la funzione costantemente 1 (che non è in generale a supporto compatto). \square

Corollario 3.4. *Lo spazio $\mathcal{P}(\Omega)$ è compatto se Ω lo è.*

Dimostrazione. La relativa compattezza delle misure di probabilità nelle misure di Radon a valori reali, munite della convergenza debole nella dualità con C_c è nota e viene dal teorema di Alaoglu, ma in un compatto $C_c = C_b = C$ e il teorema precedente ci assicura anche la chiusura di $\mathcal{P}(\Omega) \subset \mathcal{M}(\Omega)$. \square

Lemma 3.5. *Il funzionale $T_c : \mathcal{P}(\Omega)^2 \rightarrow \mathbb{R}^+$ è semicontinuo inferiormente rispetto alla convergenza debole.*

Dimostrazione. Ci basterà dimostrare la tesi nel caso che il costo c sia continuo e limitato, perché se è solo s.c.i. e non negativo possiamo scriverlo come estremo superiore di funzioni lipschitziane e limitate (vedi 1.1) e poi usare il fatto che il sup di funzionali s.c.i. è anch'esso s.c.i..

Supponiamo $\mu_h \rightharpoonup \mu$, $\nu_h \rightharpoonup \nu$ e consideriamo dei piani di trasporto ottimi γ_h fra μ_h e ν_h rispetto al costo c . Mostriamo che la successione di misure $(\gamma_h)_h$ è tesa: siccome $(\mu_h)_h, (\nu_h)_h$ convergono, esse lo sono; fissato $\varepsilon > 0$ esistono allora dei compatti $K^X, K^Y \subset \Omega$, tali che, per ogni h ,

$$\mu_h(\Omega \setminus K^X) < \frac{\varepsilon}{2}, \quad \nu_h(\Omega \setminus K^Y) < \frac{\varepsilon}{2}.$$

Siccome le misure marginali di γ_h sono proprio μ_h, ν_h , per ogni h vale allora

$$\gamma_h((\Omega \times \Omega) \setminus (K^X \times K^Y)) < \varepsilon.$$

Possiamo quindi estrarre una sottosuccessione γ_{h_k} convergente debolmente ad un piano di trasporto γ fra μ e ν (perché le misure marginali passano al limite per la convergenza debole nella dualità con C_b). Vale

$$T_c(\mu, \nu) \leq \int_{\Omega \times \Omega} c d\gamma = \lim_k \int_{\Omega \times \Omega} c d\gamma_{h_k} = \lim_k T_c(\mu_{h_k}, \nu_{h_k}).$$

La tesi segue facilmente per assurdo, perché ci si può ridurre a considerare una successione tale che valga $\lim_h T_c(\mu_h, \nu_h) < T_c(\mu, \nu)$. \square

A questo punto è naturale aspettarsi l'enunciato del seguente teorema di esistenza.

Teorema 3.6. *Se lo spazio metrico Ω è compatto allora il funzionale \mathfrak{F} , con le scelte di F e G esposte precedentemente, ammette minimo.*

Dimostrazione. Basta applicare il metodo diretto del calcolo delle variazioni allo spazio sequenzialmente compatto $\mathcal{P}(\Omega)^2$, avendo appena mostrato la sequenziale semicontinuità inferiore dell'unico addendo di \mathfrak{F} del quale non fosse già nota dalla teoria sui funzionali locali. \square

3.3 Un risultato di esistenza senza compattezza

In questa sezione ci occuperemo di un caso molto particolare in cui manchi la compattezza dello spazio ambiente: quello in cui lo spazio Ω è un sottoinsieme convesso limitato di \mathbb{R}^n e \mathcal{L} è la misura di Lebesgue su di esso. Anche la scelta del funzionale \mathfrak{F} sarà piuttosto ristretta, soprattutto per i primi due addendi. Per il funzionale G richiederemo invece molto poco, precisamente che sia s.c.i. e che favorisca la concentrazione.

Definizione 3.1. Diremo che un funzionale $G : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ favorisce la concentrazione se vale

$$G(p_{\#}\nu) \leq G(\nu),$$

ogniqualevolta $p : \Omega \rightarrow \Omega$ sia una mappa 1-lipschitziana.

Osservazione 3.3. Il funzionale G definito dalla (3.2) favorisce la concentrazione, anzi vale $G(p_{\#}\nu) \leq G(\nu)$ per ogni $p : \Omega \rightarrow \Omega$ misurabile.

Esempio 3.2. Sia $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ crescente e poniamo $\varphi(x, y) = \phi(d(x, y))$. Il funzionale

$$\Phi(\nu) = \int_{\Omega \times \Omega} \varphi d\nu \otimes \nu$$

è semicontinuo inferiormente e favorisce la concentrazione.

Per vedere che Φ favorisce la concentrazione basta osservare che, se p è 1-lipschitziana, vale

$$\Phi(p_{\#}\nu) = \int_{\Omega \times \Omega} \varphi d p_{\#}\nu \otimes p_{\#}\nu = \int_{\Omega \times \Omega} \phi(d(p(x), p(y))) d\nu \otimes \nu \leq \int_{\Omega \times \Omega} \phi(d(x, y)) d\nu \otimes \nu = \Phi(\nu).$$

Per la semicontinuità invece si usi il seguente:

Lemma 3.7. *Il funzionale $\mathcal{P}(\Omega) \ni \nu \mapsto \int_{\Omega \times \Omega} \varphi d\nu \otimes \nu \in \mathbb{R}^+$ è semicontinuo inferiormente rispetto alla convergenza debole ogni qualvolta $\varphi : \Omega \times \Omega \rightarrow \mathbb{R}^+$ sia s.c.i..*

Dimostrazione. Come nel lemma 3.5 possiamo limitarci al caso φ continua e limitata e in questo caso mostreremo addirittura che abbiamo continuità. Per farlo è sufficiente mostrare che da $\nu_h \rightharpoonup \nu$ segue $\nu_h \otimes \nu_h \rightharpoonup \nu \otimes \nu$ nella dualità con $C_b(\Omega \times \Omega)$. Ci è sufficiente tuttavia mostrare la convergenza debole nella dualità con C_c , grazie al fatto che $\nu \otimes \nu$ è una misura di probabilità, e pertanto possiamo considerare una qualunque $\chi \in C_c(\Omega \times \Omega)$ e mostrare

$$\lim_h \int_{\Omega \times \Omega} \chi d\nu_h \otimes \nu_h = \int_{\Omega \times \Omega} \chi d\nu \otimes \nu. \quad (3.3)$$

poiché però le misure $\mu_h \otimes \nu_h$ sono limitate in norma (sono tutte di probabilità) ci è sufficiente mostrare la (3.3) per particolari scelte di χ , purché le χ considerate costituiscano un denso di $C_c(\Omega \times \Omega)$. Grazie al teorema di Stone-Weierstrass si vede che quelle della forma $\chi(x, y) = \chi_1(x)\chi_2(y)$ vanno bene e per loro la relazione (3.3) discende immediatamente dalla convergenza debole della successione $(\nu_h)_h$. \square

Come conseguenza il funzionale Φ potrebbe sostituire G per i risultati di esistenza sia nel caso di Ω compatto, che nel caso di Ω convesso e limitato in \mathbb{R}^n che ora esporremo.

Osservazione 3.4. Se un funzionale G favorisce la concentrazione esso ammette minimo assoluto realizzato dalle misure di Dirac concentrate nei vari punti di Ω .

L'enunciato che dimostreremo richiederà le seguenti ipotesi:

- Ω è un sottoinsieme convesso di \mathbb{R}^n tale che, per opportuni raggi $0 < r < R$,

$$B(0, r) \subset \Omega \subset B(0, R)$$

e $\mathcal{L} = I_\Omega \cdot \mathcal{L}^n$;

- F è definito dalla (3.1) tramite una funzione f superlineare;
- G è la restrizione a $\mathcal{P}(\Omega)$ di un funzionale, che chiameremo ancora G , debolmente s.c.i. su $\mathcal{P}(\overline{\Omega})$ che favorisca la concentrazione;
- $c(x, y) = \phi(|x - y|)$ con $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ crescente, lipschitziana di una certa costante L^+ su $[0, 2R]$ e tale che

$$\forall a > 0 \exists L^-(a) > 0 : \forall a \leq t < s \leq 2R \quad \phi(s) \geq \phi(t) + L^-(a)(s - t). \quad (3.4)$$

Osservazione 3.5. Le funzioni $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ convesse ma non costanti in un intorno di 0 soddisfano la condizione (3.4), con $L^-(a) = \phi'(a)$ (nel caso che non esista la derivata basta quella destra).

Per dimostrare in queste ipotesi l'esistenza di un minimo per \mathfrak{F} ci basterà dimostrare che, in una qualunque coppia (μ, ν) che realizzi il minimo nell'ovvia estensione di \mathfrak{F} alle misure di probabilità su $\overline{\Omega}$, nessuna delle due misure assegna massa a $\partial\Omega$.

Richiamiamo ora la definizione della funzione *jauge* di un convesso, che ci sarà utile successivamente.

Definizione 3.2. Dato un convesso $C \subset \mathbb{R}^n$ definiamo la sua funzione *jauge* tramite

$$j(y) = \inf \left\{ \alpha \mid \frac{y}{\alpha} \in C \right\}.$$

La funzione j risulta evidentemente positivamente 1-omogenea, ma anche subadditiva (quindi convessa), e si può verificare che $ri(C) = \{y \mid j(y) < 1\}$, $\overline{C} = \{y \mid j(y) \leq 1\}$ (si veda ad esempio [8]). Nel caso poi che $B(0, r) \subset C \subset B(0, R)$ varrà anche

$$\frac{|y|}{R} \leq j(y) \leq \frac{|y|}{r},$$

pertanto j risulterà a valori reali e quindi continua.

Lemma 3.8. *Se $\Omega \subset \mathbb{R}^n$ è convesso allora $|\partial\Omega| = 0$.*

Dimostrazione. $\partial\Omega = \overline{\Omega} \setminus \overset{\circ}{\Omega} = \{j = 1\}$.

$$\begin{aligned} |\{j = 1\}| &= |\{j \leq 1\}| - |\{j < 1\}| = |\{j \leq 1\}| - \left| \bigcup_{l>0} \{j \leq 1 - l\} \right| \\ &= |\{j \leq 1\}| - \sup_{l>0} (1 - l)^n |\{j \leq 1\}| \\ &= |\{j \leq 1\}| (1 - \sup_{l>0} (1 - l)^n) = 0 \quad \square \end{aligned}$$

Corollario 3.9. *La misura μ non assegna massa a $\partial\Omega$.*

Dimostrazione. È sufficiente notare che $\mu \ll \mathcal{L}^n$ e applicare il lemma 3.8. □

Teorema 3.10. *Nelle ipotesi precedenti neanche ν assegna massa a $\partial\Omega$.*

Dimostrazione. Consideriamo tre numeri $0 < l_1 < l_2 < l_3 < 1$ e poniamo, per $i = 1, 2, 3$,

$$\Omega_i = (1 - l_i)\overline{\Omega}; \quad \Omega_3 \subset \Omega_2 \subset \Omega_1 \subset \overset{\circ}{\Omega} \subset \Omega \subset \overline{\Omega}.$$

Sia p la proiezione sul convesso chiuso Ω_2 : p è 1-lipschitziana, pertanto, se sostituiamo (μ, ν) con $(\mu, p\# \nu)$, il termine F resta invariato e G non aumenta. Ci resta solo da stimare la variazione di T_c . Se t è un trasporto ottimo fra μ e ν (che esiste perché $\mu \ll \mathcal{L}^n$) certamente si avrà

$$T_c(\mu, p\# \nu) \leq \int_{\Omega} \phi(d(x, p(t(x)))) \mu(dx).$$

Dobbiamo allora stimare l'incremento di $\phi(d(x, p(t(x))))$ rispetto a $\phi(d(x, t(x)))$, tenendo conto delle ipotesi quantitative sulla monotonia crescente di ϕ e differenziando a seconda delle posizioni di x e $t(x) = y$. Questa tabella riassume le situazioni in cui si abbia un guadagno (diminuzione di ϕ) o una perdita (aumento).

$x \downarrow, y \rightarrow$	Ω_3	$\Omega_2 \setminus \Omega_3$	$\Omega_1 \setminus \Omega_2$	$\overline{\Omega} \setminus \Omega_1$
Ω_3	=	=	guadagno	guadagno*
$\Omega_2 \setminus \Omega_3$	=	=	guadagno	guadagno
$\Omega_1 \setminus \Omega_2$	=	=	perdita	perdita
$\overline{\Omega} \setminus \Omega_1$	=	=	perdita	perdita

Tabella 3.1: Guadagni e perdite

Con guadagno* si intende un guadagno che verrà quantificato, mentre gli altri saranno pessimisticamente stimati con zero. Si è tenuto conto della definizione di p , che su Ω_2 è l'identità, e della sua caratterizzazione tramite la disequazione variazionale

$$(x - p(y)) \cdot (p(y) - y) \geq 0 \quad \forall y \in \overline{\Omega}, x \in \Omega_2.$$

Questa, insieme alla semplice disuguaglianza $|b| \geq |a| + |a|^{-1}(b-a) \cdot a$ valida per ogni coppia di vettori a, b , dà, per $x \in \Omega_2, y \notin \Omega_2$,

$$|x - y| \geq |x - p(y)| + \frac{(x - p(y)) \cdot (p(y) - y)}{|x - p(y)|} \geq |x - p(y)|. \quad (3.5)$$

Dobbiamo ora, nei casi significativi, paragonare $|x - p(y)|$ a $|x - y|$. Stimiamo la perdita, con $x, y \notin \Omega_2$:

$$\begin{aligned} |x - p(y)| &\leq |x - y| + |y - p(y)| \leq |x - y| + \left| y - \frac{y}{j(y)}(1 - l_2) \right| \\ &= |x - y| + \frac{|y|}{j(y)} |j(y) - 1 + l_2| \leq Rl_2, \end{aligned}$$

che è stata ottenuta tenendo conto che $j(y)^{-1}y(1 - l_2) \in \Omega_2$, che vale la disuguaglianza $1 - l_2 \leq j(y) \leq 1$, così come $j(y)^{-1}y \leq R$. Pertanto la perdita puntuale si può stimare con L^+Rl_2 , e quindi

$$\text{perdita totale} \leq L^+Rl_2\mu((\bar{\Omega} \setminus \Omega_2) \cap p^{-1}(\bar{\Omega} \setminus \Omega_2)) \leq L^+Rl_2\mu(\bar{\Omega} \setminus \Omega_2).$$

Stimiamo ora dal basso il guadagno*, migliorando la (3.5). Consideriamo due punti $x \in \Omega_3, y \in \bar{\Omega} \setminus \Omega_1$: vale allora

$$|x - y| \geq |x - p(y)| + \frac{(x - p(y)) \cdot (p(y) - y)}{|x - p(y)|} \geq |x - p(y)| + \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1),$$

dove l'ultimo passaggio è stato giustificato da $|x - p(y)| \leq 2R$ e dal successivo lemma 3.11. Allora si ha la stima puntuale

$$\phi(|x - y|) \geq \phi(|x - p(y)|) + L^-(a) \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1),$$

con $a = r(l_3 - l_2)$, in quanto il lemma 3.11 ci assicura anche che $|x - p(y)| \geq r(l_3 - l_2)$. Pertanto

$$\begin{aligned} \text{guadagno totale} &\geq L^-(r(l_3 - l_2)) \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1) \mu(\Omega_3 \cap p^{-1}(\bar{\Omega} - \Omega_1)) \\ &\geq L^-(r(l_3 - l_2)) \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1) [1 - \mu(\Omega_3^c) - \nu(\Omega_1)] \\ &= L^-(r(l_3 - l_2)) \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1) [\nu(\bar{\Omega} \setminus \Omega_1) - \mu(\bar{\Omega} \setminus \Omega_3)] \\ &\geq L^-(r(l_3 - l_2)) \frac{r^2}{2R}(l_3 - l_2)(l_2 - l_1) [\nu(\partial\Omega) - \mu(\bar{\Omega} \setminus \Omega_3)]. \end{aligned}$$

Supponiamo ora per assurdo $\nu(\partial\Omega) > 0$ e scegliamo l_3 abbastanza piccolo in modo tale che $\nu(\partial\Omega) - \mu(\bar{\Omega} \setminus \Omega_3) = \delta > 0$, cosa che si può fare perché, in virtù del corollario 3.9, $\lim_{l_3 \rightarrow 0^+} \mu(\bar{\Omega} \setminus \Omega_3) = 0$. Siccome μ, ν erano ottime il guadagno totale deve

necessariamente essere maggiorato dalla perdita totale, quindi

$$L^-(r(l_3 - l_2)) \frac{r^2}{2R} (l_3 - l_2)(l_2 - l_1)\delta \leq L^+ R l_2 \mu((\bar{\Omega} \setminus \Omega_2)).$$

Allora, scegliendo $l_2 = l$, $l_1 = \frac{l}{2}$ e $l < \frac{l_3}{2}$, si ha

$$C_1 l \leq C_2 l \mu((\bar{\Omega} \setminus \Omega_2)),$$

per opportune costanti C_1, C_2 positive. Semplificando l e mandandolo a 0 si trova un assurdo. \square

Lemma 3.11. *Se $x \in \Omega_3$ e $y \in \bar{\Omega} \setminus \Omega_1$ valgono le stime*

$$\begin{aligned} (x - p(y)) \cdot (p(y) - y) &\geq r^2(l_3 - l_2)(l_2 - l_1) \\ |x - p(y)| &\geq r(l_3 - l_2) \end{aligned}$$

Dimostrazione. Si noti intanto che, grazie alla subadditività di j ,

$$j(x + w) \leq j(x) + j(w) \leq (1 - l_3) + \frac{|w|}{r}.$$

Se $\frac{|w|}{r} < (1 - l_2) - (1 - l_3) = l_3 - l_2$, $x + w$ appartiene allora a Ω_2 , e pertanto, per la caratterizzazione variazionale di p ,

$$(x + w - p(y)) \cdot (p(y) - y) \geq 0,$$

da cui $(x - p(y)) \cdot (p(y) - y) \geq w \cdot (p(y) - y)$ per ogni w tale che $|w| \leq r(l_3 - l_2)$. Allora $(x - p(y)) \cdot (p(y) - y) \geq |y - p(y)|r(l_3 - l_2)$. Stimando dall'alto il prodotto scalare col prodotto delle norme e semplificando si ottiene la seconda parte della tesi, per ottenere la prima basta invece stimare dal basso $|y - p(y)|$, che si può fare tramite la seguente catena di disuguaglianze:

$$|y - p(y)| \geq r j(y - p(y)) \geq r [j(y) - j(p(y))] \geq r [(1 - l_1) - (1 - l_2)] = r(l_2 - l_1),$$

in cui si è tenuto conto del fatto che per $y \notin \Omega_2$ vale $p(y) \in \partial\Omega_2$ e quindi $j(p(y)) = 1 - l_2$. \square

Un altro risultato di esistenza senza compattezza dello spazio ambiente che esporremo più avanti sarà con $\Omega = \mathbb{R}^n$. Tuttavia, per la dimostrazione che intendiamo dare, in cui peraltro dovremo tornare a ridurci a G della forma data da (3.2) con $g'(0) = +\infty$, aspetteremo di avere a disposizione alcuni risultati relativi alle condizioni necessarie di ottimalità e quindi attenderemo il termine del prossimo capitolo.

Capitolo 4

Condizioni necessarie di ottimalità

Ci occupiamo in questo capitolo delle proprietà di cui necessariamente deve godere una coppia di misure (μ, ν) che risulti minimizzante per il funzionale \mathfrak{F} .

È evidente che, se (μ, ν) minimizza \mathfrak{F} , allora μ minimizza \mathfrak{F}_ν e ν viceversa minimizza \mathfrak{F}^μ , dove questi funzionali sono definiti tramite

$$\begin{aligned}\mathfrak{F}_\nu(\mu) &= \mathfrak{F}(\mu, \nu), \\ \mathfrak{F}^\mu(\nu) &= \mathfrak{F}(\mu, \nu).\end{aligned}$$

In maniera equivalente dal punto di vista della minimizzazione, si può pensare ai funzionali

$$\begin{aligned}\mu &\longmapsto T_c(\mu, \nu) + F(\mu), \\ \nu &\longmapsto T_c(\mu, \nu) + G(\nu).\end{aligned}$$

Il capitolo presenterà quindi una prima sezione in cui si studieranno le proprietà dei minimi di \mathfrak{F}_ν e una seconda dedicata a \mathfrak{F}^μ , mentre la terza cercherà di comporre le informazioni ottenute tramite lo studio di questi due diversi problemi variazionali per scoprire informazioni sulle coppie minimizzanti l'intero funzionale \mathfrak{F} .

4.1 Ottimalità a ν fissata

Ci restringeremo al caso in cui $\Omega \subset \mathbb{R}^n$ è la chiusura di un aperto limitato, $\mathcal{L} = I_\Omega \cdot \mathcal{L}^n$, $c(x, y) = |x - y|^p$, $p \geq 1$ e la funzione f con cui si costruisce il funzionale F è superlineare, strettamente convessa e C^1 (anche se alcune delle considerazioni che faremo saranno valide anche in casi più generali).

4.1.1 Prime proprietà dei minimi

Una prima, evidente, proprietà è che, se μ minimizza \mathfrak{F}_ν , siccome questo funzionale assume valore infinito sulle misure dotate di parte singolare rispetto a \mathcal{L}^n , allora $\mu \ll \mathcal{L}^n$ e la sua densità u è integrabile se composta con la funzione convessa f (quindi u appartiene ad un opportuno spazio di Orlicz). D'ora in avanti, per comodità, indicheremo sempre con u la densità di μ , cioè $\frac{d\mu}{d\mathcal{L}^n}$, invece di indicarla con μ^a .

Esempio 4.1. Se $f(x) = x^q$, $q > 1$ allora $u \in L^q(\Omega)$.

Per le prime considerazioni il punto di partenza sarà questo facile lemma.

Lemma 4.1. *L'applicazione $\mu \mapsto T_c(\mu, \nu)$ è convessa in μ .*

Dimostrazione. La formula di dualità rende evidente questo enunciato, in quanto esprime T_c come sup di quantità lineari in μ . \square

Può essere utile notare fin da subito che da questo, con le ipotesi fatte, segue l'unicità del minimo di \mathfrak{F}_ν .

Corollario 4.2. *Il funzionale \mathfrak{F}_ν è dotato di un unico punto di minimo in $\mathcal{P}(\Omega)$.*

Dimostrazione. L'esistenza segue dal metodo diretto esposto nel capitolo precedente, mentre l'unicità nasce dal fatto che il funzionale è strettamente convesso (perché, per il lemma 4.1, il primo addendo è convesso, mentre per le ipotesi su f il secondo lo è addirittura strettamente). \square

Lemma 4.3. *Se μ minimizza \mathfrak{F}_ν allora per ogni altra misura di probabilità μ_1 tale che $\mathfrak{F}_\nu(\mu_1) < +\infty$, dotata di densità u_1 , vale*

$$T_c(\mu_1, \nu) - T_c(\mu, \nu) + \int f'(u)(u_1 - u) \geq 0.$$

Dimostrazione. Per ogni $\varepsilon > 0$ vale ovviamente

$$\begin{aligned} T_c(\mu, \nu) + F(\mu) &\leq T_c(\mu + \varepsilon(\mu_1 - \mu)) + F(\mu + \varepsilon(\mu_1 - \mu)) \\ &\leq T_c(\mu, \nu) + \varepsilon(T_c(\mu_1, \nu) - T_c(\mu, \nu)) + F(\mu + \varepsilon(\mu_1 - \mu)). \end{aligned}$$

Pertanto $T_c(\mu_1, \nu) - T_c(\mu, \nu) + \varepsilon^{-1} [F(\mu + \varepsilon(\mu_1 - \mu)) - F(\mu)]$ risulta essere una quantità non negativa. Mandando $\varepsilon \rightarrow 0$ si ottiene la tesi non appena si dimostri che

$$\lim_{\varepsilon \rightarrow 0} \int \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} = \int f'(u)(u_1 - u).$$

Utilizzando la monotonia dei rapporti incrementali delle funzioni convesse si vede che, per $\varepsilon < 1$,

$$\begin{aligned} 0 &\leq \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \leq f(u_1) - f(u) \text{ se } u_1 \geq u; \\ 0 &\geq \frac{f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \geq f(u_1) - f(u) \text{ se } u_1 \leq u. \end{aligned}$$

Pertanto l'integrando, che converge puntualmente a $f'(u)(u_1 - u)$ perché f è C^1 , si domina con $|f(u) - f(u_1)|$, che è sicuramente integrabile perché il funzionale \mathfrak{F}_ν assume valori finiti sia su μ che su μ_1 e quindi le quantità $\int f(u)$ e $\int f(u_1)$ sono finite. Per convergenza dominata si ottiene la tesi. \square

D'ora in avanti ci limiteremo a T_p in luogo di T_c .

Teorema 4.4. *Per quasi ogni $x \in \Omega$ tale che $u(x) > 0$, se t è un trasporto ottimo fra μ e ν , posto $d = |x - t(x)|$, vale $u > 0$ quasi ovunque in $B(t(x), d)$.*

Dimostrazione. Consideriamo una famiglia numerabile \mathcal{A} di aperti che siano una base della topologia di Ω e limitiamoci ai punti x che siano di Lebesgue per u e per $I_{t^{-1}(A)}$ per ogni $A \in \mathcal{A}$. Questi punti sono un insieme di misura piena e sarà per loro che varrà la tesi.

Fissiamo allora un punto x fra di essi, con $u(x) > 0$, e supponiamo che la tesi non sia verificata: troveremo un assurdo procurandoci una misura μ_1 che contraddica il lemma 4.3. Per negare la validità della tesi dobbiamo supporre che esista un insieme $Z \subset B(t(x), d)$, con $|Z| > 0$, tale che $u = 0$ su Z . Possiamo supporre $Z \subset B(t(x), d - \varepsilon)$, per un certo $\varepsilon > 0$. Scegliamo un raggio $0 < R < d - \varepsilon$ e un aperto $A \in \mathcal{A}$, tale che $x \in A \subset B(t(x), R)$ e, per un raggio $0 < r < \varepsilon$, poniamo $B = t^{-1}(A) \cap B(x, r)$. Costruiamo la misura μ_1 così:

$$\mu_1 = I_{B^c} \cdot \mu + \frac{\mu(B)}{|Z|} I_Z \cdot \mathcal{L}^n;$$

la relazione fra le densità pertanto sarà

$$u_1 = I_{B^c} u + \frac{\mu(B)}{|Z|} I_Z.$$

Per costruzione $u_1 > u$ solo su Z , cioè in un insieme in cui $u = 0$ e quindi anche $f'(u) = 0$; pertanto se $T_p(\mu_1, \nu) < T_p(\mu, \nu)$ si ha una contraddizione della disuguaglianza data dal lemma 4.3 (la misura μ_1 differisce da μ solo su $B \cup Z$, dove ha densità u_1 limitata, pertanto $\mathfrak{F}_\nu(\mu_1) < +\infty$).

Costruiamo un trasporto t_1 fra μ_1 e ν nel seguente modo:

$$t_1(x) = \begin{cases} t(x) & \text{se } x \notin Z \\ \bar{t}(x) & \text{se } x \in Z, \end{cases}$$

dove \bar{t} è un qualunque trasporto fra $\frac{\mu(B)}{|Z|}I_Z \cdot \mathcal{L}^n$ e $t_{\#}I_B\mu$ (per esempio quello ottimo). Varrà ovviamente $T_p(\mu_1, \nu) \leq \int |x - t_1(x)|^p u_1(x) dx$. Paragoniamo i costi dei trasporti t e t_1 in relazione rispettivamente a μ, μ_1 : il primo ha costo nullo su Z e il secondo su B , su $(Z \cup B)^c$ coincidono sia i trasporti che le misure quindi anche il costo, su B il primo paga almeno $\mu(B)|d - r - R|^p$ e su Z il secondo al massimo $\mu(B)|d - \varepsilon + R|^p$. Per concludere ci basta scegliere r, R tali che $\mu(B)(|d - r - R|^p - |d - \varepsilon + R|^p) > 0$. Questo è equivalente a che la stessa quantità divisa per $\omega_n r^n$ sia positiva. Vale

$$\begin{aligned} & \frac{\mu(B)(|d - r - R|^p - |d - \varepsilon + R|^p)}{\omega_n r^n} \\ & \geq \frac{u(x)|B|(|d - r - R|^p - |d - \varepsilon + R|^p)}{\omega_n r^n} - \frac{\int_{B(x,r)} |u(y) - u(x)| dy}{\omega_n r^n}. \end{aligned}$$

Al limite per $r \rightarrow 0$, tenendo conto che x è un punto di Lebesgue per u e quindi l'ultimo termine è infinitesimo e che è di densità 1 per $t^{-1}(A)$ e quindi $\lim_{r \rightarrow 0} |B|(\omega_n)^{-1} r^{-n} = 1$, si ottiene $u(x)(|d - R|^p - |d - \varepsilon + R|^p)$, che è positivo se $-R > -\varepsilon + R$, cioè $R < \frac{\varepsilon}{2}$, pertanto basta scegliere r abbastanza piccolo per dimostrare la tesi. \square

Questo teorema mostra come l'insieme $\{u > 0\}$ sia in effetti, a meno di insiemi Lebesgue-trascurabili, più regolare di quanto a priori ci si potrebbe aspettare da una qualunque u integrabile (si può, per esempio, facilmente dedurre che il supporto della misura μ coincide con la chiusura della propria parte interna). Non ci addentreremo però nelle complicate considerazioni sulle proprietà di $spt(\mu)$, che potrebbero dipendere da $spt(\nu)$, ma esamineremo solo il caso più facile in cui quest'ultimo supporto sia costituito da un numero finito di punti.

Questo non sarà restrittivo per due motivi: innanzitutto perché quando esamineremo le condizioni di ottimalità sull'intero funzionale \mathfrak{F} , facendo variare entrambe le misure, per la scelta che abbiamo operato sul funzionale G , varrà $\nu = \sum_{i=0}^{\infty} a_i \delta_{x_i}$, per degli opportuni punti $(x_i)_i$ e coefficienti a_i con $\sum_{i=0}^{\infty} a_i = 1$. Siccome dimostreremo in seguito che, grazie all'ipotesi di Ω limitato, il numero di punti sarà in realtà finito, la nostra scelta sarà giustificata. Il secondo motivo consiste nel fatto che un procedimento di approssimazione di una ν generica tramite misure con supporto finito ci permetterà di mantenere la validità di alcune considerazioni anche nello studio dei punti di minimo di \mathfrak{F}_ν , togliendo a ν la restrittiva ipotesi di essere concentrata su un insieme finito.

Nel caso che il supporto di ν sia finito è facile conseguenza del teorema 4.4 che l'insieme $\{u > 0\}$ coincida, a meno di insiemi trascurabili, con un'unione finita di palle centrate nei punti di $spt(\nu)$ (parleremo sempre di palle intendendo palle in Ω , cioè intersezioni di palle euclidee con Ω). L'unione delle chiusure di queste palle avrà un numero finito di componenti connesse, ciascuna a distanza positiva dall'unione delle altre. Detta C una di queste componenti è chiaro, sempre come conseguenza del teorema

4.4, che $I_C \cdot \mu$ e $I_C \cdot \nu$ sono mandate l'una nell'altra tramite il trasporto ottimo t fra μ e ν e pertanto hanno la stessa massa. Inoltre, a meno di riscalarle di un fattore moltiplicativo, risolvono un problema di minimo con un funzionale analogo al nostro, come più avanti spiegheremo con maggiori dettagli. Pertanto le condizioni di ottimalità possono essere studiate nel caso che ν abbia supporto finito e $\{u > 0\}$ sia equivalente ad un connesso che sia unione finita di palle chiuse centrate nei punti di $\text{spt}(\nu)$.

4.1.2 Relazione fra densità e potenziale

Il risultato principale di questa sezione sarà un'equazione che legghi la densità u di una μ minimizzante alla funzione c -concava ψ che realizza il massimo nella formula di dualità. Siccome, soprattutto nel caso di $p = 1$, questa è una funzione lipschitziana nota come potenziale di Kantorovich, chiameremo questa relazione nel modo indicato dal titolo. Ricavarla non sarebbe per nulla difficile, grazie alla potenza di uno strumento quale la formula di dualità, se non ci fossero problemi legati alla non unicità di tale ψ . Il lavoro si concentrerà pertanto sul bypassare tali problemi.

Lemma 4.5. *Se $\{u > 0\}$ è equivalente ad un aperto connesso allora esiste ed è unica, a meno di costanti additive e di insiemi μ -trascurabili, una funzione c -concava ψ rispetto al costo $c(x, y) = |x - y|^p, p > 1$ che realizzi il massimo nella formula di dualità relativa al trasporto fra μ e ν .*

Dimostrazione. L'esistenza si ha perché il teorema 1.6 la garantisce nel caso di costi continui. Il teorema 1.14 invece garantisce l'unicità di t e, tramite la formula (1.10) anche di $\nabla\psi$ (entrambe le unicità valgono μ -q.o.) visto che l'ottimalità nella formula di dualità è equivalente per l'osservazione 1.3 al fatto che il piano di trasporto ottimo γ sia concentrato su $G\partial_c\psi$. Essendo la ψ lipschitziana ed essendo il suo gradiente determinato \mathcal{L}^n -q.o. su un aperto connesso, essa risulta determinata univocamente su tale aperto (e quindi μ -q.o.) a meno di costanti additive. \square

Lemma 4.6. *Se $c(x, y) = |x - y|^p$ e $D = \text{diam}(\Omega)$ allora ogni funzione c -concava è anche pD^{p-1} -lipschitziana.*

Dimostrazione. Se ψ è c -concava allora per una certa χ vale

$$\psi(x) = \inf_y |x - y|^p - \chi(y)$$

e la tesi segue dalla pD^{p-1} -lipschitzianità delle funzioni $x \mapsto |x - y|^p - \chi(y)$. \square

Indicheremo per comodità con k la funzione inversa della funzione continua e crescente $f' : \mathbb{R}^+ \rightarrow \mathbb{R}^+$.

Teorema 4.7 (Relazione densità-potenziale). *Supponiamo che la coppia (μ, ν) sia tale che sia determinata univocamente, a meno di costanti additive e di insiemi μ -trascurabili, la ψ che realizza il massimo nella formula di dualità. Allora se μ è ottima nel problema a ν fissata, esisterà una costante $l \in \mathbb{R}$ tale che*

$$u = k((l - \psi)^+). \quad (4.1)$$

Dimostrazione. Consideriamo un'arbitraria misura μ_1 su cui il funzionale \mathfrak{F}_ν assuma valore finito (che è equivalente a $\int f(u_1) < +\infty$). Come nella dimostrazione del lemma 4.3 confrontiamo con le misure $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$ per $\varepsilon \in [0, 1]$. Varrà

$$\frac{T_p(\mu_\varepsilon) - T_p(\mu)}{\varepsilon} + \frac{\int f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \geq 0.$$

Scriviamo i valori di T_p con l'ausilio della formula di dualità, scegliendo delle funzioni ψ_ε c -concave che realizzino il massimo di $\int \psi d\mu_\varepsilon + \int \psi^c d\nu$. Possiamo stimare dal basso $T_p(\mu, \nu)$ con $\int \psi_\varepsilon d\mu + \int \psi_\varepsilon^c d\nu$ per qualsiasi ε , per cui

$$\frac{\int \psi_\varepsilon d\mu_\varepsilon + \int \psi_\varepsilon^c d\nu - \int \psi_\varepsilon d\mu - \int \psi_\varepsilon^c d\nu}{\varepsilon} + \frac{\int f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \geq 0.$$

Semplificando ed esplicitando $\mu_\varepsilon, \nu_\varepsilon$ si ottiene

$$\int \psi_\varepsilon d(\mu_1 - \mu) + \frac{\int f(u + \varepsilon(u_1 - u)) - f(u)}{\varepsilon} \geq 0. \quad (4.2)$$

Ora passiamo al limite per $\varepsilon \rightarrow 0$: del secondo addendo possiamo calcolare il limite esattamente come nel lemma 4.3. Per il primo teniamo invece conto del fatto che le funzioni ψ_ε sono equilipschitziane per il lemma 4.6 quindi, almeno su una successione $\varepsilon_i \rightarrow 0$ varrà $\psi_{\varepsilon_i} \rightarrow \bar{\psi}$ uniformemente per una certa $\bar{\psi}$ (l'equilimitatezza è facilmente ottenibile modificando le ψ_ε per costanti additive). Al limite quindi la (4.2) diventa

$$\int \bar{\psi} d(\mu_1 - \mu) + \int f'(u)(u_1 - u) \geq 0 \quad (4.3)$$

Vorremmo trovare una relazione fra $\bar{\psi}$ e la ψ di cui si dichiara un certo tipo di unicità nelle ipotesi, e per farlo abbiamo ancora un certo grado di libertà nella scelta delle ψ_ε , qualora esse non siano univocamente determinate; qualora invece lo siano (cosa dovuta a condizioni sul supporto delle μ_ε) ci limiteremo a esplicitarne la natura. Si vede comunque facilmente che $\bar{\psi}$ realizza il massimo nella formula di dualità per (μ, ν) (vedi il teorema B.2, le cui ipotesi sono largamente verificate). Per l'ipotesi fatta, allora, $\bar{\psi}$ coincide con la ψ ottima nella formula di dualità per (μ, ν) quasi ovunque sull'insieme $\{u > 0\}$ (a meno di decidere le costanti additive). A noi però serve controllare il comportamento di $\bar{\psi}$ anche fuori da tale insieme.

Imponiamoci ora di scegliere in modo canonico le ψ_ε , utilizzando il teorema di Ruschen-
dorf (teorema A.3). Sia $\Gamma_\varepsilon = \text{spt}(\gamma_\varepsilon)$, dove γ_ε è il piano di trasporto ottimo per la coppia
(μ_ε, ν); scegliamo $(x_\varepsilon, y_\varepsilon) \in \Gamma_\varepsilon$: a questo punto è naturalmente definita una scelta di ψ_ε ,
che indicheremo tramite $\psi_{\Gamma_\varepsilon, x_\varepsilon, y_\varepsilon}$, data dalla formula (A.4). A meno di sottosuccessioni
possiamo supporre, come prima abbiamo già fatto notare, che

$$\begin{aligned}\psi_{\Gamma_\varepsilon, x_\varepsilon, y_\varepsilon} &\rightarrow \bar{\psi} \text{ uniformemente,} \\ \Gamma_\varepsilon &\xrightarrow{\mathcal{H}} \Lambda \subset \Omega \times \Omega, \\ x_\varepsilon &\rightarrow \bar{x} \in \Omega, \quad y_\varepsilon \rightarrow \bar{y} \in \Omega\end{aligned}$$

(per la convergenza di Hausdorff e la compattezza dell'insieme dei chiusi di $\Omega \times \Omega$ si
veda il corollario A.22 e tutta la parte dell'appendice A dedicata a tale argomento). Se
chiamiamo Γ il supporto di γ , piano di trasporto ottimo per (μ, ν) , vale $\Gamma \subset \Lambda$: questo
viene come conseguenza del fatto che $\gamma_\varepsilon \rightarrow \gamma$, cosa che è dimostrata nel successivo lemma
4.8. Possiamo esprimere $\psi_{\Gamma_\varepsilon, x_\varepsilon, y_\varepsilon}(x)$ come $\inf_n \inf_{z \in (\Omega \times \Omega)^n} F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}$, dove i funzionali
 $F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}$ sono definiti nel lemma 4.9, ove si mostra che il loro Γ -limsup per $\varepsilon \rightarrow 0$ è
maggiorato da $F_{n, \Lambda, \bar{x}, \bar{y}, x}$. Pertanto

$$\psi_{\Gamma_\varepsilon, x_\varepsilon, y_\varepsilon}(x) \leq \inf_{z \in (\Omega \times \Omega)^n} F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}$$

e, passando al limsup in ε ,

$$\bar{\psi}(x) \leq \inf_{z \in (\Omega \times \Omega)^n} F_{n, \Lambda, \bar{x}, \bar{y}, x}.$$

Passando poi all'inf in n si ottiene $\bar{\psi}(x) \leq \psi_{\Lambda, \bar{x}, \bar{y}}(x)$. poiché $\Lambda \supset \Gamma$ vale $\psi_{\Lambda, \bar{x}, \bar{y}} \leq \psi_{\Gamma, \bar{x}, \bar{y}}$,
purchè $(\bar{x}, \bar{y}) \in \Gamma$. Non è un problema però garantire questo, grazie all'arbitrarietà nella
scelta di $(x_\varepsilon, y_\varepsilon)$: per ogni $(\bar{x}, \bar{y}) \in \Gamma \subset \Lambda$, in virtù della convergenza anche nel senso di
Kuratowski di Γ_ε a Λ , è possibile scegliere una successione di punti $(x_\varepsilon, y_\varepsilon) \in \Gamma_\varepsilon$ che vi
tenda. Abbiamo quindi infine ottenuto

$$\bar{\psi} \leq \psi_{\Gamma, \bar{x}, \bar{y}} = \psi, \tag{4.4}$$

dove l'ultima uguaglianza è stata ottenuta per l'unicità supposta per ipotesi, utilizzando
come ψ quella che si annulli sul punto \bar{x} (puntualizzazione necessaria perché ψ era
definita in maniera univoca solo a meno di costanti additive). Allora si può sostituire
nella (4.3) ψ al posto di $\bar{\psi}$ e mantenere la validità della disuguaglianza, grazie alla
disuguaglianza (4.4) valida ovunque e all'uguaglianza $\psi = \bar{\psi}$, valida μ -quasi ovunque.
Si ottiene allora

$$\int \psi d(\mu_1 - \mu) + \int f'(u)(u_1 - u) \geq 0,$$

che può essere riscritta come

$$\int \psi + f'(u) d\mu_1 \geq \int \psi + f'(u) d\mu.$$

L'arbitrarietà di μ_1 permette di concludere la tesi, in quanto il primo membro dell'ultima disuguaglianza può essere reso arbitrariamente vicino a $\text{ess inf } \psi + f'(u) = l$, con oculate scelte di μ_1 fra le misure assolutamente continue. Allora l'integrale rispetto alla misura μ della funzione $\psi + f'(u)$ è maggiorato da quello della costante l , di cui $\psi + f'(u)$ è quasi ovunque maggiore. Da questo si ricava $\psi + f'(u) = l$ a meno di insiemi μ -trascurabili, che è la tesi. Infatti, modificando u in un insieme trascurabile, possiamo assumere che $\psi + f'(u) = l$ valga su $\{u > 0\}$, mentre altrove continua a valere $\psi + f'(u) \geq l$. Questo è come dire $f'(u) = (l - \psi)^+$, che, componendo con k dà la relazione (4.1). \square

Osservazione 4.1. La costante l del teorema 4.7 è determinata univocamente dal fatto che $k((l - \psi)^+)$ abbia integrale 1.

Esempio 4.2. L'ipotesi del teorema 4.7 è verificata se

$$|\Omega|f\left(\frac{1}{|\Omega|}\right) > D^p.$$

Infatti grazie al lemma 4.5 ci basta verificare che $u > 0$ q.o., per le ipotesi fatte su Ω (chiusura di un aperto connesso con bordo trascurabile). Se così non fosse potremmo definire la funzione $u_1 = \frac{1}{|\{u=0\}|} I_{\{u=0\}}$. Il lemma 4.3 ci darebbe allora

$$0 \leq T_p(\mu_1, \nu) - T_p(\mu, \nu) + \int f'(u)(u_1 - u) \leq D^p - \int f'(u)u,$$

dove si è tenuto conto del fatto che, per costruzione, laddove $u_1 > 0$, u si annulla e con essa anche $f'(u)$. Dalla disuguaglianza $\frac{f(x)}{x} \leq f'(x)$ valida per funzioni convesse segue

$$0 \leq D^p - \int f(u) \leq D^p - |\Omega|f\left(\frac{1}{|\Omega|}\right),$$

dove l'ultima stima segue dalla disuguaglianza di Jensen e contraddice l'ipotesi.

Nell'espore questa teoria abbiamo avuto bisogno, e ne avremo ancora, di alcuni lemmi di Γ -convergenza, dei quali presentiamo ora quelli che ci sono serviti nella dimostrazione del teorema 4.7.

Lemma 4.8. *Sia T_ε il funzionale definito su $\mathcal{P}(\Omega \times \Omega)$ munito della topologia derivante dalla convergenza debole, dato da*

$$T_\varepsilon(\gamma) = \begin{cases} \int c d\gamma & \text{se } (\pi_X)_\# \gamma = \mu_\varepsilon, (\pi_Y)_\# \gamma = \nu, \\ +\infty & \text{altrimenti,} \end{cases}$$

dove $\mu_\varepsilon = \mu + \varepsilon(\mu_1 - \mu)$: allora $T_\varepsilon \xrightarrow{\Gamma} T$ per $\varepsilon \rightarrow 0$ (per la definizione di T si sostituisca μ a μ_ε).

Dimostrazione. La disuguaglianza del Γ -limsup ($T \geq \Gamma - \limsup T_\varepsilon$) è facile, basta scegliere, fissato $\gamma \in \mathcal{P}(\Omega \times \Omega) : T(\gamma) < +\infty$ una successione $\gamma_\varepsilon \rightarrow \gamma$ data da $\gamma_\varepsilon = \gamma + \varepsilon(\mu_1 \otimes \nu - \gamma)$. Allora $T_\varepsilon(\gamma_\varepsilon) = \int cd\gamma_\varepsilon \rightarrow \int cd\gamma = T(\gamma)$. Per il Γ -liminf si supponga per assurdo di trovare una successione $\gamma_\varepsilon \rightarrow \gamma : \liminf_\varepsilon T_\varepsilon(\gamma_\varepsilon) < T(\gamma)$. In particolare $\liminf_\varepsilon T_\varepsilon(\gamma_\varepsilon) < +\infty$ per cui, definitivamente, $T_\varepsilon(\gamma_\varepsilon) = \int cd\gamma_\varepsilon \rightarrow \int cd\gamma$. Dal fatto che γ_ε è un piano di trasporto fra μ_ε e ν segue che γ lo è fra μ e ν , il che dà $\int cd\gamma = T(\gamma)$, che è un assurdo.

Come conseguenza se indichiamo con γ_ε il piano di trasporto ottimo per la coppia (μ_ε, ν) e con γ quello relativo a (μ, ν) , vale $\gamma_\varepsilon \rightarrow \gamma$. \square

Lemma 4.9. *Sia $F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}$ il funzionale su $(\Omega \times \Omega)^n$ munito della solita topologia, dato da*

$$F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}(z) = \begin{cases} c(x, z_n^y) - c(z_n^x, z_n^y) + \dots + c(z_1^x, y_\varepsilon) - c(x_\varepsilon, y_\varepsilon) & \text{se } z \in \Gamma_\varepsilon^n, \\ +\infty & \text{altrimenti,} \end{cases}$$

dove z_i^x, z_i^y rappresentano le componenti di z : allora vale la disuguaglianza

$$\Gamma - \limsup_{\varepsilon \rightarrow 0} F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x} \leq F_{n, \Lambda, \bar{x}, \bar{y}},$$

dove $F_{n, \Lambda, \bar{x}, \bar{y}}$ è definito con le ovvie sostituzioni.

Dimostrazione. Ci basterà considerare $z \in \Lambda^n$ e produrre una successione $z_\varepsilon \rightarrow z$ per la quale $\lim_\varepsilon F_{n, \Gamma_\varepsilon, x_\varepsilon, y_\varepsilon, x}(z_\varepsilon) = F_{n, \Lambda, \bar{x}, \bar{y}}(z)$. Siccome però tutte le funzioni che compaiono sono continue e $x_\varepsilon \rightarrow \bar{x}, y_\varepsilon \rightarrow \bar{y}$, ci basterà produrre una successione $(z_\varepsilon)_\varepsilon$ tale che $z_\varepsilon \rightarrow z, z_\varepsilon \in \Gamma_\varepsilon^n$. Scegliamo allora per $(z_\varepsilon)_i$ un punto di Γ_ε di distanza minima da z_i . Ci basta solo verificare $z_\varepsilon \rightarrow z$, ma $d(z_i, (z_\varepsilon)_i) = d(z_i, \Gamma_\varepsilon) \leq \mathcal{H}(\Lambda, \Gamma_\varepsilon) \rightarrow 0$. \square

Osservazione 4.2. La relazione densità-potenziale (4.1) in generale non è in grado di identificare immediatamente u perché ψ stessa dipende da μ : è pertanto da intendersi come una equazione di punto fisso.

L'applicazione di cui bisogna cercare un punto fisso è quella che a u associa la funzione $k((l - \psi)^+)$, dove ψ è il potenziale ottimo nella formula di dualità relativa a μ (dipendente da u) e alla ν fissata e l è tale che il tutto abbia integrale 1. In generale questa però non è una buona definizione perché l'insieme dei potenziali ottimi potrebbe non essere costituito da una sola ψ , nemmeno a meno di costanti additive. Si potrebbe allora pensare alla ricerca di un punto fisso di una multifunzione, oppure la si potrebbe restringere ai casi in cui si abbia a che fare con un'applicazione univoca. In ogni caso non ci interessa trovare condizioni per garantire l'esistenza di un punto fisso, perché il teorema 4.7 già svolge questo ruolo, ma eventualmente sarebbe interessante trovare modi per approssimarlo (contrazioni...).

Un caso facile in cui la relazione densità-potenziale permette di identificare la μ nel problema a ν fissato è quello in cui $\nu = \delta_{x_0}$, per un punto $x_0 \in \Omega$.

Esempio 4.3. Se $\nu = \delta_{x_0}$ allora $u(x) = k((l - |x - x_0|^p)^+)$, per una certa costante l che può dipendere dalla forma di Ω e dalla posizione di x_0 .

Infatti dalla relazione $\psi(x) + \psi^c(y) = |x - y|^p$ per ogni $(x, y) \in \text{spt}(\gamma)$, che segue da $\text{spt}(\gamma) \subset G\partial_c\psi$, siccome in questo caso $(x, y) \in \text{spt}(\gamma)$ se e solo se $x \in \text{spt}(\mu)$ e $y = x_0$, segue $\psi(x) = |x - x_0|^p$ a meno di costanti additive sul supporto di μ . Nel caso in cui x_0 sia abbastanza lontano dal bordo di Ω , il valore di l si può trovare facilmente (ed in alcuni casi esplicitamente), scrivendo l'integrale della densità u sulla palla in Ω , che è una vera palla euclidea, in coordinate polari ed imponendo che sia uguale a 1.

4.1.3 Regolarità e PDE

Una prima, immediata conseguenza del teorema 4.7, è la continuità di u , che è peraltro una continuità quantificata, in quanto si tratta di k composta con una funzione pD^{p-1} -lipschitziana. Cerchiamo per prima cosa di estendere la nostra analisi al di là dell'ipotesi di connessione, pur restando nel comodo caso di ν con supporto finito.

Restringiamo μ e ν ad un insieme C che sia un aperto che contiene una sola componente connessa della chiusura dell'insieme $\{u > 0\}$, che possiamo supporre, a meno di modificare u su insiemi trascurabili, essere un'unione di palle, cosa che possiamo fare grazie alla finitezza e compattezza di tali componenti, che pertanto risultano essere a distanze positive l'una dalle altre. Chiamiamo μ_C, ν_C le restrizioni e, come abbiamo già accennato, scopriamo che queste misure, come conseguenza del teorema 4.4, sono trasportate da t l'una nell'altra (t non può connettere punti di componenti connesse diverse) e hanno pertanto la stessa massa m . Le misure $\frac{\mu_C}{m}, \frac{\nu_C}{m}$ risolvono allora un problema di natura simile al nostro, ambientato in \bar{C} , che si può verificare essere quello con lo stesso costo di trasporto $c(x, y) = |x - y|^p$, ma con la funzione f sostituita da f_m , data da $f_m(x) = f(mx)/m$. Vale $f'_m(x) = f'(mx)$ e $k_m(y) = (f'_m)^{-1}(y) = k(y)/m$. Il teorema 4.7, ci darà allora

$$\frac{u}{m} = \frac{k((l - \psi)^+)}{m},$$

in cui si può chiaramente semplificare m . Si noti che in questo caso la ψ è il potenziale relativo alla coppia $(\frac{\mu_C}{m}, \nu_C m)$, che comunque risulterà in realtà legato a quello per la coppia μ, ν , grazie al fatto che i trasporti ottimi per le due coppie devono essere l'uno la restrizione dell'altro. In ogni caso la c -concavità di questo potenziale sarebbe da intendere restringendo c a $\bar{C} \times \bar{C}$.

Il vantaggio dell'aver dimostrato di poter sostituire $\bar{\psi}$ con ψ anche fuori da $\{u > 0\}$ sta ora nel fatto che la relazione trovata non vale solo su tale insieme, ma si prolunga anche al di fuori, mostrando per esempio la continuità di u anche nei punti di contatto

con l'insieme $\{u = 0\}$. In questo modo si possono mettere insieme i risultati sui vari C (ingrossamenti delle componenti connesse) per ottenere la continuità di u senza ipotesi di connessione, nel caso di $spt(\nu)$ finito.

Con i prossimi lemmi di Γ -convergenza è possibile ottenere la stessa continuità anche per ν qualsiasi e $p = 1$.

Lemma 4.10. *Il funzionale $\mathfrak{F}_{\nu,p}$ definito per $p \leq 1$ come \mathfrak{F}_{ν} utilizzando il costo $c(x, y) = |x - y|^p$, Γ -converge rispetto alla convergenza debole su $\mathcal{P}(\Omega)$ per $p \rightarrow 1^+$ a $\mathfrak{F}_{\nu,1}$.*

Dimostrazione. Mostriamo la disuguaglianza del Γ -liminf supponendo per assurdo che esista una successione $\mu_p \rightharpoonup \mu_1$ tale che $\lim_{p \rightarrow 1^+} \mathfrak{F}_{\nu,p}(\mu_p) < \mathfrak{F}_{\nu,1}(\mu_1)$. Detto γ_p un piano di trasporto ottimo per il costo $|x - y|^p$ fra μ_p e ν a meno di sottosuccessioni si può supporre che $\gamma_p \rightharpoonup \gamma_1$, con γ_1 che sarà un piano di trasporto fra μ_1 e ν . Abbiamo allora

$$\liminf_p F(\mu_p) \geq F(\mu_1),$$

per la semicontinuità di F , ma anche

$$T_p(\mu_p, \nu) = \int |x - y|^p \gamma_p(d(x, y)) \rightarrow \int |x - y| \gamma_1(d(x, y)) \geq T_1(\mu_1, \nu),$$

per la convergenza debole di γ_p , ma forte delle funzioni continue $|x - y|^p$ verso la funzione $|x - y|$. Questo contraddice l'ipotesi fatta e mostra la disuguaglianza del Γ -liminf. Per la disuguaglianza del Γ -limsup utilizziamo la successione $(\mu_p)_p$ data da $\mu_p = \mu_1$ per ogni valore di p . Scelto un γ piano di trasporto qualunque fra μ_1 e ν vale

$$\begin{aligned} \limsup_p T_p(\mu_1, \nu) + F(\mu_1) &\leq \limsup_p \int |x - y|^p \gamma(d(x, y)) + F(\mu_1) \\ &= \int |x - y| \gamma(d(x, y)) + F(\mu_1). \end{aligned}$$

Passando all'inf in γ si ottiene

$$\limsup_p T_p(\mu_1, \nu) + F(\mu_1) \leq T_1(\mu_1, \nu) + F(\mu_1). \quad \square$$

Lemma 4.11. *Per ogni $p \geq 1$ il funzionale \mathfrak{F}_{ν_h} Γ -converge rispetto alla convergenza debole su $\mathcal{P}(\Omega)$ a \mathfrak{F}_{ν} ogni qualvolta $\nu_h \rightharpoonup \nu$.*

Dimostrazione. Grazie al fatto di lavorare su uno spazio compatto $\mathcal{P}(\Omega)$ si può identificare con $W_p(\Omega)$ (non ci sono condizioni sui momenti e le due convergenze, debole e Wasserstein, sono le stesse): si noti anche che $T_p(\mu, \nu_h) = \delta_p(\mu, \nu_h)^p$. Seguiamo una linea dimostrativa analoga al teorema precedente: date $\mu_h \rightarrow \mu$ vale $\liminf_h F(\mu_h) \geq F(\mu)$ e $T_p(\mu_h, \nu_h) \rightarrow T_p(\mu, \nu)$ (per la continuità della distanza). Con gli stessi ragionamenti di prima si ottiene la prima parte della dimostrazione. La seconda si ottiene ancora considerando $\mu_h = \mu$, perché questa volta addirittura $T_p(\mu, \nu_h) + F(\mu) \rightarrow T_p(\mu, \nu) + F(\mu)$. \square

Teorema 4.12. *Per ogni $p \geq 1$ e ogni $\nu \in \mathcal{P}(\Omega)$ se μ è minimizzante per $\mathfrak{F}_{\nu,p}$ allora μ ha una densità u continua, della forma $u = k \circ \theta$, con θ funzione pD^{p-1} -lipschitziana.*

Dimostrazione. Sappiamo già che questo è valido per $p > 1$ e ν con supporto finito. Per passare a $p = 1$ basta usare il lemma 4.10, che, avendo il funzionale $\mathfrak{F}_{\nu,1}$ un unico minimo ed essendo i $\mathfrak{F}_{\nu,p}$ banalmente equicoercivi perché $\mathcal{P}(\Omega)$ è compatto, garantisce la convergenza debole delle μ_p ottime per $\mathfrak{F}_{\nu,p}$ alla μ_1 ottima per $\mathfrak{F}_{\nu,1}$. Tuttavia le μ_p sono tutte misure assolutamente continue e le loro densità formano una successione di funzioni equicontinue ed equilimitate. L'equilimitatezza viene dal fatto che non possono assumere valori troppo alti, altrimenti a causa della loro lipschitzianità sarebbero costrette ad avere un inf positivo e alto, il che impedirebbe loro di avere integrale 1. Grazie all'equilimitatezza si può restringere la funzione continua k ad un intervallo compatto su cui è quindi uniformemente continua, il che, insieme all'equicontinuità delle funzioni composte con k dà l'equicontinuità della successione. Pertanto il limite debole viene ad essere, usando Ascoli-Arzelà, una misura la cui densità è limite uniforme delle densità delle approssimanti, e quindi anch'essa continua. Componendo con l'inversa di k , che è la funzione f' , si scopre che anch'essa è della forma $k \circ \theta$, con θ lipschitziana, la cui costante di Lipschitz si può stimare con $\lim_{p \rightarrow 1+} pD^{p-1} = 1$.

Un analogo ragionamento si può fare per ottenere il caso generale, per ν arbitraria, sempre sfruttando l'equicontinuità, approssimando una qualsiasi ν con una successione di misure a supporto finito ed infine applicando il lemma 4.11. \square

Nel caso $p > 1$, la relazione (1.10) fra il potenziale e il trasporto ottimo, insieme alla relazione densità-potenziale e alla relazione (1.1) fra il trasporto e le densità di μ, ν , dà luogo ad un'equazione differenziale. Infatti $\nabla\psi = -\nabla(f'(u)) = -f''(u)\nabla u$ μ -q.o., quindi $t(x) = x - (\partial h)^{-1}(-f''(u)\nabla u) = x + (\partial h)^{-1}(f''(u)\nabla u)$, per via della particolare forma della funzione $h(z) = |z|^p$, sempre μ -q.o.. Questa relazione ha senso perché la u sarà differenziabile Lebesgue-quasi ovunque, grazie al fatto che θ lo è, se si suppone che k lo sia almeno fuori dal valore 0. L'uso della (1.1) richiede invece un'ulteriore differenziabilità (oltre che l'essenziale iniettività di t , che però è garantita qualora anche la misura ν sia assolutamente continua rispetto alla misura di Lebesgue, perché l'unico piano di trasporto ottimo si deve esprimere per mezzo di una mappa di trasporto in entrambi i versi). L'equazione che viene, valida puntualmente q.o. nel caso che si possa supporre $u \in C^{1,1}$ e che $\nu \ll \mathcal{L}^n$ con densità v , è piuttosto complicata e la diamo solo nel caso $p = 2, f(x) = x^2$:

$$u(x) = v(x + \nabla u(x)) \det(I + \nabla^2 u(x)).$$

Col cambio di variabile $\tilde{u}(x) = u(x) + \frac{x^2}{2}$, diventa

$$\tilde{u}(x) = \frac{x^2}{2} + v(\nabla \tilde{u}(x)) \det \nabla^2 \tilde{u}(x),$$

cioè un caso particolare della generica equazione di Monge-Ampère (si veda [14], capitolo 4, per le relazioni coi problemi di trasporto, i vari tipi di soluzioni deboli di queste equazioni ed alcuni risultati di regolarità):

$$\det \nabla^2 u(x) = F(x, u(x), \nabla u(x)).$$

4.1.4 Relazione densità-potenziale per $p = 1$

Abbiamo già mostrato che si può estendere qualche risultato di regolarità (essenzialmente lipschitziana) anche al caso $p = 1$, grazie all'uso della Γ -convergenza, con cui alla fine si giungeva a mostrare la convergenza di u_p , densità della misura ottima per $p > 1$, a u_1 . Per ottenere $u_1 = k((l - \psi)^+)$, per una certa l e un potenziale ψ , ottimo nella formula di dualità relativa al costo $|x - y|$ e alle misure μ_1 e ν , bisogna passare al limite, almeno su sottosuccessioni, anche nei potenziali $l_p - \psi_p$. Ci limiteremo a ragionare sulle componenti connesse di $\overline{\{u > 0\}}$, così come nel teorema 4.7. Consideriamo pertanto le misure μ_p che minimizzano $\mathfrak{F}_{\nu,p}$ e definiamo i funzionali

$$X_p(\psi) = - \int \psi d\mu_p - \int \psi^{c,p} d\nu,$$

sulle funzioni ψ che siano, per esempio $2D$ -lipschitziane e limitate in modulo dalla costante $\frac{1}{|\Omega|} + 2D^2$ (costante da cui sono limitate tutte le funzioni $2D$ -lipschitziane la cui parte positiva abbia integrale 1); con l'apice c,p si intende la c -coniugata calcolata rispetto al costo $c(x, y) = |x - y|^p$. Lo spazio di tali ψ è compatto con metrica della convergenza uniforme. poiché le funzioni $l_p - \psi_p$ minimizzano ciascuna il relativo funzionale X_p in questo spazio, se mostriamo la Γ -convergenza di X_p a X_1 otteniamo che i limiti di sottosuccessioni delle $l_p - \psi_p$ minimizzano X_1 e pertanto realizzano l'ottimo nella formula di dualità per $p = 1$. La c -concavità di tali limiti segue dal fatto che per $p = 1$ la c -concavità coincide con la proprietà di 1-lipschitzianità e tali funzioni sono state ottenute come limite per una successione di numeri p tendenti a 1^+ di funzioni pD^{p-1} -lipschitziane.

Lemma 4.13. *Sia $(\psi_p)_p$ una successione di funzioni nello spazio di definizione dei funzionali X_p e supponiamo $\psi_p \rightarrow \psi_1$ uniformemente: allora anche $\psi_p^{c,p} \rightarrow \psi_1^{c,1}$ uniformemente.*

Dimostrazione. Ci basta verificare la convergenza puntuale grazie all'equicontinuità ed equilimitatezza delle ψ_p . Fissiamo allora un punto y e abbiamo

$$\psi_p^{c,p}(y) = \inf_x [|x - y|^p - \psi_p(x)].$$

Se mostriamo la Γ -convergenza delle funzioni di cui facciamo l'inf a destra alla funzione $x \rightarrow |x - y| - \psi_1(x)$ otteniamo la convergenza dei relativi inf e quindi la convergenza

puntuale cercata. Di queste funzioni tuttavia abbiamo la convergenza uniforme (il primo addendo converge uniformemente perché siamo su un limitato, il secondo per ipotesi) e il limite uniforme è continuo. Questo implica la Γ -convergenza. \square

Teorema 4.14. Per $p \rightarrow 1^+$, se $\mu_p \rightharpoonup \mu$ allora si ha $X_p \xrightarrow{\Gamma} X_1$.

Dimostrazione. Consideriamo un'arbitraria successione $\psi_p \rightarrow \psi_1$: se dimostriamo la convergenza $X_p(\psi_p) \rightarrow X_1(\psi_1)$, avremo contemporaneamente verificato le due disuguaglianze del Γ -liminf e del Γ -limsup. Si ottiene il risultato voluto tramite

$$\int \psi_p d\mu_p \rightarrow \int \psi_1 d\mu_1,$$

per convergenza debole delle μ_p e forte delle ψ_p ;

$$\int \psi_p^{c,p} d\nu \rightarrow \int \psi_1^{c,1} d\nu,$$

per la convergenza uniforme dimostrata nel lemma 4.13. \square

4.2 Ottimalità a μ fissata

4.2.1 Cardinalità del supporto

Considereremo sempre il caso in cui g è una funzione subadditiva s.c.i., con $g(0) = 0$ e $\lim_{x \rightarrow 0^+} \frac{g(x)}{x} = +\infty$, pertanto il funzionale da minimizzare sarà

$$\mathfrak{F}^\mu(\nu) = \begin{cases} T_p(\mu, \nu) + \sum_{i=0}^{\infty} g(a_i) & \text{se } \nu = \sum_{i=0}^{\infty} a_i \delta_{x_i} \text{ per certi } a_i, x_i, \\ +\infty & \text{altrimenti.} \end{cases}$$

Sulla misura fissata μ in generale non saranno necessarie molte ipotesi, ma possiamo limitarci al caso in cui sia assolutamente continua, che sarà quello da affrontare nello studio dei minimi dell'intero funzionale \mathfrak{F} .

Una ν ottima sarà concentrata a priori su una quantità numerabile di punti, mentre a noi piacerebbe trovare condizioni sotto cui il suo supporto sia finito. Il successivo esempio mostra come, in effetti, per garantire ciò serva proprio l'ottimalità della ν , e non solo il fatto che il funzionale assuma su di essa valore finito.

Esempio 4.4. Sia $a_i = 2^{-(i+1)}$, $i \in \mathbb{N}$ e $g(x) = \sqrt{x}$: allora $\sum_{i=0}^{\infty} a_i = 1$, ma $\sum_{i=0}^{\infty} g(a_i) = \frac{1}{\sqrt{2}-1} < +\infty$.

L'ipotesi sul comportamento di g in 0 dice esattamente che per ogni $L \in \mathbb{R}^+$ esiste $\delta(L) \in (0, 1)$ tale che $g(x) > Lx$ per ogni $x < \delta$: useremo spesso questa applicazione $\delta : \mathbb{R}^+ \rightarrow (0, 1)$. Come al solito indicheremo con D il diametro di Ω .

Lemma 4.15. *Se ν è ottima almeno un atomo di ν ha massa pari almeno a $\delta_0 = \delta(D^p + g(1))$.*

Dimostrazione. Supponiamo per assurdo che tutti gli atomi abbiano massa inferiore a δ_0 e siano $(a_i)_i$ le loro masse. Allora $\sum_i g(a_i) > (D^p + g(1)) \sum_i a_i = D^p + g(1)$. Paragoniamo la coppia (μ, ν) alla coppia $(\mu, \delta_{\bar{x}})$, per un generico punto $\bar{x} \in \Omega$: otteniamo

$$\begin{aligned} D^p + g(1) &< G(\nu) \leq T_p(\mu, \nu) + G(\nu) \\ &\leq T_p(\mu, \delta_{\bar{x}}) + G(\delta_{\bar{x}}) = T_p(\mu, \delta_{\bar{x}}) + g(1) \leq D^p + g(1), \end{aligned}$$

che è un assurdo. □

Teorema 4.16. *Supponiamo che g sia localmente lipschitziana in $(0, 1]$, sia $L_0 = \text{Lip}(g|_{(\delta_0, 1]})$ e definiamo $\delta_1 = \delta(D^p + L_0)$. Allora se ν è ottima tutti gli atomi di ν hanno massa almeno pari a $\delta_0 \wedge \delta_1$.*

Dimostrazione. Supponiamo per assurdo che non sia così: troviamo allora un atomo x_0 con massa a_0 tale che $0 < a_0 < \delta_0 \wedge \delta_1$. poiché $a_0 < \delta_1$ vale

$$\frac{g(a_0)}{a_0} > D^p + L_0 \tag{4.5}$$

Sia $x_1 \neq x_0$ un atomo di ν di massa $a_1 \geq \delta_0$. Definiamo $\nu' = \nu - a_0\delta_{x_0} + a_0\delta_{x_1}$. Si noti che $G(\nu) - G(\nu') = g(a_0) + g(a_1) - g(a_0 + a_1) \geq 0$. Inoltre si può vedere che $T_p(\mu, \nu') \leq T_p(\mu, \nu) + a_0D^p$: infatti un trasporto ammissibile fra μ e ν' , non necessariamente ottimo, è il seguente:

$$t'(x) = \begin{cases} t(x) & \text{se } t(x) \neq x_0, \\ x_1 & \text{se } t(x) = x_0 \end{cases}$$

e per questo vale

$$\begin{aligned} T_p(\mu, \nu') &\leq \int_{\Omega} |x - t'(x)|^p \mu(dx) \\ &\leq D^p \mu(t^{-1}(x_0)) + \int_{\Omega \setminus t^{-1}(x_0)} |x - t'(x)|^p \mu(dx) \\ &= D^p a_0 + \int_{\Omega \setminus t^{-1}(x_0)} |x - t(x)|^p \mu(dx) \\ &\leq D^p a_0 + \int_{\Omega} |x - t(x)|^p \mu(dx) \\ &= D^p a_0 + T_p(\mu, \nu). \end{aligned}$$

L'ottimalità di ν darà allora $T_p(\mu, \nu') + G(\nu') \geq T_p(\mu, \nu) + G(\nu)$, da cui $a_0D^p \geq g(a_0) + g(a_1) - g(a_0 + a_1)$. Dividendo per a_0 si ottiene

$$D^p \geq \frac{g(a_0)}{a_0} - \frac{g(a_0 + a_1) - g(a_1)}{a_0} \geq \frac{g(a_0)}{a_0} - L_0,$$

che contraddice la (4.5). □

Corollario 4.17. *Nelle ipotesi del teorema precedente, il numero di atomi di una ν ottima non può superare $\frac{1}{\delta_0 \wedge \delta_1}$ e quindi è finito.*

Esempio 4.5. L'ipotesi di locale lipschitzianità del teorema 4.16 è verificata se g è concava.

Esempio 4.6. Sia $\Omega = B(0,1)$, $g(x) = \sqrt{x}$, $p = 1$: allora ogni ν ottima ha al più 11 atomi.

Infatti per la funzione $g(x) = \sqrt{x}$ vale $\delta(L) = \frac{1}{L^2}$ e $Lip(g|_{(x,1]}) = \frac{1}{2\sqrt{x}}$, tenendo conto di $D = 2$ e $g(1) = 1$, si ottiene $\delta_0 = \delta(2+1) = \frac{1}{9}$, $\delta_1 = \delta(2 + \frac{3}{2}) = \frac{4}{49}$. Pertanto ogni atomo di ν ha massa almeno pari a $\frac{4}{49}$, ma ce ne è anche almeno uno di massa almeno $\frac{1}{9}$. Oltre a quell'atomo pertanto ce ne sono al più $\frac{8}{9} : \frac{4}{49} < 11$. Naturalmente tutte queste stime non hanno alcuna pretesa di essere fini.

4.2.2 Posizione e valore delle masse

Consideriamo uno degli atomi di ν , chiamiamolo x_0 , e poniamo $\Omega_0 = t^{-1}(x_0)$. Vorremo innanzitutto dimostrare che il punto x_0 è in un qualche senso il baricentro dell'insieme Ω_0 .

Teorema 4.18. *Siano $p \geq 1$, x_0, Ω_0 scelti come sopra: allora vale*

$$\int_{\Omega_0} (x - x_0)|x - x_0|^{p-2} \mu(dx) = 0.$$

Dimostrazione. Il teorema 3.10 ci assicura che $x_0 \notin \partial\Omega$. Consideriamo ora un arbitrario $v \in S^{n-1}$ e paragoniamo la misura ν a quella in cui si sposta la massa di x_0 nel punto $x_0 - \varepsilon v$, che, per ε piccolo, appartiene a Ω . poiché non cambia il termine G , dobbiamo solo stimare la variazione del costo del trasporto e lo possiamo fare nel solito modo (come nella dimostrazione del teorema 4.16). Otteniamo allora

$$\int_{\Omega_0} \frac{|x - x_0 + \varepsilon v|^p - |x - x_0|^p}{\varepsilon} \mu(dx) \geq 0. \quad (4.6)$$

Nella (4.6) si può passare al limite sotto il segno di integrale: infatti l'integrando si può stimare con $pD^{p-1}|v| = pD^{p-1}$ indipendentemente da ε . Il limite puntuale si ottiene derivando rispetto a ε l'espressione $|x - x_0 + \varepsilon v|^p$, tranne nel caso $p = 1$, dove nel punto x_0 non si ha derivabilità, in cui il limite è $|v| = 1$, ma, essendo il singolo punto x_0 μ -trascurabile, questo non inficia il passaggio al limite. Si ha allora

$$\int_{\Omega_0} p|x - x_0|^{p-2}(x - x_0) \cdot v \mu(dx) \geq 0,$$

e questo vale per ogni $v \in S^{n-1}$, da cui segue la tesi. \square

Osservazione 4.3. Si noti che solo se $p = 2$ si può dire che x_0 sia il baricentro della massa distribuita da μ in Ω_0 , altrimenti si tratta di un baricentro in cui bisogna pesare la misura μ con un coefficiente $|x - x_0|^{p-2}$.

Adesso, con una tecnica dimostrativa che ricalcherà quella utilizzata per il teorema 4.7, abbandonando pertanto le variazioni della ν fatte tramite trasporti vicini all'identità (nel teorema di prima la traslazione di εv), ci dedicheremo ai valori delle masse che ν ha nei suoi atomi, e non solo alle posizioni di questi ultimi. Per questo teorema supporremo $g \in C^1((0, 1])$, ipotesi che garantisce anche la locale lipschitzianità di g nello stesso intervallo e permette di dire che il supporto di ν è finito.

Teorema 4.19. *Sia $p > 1$ e supponiamo che $u > 0$ q.o. in Ω : allora se ν minimizza \mathfrak{F}^μ , detta ψ l'unica (a meno di costanti additive) funzione c -concava che realizza il massimo nella formula di dualità per μ, ν , esisterà una costante l tale che per ogni atomo x di ν valga*

$$g'(\nu(x)) = l - \psi^c(x).$$

Dimostrazione. Consideriamo un'arbitraria misura di probabilità ν_1 concentrata sugli stessi atomi di ν : $\nu = \sum_{i=0}^m a_i \delta_{x_i}$; $\nu_1 = \sum_{i=0}^m a_i^1 \delta_{x_i}$. Posto $\nu_\varepsilon = \nu + \varepsilon(\nu_1 - \nu)$, paragoniamo ν a ν_ε e, scegliendo delle funzioni c -concave ψ_ε che realizzino il massimo nella formula di dualità fra μ e ν_ε , otteniamo

$$\int \psi_\varepsilon^c d(\nu_1 - \nu) + \frac{\sum_{i=0}^m g(a_i + \varepsilon(a_i^1 - a_i)) - g(a_i)}{\varepsilon} \geq 0,$$

relazione a cui si arriva dopo alcuni passaggi che omettiamo perché seguono esattamente il procedimento esposto nel teorema 4.7. Passando al limite per $\varepsilon \rightarrow 0^+$, utilizzando il fatto che le ψ_ε convergono uniformemente a ψ (ammettono limiti per sottosuccessioni, ma l'unico limite possibile è ψ , come già visto nel teorema 4.7) e che allora la convergenza uniforme passa anche alle loro coniugate (come mostrato nella dimostrazione del teorema A.2), si ottiene, scrivendo gli integrali come somme,

$$\sum_{i=0}^m (\psi^c(x_i) + g'(a_i)) a_i^1 \geq \sum_{i=0}^m (\psi^c(x_i) + g'(a_i)) a_i.$$

Utilizzando l'arbitrarietà di ν_1 si può ottenere, posto $l = \min_i \psi^c(x_i) + g'(a_i)$,

$$l \geq \sum_{i=0}^m (\psi^c(x_i) + g'(a_i)) a_i,$$

da cui si ottiene che per ogni i vale $l = \psi^c(x_i) + g'(a_i)$, che è la tesi. \square

Osservazione 4.4. Nelle ipotesi del teorema precedente, se la coppia (μ, ν) minimizza tutto \mathfrak{F} , si può utilizzare la relazione densità-potenziale per μ e ricavare ψ^c da $-u$ invece che da ψ .

Osservazione 4.5. Il teorema 4.19 può essere applicato alle singole componenti connesse di $\overline{\{u > 0\}}$, nel caso che in esse le masse di μ e di ν coincidano (cosa verificata se (μ, ν) è una coppia ottima per \mathfrak{F}).

Il motivo per cui è stato molto meno faticoso ottenere il teorema 4.19 rispetto a quanto successo con il 4.7 sta nel fatto che non ci interessava qui sostituire la ψ ad un'eventuale $\bar{\psi}$, limite di sottosuccessione, su un insieme più grande di $\{u > 0\}$, pertanto abbiamo messo nelle ipotesi $u > 0$ q.o.. Della u ci interessava infatti anche il comportamento al bordo di $\{u > 0\}$, essenzialmente la sua continuità, problema che non si presenta nello studio delle masse degli atomi di ν .

4.3 Considerazioni sui minimi globali

Le considerazioni delle precedenti due sezioni ci permettono di esibire varie informazioni sulle misure (μ, ν) che minimizzano \mathfrak{F} . Riassumiamo le ipotesi che ci servono

- $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ strettamente convessa, C^1 , superlineare (poniamo $k = (f')^{-1}$);
- $g : [0, 1] \rightarrow \mathbb{R}^+$, $g(0) = 0$, $g \in C^1((0, 1])$, $\lim_{x \rightarrow 0^+} \frac{g(x)}{x} = +\infty$;
- $\Omega \subset \mathbb{R}^n$ chiusura di un aperto limitato (di diametro D), con bordo trascurabile rispetto alla misura di Lebesgue, che è la nostra misura di riferimento \mathcal{L} ;
- $c(x, y) = |x - y|^p$, $p \geq 1$.

Riassumiamo ciò che sotto queste ipotesi possiamo dire fin'ora:

- $\mu \ll \mathcal{L}^n$, con densità continua della forma $k \circ \theta$, con θ funzione pD^{p-1} -lipschitziana;
- ν ha supporto finito $\{x_i | i = 1, \dots, m\}$ e nessuno degli x_i appartiene a $\partial\Omega$;
- l'insieme dove $u > 0$ è un'unione di palle di centro i punti x_i .

In realtà, però, il teorema 4.7 ci dà ancora più informazioni sulla densità u , che ora esplicheremo.

Teorema 4.20. *Sia C una componente connessa di $\overline{\{u > 0\}}$ e supponiamo $x_i \in C$ per $i = 1, \dots, m_C$. Allora per ogni $i = 1, \dots, m_C$ esiste un numero c_i tale che per ogni $x \in C$*

$$u(x) = k \left((c_1 - |x - x_1|^p \vee \dots \vee c_{m_C} - |x - x_{m_C}|^p)^+ \right).$$

Dimostrazione. Grazie alla relazione densità-potenziale, ci basta dimostrare $l - \psi(x) = c_1 - |x - x_1|^p \vee \dots \vee c_m - |x - x_{m_C}|^p$. Scegliamo $c_i = l + \psi^c(x_i)$. Siccome per quasi ogni x vale $\psi(x) + \psi^c(t(x)) = |x - y|^p$ e $\psi(x) + \psi^c(x_i) \leq |x - x_i|^p$ la tesi è dimostrata, tenendo

conto che, se $x \in C$, il punto $t(x)$ è uno dei punti $x_i \in C$. In particolare si noti che il massimo, per ciascun punto x tale che $(x, t(x)) \in \text{spt}(\gamma)$, è realizzato scegliendo i in modo che $t(x) = x_i$. \square

Questo permette di intuire il grafico della densità u ottima: tante “gobbe” centrate nei punti x_i , che potrebbero sovrapporsi nel qual caso la funzione u assume il massimo dei valori che le sono consentiti. Se consideriamo $f' \circ u$ queste gobbe hanno una forma modellata su quella di una costante cui si sottragga $|x - x_i|^p$. Il valore massimo abbiamo visto essere c_i , che coincide con $\psi^c(x_i)$ a meno di una costante additiva su ogni componente connessa C e che, per il teorema 4.19 e l’osservazione 4.5, coincide (sempre a meno di costanti additive valide su ogni componente connessa) con $g'(\nu(x_i))$, il che lega l’altezza raggiunta da u nei punti x_i alla massa ivi posta da ν .

Si noti che, se ci si limita al caso $p = 1$, grazie al fatto che è molto più chiaro di quanto non lo sia per $p > 1$ chi siano le funzioni c -concave, si possono evitare molti ragionamenti di componenti connesse. Mettendo insieme le varie funzioni $(l - \psi)^+$ ottenute nelle varie componenti connesse di $\overline{\{u > 0\}}$ leggermente ingrossate (grazie al fatto che hanno distanza positiva l’una dalle altre), fino a supporre che la loro unione si a tutto Ω , si ottiene un’unica funzione che è ancora c -concava (1-lipschitziana) ed è ancora un potenziale di Kantorovich perché la relazione importante è $\psi(x) + \psi^c(y) = |x - y|$ per ogni coppia $(x, y) \in \text{spt}(\gamma)$, ma x e $t(x)$ appartengono sempre alla stessa componente connessa e quindi questo fatto era già verificato. L’ultimo teorema può pertanto essere rinunciato per $p = 1$ in questo modo:

Teorema 4.21. *Sia $p = 1$ e (μ, ν) una coppia ottima per \mathfrak{F} : allora per ogni $i = 1, \dots, m$ esiste un numero c_i tale che per ogni $x \in \Omega$*

$$u(x) = k \left((c_1 - |x - x_1|^p \vee \dots \vee c_m - |x - x_m|^p)^+ \right).$$

Concludiamo il capitolo utilizzando l’unica proprietà di ottimalità che ancora non abbiamo sfruttato, quella di baricentralità, per mostrare come a volte si possano escludere le sovrapposizioni fra le diverse gobbe. Se la forma e la grandezza di Ω lo permettono, infatti, la coppia (μ, ν) ottima tende a separare le varie gobbe. Ogni volta in cui questo accade, il problema si riduce (come nel prossimo capitolo vedremo che potremo fare nel caso $\Omega = \mathbb{R}^n$) a decidere il numero di gobbe e quanta massa in ciascuna di esse. L’idea del motivo per cui le gobbe preferiscono dividersi dovrebbe risultare chiara da questo esempio, facilmente generalizzabile.

Esempio 4.7. Sia $p = 1$, $f(x) = \frac{x^2}{2}$, $R_1 = \sqrt[n+1]{\frac{n+1}{2\omega_n}}$, $R_2 = 2 \sqrt[n+1]{\frac{n+1}{\omega_n}}$, $\Omega = B(0, R_2) \subset \mathbb{R}^n$. Allora non può essere ottima alcuna coppia (μ, ν) con $\nu = \frac{\delta_{x_0} + \delta_{x_1}}{2}$, $x_0 + x_1 = 0, |x_0| < R_1$

Infatti se vi fosse una coppia ottima di tal genere intorno a x_0 e x_1 si creerebbero delle gobbe costruite su palle centrate in tali punti, che però dovrebbero intersecarsi.

Riuscirebbero ad evitarlo solo se tali palle non raggiunessero l'origine. Tuttavia in questo caso, guardando per esempio x_0 , si avrebbe un raggio $R < |x_0|$ tale che

$$\int_0^R (R_r) n \omega_n r^{n-1} dr = \frac{1}{2},$$

perché ormai sappiamo costruire la densità u di μ . Tuttavia svolgendo i calcoli si scopre che l'ipotesi su $|x_0|$ lo impedisce. D'altra parte le due palle che verranno a crearsi non deborderanno da Ω perché già con raggio $R = \sqrt[n+1]{\frac{n+1}{\omega_n}} < R_2 - R_1 < R_2 - |x_0|$, si riesce a raggiungere massa $\frac{1}{2}$ solo utilizzando la metà della palla che si trova dalla parte opposta all'origine, pertanto i raggi della palle che si formeranno saranno minori di questo valore e le palle non usciranno da Ω . L'assurdo nasce ora dal fatto che x_0 non sarebbe il baricentro di Ω_0 pesato secondo quanto dice il teorema 4.18, perché da un lato vi sarebbe metà palla con densità u , mentre dall'altro la densità sarebbe simmetrica ma parte della palla apparterebbe a Ω_1 invece che ad Ω_0 , precisamente quella costituita dai punti più vicini a x_1 che ad x_0 .

Capitolo 5

Esistenza in \mathbb{R}^n

In tutto il capitolo ci dedicheremo alla ricerca di un risultato di esistenza per la minimizzazione del funzionale \mathfrak{F} fra le misure di probabilità in \mathbb{R}^n . Manterremo su f e g le stesse ipotesi che ci hanno permesso di ottenere le condizioni necessarie di ottimalità riassunte al termine del capitolo precedente. Lavoreremo con costi di trasporto $|x - y|^p$, $p \geq 1$. A differenza di quanto succedeva per Ω limitato qui, quando affronteremo il problema globale, ci limiteremo alla minimizzazione di \mathfrak{F} fra le misure che appartengono a $\mathcal{W}_p(\mathbb{R}^n)$.

5.1 I problemi parziali

Cominciamo col considerare il problema di minimizzare i funzionali \mathfrak{F}_ν e \mathfrak{F}^μ , fissando pertanto una delle misure.

Teorema 5.1. *Per ogni coppia (μ, ν) di misure di probabilità in \mathbb{R}^n , entrambi i funzionali \mathfrak{F}_ν e \mathfrak{F}^μ sono dotati di minimo nello spazio $\mathcal{P}(\mathbb{R}^n)$.*

Dimostrazione. Abbiamo la semicontinuità rispetto alla convergenza debole di tutti gli addendi di tali funzionali (siccome non siamo più su un Ω limitato non possiamo più dire che la distanza di Wasserstein metrizza la convergenza debole e pertanto ottenere la continuità di T_p , tuttavia sappiamo comunque che è semicontinuo, grazie al lemma 3.5). Per ottenere l'esistenza ci basta pertanto mostrare che una qualunque successione minimizzante è tesa. Siccome utilizzeremo di tale successione solo il fatto che su di essa il termine T_p è limitato, siamo in una situazione di simmetria fra le due variabili e analizzeremo solo uno dei due casi: mostreremo che per una qualsiasi μ , se $\sup_h T_p(\mu, \nu_h) < +\infty$, allora $(\nu_h)_h$ è tesa.

Supponiamo per assurdo che non sia così. Allora esiste $\varepsilon > 0$ tale che, per ogni raggio R , esiste un indice $h(R)$ tale che $\nu_{h(R)}(B(0, R)^c) \geq \varepsilon$. Per la singola misura μ esisterà

un raggio r tale che $\mu(B(0, r)^c) < \varepsilon/2$. Preso allora un qualunque piano di trasporto ottimo $\gamma_{h(R)}$ fra μ e $\nu_{h(R)}$, si avrà

$$\begin{aligned}
\gamma_{h(R)}(B(0, r) \times B(0, R)^c) &= 1 - \gamma_{h(R)}(B(0, r)^c \times \mathbb{R}^n \cup \mathbb{R}^n \times B(0, R)) \\
&\geq 1 - \gamma_{h(R)}(B(0, r)^c \times \mathbb{R}^n) - \gamma_{h(R)}(\mathbb{R}^n \times B(0, R)) \\
&= 1 - \mu(B(0, r)^c) - \nu_{h(R)}(B(0, R)) \\
&= \nu_{h(R)}(B(0, R)^c) - \mu(B(0, r)^c) \\
&\geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.
\end{aligned}$$

Varrà inoltre, se $R > r$

$$\begin{aligned}
T_p(\mu, \nu_{h(R)}) &= \int c d\gamma_{h(R)} \geq \int_{B(0, r) \times B(0, R)^c} c d\gamma_{h(R)} \\
&\geq |R - r|^p \gamma_{h(R)}(B(0, r) \times B(0, R)^c) \geq |R - r|^p \frac{\varepsilon}{2},
\end{aligned}$$

da cui si deduce

$$\sup_h T_p(\mu, \nu_h) \geq |R - r|^p \frac{\varepsilon}{2},$$

che è assurdo, data l'arbitrarietà di R . □

Osservazione 5.1. La misura che realizza il minimo (per esempio ν) che si trova in questo modo appartiene a $\mathcal{W}_p(\mathbb{R}^n)$ se e solo se vi appartiene anche quella fissata (diciamo μ). Se $\mu \notin \mathcal{W}_p(\mathbb{R}^n)$ allora $\mathfrak{F}^\mu|_{\mathcal{W}_p} \equiv +\infty$.

Data una misura positiva σ di massa m definiamo $\mathfrak{F}_{\sigma, m}(\mu) = T_p(\mu, \frac{\sigma}{m}) + F_m(\mu)$, dove il funzionale f_m è definito utilizzando la funzione f_m introdotta a pagina 64 $f_m(x) = f(mx)/m$.

Lemma 5.2. *Sia $C \subset \mathbb{R}^n$ e indichiamo con μ_C la misura $I_C \cdot \mu$, restrizione di μ a C . Sia t un trasporto ottimo fra μ e ν e supponiamo che $t_{\#}\mu_C$ sia anch'essa una misura concentrata su C . Detta m la massa comune a μ_C e $t_{\#}\mu_C$, misura che indicheremo con ν'_C , allora, se μ minimizza \mathfrak{F}_ν , $\frac{\mu_C}{m}$ minimizza $\mathfrak{F}_{\nu'_C, m}$ sull'insieme C .*

Dimostrazione. Se così non fosse troveremmo allora una misura che indicheremo con $\frac{\mu'}{m}$ concentrata su C che dia un risultato migliore di $\frac{\mu_C}{m}$ nel funzionale $\mathfrak{F}_{\nu'_C, m}$. Dimostriamo allora che $\mu - \mu_C + \mu'$ è migliore di μ nel funzionale \mathfrak{F}_ν . Abbiamo a nostra disposizione un trasporto ottimo t fra μ e ν e un trasporto ottimo t' fra $\frac{\mu'}{m}$ e $\frac{\nu'_C}{m}$ (μ e μ' sono necessariamente assolutamente continue). Definendo

$$t''(x) = \begin{cases} t(x) & \text{se } x \notin C \\ t'(x) & \text{se } x \in C, \end{cases}$$

si può verificare che $t''_{\sharp}(\mu - \mu_C + \mu') = \nu - \nu_C + \nu_C = \nu$ e stimare pertanto anche il termine $T_p(\mu - \mu_C + \mu', \nu)$:

$$\begin{aligned} T_p(\mu - \mu_C + \mu', \nu) &\leq \int_{C^c} |x - t(x)|^p d(\mu - \mu_C) + \int_C |x - t'(x)|^p d\mu' \\ &= \int_{C^c} |x - t(x)|^p d(\mu - \mu_C) + mT_p\left(\frac{\mu'}{m}, \frac{\nu'_C}{m}\right); \\ F(\mu - \mu_C + \mu') &= F(\mu - \mu_C) + F(\mu') = F(\mu - \mu_C) + mF_m\left(\frac{\mu'}{m}\right). \end{aligned}$$

Allora

$$\begin{aligned} \mathfrak{F}_\nu(\mu - \mu_C + \mu') &\leq \int_{C^c} |x - t(x)|^p d(\mu - \mu_C) + mT_p\left(\frac{\mu'}{m}, \frac{\nu'_C}{m}\right) + F(\mu - \mu_C) + mF_m\left(\frac{\mu'}{m}\right) \\ &< \int_{C^c} |x - t(x)|^p d(\mu - \mu_C) + mT_p\left(\frac{\mu_C}{m}, \frac{\nu'_C}{m}\right) + F(\mu - \mu_C) + mF_m\left(\frac{\mu_C}{m}\right) \\ &= T_p(\mu, \nu) + F(\mu), \end{aligned}$$

il che dà l'assurdo. □

Corollario 5.3. *Se ν è a supporto compatto, allora μ ha densità u continua.*

Dimostrazione. Basta verificare che u è continua su ogni palla $B(0, R)$ che contenga il supporto di ν e questo si ottiene restringendo μ a una di tali palle, applicando il lemma precedente, e usando il teorema 4.12. □

In realtà nel caso che ν abbia supporto compatto si può ottenere molto di più.

Teorema 5.4. *Supponiamo che ν abbia supporto compatto e che μ minimizzi \mathfrak{F}_ν : allora anche μ ha supporto compatto.*

Dimostrazione. Supponiamo per assurdo che μ non lo sia e consideriamo due palle $B(0, r), B(0, R)$ con $r < R$ tali che $\text{spt}(\nu) \subset B(0, r)$. Qualunque sia R , la misura μ dà massa a $B(0, R)^c$. Definiamo

$$u' = I_{B(0, R)}u + \frac{\mu(B(0, R)^c)}{|B(0, r)|} I_{B(0, r)},$$

che è chiaramente la densità di una nuova misura di probabilità μ' . Dimostreremo che, per R grande, vale $\mathfrak{F}_\nu(\mu') < \mathfrak{F}_\nu(\mu)$, trovando pertanto un assurdo. È chiaro che nel termine di trasporto tutta la massa che con μ si trovava in $B(0, R)^c$ faceva spendere puntualmente almeno $R - r$, mentre ora la stessa massa si trova in $B(0, r)$ e fa spendere

al più $2r$, creando pertanto un risparmio di almeno $\mu(B(0, R)^c)(R - 3r)$. Per il termine F invece si ha

$$F(\mu') - F(\mu) = \int_{B(0, r)} f\left(u + \frac{\mu(B(0, R)^c)}{|B(0, r)|}\right) - f(u) - \int_{B(0, R)^c} f(u).$$

D'altra parte vale la stima

$$f\left(u + \frac{\mu(B(0, R)^c)}{|B(0, r)|}\right) - f(u) \leq f'\left(u + \frac{\mu(B(0, R)^c)}{|B(0, r)|}\right) \frac{\mu(B(0, R)^c)}{|B(0, r)|}$$

e si può tralasciare l'ultimo termine, per cui in totale si ha

$$\mathfrak{F}_\nu(\mu') - \mathfrak{F}_\nu(\mu) \leq \mu(B(0, R)^c) \left[\int f'\left(u + \frac{\mu(B(0, R)^c)}{|B(0, r)|}\right) \frac{1}{|B(0, r)|} - (R - 3r) \right].$$

Basta allora mostrare che l'espressione in parentesi quadre è negativa, per R abbastanza grande. Per vedere che ciò è vero basta accorgersi del fatto che, per ogni R , l'integrale è finito, in virtù del fatto che la u è continua per il corollario 5.3 e quindi limitata, e decresce per $R \nearrow +\infty$. Siccome l'altro termine diverge a $-\infty$ la tesi è dimostrata. \square

5.2 Esistenza per il problema globale

Avviciniamoci lentamente alla meta, cominciando da questo risultato.

Lemma 5.5. *Per il funzionale \mathfrak{F} è possibile trovare una successione minimizzante $((\mu_h, \nu_h))_h$ in cui per ogni h il supporto di ν_h è finito.*

Dimostrazione. Ci basta mostrare che, date ad arbitrio $\mu, \nu \in \mathcal{W}_p(\mathbb{R}^n)$ su cui \mathfrak{F} assuma valore finito, e $\varepsilon > 0$, è possibile sostituire ν con una misura ν' di supporto finito, senza innalzare di oltre ε il valore del funzionale. Sia $\nu = \sum_{i=0}^{\infty} a_i \delta_{x_i}$ e definiamo $\nu_I = \sum_{i=0}^I a_i \delta_{x_i} + (\sum_{i=I+1}^{\infty} a_i) \delta_0$. Ponendo $\nu' = \nu_I$ il valore del termine G non aumenta, grazie alla subadditività di g , pertanto ci basta vedere che $\lim_{I \rightarrow +\infty} T_p(\mu, \nu_I) = T_p(\mu, \nu)$ per ottenere la tesi. Questo è vero perché T_p è la distanza di \mathcal{W}_p elevata alla p -esima potenza e $\nu_I \rightarrow \nu$: infatti basta usare il teorema 1.10 e verificare che si ha convergenza debole e convergenza dei momenti. Per questo secondo fatto basta notare $m_p^p(\nu_I, 0) = \sum_{i=0}^I |x_i|^p a_i$, quantità che converge a $\sum_{i=0}^{\infty} |x_i|^p a_i = m_p^p(\nu, 0)$, mentre per la convergenza debole si utilizzi la relazione, valida per un'arbitraria funzione $\varphi \in C_c(\mathbb{R}^n)$,

$$\left| \int_{\mathbb{R}^n} \varphi d(\nu - \nu_I) \right| \leq 2 \|\varphi\|_{\infty} \sum_{i=I+1}^{\infty} a_i,$$

dove il secondo membro è infinitesimo perché la serie $\sum_{i=0}^{\infty} a_i$ è convergente. \square

Un altro modo di rendere più regolari le successioni minimizzanti è il seguente.

Lemma 5.6. *Data una qualunque successione minimizzante $((\mu_h, \nu_h))_h$ con le ν_h che hanno supporto finito, se ne può trovare una $((\mu_h''', \nu_h'))_h$ in cui le ν_h' hanno supporto finito e il supporto di ogni misura μ_h''' è costituito da un'unione disgiunta di palle, in modo tale che ogni atomo di ν_h' sia centro di una di esse, e su ognuna di queste palle la densità di μ_h''' sia del tipo a gobba precedentemente descritto.*

Dimostrazione. Si migliori in tre passaggi la successione minimizzante data. Al primo si sostituisce ad ogni μ_h la misura che minimizza $\mu_h' \mathfrak{F}_{\nu_h}$, che sarà a supporto compatto per il lemma 5.4. Si considerino gli insiemi $t^{-1}(x_i)$, dove t è un trasporto ottimo fra questa misura e ν_h e gli x_i sono gli atomi di ν_h : il secondo passaggio consiste nel traslarli insieme con i corrispondenti atomi di ν_h in modo da distanziare sufficientemente gli x_i : questo è possibile perché si tratta di un numero finito di insiemi limitati. In questa fase il funzionale non cambia il proprio valore e otteniamo delle nuove misure μ_h'', ν_h' . Nel terzo passaggio cambiamo nuovamente μ_h'' , ottimizzando $\mathfrak{F}_{\nu_h'}$ e otteniamo una misura μ_h''' . Se gli atomi di ν_h' distano almeno $2R_0 + 1$ fra di loro (con il valore di R_0 dato dal lemma successivo) si è ottenuto ciò che si cercava. I vari passaggi sono illustrati dalla figura 5.1. \square

Lemma 5.7. *Esiste un R_0 tale che, in una coppia (μ, ν) ottima, col relativo trasporto ottimo t , per μ -q.o. x vale $|x - t(x)| \leq R_0$.*

Dimostrazione. Definiamo R_0 in modo che

$$\int_0^{R_0} k(R_0^p - r^p) n \omega_n r^{n-1} dr = 1.$$

Consideriamo gli x per cui vale la tesi del teorema 4.4, per cui $(x, t(x)) \in spt(\gamma)$ e tali che $u(x) > 0$, che sono μ -quasi tutti. Se per uno di questi valesse $|x - t(x)| = R > R_0$ si avrebbe che la palla di centro $t(x)$ e raggio R sarebbe compresa nella componente connessa di $\overline{\{u > 0\}}$ comprendente il punto $t(x)$, che è uno degli atomi x_i di ν . Utilizzando il risultato del teorema 4.20 si ottiene che esiste un punto x_i (proprio $t(x)$) tale che il relativo c_i deve essere pari ad almeno R^p , perché $u(x) > 0$ e $u(x) = k((c_i - R^p)^+)$. Allora su tutta la componente connessa, e quindi anche su tutta $B(t(x), R)$, vale $u(y) \geq k(R^p - |y - t(x)|^p)$, ma allora

$$\begin{aligned} \int_{\Omega} u &\geq \int_{B(t(x), R)} u \geq \int_{B(t(x), R)} k(R^p - |y - t(x)|^p) dy \\ &> \int_{B(t(x), R_0)} k(R_0^p - |y - t(x)|^p) dy = \int_0^{R_0} k(R_0^p - r^p) n \omega_n r^{n-1} dr = 1, \end{aligned}$$

che è un assurdo. \square

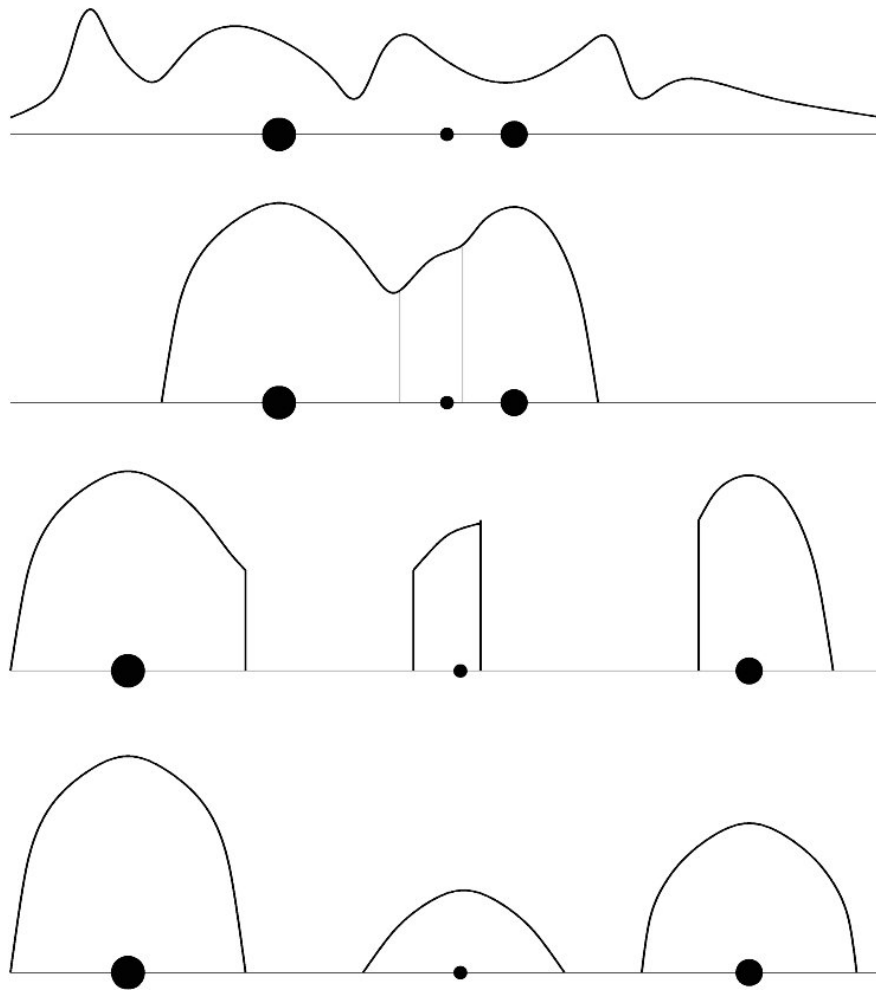


Figura 5.1: Successivi miglioramenti della successione minimizzante

Tutti questi lemmi ci permettono di arrivare ad enunciare questo criterio per assicurare l'esistenza di un minimo.

Teorema 5.8. *Se si trova una successione minimizzante $((\mu_h, \nu_h))_h$ in cui il numero di atomi di ν_h rimane limitato, allora esiste il minimo di \mathfrak{F} in $\mathcal{W}_p(\mathbb{R}^n)$.*

Dimostrazione. Applicando a questa successione il lemma 5.6 si ottengono due successioni di misure concentrate ciascuna su al più N palle di raggio al più R_0 , dove N è un numero che limita la cardinalità di $\text{spt}(\nu_h)$. Traslando queste palle, si ottiene, senza variare il valore del funzionale, una successione minimizzante in cui sia μ_h che ν_h , per ogni h , sono concentrate su $B(0, NR_0)$. L'estremo inferiore di \mathfrak{F} su $\mathcal{W}_p(\mathbb{R}^n)$ coinciderà pertanto con quello su $\mathcal{P}(B(0, NR_0))$, che è un minimo per quanto dimostrato nel capitolo 3. \square

Un primo risultato di esistenza è il seguente:

Esempio 5.1. Per il funzionale \mathfrak{F} fatto con $G(\nu) = \sharp(\text{spt}(\nu))$ esiste un minimo in $\mathcal{W}_p(\mathbb{R}^n)$.

Il funzionale G si ottiene a partire dalla funzione

$$g(x) = \begin{cases} 0 & \text{se } x = 0 \\ 1 & \text{se } x > 0 \end{cases}$$

che è subadditiva e s.c.i.. Il minimo esiste perché su una qualsiasi successione minimizzante vale

$$\sup_h \sharp(\text{spt}(\nu_h)) \leq \sup_h \mathfrak{F}(\mu_h n u_h) < +\infty.$$

Naturalmente questo esempio si generalizza facilmente al caso $\inf_{x>0} g(x) > 0$. Noi siamo però interessati a cosa succede per funzioni g quali $g(x) = x^r$, $r < 1$. Per arrivare ad un risultato di esistenza che copra anche questi casi abbiamo bisogno di definire l'effetto totale di un atomo di massa m sul funzionale.

Definizione 5.1. Definiamo l'effetto totale di un atomo di massa m sul funzionale \mathfrak{F} come la quantità

$$\begin{aligned} E(m) &= g(m) + \inf_{\mu \in m\mathcal{P}(\mathbb{R}^n)} T_p(\mu, m\delta_{x_0}) + F(\mu) \\ &= g(m) + m \inf_{\mu' \in \mathcal{P}(\mathbb{R}^n)} \mathfrak{F}_{m\delta_{x_0}, m}(\mu'), \end{aligned}$$

Questa quantità corrisponde all'effettivo apporto dell'atomo qualora la sua zona di influenza ($t^{-1}(x_0)$) sia costituita da una palla su cui la μ ottima sia del tipo a gobba, staccata dalle zone di influenza degli altri atomi. Utilizzando ciò che abbiamo scoperto per le μ ottime, e in particolare l'esempio 4.3, tenendo conto che qui non ci sono effetti

dovuti al bordo di Ω e che quindi questa gobba arriva a valere 0 sul bordo della palla, si può anche calcolare esplicitamente il valore di $E(m)$:

$$E(m) = g(m) + \int_0^{R(m)} [f(k(R(m)^p - r^p)) + k(R(m)^p - r^p)r^p] n\omega_n r^{n-1} dr,$$

dove $R(m)$ è il raggio della palla di influenza dell'atomo di massa m ed è caratterizzato da

$$m = \int_0^{R(m)} k(R(m)^p - r^p) n\omega_n r^{n-1} dr,$$

relazione invertibile che permette di trovare $m(R)$ o $R(m)$ equivalentemente.

Lemma 5.9. *Se la funzione E è subadditiva in un segmento $[0, m_0]$ allora si trova una successione minimizzante $((\mu_h, \nu_h))_h$ in cui per ogni h al più un atomo di ν_h ha massa inferiore a $\frac{m_0}{2}$.*

Dimostrazione. Consideriamo una successione minimizzante ottenuta prendendone una con le ν_h di supporto finito e applicando il lemma 5.6. Si ha allora

$$\mathfrak{F}(\mu_h, \nu_h) = \sum_{i=1}^{N(h)} E(a_i),$$

se $\nu_h = \sum_{i=1}^{N(h)} a_i \delta_{x_i}$. Adesso miglioriamo questa successione minimizzante sostituendo per ogni h alla coppia (μ_h, ν_h) una coppia sempre formata da una μ con gobbe disgiunte intorno agli atomi della ν , ma con ν ottenuta fondendo insieme alcuni atomi di ν_h ed eventualmente riposizionando lontani gli atomi ottenuti. Se si fondono due atomi di massa non superiore a $\frac{m_0}{2}$ si migliora il valore del funzionale, per l'ipotesi di subadditività di E . Possiamo iterare questo procedimento per ogni h un numero sufficiente di volte finché non rimanga al più un atomo di massa inferiore a $\frac{m_0}{2}$. \square

Chiaramente da questo lemma segue una stima sul numero di atomi di ν : se ne è al più uno di massa inferiore a $\frac{m_0}{2}$ tale numero non supererà N definito come il successivo della parte intera di $\frac{m_0}{2}$.

Il nostro obiettivo sarà pertanto mostrare che, almeno per valori piccoli di m , si ha subadditività di E . Siccome ovviamente $R(0) = 0$ e quindi $E(0) = 0$, potrà esserci utile questo lemma.

Lemma 5.10. *Sia $E : [0, b] \rightarrow \mathbb{R}^+$ una funzione concava con $E(0) = 0$: allora E è subadditiva.*

Dimostrazione. Scegliamo $x, y \in [0, b]$ e sia $T = x + y$. Supponiamo $T \leq b$: dobbiamo dimostrare $E(T) \leq E(x) + E(y)$. Consideriamo la funzione $\tilde{E} : [0, T] \rightarrow \mathbb{R}^+$ data da

$\tilde{E}(z) = E(z) + E(T - z)$. \tilde{E} è concava perché somma di concave, ed è s.c.i. come tutte le funzioni concave su un intervallo di \mathbb{R} : pertanto ha minimo e sappiamo che lo assume in uno degli estremi, per la concavità: tuttavia $\tilde{E}(0) = \tilde{E}(T)$, quindi $E(x) + E(y) = \tilde{E}(x) \geq \tilde{E}(0) = 0 + E(T) = E(T)$, che è la tesi. \square

Teorema 5.11. *Supponiamo, oltre alle solite ipotesi su f, g che $f \in C^2((0, +\infty))$, $g \in C^2((0, 1]) \cap C^0([0, 1])$. Allora il problema di minimizzazione di \mathfrak{F} in $\mathcal{W}_p(\mathbb{R}^n)$, per $p \geq 1$, ha soluzione se*

$$\lim_{R \rightarrow 0^+} g'' \left(\int_0^R k(R^p - r^p) n \omega_n r^{n-1} dr \right) \int_0^R k'(R^p - r^p) n \omega_n r^{n-1} dr = -\infty.$$

Dimostrazione. In virtù di tutti i lemmi precedenti ci basterà calcolare la derivata seconda di E e mostrare che vicino a zero è negativa, trovando pertanto un intervallo di concavità e quindi subadditività di E . Scriviamo $E(m) = g(m) + K(R(m))$ e deriviamo il secondo addendo.

$$\frac{dK(R(m))}{dm}(m) = \frac{dK}{dR}(R(m)) \frac{dR}{dm}(m) = \frac{\frac{dK}{dR}(R(m))}{\frac{dm}{dR}(R(m))}.$$

In $\frac{dK}{dR}$ ci sarebbero due addendi, quello in cui si usa il teorema fondamentale del calcolo e si sostituisce l'estremo di integrazione nell'integrando e quello in cui si usa il teorema di derivazione sotto il segno di integrale, lasciando invariato l'intervallo di integrazione; tuttavia il primo si annulla a causa di $k(R^p - R^p) = k(0) = 0$. Pertanto

$$\begin{aligned} \frac{dK}{dR}(R) &= \int_0^R pR^{p-1} [f'(k(R^p - r^p))k'(R^p - r^p) + k'(R^p - r^p)r^p] n \omega_n r^{n-1} dr \\ &= \int_0^R pR^{p-1} [(R^p - r^p)k'(R^p - r^p) + k'(R^p - r^p)r^p] n \omega_n r^{n-1} dr \\ &= \int_0^R pR^{p-1} R^p k'(R^p - r^p) n \omega_n r^{n-1} dr, \end{aligned}$$

dove si è tenuto conto di $f' \circ k = id$. D'altra parte, tenendo sempre conto del modo di derivare queste espressioni integrali e di $k(R^p - R^p) = k(0) = 0$, si ha

$$\frac{dm}{dR}(R) = \int_0^R pR^{p-1} k'(R^p - r^p) n \omega_n r^{n-1} dr,$$

pertanto semplificano si ottiene

$$\frac{dK(R(m))}{dm}(m) = R(m)^p.$$

Derivando ancora deduciamo

$$\begin{aligned} \frac{d^2 K(R(m))}{dm^2}(m) &= pR(m)^{p-1} \frac{dR}{dm}(m) \\ &= \frac{pR(m)^{p-1}}{\int_0^R(m) pR(m)^{p-1} k'(R(m)^p - r^p) n\omega_n r^{n-1} dr} \\ &= \frac{1}{\int_0^R(m) k'(R(m)^p - r^p) n\omega_n r^{n-1} dr}. \end{aligned}$$

Pertanto si ha

$$E''(m) = g''(m) + \frac{1}{\int_0^R(m) k'(R(m)^p - r^p) n\omega_n r^{n-1} dr}$$

e, per ottenere la tesi, basta che nelle vicinanze di 0 quest'espressione sia negativa, ed è perciò sufficiente che lo sia se moltiplicata per il denominatore del secondo addendo. L'ipotesi ci garantisce che in quest'ultima quantità il primo addendo diverga a $-\infty$, il che è sufficiente per ottenere la tesi. \square

Questo teorema ci dà l'esistenza del minimo in $\mathcal{W}_p(\mathbb{R}^n)$ sotto un'ipotesi che lega il comportamento di f e di g . Vediamo che, per i tipici esempi che si possono fare di funzioni che rispettino le varie ipotesi, questo è sempre verificato.

Esempio 5.2. Se $f(x) = x^q, g(x) = x^r, q > 1, r < 1$, l'ipotesi del teorema 5.11 è verificata.

Cominciamo con lo scrivere esplicitamente le funzioni con cui abbiamo a che fare, a meno di costanti moltiplicative positive che chiameremo sempre C :

$$f'(x) = qx^{q-1}, \text{ per cui } k(y) = Cy^{\frac{1}{q-1}} \text{ e } k'(y) = Cy^{\frac{2-q}{q-1}}; \quad g''(x) = \frac{-C}{x^{2-r}}.$$

La funzione g'' è crescente e negativa, mentre k' può essere crescente o decrescente a seconda di $q < 2$ o $q > 2$, ma in ogni caso si può stimare dal basso il valore (positivo) dell'integrale in cui compare k' con l'integrale fra $R/3$ e $2R/3$ e, a seconda della crescita o decrescenza, stimare puntualmente k' col valore in $R/3$ o $2R/3$, mentre r^{n-1} si stima comunque dal basso con $(R/3)^{n-1}$. Pertanto

$$\begin{aligned} g'' \left(\int_0^R k(R^p - r^p) n\omega_n r^{n-1} dr \right) &\leq g''(Rk(R^p) n\omega_n R^{n-1}) = \frac{-C}{R^{(n+\frac{p}{q-1})(2-r)}}; \\ \int_0^R k'(R^p - r^p) n\omega_n r^{n-1} dr &\leq CR^{n+p\frac{2-q}{q-1}}. \end{aligned}$$

È sufficiente che valga

$$\lim_{R \rightarrow 0^+} \frac{R^{n+p\frac{2-q}{q-1}}}{R^{(n+\frac{p}{q-1})(2-r)}} = +\infty,$$

condizione, questa, vera perché

$$n + p \frac{2-q}{q-1} - \left(n + \frac{p}{q-1} \right) (2-r) = n(r-1) + \frac{p}{q-1} (r-q) < 0.$$

Se nell'affrontare il problema in tutto \mathbb{R}^n abbiamo avuto qualche difficoltà relativa all'esistenza, non è difficile convincersi invece che, seguendo l'idea dell'esempio 4.7, risulta abbastanza chiaro come siano fatte le coppie (μ, ν) minimizzanti. In \mathbb{R}^n non vi sono infatti sgradevoli effetti legati alla forma del bordo di Ω e quindi ogni gobba ha sufficiente spazio per staccarsi dalle altre, cosa che avviene perché altrimenti vi sarebbero delle gobbe in cui è violata la condizione di baricentralità del corrispondente atomo di ν . Su coppie (μ, ν) costruite in questo modo il valore del funzionale \mathfrak{F} viene quindi a dipendere solo dai valori dei raggi R_i delle palle intorno ai punti x_i , atomi di ν , su cui si adagiano le gobbe di μ . Precisamente si ha

$$\begin{aligned} \mathfrak{F}(\mu, \nu) &= \sum_{i=0}^N E(m_i), \\ m_i &= \int_0^{R_i} k(R_i^p - r^p) n \omega_n r^{n-1} dr. \end{aligned}$$

Ci si è ridotti pertanto a minimizzare una funzione sul sottoinsieme di \mathbb{R}^N dato dal vincolo $\sum_{i=0}^N m_i = 1$, dove N è il numero costante che stima il numero di atomi di ν , ottenuto grazie alla subadditività di E su $[0, m_0]$, problema affrontabile con un semplice uso del moltiplicatore di Lagrange.

Appendice A

c -concavità e analisi convessa

A.1 c -concavità

In tutta questa sezione immagineremo assegnati due spazi X, Y e una funzione $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$. Un possibile riferimento per una brevissima trattazione di questi concetti è l'articolo originale in cui Ruschendorf dimostra il teorema che concluderà questa sezione ([13]).

A.1.1 Nozioni sulle funzioni c -concave

Definizione A.1. Una funzione $\psi : X \rightarrow \overline{\mathbb{R}}$ si dice c -concava se esiste una funzione $\chi : Y \rightarrow \overline{\mathbb{R}}$ tale che

$$\psi(x) = \inf_{y \in Y} c(x, y) - \chi(y)$$

Definizione A.2. Data una funzione $\psi : X \rightarrow \overline{\mathbb{R}}$ si definisce c -coniugata di ψ la funzione

$$\psi^c(y) = \inf_{x \in X} c(x, y) - \psi(x).$$

Teorema A.1. Per ogni funzione $\psi : X \rightarrow \overline{\mathbb{R}}$ vale $\psi^{cc} \geq \psi$, e l'uguaglianza vale se e solo se ψ è c -concava. In particolare l'applicazione $\psi \mapsto \psi^c$ è un'involutione nella classe delle funzioni c -concave.

Dimostrazione. La disuguaglianza $\psi^{cc}(x) = \inf_{y \in Y} c(x, y) - \psi^c(y) \geq \psi(x)$ è equivalente al fatto che, per ogni $y \in Y$, valga $c(x, y) - \psi^c(y) \geq \psi(x)$ (dappertutto si usa la convenzione $(+\infty) + (\pm\infty) = +\infty$). Se $c(x, y) = +\infty$ la disuguaglianza è banale, altrimenti la si può trasformare in $\psi^c(y) \leq c(x, y) - \psi(x)$, che segue dalla definizione di ψ^c .

Se vale l'uguaglianza $\psi = \psi^{cc}$ vale allora

$$\psi(x) = \inf_{y \in Y} c(x, y) - \psi^c(y) \quad (\text{A.1})$$

ed è quindi chiaro che ψ è c -concava.

Per dimostrare l'altra implicazione si noti che la c -trasformata $\psi \mapsto \psi^c$ inverte le disuguaglianze ($\psi_1 \leq \psi_2 \Rightarrow \psi_1^c \geq \psi_2^c$). Se ψ è c -concava, allora

$$\exists \chi : \quad \psi = \chi^c; \quad \psi^{cc} = \chi^{ccc} = (\chi^{cc})^c \leq \chi^c = \psi. \quad (\text{A.2})$$

Nella (A.2) si è tenuto conto del fatto generale $\chi^{cc} \geq \chi$ e dell'inversione delle disuguaglianze. \square

Definizione A.3. Data $\psi : X \rightarrow \overline{\mathbb{R}}$ e $x \in X$ diciamo che

$$y \in \partial_c \psi(x) \Leftrightarrow \psi(x) > -\infty, c(x, y) < +\infty, c(x, y) - \psi(x) = \min_{x' \in X} c(x', y) - \psi(x')$$

L'insieme $\partial_c(x)$ è detto c -superdifferenziale di ψ nel punto x . Con $G\partial_c \psi$ indicheremo invece il grafico della multifunzione $x \mapsto \partial_c \psi(x)$.

Osservazione A.1. Se $c(x, y) \in \mathbb{R}$, $\psi(x) > -\infty$, allora $y \in \partial_c \psi(x)$ se e solo se vale l'uguaglianza $\psi^c(y) = c(x, y) - \psi(x)$ (in generale vale solo una disuguaglianza).

Osservazione A.2. Se ψ è c -concava allora $\psi^{cc} = \psi$, pertanto, se $c(x, y) \in \mathbb{R}$, $\psi(x) > -\infty$, $\psi^c(y) > -\infty$,

$$y \in \partial_c \psi(x) \Leftrightarrow x \in \partial_c \psi^c(y)$$

Questo non difficile teorema ci servirà.

Teorema A.2. Se $(\psi_h)_h$ è una successione di funzioni c -concave e $\psi_h \rightarrow \psi$ uniformemente per $h \rightarrow +\infty$, allora ψ è c -concava.

Dimostrazione. Ci basterà dimostrare che per una generica successione $(\chi_h)_h$ tale che $\chi_h \rightarrow \chi$ uniformemente vale $\chi_h^c \rightarrow \chi^c$ uniformemente. Applicandolo poi con $\chi_h = \psi_h$ e successivamente con $\chi_h = \psi_h^c$, si otterrà la convergenza uniforme delle biconiugate delle ψ_h a ψ^{cc} . Infine, dal fatto che ψ è c -concava se e solo se $\psi^{cc} = \psi$, sapendo che ciò vale per le approssimanti, si otterrà la tesi. Valgono, ricordando la definizione di χ^c , le seguenti disuguaglianze

$$\begin{aligned} \chi_h^c(y) &\leq \inf_x c(x, y) - \chi(x) + \|\chi - \chi_h\|_\infty = \chi^c(y) + \|\chi - \chi_h\|_\infty \\ \chi_h^c(y) &\geq \inf_x c(x, y) - \chi(x) - \|\chi - \chi_h\|_\infty = \chi^c(y) - \|\chi - \chi_h\|_\infty, \end{aligned}$$

da cui $\|\chi^c - \chi_h^c\|_\infty \leq \|\chi - \chi_h\|_\infty$, che ci permette di concludere. \square

Esempio A.1. Sia $X = Y$ dotato di una struttura di spazio metrico e $c(x, y) = d(x, y)$. Allora le funzioni c -concave su X sono tutte e sole le funzioni 1-lipschitziane.

Infatti se ψ è c -concava è estremo inferiore di una famiglia di funzioni 1-lipschitziane, quelle del tipo $d(x, y) - \chi(y)$ a y fissato, pertanto anch'essa lo è. Viceversa se ψ è 1-lipschitziana allora vale

$$\psi(x) = \min_{y \in X} d(x, y) + \psi(y). \quad (\text{A.3})$$

Questo mostra anche che in questo caso $\psi^c = -\psi$.

A.1.2 c -monotonia e funzioni c -concave

Teorema A.3 (Ruschendorf). *Supponiamo che c sia a valori in \mathbb{R} e sia $\emptyset \neq \Gamma \subset X \times Y$ un insieme c -monotono. Allora esiste una funzione c -concava $\psi : X \rightarrow [-\infty, +\infty[$: $\Gamma \subset G\partial_c\psi$.*

Dimostrazione. Fissiamo $(x_0, y_0) \in \Gamma$ e definiamo

$$\psi(x) = \inf_{\substack{n \geq 1 \\ ((x_i, y_i))_{i=1..n} \subset \Gamma}} c(x, y_n) - c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0) \quad (\text{A.4})$$

Che ψ non assuma valore $+\infty$ deriva dal fatto che Γ è non vuoto. Si noti che, per c -monotonia, $\psi(x_0) \geq 0$, inoltre scegliendo $n = 1$ e $x_1 = x_0, y_1 = y_0$ si ottiene anche $\psi(x_0) \leq 0$ e quindi $\psi(x_0) = 0$. Ora si consideri una coppia $(x, y) \in \Gamma$ e un punto $x' \in X$. Per definizione di ψ , per ogni $\varepsilon > 0$ esistono un numero n e una n -upla $((x_i, y_i))_{i=1..n} \subset \Gamma$ tali che

$$c(x, y_n) - c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0) < \psi(x) + \varepsilon$$

Allora si avrà

$$\begin{aligned} \psi(x') &\leq c(x', y) - c(x, y) + c(x, y_n) - c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0) \\ &< c(x', y) - c(x, y) + \psi(x) + \varepsilon \end{aligned}$$

In particolare, scegliendo $x' = x_0$ si scopre che $\psi(x) > -\infty$, inoltre, data l'arbitrarietà di ε , si ottiene anche

$$\psi(x') \leq c(x', y) - c(x, y) - \psi(x) \quad \forall x' \in X,$$

relazione che ci dice $y \in \partial_c\psi(x)$. Infine, che ψ sia c -concava viene naturalmente dalla definizione data tramite la (A.4), scegliendo

$$\chi(y) = \sup_{\substack{n \geq 1 \\ ((x_i, y_i))_{i=1..n} \subset \Gamma \\ y_n = y}} -c(x_n, y_n) + \dots + c(x_1, y_0) - c(x_0, y_0)$$

□

A.2 Convessità

A.2.1 Convesse coniugate e sottodifferenziale

Siano X, Y spazi vettoriali topologici localmente convessi in dualità tramite un'applicazione bilineare \cdot . Si può allora scegliere di considerare il concetto di c -concavità rispetto alla funzione $c(x, y) = x \cdot y$. A meno dei segni (concavità invece di convessità), la teoria che ne verrebbe fuori sarebbe essenzialmente quella delle funzioni convesse.

Limitiamoci al caso di spazi normati, con $Y = X^*$; grazie al teorema di Hahn-Banach ed alla sua formulazione geometrica, si dimostra il seguente.

Teorema A.4. *Le funzioni convesse e s.c.i. da X in $\mathbb{R} \cup \{+\infty\}$ sono tutte e sole quelle che si esprimono come estremo superiore di funzioni affini, della forma cioè $x \mapsto \varphi \cdot x + b$, $\varphi \in X^*$, $b \in \mathbb{R}$.*

Questo mostra come, a meno di scambiare gli inf coi sup, le funzioni convesse s.c.i. possano essere inquadrare nella teoria della sezione precedente.

Definizione A.4. Data una funzione $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ si dice *dominio* di f l'insieme $D(f) = \{x \in X \mid f(x) < +\infty\}$; la funzione f si dice *propria* se $D(f) \neq \emptyset$.

Le seguenti definizioni sono casi particolari di quanto incontrato nello studio della c -concavità.

Definizione A.5. Data una funzione propria f si definisce sua *coniugata* (o polare) la funzione $f^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ data da

$$f^*(\varphi) = \sup_x \varphi \cdot x - f(x).$$

Si può vedere, usando il fatto che f si minora con una funzione affine, che anche f^* è propria. Inoltre f^* è sempre una funzione convessa s.c.i..

Esempio A.2. Dato $A \subset X$, definiamo allora la sua funzione indicatrice tramite

$$\delta(x|A) = \begin{cases} 0 & \text{se } x \in A \\ +\infty & \text{se } x \notin A \end{cases}.$$

Tale funzione è convessa se e solo se lo è A e s.c.i. se e solo se A è chiuso. La coniugata della funzione $\delta(\cdot|A)$ è la funzione

$$\delta^*(x|A) = \sup_{y \in A} x \cdot y,$$

che viene chiamata funzione di supporto di A .

Quest'ultima funzione è positivamente 1-omogenea, secondo la definizione che ora daremo:

Definizione A.6. Una funzione $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ si dice *positivamente 1-omogenea* (o, per brevità, omogenea) se per ogni $t > 0$ e ogni $x \in X$ vale $f(tx) = tf(x)$.

È facile vedere che, oltre al fatto che le coniugate delle funzioni indicatrici sono funzioni positivamente 1-omogenee, vale anche che le coniugate delle omogenee sono funzioni indicatrici, precisamente, se f è omogenea, vale

$$f^* = \delta(\cdot | \{y | x \cdot y \leq f(x) \forall x\}).$$

È chiaro che $\delta^*(x|A) = \delta^*(x|\text{conv}(\bar{A}))$, quindi non c'è speranza di identificare l'insieme A dalla sua funzione di supporto (o, a maggior ragione dalla biconiugata $\delta^{**}(\cdot | A)$), a meno che non ci si limiti ai convessi chiusi. Ciò è evidente se si tiene conto del fatto che la *-coniugata svolge evidentemente il ruolo che nello studio della c -concavità aveva la c -coniugata, pertanto vale sicuramente questo teorema, come corollario del teorema A.1.

Teorema A.5. Sia $J : X \rightarrow X^{**}$ l'immersione canonica nel bidual; allora data $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ vale

$$f = f^{**} \circ J \quad \Leftrightarrow \quad f \text{ è convessa e s.c.i..}$$

Quando lavoreremo in uno spazio di Hilbert sottoinderemo l'identificazione J di tale spazio col proprio duale.

Definizione A.7. Sia $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ e $x_0 \in X$ definiamo il *sottodifferenziale* di f in x_0 come l'insieme

$$\partial f(x_0) = \{\varphi \in X^* | f(x) \geq f(x_0) + \varphi \cdot (x - x_0) \forall x \in X\}$$

e diciamo che f è *sottodifferenziabile* in x_0 se $\partial f(x_0) \neq \emptyset$.

È facile vedere che l'insieme $\partial f(x_0)$ è un convesso chiuso (e debolmente chiuso) di X^* .

Teorema A.6. Se $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ è convessa e continua in un punto $x_0 \in D(f)$ allora il suo sottodifferenziale è non vuoto in ogni punto di $\text{int}(D(f))$.

Questo teorema ci garantisce la sottodifferenziabilità in molti punti di $D(f)$, è invece chiaro che il sottodifferenziale è vuoto in ogni punto x_0 tale che $f(x_0) = +\infty$, a meno che f stessa non sia costantemente $+\infty$, cosa esclusa se ci si limita a considerare funzioni proprie.

Se la funzione f è differenziabile in un punto x_0 è facile caratterizzare il sottodifferenziale di f in tale punto.

Teorema A.7. *Se esiste $\nabla f(x_0) \in X^*$ allora $\partial f(x_0) = \{\nabla f(x_0)\}$.*

Dimostrazione. La differenziabilità in x_0 ci dice che $x_0 \in \text{int}(D(f))$, pertanto in x_0 il sottodifferenziale è non vuoto. Ci basta mostrare che può contenere solo il differenziale di f . Sia $p \in \partial f(x_0)$: allora per ogni $y \in X$, $t > 0$

$$\begin{aligned} f(x_0 + ty) &\geq f(x_0) + p \cdot ty, \\ \frac{f(x_0 + ty) - f(x_0)}{t} &\geq p \cdot y \end{aligned}$$

Passando al limite in $t \rightarrow 0$ si ottiene

$$\nabla f(x_0) \cdot y \geq p \cdot y,$$

il che implica, scambiando y con $-y$, anche l'altra disuguaglianza ed infine $p = \nabla f(x_0)$. \square

Anche il sottodifferenziale gode di buone proprietà di dualità, come si vede dal seguente.

Teorema A.8. *Siano $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$, $x_0 \in X$, $\varphi \in X^*$, con f convessa e s.c.i.; allora*

$$\varphi \in \partial f(x_0) \Leftrightarrow J(x_0) \in \partial f^*(\varphi).$$

La nozione di sottodifferenziale risulta molto utile nello studio dei minimi perché si nota facilmente che una funzione ha minimo assoluto in x_0 se e solo se 0 appartiene al suo sottodifferenziale in x_0 .

Potrebbe risultare utile, per studiare i minimi di funzioni ottenute sommando funzioni convesse a funzioni regolari, la seguente nozione.

Definizione A.8. Definiamo il sottodifferenziale asintotico di f in x_0 come

$$\tilde{\partial} f(x_0) = \left\{ \varphi \in X^* \mid \liminf_{t \rightarrow 0^+} \frac{f(x_0 + t(x - x_0)) - f(x_0)}{t} \geq \varphi \cdot (x - x_0) \forall x \in X \right\}.$$

Usando la monotonia dei rapporti incrementali lungo ogni retta è possibile vedere che $\tilde{\partial} f = \partial f$ quando f è convessa, mentre in generale invece potrebbe essere $\partial f = \emptyset$, $\tilde{\partial} f \neq \emptyset$.

Sia ora $f = f_1 + f_2$, con f_1 convessa, f_2 differenziabile in x_0 : allora

$$\tilde{\partial} f(x_0) = \nabla f_2(x_0) + \partial f_1(x_0). \tag{A.5}$$

Pertanto, poiché laddove f ha minimo deve essere $0 \in \tilde{\partial} f$, se x_0 è punto di minimo assoluto per f , deve essere $\nabla f_2(x_0) \in \partial f_1(x_0)$.

A.2.2 Funzione di recessione e inf-convoluzione

D'ora in avanti per semplicità ci limiteremo a $X = \mathbb{R}^n$.

Definizione A.9. Data $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ definiamo la sua *funzione di recessione*

$$f^\infty(x) = \lim_{t \rightarrow +\infty} \frac{f(x_0 + tx)}{t},$$

dove x_0 è un qualunque punto del dominio di f .

Si può vedere che la definizione non dipende dalla scelta di $x_0 \in D(f)$ (si veda [12]) e che la funzione f^∞ risulta convessa, s.c.i., positivamente 1-omogenea. Quest'ultima proprietà è evidente dalla definizione, mentre convessità e semicontinuità seguono da

$$\lim_{t \rightarrow +\infty} \frac{f(x_0 + tx)}{t} = \lim_{t \rightarrow +\infty} \frac{f(x_0 + tx) - f(x_0)}{t} = \sup_{t > 0} \frac{f(x_0 + tx) - f(x_0)}{t}. \quad (\text{A.6})$$

Definizione A.10. Date due funzioni $\alpha, \beta : \mathbb{R}^n \rightarrow [0, +\infty]$ definiamo l'*inf-convoluzione* di α e β come

$$\alpha \square \beta(x) = \inf_y \alpha(y) + \beta(x - y).$$

Esempio A.3. Per ogni f convessa s.c.i. e propria vale

$$f \square f^\infty = f.$$

Infatti, poiché $f^\infty(0) = 0$, vale di sicuro $f \square f^\infty(x) \leq f(x)$. Per mostrare l'altra disuguaglianza si può, fissato un $y \in D(f)$, usare la (A.6) per vedere che

$$f(y) + f^\infty(x - y) = f(y) + \sup_{t > 0} \frac{f(y + t(x - y)) - f(y)}{t} \geq f(y) + f(x) - f(y) = f(x).$$

La disuguaglianza trovata è banalmente vera se $f(y) = +\infty$ e, passando all'inf in y , si trova quella fra l'inf-convoluzione e f .

Teorema A.9. Siano $\alpha, \beta, \beta_h : \mathbb{R}^n \rightarrow [0, +\infty]$, allora

1. se α, β sono convesse anche $\alpha \square \beta$ lo è;
2. $(\alpha \square \beta)^* = \alpha^* + \beta^*$;
3. se α, β sono s.c.i. e $\lim_{|x| \rightarrow +\infty} \alpha(x) = +\infty$ allora $\alpha \square \beta$ è s.c.i.;
4. se $\beta_h \nearrow \beta$, le β_h sono s.c.i. e $\lim_{|x| \rightarrow +\infty} \alpha(x) = +\infty$, allora $\alpha \square \beta_h \nearrow \alpha \square \beta$.

Dimostrazione. **1)** Presi due punti $x_0, x_1 \in \mathbb{R}^n$ ed $\varepsilon > 0$ si trovino dei punti y_0, y_1 tali che

$$\alpha(y_0) + \beta(x_0 - y_0) < \alpha \square \beta(x_0) + \varepsilon; \quad \alpha(y_1) + \beta(x_1 - y_1) < \alpha \square \beta(x_1) + \varepsilon.$$

Preso ora $t \in (0, 1)$ abbiamo

$$\begin{aligned} \alpha \square \beta((1-t)x_0 + tx_1) & \leq \alpha((1-t)y_0 + ty_1) + \beta((1-t)x_0 + tx_1 - (1-t)y_0 - ty_1) \\ & \leq (1-t)\alpha(y_0) + t\alpha(y_1) + (1-t)\beta(x_0 - y_0) + t\beta(x_1 - y_1) \\ & < (1-t)\alpha \square \beta(x_0) + (1-t)\varepsilon + t\alpha \square \beta(x_1) + t\varepsilon. \end{aligned}$$

L'arbitrarietà di $\varepsilon > 0$ dà la tesi.

2) Si noti che

$$\begin{aligned} (\alpha \square \beta)^*(p) & = \sup_y p \cdot y - \alpha \square \beta(y) = \sup_y p \cdot y + \sup_z -\alpha(z) - \beta(y - z) = \\ & = \sup_{y,z} p \cdot y - \alpha(z) - \beta(y - z) \\ & = \sup_{y-z, z} p \cdot (y - z) + p \cdot z - \alpha(z) - \beta(y - z) = \\ & = \alpha^*(p) + \beta^*(p). \end{aligned}$$

3) Supponiamo per assurdo che esista una successione $(x_h)_h$ tale che $x_h \rightarrow x$, ma $\lim_h \alpha \square \beta(x_h) < \alpha \square \beta(x)$. Scegliamo y_h tali che

$$\alpha(y_h) + \beta(x_h - y_h) < \alpha \square \beta(x_h) + \frac{1}{h}.$$

L'ipotesi su α ci garantisce che la successione $(y_h)_h$ è limitata, pertanto possiamo supporre che converga ad un certo y ; usando la semicontinuità di α e β si ha allora

$$\begin{aligned} \alpha \square \beta(x) & \leq \alpha(y) + \beta(x - y) \leq \liminf_h \alpha(y_h) + \beta(x_h - y_h) \\ & \leq \liminf_h \alpha \square \beta(x_h) + \frac{1}{h} = \lim_h \alpha \square \beta(x_h) \\ & < \alpha \square \beta(x), \end{aligned}$$

il che è chiaramente assurdo.

4) poiché l'operazione di inf-convoluzione è chiaramente crescente nei suoi argomenti, ci basta dimostrare $\sup_h \alpha \square \beta_h(x) \geq \alpha \square \beta(x)$. Fissiamo un x tale che il primo membro della disuguaglianza che vogliamo mostrare sia finito. Allora troviamo degli y_h e un $M < +\infty$ tali che

$$\alpha(y_h) + \beta_h(x - y_h) < \alpha \square \beta_h(x) + \frac{1}{h} \leq M.$$

Per l'ipotesi fatta su α la successione (y_h) risulterà limitata e pertanto possiamo supporre che converga a y . Allora per ogni k , usando la semicontinuità di α e β_k , si ha

$$\begin{aligned} \alpha(y) + \beta_k(x - y) &\leq \liminf_h [\alpha(y_h) + \beta_k(x - y_h)] \\ &\leq \liminf_h \alpha(y_h) + \beta_h(x - y_h) \leq \liminf_h \alpha \square \beta_h(x) + \frac{1}{h} \\ &= \sup_h \alpha \square \beta_h(x). \end{aligned}$$

Passando prima al sup in k e poi minorando $\alpha(y) + \beta(x - y)$ con $\alpha \square \beta(x)$ si ottiene la tesi. \square

A.2.3 Regolarità di funzioni convesse in \mathbb{R}^n

Consideriamo funzioni convesse a valori reali definite in un aperto convesso $\Omega \subset \mathbb{R}^n$, per ottenerne la continuità useremo il seguente.

Lemma A.10. *Sia X uno spazio vettoriale topologico, $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ convessa, $x_0 \in X$. allora f è continua in x_0 se e solo se è limitata in un suo intorno.*

Corollario A.11. *Se $f : \Omega \rightarrow \mathbb{R}$ è convessa allora è continua.*

Dimostrazione. Basta trovare, dato un punto $x_0 \in \Omega$ un suo intorno in cui f sia limitata. Per farlo basta scegliere dei punti $(x_i)_{i=1..n+1}$ che generino un semplice non degenero contenente x_0 nella sua parte interna e contenuto in Ω , il che è naturalmente possibile visto che siamo in un aperto. Sia V la parte interna di questo semplice: ogni punto x di V è combinazione convessa dei punti x_i , pertanto, usando la convessità di f , si può stimare $f(x)$ con una media pesata delle $f(x_i)$ e a maggior ragione con $\max_i f(x_i)$. \square

Consideriamo ora la multifunzione $x \mapsto \partial f(x)$, che sappiamo essere a valori convessi chiusi. In questo caso è addirittura a valori compatti e per mostrarlo basta la seguente stima.

Lemma A.12. *Se $B(x_0, 2r) \subset \Omega$ allora per ogni $x \in B(x_0, r)$ il sottodifferenziale di f in x è costituito da vettori di norma minore di $\text{osc}(f, B(x_0, 2r))/r$, dove con osc si intende l'oscillazione ($\sup - \inf$).*

Dimostrazione. Sia $p \in \partial f(x)$: allora, dalla definizione di sottodifferenziale, scegliendo $y = x + rp/|p|$, si ha

$$r|p| + f(x) \leq f\left(x + r\frac{p}{|p|}\right) \leq \max_{y \in \partial B(x, r)} f(y) < +\infty,$$

Stimando dall'alto tale massimo con l'estremo superiore su tutta $B(x_0, 2r) \supset \partial B(x, r)$ e dal basso $f(x)$ con l'inf sulla stessa palla si ottiene la tesi. \square

Questa stima è ovviamente utile solo quando le quantità che richiama sono finite, cosa che si verifica grazie alla continuità di f se $\overline{B(x, 2r)} \subset \Omega$.

Definizione A.11. Una *multifunzione* u da \mathbb{R}^n nei sottoinsiemi di \mathbb{R}^n si dice *monotona* se

$$(y_1 - y_2) \cdot (x_1 - x_2) \geq 0 \quad \forall x_i \in \mathbb{R}^n, y_i \in u(x_i), i = 1, 2.$$

Dalla definizione di sottodifferenziale si vede facilmente che $x \mapsto \partial f(x)$ è sempre una multifunzione monotona.

Indicando con Gu il grafico della multifunzione u diciamo che $u \subset v$ per due multifunzioni se $Gu \subset Gv$, cioè se vale $u(x) \subset v(x)$ per ogni x .

Definizione A.12. Una *multifunzione monotona* si dice *massimale* se lo è rispetto all'inclusione nella classe delle multifunzioni monotone, cioè se

$$v \supset u, v \text{ monotona} \Rightarrow v = u.$$

Si può anche vedere che la multifunzione sottodifferenziale di una funzione convessa s.c.i. è massimale. Questo fatto si trova dimostrato in [2], insieme a questo teorema, che ci permette di concludere l'argomento della differenziabilità delle funzioni convesse.

Teorema A.13. *Sia u una multifunzione monotona massimale. Allora $u(x)$ è un convesso chiuso eventualmente vuoto per ogni x e pertanto ha una sua dimensione $0 \leq k \leq n$. Inoltre, detto*

$$\Sigma^k(u) = \{x \in \mathbb{R}^n \mid \dim(u(x)) \geq k\},$$

allora per ogni $k = 1, \dots, n$ l'insieme $\Sigma^k(u)$ è numerabilmente \mathcal{H}^{n-k} -rettificabile, cioè si trovano una quantità numerabile di sottovarietà C^1 $\Gamma_i \subset \mathbb{R}^n$ di dimensione $n - k$ tali che

$$\mathcal{H}^{n-k} \left(\Sigma^k(u) \setminus \bigcup_i \Gamma_i \right) = 0.$$

Corollario A.14. *Sia $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convessa. Allora f è differenziabile fuori da un insieme \mathcal{H}^{n-1} -rettificabile.*

Dimostrazione. È sufficiente dimostrare che una funzione convessa è differenziabile in x_0 se il sottodifferenziale in x_0 di f è un singoletto, in quanto il teorema precedente ci garantisce che, fuori da un insieme \mathcal{H}^{n-1} -rettificabile, il sottodifferenziale ha dimensione nulla, e pertanto, essendo un insieme convesso, è costituito da un singolo vettore. Per vedere questo definiamo

$$g(y) = f(x_0 + y) - f(x_0) - p \cdot y, \quad h_\lambda(y) = \frac{g(\lambda y)}{\lambda}.$$

L'uso della funzione g ha il solo vantaggio di ricondurci ad una funzione nulla nell'origine, con un minimo assoluto in tale punto poiché lì l'unico vettore nel sottodifferenziale è quello nullo. Dobbiamo far vedere che $h_\lambda : \mathbb{R}^n \rightarrow \mathbb{R}$ converge uniformemente a 0 per $\lambda \rightarrow 0^+$ su $B(0, 1)$. Questa stima (che come il resto dell'idea di questa dimostrazione è presa da [12]), ci permette di ricondurci a dimostrare la convergenza puntuale. Siano $(y_i)_{i=0..n}$ punti nel cui involuppo convesso sia contenuta la palla unitaria. Ognuna delle funzioni h_λ eredita dalla g , la quale a sua volta la eredita dalla f , la proprietà di essere convessa, inoltre per ogni $y \in B(0, 1)$ esistono dei coefficienti $a_i \in (0, 1) : \sum_{i=0}^n a_i = 1$ che permettono di esprimere y come combinazione convessa degli y_i , quindi

$$h_\lambda(y) = h_\lambda\left(\sum_{i=0}^n a_i y_i\right) \leq \sum_{i=0}^n a_i h_\lambda(y_i) \leq \max_i h_\lambda(y_i).$$

Si conclude pertanto usando il successivo lemma, che garantisce appunto la convergenza a 0 dei rapporti incrementali della g in direzione degli y_i . \square

Lemma A.15. *Sia $g : \mathbb{R}^n \rightarrow \mathbb{R}$ convessa, $g(0) = 0$, $\partial g(0) = \{0\}$. Allora per ogni $y_0 \in \mathbb{R}^n$ vale*

$$\lim_{\lambda \rightarrow 0^+} \frac{g(\lambda y_0)}{\lambda} = 0.$$

Dimostrazione. Consideriamo la funzione $g' : \mathbb{R}^n \rightarrow \mathbb{R}$ data da

$$g'(x) = \lim_{\lambda \rightarrow 0^+} \frac{g(\lambda x)}{\lambda}.$$

Essa è ben definita perché sappiamo che i rapporti incrementali di funzioni convesse in \mathbb{R} sono crescenti (qui stiamo restringendo g alla retta generata da ciascun x) e pertanto hanno un limite, inoltre per le ipotesi fatte su g tale limite è finito e non negativo. Inoltre la crescita dei rapporti incrementali, confrontando il limite per $\lambda \rightarrow 0^+$ col valore per $\lambda = 1$ ci dà $g' \leq g$. Mostriamo che g' è convessa: siano dati $x_0, x_1 \in \mathbb{R}^n$, $t \in (0, 1)$; allora

$$\frac{g(\lambda((1-t)x_0 + tx_1))}{\lambda} \leq \frac{(1-t)g(\lambda x_0) + tg(\lambda x_1)}{\lambda}.$$

Passando al limite per $\lambda \rightarrow 0^+$ si ottiene $g'((1-t)x_0 + tx_1) \leq (1-t)g'(x_0) + tg'(x_1)$, che ci dà la convessità cercata. Inoltre il corollario A.11 ci garantisce la continuità di g' . Noi vorremmo dimostrare che g' è identicamente nulla e lo faremo trovandone la convessa coniugata. poiché si vede facilmente che g' è positivamente 1-omogenea la sua coniugata sarà la funzione indicatrice $\delta(\cdot | A)$ di un certo convesso chiuso non vuoto $A \subset \mathbb{R}^n$; per tali funzioni vale chiaramente $0 \in \partial \delta(\cdot | A)(x)$ per ogni $x \in A$, poiché in tali punti la funzione indicatrice è finita ed ha un minimo. Tuttavia noi sappiamo che l'unico vettore nel sottodifferenziale in g' è lo zero, infatti siccome vale $g(0) = g'(0) = 0$ e $g' \leq g$

allora $\partial g'(0) \subset \partial g(0)$. Per il teorema A.8 allora l'unico punto in cui 0 può appartenere al sottodifferenziale di $g'^* = \delta(\cdot | A)$ è lo zero stesso. Allora $A = \{0\}$. Da questo si ricava facilmente $g' = 0$, sfruttando $g' = g'^{**}$ e calcolando la coniugata dell'indicatrice dello zero. \square

La dimostrazione del fatto che ci si può ridurre a dimostrare la convergenza puntuale dei rapporti incrementali, come già accennato, viene da [12]; per dimostrare invece tale convergenza puntuale vi sono varie vie, di cui quella che ho presentato era forse la più rapida in questo contesto. Per una dimostrazione valida in un generico spazio normato (noi abbiamo usato la dimensione finita per garantire la continuità, e quindi la semicontinuit , di g') si veda ad esempio [1], laddove si affronta il legame fra sottodifferenziale e derivate di Gateaux.

Per la multifunzione sottodifferenziale   facile mostrare una propriet  di continuit :

Teorema A.16. *Sia $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convessa e s.c.i. e siano $(x_h)_h, (y_h)_h$ successioni in \mathbb{R}^n tali che $x_h \rightarrow x, y_h \in \partial f(x_h), y_h \rightarrow y$. Allora $y \in \partial f(x)$.*

Dimostrazione. La definizione di sottodifferenziale ci d 

$$f(z) \geq f(x_h) + y_h \cdot (z - x_h).$$

Passando al limite e usando la semicontinuit  di f si ottiene

$$f(z) \geq f(x) + y \cdot (z - x).$$

\square

Corollario A.17. *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$   convessa e differenziabile in tutti i punti allora   anche C^1 .*

Dimostrazione. poich  se f   differenziabile in un punto x_0 con differenziale p allora $\partial f(x_0) = \{p\}$, il teorema precedente rende evidente la continuit  del differenziale purch  si mostri che si pu  estrarre una sottosuccessione convergente dai differenziali di una successione di punti convergente. Per questo   sufficiente il lemma A.12. \square

A.3 Multifunzioni

A.3.1 Convergenza di Hausdorff

Sia X uno spazio metrico e $A, B \subset X$ due sottoinsiemi chiusi: vogliamo definire una distanza fra di essi.

Definizione A.13. Definiamo l'eccesso di A rispetto a B come

$$e(A; B) = \sup_{x \in A} d(x, B),$$

e analogamente l'eccesso di B rispetto ad A . Definiamo la *distanza di Hausdorff* fra A e B come

$$\mathcal{H}(A, B) = e(A; B) \vee e(B; A).$$

Si verifica facilmente che \mathcal{H} è una distanza, eventualmente troncadola per evitare che assuma il valore $+\infty$, che quindi dà una struttura di spazio metrico all'insieme dei chiusi di X , spazio che indicheremo con $\mathcal{P}_F(X)$. Restringendoci a classi meno ampie di insiemi possiamo anche considerare gli spazi metrici $\mathcal{P}_K(X)$, costituito dai compatti di X , $\mathcal{P}_B(X)$, dai chiusi limitati e $\mathcal{P}_{TB}(X)$, dai chiusi totalmente limitati.

Valgono i seguenti:

Teorema A.18. *Se X è completo, lo è anche $\mathcal{P}_F(X)$.*

Teorema A.19. *$\mathcal{P}_{TB}(X)$ è chiuso in $\mathcal{P}_F(X)$.*

Corollario A.20. *Se X è completo $\mathcal{P}_K(X)$ è completo.*

Teorema A.21. *Se X è totalmente limitato anche $\mathcal{P}_F(X)$ lo è.*

Corollario A.22. *Se X è compatto anche $\mathcal{P}_F(X) = \mathcal{P}_K(X) = \mathcal{P}_{TB}(X)$ lo è.*

Teorema A.23. *Se X è separabile anche $\mathcal{P}_K(X)$ lo è.*

Teorema A.24. *La topologia di $\mathcal{P}_K(X)$ è generata dalla famiglia di insiemi costituita da $\{K \in \mathcal{P}_K(X) \mid K \subset U\}$ e $\{K \in \mathcal{P}_K(X) \mid K \cap V \neq \emptyset\}$, con U, V aperti e una base di intorni di $K_0 \in \mathcal{P}_K(X)$ è data dagli insiemi $\{K \in \mathcal{P}_k \mid K \subset U, K \cap V_1 \neq \emptyset, \dots, K \cap V_n \neq \emptyset\}$ cui appartenga K_0 , con U, V_1, \dots, V_n aperti.*

Corollario A.25. *La topologia dello spazio metrico $\mathcal{P}_K(X)$ dipende solo dalla topologia di X e non dalla sua metrica.*

Teorema A.26. *Se X è separabile, la tribù borelliana di $\mathcal{P}_K(X)$ è generata dagli insiemi $\{K \in \mathcal{P}_K(X) \mid K \subset U\}$ al variare di U aperto, ma è anche generata dagli insiemi $\{K \in \mathcal{P}_K(X) \mid K \cap V \neq \emptyset\}$ al variare di V aperto.*

Tutti questi enunciati hanno una loro dimostrazione in [9]; i teoremi A.18, A.21 anche in [4]. È naturale, avendo a disposizione una struttura metrica sull'insieme dei chiusi di X , dare maggior rilievo, fra le multifunzioni a valori in X (cioè le funzioni a valori in $P(X)$), a quelle a valori chiusi. Quando sarà possibile (X spazio vettoriale topologico) ci limiteremo addirittura a quelle a valori convessi. Entrambe queste scelte

sono pienamente giustificate alla luce di ciò che finora ci è capitato di incontrare: la multifunzione sottodifferenziale e, più in generale, le multifunzioni monotone massimali. È inutile definire che cos'è una multifunzione continua, quando essa abbia come dominio uno spazio metrico Ω e sia a valori chiusi, perché semplicemente abbiamo la nozione di funzione continua da Ω in $\mathcal{P}_F(X)$. Definiamo invece il concetto di semicontinuit , che daremo in generale, anche per multifunzioni a valori non chiusi.

Definizione A.14. Sia Γ una multifunzione da Ω nei sottoinsiemi di X e $\bar{p} \in \Omega$. Diciamo che Γ   s.c.s. in \bar{p} se per ogni U aperto che contenga $\Gamma(\bar{p})$ esiste un intorno N di \bar{p} in Ω tale che $\Gamma(p) \subset U$ per ogni $p \in N$.

Definizione A.15. Diciamo che Γ   s.c.i. in \bar{p} se per ogni V aperto che intersechi $\Gamma(\bar{p})$ esiste un intorno N di \bar{p} in Ω tale che $\Gamma(p) \cap V \neq \emptyset$ per ogni $p \in N$.

Per il teorema A.24 avremo continuit  in un punto \bar{p} se Γ  , almeno in un intorno di \bar{p} , a valori compatti ed   contemporaneamente s.c.i. e s.c.s. in \bar{p} .

Teorema A.27. *Supponiamo che esista un intorno N_0 di \bar{p} tale che $\bigcup_{p \in N_0} \Gamma(p)$ sia relativamente compatta in X . Allora sono equivalenti:*

1. Γ   s.c.s. in \bar{p} ;
2. per ogni successione $p_h \rightarrow \bar{p}$, ogni $x \in X$ e ogni scelta di $x_h \in \Gamma(p_h)$ con $x_h \rightarrow x$ vale $x \in \Gamma(\bar{p})$.

Dimostrazione. (1 \Rightarrow 2) Supponiamo per assurdo che ci siano $(p_h)_h, (x_h)_h, x$ che contraddicano la tesi e siano U un arbitrario intorno di $\Gamma(\bar{p})$. Considerato l'intorno N dato dalla definizione di semicontinuit  di Γ definitivamente $p_h \in N$, quindi definitivamente $x_h \in U$. Allora $x \in \bar{U}$. In uno spazio metrico questa condizione per U arbitrario aperto contenente $\Gamma(\bar{p})$ implica $x \in \Gamma(\bar{p})$ (sarebbe bastato essere in uno spazio topologico T_3). (2 \Rightarrow 1) Supponiamo per assurdo di trovare un aperto $U \supset \Gamma(\bar{p})$ ed una successione $p_h \rightarrow \bar{p}$ di punti per cui $\Gamma(p_h)$ non sia contenuto in U . Ci  implica l'esistenza di una successione $(x_h)_h$ con $x_h \in \Gamma(p_h) \setminus U$ e possiamo supporre, per l'ipotesi di compattezza fatta, che ci sia un certo x tale che $x_h \rightarrow x$, a meno di passare eventualmente ad una sottosuccessione. Siccome $x_h \notin U$ e U   aperto, anche $x \notin U$ e quindi non pu  nemmeno appartenere a $\Gamma(\bar{p})$, contraddicendo l'ipotesi. \square

Esempio A.4. Data $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convessa, la multifunzione ∂f definita in \mathbb{R}^n a valori nei chiusi di \mathbb{R}^n   semicontinua superiormente.

Per vederlo   sufficiente verificare le ipotesi di 2 \Rightarrow 1, e questo si pu  fare grazie al lemma A.12 e al teorema A.16

Teorema A.28. *Supponendo che la multifunzione Γ sia a valori non vuoti, le seguenti proprietà sono equivalenti:*

1. Γ è s.c.i. in \bar{p} ;
2. per ogni successione $p_h \rightarrow \bar{p}$, ogni $x \in \Gamma(\bar{p})$ esiste una successione $(x_h)_h$, con $x_h \in \Gamma(p_h)$ e $x_h \rightarrow x$ ogni scelta di $x_h \in \Gamma(p_h)$ con $x_h \rightarrow x$.

Dimostrazione. (1 \Rightarrow 2) Dati $x \in \Gamma(\bar{p})$, $p_h \rightarrow \bar{p}$, costruiamo una siffatta successione. Scegliamo $V_k = B(x, r_k)$, dove $(r_k)_k$ è una data successione infinitesima. Per ogni k l'ipotesi di semicontinuità ci dice che definitivamente in h $\Gamma(p_h)$ interseca V_k , cioè esiste $h(k)$ tale che per ogni $h \geq h(k)$ troviamo un punto $y_{h,k} \in \Gamma(p_h) \cap V_k$. Definiamo per ogni $h < h(k)$ $y_{h,k}$ in un modo qualunque purché sia un punto di $\Gamma(p_h)$. Possiamo supporre ovviamente che la successione $(h(k))_k$ sia strettamente crescente. Definiamo ora

$$x_h = y_{h, \bar{k}(h)}, \text{ dove } \bar{k}(h) = \sup \{k \mid h \geq h(k)\}.$$

Per come erano definite per ogni k le successioni $(y_{h,k})_h$, ciascuno dei punti $x_h \in \Gamma(p_h)$. Che $x_h \rightarrow x$ viene da

$$x_h \in V_{\bar{k}(h)} = B(x, r_{\bar{k}(h)})$$

e dal fatto che $\lim_h \bar{k}(h) = \sup_h \bar{k}(h) = +\infty$ perché per definizione $\bar{k}(h(k)) = k$.

(2 \Rightarrow 1) Supponiamo per assurdo che esista un aperto V che interseca $\Gamma(\bar{p})$, ma che per una successione di punti $(p_h)_h$ tale che $p_h \rightarrow \bar{p}$ valga $\Gamma(p_h) \cap V = \emptyset$. Sia $x \in \Gamma(\bar{p}) \cap V$. La successione $(x_h)_h$ data dall'ipotesi creerebbe allora un assurdo in quanto dovrebbe tendere a x restando tuttavia fuori da V , che ne è un intorno. \square

Questi teoremi ci introducono ad un'altra nozione di convergenza fra insiemi chiusi, che è nota come convergenza di Kuratowski (vedi ad esempio [4]).

Definizione A.16. Diciamo che una successione $(F_h)_h$ di chiusi di uno spazio metrico X converge ad un chiuso $F \subset X$ se valgono le seguenti condizioni:

1. se $x = \lim_k x_{h_k}$ per qualche sottosuccessione di una successione $(x_h)_h$ con $x_h \in F_h$, allora $x \in F$;
2. per ogni $x \in F$ esiste una successione $(x_h)_h$ con $x_h \in F_h$, $h \rightarrow x$.

In questi casi scriveremo $F_h \xrightarrow{\mathcal{K}} F$.

Teorema A.29. *Se $\mathcal{H}(F_h, F) \rightarrow 0$ allora $F_h \xrightarrow{\mathcal{K}} F$; il viceversa vale se lo spazio X è compatto.*

Dimostrazione. Entrambe le implicazioni si dimostrano facilmente utilizzando i teoremi A.27 e A.28 applicati ad uno spazio metrico Ω che sia la metrizzazione della topologia usuale su $\mathbb{N} \cup \{+\infty\}$. Per esempio $\Omega = \{\frac{1}{h} \mid h \in \mathbb{N}\} \cup \{0\}$ con la distanza che gli deriva dall'essere sottoinsieme di \mathbb{R} e l'identificazione con $\mathbb{N} \cup \{+\infty\}$ data dalla mappa $x \mapsto \frac{1}{x}$ estesa a $\overline{\mathbb{R}}$. L'unico problema potrebbe essere dato dal fatto che nel teorema A.27 si fa riferimento ad intere successioni, mentre la 1 della definizione A.16 richiede che la stessa condizione sia valida per ogni arbitraria sottosuccessione. Siccome però se $F_h \rightarrow F$ nel senso di Hausdorff lo fa anche F_{h_k} si può utilizzare lo stesso teorema restringendosi alla sottosuccessione. \square

A.3.2 Multifunzioni misurabili

Sia X uno spazio metrico completo e separabile: avendo dato una struttura allo spazio di arrivo di una multifunzione a valori chiusi di X ora abbiamo a disposizione la nozione di multifunzione misurabile definita su un qualunque spazio misurabile (T, \mathcal{T}) . Prenderemo tuttavia in considerazione diverse nozioni di misurabilità e le implicazioni che valgono fra di esse.

Per farlo, data una multifunzione $\Gamma : T \rightarrow X$ e un sottoinsieme $B \subset X$, indicheremo con $\Gamma^{-1}(B)$ l'insieme $\{t \in T \mid \Gamma(t) \cap B \neq \emptyset\}$ e con G il grafico di Γ , cioè l'insieme $\{(t, x) \in T \times X \mid x \in \Gamma(t)\}$. Diremo infine che un'applicazione $\sigma : T \rightarrow X$ è una selezione di Γ se per ogni t $\sigma(t) \in \Gamma(t)$ e quello che ci interesserà sarà la proprietà di Γ di ammettere selezioni che siano misurabili. Consideriamo le seguenti proprietà:

1. $\forall B$ borelliano $\Gamma^{-1}(B) \in \mathcal{T}$;
2. $\forall F$ chiuso $\Gamma^{-1}(F) \in \mathcal{T}$;
3. $\forall U$ aperto $\Gamma^{-1}(U) \in \mathcal{T}$;
4. esiste una successione σ_n di selezioni misurabili tali che $\forall t, \Gamma(t) = \overline{\{\sigma_n(t) \mid n \in \mathbb{N}\}}$;
5. $\forall x \in X$ la funzione $t \rightarrow d(x, \Gamma(t))$ è \mathcal{T} -misurabile;
6. $G \in \mathcal{T} \otimes \mathcal{B}(X)$.

Si noti che nel caso di multifunzioni a valori compatti la proprietà 3 corrisponde proprio alla nozione di misurabilità per mappe a valori nello spazio metrico $\mathcal{P}_K(X)$ munito della sua tribù borelliana, grazie al teorema A.26. Utilizzeremo comunque la stessa proprietà come definizione di multifunzione chiusa misurabile, anche nel caso di valori non compatti.

Esempio A.5. Se (T, \mathcal{T}) è uno spazio metrico munito della propria tribù borelliana, ogni multifunzione s.c.i. a valori compatti è misurabile.

Questo viene facilmente verificano una proprietà addirittura più forte della 3: infatti per ogni U aperto $\Gamma^{-1}(U)$ è esso stesso aperto. È vero anche il seguente, solo leggermente più complicato.

Esempio A.6. Se (T, \mathcal{T}) è uno spazio metrico munito della propria tribù borelliana, ogni multifunzione s.c.s. a valori compatti è misurabile. Infatti per ogni chiuso F $\Gamma^{-1}(F) = \{t \in T \mid \Gamma(t) \subset F^c\}^c$ ed è pertanto aperto.

Fra le proprietà 1, . . . ,6 descritte prima valgono queste implicazioni, dimostrate su [9]:

$$1 \Rightarrow 2 \Rightarrow 3 \Leftrightarrow 4 \Leftrightarrow 5 \Rightarrow 6.$$

Risultano invece tutte equivalenti sotto l'ulteriore ipotesi che la tribù \mathcal{T} sia completa rispetto ad un'opportuna misura σ -finita su (T, \mathcal{T}) , in quanto questo permette di dimostrare che $6 \Rightarrow 1$, per mezzo del teorema di proiezione, di cui enunciamo una versione piuttosto debole.

Teorema A.30 (di proiezione). *Sia (T, \mathcal{T}, μ) uno spazio di misura completo e X uno spazio metrico completo e separabile. Allora se un sottoinsieme $H \subset T \times X$ appartiene alla tribù $\mathcal{T} \otimes \mathcal{B}(X)$ la sua proiezione sul fattore T appartiene a \mathcal{T} .*

Noi dimostreremo soltanto l'equivalenza fra le proprietà 3,4 e 5 (ma $1 \Rightarrow 2$ è banale e $2 \Rightarrow 3$ segue facilmente dal fatto che ogni aperto è unione numerabile di chiusi e dal comportamento dell'operatore controimmagine Γ^{-1} per unioni).

Teorema A.31. *Sia (T, \mathcal{T}) uno spazio misurabile, X uno spazio metrico completo e separabile e Γ una multifunzione su T a valori chiusi non vuoti di X . Allora le proprietà 3,4 e 5 sono equivalenti.*

Dimostrazione. ($4 \Rightarrow 5$) Segue da

$$d(x, \Gamma(t)) = \inf_n d(x, \sigma_n(t))$$

e dal fatto che tutte le funzioni $d(x, \sigma_n(\cdot))$ sono misurabili come conseguenza dell'ipotesi. ($5 \Rightarrow 3$) Se U è una palla $B(x, r)$ la tesi segue da

$$\Gamma^{-1}(B(x, r)) = \{t \mid d(x, \Gamma(t)) < r\}.$$

In generale U sarà un'unione numerabile di palle B_n , grazie all'ipotesi di separabilità e $\Gamma^{-1}(U) = \bigcup_n \Gamma^{-1}(B_n)$.

($3 \Rightarrow 4$) Utilizzeremo il lemma A.32 che ci permetterà di trovare, sempre sotto l'ipotesi 3, quantomeno una selezione misurabile. Ora dimostriamo che da tale lemma segue l'esistenza di una quantità numerabile di selezioni misurabili con la proprietà cercata.

sia $(x_n)_n$ una famiglia densa e numerabile di punti di X e per ogni (n, i) costruiamo una multifunzione a valori chiusi non vuoti utilizzando la seguente multifunzione:

$$\Gamma_{n,i}(t) = \begin{cases} \Gamma(t) \cap B(x_n, 2^{-i}) & \text{se } t \in \Gamma^{-1}(B(x_n, 2^{-i})) \\ \Gamma(t) & \text{altrimenti,} \end{cases}$$

e considerando $\Theta_{n,i} = \overline{\Gamma_{n,i}}$. Per ogni aperto $U \subset X$ vale

$$\begin{aligned} \Theta_{n,i}^{-1}(U) &= \{t \mid \overline{\Gamma_{n,i}(t)} \cap U \neq \emptyset\} = \{t \mid \Gamma_{n,i}(t) \cap U \neq \emptyset\} \\ &= \Gamma^{-1}(B(x_n, 2^{-i}) \cap U) \cup [\Gamma^{-1}(U) \cap \Gamma^{-1}(B(x_n, 2^{-i})^c)] \in \mathcal{T} \end{aligned}$$

Pertanto la multifunzione $\Theta_{n,i}$ soddisfa l'ipotesi del lemma A.32 e ammette una selezione misurabile $\sigma_{n,i}$. Bisogna solo dimostrare che $\Gamma(t)$ coincide con $\overline{\{\sigma_{n,i}(t) \mid (n, i) \in \mathbb{N}^2\}}$. Sia allora $x \in \Gamma(t)$, $\varepsilon > 0$, $i : 2^{-i} < \varepsilon/2$, $n : d(x_n, x) < 2^{-i}$. Allora $\Gamma(t)$ interseca $B(x_n, 2^{-i})$ perché il punto x appartiene a tutti e due e quindi $\sigma_{n,i}(t) \in \Theta_{n,i}(t) \subset \overline{B(x_n, 2^{-i})}$. allora $d(\sigma_{n,i}(t), x) \leq d(\sigma_{n,i}(t), x_n) + d(x_n, x) < \varepsilon$. \square

Lemma A.32. *Sia (T, \mathcal{T}) uno spazio misurabile, X uno spazio metrico completo e separabile e Γ una multifunzione su T a valori chiusi non vuoti di X che goda della proprietà 3. Allora Γ ammette una selezione misurabile σ .*

Dimostrazione. Consideriamo una famiglia densa e numerabile $(x_n)_n$ in X e costruiamo una successione di applicazioni misurabili $\sigma_k : (T, \mathcal{T}) \rightarrow (X, \mathcal{B}(X))$ assumenti valore fra i punti della famiglia ed aventi queste proprietà:

$$d(\sigma_k(t), \Gamma(t)) < 2^{-k} \tag{A.7}$$

$$d(\sigma_k(t), \sigma_{k+1}(t)) < 2^{-k+1}. \tag{A.8}$$

Per farlo poniamo $\sigma_0(t) = x_n$ se n è il più piccolo intero per cui $\Gamma(t)$ interseca $B(x_n, 2^0)$ (per la densità della famiglia di punti è chiaro che un tale intero c'è). L'applicazione σ_0 è misurabile, infatti

$$\sigma_0^{-1}(x_n) = \Gamma^{-1}(B(x_n, 2^0)) \setminus \bigcup_{i < n} \Gamma^{-1}(B(x_i, 2^0)) \in \mathcal{T}.$$

Per costruzione $d(\sigma_0(t), \Gamma(t)) < 2^0$. Supponiamo ora di aver definito $(\sigma_i)_{i=1..k}$ rispettando le condizioni che vogliamo imporre. Sia $T_j = \sigma_k^{-1}(x_j)$: se $t \in T_j$ allora l'intersezione $\Gamma(t) \cap B(x_j, 2^{-k})$ è non vuota, ed esisterà pertanto un minimo intero n tale che la palla di centro x_n e raggio $2^{-(k+1)}$ intersechi $\Gamma(t) \cap B(x_j, 2^{-k})$ e poniamo allora per ogni $t \in T_j$ $\sigma_{k+1}(t) = x_n$. Analogamente a prima, rincollando l'applicazione definita sui vari T_j , σ_{k+1} è misurabile e $d(\sigma_{k+1}(t), \sigma_k(t)) < 2^{-(k+1)} + 2^{-k} < 2^{-k+1}$; inoltre

$d(\sigma_{k+1}(t), \Gamma(t)) < 2^{-(k+1)}$. Una volta procuratici questa successione, notiamo che per ogni t la successione $(\sigma_k(t))_k$ è di Cauchy per la (A.8) e pertanto ha un limite. L'applicazione $\sigma : t \mapsto \lim_k \sigma_k(t)$ sarà pertanto misurabile e la condizione (A.7) ci garantisce che $\lim_k d(\sigma_k(t), \Gamma(t)) = 0$ e quindi, essendo $\Gamma(t)$ chiuso, $\sigma(t) \in \Gamma(t)$, il che ci dà una selezione misurabile della multifunzione data. \square

Corollario A.33. *Una funzione $f : T \rightarrow X$ tale che per ogni $x \in X$ la funzione $t \rightarrow d(x, f(t))$ sia misurabile è essa stessa misurabile.*

Dimostrazione. f ammette una selezione misurabile. \square

A.3.3 Multifunzioni convesse

Con il termine “multifunzione convessa” intenderemo semplicemente una multifunzione a valori convessi di un certo spazio vettoriale topologico, così come abbiamo già usato il termine “chiusa” per intendere “a valori chiusi”, senza esplicitamente sottolinearlo. Vi sono varie caratterizzazioni delle proprietà di semicontinuità e misurabilità di queste multifunzioni in termini delle funzioni supporto $p \rightarrow \delta^*(\varphi|\Gamma(p))$ con $\varphi \in X^*$.

Ci occuperemo qui solo di ciò che è essenziale per i nostri scopi, pertanto ci limiteremo al caso in cui lo spazio vettoriale sia \mathbb{R}^n e dimostreremo l'unico teorema che abbiamo utilizzato.

Teorema A.34. *Sia $\Gamma : \Omega \rightarrow \mathbb{R}^n$ una multifunzione a valori convessi compatti e supponiamo che esista un intorno N_0 di un punto $\bar{p} \in T$ tale che $\bigcup_{p \in N_0} \Gamma(\bar{p})$ sia relativamente compatta. Allora sono equivalenti:*

1. Γ è s.c.i. in \bar{p} ;
2. le funzioni $p \rightarrow \delta^*(y|\Gamma(p))$ sono s.c.i. per ogni $y \in \mathbb{R}^n$.

Dimostrazione. (1 \Rightarrow 2) Sia dato $a \in \mathbb{R} : \delta^*(y|\Gamma(\bar{p})) > a$: allora $\Gamma(\bar{p})$ interseca l'insieme $\{x|x \cdot y > a\}$ e per semicontinuità lo stesso succede anche per $p \in N$, dove N è un opportuno intorno di \bar{p} . Conseguentemente per $p \in N$ $\delta^*(y|\Gamma(p)) > a$, il che mostra la semicontinuità delle funzioni in questione.

(2 \Rightarrow 1) Consideriamo un aperto U che intersechi $\Gamma(\bar{p})$. Possiamo supporre a meno di traslazioni che $0 \in \Gamma(\bar{p}) \cap U$ e a meno di restringere l'aperto che U sia convesso. Supponiamo per assurdo che esista una successione $(p_k)_k$, con $p_k \rightarrow \bar{p}$ e $\Gamma(p_k) \cap U = \emptyset$. Allora per Hahn-Banach esistono dei vettori $y_k : y_k \cdot x \leq -1$ per ogni $x \in \Gamma(p_k)$, $y_k \cdot x \geq -1$ per ogni $x \in U$. Quest'ultima condizione, se U contiene la palla di centro 0 e raggio $r > 0$ (e questo, per un certo $r > 0$, succede) dice che $|y_k| < \frac{1}{r}$. A meno

di sottosuccessioni possiamo pertanto supporre che $y_k \rightarrow y$. Se R è un raggio tale che $\Gamma(p) \subset B(0, R)$ per ogni $p \in N_0$ vale allora

$$\delta^*(y|\Gamma(p_n)) \leq \delta^*(y_n|\Gamma(p_n)) + |y - y_n|R \leq -1 + |y - y_n|R.$$

Passando al liminf si trova $\liminf_n \delta^*(y|\Gamma(p_n)) \leq -1$, il che contraddice la semicontinuità della funzione $\delta^*(y|\Gamma(\cdot))$ perché $0 \in \Gamma(\bar{p})$ e quindi $\delta^*(y|\Gamma(\bar{p})) \geq 0$. \square

Per completezza enunciamo anche la caratterizzazione della semicontinuità superiore e della misurabilità, sempre nell'ambito di \mathbb{R}^n , anche se si potrebbe sostituire con spazi vettoriali topologici più generali (non necessariamente metrizzabili), sostituendo laddove necessario la topologia forte con quella debole.

Teorema A.35. *Sia $\Gamma : \Omega \rightarrow \mathbb{R}^n$ una multifunzione a valori nei sottoinsiemi non vuoti di \mathbb{R}^n e $\bar{p} \in \Omega$ tale che $\Gamma(\bar{p})$ sia compatto e convesso. Allora sono equivalenti*

1. Γ è s.c.s. in \bar{p} ;
2. le funzioni $p \rightarrow \delta^*(y|\Gamma(p))$ sono s.c.s. per ogni $y \in \mathbb{R}^n$.

Teorema A.36. *Sia (T, \mathcal{T}) uno spazio misurabile e $\Gamma : T \rightarrow \mathbb{R}^n$ una multifunzione a valori nei sottoinsiemi convessi compatti di \mathbb{R}^n . Allora sono equivalenti*

1. Γ è \mathcal{T} -misurabile;
2. le funzioni $p \rightarrow \delta^*(y|\Gamma(p))$ sono \mathcal{T} -misurabili per ogni $y \in \mathbb{R}^n$.

In [7] si mostra come effettivamente si possa dare una variante al teorema A.34, migliorandone un'implicazione e cambiando le ipotesi dell'altra, dando origine al seguente:

Teorema A.37. *Sia $\Gamma : \Omega \rightarrow \mathbb{R}^n$ una multifunzione a valori convessi chiusi. Definita l'applicazione $\varphi : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ data da*

$$\varphi(p, y) = \delta^*(y|\Gamma(p)),$$

sono equivalenti:

1. Γ è s.c.i.;
2. φ è s.c.i. in $\Omega \times \mathbb{R}^n$.

Dimostrazione. (1 \Rightarrow 2) Fissiamo $(\bar{p}, \bar{y}) \in \Omega \times \mathbb{R}^n$ e $r \in \mathbb{R} : r < \varphi(\bar{p}, \bar{y})$. Posto $W = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n | x \cdot y > r\}$ esiste $\bar{x} \in \Gamma(\bar{p}) : (\bar{x}, \bar{y}) \in W$, e, siccome W è aperto, anche degli intorno U, V di \bar{x}, \bar{y} rispettivamente tali che $U \times V \subset W$. poiché $\Gamma(\bar{p})$ interseca

U e Γ è s.c.i., allora esiste un intorno N di \bar{p} in Ω tale che per ogni $p \in N$ $\Gamma(p)$ interseca U in un punto x_p . Allora, se $p \in N$ e $y \in V$

$$\varphi(p, y) \geq x_p \cdot y > r,$$

che mostra la semicontinuità di φ .

(2 \Rightarrow 1) Basta ripercorrere la dimostrazione corrispondente nel teorema A.34 e usare la semicontinuità nelle due variabili nel punto (\bar{p}, y) . \square

Questa nuova caratterizzazione ci rende facile la dimostrazione di questo lemma che ci serve nello studio dei funzionali locali semicontinui.

Lemma A.38. *Se $\Gamma : \Omega \rightarrow \mathbb{R}^n$ è una multifunzione convessa s.c.i. tale che $\Gamma(\bar{p})$ contenga $B(0, r)$, allora esiste un intorno N di \bar{p} tale che per ogni $p \in N$ vale $\Gamma(p) \supset B(0, \frac{r}{2})$.*

Dimostrazione. Supponiamo per assurdo che non sia vero e troviamo una successione $(p_k)_k$ con $p_k \rightarrow \bar{p}$ e $B(0, \frac{r}{2})$ non contenuta in $\Gamma(p_k)$, pertanto per ogni k esiste un punto $x_k \in B(0, \frac{r}{2}) \setminus \Gamma(p_k)$, che può essere separato da $\Gamma(p_k)$ per Hahn-Banach, trovando quindi dei vettori $y_k \in S^{n-1}$ tali che

$$\delta^*(y_k | \Gamma(p_k)) < \frac{r}{2}.$$

Siccome da $(y_k)_k$ si può estrarre una sottosuccessione convergente ad un certo $y \in S^{n-1}$ e $\delta^*(y | \Gamma(\bar{p})) \geq r$, perché $\Gamma(\bar{p}) \supset B(0, r)$, si ha una contraddizione della semicontinuità dell'applicazione φ nota dal teorema precedente. \square

Appendice B

Γ –convergenza

B.1 Premesse

Ci è capitato più volte di usare questo fatto:

Teorema B.1. *Sia X uno spazio metrico compatto, $(f_k)_k$ una successione di funzionali s.c.i. su X , e f tale che $f_k \nearrow f$. Allora $\lim_{k \rightarrow +\infty} \min f_k = \min f$.*

Prima di dare una dimostrazione elementare di questo fatto introduciamo una teoria, nota appunto come teoria della Γ –convergenza, in grado di generalizzare risultati simili a quello del teorema B.1. Si darà una nozione di convergenza per funzionali di cui si vogliono studiare le proprietà variazionali, tale da farne passare al limite il valore del minimo; inoltre si scopriranno condizioni sotto cui a passare al limite saranno anche i punti che realizzano tali minimi (il che spesso si rivela molto più utile). La convergenza monotona di funzionali semicontinui sarà solo un caso particolare di Γ –convergenza, ma è comunque utile vedere una dimostrazione del teorema che non faccia uso di questa teoria.

Dimostrazione. È chiaro che $\min f_k$ cresce verso un certo valore non più grande di $\min f$ (f risulta semicontinua dal fatto di essere espressa come sup di funzionali s.c.i., e pertanto ha minimo). Per ottenere l'altra disuguaglianza scegliamo per ogni k un punto x_k di minimo per f_k . Estrahendo una sottosuccessione $(x_{k_h})_h$ troviamo un punto limite $x \in X$. Noi vorremmo

$$\min f \leq f(x) \leq \liminf_{h \rightarrow +\infty} f_{k_h}(x_{k_h}) = \sup_k \min f_k.$$

Ci manca ovviamente solo la seconda disuguaglianza. Si scelga $q \in \mathbb{N}$ arbitrario; definitivamente $k_h > q$ e, usando la monotonia e la semicontinuità di f_q , si ha

$$\liminf_{h \rightarrow +\infty} f_{k_h}(x_{k_h}) \geq \liminf_{h \rightarrow +\infty} f_q(x_{k_h}) \geq f_q(x).$$

La tesi segue dall'arbitrarietà di q . □

Come esempio di caso in cui serve far passare al limite i punti che realizzano l'ottimo citiamo quello del teorema 4.7, che riguarda problemi di massimo invece che di minimo ma comunque ben si adatta in questo contesto.

Teorema B.2. *Supponiamo di essere in una situazione in cui le funzioni c -concave sono continue e limitate. Allora se $\mu_\varepsilon \rightharpoonup \mu$ per $\varepsilon \rightarrow 0$ e ψ_ε massimizza nella formula di dualità per il problema del trasporto fra μ_ε e una ν fissata e $\psi_\varepsilon \rightarrow \bar{\psi}$ uniformemente allora $\bar{\psi}$ è c -concava e massimizza per (μ, ν) .*

Dimostrazione. Che $\bar{\psi}$ sia c -concava viene dall'apposito teorema A.2; siccome ψ_ε realizza il massimo abbiamo per ogni ψ c -concava

$$\int \psi d\mu_\varepsilon + \int \psi^c d\nu \leq \int \psi_\varepsilon d\mu_\varepsilon + \int \psi_\varepsilon^c d\nu.$$

Questa relazione passa al limite per $\varepsilon \rightarrow 0$ grazie alla convergenza debole delle μ_ε (convergenza per cui le ψ c -concave sono funzioni test accettabili) e a quella forte delle ψ_ε . Allora per ogni ψ c -concava vale

$$\int \psi d\mu + \int \psi^c d\nu \leq \int \bar{\psi} d\mu + \int \bar{\psi}^c d\nu,$$

che dà la massimalità di $\bar{\psi}$. □

Anche qui abbiamo dato una dimostrazione artigianale del teorema, ed in questo caso avremmo avuto delle scomodità aggiuntive per utilizzare la teoria della Γ -convergenza, dovendo cambiare i segni a causa del fatto che si ha a che fare con problemi di massimo.

B.2 Definizione ed esempi

Sia dato uno spazio metrico X e una successione di funzioni $(F_h)_h$ da X a valori in $\overline{\mathbb{R}}$.

Definizione B.1. Definiamo il Γ -lim inf F' e il Γ -lim sup F'' delle F_h come

$$F'(x) = \inf_{x_h \rightarrow x} \liminf_h F_h(x_h),$$

$$F''(x) = \inf_{x_h \rightarrow x} \limsup_h F_h(x_h).$$

Se $F' = F'' = F$ diremo che la successione $(F_h)_h$ Γ -converge a F .

Si noti il carattere asimmetrico della definizione B.1: viene dal fatto che si vogliono ottenere buone proprietà di tipo variazionale in relazione ai minimi (se le si volessero ottenere in relazione ai massimi bisognerebbe cambiare la definizione). In generale, date $(F_h)_h, F$, se si vuole dimostrare che $F_h \xrightarrow{\Gamma} F$ si opera mostrando $F \leq F', F'' \leq F$.

Esempio B.1. Supponiamo che ogni F_h sia s.c.i. rispetto alla metrica di X e che $F_h \nearrow F$: sotto queste ipotesi $F_h \xrightarrow{\Gamma} F$.

Infatti la seconda delle due disuguaglianze da dimostrare è banale: scegliendo $x_h = x$ si ha

$$F''(x) \leq \liminf_h F_h(x) = F(x).$$

La dimostrazione di $F \leq F'$ si fa fissando una successione $x_h \rightarrow x$ e poi dimostrando $F \leq \liminf_h F_h(x_h)$ come nella dimostrazione del teorema B.1 (fissando un $q \in \mathbb{N}$ e confrontando con F_q).

Il successivo esempio mostra come nel precedente non si possa omettere la condizione di semicontinuit .

Esempio B.2.

Sia $D = \{q_k | k \in \mathbb{N}\} \subset X$ un sottinsieme denso e numerabile, e

$$F_h(x) = \begin{cases} 0 & \text{se } x = q_k \text{ per qualche } k \geq h, \\ 1 & \text{altrimenti.} \end{cases}$$

La successione converge crescendo puntualmente alla costante 1, tuttavia si pu  vedere che il suo Γ -limite esiste ed   la funzione nulla. Infatti dato un qualunque punto $x \in X$ e un qualunque h   sempre possibile trovare un q_k con k arbitrariamente grande (per esempio maggiore di h) in $B(x, 1/h)$, procurandosi cos  una successione di punti $x_h = q_{k(h)}$ tali che $F_h(x_h) = 0 \forall h$.

L'esempio precedente mostra anche come non vi sia in generale nessuna implicazione fra la convergenza puntuale e la Γ -convergenza, ma di esempi simili se ne possono produrre anche di molto pi  facili e generali, come questo.

Esempio B.3. Una successione costante di funzioni su X $F_h = F_0$ Γ -converge a $\overline{F_0}$, l'involuppo semicontinuo inferiore della F_0 .

In particolare se F_0 non   s.c.i. il Γ -limite   diverso dal limite puntuale.

In [10], capitolo 5, si trovano discusse tutte le relazioni fra la convergenza puntuale e la Γ -convergenza. Vediamo qui invece un collegamento con la convergenza uniforme.

Lemma B.3. *Supponiamo che $F_h : X \rightarrow \mathbb{R}$ converga uniformemente a F e che F sia continua. Allora per ogni successione $x_h \rightarrow x$ vale $F_h(x_h) \rightarrow F(x)$.*

Dimostrazione. Vale

$$|F_h(x_h) - F(x)| \leq |F_h(x_h) - F(x_h)| + |F(x_h) - F(x)|,$$

dove l'ultimo addendo tende a 0 per la continuit  di F e il primo per la convergenza uniforme. □

Si può in realtà dimostrare che su uno spazio compatto vale anche il viceversa di questo lemma.

Corollario B.4. *Se $F_h \rightarrow F$ uniformemente e F è continua allora $F_h \xrightarrow{\Gamma} F$.*

Dimostrazione. Basta notare che nel calcolo sia di $F'(x)$ che di $F''(x)$ non si fanno né liminf né limsup, ma solo limiti e che il risultato è sempre $F(x)$, per cui non si deve nemmeno passare all'inf. \square

Per comodità futura è importante notare come si comportano i concetti di Γ -liminf e Γ -limsup per passaggio a sottosuccessioni.

Lemma B.5. *Sia $(F_h)_h$ una successione di funzioni su uno spazio metrico X a valori in $\overline{\mathbb{R}}$. Detti F' e F'' rispettivamente il suo Γ -liminf e il suo Γ -limsup e considerata una sottosuccessione $(F_{h_k})_k$, per la quale si ponga*

$$G' = \Gamma - \liminf_k F_{h_k}; \quad G'' = \Gamma - \limsup_k F_{h_k},$$

vale allora

$$F' \leq G' \leq G'' \leq F''.$$

Dimostrazione. Dimostriamo ora la prima disuguaglianza: basta considerare un arbitrario punto $x \in X$ e un'arbitraria successione $(x_k)_k$ con $x_k \rightarrow x$ e mostrare $F'(x) \leq \liminf_k F_{h_k}(x_k)$. Si consideri la successione $(y_h)_h$ data da:

$$y_h = x_k \text{ se } h_{k-1} < h \leq h_k.$$

Vale allora $F_{h_k}(y_{h_k}) = F_{h_k}(x_k)$ e $\lim_h y_h = x$. Pertanto

$$F'(x) \leq \liminf_h F_h(y_h) \leq \liminf_k F_{h_k}(y_{h_k}) = \liminf_k F_{h_k}(x_k).$$

La disuguaglianza centrale è banale, l'ultima invece viene se, considerando un qualunque punto $x \in X$ e una successione $y_h \rightarrow x$, si mostra $G''(x) \leq \limsup_h F_h(y_h)$. Si ha tuttavia che, posto $x_k = y_{h_k}$, anche $x_k \rightarrow x$, quindi

$$G''(x) \leq \limsup_k F_{h_k}(x_k) = \limsup_k F_{h_k}(y_{h_k}) \leq \limsup_h F_h(y_h).$$

\square

Corollario B.6. *Se $F_h \xrightarrow{\Gamma} F$ allora anche $F_{h_k} \xrightarrow{\Gamma} F$.*

B.3 Convergenza dei minimi

Mostriamo innanzitutto che le funzioni F', F'' che abbiamo definito sono sempre s.c.i., il che ci permetterà, quando disporremo delle giuste ipotesi di compattezza in relazione alle F_h , di concludere che possiamo parlare di minimi per F', F'' invece che solo di inf.

Teorema B.7. *Date $F_h : X \rightarrow \overline{\mathbb{R}}$, il Γ -liminf F' e il Γ -limsup F'' della successione $(F_h)_h$ sono funzioni semicontinue inferiormente sullo spazio metrico X .*

Dimostrazione. Si consideri una successione $x_n \rightarrow x \in X$. Per ogni n , per definizione di Γ -liminf, esiste una successione $(y_{n,h})_h : \lim_h y_{n,h} = x_n$ con $\liminf_h F_h(y_{n,h}) < F'(x_n) + 1/n$. Pertanto per ogni n esiste anche un numero $h(n)$ abbastanza grande (e possiamo supporre $h(n+1) > h(n)$) tale che

$$\begin{aligned} F_{h(n)}(y_{n,h(n)}) &< F'(x_n) + \frac{1}{n} \\ d(y_{n,h(n)}, x_n) &< \frac{1}{n}. \end{aligned}$$

Si noti che $d(y_{n,h(n)}, x) < 1/n + d(x_n, x) \rightarrow 0$. Considerando la sottosuccessione associata agli indici dati da $(h(n))_n$ e chiamando G' il Γ -liminf corrispondente a tale sottosuccessione, abbiamo, per il lemma B.5,

$$F'(x) \leq G'(x) \leq \liminf_n F_{h(n)}(y_{n,h(n)}) \leq \liminf_n F'(x_n) + \frac{1}{n} = \liminf_n F'(x_n),$$

che dà la semicontinuità di F' . Per la semicontinuità di F'' si considerino $x_n \rightarrow x, y_{n,h} \rightarrow x_n$ tali che

$$\limsup_h F_h(y_{n,h}) < F''(x_n) + \frac{1}{n}.$$

Esiste allora per ogni n un $h(n)$ abbastanza grande (e ancora possiamo supporre che valga la disuguaglianza $h(n+1) > h(n)$) tale che, per ogni $h \geq h(n)$,

$$\begin{aligned} F_h(y_{n,h}) &< F''(x_n) + \frac{1}{n} \\ d(y_{n,h}, x_n) &< \frac{1}{n}. \end{aligned}$$

Si definisca allora una successione $(z_h)_h$ tramite

$$z_h = y_{n(h),h} \text{ dove } n(h) \text{ è l'unico } n \text{ tale che } h(n) \leq h < h(n+1).$$

Vale allora $d(z_h, x) < d(x_n, x) + \frac{1}{n(h)} \rightarrow 0$, perché $n(h) \rightarrow +\infty$ per $h \rightarrow +\infty$. Quindi $z_h \rightarrow x$ e allora vale

$$F''(x) \leq \liminf_h F_h(z_h) = \liminf_h F_h(y_{n(h),h}) \tag{B.1}$$

$$\leq \liminf_h F''(x_{n(h)}) + \frac{1}{n(h)} = \liminf_h F''(x_{n(h)}). \tag{B.2}$$

Siccome possiamo ragionare per assurdo e supporre di avere a che fare con una successione $x_n \rightarrow x$ tale che $F''(x_n)$ abbia limite e che questo limite sia minore di $F''(x)$, all'ultimo membro della B.1 possiamo sostituire $\lim_n F''(x_n)$ e trovare una contraddizione. \square

Per proseguire avremo bisogno, come già annunciato, di richiedere alcune proprietà di compattezza alla successione $(F_h)_h$, che saranno banalmente verificate quando l'intero spazio X sarà compatto, e pertanto diamo queste definizioni:

Definizione B.2. Una funzione $F : X \rightarrow \overline{\mathbb{R}}$ si dice coerciva se i suoi sottolivelli $\{F \leq t\}$ sono relativamente compatti per ogni $t \in \mathbb{R}$.

Osservazione B.1. Se F è anche semicontinua inferiormente si può sottindere la parola 'relativamente'.

Definizione B.3. Una successione $(F_h)_h$ di funzioni su X a valori in $\overline{\mathbb{R}}$ si dice equicoerciva se per ogni $t \in \mathbb{R}$ esiste un compatto K_t tale che $\{F_h \leq t\} \subset K_t$ per ogni indice h .

Lemma B.8. Una successione $(F_h)_h$ è equicoerciva se e solo se esiste una funzione $F : X \rightarrow \overline{\mathbb{R}}$ s.c.i. e coerciva tale che $F_h \geq F$ per ogni h .

Dimostrazione. Se F esiste si ha $\{F_h \leq t\} \subset \{F \leq t\}$ e, ponendo $K_t = \{F \leq t\}$, che è un compatto per le ipotesi su F , si ottiene l'equicoercività della successione $(F_h)_h$. Per mostrare l'implicazione opposta si consideri la famiglia di compatti $(K_t)_t \in \mathbb{R}$ data dall'equicoercività e si definisca

$$F(x) = \inf \{s \in \mathbb{R} : x \in K_t \text{ per ogni } t > s\}.$$

poiché $F_h(x) < t$ per ogni $t > F_h(x)$ allora $x \in K_t$ per ogni $t > F_h(x)$, quindi $F(x) \leq F_h(x)$, per ogni x e per ogni h . Bisogna solo verificare che F sia coerciva e semicontinua, cioè che i suoi sottolivelli siano compatti, ma

$$\{F \leq s\} = \bigcap_{t>s} K_t,$$

per cui lo sono, in quanto intersezione di compatti. \square

Corollario B.9. Se la successione $(F_h)_h$ è equicoerciva, allora F', F'' sono coercive.

Dimostrazione. Basta considerare la funzione F costruita prima e dimostrare che necessariamente $F' \geq F$. Basta cioè mostrare che per ogni successione $x_h \rightarrow x$ vale $\liminf_h F_h(x_h) \geq F(x)$. Questo è vero e si vede minorando le F_h con F e usandone la semicontinuità inferiore. \square

Per arrivare ad un teorema di convergenza dei valori minimi delle (F_h) ci servono i seguenti lemmi:

Lemma B.10. *Per un'arbitraria successione $(F_h)_h$ vale*

$$\inf F' \geq \liminf_h \inf F_h$$

$$\inf F'' \geq \limsup_h \inf F_h$$

Dimostrazione. Considerato un qualunque $x \in X$ esiste una successione $x_h \rightarrow x$ tale che $F'(x) = \liminf_h F_h(x_h)$: ne si può costruire una sottosuccessione operando come nella prima parte del teorema B.7 (usando $x_n = x$) ed estenderla a tutti gli indici come nel lemma B.5. Allora

$$F'(x) = \liminf_h F_h(x_h) \geq \liminf_h \inf F_h.$$

La dimostrazione dell'altra disuguaglianza è analoga una volta che si sappia trovare una successione $x_h \rightarrow x$ tale che $F''(x) = \limsup_h F_h(x_h)$ e questo si può fare sempre con le stesse tecniche del teorema B.7, usando $x_n = x$ e scegliendo la successione z_h che lì veniva prodotta. \square

Lemma B.11. *Se esiste un compatto $K \subset X$ tale che $\inf F_h = \inf_{x \in K} F_h(x)$ per ogni h allora*

$$\min F' = \liminf_h \inf F_h.$$

Dimostrazione. Dal lemma B.10 sappiamo che $\inf F' \geq \liminf_h \inf F_h$. Inoltre vale

$$\inf F' \leq \inf_{x \in K} F'(x) \leq \liminf_h \inf_{x \in K} F_h(x) = \liminf_h \inf F_h(x).$$

Da queste disuguaglianze, di cui dobbiamo dimostrare soltanto la seconda, segue la tesi con $\inf F'$ al posto di $\min F'$, ma segue anche che $\inf F' = \inf_{x \in K} F'(x)$, il che implica che F' assume valore minimo in quanto è s.c.i. e ci si può restringere a K . Per dimostrare la disuguaglianza mancante consideriamo una successione $(x_h)_h$ tale che

$$x_h \in K; \quad F_h(x_h) < \inf_{x \in K} F_h(x) + 1/h.$$

Cominciamo con l'estrarre una sottosuccessione x_{h_k} per cui esista $\lim_k F_{h_k}(x_{h_k}) = \liminf_h F_h(x_h)$ e da questa estraiamo una sottosuccessione convergente che chiameremo ancora x_{h_k} con $x_{h_k} \rightarrow \bar{x} \in K$. In corrispondenza della stessa successione di indici consideriamo $G' = \Gamma - \liminf_k F_{h_k}$. Vale allora

$$\begin{aligned} \inf_{x \in K} F'(x) &\leq F'(\bar{x}) \leq G'(\bar{x}) \leq \liminf_k F_{h_k}(x_{h_k}) \\ &= \lim_k F_{h_k}(x_{h_k}) = \lim_h \inf F_h(x_h) \\ &\leq \liminf_h \inf_{x \in K} F_h(x) + 1/h = \liminf_h \inf_{x \in K} F_h(x). \end{aligned}$$

\square

A questo punto possiamo enunciare il teorema principale concernente la convergenza dei minimi delle F_h .

Teorema B.12. *Supponiamo che la successione $(F_h)_h$ sia equicoerciva. Allora*

$$\min F' = \liminf_h \inf F_h.$$

Se la successione $(F_h)_h$ Γ -converge ad una funzione F allora

$$\min F' = \liminf_h \inf F_h.$$

Dimostrazione. Che F' assuma minimo viene dal fatto che è coerciva grazie al corollario B.9. La disuguaglianza $\min F' \geq \liminf_h \inf F_h$ viene dal lemma B.10. Per ottenere la disuguaglianza opposta ci si può limitare al caso in cui $\liminf_h \inf F_h < +\infty$. In questo caso, scelto $t \in \mathbb{R}$ con $t > \liminf_h \inf F_h$, esiste una sottosuccessione $(F_{h_k})_k$ tale che $\inf F_{h_k} < t$ per ogni k . Per tale sottosuccessione e per il compatto K_t dato dalla definizione di equicoercività valgono allora le ipotesi del lemma B.11. Pertanto, detto G' il $\Gamma - \liminf_k F_{h_k}$, vale $\min F' \leq \min G' = \liminf_k \inf F_{h_k} \leq t$. L'arbitrarietà di $t > \liminf_h \inf F_h$ permette di concludere la prima parte dell'enunciato.

Se in più $(F_h)_h$ Γ -converge ad F allora $F = F' = F''$ e il lemma B.10 ci dice che

$$\inf F \geq \limsup_h \inf F_h,$$

che porta alla tesi. □

B.4 Convergenza dei punti di minimo

Dimostriamo in questa sezione un facile teorema relativo ai limiti di punti di minimo per le funzioni F_h , qualora la successione $(F_h)_h$ Γ -converga ad una certa F .

Su [10] viene sviluppata una teoria molto più dettagliata: viene studiato il comportamento rispetto alla convergenza nel senso di Kuratowski (qui definita nella sezione A.3.1) dell'insieme dei punti di minimo di F_h o dei punti ε -minimizzanti (minimizzanti a meno di ε), definendo opportunamente un concetto di limsup che sostituisca il limite nella convergenza di Kuratowski. Anche altre correlazioni fra la Γ -convergenza e la convergenza secondo Kuratowski sono descritte.

Qui invece ci limiteremo al seguente.

Teorema B.13. *Supponiamo che la successione $(F_h)_h$ Γ -converga ad F e sia $(x_h)_h$ una successione di punti di X tale che $F_h(x_h) = \min F_h$ per ogni h . Allora, se $x \in X$ è un punto limite di una sottosuccessione degli x_h , F assume minimo in x .*

Dimostrazione. Consideriamo la sottosuccessione $(x_{h_k})_k$ per cui $x_{h_k} \rightarrow x$ e, per gli stessi indici, il Γ -limite G della corrispondente sottosuccessione $(F_{h_k})_k$. Per il corollario B.6 si ha $F = G$. Allora vale:

$$\min F \leq F(x) \leq \liminf_k F_{h_k}(x_{h_k}) = \liminf_k \min F_{h_k} = \lim_h \min F_h = \min F,$$

dove gli ultimi passaggi sono giustificati dal teorema B.12. Da questo segue che tutte le disuguaglianze precedenti sono uguaglianze e quindi che F assume valore minimo in x . \square

Osservazione B.2. Si poteva sostituire la condizione $F_h(x_h) = \min F_h$ con la condizione più debole $\lim_h (F_h(x_h) - \min F_h) = 0$, cosa utile qualora le F_h non abbiano minimo.

Corollario B.14. *Supponiamo che la successione $(F_h)_h$ Γ -converga ad F e che F abbia un unico punto di minimo x , con $F(x) < +\infty$; supponiamo inoltre che la successione $(F_h)_h$ sia equicoerciva. Sia $(x_h)_h$ una successione di punti di X tale che $F_h(x_h) = \min F_h$ per ogni h . Allora $x_h \rightarrow x$.*

Dimostrazione. È sufficiente mostrare che la successione $(x_h)_h$ è relativamente compatta, perché il teorema precedente identifica univocamente in x i suoi eventuali punti limite. Dal teorema B.12 sappiamo $\min F_h \rightarrow \min F$ e quindi, preso $t \in \mathbb{R} : t > \min F$, definitivamente $\min F_h < t$. Allora definitivamente $F_h(x_h) < t$, da cui segue che $x_h \in K_t$, dove K_t viene dalla definizione di equicoercività. Questo dimostra la relativa compattezza. \square

Bibliografia

- [1] P. Acquistapace. *Appunti di Analisi Convessa*. SEU, Pisa, 2000.
- [2] G. Alberti and L. Ambrosio. A geometrical approach to monotone functions in \mathbb{R}^n . *Mathematische Zeitschrift*, 230:259–316, 1999.
- [3] L. Ambrosio and A. Pratelli. Existence and stability results in the l^1 theory of optimal transportation. Di prossima pubblicazione in *Lecture Notes in Mathematics*, CIME series, school in Martina Franca, 2002.
- [4] L. Ambrosio and P. Tilli. *Selected Topics on Analysis in Metric Spaces*, volume 18 of *Appunti*. Scuola Normale Superiore, Pisa, 2000.
- [5] G. Bouchitté and G. Buttazzo. New lower semicontinuity results for nonconvex functionals defined on measures. *Non linear Analysis*, 15(7):679–692, 1990.
- [6] G. Bouchitté and G. Buttazzo. Integral representation of nonconvex functionals defined on measures. *Annales de l'Institut Henry Poincaré-Analyse non Linéaire*, 9(1):101–117, 1992.
- [7] G. Bouchitté and M. Valadier. Integral representation of convex functionals on a space of measures. *Journal of Functional Analysis*, 80:398–420, 1988.
- [8] H. Brezis. *Analisi Funzionale: Teoria e Applicazioni*. Liguori Editore, Napoli, 1986.
- [9] C. Castaing and M. Valadier. *Convex Analysis and Measurable Multifunction*. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 1977.
- [10] G. Dal Maso. *An introduction to Γ -convergence*. Birkhauser, Basel, 1992.
- [11] C. Dellacherie and P.-A. Meyer. *Probabilities and Potential*. North-Holland, Amsterdam, 1978.
- [12] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.

- [13] L. Ruschendorf. On c -optimal random variables. *Statistics and Probability Letters*, 27:267–270, 1996.
- [14] C. Villani. Topics in optimal transportation. Di prossima pubblicazione da parte dell'AMS, 2002.