Βέλτιστη μεταφορά του μέτρου και γεωμετρικές ανισότητες

Διπλωματική Εργασία
Σουλτάνα Παπανικολάου

Επιβλέπων: Απόστολος Γιαννόπουλος

Τμήμα Μαθηματικών
Πανεπιστήμιο Αθηνών
Αθήνα – 2014

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Το πρόβλημα μεταφοράς μάζας

Η θεωρία της βέλτιστης μεταφοράς ανακαλύφθηκε για πρώτη φορά από τον Γάλλο γεωμέτρη Gaspard Monge. Ο Monge εφηύρε μόνος του την παραστατική γεωμετρία, η σημασία της οποίας ήταν τόσο εμφανής που διορίστηκε καθηγητής σε ηλικία 22 ετών. Το 1781 δημοσίευσε μία από τις πιο διάσημες δουλειές του: "Mémoire sur la théorie des déblais et des remblais". Το πρόβλημα που πραγματευόταν ο Monge ήταν το εξής: υποθέτουμε ότι έχουμε μία συγκεκριμένη ποσότητα ενός παραγόμενου προϊόντος και μία συγκεκριμένη ζήτηση αυτού: οι τοποθεσίες όπου παράγεται το προϊόν και εκείνες στις οποίες πρέπει να μεταφερθεί θεωρούνται γνωστές. Το ερώτημα είναι σε ποιόν προορισμό πρέπει να σταλεί το προϊόν που έχει παραχθεί σε ένα συγκεκριμένο μέρος ώστε το συνολικό κόστος μεταφοράς να είναι το ελάχιστο δυνατό.

Η ποσότητα προϊόντος που παράγεται και εκείνη που ζητείται μπορούν να μοντελοποιηθούν από μέτρα πιθανότητας, ενώ ο χώρος στον οποίο εργαζόμαστε θεωρείται εφοδιασμένος με μία φυσιολογική μετρική, όπου στην συγκεκριμένη περίπτωση θεωρούμε την απόσταση των σημείων παραγωγής και των σημείων ζήτησης.

Ο Monge μελέτησε το πρόβλημα στις τρεις διαστάσεις και στην περίπτωση που η κατανομή της μάζας αντιπροσωπεύεται από μία συνεχή συνάρτηση. Η γεωμετρική του διαίσθηση τον οδήγησε στην σημαντική παρατήρηση πως η μεταφορά θα πρέπει να γίνει κατά μήκος ευθειών γραμμών οι οποίες πρέπει να είναι ορθογώνιες σε μία οικογένεια επιφανειών. Αυτή η μελέτη τον οδήγησε στην ανακάλυψη των *γραμμών καμπυλότητας*, μία έννοια που από μόνη της αποτέλεσε τεράστια συνεισφορά στην γεωμετρία των επιφανειών. Οι ιδέες του αναπτύχθηκαν από τον Charles Dupin και αργότερα από τον Paul Appell.

Ας περιγράψουμε όμως λίγο πιο εκτενώς το πρόβλημα της βέλτιστης μεταφοράς. Όπως αναφέραμε και προηγουμένως, μπορούμε να μοντελοποιήσουμε την ποσότητα παραγωγής και την ζήτηση του προϊόντος με μέτρα πιθανότητας, έστω μ και ν, που ορίζονται αντίστοιχα σε κάποιους χώρους πιθανότητας X και Y. Συνεπώς, αν A και B είναι μετρήσιμα υποσύνολα των χώρων X και Y αντίστοιχα, τότε το $\mu(A)$ δηλώνει την ποσότητα προϊόντος που παράγεται στην περιοχή A και το $\nu(B)$ δηλώνει την ποσότητα προϊόντος που ζητείται στην περιοχή B.

Η μεταφορά όμως του προϊόντος χρειάζεται κάποια προσπάθεια, την οποία μοντελοποιούμε μέσω μιας μετρήσιμης **συνάρτησης κόστους** c ορισμένης στον $X \times Y$. Κατά κάποιον τρόπο, η $c(x, y)$ δηλώνει πόσο θα κοστίσει η μεταφορά μιας μονάδας μάζας από την τοποθεσία x στην τοποθεσία y. Μπορούμε λοιπόν να υποθέσουμε ότι η c είναι μετρήσιμη και μη αρνητική, χωρίς να απορρίπτουμε το ενδεχόμενο να λαμβάνει και την τιμή του απείρου. Συνεπώς, η c πρέπει να είναι μια μετρήσιμη απεικόνιση από τον $X \times Y$ στον $\mathbb{R}_+ \cup \{+\infty\}$

Το βασικό ερώτημα είναι πώς θα καταφέρουμε να πραγματοποιήσουμε την μεταφορά με το ελάχιστο δυνατό κόστος. Μοντελοποιώντας το πρόβλημα θεωρούμε ότι κάθε πιθανό **σχέδιο μεταφοράς** είναι ένα μέτρο πιθανότητας π στον χώρο $X \times Y$ (γράφουμε $\pi \in \mathcal{P}(X \times Y)$). Άτυπα λοιπόν, το $d\pi(x, y)$ μετρά την ποσότητα μάζας που μεταφέρεται από την τοποθεσία x στην τοποθεσία y, χωρίς να αποκλείουμε την πιθανότητα μέρος της μάζας που είναι τοποθετημένη στο x να μοιραστεί σε διάφορα μέρη. Δηλαδή, δεν απαιτούμε απαραίτητα όλη η ποσότητα μάζας που βρίσκεται στο x να μεταφερθεί σε ένα συγκεκριμένο y. Συνεπώς, ένα σχέδιο μεταφοράς $\pi \in \mathcal{P}(X \times Y)$ είναι αποδεκτό αν όλη η μάζα που παίρνουμε από το σημείο x συμπίπτει με το $d\mu(x)$ και όλη η μάζα που μεταφέρεται στο y συμπίπτει με το $d\nu(y)$. Συγκεκριμένα, έχουμε

$$\int_Y d\pi(x, y) = d\mu(x), \ \int_X d\pi(x, y) = d\nu(y)$$

Πιο αυστηρά λοιπόν, απαιτούμε

$$\pi(A \times Y) = \mu(A), \ \pi(X \times B) = \nu(B) \tag{1.1.1}$$

για όλα τα μετρήσιμα υποσύνολα A, B των X, Y αντίστοιχα. Ισοδύναμα μπορούμε να πούμε ότι για κάθε ζεύγος συναρτήσεων φ, ψ σε μία κατάλληλη κλάση συναρτήσεων δοκιμής ικανοποιείται το εξής:

$$\int_{X \times Y} [\varphi(x) + \psi(y)] d\pi(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \tag{1.1.2}$$

Το φυσικό σύνολο των αποδεκτών συναρτήσεων δοκιμής είναι οι $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$ ή ισοδύναμα οι $(\varphi, \psi) \in L^\infty(d\mu) \times L^\infty(d\nu)$. Στις πιο ενδιαφέρουσες περιπτώσεις μπορούμε να περιοριστούμε στις κλάσεις $C_b(X) \times C_b(Y)$ ή $C_0(X) \times C_0(Y)$.

Τα μέτρα πιθανότητας που ικανοποιούν την (1.1.1) λέμε ότι έχουν **περιθώρια μέτρα** μ, ν και είναι αποδεκτά σχέδια μεταφοράς. Το σύνολο αυτών των μέτρων πιθανότητας το συμβολίζουμε με

$$\Pi(\mu,\nu) = \{\pi \in P(x,y) : \text{η (1.1) ισχύει για κάθε μετρήσιμα σύνολα A,B}\} \tag{1.1.3}$$

Το σύνολο αυτό είναι μη κενό αφού $\mu \otimes \nu \in \Pi(\mu,\nu)$.

Το πρόβλημα του Monge ήρθε ξανά στο προσκήνιο πολύ αργότερα, απο τον Ρώσσο μαθηματικό Leonid Vitaliyevich Kantorovich. Ο Kantorovich ασχολήθηκε με πολλούς τομείς των μαθηματικών οι οποίοι σχετίζονταν με εφαρμογές στα οικονομικά, και αργότερα ασχολήθηκε με την θεωρητική επιστήμη των υπολογιστών. Το 1938 ένα εργαστήριο ζήτησε την βοήθειά του για την επίλυση ενός συγκεκριμένου προβλήματος βελτιστοποίησης, το οποίο ανακάλυψε πως ηταν αντιπροσωπευτικό για μία ολόκληρη κλάση γραμμικών προβλημάτων που εμφανίζονται σε διάφορες περιοχές των οικονομικών. Παρακινούμενος από αυτήν την ανακάλυψη, ανέπτυξε τα εργαλεία του γραμμικού προγραμματισμού (βλέπε [35]), τα οποία αργότερα έγιναν κυρίαρχα στα οικονομικά. Το 1975 κέρδισε το Βραβείο Νόμπελ στα οικονομικά μαζί με τον Tjalling Koopmans για την συνεισφορά τους στην θεωρία της βέλτιστης κατανομής των πηγών. Η σημαντικότερη δουλειά του στα οικονομικά παρουσιάζεται στο βιβλίο "The best use of economic resources" (βλέπε [39]) και στην επανέκδοσή του (βλέπε [35]).

Το **πρόβλημα βέλτιστης μεταφοράς του Kantorovich** είχε ως εξής: θέλουμε να ελαχιστοποιήσουμε την ποσότητα

$$I[\pi] = \int_{X\times Y} c(x,y)d\pi(x,y), \ \forall \pi \in \Pi(\mu,\nu) \tag{1.1.4}$$

το οποίο και μελέτησε ο Kantorovich την δεκαετία του σαράντα (βλέπε [40, 41]). Το πώς το πρόβλημα βέλτιστης μεταφοράς σχετίζεται με βασικές ερωτήσεις στα οικονομικά γίνεται σαφές αν κανείς σκεφτεί ότι το μ είναι η πυκνότητα μιας μονάδας παραγωγής και το ν η πυκνότητα των καταναλωτών. Συνεπώς, για ένα δοθέν σχέδιο μεταφοράς π, η μη αρνητική ποσότητα $I[\pi]$ θα καλείται το **ολικό κόστος μεταφοράς** σε σχέση με το π, ενώ το **βέλτιστο κόστος μεταφοράς** μεταξύ των μ και ν είναι η τιμή

$$T_c(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} I(\pi) \tag{1.1.5}$$

Άρα, τα π που ικανοποιούν την $I[\pi] = T_c(\mu,\nu)$, αν υπάρχουν, θα καλούνται βέλτιστα σχέδια μεταφοράς.

Το πρόβλημα του Kantorovich είναι στην ουσία μία πιο χαλαρή εκδοχή του αρχικού προβλήματος μεταφοράς μάζας που διατύπωσε ο Monge [50]. Συγκεκριμένα, το πρόβλημα του Monge είναι το ίδιο με αυτό του Kantorovich, αλλά έχει μία επιπλέον απαίτηση: *να μην διαχωρίζεται η μάζα*. Με άλλα λόγια σε κάθε τοποθεσία x αντιστοιχεί ένας μοναδικός προορισμός y. Συνεπώς, απαιτούμε από το σχέδιο μεταφοράς π στην (1.1.3) να έχει την μορφή

$$d\pi(x,y) = d\pi_T(x,y) = d\mu(x)\delta(y = T(x)), \tag{1.1.6}$$

όπου T είναι μια μετρήσιμη **απεικόνιση** από τον X στον Y. Το μέτρο πιθανότητας που εμφανίζεται στο δεξιό μέλος της σχέσης (1.1.6) μπορεί να γραφεί ως $(Id \times T)\#\mu$, δηλαδή είναι το μέτρο πιθανότητας στον $X \times Y$ που ικανοποιεί την εξής ιδιότητα:

Αν ζ είναι μία μη αρνητική μετρήσιμη συνάρτηση στον $X \times Y$ τότε,

$$\int_{X\times Y} \zeta(x,y)d\pi_T(x,y) = \int_X \zeta(x,T(x))d\mu(x). \tag{1.1.7}$$

Συγκεκριμένα, το ολικό κόστος μεταφοράς είναι

$$I[\pi_T] = \int_X c(x,T(x))d\mu(x).$$

Δεδομένης λοιπόν της σχέσης (1.1.7), η συνθήκη (1.1.2) μεταφράζεται ως εξής:

$$\int_X [\varphi(x) + \psi \circ T(x)]d\mu(x) = \int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y),$$

από όπου έπεται η συνθήκη

$$\int_X (\psi \circ T)d\mu = \int_Y \psi d\nu \tag{1.1.8}$$

Αυτή η ταυτότητα πρέπει να ισχύει για κάθε μη αρνητική συνάρτηση $\psi \in L^1(d\nu)$, η μετρήσιμη συνάρτηση $\psi \circ T$ πρέπει να ανήκει στον χώρο $L^1(d\mu)$ και οι τιμές και των δύο ολοκληρωμάτων της (1.1.8) πρέπει να συμπίπτουν. Ισοδύναμα, με όρους μετρήσιμων υποσυνόλων, για να ανήκει το π_T στο $\Pi(\mu,\nu)$ αρκεί

$$\text{για κάθε μετρήσιμο σύνολο } B \subset Y\,, \quad \nu[B] = \mu[T^{-1}(B)]. \tag{1.1.9}$$

Όποτε, λοιπόν, οι ισοδύναμες συνθήκες (1.1.8) ή (1.1.9) ικανοποιούνται, γράφουμε

$$\nu = T\#\mu$$

και λέμε ότι το ν είναι το **push forward** ή η **εικόνα μέτρου** του μ μέσω της T ή ότι η T **μεταφέρει** το μ στο ν.

Ο Monge αρχικά είχε διατυπώσει το πρόβλημα μεταφοράς στον Ευκλείδειο χώρο με συνάρτηση κόστους την $c(x, y) = |x - y|$. Πιθανώς δεν είχε συνειδητοποιήσει την τεράστια δυσκολία της αυστηρής αντιμετώπισής του. Μόλις το 1979 ο Sudakov (βλέπε [65]) ισχυρίστηκε ότι απέδειξε την ύπαρξη μεταφοράς Monge για γενικές πυκνότητες πιθανότητας με την συγκεκριμένη συνάρτηση κόστους. Όμως η απόδειξή του δεν ήταν απολύτως σωστή και τροποποιήθηκε αργότερα από τον Ambrosio (βλέπε [5]). Εν τω μεταξύ, εναλλακτικές αυστηρές αποδείξεις δόθηκαν αρχικά από τους Evans και Gangbo (βλέπε [29]), και αργότερα από τους Trudinger και Wang (βλέπε [64]) και Caffarelli, Feldman και McCann (βλέπε [17]).

Το **πρόβλημα βέλτιστης μεταφοράς του Monge** είχε ως εξής: θέλουμε να ελαχιστοποιήσουμε την ποσότητα,

$$I[\pi] = \int_X c(x, T(x))d\mu(x)$$

πάνω από το σύνολο όλων των μετρήσιμων απεικονίσεων T ώστε $T\#\mu = \nu$. Το βασικό μειονέκτημα του προβλήματος βέλτιστης μεταφοράς του Monge είναι ότι δεν υπάρχει πάντοτε αποδεκτή απεικόνιση T. Για παράδειγμα αν το μ είναι ένα μέτρο Dirac, έστω $\mu = \delta_a$ για κάποιο $a \in X$ και ν είναι ένα μέτρο που δεν είναι Dirac, τότε για κάθε $B \subset Y$ μετρήσιμο θα είχαμε

$$\nu(B) = \delta_a(T^{-1}(B)) = \begin{cases} 0, & \text{αν } T(a) \in B \\ 1, & \text{αν } T(a) \notin B \end{cases}$$

δηλαδή $\nu(B) = \delta_{T(a)}$, το οποίο αντιφάσκει με την υπόθεσή μας πως το ν δεν είναι μέτρο Dirac. Ένα ακόμη πιο προφανές παράδειγμα είναι αν επιλέξουμε $\mu = \delta_x$ για κάποιο $x \in X$ και $\nu = \frac{1}{2}(\delta_{y_1} + \delta_{y_2})$. Τότε, δεν υπάρχει απεικόνιση μεταφοράς από το μ στο ν αν δεν μοιράσουμε την μάζα που βρίσκεται στο x στις τοποθεσίες y_1 και y_2.

Το πρόβλημα του Monge ήταν διάσημο για αρκετό καιρό, μάλιστα για την λύση του η Ακαδημία του Παρισιού πρόσφερε βραβείο [23]. Λύθηκε το 1887 από τον Appell ο οποίος έχει γράψει μία διατριβή πάνω σε αυτό το θέμα [7]. Εν αντιθέσει προς το πρόβλημα του Monge, το πρόβλημα του Kantorovich έχει αρκετά πλεονεκτήματα. Κατ' αρχήν, το σύνολο $\Pi(\mu, \nu)$ είναι μη κενό ($\mu \otimes \nu \in \Pi(\mu, \nu)$), κυρτό και συμπαγές με την w^*-τοπολογία στον $\mathcal{P}(X \times Y)$. Επιπλέον, η απεικόνιση $\pi \mapsto \int c d\pi$ είναι γραμμική, το ελάχιστο κάτω από χαλαρές υποθέσεις για την c υπάρχει πάντα, και τέλος τα σχέδια μεταφοράς «περιλαμβάνουν» τις απεικονίσεις μεταφοράς, αφού από την $T\#\mu = \nu$ έπεται ότι $\pi := (Id \times T)\#\mu \in \Pi(\mu, \nu)$.

Όσον αφορά το πρόβλημα βέλτιστης μεταφοράς, ο Kantorovich διατύπωσε και απέδειξε, χρησιμοποιώντας αναλυτικά, συναρτησιακά εργαλεία, ένα θεώρημα δυϊσμού το οποίο θα έπαιζε πολύ σημαντικό ρόλο αργότερα (βλέπε [36]) και επινόησε μιά βολική αντίληψη της απόστασης μεταξύ μέτρων πιθανότητας που ήταν η εξής: η απόσταση μεταξύ δύο μέτρων πιθανότητας πρέπει να είναι το βέλτιστο κόστος μεταφοράς από το ένα στο άλλο, αν θεωρήσουμε το κόστος ως την συνάρτηση απόστασης αυτών. Αυτή η απόσταση των μέτρων πιθανότητας είναι γνωστή σήμερα ως η απόσταση Kantorovich-Rubinstein. Ήταν μετά από λίγα χρόνια από την δημοσίευση των κύριων αποτελεσμάτων της δουλειάς του, που ο Kantorovich έκανε την σύνδεση με την δουλειά του Monge (βλέπε [38]).

Έκτοτε, το πρόβλημα βέλτιστης μεταφοράς λέγεται **πρόβλημα Monge-Kantorovich**. Στην διάρκεια του δεύτερου μισού του εικοστού αιώνα, οι τεχνικές της βέλτιστης μεταφοράς και οι παραλλαγές της απόστασης Kantorovich-Rubinstein (σήμερα καλούνται αποστάσεις Wasserstein), χρησιμοποιήθηκαν από στατιστικούς και πιθανοθεωρητικούς. Αξιοσημείωτες συνεισφορές, στην δεκαετία του 1970, έγιναν από τον Roland Dobrushin, ο οποίος χρησιμοποίησε τέτοιες αποστάσεις στη μελέτη των μερικών συστημάτων και από τον Hiroshi Tanaka, ο οποίος εφήρμοσε αυτές τις αποστάσεις στην μελέτη της συμπεριφοράς σε σχέση με τον χρόνο μίας παραλλαγής της εξίσωσης Boltzman. Στα μέσα της δεκαετίας του 1980, ειδικοί στον τομέα αυτό όπως οι Svetlozar Rachev και Ludger Rüschendorf διέθεταν πολλές ιδέες, εργαλεία, τεχνικές και εφαρμογές σχετικές με την βέλτιστη μεταφορά.

Παράλληλα, πολλοί ερευνητές που ασχολούνταν με ανισότητες που είχαν να κάνουν με όγκους και ολοκληρώματα, χρησιμοποιούσαν τεχνικές αναπαραμέτρησης (αλλαγή μεταβλητών). Πολύ αργότερα, συνειδητοποίησαν και κατανόησαν ότι η βέλτιστη μεταφορά παρέχει χρήσιμες αναπαραμετρήσεις.

Στα τέλη του 1980 τρεις κατευθύνσεις έρευνας εμφανίστηκαν ανεξάρτητα και σχεδόν ταυτόχρονα, οι οποίες αναδιαμόρφωσαν πλήρως την εικόνα της βέλτιστης μεταφοράς.

Η πρώτη εμφανίστηκε με την δουλειά του John Mather πάνω στα Lagrangian δυναμικά συστήματα. Οι action-minimizing καμπύλες είναι πολύ σημαντικά αντικείμενα στην θεωρία των δυναμικών συστημάτων, και η κατασκευή κλειστών action-minimizing καμπυλών που έχουν συγκεκριμένες ποιοτικές ιδιότητες είναι ένα κλασικό πρόβλημα. Στα τέλη της δεκαετίας του 1980, ο Mather θεώρησε βολικό να μελετήσει όχι μόνο action-minimizing καμπύλες αλλά και στάσιμα action-minimizing μέτρα σε phase χώρους, τα οποία και εισήγαγε στο άρθρο του "Minimal measures." (βλέπε [44]). Τα μέτρα αυτά του Mather αποτελούν γενικεύσεις των action-minimizing καμπυλών και έλυσαν ένα πρόβλημα μεταβολής το οποίο στην ουσία αποτελεί ένα πρόβλημα μεταφοράς Monge-Kantorovich. Κάτω από κάποιες συνθήκες για την Lagrangian. ο Mather απέδειξε ότι συγκεκριμένα action-minimizing μέτρα είναι αυτομάτως συγκεντρωμένα σε γραφήματα Lipschitz συναρτήσεων. Επιπλέον, απέδειξε το θεώρημα γραφήματος των Lipschitz συναρτήσεων στο άρθρο του "Action minimizing invariant measures for positive define Lagrangian systems" (βλέπε [45]). Η σαφής σύνδεση με το πρόβλημα των Monge-Kantorovich έγινε πρόσφατα στο άρθρο των Bernard και Buffoni, (βλέπε [11], κεφάλαιο 8).

Η δεύτερη κατεύθυνση έρευνας προέκυψε από την δουλειά του Yann Brenier. Μελετώντας προβλήματα στην ασυμπίεστη μηχανική ρευστών, ο Brenier χρειάστηκε να κατασκευάσει έναν τελεστή ο οποίος θα μπορούσε να συμπεριφέρεται όπως η προβολή στο σύνολο των απεικονίσεων που διατηρούν το μέτρο σε ένα ανοιχτό σύνολο. Ο τρόπος κατασκευής ήταν ο εξής: αν u είναι η απεικόνιση της οποίας θέλουμε να υπολογίσουμε την προβολή, εισάγουμε ένα ζεύγος (**coupling**) του μέτρου Lebesgue λ με το $u\#\lambda$. Αυτή η μελέτη αποκάλυψε έναν σύνδεσμο μεταξύ της βέλτιστης μεταφοράς και της μηχανικής ρευστών, ενώ ταυτόχρονα εξέφρασε την σχέση της βέλτιστης μεταφοράς με την θεωρία των εξισώσεων Monge-Ampère. Ο Brenier ανακοίνωσε τα κύρια αποτελέσματά του σε σύντομες σημειώσεις (βλέπε [15]) και αργότερα δημοσίευσε λεπτομερείς αποδείξεις στο [14]. Το κεφάλαιο 3 του βιβλίου του Villani [59] είναι αφιερωμένο στο **πολικό θεώρημα παραγοντοποίησης του Brenier**, την ερμηνεία του και τις συνέπειές του.

Η τρίτη κατεύθυνση έρευνας προήλθε έξω από τον χώρο των μαθηματικών. Ο Mike Cullen ήταν μέλος μίας ομάδας μετεωρολόγων, οι οποίοι είχαν καλά ανεπτυγμένες γνώσεις μαθηματικών, και δούλευαν πάνω σε ημι-γεωστροφικές εξισώσεις, οι οποίες χρησιμοποιούνταν στην μετεωρολογία για να μοντελοποιήσουν ατμοσφαιρικά fronts. Το ημι-γεωστροφικό σύστημα εισήχθη από τους Eliassen (βλέπε [26]) και Hoskins (βλέπε [31, 32, 33]). Ο Cullen και οι συνεργάτες του έδειξαν ότι μία συγκεκριμένη διάσημη αλλαγή αγνώστου, η οποία οφείλεται στον Brian Hoskins, μπορεί να μεταφραστεί με όρους ενός προβλήματος βέλτιστης μεταφοράς, και ταυτοποίησαν την ιδιότητα ελαχιστοποίησης ως μία συνθήκη σταθερότητας. Ένα μεγάλο αποτέλεσμα αυτής της δουλειάς ήταν ότι η βέλτιστη μεταφορά μπορούσε να προκύψει φυσικά μέσω μερικών διαφορικών εξισώσεων οι οποίες φαίνονταν να μην έχουν καμία σχέση με αυτήν. Ο Cullen και οι συνεργάτες του έγραψαν πολλά άρθρα πάνω σε αυτό το θέμα (βλέπε π.χ. [19], την δουλειά των Cullen και Gangbo ([20], των Cullen και Feldman ([21], και το πρόσφατο βιβλίο του Cullen ([22]).

Και οι τρεις συνεισφορές έκαναν σαφές ότι *σημαντική πληροφορία μπορεί να κερδηθεί από μία ποιοτική περιγραφή της βέλτιστης μεταφοράς*. Με αυτές τις νέες κατευθύνσεις ασχολήθηκαν πολλοί μαθηματικοί (όπως οι Luis Caffarelli, Craig Enans, Wilfrid Gangbo, Robert McCann και πολλοί άλλοι), οι οποίοι δούλεψαν πάνω σε μία καλύτερη περιγραφή της δομής της βέλτιστης μεταφοράς και βρήκαν νέες εφαρμογές.

Ένα σημαντικό εννοιολογικό βήμα επιτεύχθηκε από τον Felix Otto, ο οποίος ανακάλυψε έναν ελκυστικό φορμαλισμό εισάγοντας μία διαφορική ματιά στην θεωρία της βέλτιστης μεταφοράς. Αυτό άνοιξε τον δρόμο για μία πιο γεωμετρική περιγραφή του χώρου των μέτρων πιθανότητας και συνέδεσε την βέλτιστη μεταφορά με την θεωρία των εξισώσεων διάχυσης, γεγονός το οποίο οδήγησε στην στενή αλληλεπίδραση της γεωμετρίας με την συναρτησιακή ανάλυση και τις μερικές διαφορικές εξισώσεις.

Σήμερα η βέλτιστη μεταφορά ακμάζει ολοένα και περισσότερο, με πολλούς ερευνητές από διάφορους τομείς να σχολούνται με αυτή. Ιδιαίτερα τελευταία έχει κεντρίσει το ενδιαφέρον λόγω της σημαντικής συμβολής της σε προβλήματα μη γραμμικών διαφορικών εξισώσεων, στην γεωμετρία Riemann, σε προβλήματα μεταβολών, και σε πολλές ενδιαφέρουσες ανισότητες. Βλέπε συγκεκριμένα τις εξαιρετικές παρουσιάσεις των Ambrosio (2003), Villani (2003, 2006), Ambrosio, Gigli και Savaré (2005). Ήδη από το 1987, με το έργο του Brenier [15] είχε γίνει εμφανής η αλληλεπίδραση μεταξύ των μερικών διαφορικών εξισώσεων, της μηχανικής ρευστών, της γεωμετρίας, της θεωρίας πιθανοτήτων και της συναρτησιακής ανάλυσης, η οποία έχει αναπτυχθεί κατά το πέρασμα των τελευταίων δέκα χρόνων από τις συμβολές πολλών συγγραφέων που είχαν ως κοινή βάση προβλήματα βέλτιστης μεταφοράς.

Τα προβλήματα μεταφοράς μάζας εμφανίζονται με ποικίλες μορφές, σε ποικίλες περιοχές των μαθηματικών, και έχουν διατυπωθεί σε διαφορετικά επίπεδα γενικότητας. Για παράδειγμα, ενώ η συνεχής περίπτωση του προβλήματος μεταφοράς μπορεί να αποτιμηθεί με μετροθεωρητικούς όρους, η διακριτή περίπτωση έχει να κάνει με βελτιστοποίηση πάνω σε γενικές μεταφορές πολυέδρων.

Η κλασική πιθανοθεωρητική πλευρά της βέλτιστης μεταφοράς έχει εξεταστεί εκτενώς από τους Rachev και Rüschendorf (βλέπε [55]), με αξιοσημείωτες εφαρμογές που συμπεριλαμβάνουν οριακά θεωρήματα για διάφορες τυχαίες διαδικασίες. Επιπλέον, στον ίδιο τομέα ανήκουν και οι σχέσεις με την θεωρία παιγνίων, τα οικονομικά, την στατιστική και τον έλεγχο υποθέσεων.

Η συνεισφορά του Tanaka στην κινητική θεωρία έγινε στα μέσα της δεκαετίας του εβδομήντα (βλέπε [62], [63], [51]). Η έρευνά του συνεχίστηκε από τον Toscani και τους συνεργάτες του (βλέπε [13], [54]), όπου με την πρώτη ματιά βλέπει κανείς την σύνδεση με το πρόβλημα βέλτιστης μεταφοράς.

1.2 Περιγραφή της ροής της εργασίας

Σκοπός της εργασίας αυτής είναι να δώσει μία, όσο το δυνατόν, λεπτομερή και πλήρη παρουσιάση του προβλήματος της βέλτιστης μεταφοράς όπως αυτό πρωτοδιατυπώθηκε το 1781 από τον Monge, το πώς το πρόβλημα αυτό τροποποιήθηκε και επαναδιατυπώθηκε στο πέρασμα των χρόνων, αλλά και το πώς έγινε δυνατό να επιλυθεί αρκετό καιρό μετά από την διατύπωσή του. Επιπλέον, γίνεται αναφορά στο μεγάλο πλήθος των εφαρμογών της βέλτιστης μεταφοράς, χωρίς όμως να προχωρούμε σε λεπτομέρειες. Παραθέτουμε όμως πλούσια βιβλιογραφία όπου μπορεί κανείς να βρει εκτενείς παρουσιάσεις των εφαρμογών αυτών.

Εστιάζουμε περισσότερο στις γεωμετρικές ανισότητες, στο πώς αυτές μπορούν να προκύψουν με την χρήση αποδεικτικών εργαλείων της βέλτιστης μεταφοράς, και στο ποιοί νέοι δρόμοι ανοίγονται μέσω αυτών. Για την εργασία αυτή χρησιμοποιήθηκαν διάφορα βιβλία και άρθρα, και ιδιαίτερως τα βιβλία του Villani, "Topics in optimal transportation" (βλέπε [59]) και "Optimal transport old and new" (βλέπε [60]), όπως επίσης και τα άρθρα των McCann (βλέπε [46], [49], [48]) και Aleksandrov (βλέπε [3]).

Κεφάλαιο 2. Το κεφάλαιο αυτό αποτελεί ένα εισαγωγικό κεφάλαιο όπου διατυπώνονται βασικές έννοιες και εγαλεία από την θεωρία πιθανοτήτων, την θεωρία μέτρου και τα μέτρα Haussdorf, τα οποία χρησιμοποιούμε πολύ συχνά στην συνέχεια της εργασίας.

Πιο συγκεκριμένα, στην πρώτη ενότητα παρουσιάζονται οι βασικές ιδιότητες που ικανοποιούν τα μέτρα πιθανότητας σε Πολωνικούς χώρους, δηλαδή πλήρεις και διαχωρίσιμους χώρους, και διατυπώνεται το πολύ χρήσιμο θεώρημα του Ascoli.

Στη δεύτερη ενότητα ορίζονται τα μέτρα Haussdorf, παρουσιάζεται η κατασκευή του Καραθεοδωρή και τέλος ορίζεται η διάσταση Haussdorf και οι βασικές της ιδιότητες.

Τέλος, στην τρίτη ενότητα παρουσιάζονται οι βασικές ιδιότητες που χαρακτηρίζουν την οικογένεια των κυρτών συναρτήσεων. Ορίζουμε το υποδιαφορικό μίας κυρτής συνάρτησης, τον μετασχηματισμό Legendre της, εξετάζουμε την διαφορισιμότητά της και τέλος διατυπώνουμε τον δυϊσμό Legendre για κάτω ημισυνεχείς συναρτήσεις. Ο λόγος που ενδιαφερόμαστε για τις ιδιότητες και την συμπεριφορά αυτών των συναρτήσεων είναι ότι τα ζεύγη των συναρτήσεων που ελαχιστοποιούν το δυϊκό πρόβλημα του Kantorovich είναι ζεύγη κυρτών και κάτω ημισυνεχών συναρτήσεων.

Κεφάλαιο 3. Στο κεφάλαιο αυτό παρουσιάζουμε το πρόβλημα της βέλτιστης μεταφοράς όπως αυτό διατυπώθηκε από τον Kantorovich και διατυπώνουμε το θεώρημα δυϊσμού του Kantorovich (βλέπε Θεώρημα 3.2.1). Το θεώρημα δυϊσμού του Kantorovich αποτελεί ένα πολύ σημαντικό εργαλείο, το οποίο χρησιμοποιείται εκτενώς σε διάφορα προβλήματα βελτιστοποίησης. Ο Kantorovich, στην ουσία μετέτρεψε το πρόβλημα που μελετούσε σε ένα πρόβλημα supremum/infimum και χρησιμοποιώντας εργαλεία γραμμικού προγραμματισμού κατάφερε να το επιλύσει.

Στη συνέχεια, παρουσιάζουμε το θεώρημα των Kantorovich-Rubinstein, το οποίο στην ουσία αποτελεί εφαρμογή του θεωρήματος δυϊσμού σε έναν συμπαγή χώρο εφοδιασμένο με την μετρική ως συνάρτηση κόστους (βλέπε Θεώρημα 3.3.1). Το θεώρημα των Kantorovich-Rubinstein επαναδιατυπώθηκε την δεκαετία του εβδομήντα από τον Dudley (βλέπε [24]), ο οποίος παρακινήθηκε από προβλήματα της μαθηματικής στατιστικής (βλέπε [25]) και αργότερα διατυπώθηκε σε ακόμη πιο γενικό πλαίσιο από διάφορους συγγραφείς (βλέπε [55]). Στην εργασία αυτή το πλαίσιό μας είναι αρκετά γενικό, δουλεύουμε σε έναν Πολωνικό χώρο και η συνάρτηση κόστους είναι τυχούσα κάτω ημισυνεχής συνάρτηση.

Κεφάλαιο 4. Στο κεφάλαιο αυτό μελετάμε τι πληροφορίες μπορεί κανείς να λάβει για τα μέτρα μ, ν αν γνωρίζει την τιμή $T_c(\mu,\nu)$ του βέλτιστου κόστους μεταφοράς. Η απάντηση σε αυτό το ερώτημα είναι ακριβώς το Θεώρημα 4.2.1. Στο Θεώρημα 4.1.2 παρουσιάζουμε τις ιδιότητες των αποστάσεων Monge-Kantorovich ή αλλιώς των αποστάσεων Wasserstein, οι οποίες είναι αποστάσεις που επάγονται από την τιμή του βέλτιστου κόστους μεταφοράς, όταν το κόστος είναι μία δύναμη της απόστασης. Το σημαντικό σημείο εδώ είναι πως οι ιδιότητες των αποστάσεων Wasserstein δεν εξαρτόνται από την εκάστοτε γεωμετρική δομή και συνεπώς τα αποτελέσματα του κεφαλαίου αυτού ισχύουν σε ένα πιο γενικό πλαίσιο (Πολωνικούς Χώρους).

Κεφάλαιο 5. Στο κεφάλαιο αυτό παρουσιάζουμε τα πιο σημαντικά αποτελέσματα αναφορικά με την ύπαρξη και των χαρακτηρισμό του βέλτιστου σχέδιου μεταφοράς. Όπως θα δούμε στην πορεία, είναι συνήθως πιο εύκολο να δοθεί μία απόδειξη για την ύπαρξη ενός σχεδίου μεταφοράς που ελαχιστοποιεί το πρόβλημα του Kantorovich σε αντίθεση με εκείνο του Monge. Τα τελευταία πενήντα χρόνια, έχει αναπτυχθεί μία συστηματική και κομψή θεωρία για το πρόβλημα του Monge από τους Brenier, Evans, Gangbo, Knott και Smith, McCann, Rachev, Rüschendorf και άλλους, ιδιαιτέρως στην περίπτωση όπου έχουμε τετραγωνική συνάρτηση κόστους, $c(x,y) = |x-y|^2$ στον $\mathbb{R}^n$.

Το βασικό αποτέλεσμα για την τετραγωνική συνάρτηση κόστους (βλέπε Θεώρημα 5.2.3) ανακαλύφθηκε τουλάχιστον δύο φορές, αρχικά από τους Knott και Smith και αργότερα από τον Brenier, ο οποίος παρακινήθηκε από την μελέτη της μηχανικής των ρευστών. Οι Knott και Smith, αλλά και ο Brenier απαιτούν στο θεώρημα βέλτιστης μεταφοράς για τετραγωνικό κόστος (βλέπε Θεώρημα 5.2.3), τα μέτρα πιθανότητας μ, ν στον $\mathbb{R}^n$ να έχουν πεπερασμένες ροπές δεύτερης τάξης. Η υπόθεση αυτή είναι πολύ σημαντική γιατί τότε μπορούμε να εφαρμόσουμε το λήμμα διπλής κυρτοποίησης (βλέπε Λήμμα 5.2.2) και να αποδείξουμε το θεώρημα ύπαρξης βέλτιστου ζεύγους συζυγών κυρτών συναρτήσεων (φ, φ^*) (βλέπε Θεώρημα 5.2.1) οι οποίες ελαχιστοποιούν την ποσότητα

$$J(\varphi,\psi) = \int_X \varphi(x)d\mu(x) + \int_Y \psi d\nu(y)$$

πάνω από το σύνολο $\tilde{\Phi}$, το οποίο αποτελείται από όλα τα ζεύγη $(\varphi,\psi) \in L^1(\mu) \times L^1(\nu)$ με τιμές στο $\mathbb{R} \cup \{+\infty\}$, τα οποία έχουν την ιδιότητα

$$\langle x, y\rangle \leqslant \varphi(x) + \psi(y).$$

Το θεώρημα αυτό, με την σειρά του χρησιμοποιείται σε συνδυασμό με τον δυϊσμό του Kantorovich:

$$\inf I(\pi) = \sup_{\tilde{\Phi}} J(\varphi, \varphi^*)$$

για την απόδειξη του κριτηρίου των Knott και Smith, αλλά και του θεωρήματος του Brenier (βλέπε Θεώρημα 5.2.3).

Από την άλλη πλεψρ·α, ο McCann επαναδιατύπωσε το θεώρημα του Brenier χωρίς να χρησιμοποιήσει τον δυϊσμό του Kantorovich. Τα εργαλεία του είναι γεωμετρικά, χρησιμοποιεί την έννοια της κυκλικής μονοτονίας και τα θεωρήματα των Rockafellar και Aleksandrov. Δεν προϋποθέτει τις πεπερασμένες ροπές δεύτερης τάξης όπως οι Knott και Smith, παρά μόνο υποθέτει ότι το μέτρο πιθανότητας μ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue.

Εκτός από τον McCann, τα αποτελέσματα των Knott-Smith και Brenier γενικεύτηκαν αργότερα και από τους Rachev και Rüschendorf. Για πολλούς ερευνητές όμως το θεώρημα αυτό είναι συνδεδεμένο με το όνομα του Brenier, γιατί ήταν ο πρώτος που παρουσίασε εφαρμογές της μεταφοράς της μάζας σε προβλήματα κλασικής μηχανικής και μαθηματικής φυσικής.

Όμοια αποτελέσματα με εκείνα της τετραγωνικής συνάρτησης κόστους έχουμε και στην περίπτωση όπου η συνάρτηση κόστους είναι της μορφής $|x-y|^p$, ή πιο γενικά $d(x,y)^p$ πάνω σε μία πολλαπλότητα Riemann. Στην εργασία αυτή πραγματευόμαστε αποκλειστικά την περίπτωση της τετραγωνικής συνάρτησης κόστους. Για την συνάρτηση κόστους $c(x,y) = |x-y|$ μπορεί κανείς συμβουλευτεί τις σημειώσεις του Evans (βλέπε [28]) και ιδιαιτέρως εκείνες του Ambrosio (βλέπε [5]) και των Ambrosio και Pratelli (βλέπε [6]).

Κλείνουμε το κεφάλαιο αυτό παρουσιάζοντας την βέλτιστη μεταφορά στο $\mathbb{R}$ και διατυπώνοντας το πολικό θεώρημα παραγοντοποίησης του Bernier (βλέπε θεώρημα 5.5.5), το οποίο ουσιαστικά είναι ισοδύναμο με το θεώρημα της βέλτιστης μεταφοράς.

Κεφάλαιο 6. Στο κεφάλαιο αυτό βλέπουμε πώς η θεωρία της μεταφοράς της μάζας παρέχει πολυ χρήσιμα εργαλεία για την μελέτη συναρτησιακών ανισοτήτων. Κύριο αντικείμενο αυτού του κεφαλαίου είναι οι γεωμετρικές ανισότητες, κοινό σημείο των οποίων είναι η ισοπεριμετρική ανισότητα. Διατυπώνουμε την ανισότητα Brunn-Minkowski και την ανισότητα Prékopa-Leindler, η οποία αποτελεί την συναρτησιακή μορφή της πρώτης, και τις αποδεικνύουμε μέσω εργαλείων της βέλτιστης μεταφοράς. Πιο συγκεκριμένα χρησιμοποιούμε την παρεμβολή του McCann και την κυρτότητα ως προς μετατόπιση.

Στην συνέχεια παρουσιάζουμε την ανισότητα Brascamp-Lieb και την αντίστροφή της έτσι όπως διατυπώθηκε και αποδείχθηκε από τον Barthe και τέλος, ασχολούμαστε με τις ανισότητες Sobolev, για τις οποίες χρησιμοποιούμε υλικό από την δουλειά των Cordero-Erausquin, Nazaret και Villani (βλέπε [18]).

Κεφάλαιο 2

Βασικά εργαλεία

Σε αυτό το Κεφάλαιο συγκεντρώνουμε κάποια βασικά εργαλεία από την θεωρία μέτρου, την θεωρία πιθανοτήτων και την κυρτή ανάλυση, τα οποία θα χρησιμοποιούνται συχνά στα επόμενα. Δίνουμε κάποιους ορισμούς και βασικά αποτελέσματα – σε μερικές περιπτώσεις συμπεριλαμβάνουμε σκιαγράφηση των αποδείξεών τους.

2.1 Εργαλεία θεωρίας πιθανοτήτων και θεωρίας μέτρου

Ένας μετρικός χώρος X ονομάζεται *Πολωνικός* αν είναι πλήρης και διαχωρίσιμος. Για κάθε μετρικό χώρο X συμβολίζουμε με $\mathcal{P}(X)$ την οικογένεια των Borel μέτρων πιθανότητας στον X. Θα χρησιμοποιούμε τις παρακάτω βασικές ιδιότητες που έχουν τα μέτρα πιθανότητας που ορίζονται σε Πολωνικούς χώρους.

(i) Ένα Borel μέτρο πιθανότητας σε έναν Πολωνικό χώρο X είναι αυτομάτως κανονικό (βλέπε [12]), δηλαδή

$$\mu(A) = \sup\{\mu(K) : K \subset A \text{ συμπαγές}\} = \inf\{\mu(U) : A \subset U \text{ ανοιχτό}\}.$$

(ii) Ένα μέτρο πιθανότητας μ σε έναν Πολωνικό χώρο X είναι συγκεντρωμένο σε ένα σ-συμπαγές σύνολο (βλέπε [12]): υπάρχει μετρήσιμο σύνολο S το οποίο μπορεί να γραφεί ως ένωση αριθμήσιμων το πλήθος συμπαγών συνόλων, ώστε $\mu(S) = 1$. Ισοδύναμα, το μ είναι **tight**: για κάθε $\varepsilon > 0$ υπάρχει $K_\varepsilon \subseteq X$ συμπαγές ώστε $\mu(K_\varepsilon^c) \leqslant \varepsilon$. Το αποτέλεσμα αυτό είναι γνωστό ως **λήμμα του Ulam**.

(iii) Μια οικογένεια P από μέτρα πιθανότητας σε έναν τοπολογικό χώρο X λέγεται tight αν για κάθε $\varepsilon > 0$ υπάρχει $K_\varepsilon \subseteq X$ συμπαγές για το οποίο

$$\sup_{\mu \in P} \mu(K_\varepsilon^c) \leqslant \varepsilon.$$

Αν X είναι ένας Πολωνικός χώρος, τότε το θεώρημα του Prokhorov ισχυρίζεται ότι κάθε tight οικογένεια P στον $\mathcal{P}(X)$ είναι σχετικά ακολουθιακά συμπαγής: για κάθε ακολουθία (μ_n) στην οικογένεια P υπάρχουν μια υπακολουθία (μ_{k_n}) και ένα μέτρο πιθανότητας μ_* στον X ώστε, για κάθε $\varphi \in C_b(X)$,

$$\lim_{n\to\infty} \int_X \varphi \, d\mu_{k_n} = \int_X \varphi \, d\mu_*,$$

όπου $C_b(X)$ είναι το σύνολο των συνεχών και φραγμένων συναρτήσεων $\varphi : X \to \mathbb{R}$. Το αποτέλεσμα αυτό μπορεί να γενικευτεί και σε πλαίσιο ευρύτερο από αυτό των Πολωνικών χώρων.

Ορισμός 2.1.1. Μια συνάρτηση F σε έναν μετρικό χώρο X καλείται **κάτω ημισυνεχής** αν για κάθε $x \in X$ ισχύει $F(x) \leqslant \liminf_{y\to x} F(y)$. Ισοδύναμα, αν για κάθε $t \in \mathbb{R}$ το σύνολο $\{x \in X : f(x) \leq t\}$ είναι κλειστό υποσύνολο του X.

Θα δείξουμε ότι η f είναι κάτω ημισυνεχής αν και μόνο αν, για κάθε ακολουθία (x_n) στον X με $(x_n) \to x \in X$, ισχύει

$$f(x) \leq \lim_{n\to\infty} \inf f(x_n).$$

Έστω ότι η f είναι κάτω ημισυνεχής και έστω (x_n) στον X με $x_n \to x \in X$. Θέτουμε $t = \lim_{n\to\infty} \inf f(x_n)$. Αν $f(x) > t$ τότε υπάρχει $\varepsilon > 0$ ώστε $x \notin F_{t+\varepsilon} = \{z \in X : f(z) \leq t+\varepsilon\}$. Αφού το $F_{t+\varepsilon}$ είναι κλειστό, υπάρχει $\delta > 0$ ώστε $F_{t+\varepsilon} \cap B(x,\delta) = \emptyset$. Όμως, υπάρχει υπακολουθία $x_{k_n} \to x$ ώστε $f(x_{k_n}) \to t$. Αυτο σημαίνει ότι τελικά θα ισχύει ότι $x_{k_n} \in B(x,\delta)$ και $f(x_{k_n}) < t+\varepsilon$, δηλαδή $x_{k_n} \in F_{t+\varepsilon}$. Έπεται ότι $F_{t+\varepsilon} \cap B(x,\delta) \neq \emptyset$, το οποίο είναι άτοπο. Αντίστροφα, έστω $F_t = \{z \in X : f(z) \leq t\}, t \in \mathbb{R}$. Θεωρούμε $x \in \bar{F}_t$. Τότε υπάρχει (x_n) στο F_t με $x_n \to x$. Για κάθε $n \in \mathbb{N}$ έχουμε $f(x_n) \leq t$, άρα $\lim_{n\to\infty} \inf f(x_n) \leq t$. Από την υπόθεση, $f(x) \leq \lim_{n\to\infty} \inf f(x_n) \leq t$, δηλαδή $x \in F_t$. Άρα, το F_t είναι κλειστό.

Αν X είναι ένας μετρικός χώρος και F μία μη αρνητική, κάτω ημισυνεχής συνάρτηση στον X, τότε η F μπορεί να γραφεί ως το supremum μιας αύξουσας ακολουθίας ομοιόμορφα συνεχών μη αρνητικών συναρτήσεων. Για παράδειγμα, μπορούμε να επιλέξουμε

$$F_n(x) = \inf_{y \in X} [F(y) + nd(x,y)]$$

όπου d είναι μια μετρική στον X.

Κλείνοντας την ενότητα αυτή θα διατυπώσουμε το θεώρημα Ascoli, το οποίο περιγράφει τα συμπαγή υποσύνολα του χώρου Banach $C(X)$, όπου X συμπαγής τοπολογικός χώρος με την νόρμα της ομοιόμορφης σύγκλισης, χρησιμοποιώντας την έννοια της ισοσυνέχειας (βλέπε [1]).

Ορισμός 2.1.2. Έστω X τοπολογικός χώρος, $x_0 \in X$ και $F \subset C(X)$. Το F λέγεται *ισοσυνεχές* στο x_0 αν για κάθε $\varepsilon > 0$ υπάρχει ανοιχτή περιοχή U του x_0 με $|f(x)-f(x_0)| < \varepsilon$ για κάθε $x \in U$ και $f \in F$. Το F λεγεται ισοσυνεχές στον X αν είναι ισοσυνεχές σε κάθε σημείο του X. Είναι σαφές ότι αν το F είναι ισοσυνεχές και $G \subset F$ τότε και το G είναι ισοσυνεχές.

Πρόταση 2.1.3. *Έστω X συμπαγής χώρος και $F \subset C(X)$. Τότε,*

(i) *αν το F είναι πεπερασμένο τότε είναι ισοσυνεχές στο X, και*

(ii) *αν το F είναι συμπαγές στην τοπολογία της ομοιόμορφης σύγκλισης τότε είναι ισοσυνεχές στο X.*

Απόδειξη. (i) Έστω $F = \{f_1, \ldots, f_k\}$, $x_0 \in X$ και $\varepsilon > 0$. Αφού κάθε f_i είναι συνεχής στο x_0, υπάρχει ανοιχτή περιοχή U_i του x_0 με $|f_i(x) - f_i(x_0)| < \varepsilon$ για κάθε $x \in U_i$. Είναι σαφές ότι αν θέσουμε $U = U_1 \cap U_2 \cap \ldots \cap U_k$ τότε ικανοποιείται ο ορισμός της ισοσυνέχειας του συνόλου F στο σημείο x_0.

(ii) Έστω $F \subset C(X)$ συμπαγές και $\varepsilon > 0$. Η οικογένεια $\{S(f, \frac{\varepsilon}{3}) : f \in F\}$ είναι ένα ανοιχτό κάλυμμα του F και άρα υπάρχουν $f_1, f_2, \ldots, f_k \in F$ ώστε $F \subset S(f_1, \frac{\varepsilon}{3}) \cup .. \cup S(f_k, \frac{\varepsilon}{3})$. Το σύνολο $A = \{f_1, \ldots, f_k\}$ είναι ισοσυνεχές στο X από το (i) και άρα για κάθε $x \in X$ υπάρχει ανοιχτή περιοχή U του x με $|f_i(y) - f_i(x)| < \frac{\varepsilon}{3}$ για κάθε $y \in U$. Συνεπώς,

$$\begin{aligned} |f(y) - f(x)| &\leqslant |f(y) - f_i(y)| + |f_i(y) - f_i(x)| + |f_i(x) - f(x)| \\ &\leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

για κάθε $y \in U$ και $f \in F$. □

Λήμμα 2.1.4. *Έστω X συμπαγής χώρος και F ομοιόμορφα φραγμένο υποσύνολο του $C(X)$. Τότε το F είναι ισοσυνεχές αν και μόνο αν το F είναι ολικά φραγμένο στη νόρμα της ομοιόμορφης σύγκλισης.*

Θεώρημα 2.1.5. (*Ascoli*). Έστω X συμπαγής χώρος, $C(X)$ ο χώρος των συνεχών πραγματικών συναρτήσεων επί του X με την τοπολογία της ομοιόμορφης σύγκλισης και $F \subset C(X)$. Το F είναι συμπαγές αν και μόνο αν το F είναι κλειστό, φραγμένο και ισοσυνεχές.

Απόδειξη. (⇒) Έστω F συμπαγές. Τότε, το F είναι φραγμένο και κλειστό, ενώ από την Πρόταση 2.1.3 έχουμε ότι είναι και ισοσυνεχές.

(⇐) Έστω F κλειστό, φραγμένο και ισοσυνεχές. Αφού το F είναι κλειστό και ο $C(X)$ είναι πλήρης μετρικός χώρος, έπεται ότι το F είναι πλήρης μετρικός χώρος. Από το Λήμμα 2.1.4 έχουμε ότι το F είναι ολικά φραγμένο. Συνεπώς, το F είναι συμπαγές. □

2.2 Μέτρα και διάσταση Hausdorff

Έστω (X,d) μετρικός χώρος, $\mathcal{F}$ μια οικογένεια υποσυνόλων του X, και ζ μια πραγματική συνάρτηση ορισμένη στην $\mathcal{F}$, με μη αρνητικές τιμές. Υποθέτουμε ότι

(1) Για κάθε $\delta > 0$ υπάρχουν $E_1, E_2, \ldots \in \mathcal{F}$ τέτοια ώστε $X = \bigcup_{i=1}^{\infty} E_i$ και $\mathrm{diam}(E_i) \leqslant \delta$.

(2) Για κάθε $\delta > 0$ υπάρχει $E \in \mathcal{F}$ τέτοιο ώστε $\zeta(E) \leqslant \delta$ και $\mathrm{diam}(E) \leqslant \delta$.

Για κάθε $0 < \delta \leqslant \infty$ και $A \subset X$ ορίζουμε

$$\psi_\delta(A) := \inf\left\{\sum_{i=1}^{\infty} \zeta(E_i) : A \subset \bigcup_{i=1}^{\infty} E_i,\ \mathrm{diam}(E_i) \leqslant \delta,\ E_i \in \mathcal{F}\right\}. \tag{2.2.1}$$

Η υπόθεση (1) παραπάνω εξασφαλίζει ότι τέτοια καλύμματα υπάρχουν πάντα. Από την (2) βλέπουμε ότι $\psi_\delta(\emptyset) = 0$. Επίσης, λόγω της (2) μπορούμε να χρησιμοποιούμε καλύμματα $\{E_i\}_{i\in I}$ με το σύνολο δεικτών I πεπερασμένο ή αριθμήσιμο, χωρίς να αλλάζει η τιμή του $\psi_\delta(A)$.

Παρατηρούμε ότι η συνάρτηση ψ_δ είναι μονότονη και υποπροσθετική, οπότε είναι ένα εξωτερικό μέτρο. Γενικά ωστόσο αποτυγχάνει να είναι προσθετική. Παρατηρούμε όμως ότι για $0 < \varepsilon < \delta \leqslant \infty$ έχουμε $\psi_\delta(A) \leqslant \psi_\varepsilon(A)$. Μπορούμε λοιπόν να ορίσουμε μια συνολοσυνάρτηση $\psi = \psi(\mathcal{F}, \zeta)$ ως εξής:

$$\psi(A) := \lim_{\delta\to 0} \psi_\delta(A) = \sup_{\delta>0} \psi_\delta(A). \tag{2.2.2}$$

Η μετροθεωρητική συμπεριφορά της ψ είναι πολύ καλύτερη από αυτή της ψ_δ. Συγκεκριμένα, ισχύει το επόμενο θεώρημα.

Θεώρημα 2.2.1.

(1) *Η ψ είναι μέτρο Borel.*

(2) *Αν τα στοιχεία της $\mathcal{F}$ είναι σύνολα Borel, τότε το ψ είναι κανονικό μέτρο Borel.*

Απόδειξη. (1) Η μονοτονία και η υποπροσθετικότητα της ψ έπονται άμεσα από τις αντίστοιχες ιδιότητες της ψ_δ. Έστω τώρα $A, B \subseteq X$ με $d(A,B) > 0$. Επιλέγουμε δ τέτοιο ώστε $0 < \delta < d(A,B)/2$. Αν τα σύνολα $E_1, E_2, \ldots \in \mathcal{F}$ καλύπτουν το $A \cup B$ και ικανοποιούν την $\mathrm{diam}(E_i) \leqslant \delta$, τότε κανένα τους δεν τέμνει ταυτόχρονα τα A και B. Έτσι,

$$\sum_{i=1}^{\infty} \zeta(E_i) \geqslant \sum_{\substack{i=1 \\ A\cap E_i \neq \emptyset}}^{\infty} \zeta(E_i) + \sum_{\substack{i=1 \\ B\cap E_i \neq \emptyset}}^{\infty} \zeta(E_i) \geqslant \psi_\delta(A) + \psi_\delta(B).$$

Παίρνοντας το infimum πάνω από όλα αυτά τα καλύμματα έχουμε $\psi_\delta(A \cup B) \geqslant \psi_\delta(A) + \psi_\delta(B)$. Όμως η ψ_δ είναι και υποπροσθετική, άρα $\psi_\delta(A\cup B) = \psi_\delta(A) + \psi_\delta(B)$ και αφήνοντας το $\delta \to 0$ συμπεραίνουμε ότι $\psi(A \cup B) = \psi(A) + \psi(B)$.

Για την (2) θα δείξουμε ότι για κάθε $A \subseteq X$ υπάρχει Borel σύνολο B τέτοιο ώστε $A \subseteq B$ και $\psi(A) = \psi(B)$. Έστω λοιπόν $A \subseteq X$. Για κάθε $n \in \mathbb{N}$ επιλέγουμε σύνολα $E_{n,1}, E_{n,2}, \ldots \in \mathcal{F}$ τέτοια ώστε

$$A \subseteq \bigcup_{i=1}^{\infty} E_{n,i}, \quad \mathrm{diam}(E_{n,i}) \leqslant \frac{1}{n}$$

και

$$\sum_{i=1}^{\infty} \zeta(E_{n,i}) \leqslant \psi_{\frac{1}{n}}(A) + \frac{1}{n}.$$

Τότε το $B = \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{\infty} E_{n,i}$ είναι Borel σύνολο, και ικανοποιεί τις $A \subseteq B$ και $\psi(A) = \psi(B)$. □

Έστω τώρα X διαχωρίσιμος μετρικός χώρος. Θεωρούμε την οικογένεια $\mathcal{F} = 2^X$ όλων των υποσυνόλων του X, και για δοθέν $0 \leqslant s < \infty$ ορίζουμε $\zeta(E) = \zeta_s(E) = \mathrm{diam}(E)^s$ (συμφωνούμε ότι $0^0 = 1$ και $\mathrm{diam}(\emptyset)^s = 0$). Το μέτρο ψ που προκύπτει με την διαδικασία της προηγούμενης παραγράφου λέγεται *s-διάστατο μέτρο Hausdorff* και συμβολίζεται με $\mathcal{H}^s$. Με άλλα λόγια,

$$\mathcal{H}^s(A) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(A) = \sup_{\delta > 0} \mathcal{H}^s_\delta(A),$$

όπου

$$\mathcal{H}^s_\delta(A) = \inf \left\{ \sum_{i=1}^{\infty} \mathrm{diam}(E_i)^s : A \subseteq \bigcup_{i=1}^{\infty} E_i, \, \mathrm{diam}(E_i) \leqslant \delta \right\}.$$

Θα λέμε επίσης ότι το $\mathcal{H}^s_\infty(A)$ είναι το *s-Hausdorff περιεχόμενο* του A. Για $s = 0$, είναι προφανές ότι το $\mathcal{H}^s$ ταυτίζεται με το αριθμητικό μέτρο. Στην περίπτωση που $X = \mathbb{R}^n$ και $s = n$ είναι εύκολο να δούμε ότι $\mathcal{H}^n = c \cdot \mathrm{vol}_n$ για κάποια σταθερά $c > 0$, όπου με vol_n συμβολίζουμε το n-διάστατο μέτρο Lebesgue. Αυτό προκύπτει από το γεγονός ότι τόσο το $\mathcal{H}^n$ όσο και το vol_n είναι ομοιόμορφα κατανεμημένα και κανονικά Borel μέτρα.

Το επόμενο θεώρημα δίνει κάποιες βασικές ιδιότητες του $\mathcal{H}^s$. Ειδικότερα, εξασφαλίζει ότι είναι κανονικό Borel μέτρο.

Θεώρημα 2.2.2. *Έστω X διαχωρίσιμος μετρικός χώρος. Έστω $0 \leqslant s < n$ και $\zeta(E) = \mathrm{diam}(E)^s$, για $E \subseteq X$. Αν είτε*

(1) $\mathcal{F} = \{F \subseteq X : F$ κλειστό $\}$, είτε

(2) $\mathcal{F} = \{U \subseteq X : U$ ανοικτό $\}$, είτε

(3) $X = \mathbb{R}^n$ και $\mathcal{F} = \{K \subseteq X : K$ κυρτό $\}$,

τότε $\psi(\mathcal{F}, \zeta) = \mathcal{H}^s$.

Οι ιδιότητες (1) και (3) έπονται από το γεγονός ότι η κλειστή θήκη και η κυρτή θήκη ενός $E \subseteq X$ έχουν την ίδια διάμετρο με το E, ενώ η (2) από το γεγονός ότι, για κάθε $\varepsilon > 0$, το $\{x : d(x, E) < \varepsilon\}$ είναι ανοικτό και έχει διάμετρο το πολύ ίση με $\mathrm{diam}(E) + 2\varepsilon$. Είναι άμεσο τώρα, από το Θεώρημα 2.2.1(2), ότι

Πόρισμα 2.2.3. *Το $\mathcal{H}^s$ είναι κανονικό Borel μέτρο.*

Μια ενδιαφέρουσα παρατήρηση, που θα χρησιμοποιηθεί στον ορισμό της διάστασης Hausdorff αμέσως παρακάτω, παρουσιάζεται στο επόμενο θεώρημα.

Θεώρημα 2.2.4. *Για κάθε $0 \leqslant s < t < \infty$ και $A \subseteq X$,*

(1) *Αν $\mathcal{H}^s(A) < \infty$, τότε $\mathcal{H}^t(A) = 0$.*

(2) *Αν $\mathcal{H}^t(A) > 0$, τότε $\mathcal{H}^s(A) = \infty$.*

Απόδειξη. Είναι σαφές ότι οι (1) και (2) είναι ισοδύναμες. Για την απόδειξη της (1), έστω $A \subseteq \bigcup_{i=1}^{\infty} E_i$ με $\mathrm{diam}(E_i) \leqslant \delta$ και $\sum_{i=1}^{\infty} \mathrm{diam}(E_i)^s \leqslant \mathcal{H}^s_\delta(A) + 1$. Τότε

$$\mathcal{H}^t_\delta(A) \leqslant \sum_{i=1}^{\infty} \mathrm{diam}(E_i)^t \leqslant \delta^{t-s} \sum_{i=1}^{\infty} \mathrm{diam}(E_i)^s \leqslant \delta^{t-s}(\mathcal{H}^s_\delta(A) + 1),$$

και παίρνοντας το όριο καθώς $\delta \to 0$, βλέπουμε ότι αν $\mathcal{H}^s(A) < \infty$, τότε $\mathcal{H}^t(A) = 0$. □

Βασιζόμενοι στο Θεώρημα 2.2.4 δίνουμε τον παρακάτω ορισμό.

Ορισμός 2.2.5 (διάσταση Hausdorff). *Η διάσταση Hausdorff ενός $A \subseteq X$ ορίζεται ως εξής:*

$$\begin{aligned} \dim_H(A) &:= \sup\{s : \mathcal{H}^s(A) > 0\} = \sup\{s : \mathcal{H}^s(A) = \infty\} \\ &= \inf\{t : \mathcal{H}^t(A) < \infty\} = \inf\{t : \mathcal{H}^t(A) = 0\}. \end{aligned}$$

Εξ' ορισμού της, η διάσταση Hausdorff έχει τις ιδιότητες της μονοτονίας και «ευστάθειας» ως προς τις αριθμήσιμες ενώσεις:

$$\begin{aligned} \dim_H(A) &\leqslant \dim_H(B) && \text{για } A \subseteq B \subseteq X, \\ \dim_H\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sup_{n\in\mathbb{N}} \dim_H(A_n) && \text{για } A_n \subseteq X, n \in \mathbb{N}. \end{aligned}$$

Επαναδιατυπώνοντας τον Ορισμό 2.2.5, μπορούμε να πούμε ότι η $\dim_H(A)$ είναι ο μοναδικός αριθμός (πιθανόν και $\dim_H(A) = \infty$) για τον οποίο

$$\begin{aligned} s < \dim_H(A) &\Longrightarrow \mathcal{H}^s(A) = \infty, \\ t > \dim_H(A) &\Longrightarrow \mathcal{H}^t(A) = 0. \end{aligned}$$

Στην περίπτωση που $\dim_H(A) = s$ δεν μπορούμε γενικά να γνωρίζουμε την τιμή του $\mathcal{H}^s(A)$: τα τρία ενδεχόμενα $\mathcal{H}^s(A) = 0$, $0 < \mathcal{H}^s(A) < \infty$, $\mathcal{H}^s(A) = \infty$ είναι όλα πιθανά. Αν όμως για δεδομένο A μπορούμε να βρούμε s τέτοιο ώστε $0 < \mathcal{H}^s(A) < \infty$, τότε αναγκαστικά $s = \dim_H(A)$.

2.3 Στοιχεία κυρτής ανάλυσης

Σε αυτήν την ενότητα θα δώσουμε κάποιους ορισμούς και εργαλεία της κυρτής ανάλυσης που θα χρησιμοποιηθούν στις επόμενες ενότητες, (βλέπε [57] και [59]). Στα επόμενα, όποτε αναφερόμαστε σε «μικρά» σύνολα θα εννοούμε υποσύνολα του $\mathbb{R}^n$ που έχουν διάσταση Hausdorff μικρότερη από n (μπορεί όμως κανείς να τα σκέφτεται σαν σύνολα που έχουν μέτρο Lebesgue ίσο με το 0).

Ορισμός 2.3.1. Μια **γνήσια κυρτή** συνάρτηση φ στον $\mathbb{R}^n$ είναι μια συνάρτηση $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, όχι ταυτοτικά $+\infty$, που ικανοποιεί την

$$\varphi(tx + (1-t)y) \leqslant t\varphi(x) + (1-t)\varphi(y)$$

για κάθε $x, y \in \mathbb{R}^n$ και για κάθε $t \in [0,1]$. Η φ λέγεται *αυστηρώς κυρτή* όταν από την ισότητα στην παραπάνω ανισότητα έπεται ότι $x = y$, ή $t = 0$, ή $t = 1$.

Το **πεδίο** της φ, $\mathrm{Dom}(\varphi)$, ορίζεται ως το κυρτό σύνολο των σημείων όπου η φ παίρνει πεπερασμένη τιμή. Ως κυρτό σύνολο το $\mathrm{Dom}(\varphi)$ έχει σύνορο με μηδενικό μέτρο. Πράγματι, έστω C ένα κυρτό σύνολο και έστω ότι $\lambda(\partial C) > 0$, όπου λ είναι το μέτρο Lebesgue στον $\mathbb{R}^n$. Τότε από το θεώρημα πυκνότητας του Lebesgue θα είχαμε ότι για λ-σχεδόν όλα τα $x \in \partial C$,

$$\lim_{\delta\to 0^+} \frac{\lambda(\bar{C} \cap B(x,\delta))}{\lambda(B(x,\delta))} = 1,$$

όπου $B(x,\delta)$ είναι η μπάλα με κέντρο το x και ακτίνα δ. Όμως, θεωρώντας ένα υπερεπίπεδο στήριξης του C στο σημείο x, βλέπουμε ότι

$$\frac{\lambda(\bar{C}\cap B(x,\delta))}{\lambda(B(x,\delta))} \leqslant \frac{1}{2}$$

για κάθε $\delta > 0$, το οποίο είναι άτοπο. Άρα, $\lambda(\partial C) = 0$.

Αποδεικνύεται ότι μια γνήσια κυρτή συνάρτηση φ είναι αυτομάτως συνεχής και τοπικά Lipschitz στο $\mathrm{Int}(\mathrm{Dom}(\varphi))$. Από το θεώρημα Rademacher (βλέπε [27]) έπεται ότι η κλίση της φ, $\nabla\varphi$, είναι καλά ορισμένη σχεδόν παντού και τοπικά φραγμένη. Επιπλέον, το σύνολο των σημείων όπου η $\nabla\varphi$ δεν υπάρχει είναι μικρό σύνολο (για μια σύντομη απόδειξη, βλέπε [2]).

Συνήθως, μπορεί κανείς να τροποποιήσει με διάφορους τρόπους τις τιμές της φ στο $\partial(\mathrm{Dom}(\varphi))$ χωρίς να πειραχθεί η κυρτότητα της φ. Όταν όμως η φ υποτεθεί κάτω ημισυνεχής, τότε οι τιμές της στο σύνορο καθορίζονται πλήρως. Δηλαδή, αν φ, ψ είναι δύο κυρτές, κάτω ημισυνεχείς συναρτήσεις με $\mathrm{Int}(\mathrm{Dom}(\varphi)) = \mathrm{Int}(\mathrm{Dom}(\psi))$ και οι τιμές τους συμπίπτουν στο σύνολο αυτό, τότε είναι ίσες.

Σε κάθε σημείο x στο οποίο η φ είναι διαφορίσιμη ισχύει η ανισότητα

$$\varphi(z) \geqslant \varphi(x) + \langle \nabla\varphi(x), z - x\rangle$$

για κάθε $z \in \mathbb{R}^n$. Η ανισότητα αυτή, γεωμετρικά εκφράζει το γεγονός πως το γράφημα της φ βρίσκεται πάνω από το εφαπτόμενο υπερεπίπεδό του στο σημείο x. Ειδικότερα, η $\nabla\varphi$ είναι μονότονη, δηλαδή αν x, y είναι σημεία στα οποία η φ είναι διαφορίσιμη, τότε

$$\langle \nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \geqslant 0.$$

Ορισμός 2.3.2. Έστω φ μια κυρτή συνάρτηση στον $\mathbb{R}^n$. Το διάνυσμα y λέγεται *υποκλίση* της φ στο x αν

$$\varphi(z) \geqslant \varphi(x) + \langle y, z - x\rangle$$

για κάθε $z \in \mathbb{R}^n$. Το σύνολο όλων των υποκλίσεων της φ στο x λέγεται *υποδιαφορικό* της φ στο x και συμβολίζεται με $\partial\varphi(x)$. Συνεπώς, έχουμε τον εξής περιγραφικό ορισμό του υποδιαφορικού:

$$y \in \partial\varphi(x) \iff \varphi(z) \geqslant \varphi(x) + \langle y, z - x\rangle \text{ για κάθε } z \in \mathbb{R}^n. \tag{2.3.1}$$

Το $\partial\varphi(x)$ είναι κλειστό και κυρτό σύνολο. Αν $\partial\varphi(x) \neq \emptyset$ λέμε ότι η φ είναι υποδιαφορίσιμη στο x, ενώ αν $\partial\varphi(x) = \{\nabla\varphi(x)\}$ λέμε ότι η φ είναι διαφορίσιμη στο x. Αν η φ είναι κάτω ημισυνεχής τότε το υποδιαφορικό της, $\partial\varphi$, είναι συνεχές στον $\mathbb{R}^n$. Δηλαδή, αν $(x_k), (y_k)$ είναι δύο ακολουθίες με $x_k \to x, y_k \to y$ και $y_k \in \partial\varphi(x_k)$, τότε $y \in \partial\varphi(x)$.

Μια άμεση συνέπεια του ορισμού του υποδιαφορικού μιας κυρτής συνάρτησης φ είναι ότι είναι μονότονη απεικόνιση, δηλαδή για κάθε $y_1 \in \partial\varphi(x_1)$ και $y_2 \in \partial\varphi(x_2)$ έχουμε

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geqslant 0.$$

Έστω, τώρα, μια κάτω ημισυνεχής, γνήσια κυρτή συνάρτηση, έστω x ένα σημείο στο οποίο η φ είναι διαφορίσιμη και έστω $y = \nabla\varphi(x)$. Αποδεικνύεται ότι η $\nabla\varphi$ είναι συνεχής απεικόνιση, δηλαδή για κάθε $\varepsilon > 0$ υπάρχει $\delta > 0$ ώστε

$$\nabla\varphi(B_\delta(x)) \subset \partial\varphi(B_\delta(x)) \subset B_\varepsilon(y), \tag{2.3.2}$$

όπου $B_r(x_0)$ είναι η Ευκλείδεια μπάλα με κέντρο x_0 και ακτίνα r.

Εισάγουμε τώρα την έννοια της συζυγούς συνάρτησης.

Ορισμός 2.3.3. Έστω $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ μια *proper* κυρτή συνάρτηση, όχι ταυτοτικά $+\infty$. Ορίζουμε την κυρτή συζυγή συνάρτηση (ή μετασχηματισμό Legendre) της φ ως εξής:

$$\varphi^*(y) = \sup_{x \in \mathbb{R}^n} (\langle x, y \rangle - \varphi(x)). \tag{2.3.3}$$

Τότε, η φ^* είναι μια *proper* κάτω ημισυνεχής συνάρτηση, και από τον ορισμό της έπεται ότι

$$\langle x, y \rangle \leqslant \varphi(x) + \varphi^*(y) \ \text{ για κάθε } x, y \in \mathbb{R}^n. \tag{2.3.4}$$

Από τον ορισμό της συζυγούς συνάρτησης έπεται ότι αν φ_1, φ_2 είναι δύο *proper* κάτω ημισυνεχείς συναρτήσεις με $\varphi_1 \leqslant \varphi_2$, τότε $\varphi_1^* \geqslant \varphi_2^*$. Η επόμενη πρόταση δείχνει ότι το σύνολο των ζευγών (x, y) για τα οποία ισχύει ισότητα στην σχέση (2.3.3) περιγράφεται από το υποδιαφορικό της φ.

Πρόταση 2.3.4. *Έστω φ μια γνήσια, κυρτή, κάτω ημισυνεχής συνάρτηση στον $\mathbb{R}^n$. Τότε, για κάθε $x, y \in \mathbb{R}^n$ ισχύει ότι*

$$\langle x, y \rangle = \varphi(x) + \varphi^*(y) \iff y \in \partial\varphi(x) \iff x \in \partial\varphi^*(y).$$

Απόδειξη. Παρατηρούμε ότι

$$\begin{aligned}\langle x, y \rangle = \varphi(x) + \varphi^*(y) &\iff \langle x, y \rangle \geqslant \varphi(x) + \varphi^*(y) \\ &\iff \langle x, y \rangle \geqslant \varphi(x) + \langle y, z \rangle - \varphi(z) \ \text{ για κάθε } z \in \mathbb{R}^n \\ &\iff \varphi(z) \geqslant \varphi(x) + \langle y, z - x \rangle \ \text{ για κάθε } z \in \mathbb{R}^n \\ &\iff y \in \partial\varphi(x),\end{aligned}$$

όπου η πρώτη ισοδυναμία έπεται από την (2.3.4), η δεύτερη από την (2.3.3), και η τελευταία από την (2.3.1). Λόγω συμμετρίας, η ισοδυναμία με την $x \in \partial\varphi^*(y)$ είναι συνέπεια της Πρότασης 2.3.6. □

Ορισμός 2.3.5. Έστω φ, ψ δύο γνήσιες κυρτές συναρτήσεις. Ορίζουμε την ελαχιστική συνέλιξη των φ και ψ ως εξής:

$$(\varphi \Box \psi)(z) = \inf_{x+x'=z} [\varphi(x) + \psi(x')].$$

Ισχύει ότι

$$(\varphi \Box \psi)^* = \varphi^* + \psi^*.$$

Πρόταση 2.3.6 (δυϊσμός Legendre για κάτω ημισυνεχείς συναρτήσεις)**.** *Έστω* $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *μια γνήσια συνάρτηση. Τα ακόλουθα είναι ισοδύναμα:*

(i) *Η* φ *είναι κυρτή, κάτω ημισυνεχής.*

(ii) $\varphi = \psi^*$ *για κάποια γνήσια συνάρτηση* ψ*.*

(iii) $\varphi^{**} = \varphi$*.*

Απόδειξη. Είναι σαφές ότι αν ισχύει η (iii) τότε έπεται η (ii): αρκεί να θέσουμε $\psi = \varphi^*$. Από την (ii) πάλι έπεται η (i) διότι τότε η φ γράφεται ως το supremum γραμμικών συναρτήσεων. Μένει λοιπόν να δείξουμε ότι αν η φ είναι κυρτή, κάτω ημισυνεχής, τότε $\varphi^{**} = \varphi$. Θα χωρίσουμε την απόδειξη σε τρία βήματα:

Βήμα 1. Από την (2.3.4) έχουμε ότι

$$\varphi(x) \geqslant \sup_y [\langle x, y\rangle - \varphi^*(y)] = \varphi^{**}(x).$$

Βήμα 2. Έστω $x \in \mathrm{Int}(\mathrm{Dom}(\varphi))$. Αφού $\partial\varphi(x) \neq \emptyset$, μπορούμε να επιλέξουμε $y \in \partial\varphi(x)$. Από την Πρόταση 2.3.4 έχουμε $\varphi(x) + \varphi^*(y) = \langle x, y\rangle$, άρα,

$$\varphi(x) \leqslant \sup_y [\langle x, y\rangle - \varphi^*(y)] = \varphi^{**}(x).$$

Συνεπώς, οι φ, φ^{**} ταυτίζονται στο $\mathrm{Int}(\mathrm{Dom}(\varphi))$. Ειδικότερα, $\varphi = \varphi^{**}$ αν $Dom(\varphi) = \mathbb{R}^n$.

Βήμα 3. Για την γενική περίπτωση, όπου $x \in \mathbb{R}^n$ τυχόν, θα χρησιμοποιήσουμε την ελαχιστική συνέλιξη. Για κάθε $\varepsilon > 0$ θεωρούμε την $\psi_\varepsilon(x) = \frac{|x|^2}{2\varepsilon}$ και θέτουμε $\varphi_\varepsilon = \varphi \Box \psi_\varepsilon$. Χρησιμοποιώντας το γεγονός πως η φ είναι κάτω ημισυνεχής και φραγμένη από κάτω από μια αφφινική συνάρτηση, βλέπουμε ότι

$$\varphi(x) = \lim_{\varepsilon \to 0} \varphi_\varepsilon(x),$$

και αφού $\mathrm{Dom}(\varphi_\varepsilon) = \mathbb{R}^n$, από το Βήμα 2 γνωρίζουμε ότι $(\varphi_\varepsilon)^{**} = \varphi_\varepsilon$. Όμως, $\varphi_\varepsilon \leqslant \varphi$, άρα $(\varphi_\varepsilon)^* \geqslant \varphi^*$, και άρα $(\varphi_\varepsilon)^{**} \leqslant \varphi^{**}$. Συμπεραίνουμε έτσι ότι, για κάθε $x \in \mathbb{R}^n$,

$$\varphi^{**}(x) \geqslant \liminf_{\varepsilon \to 0} \varphi_\varepsilon(x) = \varphi(x).$$

Σε συνδυασμό με το Βήμα 1 παίρνουμε το ζητούμενο. $\square$

Ο δυϊσμός Legendre έχει σημαντικές συνέπειες στην διαφορισιμότητα. Αν η φ είναι αυστηρώς κυρτή στην περιοχή ενός $x \in \mathbb{R}^n$, τότε η φ^* είναι διαφορίσιμη στο $\partial\varphi(x)$ και $\nabla\varphi^*(y) = x$ για κάθε $y \in \partial\varphi(x)$. Αν η φ είναι διαφορίσιμη και αυστηρώς κυρτή, τότε το ίδιο ισχύει για την φ^* και η $\nabla\varphi$ είναι ένα προς ένα απεικόνιση. Παραγωγίζοντας την (2.3.4) ή χρησιμοποιώντας το τελευταίο μέρος της Πρότασης 2.3.4, παίρνουμε

$$(\nabla\varphi)^{-1} = \nabla\varphi^*. \tag{2.3.5}$$

Κεφάλαιο 3

Το θεώρημα δυϊσμού του Kantorovich

3.1 Το πρόβλημα βέλτιστης μεταφοράς

Δοθέντος ενός Πολωνικού χώρου (X, d) (δηλαδή, ενός πλήρους και διαχωρίσιμου μετρικού χώρου) συμβολίζουμε με $\mathcal{P}(X)$ το σύνολο των Borel μέτρων πιθανότητας στον X και με $\mathcal{M}(X)$ το σύνολο των πεπερασμένων, προσημασμένων μέτρων στον X, δηλαδή τον διανυσματικό χώρο που παράγεται από τον $\mathcal{P}(X)$ εφοδιασμένο με τη νόρμα της ολικής κύμανσης:

$$\|\mu\|_{\mathrm{TV}} = \inf\{\mu_+(X) + \mu_-(X)\}$$

όπου το infimum παίρνεται πάνω από όλα τα μη αρνητικά μέτρα μ_+, μ_- για τα οποία το μ γράφεται στην μορφή $\mu = \mu_+ - \mu_-$.

Πρόβλημα 3.1.1 (βέλτιστης μεταφοράς, του Kantorovich)**.** Έστω $(X, \mu), (Y, \nu)$ δύο χώροι μέτρου και έστω $c : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ μια μετρήσιμη συνάρτηση κόστους. Θέλουμε να ελαχιστοποιήσουμε το γραμμικό συναρτησοειδές

$$\pi \to \int_{X\times Y} c(x,y)d\pi(x,y)$$

στο σύνολο $\Pi(\mu, \nu)$ όλων των σχεδίων μεταφοράς $\pi \in \mathcal{P}(X \times Y)$ από το μ στο ν, δηλαδή στο σύνολο όλων των Borel μέτρων πιθανότητας π στον $X \times Y$ που έχουν περιθώρια μέτρα τα μ και ν. Δηλαδή, των μέτρων που ικανοποιούν τις

$$\pi(A \times Y) = \mu(A),\ \pi(X \times B) = \nu(B) \tag{3.1.1}$$

για όλα τα μετρήσιμα υποσύνολα A, B των X, Y αντίστοιχα.

Ουσιαστικά η ποσότητα $\pi(A \times B)$ δηλώνει πόση μάζα μεταφέρεται από το A στο B, συνεπώς τα σχέδια μεταφοράς μπορούμε να τα σκεφτόμαστε ως απεικονίσεις μεταφοράς οι οποίες επιδέχονται πολλαπλές τιμές για κάθε δεδομένο x, δηλαδή

$$\pi = \int \pi_x d\mu(x), \text{ όπου } \pi_x \in \mathcal{P}(\{x\} \times Y).$$

Στην ουσία, λοιπόν, το μέτρο π_x περιγράφει το πως κατανέμεται η μάζα που βρίσκεται στο $x \in X$ στον χώρο Y. Παρατηρούμε ότι η συνολική μάζα που παίρνουμε από την περιοχή $A \subset X$ είναι,

$$\int_A \pi_x(Y) d\mu(x) = \int_X \Big(\int_Y \mathbf{1}_{A\times Y}(x,y) d\pi_x(y) \Big) d\mu(x)$$

και ισούται με την μάζα $\mu(A)$ που παράχθηκε στην περιοχή A. Επιπλέον, για να είναι το π υλοποιήσιμο πρέπει η συνολική μάζα

$$\int_X \pi_x(B) d\mu(x) = \int_X \Big(\int_Y \mathbf{1}_{X\times B}(x,y) d\pi_x(y) \Big) d\mu(x)$$

που μεταφέρεται στο $B \subset Y$ να ισούται με την ζήτηση $\nu(B)$ του προϊόντος στην περιοχή B.

Παρατηρούμε ότι το σύνολο $\Pi(\mu,\nu)$ είναι μη κενό (για παράδειγμα, το μέτρο γινόμενο $\mu \otimes \nu$ ανήκει στο $\Pi(\mu,\nu)$) και κυρτό. Η συνθήκη (3.1.1) αναγκάζει το π να είναι μέτρο πιθανότητας. Επιπλέον, αν $\pi \in \Pi(\mu,\nu)$ τότε λόγω της (3.1.1) ισχύει η εξής ισοδυναμία:

> $\pi \in \Pi(\mu,\nu) \iff$ το π είναι μη αρνητικό μέτρο στον $X \times Y$ ώστε, για κάθε ζεύγος μετρήσιμων συναρτήσεων $(\varphi,\psi) \in L^1(d\mu) \times L^1(d\nu)$,

$$\int_{X\times Y} [\varphi(x) + \psi(y)] \, d\pi(x,y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y). \tag{3.1.2}$$

Στην συνέχεια διατυπώνουμε ένα λήμμα το οποίο θα παίξει καθοριστικό ρόλο στην απόδειξη της ύπαρξης βέλτιστου σχεδίου μεταφοράς για το πρόβλημα ελαχιστοποίησης του Kantorovich.

Λήμμα 3.1.2. *Έστω X, Y Πολωνικοί χώροι και έστω μ,ν Borel μέτρα πιθανότητας στους X, Y αντίστοιχα. Το σύνολο $\Pi(\mu,\nu)$ όλων των Borel μέτρων πιθανότητας π στον $X \times Y$ που έχουν περιθώρια μέτρα τα μ και ν, δηλαδή των μέτρων που ικανοποιούν τις*

$$\pi(A \times Y) = \mu(A), \ \pi(X \times B) = \nu(B) \tag{3.1.3}$$

για όλα τα μετρήσιμα υποσύνολα A, B των X, Y αντίστοιχα, είναι w^*-συμπαγές.

Απόδειξη. Παρατηρούμε ότι η οικογένεια $\Pi(\mu,\nu)$ είναι tight στον $\mathcal{P}(X\times Y)$. Πράγματι, από το λήμμα του Ulam έχουμε ότι τα Borel μέτρα πιθανότητας μ και ν είναι tight, συνεπώς για κάθε $\varepsilon>0$ υπάρχουν $K_1\subseteq X$ και $K_2\subseteq Y$ συμπαγή ώστε $\mu(X\setminus K_1)\leqslant \varepsilon/2$ και $\nu(Y\setminus K_2)\leqslant\varepsilon/2$. Τότε, από τον εγκλεισμό

$$(X\times Y)\setminus(K_1\times K_2)\subseteq[(X\setminus K_1)\times Y]\cup[X\times(Y\setminus K_2)],$$

για κάθε $\pi\in\Pi(\mu,\nu)$ ισχύει

$$\begin{aligned}\pi((X\times Y)\setminus(K_1\times K_2))&\leqslant \pi[(X\setminus K_1)\times Y]+\pi[X\times(Y\setminus K_2)]\\&=\mu(X\setminus K_1)+\nu(Y\setminus K_2)\leqslant\varepsilon.\end{aligned}$$

Άρα, η οικογένεια $\Pi(\mu,\nu)$ είναι tight και από το θεώρημα του Prokhorov είναι σχετικά ακολουθιακά συμπαγής με την w^*-τοπολογία.

Θα δείξουμε ότι η οικογένεια $\Pi(\mu,\nu)$ είναι και αθενώς κλειστή. Θεωρούμε ακολουθία $(\pi_n)\in\Pi(\mu,\nu)$ και υποθέτουμε ότι $\pi_n\to\pi$ με την w^*-τοπολογία. Θα δείξουμε ότι $\pi\in\Pi(\mu,\nu)$. Παρατηρούμε ότι για κάθε $\varphi\in C_b(X)$ η συνάρτηση $g(x,y)=\varphi(x)$ είναι συνεχής και φραγμένη στον $X\times Y$, συνεπώς έχουμε

$$\begin{aligned}\int_X\varphi(x)d(p^X\#\pi)(x)&=\int_{X\times Y}g(x,y)d\pi(x,y)=\lim_{n\to\infty}\int_{X\times Y}g(x,y)d\pi_n(x,y)\\&=\lim_{n\to\infty}\int_X\varphi(x)d(p^X\#\pi_n)(x)=\lim_{n\to\infty}\int_X\varphi(x)d\mu(x)\\&=\int_X\varphi(x)d\mu(x),\end{aligned}$$

και αφού η φ ήταν τυχούσα στον $C_b(X)$ έχουμε $p^X\#\pi=\mu$. Όμοια αποδεικνύεται ότι $p^Y\#\pi=\nu$, συνεπώς $\pi\in\Pi(\mu,\nu)$. □

3.2 Θεώρημα δυϊσμού του Kantorovich

Θεώρημα 3.2.1. *Έστω X,Y δύο Πολωνικοί χώροι (δηλαδή, πλήρεις και διαχωρίσιμοι μετρικοί χώροι), έστω $\mu\in\mathcal{P}(X)$ και $\nu\in\mathcal{P}(Y)$, και έστω $c:X\times Y\to\mathbb{R}_+\cup\{+\infty\}$ μια κάτω ημισυνεχής συνάρτηση κόστους. Για κάθε $\pi\in\mathcal{P}(X\times Y)$ και $(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu)$ ορίζουμε*

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

Ορίζουμε $\Pi(\mu,\nu)$ το σύνολο των Borel μέτρων πιθανότητας στον $X\times Y$ με περιθώρια μέτρα τα μ,ν, και Φ_c το σύνολο των ζευγών μετρήσιμων συναρτήσεων $(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu)$ που ικανοποιούν την σχέση

$$\varphi(x)+\psi(y)\leqslant c(x,y)$$

μ-σχεδόν για κάθε $x \in X$ και ν-σχεδόν για κάθε $y \in Y$. Τότε,

$$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi). \tag{3.2.1}$$

Επιπλέον, το infimum στο αριστερό μέλος πιάνεται από ένα μέτρο $\pi \in \Pi(\mu,\nu)$. Τέλος, η τιμή του supremum στο δεξιό μέλος δεν αλλάζει αν, για τον ορισμό του Φ_c, περιοριστούμε στα ζεύγη (φ,ψ) συνεχών και φραγμένων συναρτήσεων.

Αποδεικνύουμε πρώτα ένα μέρος του ισχυρισμού του Θεωρήματος 3.2.1 το οποίο είναι απλό και δείχνει τον λόγο για τον οποίο θα περίμενε κανείς να ισχύει η (3.2.1).

Πρόταση 3.2.2. *Κάτω από τις υποθέσεις του Θεωρήματος 3.2.1 έπεται ότι*

$$\sup_{\Phi_c\cap C_b} J(\varphi,\psi) \leqslant \sup_{\Phi_c\cap L^1} J(\varphi,\psi) \leqslant \inf_{\Pi(\mu,\nu)} I[\pi].$$

Απόδειξη. Η αριστερή ανισότητα είναι προφανής αν παρατηρήσουμε ότι $C_b(X)\times C_b(Y) \subset L^1(d\mu)\times L^1(d\nu)$. Μένει να δείξουμε την δεξιά ανισότητα. Έστω $(\varphi,\psi)\in\Phi_c\cap L^1$ και έστω $\pi\in\Pi(\mu,\nu)$. Τότε, από τον ορισμό του $\Pi(\mu,\nu)$ έχουμε

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

Από την υπόθεση του θεωρήματος έχουμε

$$\varphi(x)+\psi(y) \leqslant c(x,y) \tag{3.2.2}$$

μ-σχεδόν για κάθε $x\in X$ και ν-σχεδόν για κάθε $y\in Y$. Θα δείξουμε ότι η (3.2.2) ισχύει και π-σχεδόν παντού. Πράγματι, υπάρχουν N_x, N_y μετρήσιμα υποσύνολα των X, Y αντίστοιχα, ώστε $\mu(N_x)=0$, $\nu(N_y)=0$ και η (3.2.2) να ισχύει για κάθε $(x,y)\in N_x^c\times N_y^c$. Αφού το π έχει περιθώρια μέτρα μ και ν έχουμε $\pi(N_x\times Y)=\mu(N_x)=0$ και $\pi(X\times N_y)=\nu(N_y)=0$, άρα $\pi((N_x^c\times N_y^c)^c)=0$. Συνεπώς,

$$\int_{X\times Y} [\varphi(x)+\psi(y)]d\pi(x,y) \leqslant \int_{X\times Y} c(x,y)d\pi(x,y) = I[\pi]. \tag{3.2.3}$$

Το ζητούμενο έπεται αν πάρουμε supremum στο αριστερό μέλος της (3.2.3) και infimum στο δεξιό. □

Ιδέα της απόδειξης του θεωρήματος 3.2.1. Θα κάνουμε χρήση της αρχής min-max του γραμμικού προγραμματισμού. Ισχύει ότι

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \inf_{\pi\in M_+(X\times Y)} \left(I[\pi] + \begin{Bmatrix} 0, & \text{αν } \pi\in\Pi(\mu,\nu) \\ +\infty, & \text{αλλιώς} \end{Bmatrix} \right). \tag{3.2.4}$$

όπου $M_+(X \times Y)$ είναι το σύνολο των μη αρνητικών μέτρων Borel στον $X \times Y$. Τώρα, αφού οι περιορισμοί στον ορισμό της $\Pi(\mu,\nu)$ είναι γραμμικοί, μπορούμε να γράψουμε την χαρακτηριστική συνάρτηση της $\Pi(\mu,\nu)$ που εμφανίζεται μέσα στην αγκύλη ως το supremum γραμμικών συναρτησοειδών:

$$\left\{ \begin{array}{ll} 0, & \text{αν } \pi \in \Pi(\mu,\nu) \\ +\infty, & \text{αλλιώς} \end{array} \right\} = \sup_{(\varphi,\psi)} \left[\int \varphi d\mu + \int \psi d\nu - \int [\varphi(x) + \psi(y)] d\pi(x,y) \right],$$

όπου το supremum τρέχει πάνω από όλα τα ζεύγη $(\varphi,\psi) \in C_b(X) \times C_b(Y)$. Άρα έχουμε ότι

$$\begin{aligned} \inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \inf_{\pi \in M_+(X\times Y)} \sup_{(\varphi,\psi)} \Big\{ \int_{X\times Y} c(x,y) d\pi(x,y) + \int_X \varphi\mu + \int_Y \psi d\nu \\ - \int_{X\times Y} [\varphi(x) + \psi(y)] d\pi(x,y) \Big\} \end{aligned}$$

Θεωρώντας δεδομένο ότι μπορούμε να εφαρμόσουμε την αρχή min-max ξαναγράφουμε την προηγούμενη ισότητα ως εξής:

(3.2.5)

$$\begin{aligned} \inf_{\pi \in \Pi(\mu,\nu)} I[\pi] &= \sup_{(\varphi,\psi)} \inf_{\pi \in M_+(X\times Y)} \Big\{ \int_{X\times Y} c(x,y) d\pi(x,y) + \int_X \varphi d\mu + \int \psi d\nu \\ &\qquad - \int_{X\times Y} [\varphi(x) + \psi(y)] d\pi(x,y) \Big\} \\ &= \sup_{(\varphi,\psi)} \Big\{ \int_X \varphi d\mu + \int_Y \psi d\nu - \sup_{\pi \in M_+(X\times Y)} \int_{X\times Y} [\varphi(x) + \psi(y) - c(x,y)] d\pi(x,y) \Big\}. \end{aligned}$$

Ας υπολογίσουμε το supremum μέσα στην αγκύλη. Αν η συνάρτηση $\zeta(x,y) = \varphi(x) + \psi(y) - c(x,y)$ παίρνει θετική τιμή σε κάποιο σημείο (x_0,y_0) τότε επιλέγοντας $\pi = \lambda\delta_{(x_0,y_0)}$ και αφήνοντας το $\lambda \to \infty$, συμπεραίνουμε ότι το supremun είναι το άπειρο. Από την άλλη πλευρά, αν η ζ είναι μη θετική συνάρτηση $d\pi$-σχεδόν παντού, τότε το supremum πιάνεται όταν $\pi = 0$. Άρα,

$$\sup_{\pi \in M_+(X\times Y)} \int_{X\times Y} [\varphi(x) + \psi(y) - c(x,y)]\, d\pi(x,y) = \left\{ \begin{array}{ll} 0, & \text{αν } (\varphi,\psi) \in \Phi_c \\ +\infty, & \text{αλλιώς} \end{array} \right\}.$$

Συνδυάζοντας την τελευταία ισότητα με την σχέση (3.2.5) παίρνουμε το ζητούμενο, δηλαδή

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

Για την γενική απόδειξη του Θεωρήματος 3.2.1 θα χρειαστούμε την έννοια του μετασχηματισμού Legendre-Fenchel.

Ορισμός 3.2.3. Έστω E ένας διανυσματικός χώρος με νόρμα και Θ μια κυρτή συνάρτηση στον E με τιμές στο $\mathbb{R} \cup \{+\infty\}$. Ο μετασχηματισμός Legendre-Fenchel της Θ ορίζεται στον τοπολογικό δυϊκό E^* του E από την

$$\Theta^*(z^*) = \sup_{z \in E}[\langle z^*, z\rangle - \Theta(z)]$$

για κάθε $z^* \in E^*$.

Θεώρημα 3.2.4 (δυϊσμός των Fenchel-Rockafellar)**.** *Έστω E ένας χώρος με νόρμα, E^* ο τοπολογικός δυϊκός του, και Θ, Ξ δύο κυρτές συναρτήσεις στον E με τιμές στο $\mathbb{R}\cup\{+\infty\}$. Έστω Θ^*, Ξ^* οι μετασχηματισμοί Legendre-Fenchel των Θ, Ξ αντίστοιχα. Υποθέτουμε ότι υπάρχει $z_0 \in E$ τέτοιο ώστε να ισχύουν οι $\Theta(z_0) < +\infty$, $\Xi(z_0) < +\infty$ και η Θ να είναι συνεχής στο z_0. Τότε,*

$$\inf_E [\Theta + \Xi] = \max_{z^* \in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)].$$

Απόδειξη. Παρατηρούμε ότι

$$\Theta^*(-z^*) = \sup_{x \in E}[\langle -z^*, x\rangle - \Theta(x)] = -\inf_{z \in E}[\langle z^*, x\rangle + \Theta(x)]$$

και

$$\Xi(z^*) = \sup_{y \in E}[\langle z^*, y\rangle - \Xi(y)] = -\inf_{y \in E}[\Xi(y) - \langle z^*, y\rangle]$$

Άρα, αρκεί να δείξουμε ότι:

$$\sup_{z^* \in E^*} \inf_{x,y \in E} [\Theta(x) + \Xi(y) + \langle z^*, x - y\rangle] = \inf_{x \in E} [\Theta(x) + \Xi(x)].$$

(i) Η επιλογή $x = y$ δείχνει ότι το αριστερό μέλος της ανισότητας δεν είναι μεγαλύτερο από το δεξιό μέλος. Έτσι, αυτό που έχουμε να δείξουμε είναι ότι υπάρχει ένα γραμμικό συναρτησοειδές $z^* \in E^*$ τέτοιο ώστε για κάθε $x, y \in E$ να ισχύει

$$\Theta(x) + \Xi(y) + \langle z^*, x - y\rangle \geqslant m \equiv \inf\,(\Theta + \Xi).$$

Επειδή $\Theta(z_0) + \Xi(z_0) < +\infty$, έπεται ότι το infimum m στο δεξιό μέλος είναι πεπερασμένο.
(ii) Θεωρούμε $x, y \in E$ και θέτουμε

$$C = \{(x, \lambda) \in E \times \mathbb{R} :\ \lambda > \Theta(x)\} \text{ και } C' = \{(y, \mu) \in E \times \mathbb{R} :\ \mu \leqslant m - \Xi(y)\}.$$

Είναι εύκολο να δούμε ότι αφού οι Θ, Ξ είναι κυρτές τότε και τα C, C' είναι κυρτά σύνολα. Λόγω της συνέχειας της Θ στο z_0 έπεται ότι το $(z_0, \Theta(z_0) + 1) \in \mathrm{int}(C)$. Πράγματι, αρκεί να δείξουμε ότι υπάρχει $\delta > 0$ ώστε αν $\|z - z_0\| + |\lambda - (\Theta(z_0) + 1)| < \frac{\delta}{2}$ τότε $\lambda > \Theta(z)$.

Από την συνέχεια της Θ στο z_0 μπορούμε να επιλέξουμε $0 < \delta < 1$ ώστε: αν $\|z - z_0\| < \frac{\delta}{2}$ τότε $|\Theta(z) - \Theta(z_0)| < \frac{1}{2}$. Για κάθε τέτοιο z έχουμε

$$\lambda > \Theta(z_0) + 1 - \frac{\delta}{2} = (\Theta(z_0) - \Theta(z)) + 1 - \frac{\delta}{2} + \Theta(z),$$

απ' όπου παίρνουμε

$$\begin{aligned}\frac{\delta}{2} &> 1 - \lambda + \Theta(z) + (\Theta(z_0) - \Theta(z)) \geqslant 1 - \lambda + \Theta(z) - |\Theta(z_0) - \Theta(z)| \\ &> 1 - \lambda + \Theta(z) - \frac{1}{2}\end{aligned}$$

και συνεπώς,

$$\Theta(z) < \frac{\delta}{2} - \frac{1}{2} + \lambda = \frac{\delta - 1}{2} + \lambda.$$

Έπεται ότι

$$\lambda > \Theta(z) + \frac{1-\delta}{2} \text{ δηλαδή } \lambda > \Theta(z).$$

Άρα, $\mathrm{int}(C) \neq \emptyset$. Από την κυρτότητα του C έπεται εύκολα ότι $\overline{C} = \overline{\mathrm{int}(C)}$. Ακόμα, $C \cap C' = \emptyset$, επειδή $m = \inf[\Theta + \Xi]$. Από το διαχωριστικό θεώρημα Hahn-Banach έπεται ότι υπάρχει ένα γραμμικό συναρτησοειδές $\ell \in (E \times \mathbb{R})^*$ για το οποίο ισχύει:

$$\inf_{c \in C} \ell(c) = \inf_{c \in \mathrm{int}(C)} \ell(c) \geqslant \sup_{c' \in C'} \ell(c').$$

Με άλλα λόγια, υπάρχουν $w^* \in E^*$ και $\alpha \in \mathbb{R}$ με $(w^*, \alpha) \neq (0,0)$, ώστε

$$\langle w^*, x\rangle + \alpha\lambda \geqslant \langle w^*, y\rangle + \alpha\mu$$

αν υποθέσουμε ότι $\lambda > \Theta(x)$ και $\mu \leqslant m - \Xi(y)$. Εύκολα ελέγχουμε ότι $\alpha > 0$. Πράγματι: αν υποθέσουμε ότι $\alpha \leqslant 0$ και θέσουμε $x = y = z_0$, τότε για κάθε $\lambda > \Theta(z_0)$ και $\mu \leqslant m - \Xi(z_0)$ έχουμε

$$\lambda \leqslant \mu \leqslant m - \Xi(z_0) \leqslant \Theta(z_0),$$

το οποίο είναι άτοπο. Άρα, ορίζοντας $z^* = w^*/\alpha$ έχουμε

$$\langle z^*, x\rangle + \lambda \geqslant \langle z^*, y\rangle + \mu$$

για κάθε (x, λ) με $\lambda > \Theta(x)$ και (y, μ) με $\mu \leqslant m - \Xi(y)$. Αυτό σημαίνει ότι για κάθε $x, y \in E$ έχουμε ότι

$$\langle z^*, x\rangle + \Theta(x) \geqslant \langle z^*, y\rangle + m - \Xi(y),$$

από όπου έπεται το ζητούμενο. □

Μπορούμε τώρα να δώσουμε την γενική απόδειξη του θεωρήματος του Kantorovich.

Απόδειξη του θεωρήματος 3.2.1. Η απόδειξη θα γίνει σε τρία βήματα. Σε κάθε βήμα εξασφαλίζουμε το συμπέρασμα χαλαρώνοντας τις υποθέσεις μας.

Βήμα 1. Αρχικά υποθέτουμε ότι οι χώροι X, Y είναι συμπαγείς και η συνάρτηση κόστους είναι συνεχής στον $X \times Y$, άρα και φραγμένη. Θέτουμε $E = C_b(X \times Y)$ το σύνολο των (φραγμένων) συνεχών συναρτήσεων στον $X \times Y$ εφοδιασμένο με την $\|\cdot\|_\infty$ νόρμα. Από το Θεώρημα του Riesz ο τοπολογικός δυϊκός του μπορεί να ταυτιστεί με τον χώρο των (κανονικών) μέτρων τον οποίο συμβολίζουμε με $\mathcal{M}(X \times Y)$ και τον θεωρούμε εφοδιασμένο με την νόρμα της ολικής κύμανσης. Θέτουμε

$$\Theta : u \in C_b(X \times Y) \longmapsto \begin{cases} 0, & \text{αν } u(x,y) \geqslant -c(x,y) \\ +\infty, & \text{αλλιώς} \end{cases}$$

και

$$\Xi : u \in C_b(X \times Y) \longmapsto \begin{cases} \int_X \varphi d\mu + \int_Y \psi d\nu, & \text{αν } u(x,y) = \varphi(x) + \psi(y) \\ +\infty, & \text{αλλιώς} \end{cases}$$

Παρατηρούμε ότι η Ξ είναι καλά ορισμένη: αν $\varphi(x)+\psi(y) = \tilde{\varphi}(x)+\tilde{\psi}(y)$ για κάθε x, y τότε $\varphi = \tilde{\varphi}+s$, $\psi = \tilde{\psi}-s$ για κάποιον $s \in \mathbb{R}$. Οπότε, θα είναι $\int \tilde{\varphi} d\mu + \int \tilde{\psi} d\nu = \int \varphi d\mu + \int \psi d\nu$. Οι υποθέσεις του Θεωρήματος 3.2.4 ικανοποιούνται αν θεωρήσουμε την σταθερή συνάρτηση $u_0 \equiv M$ για κάποια σταθερά M. Τώρα, θα υπολογίσουμε τα δύο μέλη της ισότητας. Το αριστερό μέλος είναι:

$$\begin{aligned} \inf_E(\Theta(u) + \Xi(u)) &= \inf\left\{\int_X \varphi\, d\mu + \int_Y \psi\, d\nu, \quad \varphi(x) + \psi(y) \geqslant -c(x,y)\right\} \\ &= -\sup\{J(\varphi,\psi), \quad (\varphi,\psi) \in \Phi_c\}. \end{aligned}$$

Υπολογίζουμε τους μετασχηματισμούς Legendre–Fenchel των Θ, Ξ. Κατ' αρχήν, για κάθε $\pi \in \mathcal{M}(X \times Y)$ έχουμε

$$\begin{aligned} \Theta^*(-\pi) &= \sup_{u \in C_b(X\times Y)} \left\{-\int_{X\times Y} u(x,y) d\pi(x,y), \quad u(x,y) \geqslant -c(x,y)\right\} \\ &= \sup_{u \in C_b(X\times Y)} \left\{\int_{X\times Y} u(x,y) d\pi(x,y), \quad u(x,y) \leqslant c(x,y)\right\}. \end{aligned}$$

Εξετάζουμε δύο περιπτώσεις:

- Αν το π δεν είναι μη αρνητικό μέτρο, τότε υπάρχει μη θετική συνάρτηση $v \in C_b(X \times Y)$ τέτοια ώστε $\int v\, d\pi > 0$. Αν επιλέξουμε $u = \lambda v$ και αφήσουμε το $\lambda \to +\infty$, βλέπουμε ότι το supremum είναι $+\infty$.

- Αν το π είναι μη αρνητικό μέτρο, τότε το supremum είναι ίσο με $\int_{X\times Y} c\, d\pi$.

Έτσι, έχουμε

$$\Theta^*(-\pi) = \begin{cases} \int_{X\times Y} c\, d\pi, & \text{αν } \pi \in M_+(X\times Y) \\ +\infty, & \text{αλλιώς.} \end{cases}$$

Ομοίως βλέπουμε ότι

$$\Xi^*(\pi) = \begin{cases} 0, & \text{αν για κάθε } (\varphi,\psi) \in C_b(X)\times C_b(Y) \text{ ισχύει} \\ & \qquad \int_{X\times Y}[\varphi(x)+\psi(y)]\, d\pi(x,y) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu, \\ +\infty, & \text{αλλιώς.} \end{cases}$$

Με άλλα λόγια οι Θ^*, Ξ^* είναι οι χαρακτηριστικές συναρτήσεις των $M_+(X\times Y)$ και $\Pi(\mu,\nu)$, αντίστοιχα. Από τα παραπάνω, βλέπουμε αμέσως ότι η

$$\inf_{C_b(X\times Y)} [\Theta(u) + \Xi(u)] = \max_{\pi\in\mathcal{M}(X\times Y)} [-\Theta^*(-\pi) - \Xi^*(\pi)]$$

παίρνει την μορφή

$$\sup\{J(\varphi,\psi) :\ (\varphi,\psi)\in\Phi_c\cap C_b\} = \inf\left\{\int_{X\times Y} c(x,y)\ d\pi(x,y) :\ \pi\in\Pi(\mu,\nu)\right\}$$

και έχουμε δείξει αυτό που θέλαμε.

Βήμα 2. Υποθέτουμε ότι οι χώροι είναι Πολωνικοί και εξακολουθούμε να υποθέτουμε ότι η c είναι φραγμένη και ομοιόμορφα συνεχής. Ορίζουμε

$$\|c\|_\infty = \sup_{X\times Y} c(x,y).$$

Από το Λήμμα 3.1.2 το $\Pi(\mu,\nu)$ είναι συμπαγές υποσύνολο του $\mathcal{M}(X\times Y)$, άρα υπάρχει μέτρο πιθανότητας $\pi_* \in \Pi(\mu,\nu)$ ώστε

$$I[\pi_*] = \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

Αφού η c είναι φραγμένη έπεται επίσης ότι το infimum είναι πεπερασμένο.

Θεωρούμε τυχόν $\delta > 0$. Εφόσον οι X, Y είναι Πολωνικοί χώροι έπεται ότι και ο $X\times Y$ είναι Πολωνικός. Όπως είδαμε στην απόδειξη του Λήμματος 3.1.2, το π_* είναι tight – ακριβέστερα, μπορούμε να βρούμε συμπαγή $X_0 \subset X$ και $Y_0 \subset Y$ τέτοια ώστε:

$$\pi_*\left((X\times Y)\setminus(X_0\times Y_0)\right) \leqslant 2\delta.$$

Ορίζουμε

$$\pi_{*0} = \frac{\mathbf{1}_{X_0\times Y_0}}{\pi_*(X_0\times Y_0)}\pi_*$$

και ελέγχουμε ότι το π_{*0} είναι μέτρο πιθανότητας στον $X_0 \times Y_0$. Γράφουμε μ_0, ν_0 για τα περιθώρια μέτρα του π_{*0} στους X_0, Y_0 αντίστοιχα. Επίσης ορίζουμε $\Pi_0(\mu_0, \nu_0)$ το σύνολο των μέτρων πιθανότητας στον $X_0 \times Y_0$ με περιθώρια μέτρα μ_0, ν_0, και I_0 την ποσότητα

$$I_0[\pi_0] = \int_{X_0 \times Y_0} c(x,y)\, d\pi_0(x,y).$$

Θεωρούμε ένα μέτρο $\tilde{\pi}_0 \in \Pi_0(\mu_0, \nu_0)$ για το οποίο ισχύει

$$I_0[\tilde{\pi}_0] = \inf_{\pi_0 \in \Pi_0(\mu_0, \nu_0)} I_0[\pi_0].$$

Επεκτείνουμε το $\tilde{\pi}_0$ σε ένα μέτρο $\tilde{\pi} \in \Pi(\mu, \nu)$ θέτοντας

$$\tilde{\pi} = \pi_*(X_0 \times Y_0)\tilde{\pi}_0 + \mathbf{1}_{(X_0 \times Y_0)^c}\pi_*$$

(δεν είναι δύσκολο να ελέγξουμε ότι, πράγματι, $\tilde{\pi} \in \Pi(\mu, \nu)$). Έτσι, έχουμε

$$\begin{aligned} I[\tilde{\pi}] &= \pi_*(X_0 \times Y_0) I_0[\tilde{\pi}_0] + \int_{(X_0 \times Y_0)^c} c(x,y)\ d\pi_*(x,y) \\ &\leqslant I_0[\tilde{\pi}_0] + 2\delta\|c\|_\infty = \inf I_0 + 2\delta\|c\|_\infty. \end{aligned}$$

Τώρα θεωρούμε το συναρτησοειδές

$$J_0(\varphi_0, \psi_0) = \int_{X_0} \varphi_0\ d\mu_0 + \int_{Y_0} \psi_0\ d\nu_0,$$

που ορίζεται στον $L^1(d\mu_0) \times L^1(d\nu_0)$. Από το Βήμα 1 της απόδειξης έπεται ότι $\inf I_0 = \sup J_0$, όπου το supremum τρέχει πάνω από τα ζεύγη $(\varphi_0, \psi_0) \in L^1(d\mu_0) \times L^1(d\nu_0)$, που ικανοποιούν την ανισότητα $\varphi_0(x) + \psi_0(y) \leqslant c(x,y)$ σχεδόν για κάθε x, y. Έτσι, υπάρχει ένα ζεύγος συναρτήσεων $(\tilde{\varphi}_0, \tilde{\psi}_0)$ τέτοιο ώστε

$$J_0(\tilde{\varphi}_0, \tilde{\psi}_0) \geqslant \sup J_0 - \delta.$$

Το πρόβλημά μας τώρα ανάγεται στο να κατασκευάσουμε ένα ζεύγος (φ, ψ) κατάλληλο για το πρόβλημα μεγιστοποίησης του $J(\varphi, \psi)$. Είναι χρήσιμο να εξασφαλίσουμε ότι η ανισότητα $\tilde{\varphi}_0(x) + \tilde{\psi}_0(y) \leqslant c(x,y)$ ισχύει για κάθε x, y. Αυτό γίνεται, γιατί μπορούμε να βρούμε σύνολα N_x, N_y ώστε $\tilde{\varphi}_0(x) + \tilde{\psi}_0(y) \leqslant c(x,y)$ για κάθε $(x,y) \in N_x^c \times N_y^c$ και μετά να δώσουμε στις $\tilde{\varphi}_0, \tilde{\psi}_0$ την τιμή $-\infty$ στα N_x, N_y αντίστοιχα.

Χωρίς βλάβη της γενικότητας, υποθέτουμε ότι $\delta \leqslant 1$. Αφού $J_0(0,0) \geqslant 0$, ξέρουμε ότι $\sup J_0 \geqslant 0$, οπότε $J_0(\tilde{\varphi}_0, \tilde{\psi}_0) \geqslant -1$. Γράφοντας

$$J_0(\tilde{\varphi}_0, \tilde{\psi}_0) = \int_{X \times Y} [\tilde{\varphi}_0(x) + \tilde{\psi}_0(y)]\, d\pi_0(x,y)$$

για κάποιο στοιχείο π_0 του $\Pi_0(\mu_0,\nu_0)$, συμπεραίνουμε ότι υπάρχει $(x_0,y_0)\in X_0\times Y_0$ τέτοιο ώστε

$$\tilde{\varphi}_0(x_0)+\tilde{\psi}_0(y_0)\geqslant -1.$$

Αν αντικαταστήσουμε τις $(\tilde{\varphi}_0,\tilde{\psi}_0)$ με τις $(\tilde{\varphi}_0+s,\tilde{\psi}_0-s)$, όπου s είναι τυχών πραγματικός αριθμός, παρατηρούμε ότι η τιμή $J_0(\tilde{\varphi}_0,\tilde{\psi}_0)$ δεν αλλάζει. Με κατάλληλη επιλογή του s μπορούμε να υποθέσουμε ότι

$$\tilde{\varphi}_0(x_0)\geqslant -\frac{1}{2}\quad \text{και}\quad \tilde{\psi}_0(y_0)\geqslant -\frac{1}{2}.$$

Αυτό σημαίνει ότι, για κάθε $(x,y)\in X_0\times Y_0$,

$$\tilde{\varphi}_0(x)\leqslant c(x,y_0)-\tilde{\psi}_0(y_0)\leqslant c(x,y_0)+\frac{1}{2}$$

και

$$\tilde{\psi}_0(y)\leqslant c(x_0,y)-\tilde{\varphi}_0(x_0)\leqslant c(x_0,y)+\frac{1}{2}.$$

Ορίζουμε μια συνάρτηση $\bar{\varphi}_0$ στον X ως εξής: για κάθε $x\in X$ θέτουμε

$$\bar{\varphi}_0(x)=\inf_{y\in Y_0}[c(x,y)-\tilde{\psi}_0(y)].$$

Από την ανισότητα $\tilde{\varphi}_0(x)\leqslant c(x,y)-\tilde{\psi}_0(y)$ έπεται ότι $\tilde{\varphi}_0\leqslant\bar{\varphi}_0$ στον X_0. Αυτό σημαίνει ότι $J_0(\bar{\varphi}_0,\tilde{\psi}_0)\geqslant J_0(\tilde{\varphi}_0,\tilde{\psi}_0)$. Επιπλέον, για κάθε $x\in X$ έχουμε άνω και κάτω φράγμα για την $\bar{\varphi}_0(x)$ συναρτήσει της c:

$$\bar{\varphi}_0(x)\geqslant \inf_{y\in Y_0}[c(x,y)-c(x_0,y)]-\frac{1}{2}$$

και

$$\bar{\varphi}_0(x)\leqslant c(x,y_0)-\tilde{\psi}_0(y_0)\leqslant c(x,y_0)+\frac{1}{2}.$$

Τέλος, για κάθε $y\in Y$ ορίζουμε

$$\bar{\psi}_0(y)=\inf_{x\in X}[c(x,y)-\bar{\varphi}_0(x)],$$

και παρατηρούμε ότι $(\bar{\varphi}_0,\bar{\psi}_0)\in\Phi_c$. Έτσι, έχουμε $J_0(\bar{\varphi}_0,\bar{\psi}_0)\geqslant J_0(\bar{\varphi}_0,\tilde{\psi}_0)\geqslant J_0(\tilde{\varphi}_0,\tilde{\psi}_0)$. Επιπλέον, για κάθε $y\in Y$ έχουμε

$$\bar{\psi}_0(y)\geqslant \inf_{x\in X}[c(x,y)-c(x,y_0)]-\frac{1}{2}$$

και

$$\bar{\psi}_0(y)\leqslant c(x_0,y)-\bar{\varphi}_0(x_0)\leqslant c(x_0,y)-\tilde{\varphi}_0(x_0)\leqslant c(x_0,y)+\frac{1}{2}.$$

Ειδικότερα,

$$\bar{\varphi}_0(x) \geqslant -\|c\|_\infty - \frac{1}{2}$$

και

$$\bar{\psi}_0(y) \geqslant -\|c\|_\infty - \frac{1}{2}.$$

Τώρα, μπορούμε να γράψουμε

$$\begin{aligned}
J(\bar{\varphi}_0, \bar{\psi}_0) &= \int_X \bar{\varphi}_0 \, d\mu + \int_Y \bar{\psi}_0 \, d\nu = \int_{X\times Y} [\bar{\varphi}_0(x) + \bar{\psi}_0(y)] \, d\pi_*(x,y) \\
&= \pi_*(X_0 \times Y_0) \int_{X_0\times Y_0} [\bar{\varphi}_0(x) + \bar{\psi}_0(y)] \, d\pi_{*0}(x,y) \\
&\qquad + \int_{(X_0\times Y_0)^c} [\bar{\varphi}_0(x) + \bar{\psi}_0(y)] \, d\pi_*(x,y) \\
&\geqslant (1-2\delta) \left(\int_{X_0} \bar{\varphi}_0 \, d\mu_0 + \int_{Y_0} \bar{\psi}_0 \, d\nu_0 \right) - (2\|c\|_\infty + 1)\pi_* \left((X_0 \times Y_0)^c \right) \\
&\geqslant (1-2\delta) J_0(\bar{\varphi}_0, \bar{\psi}_0) - 2(2\|c\|_\infty + 1)\delta \\
&\geqslant (1-2\delta) J_0(\tilde{\varphi}_0, \tilde{\psi}_0) - 2(2\|c\|_\infty + 1)\delta \\
&\geqslant (1-2\delta)(\inf I_0 - \delta) - 2(2\|c\|_\infty + 1)\delta \\
&\geqslant (1-2\delta) \left(\inf I - (2\|c\|_\infty + 1)\delta \right) - 2(2\|c\|_\infty + 1)\delta.
\end{aligned}$$

Αφού το $\delta > 0$ ήταν τυχόν, συμπεραίνουμε ότι $\sup J(\varphi, \psi) \geqslant \inf I$, άρα ισχύει η ισότητα.

Σημείωση. Οι συναρτήσεις $\bar{\varphi}_0, \bar{\psi}_0$ είναι ομοιόμορφα συνεχείς (άρα και μετρήσιμες) στους X, Y αντίστοιχα, διότι η c είναι ομοιόμορφα συνεχής. Συνεπώς, δεν έχει σημασία αν θα θεωρήσουμε το supremum της J στο $\Phi_c \cap L^1$ ή στο $\Phi_c \cap C_b$.

Βήμα 3. Εξετάζουμε τώρα τη γενική περίπτωση, όπου η c είναι κάτω ημισυνεχής. Όπως έχουμε ήδη αναφέρει, η c γράφεται στη μορφή $c = \sup_n c_n$, όπου (c_n) είναι μια αύξουσα ακολουθία ομοιόμορφα συνεχών συναρτήσεων κόστους. Επίσης, επειδή μπορούμε να αντικαταστήσουμε την c_n από την $\min(c_n, n)$, μπορούμε να υποθέσουμε ότι όλες οι c_n είναι φραγμένες.

Ορίζουμε

$$I_n[\pi] = \int_{X\times Y} c_n \, d\pi \quad \text{για κάθε } \pi \in \Pi(\mu, \nu).$$

Από το Βήμα 2 έχουμε ότι

$$\inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi] = \sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi, \psi). \tag{3.2.6}$$

Τώρα, αν δείξουμε ότι για κάθε n ισχύει

$$\sup_{(\varphi,\psi)\in\Phi_{c_n}} J(\varphi,\psi) \leqslant \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi), \tag{3.2.7}$$

και ότι

$$\sup_n \inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi] = \inf_{\pi\in\Pi(\mu,\nu)} I[\pi] \tag{3.2.8}$$

έχουμε τελειώσει. Πράγματι, συνδυάζοντας τις (3.2.8), (3.2.6) και (3.2.7) παίρνουμε

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

και αφού η αντίστροφη ανισότητα ισχύει πάντα από την Πρόταση 3.2.2, το ζητούμενο έπεται.

Η (3.2.7) είναι προφανής: αφού $c_n \leq c$, το Φ_{c_n} είναι υποσύνολο του Φ_c στο οποίο οι J_n και J συμπίπτουν.

Θέτουμε $A_n := \inf_{\pi\in\Pi(\mu,\nu)} I_n[\pi]$. Παρατηρούμε ότι η ακολουθία (I_n) είναι αύξουσα, αφού η (c_n) είναι αύξουσα. Άρα και η (A_n) είναι αύξουσα. Επομένως, το μόνο που έχουμε να δείξουμε είναι ότι

$$\lim_{n\to\infty} A_n \geqslant \inf_{\pi\in\Pi(\mu,\nu)} I[\pi]. \tag{3.2.9}$$

Από το Λήμμα 3.1.2 γνωρίζουμε ότι το $\Pi(\mu,\nu)$ είναι w^*-συμπαγές. Συνεπώς, αν $(\pi_n^k)_k$ είναι μια ακολουθία που ικανοποιεί την $I[\pi_n^k] \to A_n$, μπορούμε να βρούμε υπακολουθία $(\pi_n^{s_k}) \in \Pi(\mu,\nu)$ και ένα μέτρο πιθανότητας $\pi_n \in \Pi(\mu,\nu)$ ώστε $\pi_n^{s_k} \xrightarrow{w} \pi_n$. Αυτό σημαίνει ότι για κάθε συνεχή συνάρτηση θ στον $X\times Y$,

$$\int \theta(x,y)\, d\pi_n^{s_k}(x,y) \longrightarrow \int \theta(x,y)\, d\pi_n(x,y)$$

καθώς $k\to\infty$. Αφού $\pi_n \in \Pi(\mu,\nu)$, συμπεραίνουμε ότι

$$\inf I_n[\pi] = A_n = \lim_{k\to\infty} \int c_n\, d\pi_n^{s_k} = \int c_n\, d\pi_n,$$

δηλαδή για κάθε n υπάρχει $\pi_n \in \Pi(\mu,\nu)$ ώστε $\inf I_n = I[\pi_n]$.

Όμοια, η ακολουθία $(\pi_n)_{n\in\mathbb{N}}$ έχει ένα σημείο συσσώρευσης, έστω π_*, από τη συμπάγεια του $\Pi(\mu,\nu)$. Δηλαδή, υπάρχει υπακολουθία (π_{k_n}) ώστε $\pi_{k_n} \xrightarrow{w^*} \pi_*$. Παρατηρούμε ότι για κάθε $n \geqslant m$ ισχύει ότι $I_{k_n}[\pi_{k_n}] \geqslant I_{k_m}[\pi_{k_n}]$. Από τη συνέχεια του I_{k_m} έπεται ότι

$$\lim_{n\to\infty} I_{k_n}[\pi_{k_n}] \geqslant \limsup_{n\to\infty} I_{k_m}[\pi_{k_n}] = I_{k_m}[\pi_*].$$

Από το θεώρημα μονότονης σύγκλισης έπεται ότι $I_{k_m}[\pi_*] \to I[\pi_*]$ καθώς $m \to \infty$, οπότε

$$\lim_{n\to\infty} I_{k_n}[\pi_{k_n}] \geqslant \lim_{m\to\infty} I_{k_m}[\pi_*] = I[\pi_*] \geqslant \inf_{\pi\in\Pi(\mu,\nu)} I[\pi],$$

και έχουμε δείξει αυτό που θέλαμε.

Τέλος το infimum πιάνεται από ένα μέτρο $\pi_* \in \Pi(\mu,\nu)$ λόγω της συμπάγειας του $\Pi(\mu,\nu)$. Πράγματι, αν $(\pi_k)_k$ είναι μια ακολουθία στο $\Pi(\mu,\nu)$ ώστε $I[\pi_k] \to \inf_{\pi\in\Pi(\mu,\nu)} I[\pi]$, τότε μπορούμε να βρούμε υπακολουθία $(\pi_{s_k})_k$ της $(\pi_k)_k$ και μέτρο πιθανότητας π_* ώστε $\pi_{s_k} \xrightarrow{w^*} \pi_*$. Τότε,

$$I[\pi_*] = \lim_{n\to\infty} I_n[\pi_*] \leqslant \lim_{n\to\infty} \limsup_{k\to\infty} I_n[\pi_{s_k}] \leqslant \limsup_{k\to\infty} I[\pi_{s_k}] = \inf I,$$

όπου η πρώτη ισότητα και η τρίτη ανισότητα είναι συνέπειες του θεωρήματος μονότονης σύγκλισης. □

Παρατήρηση 3.2.5 (c-κοίλες συναρτήσεις)**.** Από την απόδειξη του θεωρήματος του Kantorovich έπεται ότι όταν η c είναι φραγμένη τότε το supremum στο δεξιό μέλος της ισότητας

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

μπορεί να θεωρηθεί πάνω από τα ζεύγη της μορφής (φ^{cc},φ^c), όπου η φ είναι φραγμένη και

$$\varphi^c(y) = \inf_{x\in X}[c(x,y)-\varphi(x)],\ \varphi^{cc}(x) = \inf_{y\in Y}[c(x,y)-\varphi^c(y)]. \tag{3.2.10}$$

Το ζεύγος (φ^{cc},φ^c) λέγεται ζεύγος **συζυγών c-κοίλων** συναρτήσεων. Παρατηρήστε ότι οι φ^c,φ^{cc} είναι μετρήσιμες συναρτήσεις. Πράγματι, η φ^c μπορεί να γραφεί ως $\lim_{\ell\to+\infty}\psi_\ell$, όπου

$$\psi_\ell(y) = \inf_{x\in X}[c_\ell(x,y)-\varphi(x)],$$

και (c_ℓ) είναι μια αύξουσα οικογένεια φραγμένων και ομοιόμορφα συνεχών συναρτήσεων που συγκλίνουν κατά σημείο στην c. Συνεπώς, κάθε ψ_ℓ είναι ομοιόμορφα συνεχής και άρα η φ^c είναι μετρήσιμη. Όμοια, η φ^{cc} είναι μετρήσιμη. Είναι σαφές ότι με την αντικατάσταση του ζεύγους (φ,ψ) από το (φ^{cc},φ^c) παίρνουμε καλύτερο αποτέλεσμα: δηλαδή $J(\varphi,\psi) \leqslant J(\varphi^{cc},\varphi^c)$.

Πράγματι, έχοντας υπόψιν ότι $\varphi(x)+\psi(y) \leqslant c(x,y)$ $\pi-$σχεδόν παντού, βλέπουμε ότι $\psi(y) \leqslant \varphi^c(y)$. Συνεπώς η φ^c είναι η μέγιστη συνάρτηση $\upsilon : Y \to [-\infty,+\infty]$ τέτοια ώστε το ζεύγος (φ,υ) να είναι c-αποδεκτό, δηλαδή να ισχύει $\varphi(x)+\upsilon(y) \leqslant c(x,y)$. Άρα ισχύει $J(\varphi,\psi) \leqslant J(\varphi,\varphi^c)$. Με παρόμοιο επιχείρημα έχουμε ότι $\varphi(x) \leqslant \inf[c(x,y)-\varphi^c(y) = \varphi^{cc}(x)$, συνεπώς ισχύει ότι $J(\varphi,\varphi^c) \leqslant J(\varphi^{cc},\varphi^c)$. Καταλήγουμε, λοιπόν, στο συμπέρασμα ότι το $sup_{\Phi_c} J(\varphi,\psi)$ μπορεί να περιοριστεί στα ζέυγη των συζυγών c-κοίλων συναρτήσεων της μορφής (φ^{cc},φ^c).

Παρατήρηση 3.2.6 (υπολογισμοί για φραγμένες συναρτήσεις κόστους)**.** Αν $(\varphi^{cc}, \varphi^c)$ είναι ένα ζεύγος συζυγών c-κοίλων συναρτήσεων τότε

$$\begin{cases} -\sup\varphi \leqslant \varphi^c \leqslant \|c\|_\infty - \sup\varphi \\ -\sup\varphi^c \leqslant \varphi = \varphi^{cc} \leqslant \|c\|_\infty - \sup\varphi^c \end{cases} \tag{3.2.11}$$

Επιπλέον, αφού $J(\varphi+s, \psi-s) = J(\varphi,\psi)$ για κάθε $s \in \mathbb{R}$ και $(\varphi+s)^c = \varphi^c - s$, μπορούμε να υποθέσουμε χωρίς βλάβη της γενικότητας ότι $\sup\varphi = \|c\|_\infty$. Τότε, από την (3.2.11) έπεται ότι $-\|c\|_\infty \leqslant \varphi^c \leqslant 0$ και αυτό δείχνει με την σειρά του ότι $\inf\varphi \geqslant 0$. Άρα, όταν η c είναι φραγμένη, το supremun μπορεί να περιοριστεί περαιτέρω ως εξής:

$$\begin{aligned} &\sup\{J(\varphi,\psi) : (\varphi,\psi) \in \Phi_c\} \\ &\qquad = \sup\{J(\varphi,\psi) : (\varphi,\psi) \in \Phi_c, 0 \leqslant \varphi \leqslant \|c\|_\infty, -\|c\|_\infty \leqslant \psi \leqslant 0\}. \end{aligned}$$

3.3 Θεώρημα Kantorovich-Rubinstein

Σε αυτήν την παράγραφο υποθέτουμε ότι $X = Y$ και ότι η συνάρτηση κόστους είναι μια μετρική $c(x,y) = d(x,y)$ στον X, χωρίς απαραίτητα αυτή η απόσταση να είναι η απόσταση που ορίζει την τοπολογία του χώρου.

Θεώρημα 3.3.1 (Kantorovich-Rubinstein)**.** *Έστω X ένας Πολωνικός χώρος και έστω d μια κάτω ημισυνεχής μετρική στον X. Έστω $\mathcal{T}_d$ το κόστος της βέλτιστης μεταφοράς για την συνάρτηση κόστους $c(x,y) = d(x,y)$. Δηλαδή,*

$$\mathcal{T}_d(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x,y)\, d\pi(x,y).$$

Συμβολίζουμε με $\mathrm{Lip}(X)$ τον χώρο των Lipschitz συναρτήσεων στον X, όπου

$$\|\varphi\|_{\mathrm{Lip}} \equiv \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)}.$$

Τότε,

$$\mathcal{T}_d(\mu,\nu) = \sup\left\{ \int_X \varphi\, d(\mu-\nu) : \varphi \in \cap L^1(d|\mu-\nu|) \text{ και } \|\varphi\|_{\mathrm{Lip}} \leqslant 1 \right\}.$$

Επιπλέον η τιμή του supremum δεν αλλάζει αν περιοριστούμε σε φραγμένες Lipschitz συναρτήσεις φ.

Απόδειξη. Για κάθε $n \geqslant 1$ ορίζουμε $d_n = \frac{d}{1+n^{-1}d}$. Αυτή είναι μια συνάρτηση απόστασης που ικανοποιεί την $d_n \leqslant d$ και για κάθε $x, y \in X$ η ακολουθία $d_n(x,y)$ συγκλίνει μονότονα στην $d(x,y)$ όταν $n \to \infty$. Ειδικότερα, το σύνολο των 1-Lipschitz συναρτήσεων ως προς την d_n περιέχεται στο σύνολο των 1-Lipschitz συναρτήσεων ως προς την d, αφού

$$\sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)} \leqslant \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_n(x,y)} \leqslant 1.$$

Με βάση τον συλλογισμό του Βήματος 3 στην γενική απόδειξη του θεωρήματος του Kantorovich, αρκεί να αποδείξουμε το θεώρημα για την d_n στην θέση της d. Από εδώ και στο εξής μπορούμε λοιπόν να υποθέσουμε ότι η d είναι φραγμένη. Συνεπώς, όλες οι Lipschitz συναρτήσεις είναι φραγμένες και άρα ολοκληρώσιμες ως προς μ, ν. Με βάση λοιπόν το θεώρημα δυισμού του Kantorovich αρκεί να ελέγξουμε ότι

$$\sup_{(\varphi,\psi)\in\Phi_d} J(\varphi,\psi) = \sup\left\{ \int_X \varphi d(\mu-\nu) : \|\varphi\|_{\mathrm{Lip}} \leqslant 1 \right\},$$

όπου $J(\varphi,\psi) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$.

Από την Παρατήρηση 3.2.5 έχουμε ότι

$$\sup_{(\varphi,\psi)\in\Phi_d} J(\varphi,\psi) = \sup_{\varphi\in L^1(d\mu)} J(\varphi^{dd},\varphi^d),$$

όπου

$$\varphi^d(y) \equiv \inf_{x\in X}[d(x,y) - \varphi(x)],\ \varphi^{dd}(x) \equiv \inf_{y\in X}[d(x,y) - \varphi^d(y)].$$

Η συνάρτηση φ^d είναι 1-Lipschitz ως το infimum 1-Lipschitz συναρτήσεων οι οποίες είναι φραγμένες από κάτω σε κάποιο σημείο x_0. Άρα,

$$-\varphi^d(x) \leqslant \inf_{y\in X}[d(x,y) - \varphi^d(y)] \leqslant -\varphi^d(x), \tag{3.3.1}$$

όπου η δεξιά ανισότητα προκύπτει αν επιλέξουμε $x = y$ ενώ η αριστερή έπεται από το γεγονός ότι η φ^d είναι 1-Lipschitz: πράγματι, έχουμε

$$d(x,y) - \varphi^d(y) \geqslant |\varphi^d(y) - \varphi^d(x)| - \varphi^d(y) \geqslant -\varphi^d(x).$$

Συνεπώς έχουμε ότι $\varphi^{dd} = -\varphi^d$, και

$$\begin{aligned}\sup_{\Phi_c} J(\varphi,\psi) &\leqslant \sup_{\varphi\in L^1(d\mu)} J(\varphi^{dd},\varphi^d) = \sup_{\varphi\in L^1(d\mu)} J(-\varphi^d,\varphi^d)\\ &\leqslant \sup_{\|\varphi\|_{\mathrm{Lip}}\leqslant 1} J(\varphi,-\varphi) \leqslant \sup_{\Phi_c} J(\varphi,\psi)\end{aligned}$$

όπου η πρώτη ανισότητα έπεται από την Παρατήρηση 3.2.5, η δεύτερη από την σχέση (3.3.1) και η τρίτη από το γεγονός ότι η κλάση $J(\varphi,\psi)$ με $(\varphi,\psi) \in \Phi_d$ περιέχει την $J(\varphi,-\varphi)$ με $\|\varphi\|_{Lip} \leqslant 1$ (ισχύει $(\varphi,-\varphi) \in \Phi_d$), συνεπώς το supremum της πρώτης κλάσης θα είναι μεγαλύτερο ή ίσο από εκείνο της δεύτερης. Συνεπώς, έχουμε παντού ισότητα και το ζητούμενο έπεται. □

Εφαρμογή 3.3.2 (ο τύπος της ολικής κύμανσης)**.** Θεωρούμε έναν Πολωνικό χώρο X και την συνάρτηση κόστους $c(x,y) = \mathbf{1}_{\{x \neq y\}}$. Παρατηρούμε ότι η συνάρτηση κόστους c είναι μια κάτω ημισυνεχής μετρική στον X. Συνεπώς, για κάθε ζεύγος Borel μέτρων πιθανότητας μ και ν στον X, από το Θεώρημα 3.3.1 έχουμε

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} \mathbf{1}_{\{x \neq y\}} \, d\pi(x,y) = \sup \left\{ \int_X f \, d(\mu - \nu) : \sup_{x \neq y} |f(x) - f(y)| \leqslant 1 \right\},$$

δηλαδή

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi[\{x \neq y\}] = \sup_{0 \leqslant f \leqslant 1} \int_X f \, d(\mu - \nu).$$

Επιπλέον, ισχύει το εξής:

$$\begin{aligned}\sup_{0 \leqslant f \leqslant 1} \int_X f \, d(\mu - \nu) &= \sup_{0 \leqslant f \leqslant 1} \left[\int_X f \, d(\mu - \nu)_+ - \int_X f \, d(\mu - \nu)_- \right] \\ &= (\mu - \nu)_+(X) = (\mu - \nu)_-(X) = \frac{1}{2} \|\mu - \nu\|_{\mathrm{TV}},\end{aligned}$$

όπου χρησιμοποιήσαμε την ανάλυση: $(\mu - \nu) = (\mu - \nu)_+ - (\mu - \nu)_-$ και το γεγονός πως $(\mu - \nu)(X) = 0$, αφού τα μ, ν είναι Borel μέτρα πιθανότητας. Τώρα, χρησιμοποιώντας την κανονικότητα του $(\mu - \nu)_+$ έχουμε

$$(\mu - \nu)_+(X) = \sup_K (\mu - \nu)_+(K) = \sup_K (\mu(K) - \nu(K)),$$

όπου το supremum παίρνεται πάνω από όλα τα συμπαγή υποσύνολα K του X.

Παρατήρηση 3.3.3. Έστω $c(x,y) = |x-y|^2$ στον $\mathbb{R}^n$ και έστω μ, ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$ ώστε $\mathcal{T}_c(\mu,\nu) < +\infty$. Έστω επίσης $\pi \in \Pi(\mu,\nu)$ ένα σχέδιο μεταφοράς ώστε $I[\pi] < +\infty$. Στόχος μας είναι να ελαχιστοποιήσουμε όσο είναι δυνατόν το κόστος μεταφοράς. Ακολουθούμε την εξής στρατηγική: όποτε χρειάζεται να γίνει μια μεταφορά από ένα αρχικό σημείο x σε ένα τελικό σημείο y, εμείς κάνουμε τη μεταφορά από το x στο $\frac{x+y}{2}$ και *ταυτόχρονα* από το $\frac{x+y}{2}$ στο y. Η ποσότητα που θα χρειαζόταν να ελαχιστοποιήσουμε εάν μεταφέραμε κατευθείαν το x στο y θα ήταν η εξής:

$$\inf_{\Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 d\pi(x,y)$$

Από την άλλη αν ακολουθούσαμε την παραπάνω στρατηγική θα είχαμε να ελαχιστοποιήσουμε την εξής ποσότητα:

$$\inf_{\Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \left[|x - \frac{x+y}{2}|^2 + |\frac{x+y}{2} - y|^2 \right] d\pi = \inf_{\Pi(\mu,\nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2} d\pi.$$

Κατ' αυτόν τον τρόπο παρατηρούμε ότι το κόστος μεταφοράς μειώνεται κατά έναν παράγοντα 2. Συνεχίζοντας επαγωγικά κατά τον ίδιο τρόπο καταλήγουμε στο συμπέρασμα πως το βέλτιστο κόστος μεταφοράς είναι 0, το οποίο βέβαια μπορεί να συμβεί μόνο όταν $\mu = \nu$. Αποδεικνύεται ότι το συμπέρασμα αυτό ισχύει για κάθε δύναμη $|x-y|^p, p > 1$ αλλά και γενικότερα, όταν $c(x,y) = \varphi(|x-y|)$, όπου η φ είναι αύξουσα στον $\mathbb{R}_+$ με $\varphi(0) = 0$ και $\varphi'(0) = 0$.

3.4 Συνάρτηση κόστους με τιμές στο {0,1}

Το θεώρημα δυϊσμού του Kantorovich παίρνει πολύ συγκεκριμένη μορφή όταν η συνάρτηση κόστους παίρνει μόνο τις τιμές 0 και 1, δηλαδή όταν είναι της μορφής $\mathbf{1}_C(x,y)$ για κάποιο $C \subset X \times Y$.

Θεώρημα 3.4.1. *Έστω X και Y δύο Πολωνικοί χώροι. Έστω $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, και C ένα μη κενό ανοιχτό υποσύνολο του $X \times Y$. Τότε,*

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi(C) = \sup\{\mu(A) - \nu(A_C) : A \subset X,\ A \text{ κλειστό}\},$$

όπου

$$A_C = \{y \in Y : \exists x \in A, (x,y) \notin C\}.$$

Απόδειξη. Αφού το C είναι ανοιχτό, η συνάρτηση κόστους $c(x,y) = \mathbf{1}_C(x,y)$ είναι κάτω ημισυνεχής στον $X \times Y$. Άρα, μπορεί να προσεγγιστεί κατά σημείο από μια αύξουσα ακολουθία (c_k) συνεχών συναρτήσεων με $0 \leqslant c_k \leqslant c$. Συνεπώς, έχουμε ότι

$$\begin{aligned}\inf_{\pi \in \Pi(\mu,\nu)} \pi(C) &= \inf_{\pi \in \Pi(\mu,\nu)} \int 1_C d\pi = \lim_{k\to\infty} \inf_{\pi \in \Pi(\mu,\nu)} \int c_k\, d\pi \\ &= \lim_{k\to\infty} \sup\left\{\int_X \varphi\, d\mu + \int_Y \psi\, d\nu : (\varphi,\psi) \in \Phi_{c_k}\right\}.\end{aligned}$$

Από την Παρατήρηση 3.2.6, για κάθε k μπορούμε να περιορίσουμε το supremun σε εκείνα τα ζεύγη (φ,ψ) άνω ημισυνεχών συναρτήσεων που ικανοποιούν τις

$$0 \leqslant \varphi \leqslant 1,\ -1 \leqslant \psi \leqslant 0,$$

και ανήκουν στο Φ_{c_k}, δηλαδή ικανοποιούν την $\varphi(x) + \psi(y) \leqslant c_k(x,y) \leqslant c(x,y)$ για κάθε x, y. Έπεται ότι

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi(C) = \sup\left\{\int_X \varphi\, d\mu + \int_Y \psi\, d\nu : (\varphi,\psi) \in \tilde{\Phi}_c\right\}, \tag{3.4.1}$$

όπου $\tilde{\Phi}_c$ είναι το σύνολο όλων των ζευγών $(\varphi,\psi) \in L^1(d\mu) \times L^1(d\nu)$ ώστε

$$
(3.4.2) \qquad \begin{cases} \varphi(x)+\psi(y) \leqslant c(x,y) = \mathbf{1}_C(x,y) \text{ για κάθε } (x,y) \\ 0 \leqslant \varphi \leqslant 1, \ -1 \leqslant \psi \leqslant 0 \\ \text{η } \varphi \text{ είναι πάνω ημισυνεχής.} \end{cases}
$$

Παρατηρούμε ότι το $\tilde{\Phi}_c$ είναι κυρτό σύνολο.

Ισχυριζόμαστε ότι κάθε $(\varphi,\psi) \in \tilde{\Phi}_c$ μπορεί να αναπαρασταθεί ως κυρτός συνδυασμός ζευγών της μορφής $(\mathbf{1}_A, -\mathbf{1}_B)$, όπου το A είναι κλειστό και τα ζεύγη $(\mathbf{1}_A, -\mathbf{1}_B)$ ανήκουν στο $\tilde{\Phi}_c$ (δηλαδή, $\mathbf{1}_A(x) - \mathbf{1}_B(y) \leqslant \mathbf{1}_C(x,y)$ για κάθε x,y). Αν υποθέσουμε αυτόν τον ισχυρισμό τότε, αφού το συναρτησοειδές $J(\varphi,\psi) = \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$ που θέλουμε να μεγιστοποιήσουμε είναι γραμμικό, έπεται ότι για κάθε $(\varphi,\psi) \in \tilde{\Phi}_c$ υπάρχει ένα ζευγάρι $(\mathbf{1}_A, -\mathbf{1}_B)$ με $J(\mathbf{1}_A, -\mathbf{1}_B) \geqslant J(\varphi,\psi)$. Ειδικότερα, η τιμή στο δεξιό μέλος της (3.4.1) δεν αλλάζει εάν περιορίσουμε το supremum στα ζευγάρια της μορφής $(\mathbf{1}_A, -\mathbf{1}_B)$. Από αυτήν την παρατήρηση έπεται το θεώρημα. Πράγματι, η σχέση $\mathbf{1}_A - \mathbf{1}_B \leqslant \mathbf{1}_C$ συνεπάγεται ότι για κάθε $y \in Y$,

$$
\mathbf{1}_B(y) \geqslant \sup_{x\in X}[\mathbf{1}_A(x) - \mathbf{1}_C(x,y)] = \mathbf{1}_{A_C}(y),
$$

το οποίο σημαίνει ότι $A_C \subset B$, και άρα

$$
\mu(A) - \nu(B) \leqslant \mu(A) - \nu(A_C).
$$

Μένει να αποδείξουμε τον ισχυρισμό. Θα χρησιμοποιήσουμε το γεγονός ότι κάθε μετρήσιμη απεικόνιση $u : X \to [0,1]$ μπορεί να γραφεί ως $u = \int_0^1 \mathbf{1}_{\{u\geqslant s\}} ds$. Ειδικότερα, αν $(\varphi,\psi) \in \tilde{\Phi}_c$, μπορούμε να γράψουμε

$$
(\varphi,\psi) = \int_0^1 \left(\mathbf{1}_{\{\varphi\geqslant s\}}, \mathbf{1}_{\{\psi\geqslant s\}}\right) ds = \int_0^1 \left(\mathbf{1}_{\{\varphi\geqslant s\}}, -\mathbf{1}_{\{\psi\leqslant -s\}}\right) ds.
$$

Τώρα, αφού η φ είναι άνω ημισυνεχής, το σύνολο $\mathbf{1}_{\{\varphi\geqslant s\}}$ είναι κλειστό για κάθε $s \in [0,1]$. Άρα, το μόνο που έχουμε να ελέγξουμε είναι ότι, για κάθε s,

$$
\mathbf{1}_{\{\varphi\geqslant s\}}(x) - \mathbf{1}_{\{\psi\leqslant -s\}}(y) \leqslant \mathbf{1}_C(x,y).
$$

Η μόνη μη τετριμμένη περίπτωση είναι όταν $\varphi(x) \geqslant s$ και $\psi(y) > -s$, όπου και πρέπει να δείξουμε ότι $(x,y) \in C$. Σε αυτήν την περίπτωση όμως έπεται ότι $\varphi(x)+\psi(y) > 0$, άρα

$$
c(x,y) \geqslant \varphi(x) + \psi(y) > 0,
$$

και αφού η συνάρτηση c παίρνει τιμές μόνο στο $\{0,1\}$ έπεται ότι $c(x,y) = 1$, δηλαδή $(x,y) \in C$. Η απόδειξη είναι πλήρης. $\square$

Ένα πόρισμα του Θεωρήματος 3.4.1 είναι το εξής χρήσιμο θεώρημα του Strassen.

Πόρισμα 3.4.2 (το θεώρημα του Strassen). *Έστω X ένας Πολωνικός χώρος, $\mu, \nu \in \mathcal{P}(X)$ και $\varepsilon \geqslant 0$. Τότε,*

$$\inf_{\pi \in \Pi(\mu,\nu)} \pi(\{d(x,y) > \varepsilon\}) = \sup_{A \text{ κλειστό}} \{\mu(A) - \nu(A^{\epsilon})\},$$

όπου $A^{\epsilon} = \{y \in X : d(y, A) \leqslant \varepsilon\}$.

Πράγματι το σύνολο C του θεωρήματος 3.4.1 είναι σε αυτήν την περίπτωση το $C = \{d(x,y) > \varepsilon\}$ και $A_C = A^{\varepsilon} = \{y \in X : d(y, A) \leq \varepsilon\}$.

Κεφάλαιο 4

Αποστάσεις Wasserstein

Στο προηγούμενο κεφάλαιο ασχοληθήκαμε με το πρόβλημα της ύπαρξης και του χαρακτηρισμού ενός βέλτιστου σχεδίου μεταφοράς. Σε αυτό το κεφάλαιο εστιάζουμε στις πληροφορίες που μπορούμε να πάρουμε για τα μέτρα μ, ν όταν γνωρίζουμε την τιμή $T_c(\mu, \nu)$ του βέλτιστου κόστους μεταφοράς.

4.1 Αποστάσεις Monge-Kantorovich

Ορισμός 4.1.1. Έστω X ένας Πολωνικός χώρος εφοδιασμένος με μία απόσταση d και έστω $p \geqslant 0$ ένας μη αρνητικός πραγματικός αριθμός. Από εδώ και στο εξής θα θεωρούμε την συνάρτηση κόστους $c(x, y) = d(x, y)^p$, με την σύμβαση $d(x, y)^0 = \mathbf{1}_{\{x \neq y\}}$, και θα χρησιμοποιούμε τον συμβολισμό $T_p(\mu, \nu) = T_{d^p}(\mu, \nu)$ για το αντίστοιχο βέλτιστο κόστος μεταφοράς μεταξύ των μέτρων πιθανότητας μ, ν στον X.

Θα συμβολίζουμε με $\mathcal{P}_p(X)$ το σύνολο των μέτρων πιθανότητας με πεπερασμένες ροπές τάξης p, δηλαδή το σύνολο εκείνων των μέτρων μ με την ιδιότητα ότι για κάποιο x_0 (και συνεπώς για κάθε $x \in X$),

$$\int_X d(x_0, y)^p d\mu(y) < +\infty.$$

Βέβαια, αν η d είναι φραγμένη τότε το σύνολο $\mathcal{P}_p(X)$ συμπίπτει με το σύνολο $\mathcal{P}(X)$ όλων των μέτρων πιθανότητας στον X.

Θεώρημα 4.1.2 (αποστάσεις Wasserstein). (i) *Για κάθε $p \in [1, \infty)$ η $W_p = T_p^{\frac{1}{p}}$ ορίζει μία μετρική στον $\mathcal{P}_p(X)$.*

(ii) *Για κάθε $p \in [0, 1)$ η $W_p = T_p$ ορίζει μία μετρική στον $\mathcal{P}_p(X)$.*

Παρατηρήσεις 4.1.3. (i) Αν η d είναι φραγμένη τότε ως πόρισμα του προηγούμενου θεωρήματος έχουμε ότι η W_p ορίζει μία μετρική στον $\mathcal{P}(X)$. Στην περίπτωση που η d δεν είναι φραγμένη, μπορούμε να την αντικαταστήσουμε με την μετρική $\tilde{d} = \min\{d, 1\}$ η οποία επάγει την ίδια τοπολογία που επάγει η d.

(ii) Θα ονομάζουμε την W_p *απόσταση Monge-Kantorovich τάξης* p. Η απόσταση Monge-Kantorovich με εκθέτη 2, $W_2 = T_2^{\frac{1}{2}}$, θα ονομάζεται **τετραγωνική απόσταση Wasserstein**, ενώ η απόσταση Monge-Kantorovich με εκθέτη 1, $W_1 = T_1$, θα ονομάζεται **απόσταση Kantorovich-Rubinstein**. Τέλος, η απόσταση Monge-Kantorovich τάξης 0, $W_0 = T_0$, είναι το μισό της νόρμας της ολικής κύμανσης: υπενθυμίζουμε ότι όταν $c(x,y) = \mathbf{1}_{\{x\neq y\}}$, τότε το κόστος της ολικής μεταφοράς δίνεται από την

$$T_c(\mu,\nu) = \frac{1}{2}\|\mu-\nu\|_{\mathrm{TV}} \tag{4.1.1}$$

όπου

$$\|\mu\|_{\mathrm{TV}} = \inf\{\mu_+(X) + \mu_-(X)\}$$

πάνω από όλα τα ζεύγη μη αρνητικών μέτρων μ_+, μ_- για τα οποία το μ γράφεται στην μορφή $\mu = \mu_+ - \mu_-$.

(iii) Η απόσταση Kantorovich-Rubinstein μπορεί να οριστεί διαφορετικά, μέσω του δυϊκού τύπου Kantorovich-Rubinstein, ως εξής:

$$W_1(\mu,\nu) = \sup_{\|\varphi\|_{Lip}\leqslant 1} \int_X \varphi d(\mu-\nu).$$

Παρατηρούμε ότι αν η d είναι φραγμένη από κάποια θετική σταθερά M και η φ είναι 1-Lipschitz ως προς d, τότε $\sup\varphi - \inf\varphi \leqslant M$ και συνεπώς, χωρίς βλάβη της γενικότητας, μπορούμε να υποθέσουμε ότι $0 \leqslant \varphi \leqslant M$. Επιπλέον, το supremum μπορεί να περιοριστεί στις φραγμένες Lipschitz συναρτήσεις.

Πριν προχωρήσουμε στην απόδειξη του Θεωρήματος 4.1.2 θα διατυπώσουμε ένα λήμμα, μέσω του οποίου μπορούμε να «συγκολλήσουμε» δύο σχέδια μεταφοράς που έχουν κοινά περιθώρια μέτρα.

Λήμμα 4.1.4 (λήμμα συγκόλλησης). *Έστω μ_1, μ_2, μ_3 μέτρα πιθανότητας με φορείς τους Πολωνικούς χώρους X_1, X_2, X_3 αντίστοιχα, και έστω $\pi_{12} \in \Pi(\mu_1,\mu_2)$, $\pi_{23} \in \Pi(\mu_2,\mu_3)$ δύο σχέδια μεταφοράς. Τότε, υπάρχει ένα μέτρο πιθανότητας $\pi \in P(X_1 \times X_2 \times X_3)$ με περιθώρια μέτρα π_{12} στον $X_1 \times X_2$ και π_{23} στον $X_2 \times X_3$.*

Απόδειξη. Αν X, Y είναι δύο Πολωνικοί χώροι, το θεώρημα διάσπασης του μέτρου (βλέπε [30]) μας επιτρέπει να γράψουμε κάθε μέτρο πιθανότητας στον $X \times Y$ ως ένα μέσο όρο από μέτρα πιθανότητας στον $\{x\} \times Y$, για $x \in X$. Ειδικότερα, αν π είναι ένα μέτρο πιθανότητας στον $X \times Y$ με περιθώριο μέτρο μ στον X, τότε υπάρχει μετρήσιμη απεικόνιση $x \mapsto \pi_x$, από τον X στον $\mathcal{P}(Y)$, μονοσήμαντα ορισμένη μ-σχεδόν παντού, ώστε

$$\pi = \int_X (\delta_x \otimes \pi_x) d\mu(x). \tag{4.1.2}$$

Με την ισότητα (4.1.2) εννοούμε ότι, για κάθε $u \in C_b(X \times Y)$,

$$\int_{X \times Y} u(x,y) d\pi(x,y) = \int_X \left[\int_Y u(x,y) d\pi_x(y) \right] d\mu(x),$$

ή ισοδύναμα ότι για κάθε μετρήσιμο σύνολο $A \subset X \times Y$,

$$\pi(A) = \int_X \pi_x(A_x) d\mu(x),$$

όπου

$$A_x = \{y \in Y : (x,y) \in A\}.$$

Τώρα, θεωρούμε τα σχέδια μεταφοράς π_{12} και π_{23} που δίνονται στην υπόθεση του Λήμματος, και τα διασπούμε ως προς το κοινό τους περιθώριο μέτρο μ_2. Δηλαδή, βρίσκουμε μετρήσιμες απεικονίσεις $\pi_{12:2} : X_2 \to P(X_1)$ και $\pi_{23:2} : X_2 \to P(X_3)$ τέτοιες ώστε

$$\pi_{12} = \int_{X_2} \pi_{12:2} \otimes \delta_{x_2} d\mu_2(x_2)$$

και

$$\pi_{23} = \int_{X_2} \delta_{x_2} \otimes \pi_{23:2} d\mu_2(x_2).$$

Τέλος, κατασκευάζουμε το μέτρο $\pi \in \mathcal{P}(X_1 \times X_2 \times X_3)$ θέτοντας

$$\pi = \int_{X_2} (\pi_{12:2} \otimes \delta_{x_2} \otimes \pi_{23:2}) d\mu_2(x_2).$$

Παρατηρούμε ότι, για κάθε $A \subset X_1 \times X_2$,

$$\pi(A \times X_3) = \int_{X_2} \pi_{12:2}(A_{x_2}) \pi_{23:2}(X_3) d\mu_2(x_2) = \int_{X_2} \pi_{12:2}(A_{x_2}) d\mu_2(x_2) = \pi_{12}(A),$$

όπου $A_{x_2} = \{x_1 \in X_1 : (x_1, x_2) \in A\}$, και για κάθε $B \subset X_2 \times X_3$,

$$\pi(X_1 \times B) = \int_{X_2} \pi_{12:2}(X_1) \pi_{23:2}(B_{x_2}) d\mu_2(x_2) = \int_{X_2} \pi_{23:2}(B_{x_2}) d\mu_2(x_2) = \pi_{23}(B),$$

όπου $B_{x_2} = \{x_3 \in X_3 : (x_2, x_3) \in B\}$, δηλαδή το μέτρο π έχει περιθώρια μέτρα το π_{12} στον $X_1 \times X_2$ και το π_{23} στον $X_2 \times X_3$. Η απόδειξη του λήμματος είναι πλήρης. $\square$

Απόδειξη του Θεωρήματος 4.1.2. Αρκεί να αποδείξουμε την (i), διότι η (ii) αποτελεί ειδική περίπτωση της (i). Πράγματι, αν $0 < p < 1$ τότε αντικαθιστώντας την d με την τοπολογικά ισοδύναμή της d^p έχουμε το ζητούμενο λόγω της (i), ενώ αν $p = 0$ το συμπέρασμα είναι συνέπεια της (4.1.1). Συνεπώς, μπορούμε να υποθέσουμε ότι $p \geqslant 1$.

Είναι σαφές ότι η W_p είναι συμμετρική, μη αρνητική και πεπερασμένη στον $\mathcal{P}_p(X)$. Επιπλέον, $W_p(\mu, \mu) = 0$. Αντίστροφα, έστω μ, ν δύο μέτρα πιθανότητας τέτοια ώστε $W_p(\mu, \nu) = 0$. Θα δείξουμε ότι $\mu = \nu$. Αν π είναι ένα βέλτιστο σχέδιο μεταφοράς, τότε ο φορέας του π βρίσκεται στην διαγώνιο $\{y = x\}$. Συνεπώς, για κάθε $\varphi \in C_b(X)$ έχουμε

$$\int \varphi d\mu = \int \varphi(x) d\pi(x, y) = \int \varphi(y) d\pi(x, y) = \int \varphi d\nu,$$

από όπου έπεται ότι $\mu = \nu$.

Μένει λοιπόν να αποδείξουμε την τριγωνική ανισότητα για να ολοκληρώσουμε την απόδειξη. Έστω $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(X)$ και έστω τα βέλτιστα σχέδια μεταφοράς π_{12}, μεταξύ των μ_1, μ_2, και π_{23}, μεταξύ των μ_2, μ_3. Θεωρούμε X_i να είναι οι φορείς των μ_i, π ένα μέτρο όπως περιγράφεται στο λήμμα συγκόλλησης και π_{13} το περιθώριο μέτρο του π στον $X_1 \times X_3$. Είναι σαφές ότι $\pi_{13} \in \Pi(\mu_1, \mu_3)$. Χρησιμοποιώντας διαδοχικά τον ορισμό του T_p, την ιδιότητα των περιθώριων μέτρων, την τριγωνική ανισότητα και την ανισότητα Minkowski στον L^p, έχουμε

$$\begin{aligned}
W_p(\mu_1, \mu_3) &\leqslant \Big(\int_{X_1 \times X_3} d(x_1, x_3)^p d\pi_{13}(x_1, x_3) \Big)^{\frac{1}{p}} \\
&= \Big(\int_{X_1 \times X_2 \times X_3} d(x_1, x_3)^p d\pi(x_1, x_2, x_3) \Big)^{\frac{1}{p}} \\
&\leqslant \Big(\int_{X_1 \times X_2 \times X_3} (d(x_1, x_2) + d(x_2, x_3))^p d\pi(x_1, x_2, x_3) \Big)^{\frac{1}{p}} \\
&\leqslant \Big(\int_{X_1 \times X_2 \times X_3} d(x_1, x_2)^p d\pi(x_1, x_2, x_3) \Big)^{\frac{1}{p}} \\
&\qquad + \Big(\int_{X_1 \times X_2 \times X_3} d(x_2, x_3)^p d\pi(x_1, x_2, x_3) \Big)^{\frac{1}{p}} \\
&= \Big(\int_{X_1 \times X_2} d(x_1, x_2)^p d\pi_{12}(x_1, x_2) \Big)^{\frac{1}{p}} \\
&\qquad + \Big(\int_{X_2 \times X_3} d(x_2, x_3)^p d\pi_{23}(x_2, x_3) \Big)^{\frac{1}{p}} \\
&= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3).
\end{aligned}$$

Η απόδειξη είναι πλήρης. □

Παρατήρηση 4.1.5. Μια σημαντική ιδιότητα των αποστάσεων W_p είναι ότι είναι διατεταγμένες. Πράγματι, από την ανισότητα Hölder έπεται ότι

$$p_1 \geqslant p_2 \geqslant 1 \quad \Longrightarrow \quad W_{p_1} \geqslant W_{p_2}. \tag{4.1.3}$$

Δεν είναι γενικά εφικτό να συγκρίνουμε τις W_{p_1} και W_{p_2} προς την αντίθετη κατεύθυνση, εκτός αν η d είναι φραγμένη. Τότε, μπορούμε να δείξουμε χρησιμοποιώντας παρεμβολή ότι

$$p_1 \geqslant p_2 \geqslant 1 \quad \Longrightarrow \quad W_{p_1} \leqslant W_{p_2}^{\frac{p_2}{p_1}} \cdot \operatorname{diam}(X)^{1-\frac{p_2}{p_1}}, \tag{4.1.4}$$

όπου $\operatorname{diam}(X) = \sup\{d(x,y) : x, y \in X\}$ είναι η διάμετρος του X. Συνεπώς, σε αυτήν την περίπτωση όλες οι αποστάσεις W_p $(p \geqslant 1)$ ορίζουν την ίδια τοπολογία στον $\mathcal{P}(X)$.

Πρόταση 4.1.6 (έλεγχος της απόστασης Wasserstein μέσω σταθμισμένης ολικής κύμανσης)**.** *Έστω μ, ν δύο μέτρα πιθανότητας σε έναν Πολωνικό χώρο X και έστω d μία απόσταση στον X. Τότε, για κάθε $p \geqslant 0$ και για κάθε $x_0 \in X$ ισχύει*

$$T_p(\mu,\nu) \leqslant \max\{1, 2^{p-1}\} \int d(x_0,x)^p d|\mu-\nu|(x) = \max\{1, 2^{p-1}\} \|d(x_0,\cdot)^p(\mu-\nu)\|_{TV}.$$

Απόδειξη. Έστω π το σχέδιο μεταφοράς που προκύπτει αν κρατήσουμε σταθερή την κοινή μάζα των μ, ν και κατανείμουμε ομοιόμορφα την υπόλοιπη, δηλαδή

$$\pi = (Id \times Id)\#(\mu \wedge \nu) + \frac{1}{\alpha}(\mu-\nu)_+ \otimes (\mu-\nu)_-,$$

όπου $\mu \wedge \nu = \mu - (\mu-\nu)_+$ και $\alpha = (\mu-\nu)_-(X) = (\mu-\nu)_+(X)$. Ένας άλλος τρόπος γραφής του π είναι ο εξής:

$$d\pi(x,y) = d(\mu \wedge \nu)(x)\delta_{y=x} + \frac{1}{\alpha} d(\mu-\nu)_+(x) d(\mu-\nu)_-(y).$$

Χρησιμοποιώντας διαδοχικά τον ορισμό του T_p, τον ορισμό του π, την τριγωνική ανισότητα για την d, την στοιχειώδη ανισότητα $(A+B)^p \leqslant \max\{1, 2^{p-1}\}\,(A^p + B^p)$ και τον ορισμό του α, έχουμε

$$\begin{aligned}
T_p(\mu,\nu) &\leqslant \int d(x,y)^p d\pi(x,y) \\
&= \frac{1}{\alpha}\int d(x,y)^p d(\mu-\nu)_+(x) d(\mu-\nu)_-(y) \\
&\leqslant \frac{\max\{1,2^{p-1}\}}{\alpha}\int (d(x,x_0)^p + d(x_0,y)^p) d(\mu-\nu)_+(x) d(\mu-\nu)_-(y) \\
&\leqslant \max\{1,2^{p-1}\}\left[\int d(x,x_0)^p d(\mu-\nu)_+(x) + \int d(x_0,y)^p d(\mu-\nu)_-(y)\right] \\
&= \max\{1,2^{p-1}\}\int d(x,x_0)^p d[(\mu-\nu)_+ + (\mu-\nu)_-](x) \\
&= \max\{1,2^{p-1}\}\int d(x,x_0)^p d|\mu-\nu|(x).
\end{aligned}$$

Αυτό αποδεικνύει την πρόταση. □

4.2 Η τοπολογία των αποστάσεων Wasserstein

Σε αυτήν την ενότητα θα ασχοληθούμε με τις τοπολογικές ιδιότητες των αποστάσεων Monge-Kantorovich W_p. Στα επόμενα ο χώρος X υποτίθεται Πολωνικός και εφοδιασμένος με μία απόσταση d.

Θεώρημα 4.2.1 (οι αποστάσεις Wasserstein μετρικοποιούν την w^* σύγκλιση)**.** *Έστω $p \in (0,\infty)$, $(\mu_k)_{k\in\mathbb{N}}$ μία ακολουθία μέτρων πιθανότητας στον $\mathcal{P}_p(X)$ και $\mu \in \mathcal{P}(X)$. Τα ακόλουθα είναι ισοδύναμα:*

(i) $W_p(\mu_k,\mu) \longrightarrow 0$, *όταν* $k \to \infty$.

(ii) $\mu_k \overset{w^*}{\to} \mu$, *όταν* $k\to\infty$ *και η* μ_k *ικανοποιεί την εξής συνθήκη* tightness*: για κάποιο* $x_0 \in X$ *(και λόγω της τριγωνικής ανισότητας, για κάθε* x*)*

$$\lim_{R\to\infty}\limsup_{k\to\infty}\int_{d(x_0,y)\geqslant R} d(x_0,y)^p d\mu_k(y) = 0. \tag{4.2.1}$$

(iii) $\mu_k \overset{w^*}{\longrightarrow} \mu$, *όταν* $k\to\infty$ *και οι ροπές τάξης* p *συγκλίνουν, δηλαδή για κάποιο* $x_0 \in X$ *(άρα και για κάθε* $x \in X$*),*

$$\int d(x_0,y)^p d\mu_k(y) \overset{k\to\infty}{\longrightarrow} \int d(x_0,y)^p d\mu(y). \tag{4.2.2}$$

(iv) *Για κάθε συνεχή συνάρτηση φ στον X που ικανοποιεί την* αυξητική συνθήκη $|\varphi(x)| \leqslant C(1 + d(x_0, x)^p)$ *για κάποιο $x_0 \in X$ και κάποια σταθερά $C > 0$, έχουμε*

$$\int \varphi d\mu_k \overset{k\to\infty}{\longrightarrow} \int \varphi d\mu.$$

Απόδειξη. Θα χωρίσουμε την απόδειξη σε βήματα.

1. Έστω ότι ισχύουν οι υποθέσεις του θεωρήματος. Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $p \geqslant 1$ (βλέπε την απόδειξη του Θεωρήματος 4.1.2). Είναι σαφές ότι από την (iv) έπεται η (iii), καθώς η $d(x_0, x)^p$ είναι μία συνεχής συνάρτηση που ικανοποιεί την αυξητική συνθήκη στην (iv). Αρχικά, λοιπόν θα δείξουμε την ισοδυναμία των (ii),(iii) και (iv).

(ii) $\Rightarrow$ *(iv).* Έστω ότι η (ii) ικανοποιείται για κάποιο $x_0 \in X$ και έστω φ τυχούσα συνεχής συνάρτηση που ικανοποιεί την αυξητική συνθήκη στην (iv). Τότε, για κάθε $R > 1$ γράφουμε

$$\varphi = \varphi_R + \psi_R,$$

όπου οι $\varphi_R(x) = \inf\{\varphi(x), C(1 + R^p)\}$ και $\psi_R(x) = \varphi(x) - \varphi_R(x)$ είναι κατά σημείο φραγμένες από την ποσότητα $Cd(x_0, x)^p \cdot \mathbf{1}_{\{d(x_0,x)\geqslant R\}}$. Τότε, έχουμε

$$\begin{aligned}\left|\int \varphi d\mu_k - \int \varphi d\mu\right| \leqslant \left|\int \varphi_R d(\mu_k - \mu)\right| &+ C \int_{\{d(x_0,x)\geqslant R\}} d(x_0, x)^p d\mu_k(x) \\ &+ C \int_{\{d(x_0,x)\geqslant R\}} d(x_0, x)^p d\mu(x).\end{aligned}$$

Συνεπώς,

$$\limsup_{k\to\infty} \left|\int \varphi d\mu_k - \int \varphi d\mu\right| \leqslant C \limsup_{k\to\infty} \int_{\{d(x_0,x)\geqslant R\}} d(x_0, x)^p (d\mu_k + d\mu)(x),$$

και αφήνοντας το $R \to \infty$ παρατηρούμε ότι το δεξιό μέλος συγκλίνει στο 0 και άρα παίρνουμε την (iv).

(iii) $\Rightarrow$ *(ii).* Έστω ότι ισχύει η (iii) για κάποιο $x_0 \in X$. Με τον συμβολισμό $\alpha \wedge \beta = \min\{\alpha, \beta\}$, έχουμε

$$\int (d(x_0, x) \wedge R)^p d\mu_k(x) \overset{k\to\infty}{\longrightarrow} \int (d(x_0, x) \wedge R)^p d\mu(x).$$

Από την άλλη πλευρά, από το θεώρημα μονότονης σύγκλισης έχουμε

$$\lim_{R\to\infty} \int (d(x_0, x) \wedge R)^p d\mu(x) = \int d(x_0, x)^p d\mu(x),$$

ενώ από την υπόθεση της (iii) ισχύει

$$\int d(x_0,x)^p d\mu_k(x) \overset{k\to\infty}{\longrightarrow} \int d(x_0,x)^p d\mu(x).$$

Συνδυάζοντας τα παραπάνω, έχουμε

$$\lim_{R\to\infty}\lim_{k\to\infty}\int [d(x_0,x)^p - (d(x_0,x)\wedge R)^p] d\mu_k(x) = 0.$$

Τώρα, αν $d(x_0,x) \geqslant 2R$ τότε $d(x_0,x)^p - R^p \geqslant (1-\frac{1}{2^p})d(x_0,x)^p$, και έπεται ότι

$$\lim_{R\to\infty}\limsup_{k\to\infty}\int_{\{d(x_0,x)\geqslant 2R\}} d(x_0,x)^p d\mu_k(x) = 0,$$

που είναι ακριβώς η (ii).

2. Για να ολοκληρώσουμε την απόδειξη του θεωρήματος αρκεί να αποδείξουμε την ισοδυναμία των (i) και (iii). Παρατηρούμε ότι, από την w^* σύγκλιση της (μ_k) στο μ, έχουμε

$$\begin{aligned}\int d(x_0,x)^p d\mu(x) &= \lim_{R\to\infty}\lim_{k\to\infty}\int (d(x_0,x)\wedge R)^p d\mu_k(x)\\ &\leqslant \liminf_{k\to\infty}\int d(x_0,x)^p d\mu_k(x),\end{aligned}$$

συνεπώς η συνθήκη σύγκλισης των ροπών τάξης p στην (iii) είναι ισοδύναμη με την

$$\limsup_{k\to\infty}\int d(x_0,x)^p d\mu_k(x) \leqslant \int d(x_0,x)^p d\mu(x). \tag{4.2.3}$$

3. Θα δείξουμε ότι η W_p-σύγκλιση συνεπάγεται την (4.2.3). Για τον σκοπό αυτό θα χρησιμοποιήσουμε την εξής ανισότητα: για κάθε $\varepsilon > 0$ υπάρχει σταθερά $C_\varepsilon > 0$ τέτοια ώστε για οποιουσδήποτε μη αρνητικούς πραγματικούς αριθμούς α,β να ισχύει

$$(\alpha+\beta)^p \leqslant (1+\varepsilon)\alpha^p + C_\varepsilon \beta^p.$$

Συνδυάζοντας αυτήν την ανισότητα με την συνήθη τριγωνική ανισότητα παρατηρούμε ότι, για κάθε $x_0,x,y\in X$,

$$d(x_0,x)^p \leqslant (1+\varepsilon)d(x_0,y)^p + C_\varepsilon d(x,y)^p. \tag{4.2.4}$$

Έστω τώρα $(\mu_k)\subset \mathcal{P}_p(x)$ μία ακολουθία μέτρων πιθανότητας τέτοια ώστε $W_p(\mu_k,\mu)\to 0$, και έστω π_k ένα βέλτιστο σχέδιο μεταφοράς μεταξύ των μ_k και μ. Ολοκληρώνοντας την (4.2.4) ως προς π_k και χρησιμοποιώντας την ιδιότητα των περιθώριων μέτρων έχουμε

$$\int d(x_0,x)^p d\mu_k(x) \leqslant (1+\varepsilon)\int d(x_0,y)^p d\mu(y) + C_\varepsilon \int d(x,y)^p d\pi_k(x,y).$$

Όμως, λόγω της W_p-σύγκλισης έχουμε

$$\int d(x,y)^p d\pi_k(x,y) = W_p(\mu_k,\mu)^p \overset{k\to\infty}{\longrightarrow} 0.$$

Συνεπώς,

$$\limsup_{k\to\infty} \int d(x_0,x)^p d\mu_k(x) \leqslant (1+\varepsilon) \int d(x_0,x)^p d\mu(x),$$

και αφήνοντας το $\varepsilon \to 0$ παίρνουμε την (4.2.3).

4. Μένει να δείξουμε ότι η W_p-σύγκλιση έπεται την w^*-σύγκλιση του μ_k στο μ και ότι η (iii) έπεται την (i). Μπορούμε να υποθέσουμε ότι η d είναι φραγμένη: πράγματι, επιλέγουμε $\tilde{d} = \inf\{d,1\}$ (επάγει την ίδια τοπολογία με την d) και θέτουμε $\tilde{W}_p$ την μετρική Monge-Kantorovich τάξης p που κατασκευάζεται από την $\tilde{d}$. Παρατηρούμε ότι $W_p \geqslant \tilde{W}_p$, συνεπώς αν θέλουμε να δείξουμε ότι η W_p-σύγκλιση έπεται την w^*-σύγκλιση αρκεί να δείξουμε ότι η $\tilde{W}_p$-σύγκλιση έπεται την w^*-σύγκλιση. Πράγματι, έστω ότι η (iii) ισχύει και ότι $\tilde{W}_p(\mu_k,\mu) \to 0$, θα δείξουμε ότι $W_p(\mu_k,\mu)\to 0$. Για τον σκοπό αυτό θα χρησιμοποιήσουμε την εξής ανισότητα,

$$d(x,y) \leqslant d(x,y)\wedge R + 2d(x,x_0)\cdot 1_{d(x,x_0)\geqslant \frac{R}{2}} + 2d(x_0,y)\cdot 1_{d(x_0,y)\geqslant \frac{R}{2}},$$

και το πόρισμά της,

$$d(x,y)^p \leqslant C_p\Big((d(x,y)\wedge R)^p + d(x,x_0)^p\cdot 1_{d(x,x_0)\geqslant \frac{R}{2}} + d(x_0,y)^p\cdot 1_{d(x_0,y)\geqslant \frac{R}{2}}\Big),$$

για κάποια σταθερά C_p που εξαρτάται μόνο από το p. Έστω π_k ένα βέλτιστο σχέδιο μεταφοράς μεταξύ των μ_k και μ για την συνάρτηση κόστους d^p. Από την παραπάνω ανισότητα για $R\geqslant 1$, έχουμε

$$\begin{aligned} W_p(\mu_k,\mu)^p = \int d(x,y)^p d\pi_k(x,y) &\leqslant C_p \int (d(x,y)\wedge R)^p d\pi_k(x,y) + \\ &C_p \int_{d(x,x_0)\geqslant \frac{R}{2}} d(x,x_0)^p \pi_k(x,y) + C_p \int_{d(x_0,y)\geqslant\frac{R}{2}} d(x_0,y)^p d\pi_k(x,y) \\ &\leqslant C_p R^p \tilde{W}_p^p(\mu_k,\mu) + C_p\int_{d(x,x_0)\geqslant\frac{R}{2}} d(x,x_0)^p d\mu_k(x) + C_p\int_{d(x_0,y)\geqslant\frac{R}{2}} d(x_0,y)^p d\mu(y). \end{aligned}$$

Αφήνοντας το $k\to\infty$ και λαμβάνοντας υπόψιν ότι $\tilde{W}_p(\mu_k,\mu)\to 0$ παίρνουμε,

$$\limsup_{k\to\infty} W_p^p(\mu_k,\mu) \leqslant C_p \limsup_{k\to\infty} \int_{d(x,x_0)\geqslant\frac{R}{2}} d(x,x_0)^p d\mu_k(x) + C_p \int_{d(x_0,y)\geqslant\frac{R}{2}} d(x_0,y)^p d\mu(y).$$

Τέλος, λόγω της (iii) η (μ_k) έχει ομοιόμορφα ολοκληρώσιμη ροπή τάξης p, συνεπώς αφήνοντας το $R\to\infty$ παίρνουμε ότι $W_p^p(\mu_k,\mu)\to 0$.

5. Από εδώ και στο εξής λοιπόν, υποθέτουμε ότι η d είναι φραγμένη, έστω $d \leqslant 1$. Συνεπώς οι αποστάσεις W_p είναι ισοδύναμες και λόγω του βήματος 4 αρκεί να δείξουμε το θεώρημα για $p = 1$. Χρησιμοποιώντας τον δυϊκό τύπο των Kantorovich-Rubinstein βλέπουμε ότι η (i) παίρνει τη μορφή

$$\sup_{\|\varphi\|_{\mathrm{Lip}} \leqslant 1} \int \varphi d(\mu_k - \mu) \overset{k\to\infty}{\longrightarrow} 0, \tag{4.2.5}$$

όπου η φ μπορεί να υποτεθεί απολύτως φραγμένη απο 1 (βλέπε παρατήρηση 4.1.3 (iii).)

6. Υποθέτουμε λοιπόν ότι $W_1(\mu_k, \mu) \to 0$ και θα αποδείξουμε ότι $\mu_k \overset{w^*}{\longrightarrow} \mu$, δηλαδή για κάθε $\varphi \in C_b(X)$,

$$\int \varphi d\mu_k \overset{k\to\infty}{\longrightarrow} \int \varphi d\mu. \tag{4.2.6}$$

Από την (4.2.5) γνωρίζουμε ότι η (4.2.6) ισχύει αν η φ είναι 1-Lipschitz. Αντικαθιστώντας την φ με την $\frac{\varphi}{\|\varphi\|_{Lip}}$ αν $\varphi \neq 0$, βλέπουμε ότι η (4.2.6) εξακολουθεί να ισχύει αν η φ είναι τυχούσα Lipschitz συνάρτηση. Τώρα, για να αποδείξουμε τον ισχυρισμό μας θα χρησιμοποιήσουμε μία ιδιότητα των μετρικών χώρων: κάθε συνεχής και φραγμένη συνάρτηση σε έναν μετρικό χώρο προσεγγίζεται από κάτω και από πάνω από μία ακολουθία Lipschitz συναρτήσεων. Πιο συγκεκριμένα, υπάρχουν ακολουθίες $(\alpha_n)_{n\in\mathbb{N}}$ και $(\beta_n)_{n\in\mathbb{N}}$ από Lipschitz συναρτήσεις, ομοιόμορφα φραγμένες, τέτοιες ώστε η (α_n) να είναι αύξουσα, η (β_n) φθίνουσα και

$$\lim_{n\to\infty} \alpha_n = \varphi = \lim_{n\to\infty} \beta_n.$$

Τότε,

$$\limsup_{k\to\infty} \int \varphi d\mu_k \leqslant \liminf_{n\to\infty} \limsup_{k\to\infty} \int b_n d\mu_k = \liminf_{n\to\infty} \int b_n d\mu = \int \varphi d\mu,$$

όπου στην τελευταία ισότητα χρησιμοποιήσαμε το θεώρημα κυριαρχημένης σύγκλισης του Lebesgue. Όμοια, $\liminf_{k\to\infty} \int \varphi d\mu_k \geqslant \int \varphi d\mu$, το οποίο ολοκληρώνει την απόδειξη της (4.2.6).

7. Υποθέτουμε τώρα ότι $\mu_k \overset{w^*}{\longrightarrow} \mu$ και θα αποδείξουμε την (4.2.5), δηλαδή ότι $W_1(\mu_k, \mu) \to 0$. Έστω $x_0 \in X$ και έστω $\mathrm{Lip}_{1,x_0}(X)$ ο χώρος όλων των Lipschitz συναρτήσεων φ στον X με σταθερά το πολύ 1, που επιπλέον ικανοποιούν την $\varphi(x_0) = 0$. Για να αποδείξουμε την (4.2.5) αρκεί να δείξουμε ότι

$$\sup_{\varphi \in \mathrm{Lip}_{1,x_0}} \int \varphi d(\mu_k - \mu) \overset{k\to\infty}{\longrightarrow} 0.$$

Από το θεώρημα Prokhorov γνωρίζουμε ότι υπάρχει αύξουσα ακολουθία συμπαγών συνόλων $(K_n)_{n\geqslant 1}$ τέτοια ώστε για κάθε $n \geqslant 1$ να έχουμε $\sup_k \mu_k(K_n^c) \leqslant \frac{1}{n}$ και $\mu(K_n^c) \leqslant \frac{1}{n}$. Χωρίς βλάβη της γενικότητας υποθέτουμε ότι $x_0 \in K_1$. Τότε, για κάθε $n \geqslant 1$ το

$$\{\varphi \cdot \mathbf{1}_{K_n} : \ \varphi \in \mathrm{Lip}_{1,x_0}(X)\}$$

είναι ένα υποσύνολο του $\mathrm{Lip}_{1,x_0}(K_n)$ και από το θεώρημα Ascoli είναι ένα συμπαγές υποσύνολο του $C_b(K_n)$ (εφοδιασμένο με την $\|\cdot\|_\infty$). Συνεπώς, για κάθε n και για κάθε ακολουθία στον Lip_{1,x_0} υπάρχει υπακολουθία που συγκλίνει ομοιόμορφα στο K_n. Με ένα διαγώνιο επιχείρημα έχουμε το εξής: για κάθε ακολουθία $(\varphi_k)_{k\in\mathbb{N}} \in \mathrm{Lip}_{1,x_0}$ υπάρχει υπακολουθία (φ_{k_n}) η οποία συγκλίνει ομοιόμορφα στο K_n σε μία μετρήσιμη συνάρτηση φ_∞, ορισμένη στο $S = \cup K_n$, το οποίο είναι ένα Lipschitz φραγμένο σύνολο αφού η οικογένεια (φ_k) είναι ομοιόμορφα φραγμένη και ομοιόμορφα Lipschitz.

Εφαρμόζουμε το συμπέρασμα αυτό σε μία οικογένεια (φ_k) που ικανοποιεί την

$$\sup_{\varphi\in\mathrm{Lip}_{1,x_0}} \int \varphi d(\mu_k - \mu) \leqslant \int \varphi_k d(\mu_k - \mu) + \frac{1}{k}.$$

Συνεπώς, υπάρχει υπακολουθία (φ_{k_s}), η οποία συγκλίνει ομοιόμορφα σε κάθε K_n, σε μία $1-$Lipschitz συνάρτηση φ_∞ επί του $S = \cup K_n$.

Τώρα, θα χρησιμοποιήσουμε το γεγονός ότι κάθε $1-$Lipschitz συνάρτηση F ορισμένη σε ένα υποσύνολο S ενός μετρικού χώρου X μπορεί να επεκταθεί σε μία $1-$Lipschitz συνάρτηση $\tilde{F}$ σε ολόκληρο το X. Για παράδειγμα, μπορούμε να χρησιμοποιήσουμε την $\tilde{F}(x) = \inf_{y\in S}(F(y) + d(x,y))$. Συνεπώς, μπορούμε να επεκτείνουμε την φ_∞ σε ένα στοιχείο του συνόλου $\mathrm{Lip}_{1,x_0}(X)$: ειδικότερα, η φ_∞ είναι συνεχής και φραγμένη.

Για να ολοκληρώσουμε την απόδειξη μένει να δείξουμε ότι $\int \varphi_{k_s} d(\mu_{k_s} - \mu) \to 0$ καθώς το $k \to \infty$. Γράφουμε

$$\begin{aligned}\int \varphi_{k_s} d(\mu_{k_s} - \mu) \leqslant &\left| \int_{K_n} (\varphi_{k_s} - \varphi_\infty) d(\mu_{k_s} - \mu) \right| \\ &+ \left| \int_{K_n^c} (\varphi_{k_s} - \varphi_\infty) d(\mu_{k_s} - \mu) \right| + \left| \int_X \varphi_\infty d(\mu_{k_s} - \mu) \right|.\end{aligned}$$

Όμως, για κάθε n, το $\int_{K_n} (\varphi_{k_s} - \varphi_\infty) d(\mu_{k_s} - \mu) \to 0$ καθώς $s \to \infty$, διότι η φ_{k_s} συγκλίνει ομοιόμορφα στην φ_∞ στο K_n καθώς $s \to \infty$. Επιπλέον αφού οι φ_{k_s} και φ_∞ είναι ομοιόμορφα φραγμένες, ο δεύτερος όρος του δεξιού μέλους της ανισότητας είναι φραγμένος από $C(\mu_{k_s}(K_n^c) + \mu(K_n^c)) \leqslant 2\frac{C}{n}$, για κάποια σταθερά $C > 0$ άρα συγκλίνει ομοιόμορφα ως προς s στο 0 καθώς $n \to \infty$. Τέλος, $\int_X \varphi_\infty d(\mu_{k_s} - \mu) \to 0$ καθώς $s \to \infty$ διότι $\mu_k \xrightarrow{w^*} \mu$. Για να ολοκληρώσουμε την απόδειξη του θεωρήματος, αφήνουμε το $n \to \infty$ και έπειτα το $s \to \infty$. □

Παρατηρήσεις 4.2.2. (i) Ως συνέπεια του θεωρήματος έχουμε ότι η W_p μετρικοποιεί την w^* τοπολογία των μέτρων σε κάθε υποσύνολο του $\mathcal{P}_p(X)$ το οποίο ικανοποιεί την συνθήκη tightness στην (ii).

(ii) Ειδικότερα, αν η d είναι φραγμένη τότε η W_p μετρικοποιεί την w^* σύγκλιση σε όλο το $\mathcal{P}(X)$. Όμως, αφού μπορούμε πάντα να αντικαταστήσουμε την d με την τοπολογικά ισοδύναμη και φραγμένη $\tilde{d} = \min\{d, 1\}$, έπεται ότι ο $\mathcal{P}(X)$ εφοδιασμένος με την w^* τοπολογία είναι ένας μετρικός χώρος.

(iii) Από την τριγωνική ανισότητα έπεται ότι αν $W_p(\mu_k, \mu) \to 0$ και $W_p(\nu_k, \nu) \to 0$, τότε $W_p(\mu_k, \nu_k) \to 0$.

4.3 Διαβαθμισμένη συνέλιξη

Στην ενότητα αυτή θα διατυπώσουμε κάποιες στοιχειώδεις και πολύ χρήσιμες ιδιότητες, σχετικές με την κυρτότητα, που ικανοποιούν οι αποστάσεις Monge-Kantorovich. Λόγω της κυρτότητας του προβλήματος ελαχιστοποίησης Monge-Kantorovich, έχουμε ως άμεση συνέπεια ότι για κάθε συνάρτηση κόστους c, για οποιαδήποτε μέτρα πιθανότητας $\mu_1, \mu_2, \nu_1, \nu_2$ και για κάθε $\alpha \in [0, 1]$ ισχύει

$$T_c(\alpha\mu_1 + (1-\alpha)\mu_2, \alpha\nu_1 + (1-\alpha)\nu_2) \leqslant \alpha T_c(\mu_1, \nu_1) + (1-\alpha)T_c(\mu_2, \nu_2).$$

Όταν ο χώρος X στον οποίο δουλεύουμε είναι χώρος με νόρμα και η συνάρτηση κόστους είναι η p-δύναμη της νόρμας, τότε έχουμε κάποιες επιπλέον ιδιότητες. Πριν προχωρήσουμε στην διατύπωση αυτών των ιδιοτήτων ας υπενθυμίσουμε μία βασική ιδιότητα για τυχαίες μεταβλητές που θα χρησιμοποιήσουμε στην συνέχεια:

> Αν U και V είναι δύο τυχαίες μεταβλητές με αντίστοιχες κατανομές μ, ν, γράφουμε
> $$T_p(U, V) = T_p(\mu, \nu), \; W_p(U, V) = W_p(\mu, \nu)$$
> για το βέλτιστο κόστος μεταφοράς και την απόσταση Wasserstein, όπου χρησιμοποιούμε ως συνάρτηση κόστους την συνάρτηση $\|\cdot\|^p$.

Εξ ορισμού, μία τυχαία μεταβλητή U είναι στοιχείο του χώρου L^p αν η μέση τιμή $\mathbb{E}\|U\|^p < +\infty$.

Πρόταση 4.3.1 (συμπεριφορά των αποστάσεων Wasserstein ως προς διαβαθμισμένη συνέλιξη)**.** *Έστω X χώρος με νόρμα και $p \geqslant 1$.*

(i) *Αν οι τυχαίες μεταβλητές U, V, με τιμές στον X, ανήκουν στον L^p και $\alpha \in \mathbb{R}$, τότε*

$$T_p(\alpha U, \alpha V) = |\alpha|^p T_p(U, V), \quad W_p(\alpha U, \alpha V) = |\alpha| W_p(U, V).$$

(ii) *Αν οι τυχαίες μεταβλητές U_1, U_2, V_1, V_2, με τιμές στον X, ανήκουν στον L^p και $\alpha_1, \alpha_2 \in \mathbb{R}$, τότε*

$$T_p(\alpha_1 U_1 + \alpha_2 U_2, \alpha_1 V_1 + \alpha_2 V_2) \leqslant 2^{p-1}(|\alpha_1|^p T_p(U_1, V_1) + |\alpha_2|^p T_p(U_2, V_2)).$$

Απόδειξη. Είναι πιο βολικό να δουλέψουμε με τυχαίες μεταβλητές.

(i) Γράφουμε,

$$\mathrm{E}\|\alpha U - \alpha V\|^p = |\alpha|^p \mathrm{E}\|U - V\|^p.$$

Παίρνοντας, τώρα infimum στο αριστερό μέλος (πάνω από όλα τα πιθανά ζεύγη αU και αV) και έπειτα στο δεξιό μέλος, παίρνουμε το ζητούμενο.

(ii) Θα εργαστούμε όπως στο (i). Γράφουμε

$$\begin{aligned}\mathbb{E}\|(\alpha_1 U_1 + \alpha_2 U_2) - (\alpha_1 V_1 + \alpha_2 V_2)\|^p &= \mathbb{E}\|\alpha_1(U_1 - V_1) + \alpha_2(U_2 - V_2)\|^p \\ &\leqslant 2^{p-1}(|\alpha_1|^p \mathbb{E}\|U_1 - V_1\|^p + |\alpha_2|^p \mathbb{E}\|U_2 - V_2\|^p),\end{aligned}$$

όπου και πάλι χρησιμοποιήσαμε την ανισότητα $(\alpha + \beta)^p \leqslant 2^{p-1}(\alpha^p + \beta^p)$, και παίρνοντας infimum (πάνα από όλα τα πιθανά ζεύγη $\alpha_1 U_1 + \alpha_2 U_2$ και $\alpha_1 V_1 + \alpha_2 V_2$) και στα δύο μέλη έχουμε το ζητούμενο.

□

Όταν $p = 2$ και ο X είναι χώρος Hilbert, το προηγούμενο φράγμα μπορεί να βελτιωθεί αν προσθέσουμε μια συνθήκη για την μέση τιμή, και αυτό είναι πολύ σημαντικό σε πολλές εφαρμογές. Έχουμε, λοιπόν την εξής πρόταση:

Πρόταση 4.3.2 (υποπροσθετικότητα της T_2 ως προς διαβαθμισμένη συνέλιξη)**.** *Έστω $U_1, U_2, V_1, V_2 \in L^2$ τυχαίες μεταβλητές, με τιμές στον χώρο Hilbert X, τέτοιες ώστε $\mathbb{E}U_1 = \mathbb{E}V_1$ ή $\mathbb{E}U_2 = \mathbb{E}V_2$, και έστω $\alpha_1, \alpha_2 \in \mathbb{R}$. Τότε,*

$$T_2(\alpha_1 U_1 + \alpha_2 U_2, \alpha_1 V_1 + \alpha_2 V_2) \leqslant \alpha_1^2 T_2(U_1, V_1) + \alpha_2^2 T_2(U_2, V_2). \tag{4.3.1}$$

Απόδειξη. Χωρίς βλάβη της γενικότητας μπορούμε να επιλέξουμε το ζεύγος (U_1, V_1) ανεξάρτητο από το (U_2, V_2) αφού μόνο οι κατανομές των $\alpha_1 U_1 + \alpha_2 U_2$ και $\alpha_1 V_1 + \alpha_2 V_2$ παίζουν ρόλο στον ορισμό του αριστερού μέλους της (4.3.1). Συνεπώς, για την τετραγωνική μέση τιμή έχουμε

$$\begin{aligned}\mathbb{E}\|\alpha_1(U_1 - V_1) + \alpha_2(U_2 - V_2)\|^2 =& \alpha_1^2 \mathbb{E}\|U_1 - V_1\|^2 + \alpha_2^2 \mathbb{E}\|U_2 - V_2\|^2 \\ &+ \alpha_1 \alpha_2 \mathbb{E}\langle U_1 - V_1, U_2 - V_2\rangle,\end{aligned}$$

όπου παρατηρούμε ότι ο τελευταίος όρος μηδενίζεται, διότι λόγω της ανεξαρτησίας των ζευγών (U_1, V_1) και (U_2, V_2) είναι ίσος με $2\alpha_1\alpha_2\langle \mathbb{E}U_1 - \mathbb{E}U_2, \mathbb{E}V_1 - \mathbb{E}V_2\rangle$ και, λόγω της υπόθεσης που έχουμε κάνει για τη μέση τιμή, μηδενίζεται. Παίρνοντας το infimum όπως πριν, έχουμε το ζητούμενο. □

Παρατήρηση 4.3.3. *Έστω X_1, X_2 δύο τυχαίες μεταβλητές με $\mathrm{law}(X_1) = \mu_1$ και $\mathrm{law}(X_2) = \mu_2$. Τότε,*

$$\mathrm{law}(\alpha X_1 + (1-\alpha)X_2) = (m_\alpha \# \mu_1) * (m_{1-\alpha} \# \mu_2),$$

όπου με $*$ συμβολίζουμε την συνέλιξη και με m_λ τον πολλαπλασιασμό με λ. Αυτό εξηγεί και την ορολογία «διαβαθμισμένη συνέλιξη». Τώρα, με όρους πυκνότητας, αν $X = \mathbb{R}^n$ έχουμε

$$d\mu(x) = f(x)dx \Rightarrow d(m_\alpha \# \mu)(x) = \frac{1}{|\alpha|^n} f(\frac{x}{\alpha})dx.$$

Μία από τις πολλές εφαρμογές αυτών των ιδιοτήτων κυρτότητας είναι η μελέτη **οριακών θεωρημάτων** για αθροίσματα τυχαίων μεταβλητών. Ας πάρουμε για παράδειγμα την περίπτωση όπου $(V_i)_{i\in\mathbb{N}}$ είναι μία ακολουθία ανεξάρτητων τυχαίων μεταβλητών στον $\mathbb{R}^n$, με πεπερασμένη διασπορά, οι οποίες έχουν την ίδια κατανομή, και ας θεωρήσουμε την

$$S_m = \frac{V_1 + \cdots + V_m}{\sqrt{m}}.$$

Τότε, η S_{2^k} έχει την ίδια κατανομή με την

$$\frac{S_{2^{k-1}} + \tilde{S}_{2^{k-1}}}{\sqrt{2}},$$

όπου $\tilde{S}$ είναι ένα ανεξάρτητο αντίγραφο της S, δηλαδή μία ανεξάρτητη προς την S τυχαία μεταβλητή η οποία έχει την ίδια κατανομή με την S. Έστω G μία Gaussian τυχαία μεταβλητή που έχει την ίδια μέση τιμή και τον ίδιο πίνακα συνδιακυμάνσεων με την V_1. Είναι γνωστό ότ η G έχει την ίδια κατανομή με την

$$\frac{G + \tilde{G}}{\sqrt{2}},$$

όπου $\tilde{G}$ είναι ξανά ένα ανεξάρτητο αντίγραφο της G. Εφαρμόζοντας την Πρόταση 4.3.2 παίρνουμε

$$T_2(S_{2^k}, G) \leqslant T_2(S_{2^{k-1}}, G).$$

Λόγω αυτής της ιδιότητας μπορούμε να πάρουμε, με λίγη δουλειά, το εξής:

$$W_2(S_m, G) \overset{m\to\infty}{\longrightarrow} 0, \tag{4.3.2}$$

το οποίο είναι το κλασικό **κεντρικό οριακό θεώρημα**.

Παρατήρηση 4.3.4. *Ο υπολογισμός στην* (4.3.2) *φαίνεται να ισχυροποιεί το κεντρικό οριακό θεώρημα, το οποίο ισχυρίζεται μόνο ότι* $\mathrm{law}(S_m) \xrightarrow{w^*} \mathrm{law}(G)$. *Όμως αφού* $\mathbb{E}|S_m|^2 = \mathbb{E}|G|^2$, *για κάθε* m, *η* w^**-σύγκλιση είναι ισοδύναμη, στην περίπτωση αυτή, με την* W_2*-σύγκλιση.*

Αυτήν την απόδειξη του κεντρικού οριακού θεωρήματος μπορεί κανείς να την βρει στο άρθρο των Murata και Tanaka [51]. Η απόδειξη αυτή φαίνεται πιο πολύπλοκη από την κλασική η οποία βασίζεται στον μετασχηματισμό Fourier, αλλά το ενδιαφέρον σε αυτήν είναι ακριβώς το ότι αποφεύγει να χρησιμοποιήσει ανάλυση Fourier.

Κεφάλαιο 5

Γεωμετρία της βέλτιστης μεταφοράς

Σε αυτό το Κεφάλαιο θα ασχοληθούμε με τετραγωνικές συναρτήσεις κόστους στον $\mathbb{R}^n$. Τα αποτελέσματα εδώ είναι πιο απλά, αλλά τα πιο σημαντικά για τις εφαρμογές. Το θεμελιώδες αποτέλεσμα είναι το εξής: αν τα μέτρα πιθανότητας μ και ν έχουν πεπερασμένες ροπές δεύτερης τάξης, τότε ένα σχέδιο μεταφοράς π είναι βέλτιστο αν και μόνο αν ο φορέας του περιέχεται στο υποδιαφορικό μιας κυρτής συνάρτησης.

5.1 Τετραγωνική συνάρτηση κόστους

Έστω $X = Y = \mathbb{R}^n$ και $c(x,y) = \frac{|x-y|^2}{2}$. Η ποσότητα που θέλουμε να ελαχιστοποιήσουμε στο πρόβλημα του Kantorovich παίρνει τώρα την εξής μορφή:

$$I[\pi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2}\, d\pi(x,y). \tag{5.1.1}$$

Υποθέτουμε ότι μ και ν είναι δύο Borel μέτρα πιθανότητας στον $\mathbb{R}^n$ με πεπερασμένες ροπές δεύτερης τάξης. Δηλαδή,

$$M_2 = \int_{\mathbb{R}^n} \frac{|x|^2}{2}\, d\mu(x) + \int_{\mathbb{R}^n} \frac{|y|^2}{2}\, d\nu(y) < \infty. \tag{5.1.2}$$

Η συνθήκη αυτή διασφαλίζει ότι το συναρτησοειδές $I[\pi]$ είναι φραγμένο στο σύνολο $\Pi(\mu,\nu)$. Πράγματι, αν $\pi \in \Pi(\mu,\nu)$, τότε

$$I[\pi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2}\, d\pi(x,y) \leqslant \int_{\mathbb{R}^n \times \mathbb{R}^n} (|x|^2 + |y|^2)\, d\pi(x,y) = 2M_2 < \infty.$$

Υπενθυμίζουμε πως η αρχή δυϊσμού του Kantorovich είναι η εξής:

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi), \tag{5.1.3}$$

όπου Φ_c είναι το σύνολο των ζευγών μετρήσιμων συναρτήσεων $(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu)$ που ικανοποιούν την σχέση $\varphi(x)+\psi(y)\leqslant c(x,y)$ μ-σχεδόν για κάθε $x\in X$ και ν-σχεδόν για κάθε $y\in Y$, και

$$J(\varphi,\psi) = \int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y).$$

Στην περίπτωσή μας, η συνθήκη για να ανήκει ένα ζεύγος (φ,ψ) στο Φ_c είναι

$$\varphi(x)+\psi(y)\leqslant \frac{|x-y|^2}{2} \tag{5.1.4}$$

μ-σχεδόν για κάθε x και ν-σχεδόν για κάθε y.

Ας θυμηθούμε το **αρχικό** πρόβλημα 3.1.1 του Kantorovich. Στο Κεφάλαιο 3 αποδείξαμε την ύπαρξη ελαχίστου για το πρόβλημα $\inf\{I[\pi] : \pi\in\Pi(\mu,\nu)\}$:

Πρόταση 5.1.1 (ύπαρξη βέλτιστου μέτρου)**.** *Έστω μ και ν δύο Borel μέτρα πιθανότητας στον $\mathbb{R}^n$ με πεπερασμένες ροπές δεύτερης τάξης. Το πρόβλημα ελαχιστοποίησης της ποσότητας*

$$I[\pi] = \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{|x-y|^2}{2}\, d\pi(x,y)$$

δέχεται ελάχιστο στο $\Pi(\mu,\nu)$.

Απόδειξη. Θεωρούμε μια ακολουθία (π_k) με την ιδιότητα

$$I[\pi_k] \to \inf_{\pi\in\Pi(\mu,\nu)} I[\pi].$$

Όπως έχουμε δείξει στο προηγούμενο κεφάλαιο, το σύνολο $\Pi(\mu,\nu)$ είναι ένα συμπαγές σύνολο για την w^*-τοπολογία των μέτρων πιθανότητας, συνεπώς η ακολουθία π_k δέχεται σημείο συσσώρευσης $\pi_*\in\Pi(\mu,\nu)$. Τώρα, αφού η τετραγωνική συνάρτηση κόστους c είναι συνεχής (άρα και κάτω ημισυνεχής) μπορούμε να την γράψουμε ως το supremum μιας αύξουσας ακολουθίας (c_ℓ) από φραγμένες συνεχείς συναρτήσεις. Εφαρμόζοντας το θεώρημα μονότονης σύγκλισης, χρησιμοποιώντας το γεγονός ότι το π_* είναι σημείο συσώρευσης, την ανισότητα $c_\ell\leqslant c$ και το ότι η ακολουθία $I[\pi_k]$ συγκλίνει στο infimum του I, παίρνουμε το εξής:

$$\begin{aligned}\int c(x,y)d\pi_*(x,y) &= \lim_{\ell\to\infty}\int c_\ell(x,y)d\pi_*(x,y)\\ &\leqslant \lim_{\ell\to\infty}\limsup_{k\to\infty}\int c_\ell(x,y)d\pi_k(x,y)\\ &\leqslant \limsup_{k\to\infty}\int c(x,y)d\pi_k(x,y) = \inf I.\end{aligned}$$

Συνεπώς το π_* ελαχιστοποιεί το I. □

Ας ασχοληθούμε τώρα με το **δυϊκό** πρόβλημα. Χρησιμοποιώντας την ειδική μορφή της τετραγωνικής συνάρτησης κόστους, μπορούμε να γράψουμε την (5.1.4) στην εξής μορφή:

$$\langle x,y\rangle \leqslant \left[\frac{|x|^2}{2}-\varphi(x)\right]+\left[\frac{|y|^2}{2}-\psi(y)\right]. \tag{5.1.5}$$

Εισάγοντας τις νέες συναρτήσεις

$$\tilde{\varphi}(x)=\frac{|x|^2}{2}-\varphi(x),\ \tilde{\psi}(y)=\frac{|y|^2}{2}-\psi(y)$$

και χρησιμοποιώντας την σχέση (5.1.2), έχουμε

$$\begin{aligned}\inf_{\Pi(\mu,\nu)} I[\pi] &= \inf_{\Pi(\mu,\nu)}\int\frac{|x-y|^2}{2}d\pi = \inf_{\Pi(\mu,\nu)}\left[\int\Big(\frac{|x|^2}{2}+\frac{|y|^2}{2}-\langle x,y\rangle\Big)d\pi\right]\\ &= \inf_{\Pi(\mu,\nu)}\left[\int\frac{|x|^2}{2}d\mu+\int\frac{|y|^2}{2}d\nu-\int\langle x,y\rangle d\pi\right]\\ &= M_2-\sup_{\Pi(\mu,\nu)}\int\langle x,y\rangle d\pi\end{aligned} \tag{5.1.6}$$

και

(5.1.7)

$$\begin{aligned}\sup_{\Phi_c} J(\varphi,\psi) &= \sup_{\Phi_c}\left[\int\frac{|x|^2}{2}d\mu+\int\frac{|y|^2}{2}d\nu-\Big(\int\frac{|x|^2}{2}d\mu-\int\varphi d\mu+\int\frac{|y|^2}{2}d\nu-\int\psi d\nu\Big)\right]\\ &= M_2-\inf_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}}\left[\int\tilde{\varphi}d\mu+\int\tilde{\psi}d\nu\right],\end{aligned}$$

όπου $\tilde{\Phi}$ είναι το σύνολο όλων των ζευγαριών $(\tilde{\varphi},\tilde{\psi})\in L^1(d\mu)\times L^1(d\nu)$ με τιμές στο $\mathbb{R}\cup\{+\infty\}$ και με την ιδιότητα ότι σχεδόν για κάθε x,y ισχύει

$$\langle x,y\rangle \leqslant \tilde{\varphi}(x)+\tilde{\psi}(y). \tag{5.1.8}$$

Συνεπώς η αρχή δυϊσμού του Kantorovich (5.1.3) δίνει

$$\sup_{\Pi(\mu,\nu)} \int \langle x, y\rangle d\pi(x,y) = \inf_{\tilde{\Phi}} J(\varphi,\psi), \tag{5.1.9}$$

όπου πλέον συμβολίζουμε για απλότητα με (φ,ψ) (αντί για $(\tilde{\varphi},\tilde{\psi})$) ένα ζεύγος συναρτήσεων της κλάσης $\tilde{\Phi}$.

Θα προσπαθήσουμε τώρα να βελτιώσουμε τα αποδεκτά ζευγάρια (φ,ψ) μέσω ενός τεχνάσματος το οποίο έχουμε χρησιμοποιήσει ξανά στο Κεφάλαιο 3. Μπορούμε να επιτύχουμε η σχέση (5.1.8) να ισχύει για κάθε x και για κάθε y. Πράγματι, έστω $(\varphi,\psi)\in\tilde{\Phi}$. Τότε, υπάρχουν μετρήσιμα σύνολα N_x, N_y με $\mu(N_x)=0, \nu(N_y)=0$ ώστε η ανισότητα (5.1.8) να ισχύει για κάθε $(x,y)\in N_x^c\times N_y^c$. Τροποποιούμε τις τιμές των φ,ψ θέτοντας $\varphi=+\infty$ στο N_x και $\psi=+\infty$ στο N_y. Το νέο ζευγάρι συναρτήσεων που προκύπτει, συνεχίζει να ανήκει στο $\tilde{\Phi}$, και η τιμή της ποσότητας $J(\varphi,\psi)$ δεν αλλάζει, αφού οι αλλαγές επηρεάζουν μόνο σύνολα μηδενικού μέτρου. Θέτουμε

$$\left\{\begin{array}{l} \varphi^*(y) = \sup_x[\langle x,y\rangle - \varphi(x)] \\ \varphi^{**}(x) = \sup_y[\langle x,y\rangle - \varphi^*(y)], \end{array}\right\} \tag{5.1.10}$$

και παρατηρούμε ότι $\varphi^*(y)\leqslant\psi(y)$, άρα $J(\varphi,\psi)\geqslant J(\varphi,\varphi^*)$ και $\varphi^{**}(x)\leqslant\varphi(x)$, άρα $J(\varphi,\varphi^*)\geqslant J(\varphi^{**},\varphi^*)$. Συνεπώς, ισχύει ότι

$$\inf_{\tilde{\Phi}} J(\varphi,\psi) \geqslant \inf_{\varphi\in L^1(d\mu)} J(\varphi^{**},\varphi^*). \tag{5.1.11}$$

Παρατηρούμε λοιπόν ότι το infimum της J πάνω στο $\tilde{\Phi}$ δεν αλλάζει αν το περιορίσουμε σε ένα υποσύνολο του $\tilde{\Phi}$ που αποτελείται από τα ζευγάρια της μορφής (φ^{**},φ^*), όπου οι φ^{**},φ^* είναι κυρτές κάτω ημισυνεχείς συναρτήσεις, αφού καθεμία έχει οριστεί ως το supremum μιας οικογένειας γραμμικών συναρτήσεων.

5.2 Κριτήριο Knott-Smith και θεώρημα Brenier

Σκοπός μας είναι να δείξουμε την ύπαρξη ενός βέλτιστου ζεύγους συζυγών κυρτών συναρτήσεων για το δυϊκό πρόβλημα Kantorovich. Συγκεκριμένα, θα αποδείξουμε το εξής:

Θεώρημα 5.2.1 (ύπαρξη βέλτιστου ζεύγους συζυγών κυρτών συναρτήσεων)**.** *Έστω μ,ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$ με πεπερασμένες ροπές δεύτερης τάξης. Έστω $\tilde{\Phi}$ το σύνολο που ορίσαμε στην* (5.1.8). *Τότε, υπάρχει ένα ζευγάρι (φ,φ^*) από κάτω ημισυνεχείς, συζυγείς, γνήσιες κυρτές συναρτήσεις στον $\mathbb{R}^n$ ώστε*

$$\inf_{\tilde{\Phi}} J = J(\varphi,\varphi^*).$$

Θα δώσουμε δύο αποδείξεις του Θεωρήματος 5.2.1. Η πρώτη είναι πιο απλή, αλλά περιορίζεται στην περίπτωση όπου τα μ, ν έχουν φορείς φραγμένα υποσύνολα του $\mathbb{R}^n$. Η δεύτερη (βλέπε [55]) είναι σε πιο γενικό πλαίσιο, αφού τα μέτρα μ, ν μπορούν να έχουν φορέα ολόκληρο τον $\mathbb{R}^n$, και ο $\mathbb{R}^n$ μπορεί να αντικατασταθεί από τυχόντα χώρο Banach.

Θα χρειαστούμε το ακόλουθο Λήμμα.

Λήμμα 5.2.2 (διπλή κυρτοποίηση)**.** *Έστω μ, ν δύο μέτρα πιθανότητας με φορείς τα υποσύνολα X, Y του $\mathbb{R}^n$ αντίστοιχα, που ικανοποιούν το εξής:*

$$M_2 \equiv \int_X \frac{|x|^2}{2} d\mu(x) + \int_Y \frac{|y|^2}{2} d\nu(y) < +\infty.$$

Αν φ, ψ είναι μετρήσιμες συναρτήσεις με τιμές στο $\mathbb{R} \cup \{+\infty\}$, τότε ορίζουμε

$$\varphi^*(y) = \sup_{x \in X} [\langle x, y \rangle - \varphi(x)], \tag{5.2.1}$$

$$\psi^*(x) = \sup_{y \in Y} [\langle x, y \rangle - \psi(y)]. \tag{5.2.2}$$

Έστω $\tilde{\Phi}$ το σύνολο που ορίστηκε στην (5.1.8) και έστω $(\varphi_k, \psi_k)_{k \in \mathbb{N}}$ μια ακολουθία που ελαχιστοποιεί το J πάνω στο σύνολο $\tilde{\Phi}$, δηλαδή $J(\varphi_k, \psi_k) \to \inf J$. Τότε:

(i) *Τροποποιώντας τις τιμές των φ_k, ψ_k σε σύνολα μηδενικού μέτρου ως προς τα μ και ν αντίστοιχα, μπορούμε να επιτύχουμε η σχέση (5.1.8) να ισχύει για κάθε $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ χωρίς να αλλάξουν οι τιμές των $J(\varphi_k, \psi_k)$.*

(ii) *Υπάρχει μια ακολουθία πραγματικών αριθμών $(\alpha_k)_{k \in \mathbb{N}}$ ώστε η ακολουθία*

$$(\bar{\varphi}_k, \bar{\psi}_k) = (\varphi_k^{**} - \alpha_k, \varphi_k^* + \alpha_k)$$

να συνεχίζει να ελαχιστοποιεί το J πάνω στο σύνολο $\tilde{\Phi}$ και να ικανοποιεί τα κάτω φράγματα

$$\bar{\varphi}_k(x) \geqslant -\frac{|x|^2}{2}, \ \bar{\psi}_k(y) \geqslant -\frac{|y|^2}{2} \ \text{ για κάθε } x \in X \text{ και } y \in Y, \tag{5.2.3}$$

καθώς και τα άνω φράγματα

$$\liminf_{k \to \infty} \inf_{x \in X} \left(\bar{\varphi}_k(x) + \frac{|x|^2}{2} \right) \leqslant \inf_{\tilde{\Phi}} J + M_2 \tag{5.2.4}$$

$$\liminf_{k \to \infty} \inf_{y \in Y} \left(\bar{\psi}_k(y) + \frac{|y|^2}{2} \right) \leqslant \inf_{\tilde{\Phi}} J + M_2. \tag{5.2.5}$$

(iii) *Ειδικότερα, αν $X = Y = \mathbb{R}^n$ τότε ο τελεστής $*$ συμπίπτει με τον μετασχηματισμό Legendre, και*

$$\inf_{\tilde{\Phi}} J = \inf_{\varphi \in L^1(d\mu)} J(\varphi^{**}, \varphi^*).$$

Συνεπώς, το infimum του J πάνω στο $\tilde{\Phi}$ δεν αλλάζει αν το περιορίσουμε στο μικρότερο υποσύνολο του $\tilde{\Phi}$ το οποίο αποτελείται από ζεύγη συζυγών γνησίων κυρτών συναρτήσεων.

Απόδειξη. Η απόδειξη για το (i) έχει δοθεί στην Ενότητα 4.1, ενώ το (iii) είναι άμεση συνέπεια του (ii). Οπότε, αρκεί να αποδείξουμε το (ii).

1. Έστω (φ_k, ψ_k) μια ακολουθία που ελαχιστοποιεί το J. Από το (i) μπορούμε να υποθέσουμε ότι η σχέση $\langle x, y\rangle \leqslant \varphi_k(x) + \psi_k(y)$ ισχύει για κάθε x και y. Τώρα, αφού η ψ_k δεν είναι ταυτοτικά $+\infty$, υπάρχει $y_0 \in Y$ ώστε $\psi_k(y_0) < +\infty$. Θέτοντας $b_0 = -\psi_k(y_0)$ έχουμε (για κάθε x) $\varphi_k(x) \geqslant \langle x, y_0\rangle + b_0$, δηλαδή η φ_k είναι φραγμένη από κάτω από μια αφφινική συνάρτηση. Έπεται ότι

$$\varphi_k^*(y_0) = \sup_{x \in X} [\langle x, y_0\rangle - \varphi_k(x)] \leqslant -b_0.$$

Ειδικότερα, η φ_k^* δεν είναι ταυτοτικά $+\infty$. Όμοια, η φ_k δεν είναι ταυτοτικά $+\infty$. Άρα, υπάρχει $x_0 \in X$ ώστε $\varphi_k(x_0) < +\infty$, άρα $\varphi_k^*(y) \geqslant \langle x_0, y\rangle + c_0$, όπου $c_0 := -\varphi_k(x_0)$, δηλαδή η φ_k^* είναι φραγμένη από κάτω από μία αφφινική συνάρτηση. Από αυτό έπεται ότι αν θέσουμε

$$\alpha_k := \inf_{y \in Y} \left(\varphi_k^*(y) + \frac{|y|^2}{2} \right)$$

τότε $\alpha_k > -\infty$. Επιλέγουμε

$$(\bar{\varphi}_k, \bar{\psi}_k) = (\varphi_k^{**} + \alpha_k, \varphi_k^* - \alpha_k)$$

και παρατηρούμε ότι $\bar{\varphi}_k = (\bar{\psi}_k)^*$. Πράγματι, έχουμε

$$\begin{aligned}(\bar{\psi}_k)^*(x) &= (\varphi_k^* - \alpha_k)^*(x) = \sup_y [\langle x, y\rangle - (\varphi_k^*(y) - \alpha_k)] \\ &= \sup_y [\langle x, y\rangle - \varphi_k^*(y)] + \alpha_k = \varphi_k^{**}(x) + \alpha_k = \bar{\varphi}_k(x).\end{aligned}$$

Από τον ορισμό του α_k και της $\bar{\psi}_k$ προκύπτει ότι

$$\inf_{y \in Y} \left(\bar{\psi}_k(y) + \frac{|y|^2}{2} \right) = 0.$$

Επιπλέον, έχουμε

$$\begin{aligned}\bar{\varphi}_k(x) + \frac{|x|^2}{2} &= (\bar{\psi}_k)^*(x) + \frac{|x|^2}{2} = \sup_{y\in Y}\left(\langle x,y\rangle - \bar{\psi}_k(y) + \frac{|x|^2}{2}\right)\\ &\geqslant \sup_{y\in Y}\left(-\frac{|y|^2}{2} - \bar{\psi}_k(y)\right) = -\inf_{y\in Y}\left(\frac{|y|^2}{2} + \bar{\psi}_k(y)\right) = 0.\end{aligned}$$

Συνεπώς, το ζευγάρι $(\bar{\varphi}_k, \bar{\psi}_k)$ ικανοποιεί την (5.2.3).

2. Στην συνέχεια ελέγχουμε την ολοκληρωσιμότητα των $\bar{\varphi}_k$ και $\bar{\psi}_k$. Παρατηρούμε ότι λόγω της $J(\varphi+s,\psi-s) = J(\varphi,\psi)$ αλλά και του συμπεράσματος της Ενότητας 4.1, έχουμε

$$J(\bar{\varphi}_k, \bar{\psi}_k) = J(\varphi_k^{**}, \varphi_k^*) \leqslant J(\varphi_k, \psi_k) < +\infty. \tag{5.2.6}$$

Ειδικότερα, αφού τα μ, ν είναι μέτρα πιθανότητας με πεπερασμένες δεύτερες ροπές, έχουμε

$$\int_X \left(\frac{|x|^2}{2} + \bar{\varphi}_k(x)\right) d\mu(x) + \int_Y \left(\frac{|y|^2}{2} + \bar{\psi}_k(y)\right) d\nu(y) < +\infty.$$

Αφού οι δύο ποσότητες μέσα στα ολοκληρώματα είναι μη αρνητικές, έπεται ότι είναι ολοκληρώσιμες, και λόγω των υποθέσεων για τα μέτρα μ, ν συμπεραίνουμε ότι $\bar{\varphi}_k \in L^1(d\mu)$ και $\bar{\psi}_k \in L^1(d\nu)$. Επιπλέον,

$$\bar{\varphi}_k(x) + \bar{\psi}_k(y) = \varphi_k^{**}(x) + \varphi_k^*(y) = \sup_z[\langle x,z\rangle - \varphi_k^*(z)] + \varphi_k^*(y) \geqslant \langle x,y\rangle,$$

άρα $(\bar{\varphi}_k, \bar{\psi}_k) \in \tilde{\Phi}$, και λόγω της ανισότητας (5.2.6) έχουμε ότι η ακολουθία $(\bar{\varphi}_k, \bar{\psi}_k)$ συνεχίζει να ελαχιστοποιεί το J.

3. Το μόνο που μένει είναι να αποδείξουμε την (5.2.4) και την (5.2.5). Παρατηρούμε ότι

$$\begin{aligned}J(\varphi_k,\psi_k) + M_2 &= \int_X \left(\varphi_k(x) + \frac{|x|^2}{2}\right) d\mu(x) + \int_Y \left(\psi_k(y) + \frac{|y|^2}{2}\right) d\nu(y)\\ &\geqslant \inf_{x\in X}\left(\varphi_k(x) + \frac{|x|^2}{2}\right)\mu(X) + \inf_{y\in Y}\left(\psi_k(y) + \frac{|y|^2}{2}\right)\nu(Y)\\ &= \inf_{x\in X}\left(\varphi_k(x) + \frac{|x|^2}{2}\right) + \inf_{y\in Y}\left(\psi_k(y) + \frac{|y|^2}{2}\right).\end{aligned}$$

Αφού και οι δύο εκφράσεις μέσα στις παρενθέσεις είναι μη αρνητικές, έπεται ότι

$$\liminf_{k\to\infty}\inf_{x\in X}\left(\varphi_k(x) + \frac{|x|^2}{2}\right) \leqslant \liminf_{k\to\infty} J(\varphi_k,\psi_k) + M_2 = \inf_{\tilde{\Phi}} J + M_2.$$

Όμοια έπεται η (5.2.5). □

Πρώτη απόδειξη του Θεωρήματος 5.2.1. Υποθέτουμε ότι τα μ, ν έχουν φορείς τα συμπαγή υποσύνολα X, Y του $\mathbb{R}^n$ αντίστοιχα. Έστω (φ_k, ψ_k) μια ακολουθία που ελαχιστοποιεί το J πάνω στο $\tilde{\Phi}$. Έστω $(\bar{\varphi}_k, \bar{\psi}_k)$ η ελαχιστοποιούσα ακολουθία η οποία κατασκευάστηκε στο Λήμμα 5.2.2. Από την σχέση

$$\bar{\psi}_k(y) = \sup_{x \in X} [\langle x, y\rangle - \varphi_k(x) - \alpha_k]$$

έπεται ότι

$$\|\bar{\psi}_k\|_{\mathrm{Lip}(Y)} \leqslant \sup_{x \in X} |x| \text{ και } \|\bar{\varphi}_k\|_{\mathrm{Lip}(X)} \leqslant \sup_{y \in Y} |y|.$$

Συνεπώς, οι $\bar{\varphi}_k, \bar{\psi}_k$ είναι ομοιόμορφα Lipschitz. Από την άλλη πλευρά, από το Λήμμα 5.2.2 γνωρίζουμε ότι, για k αρκετά μεγάλο, υπάρχουν $x_k \in X, y_k \in Y$ τέτοια ώστε

$$-\sup_{x \in X} \frac{|x|^2}{2} \leqslant \bar{\varphi}_k(x_k) \leqslant \sup_{x \in X} \frac{|x|^2}{2} + \inf J + M_2 + 1$$

και

$$-\sup_{y \in Y} \frac{|y|^2}{2} \leqslant \bar{\psi}_k(y_k) \leqslant \sup_{y \in Y} \frac{|y|^2}{2} + \inf J + M_2 + 1.$$

Συνδυάζοντας τα παραπάνω με την συμπάγεια των X, Y, βλέπουμε ότι οι $\bar{\varphi}_k, \bar{\psi}_k$ είναι ομοιόμορφα φραγμένες. Συνεπώς, από το θεώρημα Ascoli, υπάρχει υπακολουθία που συγκλίνει ομοιόμορφα (στους $C_b(X), C_b(Y)$) σε κάποιες συνεχείς συναρτήσεις $\bar{\varphi}, \bar{\psi}$ αντίστοιχα. Από την ομοιόμορφη σύγκλιση έπεται ότι $J(\bar{\varphi}, \bar{\psi}) = \lim_{k\to\infty} J(\bar{\varphi}_k, \bar{\psi}_k)$, άρα το $(\bar{\varphi}, \bar{\psi})$ είναι ένα βέλτιστο ζευγάρι. Μπορούμε τώρα να επεκτείνουμε τις $\bar{\varphi}, \bar{\psi}$ ώστε να παίρνουν την τιμή $+\infty$ έξω από τα X, Y, και τέλος κάνοντας διπλή κυρτοποίηση σε αυτές (επιλέγοντας π.χ $X = Y = \mathbb{R}^n$) βρίσκουμε ένα ζευγάρι από κυρτές συζυγείς συναρτήσεις που έχουν τις επιθυμητές ιδιότητες.

Δεύτερη απόδειξη του Θεωρήματος 5.2.1.

Βήμα 1. Αντικαθιστούμε την συνάρτηση $\langle x, y\rangle$ με μια μη αρνητική συνάρτηση: για παράδειγμα,

$$(\varphi, \psi) \in \tilde{\Phi} \iff \left(\varphi(x) + \frac{|x|^2}{2}\right) + \left(\psi(y) + \frac{|y|^2}{2}\right) \geqslant \frac{|x|^2}{2} + \frac{|y|^2}{2} + \langle x, y\rangle = \frac{|x+y|^2}{2}.$$

Έστω $(\varphi_k, \psi_k)_{k\in\mathbb{N}}$ μία ελαχιστοποιούσα ακολουθία για το J. Από το Λήμμα 5.2.2 μπορούμε να υποθέσουμε ότι

$$0 \leqslant \varphi_k(x) + \frac{|x|^2}{2}, \ 0 \leqslant \psi_k(y) + \frac{|y|^2}{2},$$

και

$$\left(\varphi_k(x) + \frac{|x|^2}{2}\right) + \left(\psi_k(y) + \frac{|y|^2}{2}\right) \geqslant \frac{|x+y|^2}{2} \geqslant 0.$$

Βήμα 2. Για κάθε $\ell \in \mathbb{N}$ ορίζουμε

$$\varphi_k^{(\ell)}(x) + \frac{|x|^2}{2} = \min\Big(\varphi_k(x) + \frac{|x|^2}{2}, \ell\Big)$$

και

$$\psi_k^{(\ell)}(y) + \frac{|y|^2}{2} = \min\Big(\psi_k(y) + \frac{|y|^2}{2}, \ell\Big).$$

Εύκολα ελέγχουμε τις εξής ιδιότητες:

$$\left\{ \begin{array}{l} 0 \leqslant \varphi_k^{(\ell)}(x) + \frac{|x|^2}{2} \leqslant \ell, \\ 0 \leqslant \psi_k^{(\ell)}(y) + \frac{|y|^2}{2} \leqslant \ell \end{array} \right\} \tag{5.2.7}$$

$$\left\{ \begin{array}{l} \varphi_k^{(1)} \leqslant \varphi_k^{(2)} \leqslant \cdots \leqslant \varphi_k^{(\ell)} \leqslant \cdots \\ \psi_k^{(1)} \leqslant \psi_k^{(2)} \leqslant \cdots \leqslant \psi_k^{(\ell)} \leqslant \cdots \end{array} \right\} \tag{5.2.8}$$

$$J(\varphi_k^{(\ell)}, \psi_k^{(\ell)}) \leqslant J(\varphi_k, \psi_k), \tag{5.2.9}$$

$$\varphi_k^{(\ell)}(x) + \psi_k^{(\ell)}(y) \geqslant \min\left(\frac{|x+y|^2}{2}, \ell\right) - \Big(\frac{|x|^2}{2} + \frac{|y|^2}{2}\Big). \tag{5.2.10}$$

Βήμα 3. Από την (5.2.7) έχουμε ότι, για κάθε ℓ,

$$\varphi_k^{(\ell)}(x) = -\frac{|x|^2}{2} + O(\ell)$$

και αφού η $-\frac{|x|^2}{2}$ είναι μία συγκεκριμένη συνάρτηση στον $L^1(d\mu)$, έπεται ότι η $(\varphi_k^{(\ell)})_{k\in\mathbb{N}}$ έχει ασθενώς συγκλίνουσα υπακολουθία, έστω $\varphi_{k_n}^{(\ell)}$, στον $L^1(d\mu)$. Δηλαδή έχουμε,

$$\varphi_{k_n}^{(\ell)} \to \varphi^{(\ell)}$$

ασθενώς στον $L^1(d\mu)$ καθώς $k \to \infty$. Όμοια, για κάθε ℓ υπάρχει υπακολουθία $\psi_{k_n}^{(\ell)}$ της $(\psi_k^{(\ell)})_{k\in\mathbb{N}}$ η οποία συγκλίνει ασθενώς στον $L^1(d\nu)$ σε κάποια $\psi^{(\ell)} \in L^1(d\nu)$. Με ένα διαγώνιο επιχείρημα μπορούμε να βρούμε μία ακολουθία $(k_n) \in \mathbb{N}$ για την οποία η σύγκλιση να ισχύει για κάθε ℓ. Τότε, αφού η ασθενής σύγκλιση διατηρεί την διάταξη έπεται από τις (5.2.8), (5.2.9), (5.2.10) ότι,

$$\left\{ \begin{array}{l} \varphi^{(1)} \leqslant \varphi^{(2)} \leqslant \cdots \leqslant \varphi^{(\ell)} \leqslant \cdots, \\ \psi^{(1)} \leqslant \psi^{(2)} \leqslant \cdots \leqslant \psi^{(\ell)} \leqslant \cdots \end{array} \right\} \tag{5.2.11}$$

$$J(\varphi^{(\ell)},\psi^{(\ell)})=\lim_{n\to\infty}J(\varphi^{(\ell)}_{k_n},\psi^{(\ell)}_{k_n})\leqslant\liminf_{n\to\infty}J(\varphi_{k_n},\psi_{k_n})=\inf_{\tilde{\Phi}}J, \tag{5.2.12}$$

$$\varphi^{(\ell)}(x)+\psi^{(\ell)}(y)\geqslant\min\left(\frac{|x+y|^2}{2},\ell\right)-\Big(\frac{|x|^2}{2}+\frac{|y|^2}{2}\Big). \tag{5.2.13}$$

Βήμα 4. Οι ακολουθίες $(\varphi^{(\ell)}),(\psi^{(\ell)})$ είναι φραγμένες στον L^1, αύξουσες και φραγμένες από κάτω από μία συγκεκριμένη συνάρτηση στον L^1. Συνεπώς, μπορούμε να εφαρμόσουμε το θεώρημα μονότονης σύγκλισης από το οποίο έπεται ότι υπάρχουν $\varphi,\psi\in L^1$, ορισμένες σχεδόν παντού, ώστε

$$\varphi=\lim_{\ell\to\infty}\varphi^{(\ell)},\ \ \psi=\lim_{\ell\to\infty}\psi^{(\ell)},$$

και οι οποίες ικανοποιούν την

$$J(\varphi,\psi)=\lim_{\ell\to\infty}J(\varphi^{(\ell)},\psi^{(\ell)})\leqslant\inf_{\tilde{\Phi}}J.$$

Τώρα, παίρνοντας όριο καθως $\ell\to\infty$ στην (5.2.13) βλέπουμε ότι $(\varphi,\psi)\in\tilde{\Phi}$, συνεπώς είναι ένα βέλτιστο ζευγάρι.

Βήμα 5. Κάνοντας διπλή κυρτοποίηση στο ζευγάρι (φ,ψ) παίρνουμε ένα βέλτιστο ζευγάρι από συζυγείς, γνήσιες κυρτές συναρτήσεις. Η απόδειξη είναι πλήρης. □

Θεώρημα 5.2.3 (θεώρημα βέλτιστης μεταφοράς για τετραγωνικό κόστος)**.** *Έστω μ,ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, με πεπερασμένες ροπές δεύτερης τάξης. Θεωρούμε το πρόβλημα βέλτιστης μεταφοράς των Monge-Kantorovich με την τετραγωνική συνάρτηση κόστους $c(x,y)=|x-y|^2$.*

(i) (***Κριτήριο Knott-Smith***) *Ένα μέτρο $\pi\in\Pi(\mu,\nu)$ είναι βέλτιστο αν και μόνο αν υπάρχει κυρτή, κάτω ημισυνεχής συνάρτηση φ ώστε*

$$\mathrm{Supp}(\pi)\subseteq\mathrm{Graph}(\partial\varphi), \tag{5.2.14}$$

ή ισοδύναμα,

$$\pi-\textit{σχεδόν για κάθε }(x,y),\ y\in\partial\varphi(x). \tag{5.2.15}$$

Επιπλέον, σε αυτήν την περίπτωση το ζευγάρι (φ,φ^*) πρέπει να ελαχιστοποιεί το πρόβλημα,

$$\inf\left\{\int_{\mathbb{R}^n}\varphi d\mu+\int_{\mathbb{R}^n}\psi d\nu:\forall(x,y),\ \langle x,y\rangle\leqslant\varphi(x)+\psi(y)\right\}.$$

(ii) (***Θεώρημα Brenier***) *Αν το μ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue,$\mu\ll\lambda$, τότε υπάρχει μοναδικό βέλτιστο μέτρο π, το οποίο προσδιορίζεται από την*

$$d\pi(x,y)=d\mu(x)\delta[y=\nabla\varphi(x)], \tag{5.2.16}$$

ή ισοδύναμα

$$\pi = (Id \times \nabla\varphi)\#\mu, \tag{5.2.17}$$

όπου $\nabla\varphi$ *είναι η* μ*-σχεδόν παντού μονοσήμαντα ορισμένη κλίση μίας κυρτής συνάρτησης που προωθεί το* μ *στο* ν*:* $\nabla\varphi\#\mu = \nu$*. Επιπλέον,*

$$\mathrm{Supp}(\nu) = \overline{\nabla\varphi(\mathrm{Supp}(\mu))}.$$

(iii) *Ως πόρισμα έχουμε ότι, κάτω από τις υποθέσεις του* (ii), *η* $\nabla\varphi$ *είναι η μοναδική λύση του προβλήματος μεταφοράς του Monge:*

$$\int_{\mathbb{R}^n} |x - \nabla\varphi(x)|^2 d\mu(x) = \inf_{T\#\mu=\nu} \int_{\mathbb{R}^n} |x - T(x)|^2 d\mu(x),$$

ή ισοδύναμα,

$$\int_{\mathbb{R}^n} \langle x, \nabla\varphi(x)\rangle d\mu(x) = \sup_{T\#\mu=\nu} \int_{\mathbb{R}^n} \langle x, T(x)\rangle d\mu(x).$$

Τέλος, αν και το ν *είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue,* $\nu \ll \lambda$*, τότε* μ*-σχεδόν για κάθε* x *και* ν*-σχεδόν για κάθε* y *έχουμε*

$$\nabla\varphi^* \circ \nabla\varphi(x) = x,\ \nabla\varphi \circ \nabla\varphi^*(y) = y,$$

και $\nabla\varphi^*$ *είναι η* ν*-σχεδόν παντού μονοσήμαντα ορισμένη ρκλίση μιας κυρτής συνάρτησης, η οποία προωθεί το* ν *στο* μ *και επίσης είναι η μοναδική λύση του προβλήματος Monge για την μεταφορά του* ν *στο* μ *με τετραγωνική συνάρτηση κόστους.*

Παρατήρηση 5.2.4. Όταν οι υποθέσεις της (ii) ισχύουν, θα αναφερόμαστε στην μοναδικά ορισμένη μ-σχεδόν παντου απεικόνιση $\nabla\varphi$ ως την *απεικόνιση Brenier* η οποία προωθεί το μέτρο μ στο ν.

Απόδειξη. (του Θεωρήματος 5.2.3) Χρησιμοποιώντας την ειδική (τετραγωνική) μορφή της συνάρτησης κόστους μετατρέπουμε το πρόβλημα ελαχιστοποίησης των Monge-Kantorovich στην μορφή που δίνεται από τις (5.1.6)–(5.1.8).

(i) Από την πρόταση 5.1.1 υπάρχει βέλτιστο σχέδιο μεταφοράς π. Από το Θεώρημα 5.2.1 υπάρχει ζεύγος κυρτών συζυγών συναρτήσεων (φ, φ^*), το οποίο είναι βέλτιστο για το δυϊκό πρόβλημα. Τότε, από τον δυϊσμό του Kantorovich (την (5.1.9)), και από τον ορισμό του $\Pi(\mu,\nu)$, έχουμε

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n} (\langle x, y\rangle) d\pi(x,y) &= \int_{\mathbb{R}^n} \varphi d\mu + \int_{\mathbb{R}^n} \varphi^* d\nu \\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} [\varphi(x) + \varphi^*(y)]\, d\pi(x,y).\end{aligned}$$

Ισοδύναμα,

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} [\varphi(x)+\varphi^*(y)-\langle x,y\rangle]\, d\pi(x,y)=0.$$

Όμως, η ποσότητα μέσα στην παρένθεση είναι μη αρνητική από τον ορισμό της φ^* (πιο συγκεκριμένα, την (2.3.4)), συνεπώς μηδενίζεται π-σχεδόν για κάθε (x,y) και άρα από την Πρόταση 2.3.4 έπεται η (5.2.15).

(ii) Αντίστροφα, έστω $\pi\in\Pi(\mu,\nu)$ που ικανοποιεί την (5.2.15). Τότε, με το ίδιο επιχείρημα βλέπουμε ότι

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\, d\pi(x,y)=\int_{\mathbb{R}^n}\varphi d\mu+\int_{\mathbb{R}^n}\varphi^* d\nu.$$

Άρα, το μέτρο π είναι βέλτιστο για το αριστερό μέλος της (5.1.9) και το ζεύγος (φ,φ^*) είναι βέλτιστο για το δεξιό. Η απόδειξη του (i) είναι πλήρης.

(iii) Υποθέτουμε τώρα ότι το μ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue, $\mu\ll\lambda$, και έστω φ όπως πριν. Αφού η $\varphi\in L^1(d\mu)$ έπεται ότι είναι μ-σχεδόν παντού πεπερασμένη, άρα $\mu(\mathrm{Dom}(\varphi))=1$. Από την άλλη πλευρά, το $\mathrm{Dom}(\varphi)$ είναι κυρτό σύνολο, άρα $\lambda(\partial\mathrm{Dom}(\varphi))=0$, συνεπώς $\mu(\partial\mathrm{Dom}(\varphi))=0$, απ' όπου έπεται ότι $\mu(\mathrm{Int}(\mathrm{Dom}(\varphi)))=1$. Όμως, στο $\mathrm{Int}(\mathrm{Dom}(\varphi))$, το σύνολο όπου η φ δεν είναι διαφορίσιμη έχει μέτρο Lebesgue μηδέν (θεώρημα Rademacher), άρα είναι και μ-μηδενικό. Συνεπώς, μ-σχεδόν για κάθε $x\in X$ η φ είναι διαφορίσιμη, άρα μ-σχεδόν για κάθε x, $\partial\varphi(x)=\{\nabla\varphi(x)\}$. Γνωρίζοντας όμως ότι όταν κάτι ισχύει μ-σχεδόν για κάθε x ισχύει και π-σχεδόν για κάθε (x,y), παίρνουμε ότι $y=\nabla\varphi(x)$ π-σχεδόν για κάθε (x,y).

(iv) Έως τώρα έχουμε δείξει ότι κάθε βέλτιστο σχέδιο μεταφοράς έχει την μορφή $(Id\times\nabla\varphi)\#\mu$ για κάποια κυρτή συνάρτηση φ τέτοια ώστε $\nabla\varphi\#\mu=\nu$ και ότι υπάρχει τουλάχιστον ένα τέτοιο σχέδιο μεταφοράς. Τώρα θα δείξουμε την μοναδικότητα. Έστω $\bar\varphi$ μια άλλη κυρτή συνάρτηση τέτοια ώστε $\nabla\bar\varphi\#\mu=\nu$. Θέλουμε να δείξουμε ότι $\nabla\varphi=\nabla\bar\varphi$ μ-σχεδόν παντού. Από το (i) έχουμε ότι το $(Id\times\nabla\bar\varphi)\#\mu$ είναι ένα βέλτιστο σχέδιο μεταφοράς και το ζεύγος $(\bar\varphi,\bar\varphi^*)$ είναι ένα βέλτιστο ζεύγος για το δυϊκό πρόβλημα, όπως το (φ,φ^*). Συνεπώς,

$$\int_{\mathbb{R}^n}\bar\varphi d\mu+\int_{\mathbb{R}^n}\bar\varphi^* d\nu=\int_{\mathbb{R}^n}\varphi d\mu+\int_{\mathbb{R}^n}\varphi^* d\nu. \tag{5.2.18}$$

Έστω π το βέλτιστο σχέδιο μεταφοράς που σχετίζεται με την φ. Από την (5.2.18) έχουμε

$$\begin{aligned}\int_{\mathbb{R}^n\times\mathbb{R}^n}[\bar\varphi(x)+\bar\varphi^*(y)]\, d\pi(x,y)&=\int_{\mathbb{R}^n\times\mathbb{R}^n}[\varphi(x)+\varphi^*(y)]\, d\pi(x,y)\\&=\int_{\mathbb{R}^n\times\mathbb{R}^n}\langle x,y\rangle\, d\pi(x,y),\end{aligned}$$

και αφού $\pi = (Id \times \nabla\varphi)\#\mu$, μπορούμε να γράψουμε

$$\int_{\mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x))]\, d\mu(x) = \int_{\mathbb{R}^n} \langle x, \nabla\varphi(x)\rangle\, d\mu(x),$$

δηλαδή

$$\int_{\mathbb{R}^n} [\bar{\varphi}(x) + \bar{\varphi}^*(\nabla\varphi(x)) - \langle x, \nabla\varphi(x)\rangle]\, d\mu(x) = 0,$$

και αφού η ολοκληρώσιμη ποσότητα είναι μη αρνητική, μηδενίζεται μ-σχεδόν παντού. Χρησιμοποιώντας την Πρόταση 2.3.4 ξανά, παίρνουμε

$$\nabla\varphi(x) \in \partial\bar{\varphi}(x),\ \mu - \text{σχεδόν για κάθε } x,$$

και αφού η $\bar{\varphi}$ είναι διαφορίσιμη μ-σχεδόν για κάθε x, έπεται ότι

$$\nabla\varphi(x) = \nabla\bar{\varphi}(x),\ \mu - \text{σχεδόν για κάθε } x.$$

Η απόδειξη του (ii) είναι πλήρης. Παρατηρούμε ότι αποδείξαμε μοναδικότητα όχι μόνο για την λύση του προβλήματος Monge-Kantorovich αλλά και για την κλίση $\nabla\varphi$ της κυρτής συνάρτησης φ για την οποία ισχύει $\nabla\varphi\#\mu = \nu$.

(v) Τώρα θα δείξουμε ότι $\mathrm{Supp}(\nu) = \overline{\nabla\varphi(\mathrm{Supp}(\mu))}$. Έστω $x \in \mathrm{Supp}(\mu)$ ένα σημείο στο οποίο η φ είναι διαφορίσιμη και έστω $y = \nabla\varphi(x)$. Από την συνέχεια της κλίσης της φ έχουμε ότι για κάθε $\varepsilon > 0$ υπάρχει $\delta > 0$ με $\nabla\varphi(B_\delta(x)) \subset B_\varepsilon(y)$, και ειδικότερα,

$$\nu(B_\varepsilon(y)) \geqslant \mu(\nabla\varphi^{-1}(\nabla\varphi(B_\delta(x)))) \geqslant \mu(B_\delta(x)).$$

Όμως, $\mu(B_\delta(x)) > 0$ για $x \in \mathrm{Supp}(\mu)$, συνεπώς $\nu(B_\varepsilon(y)) > 0$, και αφού το ε ήταν τυχόν έπεται ότι $y \in \mathrm{Supp}(\nu)$. Συμπεραίνουμε λοιπόν ότι

$$\nabla\varphi(\mathrm{Supp}(\mu)) \subset \mathrm{Supp}(\nu). \tag{5.2.19}$$

(vi) Από την άλλη πλευρά,

$$\nu(\nabla\varphi(\mathrm{Supp}(\mu))) = \mu(\nabla\varphi^{-1}(\nabla\varphi(\mathrm{Supp}(\mu)))) \geqslant \mu((\mathrm{Supp}(\mu)) = 1.$$

Άρα, το ν είναι συγκεντρωμένο στο $\nabla\varphi(\mathrm{Supp}(\mu))$, και συνεπώς

$$\mathrm{Supp}(\nu) \subset \overline{\nabla\varphi(\mathrm{Supp}(\mu))}.$$

Συνδυάζοντας αυτό με την (5.2.19) βλέπουμε ότι $\mathrm{Supp}(\nu) = \overline{\nabla\varphi(\mathrm{Supp}(\mu))}$.

(vii) Η απόδειξη του (iii) είναι πλέον προφανής.

(viii) Αν το π είναι βέλτιστο τότε π-σχεδόν παντού $y = \nabla\varphi(x)$, το οποίο είναι ισοδύναμο με το ότι $x \in \partial\varphi^*(y)$ (Πρόταση 2.3.4). Αλλά, αφού η φ^* είναι πεπερασμένη ν-σχεδόν παντού, έπεται ότι είναι και διαφορίσιμη ν-σχεδόν παντού. Άρα π-σχεδόν παντού έχουμε $x = \nabla\varphi^*(y) = \nabla\varphi^*(\nabla\varphi(x))$, και παίρνοντας το περιθώριο μέτρο μ έχουμε ότι η σχέση αυτή ισχύει και μ-σχεδόν παντου. Για τον άλλο ισχυρισμό έχουμε ότι αν το π είναι βέλτιστο, τότε π-σχεδόν παντού ισχύει $y = \nabla\varphi(x)$, άρα $\nabla\varphi^*(y) = \nabla\varphi^*(\nabla\varphi(x)) = x$, και συνεπώς $\nabla\varphi(\nabla\varphi^*(y)) = \nabla\varphi(x) = y$. □

Θα κλείσουμε αυτήν την ενότητα διατυπώνοντας ένα λήμμα για το προωθητικό μέτρο μέ όρους υποδιαφορικών. Υπενθυμίζουμε ότι αν X, Y Πολωνικοί χώροι, μ μέτρο πιθανότητας στον X, ν μέτρο πιθανότητας στον Y και $T : X \to Y$ μετρήσιμη απεικόνιση ώστε $T\#\mu = \nu$, τότε το μέτρο ν καλείται το προωθητικό μέτρο ή η εικόνα του μέτρου μ μέσω της T.

Λήμμα 5.2.5. (i) *Έστω φ μία κυρτή συνάρτηση και μ ένα μέτρο πιθανότητας στον $\mathbb{R}^n$, απολύτως συνεχές ως προς το μέτρο Lebesgue. Τότε, για κάθε Borel σύνολο $A \subset \mathbb{R}^n$ έχουμε,*

$$\nabla\varphi\#\mu(A) = \mu(\partial\varphi^*(A)).$$

(ii) *Αν επιπλέον υποθέσουμε ότι το μέτρο $\nabla\varphi\#\mu = \nu$ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue, και συμβολίσουμε με f και g τις αντίστοιχες πυκνότητες των μ, ν, τότε για κάθε Borel υποσύνολο $A \subset \mathbb{R}^n$ έχουμε*

$$\int_{\partial\varphi(A)} g(y)dy = \int_A f(x)dx.$$

Απόδειξη. Για να αποδείξουμε το (i) αρκεί να δείξουμε ότι για κάθε μετρήσιμο σύνολο A, ισχύει ότι $\mu[(\nabla\varphi)^{-1}(A)] = \mu(\partial\varphi^*(A))$. Από τις ιδιότητες των κυρτών συναρτήσεων έχουμε ότι $\nabla\varphi(x) = y \Longrightarrow x \in \partial\varphi^*(y)$, συνεπώς $(\nabla\varphi)^{-1}(A) \subset \partial\varphi^*(A)$. Επιπλέον, αφου μ απόλυτα συνεχές, αρκεί να ελέγξουμε ότι το σύνολο

$$Z = \partial\varphi^*(A) \setminus (\nabla\varphi)^{-1}(A)$$

έχει μέτρο Lebesgue 0.

Έστω $z \in \partial\varphi^*(A)$. Τότε, υπάρχει $x \in A$ τέτοιο ώστε $z \in \partial\varphi^*(x)$, το οποίο σημαίνει ότι $x \in \partial\varphi(z)$. Αν η φ είναι διαφορίσιμη στο z, τότε $\nabla\varphi(z) = x \in A$, άρα $z \in (\nabla\varphi)^{-1}(A)$. Συνεπώς το Z περιέχεται στο σύνολο των σημείων όπου η φ δεν είναι διαφορίσιμη, το οποίο έχει μέτρο Lebesgue 0, άρα και το Z έχει μέτρο Lebesgue 0.

Από το Θεώρημα 5.2.3 γνωρίζουμε ότι $\nabla\varphi^*\#\nu = \mu$ όταν το ν είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue. Συνεπώς για κάθε Borel υποσύνολο $A \subset \mathbb{R}^n$ έχουμε,

$$\begin{aligned}\int_{\partial\varphi(A)} g(y)dy &= \int_{\partial\varphi(A)} d\nu(y) = \int_{\partial\varphi(A)} d(\nabla\varphi\#\mu) \\ &= \int_{\partial\varphi(A)} d\mu(\partial\varphi^*(y)) = \int_A d\mu(x) \\ &= \int_A f(x)dx.\end{aligned}$$

□

5.3 Το θεώρημα του McCann

Στην ενότητα αυτή θα ασχοληθούμε και πάλι με το πρόβλημα βέλτιστης μεταφοράς στην περίπτωση της τετραγωνικής συνάρτησης κόστους, όμως δεν θα χρησιμοποιήσουμε την μέθοδο δυϊσμού του Kantorovich, αλλά γεωμετρικά επιχειρήματα. Το πλεονέκτημα είναι πως δεν απαιτούνται περιορισμοί για τις ροπές των μέτρων μ, ν και πως ο δυϊσμός του Kantorovich προκύπτει ως συνέπεια αυτών των μεθόδων. Τα βασικά εργαλεία θα είναι το θεώρημα του Rockafellar (βλέπε [56]) για τον χαρακτηρισμό των κυκλικά μονότονων συνόλων και ένα λήμμα του Aleksandrov (βλέπε [3]), το οποίο γενίκευσε ο McCann.

Ορισμός 5.3.1 (κυκλική μονοτονία). Ένα υποσύνολο $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ λέγεται *κυκλικά μονότονο* αν ικανοποιεί την εξής συνθήκη: για κάθε $m \geqslant 1$ και για κάθε $(x_1, y_1), \dots, (x_m, y_m) \in \Gamma$,

$$\sum_{i=1}^m |x_i - y_i|^2 \leqslant \sum_{i=1}^m |x_i - y_{i-1}|^2, \tag{5.3.1}$$

με την σύμβαση $y_0 = y_m$, ή ισοδύναμα,

$$\sum_{i=1}^m \langle y_i, x_{i+1} - x_i\rangle \leqslant 0, \tag{5.3.2}$$

με την σύμβαση $x_{m+1} = x_1$.

Παρατήρηση 5.3.2. Αν $\pi_k \to \pi$ ασθενώς, τότε κάθε σημείο $(x, y) \in \mathrm{Supp}(\pi)$ μπορεί να προσεγγισθεί από μια ακολουθία $(x_k, \psi_k) \in \mathrm{Supp}(\pi_k)$. Συνεπώς, το ασθενές όριο μιας ακολουθίας μέτρων πιθανότητας που είναι συγκεντρωμένα σε κυκλικά μονότονα σύνολα είναι επίσης συγκεντρωμένο σε ένα κυκλικά μονότονο σύνολο.

Πρόταση 5.3.3 (τα βέλτιστα σχέδια μεταφοράς έχουν κυκλικά μονότονους φορείς). *Έστω μ, ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$ και έστω $\pi \in \Pi(\mu, \nu)$ βέλτιστο σχέδιο για το* *πρόβλημα μεταφοράς μάζας από το μ στο ν του Kantorovich, με τετραγωνική συνάρτηση κόστους $c(x,y)=|x-y|^2$. Τότε, το $\mathrm{Supp}(\pi)$ είναι κυκλικά μονότονο.*

Για την απόδειξη της Πρότασης 5.3.3 θα χρησιμοποιήσουμε την παρατήρηση ότι, αν $\mu_j \in \mathcal{P}(X)$, $j=1,\dots,n$ (όπου $\mathcal{P}(X)$ η οικογένεια των Borel μέτρων πιθανότητας στον χώρο X) τότε υπάρχει ένας χώρος πιθανότητας $(\Omega,\mathcal{B},\eta)$ ώστε κάθε μ_j να είναι το μέτρο εικόνα του η μέσω μιας Borel μετρήσιμης απεικόνισης $\pi_j:\Omega\to X$: παίρνουμε $\eta := \mu_1\otimes\mu_2\otimes\cdots\otimes\mu_n$ το μέτρο γινόμενο πάνω στον χώρο γινόμενο $\Omega := X^n$ και $\pi_j(x_1,..,x_n) := x_j$ να είναι η προβολή στην $j-$συνιστώσα του X. Τότε, $\pi_j\#\eta=\mu_j$.

Απόδειξη της Πρότασης 5.3.3. Θα αποδείξουμε την πρόταση σε ένα πιο γενικό πλαίσιο. Θεωρούμε μια συνεχή συνάρτηση $c(x,y)\geqslant 0$ στον $X\times Y$. Αν το $\gamma\in\mathcal{P}(X\times Y)$ είναι βέλτιστο και $I[\gamma]<\infty$, τότε το γ έχει $c-$κυκλικά μονότονο φορέα, δηλαδή για κάθε $(x_i,y_i)\in\mathrm{Supp}(\gamma)$, $i=1,\dots,n$ και για κάθε μετάθεση σ του $\{1,\dots,n\}$ ισχύει

$$\sum_{j=1}^n c(x_j,y_j)\leqslant\sum_{j=1}^n c(x_{\sigma(j)},y_j).$$

Έστω λοιπόν βέλτιστο $\gamma\in\Pi(\mu,\nu)$, δηλαδή $I[\gamma]=\inf_{\Pi(\mu,\nu)}I[\pi]$. Υποθέτουμε ότι το $\mathrm{Supp}(\gamma)$ δεν είναι κυκλικά μονότονο. Τότε υπάρχουν ακέραιος n και μετάθεση σ του $\{1,\dots,n\}$ ώστε η συνάρτηση

$$f(x_1,\dots,x_n;y_1,\dots,y_n):=\sum_{j=1}^n c(x_{\sigma(j)},y_j)-c(x_j,y_j)$$

να παίρνει αρνητική τιμή για κάποια σημεία $(x_1,y_1),\dots,(x_n,y_n)\in\mathrm{Supp}(\gamma)$. Τώρα, αφού η f είναι συνεχής υπάρχουν συμπαγείς περιοχές $U_j\subset X$ του x_j και $V_j\subset Y$ του y_j ώστε, $f(u_1,\dots,u_n;v_1,\dots,v_n)<0$ για κάθε $u_j\in U_j$ και $v_j\in V_j$. Επίσης,

$$\lambda:=\inf_j\gamma(U_j\times V_j)>0,$$

διότι $(x_j,y_j)\in\mathrm{Supp}(\gamma)$. Έστω $\gamma_j\in\mathcal{P}(X\times Y)$ ο περιορισμός του γ στο $U_j\times V_j$. Εισάγουμε έναν παράγοντα n σε περίπτωση που τα γ_j δεν έχουν ξένους ανά δύο φορείς και αφαιρώντας την ποσότητα $\sum_j\frac{\lambda\gamma_j}{n}$ από το γ μένει μία ποσότητα με θετικό μέτρο. Για κάθε j επιλέγουμε μία Borel απεικόνιση $\omega\to(u_j(\omega),v_j(\omega))$ από το Ω στο $X\times Y$ ώστε $\gamma_j=(u_j\times v_j)\#\eta$, η οποία παίρνει τιμές στο συμπαγές σύνολο $U_j\times V_j$. Ορίζουμε το θετικό μέτρο,

$$\gamma':=\gamma+\frac{\lambda}{n}\sum_{j=1}^n(u_{\sigma(j)}\times v_j)\#\eta-(u_j\times v_j)\#\eta.$$

Παρατηρούμε ότι $\gamma' \in \Pi(\mu, \nu)$ και

$$I[\gamma'] - I[\gamma] = \frac{\lambda}{n} \int_\Omega \left[\sum_{j=1}^{n} c(u_{\sigma(j)}, v_j) - c(u_j, v_j) \right] d\eta < 0.$$

Συνεπώς, το γ δεν είναι βέλτιστο, άτοπο.

Άρα, το γ πρέπει να έχει c-κυκλικό φορέα. □

Θεώρημα 5.3.4 (Θεώρημα Rockafellar). *Ένα μη κενό σύνολο $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ είναι κυκλικά μονότονο αν και μόνο αν περιέχεται στο υποδιαφορικό μιας γνήσιας κυρτής, κάτω ημισυνεχούς συνάρτησης φ στον $\mathbb{R}^n$. Επιπλέον, τα μεγιστικά (με την έννοια του περιέχεσθαι) κυκλικά μονότονα σύνολα είναι ακριβώς τα υποδιαφορικά των γνησίων κυρτών, κάτω ημισυνεχών συναρτήσεων.*

Απόδειξη. (α) Έστω φ μια κυρτή συνάρτηση. Θα αποδείξουμε ότι το υποδιαφορικό της είναι ένα κυκλικά μονότονο υποσύνολο του $\mathbb{R}^n \times \mathbb{R}^n$. Αν δειχθεί αυτό τότε έπεται ότι κάθε υποσύνολό του είναι επίσης κυκλικά μονότονο.

Έστω $(x_1, y_1), \ldots, (x_m, y_m)$ τέτοια ώστε $y_i \in \partial\varphi(x_i)$ για κάθε $1 \leqslant i \leqslant m$. Από τον ορισμό του υποδιαφορικού έπεται ότι, για κάθε $z \in \mathbb{R}^n$,

$$\varphi(z) \geqslant \varphi(x_i) + \langle \psi_i, z - x_i \rangle.$$

Ειδικότερα,

$$\left\{ \begin{array}{l} \varphi(x_2) \geqslant \varphi(x_1) + \langle y_1, x_2 - x_1 \rangle \\ \varphi(x_3) \geqslant \varphi(x_2) + \langle y_2, x_3 - x_2 \rangle \\ \cdots \\ \varphi(x_1) \geqslant \varphi(x_m) + \langle y_m, x_1 - x_m \rangle \end{array} \right\}$$

Προσθέτοντας όλες αυτές τις ανισότητες παίρνουμε την (5.3.2).

(β) Έστω τώρα ένα κυκλικά μονότονο σύνολο $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$. Θα κατασκευάσουμε μια γνήσια κυρτή, κάτω ημισυνεχή συνάρτηση φ ώστε $\Gamma \subset \mathrm{Graph}(\partial\varphi)$, και αυτό θα ολοκληρώσει την απόδειξη.

Επιλέγουμε ένα $(x_0, y_0) \in \Gamma$ και ορίζουμε

$$\begin{aligned} \varphi(x) = \sup\{ &\langle y_m, x - x_m \rangle + \cdots + \langle y_0, x_1 - x_0 \rangle : m \in \mathbb{N}, \\ &(x_1, y_1), \ldots, (x_m, y_m) \in \Gamma \}. \end{aligned} \tag{5.3.3}$$

Η φ είναι κάτω ημισυνεχής κυρτή συνάρτηση ως το supremum αφφινικών συναρτήσεων. Επιπλέον, λόγω της κυκλικής μονοτονίας του συνόλου Γ έχουμε $\varphi(x_0) \leqslant 0$. Πιο συγκεκριμένα, $\varphi(x_0) = 0$: επιλέγουμε στον ορισμό $m = 1, x_1 = x_0, y_1 = y_0$. Συνεπώς, η φ είναι γνήσια κυρτή. Έστω $(x,y)\in\Gamma$. Για να δείξουμε ότι $\Gamma\subset \mathrm{Graph}(\partial\varphi)$ αρκεί να ελέγξουμε ότι

$$\varphi(z)\geqslant \varphi(x)+\langle y,z-x\rangle \text{ για κάθε } z\in\mathbb{R}^n.$$

Ισοδύναμα, αρκεί να ελέγξουμε ότι για κάθε $z\in\mathbb{R}^n$ και για κάθε $\alpha<\varphi(x)$,

$$\varphi(z)\geqslant \alpha+\langle y,z-x\rangle. \tag{5.3.4}$$

Αν $\alpha<\varphi(x)$, τότε από τον ορισμό της φ υπάρχουν m και $x_1,y_1,\dots,x_m,y_m$ ώστε

$$\alpha\leqslant \langle y_m,x-x_m\rangle+\cdots+\langle y_0,x_1-x_0\rangle.$$

Άρα,

$$\alpha+\langle y,z-x\rangle\leqslant \langle y,z-x\rangle+\langle y_m,x-x_m\rangle+\cdots+\langle y_0,x_1-x_0\rangle.$$

Συνεπώς, αν θέσουμε $x=x_{m+1}, y=y_{m+1}$ και εφαρμόσουμε τον ορισμό της φ παίρνουμε την (5.3.4). □

Συνδυάζοντας την Πρόταση 5.3.3 και το Θεώρημα 5.3.4 παίρνουμε το εξής θεώρημα:

Θεώρημα 5.3.5 (τα βέλτιστα σχέδια μεταφοράς είναι συγκεντρωμένα σε υποδιαφορικά)**.** *Έστω μ,ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, και έστω $\pi\in\Pi(\mu,\nu)$ ένα σχέδιο μεταφοράς. Αν το π είναι βέλτιστο για το πρόβλημα του Kantorovich με τετραγωνική συνάρτηση κόστους, $c(x,y)=|x-y|^2$, τότε το π είναι συγκεντρωμένο στο υποδιαφορικό μίας γνήσιας κυρτής, κάτω ημισυνεχούς συνάρτησης.*

Η απόδειξη αυτού του θεωρήματος δεν χρησιμοποιεί την Ευκλείδεια δομή του χώρου, συνεπώς το ίδιο θεώρημα ισχύει και για τυχόντα χώρο Hilbert. Αν όμως εκμεταλλευτούμε τις ιδιότητες διαφορισιμότητας που έχουν οι κυρτές συναρτήσεις στον $\mathbb{R}^n$ μπορούμε να πάρουμε πιο ακριβή αποτελέσματα. Αυτό ακριβώς χρησιμοποίησε ο McCann και απέδειξε το εξής:

Πόρισμα 5.3.6 (αναθεωρημένο θεώρημα Brenier: ύπαρξη)**.** *Έστω μ,ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, ώστε το μ να είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue, $\mu\ll\lambda$. Τότε, υπάρχει κυρτή συνάρτηση φ στον $\mathbb{R}^n$ ώστε*

$$\nabla\varphi\#\mu=\nu.$$

Παρατηρούμε ότι οι υποθέσεις εδώ είναι πιο γενικές από εκείνες του Θεωρήματος 5.2.3, καθώς δεν απαιτούνται περιορισμοί για τις ροπές δεύτερης τάξης των μέτρων μ,ν. Για να ολοκληρωθεί η αναθεωρημένη εκδοχή του θεωρήματος Bernier, μένει να αποδείξουμε και την μοναδικότητα του βέλτιστου σχεδίου μεταφοράς, υπό την υπόθεση της απόλυτης συνέχειας του μ ως προς το μέτρο Lebesgue. Στην απόδειξη αυτή, η οποία δόθηκε από τον McCann, δεν χρησιμοποιείται η θεωρία δυϊσμού, αλλά γεωμετρικά επιχειρήματα. Πιο συγκεκριμένα, ο McCann χρησιμοποίησε το εξής λήμμα του Aleksandrov:

Λήμμα 5.3.7 (λήμμα Aleksandrov)**.** *Έστω φ και $\bar{\varphi}$ δύο κυρτές συναρτήσεις ώστε $\varphi(x_0) = \bar{\varphi}(x_0)$, αλλά $\nabla\varphi(x_0) \neq \nabla\bar{\varphi}(x_0)$. Θέτουμε $V = \{\varphi > \bar{\varphi}\}$ και*

$$Z = \nabla\bar{\varphi}^{-1}(\nabla\varphi(V)).$$

Τότε, $x_0 \in \bar{V}, Z \subset V$, αλλά η απόσταση του x_0 από το Z είναι θετική.

Απόδειξη. Μπορούμε να υποθέσουμε ότι $\varphi(x_0) = \bar{\varphi}(x_0) = 0$ και $\nabla\varphi(x_0) \neq \nabla\bar{\varphi}(x_0) = 0$. Έστω $x \in Z$ και $y = \nabla\bar{\varphi}(x)$. Τότε υπάρχει $m \in V$ ώστε $y \in \partial\varphi(m)$. Συνεπώς για κάθε $z \in \mathbb{R}^n$,

$$\varphi(z) \geqslant \langle y, z - m\rangle + \varphi(m) \text{ και } \bar{\varphi}(m) \geqslant \langle y, z - x\rangle + \bar{\varphi}(x).$$

Όμως, $\varphi(m) > \bar{\varphi}(m)$, συνεπώς οι παραπάνω ανισότητες δίνουν

$$\varphi(z) > \langle y, z - x\rangle + \bar{\varphi}(x). \tag{5.3.5}$$

Παίρνοντας $z = x$ έπεται ότι $x \in V$, και αφού x τυχόν στο Z έχουμε $Z \subset V$. Τώρα, έστω $(x_n) \in Z$ ώστε $x_n \to x_0$. Τότε για κάθε n υπάρχει $m_n \in V$ ώστε $\nabla\bar{\varphi}(x_n) \in \partial\varphi(m_n)$. Από υπόθεση $\nabla\bar{\varphi}(x_0) = 0$, συνεπώς λόγω κυρτότητας της $\bar{\varphi}$ έπεται ότι $\bar{\varphi} \geqslant 0$ και λόγω συνέχειας της $\nabla\bar{\varphi}$ έπεται ότι $\nabla\bar{\varphi}(x_n) \to 0$. Από την άλλη, $\nabla\varphi(x_0) \neq 0$ από όπου έπεται ότι $\varphi(z) < 0$ για κάποιο z κοντά στο x_0. Χρησιμοποιώντας την (5.3.5) έχουμε

$$\begin{aligned} 0 > \varphi(z) &> \langle \nabla\bar{\varphi}(x_n), z - x_n\rangle + \bar{\varphi}(x_n) \\ &\geqslant -|\nabla\bar{\varphi}(x_n)|\,|z - x_n|. \end{aligned}$$

Συνεπώς έχουμε $x_n \to x_0$ και $\nabla\bar{\varphi}(x_n) \to 0$, το οποίο είναι άτοπο. Άρα, $x \notin \bar{Z}$. □

Κάνοντας χρήση αυτού του λήμματος, ο McCann διατύπωσε την εξής αναθεωρημένη εκδοχή του θεωρήματος Brenier:

Θεώρημα 5.3.8 (αναθεωρημένη εκδοχή του θεωρήματος Brenier)**.** *Έστω μ, ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, ώστε $\mu \ll \lambda$. Τότε, υπάρχει μοναδική μετρήσιμη απεικόνιση T ώστε $T\#\mu = \nu$ και $T = \nabla\varphi$ για κάποια κυρτή συνάρτηση φ, με την έννοια ότι κάθε δύο τέτοιες απεικονίσεις συμπίπτουν μ-σχεδόν παντού.*

Απόδειξη. Έχουμε ήδη αποδείξει την ύπαρξη βέλτιστου σχεδίου, μένει λοιπόν να αποδείξουμε την μοναδικότητα. Υποθέτουμε ότι υπάρχουν δύο διαφορετικές κυρτές συναρτήσεις $\varphi, \bar{\varphi}$ ώστε $\nabla\varphi\#\mu = \nabla\bar{\varphi}\#\mu = \nu$, αλλά οι $\nabla\varphi, \nabla\bar{\varphi}$ δεν ταυτίζονται στο $\mathrm{Supp}(\mu)$. Έστω $x_0 \in \mathrm{Supp}(\mu)$ ώστε $\nabla\varphi(x_0) \neq \nabla\bar{\varphi}(x_0)$. Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι $\varphi(x_0) = \bar{\varphi}(x_0)$ και λόγω ενός θεωρήματος για κυρτές συναρτήσεις, το οποίο έχει αποδείξει ο McCann, μπορεί να αποδειχθεί ότι το σύνολο $\{\varphi = \bar{\varphi}\}$ έχει μέτρο Lebesgue 0. Τώρα, αφού $x_0 \in \mathrm{Supp}(\mu)$ και αφού $\mu \ll \lambda$, κάθε μικρή περιοχή του x_0 έχει τομή είτε με το σύνολο $\{\varphi < \bar{\varphi}\}$ είτε με το σύνολο $\{\varphi > \bar{\varphi}\}$, η οποία έχει θετικό μέτρο ως προς το μ, έστω το πρώτο. Τώρα εφαρμόζουμε το Λήμμα 5.3.7. Αφού η απόσταση του x_0 από το Z είναι θετική έπεται ότι $\mu(Z) < \mu(V)$. Συνεπώς, $\nabla\varphi\#\mu \neq \nabla\bar{\varphi}\#\mu$. Πράγματι,

$$\begin{aligned}\nabla\bar{\varphi}\#\mu(\nabla\varphi(V)) &= \mu(\nabla\bar{\varphi}^{-1}(\nabla\varphi(V))) = \mu(Z) < \mu(V) \leqslant \mu(\nabla\varphi^{-1}(\nabla\varphi(V))) \\ &= \nabla\varphi\#\mu(\nabla\varphi(V)).\end{aligned}$$

Η απόδειξη είναι πλήρης. □

Για την απόδειξη της μοναδικότητας στο προηγούμενο θεώρημα ήταν απαραίτητο να γνωρίζουμε ότι όταν δύο κυρτές συναρτήσεις ψ, φ συμπίπτουν σε ένα σημείο $p \in \mathbb{R}^n$ όπου είναι και οι δύο διαφορίσιμες αλλά $\nabla\psi(p) \neq \nabla\varphi(p)$, τότε τοπικά τουλάχιστον, $\psi = \varphi$ μόνο σε ένα σύνολο που έχει πεπερασμένο μέτρο Hausdorf διάστασης $d-1$. Στο ίδιο συμπέρασμα θα μπορούσαμε να φτάσουμε χρησιμοποιώντας το θεώρημα πεπλεγμένης συνάρτησης, όμως οι φ, ψ δεν είναι απαραίτητα συνεχώς διαφορίσιμες σε κάθε περιοχή του p, η οποία είναι η συνήθης υπόθεση του θεωρήματος. Θα διατυπώσουμε, λοιπόν, μία εκδοχή του θεωρήματος των πεπλεγμένων συναρτήσεων η οποία εφαρμόζεται στην $\psi - \varphi$ και διατυπώθηκε από τον MaCann (βλέπε [46]).

Θεώρημα 5.3.9 (*θεώρημα πεπλεγμένης συνάρτησης*)**.** *Έστω φ, ψ κυρτές συναρτήσεις στον $\mathbb{R}^n$, διαφορίσιμες σε ένα σημείο $p \in \mathbb{R}^n$ με $\varphi(p) = \psi(p)$ αλλά $\nabla\varphi(p) \neq \nabla\psi(p)$. Θεωρούμε την $\nabla\psi(p) - \nabla\varphi(p)$ κατα μήκος του x_1 άξονα (χωρίς βλάβη της γενικότητας). Τότε υπάρχει μια συνάρτηση $f : \mathbb{R}^{n-1} \to \mathbb{R}$ η οποία είναι Lipschitz με σταθερά 1, και μία περιοχή U του p στην οποία ισχύει ότι:*

$$\psi(x) = \varphi(x) \Longleftrightarrow x_1 = f(x_2, \ldots, x_n).$$

Το επόμενο πόρισμα αυτού του θεωρήματος μας δίνει το επιθυμητό αποτέλεσμα.

Πόρισμα 5.3.10. *Έστω φ, ψ δύο κυρτές συναρτήσεις στον $\mathbb{R}^n$, διαφορίσιμες σε ένα σημείο p με $\varphi(p) = \psi(p)$, αλλά $\nabla\varphi(p) \neq \nabla\psi(p)$. Τότε σε μία μικρή περιοχή U του p, το μέτρο Hausdorf διάστασης $d-1$ του συνόλου $\{x \in U : \varphi(x) = \psi(x)\}$ είναι πεπερασμένο.*

Απόδειξη. Ένας τυπικός υπολογισμός στην γεωμετρική θεωρία μέτρου φράσει το μέτρο Hausdorf διάστασης $d-1$, H^{d-1}, της εικόνας $g(M)$ ενός συνόλου M μέσω μίας Lipschitz απεικόνισης g ως εξής:

$$H^{d-1}(g(M)) \leqslant k^{d-1} \cdot H^{d-1}(M), \tag{5.3.6}$$

όπου η $g : \mathbb{R}^n \to \mathbb{R}^m$ ικανοποιεί την $|g(w) - g(z)| \leqslant k \cdot |w - z|$ και $M \subset \mathbb{R}^n$. Παίρνοντας $g(w) := (f(w), w)$ στον $\mathbb{R}^{n-1}$, θεωρώντας το φραγμένο σύνολο $M := \{w : g(w) \in U\}$ και εφαρμόζοντας την (5.3.6), έχουμε το επιθυμητό αποτέλεσμα, δηλαδη $H^{d-1}(g(M)) < \infty$. □

5.4 Βέλτιστη μεταφορά στο $\mathbb{R}$

Ορισμός 5.4.1. Ένα υποσύνολο Γ του $\mathbb{R}^2$ λέγεται *μονότονο* αν,

$$(x_1, y_1), (x_2, y_2) \in \Gamma \Longrightarrow (x_1 \leqslant x_2 \text{ και } y_1 \leqslant y_2) \text{ ή } (x_1 \geqslant x_2 \text{ και } y_1 \geqslant y_2). \tag{5.4.1}$$

Ισοδύναμα μπορούμε να απαιτήσουμε $(x_1 - x_2)(y_1 - y_2) \geqslant 0$.

Στο $\mathbb{R}$ οι κλίσεις των κυρτών συναρτήσεων συμπίπτουν με τις αύξουσες συναρτήσεις και τα υποδιαφορικά των κυρτών συναρτήσεων με τα μέγιστα μονότονα υποσύνολα του $\mathbb{R}^2$. Γεωμετρικά, ένα τέτοιο σύνολο είναι το συνηθισμένο γράφημα μιας αύξουσας συνάρτησης, με πιθανόν κάποιες κατακόρυφες γραμμές οι οποίες προστίθενται ώστε να είναι το γράφημα συνεχές. Αυτές οι γραμμές αντιστοιχούν στα σημεία x στα οποία η αριστερή και δεξιά παράγωγος, $\varphi'_-(x)$ και $\varphi'_+(x)$ της κυρτής συνάρτησης φ δεν συμπίπτουν. Πιο συγκεκριμένα,

$$\partial\varphi(x) = [\varphi'_-(x), \varphi'_+(x)].$$

Κάθε μέτρο πιθανότητας μ στο $\mathbb{R}$ μπορεί να αναπαρασταθεί από την αθροιστική συνάρτηση κατανομής του:

$$F(x) = \int_{-\infty}^{x} d\mu = \mu((-\infty, x]),$$

η οποία, όπως γνωρίζουμε από την θεωρία πιθανοτήτων, είναι δεξιά συνεχής, αύξουσα και ικανοποιεί τις $F(-\infty) = 0$ και $F(+\infty) = 1$. Συνεπώς, μπορούμε να ορίσουμε την *γενικευμένη αντίστροφο* της F μέσω της,

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) > t\}.$$

Η γενικευμένη αντίστροφος της F είναι επίσης δεξιά συνεχής κα ισχύουν οι ανισότητες

$$F^{-1}(F(x)) \geqslant x, \text{ για κάθε } x \in \mathbb{R} \text{ και } F(F^{-1}(t)) \geqslant t, \text{ για κάθε } t \in [0,1].$$

Τώρα, όσον αφορά τα μέτρα πιθανότητας στον χώρο γινόμενο $\mathbb{R} \times \mathbb{R}$, αυτά μπορούν να αναπαρασταθούν από την από κοινού διδιάστατη αθροιστική κατανομή τους:

$$H(x_0, y_0) = \int_{R(x_0, y_0)} d\pi = \pi(R(x_0, y_0)),$$

όπου $R(x_0, y_0)$ είναι το ορθογώνιο που αποτελείται από τα σημεία $(x, y) \in \mathbb{R}^2$ με $x \leqslant x_0$ και $y \leqslant y_0$. Μία συνάρτηση H στον $\mathbb{R}^2$, η οποία είναι αύξουσα, δεξιά συνεχής ως προς τις δύο μεταβλητές x, y και έχει όρια 0 και 1 στα $(-\infty, -\infty)$ και $(+\infty, +\infty)$ αντίστοιχα, αντιστοιχεί σε ένα μοναδικό μέτρο πιθανότητας π στον $\mathbb{R}^2$ (και τότε γράφουμε $d\pi = dH$). Για να το δούμε αυτό παρατηρούμε ότι η H προσδιορίζει τη μάζα όλων των ορθογωνίων που έχουν πλευρές παράλληλες στους άξονες, και αυτά παράγουν όλα τα Borel σύνολα στον $\mathbb{R}^2$.

Θεώρημα 5.4.2 (βέλτιστη μεταφορά για τετραγωνικό κόστος στο $\mathbb{R}$)**.** *Έστω μ, ν δύο μέτρα πιθανότητας στο $\mathbb{R}$ με συναρτήσεις κατανομής F και G αντίστοιχα. Έστω π το μέτρο πιθανότητας στον $\mathbb{R}^2$ με από κοινού συνάρτηση κατανομής*

$$H(x,y) = \min(F(x), G(y)). \tag{5.4.2}$$

Τότε, το $\pi \in \Pi(\mu,\nu)$ και είναι βέλτιστο για το πρόβλημα μεταφοράς του Kantorovich από το μ στο ν, με τετραγωνική συνάρτηση κόστους $c(x,y) = |x-y|^2$. Επιπλέον, η τιμή του βέλτιστου κόστους μεταφοράς είναι

$$\mathcal{T}_2(\mu,\nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt. \tag{5.4.3}$$

Απόδειξη. Αν F είναι μια συνάρτηση κατανομής, θα συμβολίζουμε με $F(x^-)$ το αριστερό πλευρικό όριο $\lim_{z\to x} F(z)$, το οποίο πάντα υπάρχει λόγω μονοτονίας της F, ενώ αφου η F είναι δεξιά συνεχής δεν υπάρχει ανάγκη να εισάγουμε τα δεξιά όρια.

(i) Ισχυριζόμαστε ότι

$$\mathrm{Supp}(\pi) \subset \left\{(x,y) \in \mathbb{R}^2 : F(x^-) \leqslant G(y) \text{ και } G(y^-) \leqslant F(x)\right\}. \tag{5.4.4}$$

Έστω, προς άτοπο, ότι $F(x^-) > G(y)$. Από την δεξιά συνέχεια της G και από το γεγονός ότι οι F, G είναι αύξουσες συναρτήσεις, έπεται ότι για κάθε x' σε μία μικρή περιοχή του x και για κάθε y' σε μία μικρή περιοχή του y ισχύει $F(x') > G(y')$. Άρα,

$$H(x',y') = \min(F(x'), G(y')) = G(y').$$

Συνεπώς, σε ένα μικρό ορθογώνιο κέντρου (x,y) η συνάρτηση H είναι ανεξάρτητη της μεταβλητης x', άρα $\pi(x',y') = 0$ για κάθε (x',y') στο ορθογώνιο αυτό, δηλαδή $(x,y) \notin \mathrm{Supp}(\pi)$, το οποίο είναι άτοπο.

Τώρα θα αποδείξουμε ότι το $\mathrm{Supp}(\pi)$ είναι μονότονο σύνολο. Έστω $(x_1,y_1), (x_2,y_2) \in \mathrm{Supp}(\pi)$. Υποθέτουμε ότι $x_1 > x_2$ και θα δείξουμε ότι $y_1 \geqslant y_2$. Εφαρμόζοντας την (5.4.4) και χρησιμοποιώντας το γεγονός ότι η F είναι αύξουσα, έχουμε

$$G(y_1) \geqslant F(x_1^-) \geqslant F(x_2) \geqslant G(y_2^-).$$

Αν $G(y_1) > G(y_2^-)$ τότε, από την μονοτονία της G, έχουμε ότι $y_2 \leqslant y_1$ και έχουμε τελειώσει. Αν όχι, τότε έχουμε $G(y_1) = F(x_1^-) = F(x_2) = G(y_2^-)$. Αν $y_2 > y_1$, τότε αυτό σημαίνει ότι η F είναι συνεχής στο $[x_2, x_1)$ και η G στο $[y_1, y_2)$. Θα δείξουμε ότι αυτό είναι αδύνατο αποδεικνύοντας ότι τα $(x_1,y_1), (x_2,y_2)$ δεν ανήκουν στο $\mathrm{Supp}(\pi)$. Ας το εξηγήσουμε για το (x_2,y_2). Έστω $\varepsilon > 0$, θα δείξουμε ότι το ορθογώνιο R με κορυφές τα $(x_2-\varepsilon, y_2-\varepsilon), (x_2+\varepsilon, y_2+\varepsilon), (x_2-\varepsilon, y_2+\varepsilon), (x_2+\varepsilon, y_2-\varepsilon)$ έχει μέτρο 0 ως προς π. Εκφράζουμε το μέτρο του ορθογωνίου R μέσω της H, και έχουμε

$$\pi(R) = H(x_2+\varepsilon, y_2+\varepsilon) + H(x_2-\varepsilon, y_2-\varepsilon) - H(x_2-\varepsilon, y_2+\varepsilon) - H(x_2+\varepsilon, y_2-\varepsilon).$$

Χρησιμοποιώντας τον ορισμό της H, τις ανισότητες $x_2 < x_1$ και $y_2 > y_1$, και το γεγονός ότι οι F, G είναι αύξουσες, παίρνουμε για $\varepsilon \to 0$ ότι $\pi(R) = 0$. Συνεπώς η υπόθεση $y_2 > y_1$ είναι αδύνατη και άρα η (5.4.4) ισχύει.

(ii) Έχουμε δείξει λοιπόν ότι το $\mathrm{Supp}(\pi)$ περιέχεται σέ ένα μονότονο υποσύνολο του $\mathbb{R}^2$, συνεπώς στο υποδιαφορικό μίας κάτω ημισυνεχούς, κυρτής συνάρτησης. Άρα, από το κριτήριο των Knott-Smith, το π είναι ένα βέλτιστο σχέδιο μεταφοράς. Τώρα, ισχυριζόμαστε ότι

$$\pi = (F^{-1} \times G^{-1})\#\lambda, \tag{5.4.5}$$

όπου λ είναι το μέτρο Lebesgue στο $[0,1]$. Αρκεί να ελέγξουμε την (5.4.2) για το τυχόν ορθογώνιο της μορφής $R(x,y)$, και σε αυτήν την περίπτωση η (5.4.2) παίρνει τη μορφή

$$\begin{aligned}\pi(R(x,y)) &= \lambda(\{t : (F^{-1}(t), g^{-1}(t)) \in R(x,y)\}) \\ &= \lambda(\{t : F^{-1}(t) \leqslant x\} \cap \{t : G^{-1}(t) \leqslant y\}).\end{aligned} \tag{5.4.6}$$

Ανάλογα με την περίπτωση, το σύνολο $\{F^{-1}(t) \leqslant x\}$ είναι το $[0, F(x)]$ ή το $[0, F(x))$. Σε κάθε περίπτωση όμως, η ποσότητα της οποίας παίρνουμε το μέτρο Lebesgue στην (5.4.6) είναι ένα διάστημα με άκρα τα 0 και $\min(F(x), G(y))$, και άρα το μέτρο του είναι ίσο με το $\min(F(x), G(y)) = H(x,y)$. Η απόδειξη του ισχυρισμού είναι πλήρης.

(iii) Ως συνέπεια της (5.4.5), για κάθε μη αρνητική, μετρήσιμη συνάρτηση g στο $\mathbb{R}^2$, έχουμε

$$\int_{\mathbb{R}^2} g(x,y)\,d\pi(x,y) = \int_0^1 g(F^{-1}(t), G^{-1}(t))\,dt.$$

Έπεται η (5.4.3). $\square$

Παρατηρήσεις 5.4.3. (α) Το μέτρο π που κατασκευάστηκε στο Θεώρημα 5.4.2 είναι βέλτιστο, όποια κυρτή συνάρτηση και αν έχουμε ως συνάρτηση κόστους. Πιο συγκεκριμένα, το π είναι βέλτιστο όταν η συνάρτηση κόστους $c(x,y)$ έχει την μορφή $c(x-y)$, όπου c είναι μία κυρτή μη αρνητική συμμετρική συνάρτηση στο $\mathbb{R}$. Στην περίπτωση αυτή, το βέλτιστο κόστος μεταφοράς είναι

$$\mathcal{T}_c(\mu,\nu) = \int_0^1 c(F^{-1}(t) - G^{-1}(t))\,dt.$$

(β) Ιδιαίτερα, στην περίπτωση που η συνάρτηση κόστους είναι η $c(x, y) = |x - y|$, το ολικό κόστος μεταφοράς είναι

$$\mathcal{T}_1(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|dt = \int_{\mathbb{R}} |F(x) - G(x)|dx,$$

όπου η δεύτερη ισότητα είναι συνέπεια του θεωρημάτος Fubini. Συνεπώς, στην περίπτωση αυτή το βέλτιστο κόστος μεταφοράς συμπίπτει με την L^1 απόσταση των αθροιστικών συναρτήσεων κατανομής.

(γ) Αν το μέτρο μ δεν δίνει μάζα σε σημεία, τότε η συνάρτηση $T = G^{-1} \circ F$ μεταφέρει το μ στο ν και ισχύει ότι

$$\int_{-\infty}^{x} d\mu = \int_{-\infty}^{T(x)} d\nu. \tag{5.4.7}$$

Η σχέση αυτή εκφράζει το γεγονός πως η λύση του προβλήματος βέλτιστης μεταφοράς δίνεται από μία μονότονη αναδιάταξη του μ επί του ν. Παρατηρούμε ότι τα σημεία ασυνέχειας της G αντιστοιχούν σε άτομα για το ν.

(δ) Υποθέτουμε ότι τα μέτρα μ, ν έχουν πυκνότητες f και g ως προς το μέτρο Lebesgue. Έστω επιπλέον ότι οι f και g είναι συνεχείς και ότι η g είναι αυστηρώς θετική. Τότε, η T είναι C^1 συνάρτηση και παραγωγίζοντας την (5.4.7) παίρνουμε την ταυτότητα

$$f(x) = g(T(x))T'(x).$$

Στη συνέχεια θα δούμε τη μορφή που παίρνει το Θεώρημα 4.2.1 όταν $X = \mathbb{R}$. Στην περίπτωση αυτή, το θεώρημα μπορεί να διατυπωθεί μέσω των αθροιστικών συναρτήσεων κατανομής. Υπενθυμίζουμε ότι κάθε μέτρο πιθανότητας μ στο $\mathbb{R}$ μπορεί να αναπαρασταθεί από την αθροιστική συνάρτηση κατανομής του:

$$F(x) = \int_{-\infty}^{x} d\mu = \mu((-\infty, x]),$$

η οποία (από την θεωρία πιθανοτήτων) γνωρίζουμε πως είναι δεξιά συνεχής, αύξουσα και ικανοποιεί τις $F(-\infty) = 0$ και $F(+\infty) = 1$.

Ορισμός 5.4.4. Έστω $(\Omega, \mathcal{A}, \mathrm{P})$ ένας χώρος μέτρου και έστω $U : (\Omega, \mathcal{A}) \to \mathbb{R}$ πραγματική τυχαία μεταβλητή. Τότε ορίζουμε law(U) να είναι το μέτρο πιθανότητας μ στο $\mathbb{R}$ το οποίο ικανοποιεί την

$$\mu(A) = \mathrm{P}(U^{-1}(A))$$

για κάθε $A \in \mathcal{A}$.

Θα χρησιμοποιήσουμε μία βασική πρόταση από την θεωρία πιθανοτήτων.

Πρόταση 5.4.5. *Έστω (μ_k) μία ακολουθία μέτρων πιθανότητας στο $\mathbb{R}$, με αντίστοιχες αθροιστικές συναρτήσεις κατανομής F_k, και έστω μ ένα μέτρο πιθανότητας με αθροιστική συνάρτηση κατανομής F. Τότε, $\mu_k \xrightarrow{w^*} \mu$ αν και μόνο αν*

$$F_k(x) \to F(x), \text{ για κάθε } x \text{ στο οποίο η } F \text{ είναι συνεχής.}$$

Υπενθυμίζουμε ότι

$$W_1(\mu_k, \mu) = T_1(\mu_k, \mu) = \|F_k - F\|_{L^1(\mathbb{R})},$$

όταν η W_1 υπολογίζεται με την συνήθη μετρική στον $\mathbb{R}$. Για απλότητα υποθέτουμε ότι δουλεύουμε με κατανομές πιθανότητας οι οποίες έχουν συμπαγείς φορείς. Λαμβάνοντας υπόψιν και την ισοδυναμία των αποστάσεων Monge-Kantorovich στην περίπτωση αυτή, το θεώρημα 4.2.1 μεταφράζεται ως εξής:

> *Έστω (F_k) και F αύξουσες και δεξιά συνεχείς συναρτήσεις σε ένα διάστημα $[\alpha, \beta] \subset \mathbb{R}$, με τιμές 0 στο α και 1 στο β. Τότε, $F_k \xrightarrow{\|\cdot\|_{L^1}} F$ αν και μόνο αν $F_k(x) \to F(x)$ για κάθε x στο οποίο η F είναι συνεχής.*

Ορισμός 5.4.6. Η ακολουθία πραγματικών τυχαίων ματαβλητών (X_n) συγκλίνει κατά κατανομή στην τυχαία μεταβλητή X, αν για κάθε πραγματική, φραγμένη και συνεχή συνάρτηση $g : \mathbb{R} \to \mathbb{R}$ ισχύει ότι,

$$\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X)).$$

Το θεώρημα αναπαράστασης του Skorohod ισχυρίζεται το εξής: αν $(V_k)_{k\in\mathbb{N}}$ είναι μία ακολουθία πραγματικών τυχαίων μεταβλητών, η οποία συγκλίνει κατά κατανομή σε μία τυχαία μεταβλητή V (δηλαδή, το $\mu_k = \mathrm{law}(V_k) \xrightarrow{w^*} \mu = \mathrm{law}(V)$), τότε υπάρχει μία ακολουθία τυχαίων μεταβλητων (V_k') και μία τυχαία μεταβλητή V, ώστε για κάθε k, $\mathrm{law}(V_k') = \mathrm{law}(V_k)$, $\mathrm{law}(V) = \mathrm{law}(V')$ και η (V_k') συγκλίνει σχεδόν παντού στην V'.

Για να αποδείξουμε το θεώρημα αυτό αρκεί, θέτοντας

$$V' = F^{-1}(U), \; V_k' = F_k^{-1}(U),$$

όπου U τυχαία μεταβλητή ομοιόμορφα κατανεμημένη στο $[0, 1]$, να ελέγξουμε ότι ισχύει το κριτήριο της Πρότασης 5.4.5. Όμως, με την χρήση της απόστασης Wasserstein W_1 μπορούμε να διατυπώσουμε την εξής πιο ποσοτική μορφή του θεωρήματος Skorohod: με τον συμβολισμό $\mathbb{E}$ για την μέση τιμή και με την επιπρόσθετη υπόθεση ότι

$$\lim_{R\to\infty} \limsup_{k\to\infty} \mathbb{E}|V_k| \cdot \mathbf{1}_{\{|V_k|\geqslant R\}} = 0,$$

(η οποία ικανοποιείται για παράδειγμα αν $\sup \mathbb{E}|V_k|^p < +\infty$ για κάποιο $p > 1$), γνωρίζουμε από το Θεώρημα 4.2.1 ότι

$$W_1(\mu_k, \mu) \to 0,$$

και μπορούμε να γράψουμε

$$\|V_k' - V'\|_{L^1} = \|F_k^{-1} - F^{-1}\|_{L^1(0,1)} = W_1(\mu_k, \mu),$$

από όπου έπεται ότι η V_k' συγκλίνει σχεδόν παντού στην V'.

5.5 Πολικό θεώρημα παραγοντοποίησης του Brenier

Ορισμός 5.5.1 (αναδιάταξη)**.** Έστω $f : (W, \lambda) \to (X, \mu)$ μία μετρήσιμη συνάρτηση μεταξύ δύο χώρων μέτρου. Μία συνάρτηση $\tilde{f} : (W, \lambda) \to (X, \mu)$ λέγεται αναδιάταξη της f αν ισχύει το εξής: για κάθε μετρήσιμη συνάρτηση $g : X \to \mathbb{R}$ τέτοια ώστε $g \circ f \in L^1(\lambda)$, ισχύει ότι $g \circ \tilde{f} \in L^1(\lambda)$ και

$$\int_W (g \circ f) d\lambda = \int_W (g \circ \tilde{f}) d\lambda. \tag{5.5.1}$$

Παρατήρηση 5.5.2. Αν $\lambda(W) < +\infty$, τότε είναι ισοδύναμο να απαιτήσουμε τα ολοκληρώματα στην (5.5.1) να είναι ίσα για κάθε φραγμένη μετρήσιμη συνάρτηση g. Γενικά όμως είναι ισοδύναμο να απαιτήσουμε η (5.5.1) να ισχύει για κάθε μη αρνητική μετρήσιμη συνάρτηση g.

Ορισμός 5.5.3 (απεικονίσεις που διατηρούν το μέτρο)**.** Έστω (W, λ) ένας μετρήσιμος χώρος. Λέμε ότι μία μετρήσιμη συνάρτηση $s : W \to W$ διατηρεί το μέτρο αν

$$s\#\lambda = \lambda.$$

Με άλλα λόγια, αν για κάθε μετρήσιμο σύνολο $A \subset W$ έχουμε $\lambda(s^{-1}(A)) = \lambda(A)$. Συμβολίζουμε τον χώρο των συναρτήσεων $s : W \to W$ που διατηρούν το μέτρο με $S(W)$.

Πρόταση 5.5.4. *Έστω (W, λ) ένας μετρήσιμος χώρος και $f : W \to W$ μετρήσιμη συνάρτηση. Αν $s \in S(W)$ και $\tilde{f} = f \circ s$, τότε η $\tilde{f}$ είναι αναδιάταξη της f. Αντίστροφα, αν $\tilde{f}$ είναι μία ένα προς ένα αναδιάταξη της f, τότε $\tilde{f}^{-1} \circ f \in S(W)$.*

Απόδειξη. Έστω $\tilde{f} = f \circ s$, όπου $s \in S(W)$. Τότε, για κάθε μετρήσιμη συνάρτηση $g : W \to \mathbb{R}_+$, έχουμε

$$\int (g \circ \tilde{f})\, d\lambda = \int (g \circ f) \circ s\, d\lambda = \int (g \circ f)\, d(s\#\lambda) = \int (g \circ f)\, d\lambda.$$

Άρα, η $\tilde{f}$ είναι αναδιάταξη της f. Αντίστροφα, έστω $\tilde{f}$ μία ένα προς ένα αναδιάταξη της f. Ορίζουμε $s = \tilde{f}^{-1} \circ f$. Τότε, για κάθε μη αρνητική μετρήσιμη συνάρτηση $g : W \to \mathbb{R}_+$, έχουμε

$$\int g\, d(s\#\lambda) = \int (g \circ s)\, d\lambda = \int (g \circ \tilde{f}^{-1}) \circ f\, d\lambda = \int (g \circ \tilde{f}^{-1}) \circ \tilde{f}\, d\lambda = \int g\, d\lambda.$$

Συνεπώς, $s \in S(W)$. □

Θεώρημα 5.5.5 (πολικό θεώρημα παραγοντοποίησης του Brenier). *Έστω Ω ένα φραγμένο υποσύνολο του $\mathbb{R}^n$ με θετικό μέτρο Lebesgue. Έστω $h : \Omega \to \mathbb{R}^n$ μία L^2 απεικόνιση η οποία ικανοποιεί την εξής μη εκφυλιστική συνθήκη:*

$$\lambda(h^{-1}(N)) = 0 \text{ για κάθε } N \subset \mathbb{R}^n \text{ με } \lambda(N) = 0.$$

Τότε, υπάρχει μοναδική αναδιάταξη $\nabla\psi$ της h στην κλάση των L^2 κλίσεων κυρτών συναρτήσεων και μοναδική απεικόνιση $s \in S(\Omega)$ που διατηρεί το μέτρο, τέτοια ώστε

$$h = \nabla\psi \circ s.$$

Επιπλέον, η s είναι η μοναδική L^2 προβολή της h στον $S(\Omega)$.

Παρατηρήσεις 5.5.6. (i) Όταν λέμε ότι η s είναι η L^2 προβολή της h στον $S(\Omega)$ εννοούμε ότι η s ελαχιστοποιεί την ποσότητα $\|h - \sigma\|_{L^2(\Omega)}$ πάνω από όλες τις $\sigma \in S(\Omega)$.

(ii) Η μη εκφυλιστική συνθήκη σημαίνει ότι η εικόνα του μέτρου Lebesgue μέσω της h δεν δίνει μάζα σε σύνολα μηδενικού μέτρου Lebesgue. Δηλαδή, αν $\lambda(h^{-1}(N)) = 0$ για κάθε μετρήσιμο σύνολο N που έχει μέτρο 0. Αυτή ήταν η αρχική υπόθεση του Bernier (βλέπε [14]). Η μη εκφυλιστική συνθήκη δεν είναι αναγκαία για την ύπαρξη της πολικής παραγοντοποίησης (βλέπε [16]), είναι όμως αναγκαία συνθήκη για την μοναδικότητα.

Το πολικό θέωρημα παραγοντοποίησης του Brenier είναι άμεσα συνδεδεμένο με το πρόβλημα μεταφοράς του Monge. Στη συνέχεια θα διατυπώσουμε και θα αποδείξουμε ένα πιο γενικό αποτέλεσμα το οποίο είναι σχεδόν ισοδύναμο με το Θεώρημα 5.5.5.

Θεώρημα 5.5.7. *Έστω W και Y μετρήσιμα υποσύνολα του $\mathbb{R}^n$ και έστω $\lambda \in \mathcal{P}(W)$ και $\nu \in \mathcal{P}(Y)$ με $\int_Y |y|^2 d\nu(y) < +\infty$. Έστω, επιπλέον, $h : W \to X \subset \mathbb{R}^n$ μία $L^2(\lambda)$ απεικόνιση και έστω $\mu = h\#\lambda$. Υποθέτουμε ότι τα μέτρα μ, ν δεν δίνουν μάζα σε σύνολα μέτρου Lebesgue 0. Τότε, υπάρχει μοναδικό ζευγάρι $(\nabla\psi, s)$ τέτοιο ώστε,*

$$\left\{ \begin{array}{ll} \eta\ \psi : Y \to X & \text{είναι κυρτή συνάρτηση} \\ s : W \to Y & \text{με } s\#\lambda = \nu \\ h = \nabla\psi \circ s & \end{array} \right\}$$

Επιπλέον η s είναι η μοναδική L^2 προβολή της h στον $S(W,Y)=\{\sigma : W\to Y : \sigma\#\lambda=\nu\}$.

Απόδειξη. Αναζητούμε μία απεικόνιση $s\in S(W,Y)=\{\sigma : W\to Y : \sigma\#\lambda=\nu\}$ που να ελαχιστοποιεί την ποσότητα

$$\int_W |h(w)-\sigma(w)|^2 d\lambda(w)$$

πάνω από όλες τις $\sigma\in S(W,Y)$. Πιο συγκεκριμένα λοιπόν αναζητούμε την προβολή της h στον $S(W,Y)$. Για κάθε $\sigma\in S(W,Y)$ θεωρούμε το μέτρο $\pi_\sigma=(h\times\sigma)\#\lambda$. Τότε, το πρόβλημά μας παίρνει την μορφή

$$\min_{\sigma\in S(W,Y)}\left\{\int_{X\times Y}|x-y|^2 d\pi_\sigma(x,y) : \pi_\sigma=(h\times\sigma)\#\lambda,\ \sigma\#\lambda=\nu\right\}.$$

Παρατηρούμε ότι $\pi_\sigma\in\Pi(h\#\lambda,\sigma\#\lambda)=\Pi(\mu,\nu)$, όπου $\nu=h\#\lambda$. Θα δείξουμε ότι

(5.5.2)
$$\min_{\sigma\in S(W,Y)}\left\{\int_{X\times Y}|x-y|^2 d\pi : \pi=(h\times\sigma)\#\lambda,\ \sigma\#\lambda=\nu\right\}=\min_{\Pi(\mu,\nu)}\int_{X\times Y}|x-y|^2 d\pi.$$

Έχουμε ότι για κάθε $\sigma\in S(W,Y)$ ισχύει $\pi_\sigma=(h\times\sigma)\#\lambda\in\Pi(\mu,\nu)$, συνεπώς το πρώτο μέλος είναι μεγαλύτερο ή ίσο του δεύτερου. Τώρα, για την αντίστροφη ανισότητα παρατηρούμε ότι,

$$\int_X |x|^2 d\mu(x)=\int_X |x|^2 d(h\#\lambda)=\int_W |h(w)|^2 d\lambda(w)<+\infty,$$

διότι από υπόθεση η $h\in L^2(\lambda)$. Συνεπώς, τα μ,ν είναι μέτρα πιθανότητας στον $\mathbb{R}^n$ με πεπερασμένες ροπές δεύτερης τάξης. Άρα, από το Θεώρημα 5.2.1 υπάρχει ένα ζεύγος (φ,φ^*) από γνήσιες κάτω ημισυνεχείς, συζυγείς, κυρτές συναρτήσεις το οποίο αποτελεί λύση για το δυϊκό πρόβλημα των Monge-Kantorovich, δηλαδή $J(\varphi,\varphi^*)=\inf_{\tilde\Phi} J$, όπου η $\tilde\Phi$ είναι όπως ακριβώς ορίστηκε στην (5.1.8). Από το θεώρημα του Brenier (Θεώρημα 5.2.3) έχουμε ότι το μοναδικό βέλτιστο σχέδιο μεταφοράς για το πρόβλημα των Monge-Kantorovich με συνάρτηση κόστους $c(x,y)=|x-y|^2$ (δηλαδή ακριβώς το δεξιό μέλος της (5.5.2)) έχει τη μορφή $\pi_*=(Id\times\nabla\varphi)\#\mu$ και ισχύει ότι $\nu=\nabla\varphi\#\mu=(\nabla\varphi\circ h)\#\lambda$. Συνεπώς,

$$\min_{\Pi(\mu,\nu)}\int|x-y|^2 d\pi=\int|x-y|^2 d\pi_*.$$

Θέτουμε

(5.5.3)
$$s=\nabla\varphi\circ h$$

και παρατηρούμε ότι $s\in S(W,Y)$ αφού $s\#\lambda=\nu$. Συνεπώς, το $\pi_s=(h\times s)\#\lambda\equiv\pi_*$ είναι βέλτιστο σχέδιο για το αριστερό μέλος της (5.5.2). Έτσι παίρνουμε την αντίστροφη ανισότητα, και έπεται η (5.5.2). Ειδικότερα, θέτοντας $\psi = \varphi^*$ έχουμε ότι το ζεύγος $(\nabla\psi, s)$ αποτελεί λύση για το πρόβλημα.

Μένει να δείξουμε την μοναδικότητα, την οποία θα πάρουμε ως συνέπεια της μοναδικότητας της L^2 προβολής s της h στον $S(W)$. Έστω, λοιπόν προς απαγωγή σε άτοπο, ότι s' είναι μία άλλη L^2 προβολή της h στον $S(W)$. Τότε, λόγω της μοναδικότητας της λύσης του προβλήματος των Monge-Kantorovich, έχουμε ότι

$$(h \times s)\#\lambda = (h \times s')\#\lambda. \tag{5.5.4}$$

Θεωρούμε την συνάρτηση $F(x, y) = \langle \nabla\varphi(x), y\rangle$ και λόγω της (5.5.4) έχουμε ότι

$$\begin{aligned}
&\int_W F(h(w), s(w))d\lambda(w) = \int_W F(h(w), s'(w))d\lambda(w)\\
\Longrightarrow &\int_W \langle \nabla\varphi(h(w)), s(w)\rangle\, d\lambda(w) = \int_W \langle \nabla\varphi(h(w)), s'(w)\rangle\, d\lambda(w)\\
\Longrightarrow &\int_W \langle s(w), s(w)\rangle d\lambda(w) = \int_W \langle s(w), s'(w)\rangle\, d\lambda(w)\\
\Longrightarrow &\int_W |s(w)|^2\, d\lambda(w) = \int_W \langle s(w), s'(w)\rangle\, d\lambda(w),
\end{aligned}$$

όπου για τη δεύτερη συνεπαγωγή χρησιμοποιήσαμε την (5.5.3). Όμως, ο χώρος L^2 είναι Hilbert και οι s, s' είναι L^2 προβολές της h στον $S(W)$, συνεπώς έχουμε $s - h \perp h$ και $s' - h \perp h$. Άρα, από τον κανόνα του παραλληλογράμμου έχουμε ότι

$$\begin{aligned}
\int |s'|^2 d\lambda = \int |s' - h + h|^2 d\lambda &= \int |s' - h|^2 d\lambda + \int |h|^2 d\lambda\\
&= \int |s - h|^2 + \int |h|^2 d\lambda = \int |s|^2 d\lambda,
\end{aligned}$$

συνεπώς $s = s'$ λ-σχεδόν παντού. □

Το θεώρημα 5.5.5 είναι η περίπτωση όπου $W = Y = \Omega$ (εφοδιασμένο με το μέτρο Lebesgue λ).

Κεφάλαιο 6

Γεωμετρικές ανισότητες

Στο κεφάλαιο αυτό θα ασχοληθούμε με κάποιες εφαρμογές της βέλτιστης μεταφοράς στο πεδίο των συναρτησιακών ανισοτήτων με γεωμετρικό περιεχόμενο. Αρχικά θα παρουσιάσουμε μία απόδειξη της ανισότητας Brunn-Minkowski και της αντίστοιχης συναρτησιακής της μορφής, της ανισότητας Prékopa-Leindler. Έπειτα, θα ασχοληθούμε με μία πολύ γενικότερη ανισότητα, η οποία οφείλεται στον Barthe και είναι δυϊκή της ανισότητας Brascamp-Lieb. Τέλος, θα δώσουμε αποδείξεις, που δίνουν την βέλτιστη σταθερά, για ανισότητες Sobolev.

6.1 Ανισότητα Brunn-Minkowski

Ορισμός 6.1.1. Ένα σύνολο $A \subset \mathbb{R}^n$ λέγεται *κυρτό* αν για κάθε $x, y \in A$ και για κάθε $\lambda \in (0,1)$ ισχύει ότι $(1-\lambda)x + \lambda y \in A$. Το *άθροισμα Minkowski* δύο συνόλων $A, B \subset \mathbb{R}^n$ ορίζεται ως εξής:

$$A + B = \{\alpha + \beta : \alpha \in A, \beta \in B\}$$

και αν $t \geqslant 0$, τότε

$$tA = \{t\alpha : \alpha \in A\}.$$

Παρατηρούμε ότι το άθροισμα δύο κυρτών συνόλων παραμένει κυρτό και ότι ένα σύνολο A είναι κυρτό αν και μόνο αν $(1-\lambda)A + \lambda A = A$ για κάθε $\lambda \in (0,1)$.

Ορισμός 6.1.2. Έστω K ένα συμπαγές υποσύνολο του $\mathbb{R}^n$. Αν $t > 0$ ορίζουμε ως *t-περιοχή του K* το σύνολο

$$K_t = \{y \in \mathbb{R}^n : d(y, K) \leqslant t\} = K + tB_2^n$$

και ως *επιφάνεια του K* την ποσότητα

$$\partial(K) = \liminf_{t \to 0^+} \frac{|K_t| - |K|}{t}$$

όπου με $|K|$ συμβολίζουμε τον όγκο του σώματος K.

Θεώρημα 6.1.3 (*ισοπεριμετρική ανισότητα*). *Μεταξύ όλων των συμπαγών, μη κενών υποσυνόλων του $\mathbb{R}^n$ με δοσμένο όγκο, η μπάλα έχει την μικρότερη επιφάνεια.*

Έχουν δοθεί πολλές αποδείξεις για το θεώρημα αυτό (αρκετές από τις οποίες χρησιμοποιούν την μέθοδο της συμμετρικοποίησης), όμως η πιο σύντομη και αναλυτική είναι αυτή που δίνεται μέσω της ανισότητας Brunn-Minkowski.

Ανισότητα Brunn-Minkowski: Έστω A, B δύο μη κενά, συμπαγή υποσύνολα του $\mathbb{R}^n$. Τότε,

$$|A+B|^{\frac{1}{n}} \geqslant |A|^{\frac{1}{n}} + |B|^{\frac{1}{n}}. \tag{6.1.1}$$

Παρατηρήσεις 6.1.4. (i) Η ανισότητα Brunn-Minkowski μπορεί να γραφτεί στην εξής μορφή: για κάθε $\lambda \in [0,1]$,

$$|(1-\lambda)A + \lambda B|^{\frac{1}{n}} \geqslant (1-\lambda)|A|^{\frac{1}{n}} + \lambda |B|^{\frac{1}{n}} \tag{6.1.2}$$

Πράγματι, έχουμε

$$|(1-\lambda)A + \lambda B|^{\frac{1}{n}} \geqslant |(1-\lambda)A|^{\frac{1}{n}} + |\lambda B|^{\frac{1}{n}} = (1-\lambda)|A|^{\frac{1}{n}} + \lambda |B|^{\frac{1}{n}},$$

όπου η ανισότητα έπεται από την (6.1.1), ενώ η ισότητα από το γεγονός ότι, για κάθε $r \in \mathbb{R}^n$ και για κάθε $A \subset \mathbb{R}^n$ μη κενό, συμπαγές, ισχύει ότι $|rA| = r^n|A|$. Από την (6.1.2) έπεται ότι ο όγκος είναι κοίλη συνάρτηση ως προς την πρόσθεση Minkowski.

(ii) Χρησιμοποιώντας την ανισότητα αριθμητικού-γεωμετρικού μέσου έχουμε

$$|(1-\lambda)A + \lambda B|^{\frac{1}{n}} \geqslant |A^{\frac{1}{n}}|^{1-\lambda} \cdot |B^{\frac{1}{n}}|^{\lambda} = \left(|A|^{1-\lambda} \cdot |B|^{\lambda}\right)^{\frac{1}{n}}$$

απ' όπου έπεται ότι

$$|(1-\lambda)A + \lambda B| \geqslant |A|^{1-\lambda} \cdot |B|^{\lambda} \tag{6.1.3}$$

Το πλεονέκτημα της μορφής (6.1.3) που μπορεί να πάρει η ανισότητα Brunn-Minkowski είναι ότι είναι ανεξάρτητη της διάστασης.

Τώρα, για να δούμε πώς από την (6.1.1) έπεται το Θεώρημα 6.1.3, παρατηρούμε τα εξής:

(i) Έστω $r > 0$ και $t > 0$. Τότε,

$$\begin{aligned} \partial(rB_2^n) &= \lim_{t\to 0^+} \frac{|rB_2^n + tB_2^n| - |rB_2^n|}{t} = \lim_{t\to 0^+} \frac{|(r+t)B_2^n| - |rB_2^n|}{t} \\ &= |B_2^n| \lim_{t\to 0^+} \frac{(r+t)^n - r^n}{t} = |B_2^n| n r^{n-1} \\ &= n|B_2^n| r^{n-1}. \end{aligned}$$

(ii) Έστω $K \subset \mathbb{R}^n$ συμπαγές και $t > 0$ τότε,

$$\frac{|K_t| - |K|}{t} = \frac{|K + tB_2^n| - |K|}{t} \geqslant \frac{\left(|K|^{\frac{1}{n}} + |tB_2^n|^{\frac{1}{n}}\right)^n - |K|}{t}$$
$$\geqslant \frac{|K| + n|K|^{\frac{n-1}{n}}|tB_2^n|^{\frac{1}{n}} - |K|}{t} = \frac{n|K|^{\frac{n-1}{n}}t|B_2^n|^{\frac{1}{n}}}{t},$$

όπου η πρώτη ανισότητα έπεται από την (6.1.1) και η δεύτερη από το διωνυμικό θεώρημα. Τέλος, παίρνοντας $\liminf_{t\to 0^+}$ στην τελευταία σχέση, έχουμε

$$\partial(K) \geqslant n|B_2^n|^{\frac{1}{n}}|K|^{\frac{n-1}{n}}. \tag{6.1.4}$$

Συνεπώς, αν $|K| = r^n|B_2^n|$ τότε η (6.1.4) δίνει

$$\partial(K) \geqslant n|B_2^n|r^{n-1} = \partial(rB_2^n), \tag{6.1.5}$$

το οποίο είναι ακριβώς το συμπέρασμα του Θεωρήματος 6.1.3.

Δεδομένου ότι $\partial(B_2^n) = n|B_2^n|$, μπορούμε να γράψουμε την ισοπεριμετρική ανισότητα στην ισοδύναμη μορφή

$$\left(\frac{\partial(K)}{\partial(B_2^n)}\right)^{\frac{1}{n-1}} \geqslant \left(\frac{|K|}{|B_2^n|}\right)^{\frac{1}{n}} \tag{6.1.6}$$

για κάθε μη κενό, συμπαγές $K \subset \mathbb{R}^n$.

Πριν προχωρήσουμε στην απόδειξη της ανισότητας Brunn-Minkowski με τεχνικές μεταφοράς της μάζας, θα αναφερθούμε σε δύο βασικά εργαλεία που είναι απαραίτητα γι' αυτόν τον σκοπό: την παρεμβολή του MacCann (ή παρεμβολή μέσω μετατόπισης) και την κυρτότητα ως προς μετατόπιση.

Παρεμβολή του McCann. Έστω μ, ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, τα οποία είναι απόλυτα συνεχή ως προς το μέτρο Lebesgue. Τότε, από το Θεώρημα 5.3.8, υπάρχει κυρτή συνάρτηση φ τέτοια ώστε $\nabla\varphi\#\mu = \nu$. Ορίζουμε

$$\rho_t = [\mu, \nu]_t \equiv [(1-t)Id + t\nabla\varphi]\#\mu. \tag{6.1.7}$$

Η οικογένεια των μέτρων πιθανότητας $(\rho_t)_{0\leqslant t\leqslant 1}$ παρεμβάλλεται μεταξύ των μέτρων μ, ν και είναι σαφές ότι

$$[\mu, \nu]_0 = \mu, \ [\mu, \nu]_1 = \nu.$$

Επιπλέον, αφού η $(1-t)Id+t\nabla\varphi = \nabla\left[\frac{(1-t)|\cdot|^2}{2}+t\varphi\right]$ είναι η κλίση μιας κυρτής συνάρτησης για κάθε $t\in[0,1]$, έχουμε ότι το κόστος μεταφοράς από το μ στο $[\mu,\nu]_t$ είναι

$$\begin{aligned}T_2(\mu,\rho_t) &= \int_{\mathbb{R}^n} |x-[(1-t)x+t\nabla\varphi(x)]|^2 d\mu(x)\\ &= t^2\int_{\mathbb{R}^n} |x-\nabla\varphi(x)|^2 d\mu(x) = t^2T_2(\mu,\nu).\end{aligned}$$

Πρόταση 6.1.5 (βασικές ιδιότητες της παρεμβολής)**.** *Έστω μ,ν δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, τα οποία είναι απόλυτα συνεχή ως προς το μέτρο Lebesgue. Για κάθε $t\in[0,1]$, ισχύουν τα ακόλουθα:*

(i) $[\mu,\nu]_t = [\nu,\mu]_{1-t}$.

(ii) $[[\mu,\nu]_t,[\mu,\nu]_{t'}]_s = [\mu,\nu]_{(1-s)t+st'}$.

(iii) *Το μέτρο παρεμβολής $[\mu,\nu]_t$ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue.*

Απόδειξη. Για να αποδείξουμε το (i) αρκεί να παρατηρήσουμε ότι,

$$\begin{aligned}[\mu,\nu]_t &= ((1-t)Id+t\nabla\varphi)\#\mu\\ &= ((1-t)Id+t\nabla\varphi)\#(\nabla\varphi^*\#\nu)\\ &= [((1-t)Id+t\nabla\varphi)\circ\nabla\varphi^*]\,\#\nu\\ &= ((1-t)\nabla\varphi^*+tId)\#\nu\\ &= [\nu,\mu]_{1-t}\end{aligned}$$

Το (ii) επαληθεύεται με απευθείας υπολογισμό.

Για την απόδειξη του (iii) ορίζουμε

$$\varphi_t(x) = t\varphi(x)+(1-t)\frac{|x|^2}{2}$$

και παρατηρούμε ότι

$$\langle\nabla\varphi_t(x)-\nabla\varphi_t(y),x-y\rangle \geqslant (1-t)|x-y|^2.$$

Ειδικότερα,

$$|\nabla\varphi_t(x)-\nabla\varphi_t(y)| \geqslant (1-t)|x-y|. \tag{6.1.8}$$

Τώρα, αφού η φ_t είναι ομοιόμορφα κυρτή, ο μετασχηματισμός Legendre της φ_t^* είναι παντού διαφορίσιμος και από την (6.1.8) έπεται ότι η $\nabla\varphi_t^* = (\nabla\varphi_t)^{-1}$ είναι Lipschitz με σταθερά Lipschitz μικρότερη ή ίση από $\frac{1}{1-t}$. Ειδικότερα, αν A είναι ένα σύνολο μέτρου Lebesgue 0, τότε το $\nabla\varphi_t^*(A)$ έχει επίσης μέτρο Lebesgue 0. Χρησιμοποιώντας λοιπόν το Λήμμα 5.2.5 μπορούμε να γράψουμε

$$\rho_t(A) = \mu(\partial\varphi_t^*(A)) = \mu(\nabla\varphi_t^*(A)) = 0,$$

δηλαδή το $\rho_t = [\mu,\nu]_t$ είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue. □

Κυρτότητα ως προς μετατόπιση. Συμβολίζουμε με $P_{ac}(\mathbb{R}^n)$ το σύνολο όλων των μέτρων πιθανότητας μ στον $\mathbb{R}^n$ τα οποία έχουν πυκνότητα $\rho(x)$ ως προς το μέτρο Lebesgue, δηλαδή $d\mu(x) = \rho(x)dx$. Έστω τώρα $\mu,\nu \in P_{ac}(\mathbb{R}^n)$. Για κάθε $t \in [0,1]$ θεωρούμε το μέτρο παρεμβολής $(\rho_t)_{0\leqslant t\leqslant 1}$ των μ,ν, δηλαδή

$$\rho_t = [(1-t)Id + t\nabla\varphi]\#\mu,$$

όπου η φ είναι κυρτή και $\rho_1 = \nabla\varphi\#\mu = \nu$..

Ορισμός 6.1.6 (*κυρτότητα ως προς μετατόπιση*). (i) Ένα υποσύνολο P του $P_{ac}(\mathbb{R}^n)$ λέγεται *κυρτό ως προς μετατόπιση* αν παραμένει αναλλοίωτο κάτω από την παρεμβολή στοιχείων του μέσω μετατόπισης, δηλαδή

$$\text{για κάθε } \mu,\nu \in P \text{ και για κάθε } t \in [0,1] \Longrightarrow \rho_t = [\mu,\nu]_t \in P.$$

(ii) Έστω F ένα συναρτησοειδές, ορισμένο σε ένα κυρτό ως προς μετατόπιση υποσύνολο P του $P_{ac}(\mathbb{R}^n)$, το οποίο παίρνει τιμές στο $\mathbb{R}\cup\{+\infty\}$. Ένα τέτοιο συναρτησοειδές λέγεται *κυρτό ως προς μετατόπιση* αν έχει την εξής ιδιότητα:

> «Αν $\rho_0 = \mu$ και $\rho_1 = \nu$ είναι δύο στοιχεία του P και $(\rho_t)_{0\leqslant t\leqslant 1}$ τα μέτρα παρεμβολής τους, τότε η $t \mapsto F(\rho_t)$ είναι κυρτή στο $[0,1]$.

Ένα βασικό παράδειγμα συναρτησοειδούς το οποίο είναι κυρτό ως προς μετατόπιση μας δίνει το συναρτησοειδές εσωτερικής ενέργειας

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^n} U(\rho(x))dx,$$

όπου το ρ είναι ένα μέτρο πιθανότητας στον $\mathbb{R}^n$ με πυκνότητα $\rho(x)$, το οποίο είναι απόλυτα συνεχές ως προς το μέτρο Lebesgue και η $U : \mathbb{R}_+ \mapsto \mathbb{R}\cup\{+\infty\}$ είναι μετρήσιμη συνάρτηση, η οποία ονομάζεται *πυκνότητα της εσωτερικής ενέργειας*. Το $\mathcal{U}$ είναι καλά ορισμένο στο $P_{ac}(\mathbb{R}^n)$, με τιμές στο $\mathbb{R}\cup\{+\infty\}$, όταν $U \geqslant 0$. Επιπλέον το $\mathcal{U}$ δεν είναι ταυτοτικά $+\infty$ όταν $U(0) = 0$, και η U δεν είναι ταυτοτικά $+\infty$ στο $\mathbb{R}_+ \setminus \{0\}$.

Θεώρημα 6.1.7 (*κριτήριο κυρτότητας ως προς μετατόπιση*)**.** *Έστω P ένα κυρτό ως προς μετατόπιση υποσύνολο του $P_{ac}(\mathbb{R}^n)$ στο οποίο το συναρτησοειδές $\mathcal{U}$ είναι καλά ορισμένο και παίρνει τιμές στο $\mathbb{R} \cup \{+\infty\}$. Αν $U(0) = 0$ και η*

$$\Psi : r \mapsto r^n U(r^{-n}) \text{ είναι κυρτή και φθίνουσα στο } (0, +\infty),$$

τότε το $\mathcal{U}$ είναι κυρτό ως προς μετατόπιση στο P. Αντίστροφα, αν η Ψ είναι φθίνουσα και το $\mathcal{U}$ είναι κυρτό ως προς μετατόπιση, τότε η Ψ είναι κυρτή.

Μπορούμε τώρα να δούμε την απόδειξη της ανισότητας Brunn-Minkowski μέσω της θεωρίας μεταφοράς μάζας, ως συνέπεια της κυρτότητας ως προς μετατόπιση του συναρτησοειδούς

$$\mathcal{U}(\rho) = -\int_{\mathbb{R}^n} \rho(x)^{1-\frac{1}{n}} dx.$$

Αν μ είναι το ομοιόμορφο μέτρο πιθανότητας επί του X, όπου X συμπαγές μη κενό υποσύνολο του $\mathbb{R}^n$, δηλαδή

$$d\mu(x) = \frac{\mathbf{1}_X dx}{|X|},$$

τότε $\mathcal{U}(\mu) = -|X|^{\frac{1}{n}}$. Η σύνδεση με το άθροισμα Minkowski γίνεται εμφανής με το επόμενο λήμμα.

Λήμμα 6.1.8. *Έστω $\mu = \rho_0$ και $\nu = \rho_1$ τα ομοιόμορφα μέτρα πιθανότητας στα συμπαγή σύνολα X, Y αντίστοιχα. Τότε, για κάθε $t \in [0,1]$, ο φορέας του μέτρου παρεμβολής $\rho_t = [\mu, \nu]_t$ περιέχεται στο άθροισμα Minkowski $(1-t)X + tY$.*

Απόδειξη της ανισότητας Brunn-Minkowski. Έστω $t \in (0,1)$ και έστω S_t ο φορέας του ρ_t. Τότε, το $\frac{\mathbf{1}_{S_t} dx}{|S_t|}$ είναι μέτρο πιθανότητας. Συνεπώς, από την ανισότητα Jensen έχουμε

$$\mathcal{U}(\rho_t) = \int_{S_t} U\Big(\frac{d\rho_t}{dx}\Big) dx \geqslant |S_t| U\Big(\frac{1}{|S_t|}\int d\rho_t\Big) = |S_t| U\Big(\frac{\mathbf{1}_{S_t}}{|S_t|}\Big) = -|S_t|^{\frac{1}{n}}.$$

Όμως, από το Λήμμα 6.1.8 έχουμε

$$-|S_t|^{\frac{1}{n}} \geqslant -|(1-t)X + tY|^{\frac{1}{n}}.$$

Συνεπώς, από την κυρτότητα ως προς μετατόπιση του $\mathcal{U}$ (βλέπε Θεώρημα 6.1.7) έχουμε ότι

$$\mathcal{U}(\rho_t) \leqslant (1-t)\mathcal{U}(\rho_0) + t\mathcal{U}(\rho_1) = (1-t)\mathcal{U}(\mu) + t\mathcal{U}(\nu) = -(1-t)|X|^{\frac{1}{n}} - t|Y|^{\frac{1}{n}},$$

απ' όπου έπεται ότι

$$|(1-t)X + tY|^{\frac{1}{n}} \geqslant (1-t)|X|^{\frac{1}{n}} + t|Y|^{\frac{1}{n}}$$

η οποία είναι ακριβώς η ανισότητα Brunn-Minkowski (βλέπε Παρατηρήσεις 6.1.4 (i)).

6.2 Ανισότητα Prékopa-Leindler

Όπως έχουμε αναφέρει και στην εισαγωγή αυτού του κεφαλαίου η ανισότητα Prékopa-Leindler αποτελεί την συναρτησιακή εκδοχή της ανισότητας Brunn-Minkowski (βλέπε Prékopa [53], Leindler [42]) και διάφορες παραλλαγές της είχαν ήδη αποδειχθεί από διάφορους συγγραφείς στην δεκαετία του 1950.

Θεώρημα 6.2.1 (ανισότητα Prékopa-Leindler). *Έστω f, g, h τρεις μη αρνητικές ολοκληρώσιμες συναρτήσεις στον $\mathbb{R}^n$ και έστω $\lambda \in [0,1]$. Υποθέτουμε ότι για κάθε $x, y \in \mathbb{R}^n$,*

$$h((1-\lambda)x + \lambda y) \geqslant f(x)^{1-\lambda} g(y)^{\lambda}.$$

Τότε,

$$\int_{\mathbb{R}^n} h \geqslant \Big(\int_{\mathbb{R}^n} f \Big)^{1-\lambda} \cdot \Big(\int_{\mathbb{R}^n} g \Big)^{\lambda}. \tag{6.2.1}$$

Πριν προχωρήσουμε στην απόδειξη του Θεωρήματος 6.2.1 με μεθόδους βέλτιστης μεταφοράς, θα αναφερθούμε στην εξίσωση Monge-Ampère και σε ένα βασικό λήμμα που αφορά την ανισότητα αριθμητικού-γεωμετρικού μέσου.

Εξίσωση Monge-Ampère. Έστω $d\mu(x) = f(x)dx$ και $d\nu(y) = g(y)dy$ δύο μέτρα πιθανότητας στον $\mathbb{R}^n$, απολύτως συνεχή ως προς το μέτρο Lebesgue. Από το Θεώρημα 5.2.3 γνωρίζουμε πως υπάρχει μοναδική κυρτή συνάρτηση φ τέτοια ώστε, για κάθε συνάρτηση $\zeta \in C_b(\mathbb{R}^n)$,

$$\int \zeta(y) g(y) dy = \int \zeta(\nabla\varphi(x)) f(x) dx. \tag{6.2.2}$$

Υποθέτουμε ότι η $\nabla\varphi$ είναι C^1 και $1-1$. Κάνοντας την αλλαγή μεταβλητών $y = \nabla\varphi(x)$ στο αριστερό μέλος της (6.2.2), έχουμε

$$\int \zeta(y) g(y) dy = \int \zeta(\nabla\varphi(x)) g(\nabla\varphi(x)) \det(D^2\varphi(x)) dx. \tag{6.2.3}$$

Όμως η συνάρτηση ζ ήταν τυχούσα στον $C_b(\mathbb{R}^n)$, άρα από τις (6.2.2) και (6.2.3) έχουμε

$$f(x) = g(\nabla\varphi(x)) \det(D^2\varphi(x)). \tag{6.2.4}$$

Αν η g είναι θετική τότε μπορούμε να γράψουμε

$$\det(D^2\varphi(x)) = \frac{f(x)}{g(\nabla\varphi(x))}. \tag{6.2.5}$$

Αυτή είναι μία ειδική περίπτωση της εξίσωσης Monge-Ampère η οποία έχει την γενική μορφή

$$\det(D^2\varphi(x)) = F(x, \varphi(x), \nabla\varphi(x)). \tag{6.2.6}$$

Λήμμα 6.2.2 (ανισότητα αριθμητικού-γεωμετρικού μέσου). (i) *Έστω* $(x_i)_{1\leqslant i\leqslant n}$ *και* $(\lambda_i)_{1\leqslant i\leqslant n}$ *πραγματικοί αριθμοί που ικανοποιούν τα εξής,*

$$x_i \geqslant 0,\ \lambda_i \geqslant 0,\ \sum_{i=1}^{n} \lambda_i = 1.$$

Τότε με την σύμβαση $0^0 = 1$ *ισχύει*

$$\sum_{i=1}^{n} \lambda_i x_i \geqslant \prod_{i=1}^{n} x_i^{\lambda_i}.$$

(ii) *Έστω* A *και* B *δύο μη αρνητικοί συμμετρικοί* $n\times n$ *πίνακες και έστω* $\lambda\in[0,1]$. *Τότε,*

$$\det(\lambda A+(1-\lambda)B)^{\frac{1}{n}} \geqslant \lambda(\det A)^{\frac{1}{n}} + (1-\lambda)(\det B)^{\frac{1}{n}}.$$

(ii) *Έστω* A *και* B *δύο μη αρνητικοί συμμετρικοί* $n\times n$ *πίνακες και έστω* $\lambda\in[0,1]$. *Τότε,*

$$\det(\lambda A+(1-\lambda)B) \geqslant (\det A)^{\lambda}(\det B)^{1-\lambda}.$$

Απόδειξη. (i) Η απόδειξη του (i) είναι άμεση συνέπεια του ότι η λογαριθμική συνάρτηση στον $\mathbb{R}_+$ είναι κοίλη.

(ii) Δεδομένου ότι ισχύει η ταυτότητα $\det(\lambda A) = \lambda^n(\det A)$, αρκεί να δείξουμε ότι

$$\det(A+B)^{\frac{1}{n}} \geqslant (\det A)^{\frac{1}{n}} + (\det B)^{\frac{1}{n}}. \tag{6.2.7}$$

Αν δείξουμε την (6.2.7) για την ειδική περίπτωση όπου ο A είναι αντιστρέψιμος τότε η γενική περίπτωση έπεται λόγω πυκνότητας. Υποθέτουμε λοιπόν, ότι ο A είναι αντιστρέψιμος και δεδομένου ότι ισχύει η ταυτότητα $\det(MN) = (\det M)(\det N)$, η ανισότητα (6.2.7) θα είναι συνέπεια της

$$\det(I_n+C)^{\frac{1}{n}} \geqslant (\det I_n)^{\frac{1}{n}} + (\det C)^{\frac{1}{n}},\ \text{όπου } C = A^{-\frac{1}{2}}BA^{-\frac{1}{2}}. \tag{6.2.8}$$

Παρατηρούμε ότι ο C είναι συμμετρικός και μη αρνητικός, άρα αρκεί να δείξουμε την (6.2.8) για τον τυχόντα μη αρνητικό συμμετρικό $n\times n$ πίνακα C. Διαγωνοποιούμε τον C και θεωρούμε τις ιδιοτιμές του $c_1,\dots,c_n$ οι οποίες είναι μη αρνητικές. Τότε, η (6.2.8) παίρνει την μορφή

$$\prod_{i=1}^{n}(1+c_i)^{\frac{1}{n}} \geqslant 1 + \Big(\prod_{i=1}^{n} c_i\Big)^{\frac{1}{n}}.$$

Η τελευταία ανισότητα είναι άμεση συνέπεια του (i) αφού,

$$\prod_{i=1}^{n}\Big(\frac{1}{1+c_i}\Big)^{\frac{1}{n}} + \prod_{i=1}^{n}\Big(\frac{c_i}{1+c_i}\Big)^{\frac{1}{n}} \leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+c_i} + \frac{1}{n}\sum_{i=1}^{n}\frac{c_i}{1+c_i} = 1.$$

(iii) Τέλος, λόγω του (i), ισχύει ότι
$$\lambda(\det A)^{\frac{1}{n}} + (1-\lambda)(\det B)^{\frac{1}{n}} \geqslant (\det A)^{\frac{\lambda}{n}} (\det B)^{\frac{1-\lambda}{n}}.$$
Υψώνοντας την τελευταία ανισότητα στην δύναμη n και χρησιμοποιώντας το (ii) παίρνουμε το ζητούμενο.

□

Μπορούμε τώρα να προβούμε στην απόδειξη του Θεωρήματος 6.2.1, η οποία έχει δοθεί από τον Barthe (βλέπε [10]).

Απόδειξη του Θεωρήματος 6.2.1. Ορίζουμε p να είναι η πυκνότητα Lebesgue στο $[0,1]^n$, υποθέτουμε χωρίς περιορισμό της γενικότητας ότι οι f και g είναι πυκνότητες (έχουν ολοκλήρωμα ίσο με 1), τις ταυτίζουμε με τα αντίστοιχα μέτρα πιθανότητας $f(x)dx$, $g(y)dy$ και εισάγουμε τις απεικονίσεις βέλτιστης μεταφοράς $\nabla\varphi_1$ από το p στο f και $\nabla\varphi_2$ από το p στο g.

Αυτό που πρέπει να δείξουμε είναι ότι $\int h \geqslant 1$. Γράφουμε τις εξισώσεις Monge-Ampère:
$$f(\nabla\varphi_1(x))\det(D_A^2\varphi_1(x)) = 1,\ g(\nabla\varphi_2(x))\det(D_A^2\varphi_2(x)) = 1,$$
σχεδόν για κάθε $x \in [0,1]^n$, όπου D_A είναι η δεύτερη παράγωγος κατά Aleksandrov. Ορίζουμε $\varphi = (1-\lambda)\varphi_1 + \lambda\varphi_2$. Τότε,
$$\begin{aligned}\int_{\mathbb{R}^n} h &\geqslant \int_{[0,1]^n} h(\nabla\varphi(x))\det(D_A^2\varphi(x))\,dx\\ &\geqslant \int_{[0,1]^n} h((1-\lambda)\nabla\varphi_1 + \lambda\nabla\varphi_2)(\det(D_A^2\varphi_1))^{1-\lambda}(\det(D_A^2\varphi_2))^{\lambda}\\ &\geqslant \int_{[0,1]^n} f(\nabla\varphi_1)^{1-\lambda} g(\nabla\varphi_2)^{\lambda}(\det(D_A^2\varphi_1))^{1-\lambda}(\det(D_A^2\varphi_2))^{\lambda}\\ &= \int_{[0,1]^n} 1 = 1,\end{aligned}$$
το οποίο αποδεικνύει το θεώρημα. □

Πολλές γενικεύσεις της ανισότητας Prékopa-Leindler έχουν αποδειχθεί με παρόμοιο τρόπο. Συγκεκριμένα, η ακόλουθη οικογένεια ανισοτήτων έχει δειχθεί (για $\alpha > 0$) από τους Henstock και McBeath και στην γενική περίπτωση από τον Borell και από τους Brascamp και Lieb. Πραγματεύεται μία γενίκευση του αριθμητικού μέσου: για δύο μη αρνητικούς αριθμούς a, b ορίζουμε
$$M_\alpha^\lambda(a,b) = \left\{ \begin{array}{ll} [\lambda a^\alpha + (1-\lambda)b^\alpha]^{\frac{1}{\alpha}}, & \text{αν} \quad a,b>0 \\ 0, & \text{αλλιώς} \end{array} \right\}.$$
Τότε, έχουμε το επόμενο θεώρημα.

Θεώρημα 6.2.3 (ανισότητα Henstock-McBeath)**.** *Έστω f, g, h τρεις μη αρνητικές, ολοκληρώσιμες συναρτήσεις στον $\mathbb{R}^n$, και έστω $\lambda \in [0,1]$ και $\alpha \geqslant -\frac{1}{n}$. Υποθέτουμε ότι για κάθε $x, y \in \mathbb{R}^n$*

$$h(\lambda x + (1-\lambda)y) \geqslant M_\alpha^\lambda(f(x), g(y)).$$

Τότε,

$$\int_{\mathbb{R}^n} h \geqslant M^\lambda_{\frac{\alpha}{1+n\alpha}}\left(\int_{\mathbb{R}^n} f, \int_{\mathbb{R}^n} g\right).$$

Μια απόδειξη αυτού του θεωρήματος μπορεί να γίνει με τρόπο παρόμοιο με εκείνον της απόδειξης του Θεωρήματος 6.2.1, αν αντικαταστήσουμε την ανισότητα αριθμητικού-γεωμετρικού μέσου με την εξής συνέπεια της ανισότητας Hölder: αν $\alpha+\beta \geqslant 0$ και $\frac{1}{\alpha}+\frac{1}{\beta} = \frac{1}{\gamma}$, τότε

$$M_\alpha^\lambda(a,b)M_\beta^\lambda(c,d) \geqslant M_\gamma^\lambda(ac, bd).$$

6.3 Η ανισότητα Brascamp-Lieb και η αντίστροφή της

Στην ενότητα αυτή θα ασχοληθούμε με μία άλλη εφαρμογή της βέλτιστης μεταφοράς, η οποία είναι επίσης στενά συνδεδεμένη με την γεωμετρία, την ανισότητα Brascamp-Lieb. Η ανισότητα αυτή ανήκει στην κατηγορία των «Gaussian ανισοτήτων», μια οικογένεια από ανισότητες στις οποίες οι Gaussian πυκνότητες παίζουν σημαντικό ρόλο (ιδιαίτερα στις περιπτώσεις ισότητας).

Ορισμός 6.3.1 (*Gaussian πυκνότητα*)**.** Gaussian πυκνότητα στον $\mathbb{R}^n$ είναι μια συνάρτηση της μορφής

$$\gamma(x) = \gamma_0 \frac{\exp(-\frac{1}{2}\langle A^{-1}(x-x_0), x-x_0\rangle)}{(2\pi)^{\frac{n}{2}}(\det A)^{\frac{1}{2}}},$$

όπου γ_0 είναι μη αρνητική σταθερά, x_0 ένα διάνυσμα στον $\mathbb{R}^n$ και A ένας θετικά ορισμένος συμμετρικός πίνακας. Η πυκνότητα γ λέγεται *κεντραρισμένη* αν $x_0 = 0$.

Παρατήρηση 6.3.2. Παρατηρούμε ότι $\int \gamma = \gamma_0$. Συνεπώς, όταν $\gamma_0 = 1$ τότε η παραπάνω πυκνότητα γ είναι πυκνότητα πιθανότητας με μέσο το x_0 και πίνακα συνδιακυμάνσεων τον A.

Θα αποδείξουμε την ακόλουθη γενικευμένη μορφή της ανισότητας Brascamp-Lieb, η οποία διατυπώθηκε από τον Lieb (βλέπε [43]).

Θεώρημα 6.3.3 (ανισότητα Brascamp-Lieb)**.** *Έστω n και $(n_i)_{1\leqslant i\leqslant m}$ φυσικοί με $n_i \leqslant n$, $m \geqslant n$, και έστω $(c_i)_{1\leqslant i\leqslant m}$ θετικοί πραγματικοί αριθμοί ώστε*

$$\sum_{i=1}^m c_i n_i = n. \tag{6.3.1}$$

Έστω $B_i : \mathbb{R}^n \to \mathbb{R}^{n_i}$ *γραμμικές απεικονίσεις, οι οποίες είναι επί και ικανοποιούν την*

$$\bigcap_{i=1}^{m} \ker B_i = \{0\}.$$

Θεωρούμε τον τελεστή $I : L_1^+(\mathbb{R}^{n_1}) \times \cdots \times L_1^+(\mathbb{R}^{n_m}) \to \mathbb{R}$ *που ορίζεται ως εξής:*

$$I(f_1, \dots, f_m) = \int_{\mathbb{R}^n} \prod_{i=1}^{m} f_i^{c_i}(B_i x) dx. \tag{6.3.2}$$

Έστω F *η βέλτιστη σταθερά για την οποία ισχύει η ανισότητα*

$$I(f_1, \dots, f_m) \leqslant F \prod_{i=1}^{m} \Big(\int_{\mathbb{R}^{n_i}} f_i \Big)^{c_i}$$

για κάθε επιλογή συναρτήσεων $f_i \in L_1^+(\mathbb{R}^{n_i})$. *Τότε, μπορούμε να προσδιορίσουμε την* F *θεωρώντας μόνο κεντραρισμένες Gaussian συναρτήσεις:*

$$F = F_g = \sup \left\{ \frac{I(g_1, \dots, g_m)}{\prod_{i=1}^{m} (\int_{\mathbb{R}^{n_i}} g_i)^{c_i}} \ \Big| \ g_i \text{ κεντραρισμένες Gaussian συναρτήσεις} \right\}. \tag{6.3.3}$$

Θα παρουσιάσουμε την απόδειξη του Barthe για το Θεώρημα 6.3.3. Η απόδειξη αυτή χρησιμοποιεί δύο βασικά εργαλεία: την βέλτιστη μεταφορά από τη μία πλευρά, και τον δυϊσμό συναρτησιακών ανισοτήτων από την άλλη. Συνέπεια της μεθόδου απόδειξης είναι ότι, ταυτόχρονα, προκύπτει μια άλλη οικογένεια ανισοτήτων (οι λεγόμενες αντίστροφες ανισότητες Brascamp-Lieb, γνωστές και ως ανισότητες του Barthe) οι οποίες γενικεύουν την ανισότητα Prékopa-Leindler και παρουσιάζουν ενδιαφέρον από μόνες τους.

Θεώρημα 6.3.4 (αντίστροφη ανισότητα Brascamp-Lieb)**.** *Με τον συμβολισμό του Θεωρήματος* 6.3.3 *ορίζουμε έναν τελεστή* $K : L_1^+(\mathbb{R}^{n_1}) \times \cdots \times L_1^+(\mathbb{R}^{n_m}) \to \mathbb{R}$ *ως εξής:*

$$K(h_1, \dots, h_m) = \int_{\mathbb{R}^m}^{*} m(x) dx,$$

όπου $\int^*$ *συμβολίζει το εξωτερικό ολοκλήρωμα και*

$$m(x) = \sup \left\{ \prod_{i=1}^{m} h_i^{c_i}(y_i) : \ y_i \in \mathbb{R}^{n_i}, \ \sum_{i=1}^{m} c_i B_i^* y_i = x \right\}.$$

Έστω E *η μεγαλύτερη σταθερά για την οποία ισχύει η ανισότητα*

$$K(h_1, \dots, h_m) \geqslant E \prod_{i=1}^{m} \left(\int_{\mathbb{R}^{n_i}} h_i \right)^{c_i}$$

για κάθε επιλογή συναρτήσεων $h_i \in L_1^+(\mathbb{R}^{n_i})$*. Τότε, μπορούμε να προσδιορίσουμε την* E *θεωρώντας μόνο κεντραρισμένες Gaussian συναρτήσεις, και επιπλέον ισχύει*

$$F \cdot E = 1.$$

Θα αποδείξουμε την ανισότητα Brascamp-Lieb και την αντίστροφή της μαζί, χωρίζοντας την απόδειξη σε βήματα. Στα επόμενα συμβολίζουμε με $S^+(\mathbb{R}^k)$ το σύνολο των $k \times k$ θετικά ορισμένων, συμμετρικών πινάκων. Επιπλέον αν $A \in S^+(\mathbb{R}^k)$ τότε συμβολίζουμε με G_A την κεντραρισμένη Gaussian συνάρτηση $G_A : \mathbb{R}^k \to \mathbb{R}$ η οποία ορίζεται ως εξής:

$$G_A(x) = \exp(-\langle Ax, x\rangle).$$

Με βάση τους συμβολισμούς και τις υποθέσεις των Θεωρήματων 6.3.3 και 6.3.4 έχουμε το εξής θεώρημα, η απόδειξη του οποίου αποδεικνύει ταυτόχρονα την ανισότητα Brascamp-Lieb και την αντίστροφή της.

Θεώρημα 6.3.5. *Μπορούμε να υπολογίσουμε τις σταθερές* E *και* F *χρησιμοποιώντας μόνο κεντραρισμένες Gaussian συναρτήσεις:*

$$E = \inf\left\{ \frac{K(g_1, \dots, g_m)}{\prod_{i=1}^m (\int_{\mathbb{R}^{n_i}} g_i)^{c_i}} : g_i \text{ κεντραρισμένη Gaussian συνάρτηση}, \ i = 1, \dots, m \right\}$$

και

$$F = \sup\left\{ \frac{I(g_1, \dots, g_m)}{\prod_{i=1}^m \int_{\mathbb{R}^{n_i}} g_i)^{c_i}} : g_i \text{ κεντραρισμένη Gaussian συνάρτηση}, \ i = 1, \dots, m \right\}.$$

Επιπλέον, αν D *είναι ο μεγαλύτερος πραγματικός αριθμός για τον οποίο ισχύει*

$$\det\Big(\sum_{i=1}^m c_i B_i^* A_i B_i \Big) \geqslant D \prod_{i=1}^m (\det A_i)^{c_i}$$

για κάθε $A_i \in S^+(\mathbb{R}^{n_i})$*, τότε*

$$E = \sqrt{D} \text{ και } F = \frac{1}{\sqrt{D}}. \tag{6.3.4}$$

Θεωρούμε τις σταθερές

$$E_g = \inf\left\{ \frac{K(g_1, \dots, g_m)}{\prod_{i=1}^m (\int_{\mathbb{R}^{n_i}} g_i)^{c_i}} : g_i \text{ κεντραρισμένη Gaussian συνάρτηση}, \ i = 1, \dots, m \right\}$$

και

$$F_g = \sup\left\{ \frac{I(g_1, \dots, g_m)}{\prod_{i=1}^m \int_{\mathbb{R}^{n_i}} g_i)^{c_i}} : g_i \text{ κεντραρισμένη Gaussian συνάρτηση}, \ i = 1, \dots, m \right\}.$$

Θέλουμε να δείξουμε ότι

$$E = E_g = \sqrt{D} \text{ και } F = F_g = \frac{1}{\sqrt{D}}.$$

Η απόδειξη του Θεωρήματος 6.3.5 θα δοθεί μέσω των δύο επόμενων λημμάτων και της πρότασης που τα ακολουθεί.

Λήμμα 6.3.6. $F_g = \frac{1}{\sqrt{D}}$.

Απόδειξη. Έστω $g_i = G_{A_i}$, $i = 1, \dots, m$, όπου $A_i \in S^+(\mathbb{R}^{n_i})$. Τότε,

$$\begin{aligned} I(g_1, \dots, g_m) &= \int_{\mathbb{R}^n} \prod_{i=1}^m g_i^{c_i}(B_i x) dx \\ &= \int_{\mathbb{R}^n} \exp\Big(- \sum_{i=1}^m c_i \langle A_i B_i x, B_i x\rangle \Big) dx \\ &= \int_{\mathbb{R}^n} \exp\Big(- \Big\langle \Big(\sum_{i=1}^m c_i B_i^* A_i B_i \Big)(x), x \Big\rangle \Big) dx \\ &= \frac{\pi^{\frac{n}{2}}}{\sqrt{\det\left(\sum_{i=1}^m c_i B_i^* A_i B_i\right)}}, \end{aligned}$$

όπου στην τελευταία ισότητα χρησιμοποιήσαμε την ταυτότητα

$$\int e^{-\langle Ax, x\rangle} dx = \frac{\pi^{\frac{n}{2}}}{\sqrt{\det A}}.$$

Από την άλλη πλευρά,

$$\begin{aligned} \prod_{i=1}^m \Big(\int_{\mathbb{R}^{n_i}} g_i \Big)^{c_i} &= \prod_{i=1}^m \Big(\int_{\mathbb{R}^{n_i}} G_{A_i} \Big)^{c_i} = \prod_{i=1}^m \Big(\int_{\mathbb{R}^{n_i}} \exp(-\langle A_i x, x\rangle) \Big)^{c_i} \\ &= \prod_{i=1}^m \Big(\frac{\pi^{\frac{n_i}{2}}}{\sqrt{\det A_i}} \Big)^{c_i} = \frac{\pi^{\frac{n}{2}}}{\sqrt{\prod_{i=1}^m (\det A_i)^{c_i}}}, \end{aligned}$$

όπου στην τελευταία ισότητα χρησιμοποιήσαμε την $\sum_{i=1}^m c_i n_i = n$. Έπεται ότι

$$\begin{aligned} \frac{1}{F_g^2} &= \inf\left\{ \left(\frac{\prod_{i=1}^m (\int_{\mathbb{R}^{n_i}} g_i)^{c_i}}{\int_{\mathbb{R}^n} \prod_{i=1}^m g_i^{c_i}(B_i x) dx} \right)^2 : g_i \text{ κεντραρισμένη Gaussian συνάρτηση} \right\} \\ &= \inf\left\{ \frac{\det(\sum_{i=1}^m c_i B_i^* A_i B_i)}{\prod_{i=1}^m (\det A_i)^{c_i}} : A_i \in S^+(\mathbb{R}^{n_i} \right\} \\ &= D. \end{aligned}$$

□

Λήμμα 6.3.7. Ισχύει ότι $E_g \cdot F_g = 1$ και $E_g = 0$ αν και μόνο αν $F_g = +\infty$. Συνεπώς, $E_g = \sqrt{D}$.

Απόδειξη. Έστω $A_i \in S^+(\mathbb{R}^{n_i})$. Θεωρούμε την τετραγωνική μορφή Q η οποία ορίζεται ως εξής:

$$Q(y) = \Big\langle \sum_{i=1}^{m} c_i B_i^* A_i B_i y, y \Big\rangle,$$

και την συνάρτηση

$$R(x) = \inf \left\{ \sum_{i=1}^{m} c_i \langle A_i^{-1} x_i, x_i \rangle : x_i \in \mathbb{R}^{n_i} \text{ και } x = \sum_{i=1}^{m} c_i B_i^* x_i \right\}.$$

Θα δείξουμε ότι

$$R(x) = Q^*(y) = \sup\{\langle x, y\rangle^2 : Q(y) \leqslant 1\}.$$

Παρατηρούμε ότι αν $x = \sum_{i=1}^{m} c_i B_i^* x_i$, όπου $x_i \in \mathbb{R}^{n_i}$, τότε

$$\langle x, y\rangle^2 = \Big\langle \sum_{i=1}^{m} c_i B_i^* x_i, y \Big\rangle^2 = \Big(\sum_{i=1}^{m} \langle \sqrt{c_i} A_i^{-\frac{1}{2}} x_i, \sqrt{c_i} A_i^{\frac{1}{2}} B_i y\rangle \Big)^2,$$

και εφαρμόζοντας την ανισότητα Cauchy-Schwatz έχουμε

$$\begin{aligned} \langle x, y\rangle^2 &\leqslant \Big(\sum_{i=1}^{m} |\sqrt{c_i} A_i^{-\frac{1}{2}} x_i|^2 \Big) \cdot \Big(\sum_{i=1}^{m} |\sqrt{c_i} A_i^{\frac{1}{2}} B_i y|^2 \Big) \\ &= \Big(\sum_{i=1}^{m} c_i \langle x_i, A_i^{-1} x_i\rangle \Big) \cdot \Big(\Big\langle \sum_{i=1}^{m} c_i B_i^* A_i B_i y, y \Big\rangle \Big) \end{aligned}$$

για κάθε $x_i \in \mathbb{R}^{n_i}$ και για κάθε $y \in \mathbb{R}^n$, άρα

$$\langle x, y\rangle^2 \leqslant R(x) Q(y).$$

Από την άλλη πλευρά, αν επιλέξουμε $y = \Big(\sum_{i=1}^{m} c_i B_i^* A_i B_i \Big)^{-1}(x)$ και $x_i = A_i B_i y$ παίρνουμε $\langle x, y\rangle^2 \geqslant R(x)Q(y)$, δηλαδή έχουμε ισότητα. Πράγματι,

$$\langle x, y\rangle^2 = \Big\langle \sum_{i=1}^{m} c_i B_i^* A_i B_i y, y \Big\rangle^2 = Q(y)^2,$$

οπότε αρκεί να ελέγξουμε ότι $Q(y) \geqslant R(x)$, το οποίο ισχύει αφού

$$R(x) \leqslant \sum_{i=1}^{m} c_i \langle A_i^{-1} x_i, x_i\rangle = \sum_{i=1}^{m} c_i \langle B_i y, A_i B_i y\rangle = \Big\langle \sum_{i=1}^{m} c_i B_i^* A_i B_i y, y \Big\rangle = Q(y).$$

Έπεται ότι $R = Q^*$.

Με έναν άμεσο υπολογισμό έχουμε ότι αν $A_i \in S^+(\mathbb{R}^{n_i})$ τότε

$$\frac{I(G_{A_1}, \ldots, G_{A_m})}{\prod_{i=1}^m (\int G_{A_i})^{c_i}} = \Big(\frac{\prod_{i=1}^m (\det A_i)^{c_i}}{\det Q} \Big)^{\frac{1}{2}}$$

και

$$\frac{K(G_{A_1^{-1}}, \ldots, G_{A_m^{-1}})}{\prod_{i=1}^m (\int G_{A_i})^{c_i}} = \Big(\frac{\prod_{i=1}^m (\det A_i)^{-c_i}}{\det R} \Big)^{\frac{1}{2}}.$$

Επιπλέον, αφού $R = Q^*$, έχουμε $(\det Q) \cdot (\det R) = 1$, συνεπώς,

$$\frac{I(G_{A_1}, \ldots, G_{A_m})}{\prod_{i=1}^m (\int G_{A_i})^{c_i}} \cdot \frac{K(G_{A_1^{-1}}, \ldots, G_{A_m^{-1}})}{\prod_{i=1}^m (\int G_{A_i})^{c_i}} = 1.$$

Από τον ορισμό των E_g και F_g έπεται το ζητούμενο. $\square$

Πρόταση 6.3.8. *Υποθέτουμε ότι οι συναρτήσεις $h_i, f_i \in L_1^+(\mathbb{R}^{n_i})$, $i = 1, \ldots, m$, ικανοποιούν την*

$$\int_{\mathbb{R}^{n_i}} f_i = \int_{\mathbb{R}^{n_i}} h_i = 1.$$

Τότε,

$$K(h_1, \ldots, h_m) \geqslant D \cdot I(f_1, \ldots, f_m).$$

Απόδειξη. Έστω $f_1, \ldots, f_m$ και $h_1, \ldots, h_m$ πυκνότητες πιθανότητας στους $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_m}$ αντίστοιχα (ταυτίζουμε τα μέτρα πιθανότητας $f_i dx_i$, $g_i dy_i$ με τις πυκνότητές τους f_i και g_i). Έχουμε το ακόλουθο διάγραμμα:

$$\mathbb{R}^n \xrightarrow{B_i} \mathbb{R}^{n_i} \xrightarrow{B_i^*} \mathbb{R}^n.$$

Εισάγουμε τις βέλτιστες απεικονίσεις μεταφοράς $T_1, \ldots, T_m$ ώστε $T_i \# f_i = h_i$ και παίρνουμε το εξής διάγραμμα:

$$\mathbb{R}^n \xrightarrow{B_i} x_i \in \mathbb{R}^{n_i} \xrightarrow{T_i} y_i \in \mathbb{R}^{n_i} \xrightarrow{B_i^*} \mathbb{R}^n.$$

Μπορούμε τώρα να γράψουμε τις εξισώσεις Monge-Ampère

$$f_i = (h_i \circ T_i) \det(J(T_i)), \tag{6.3.5}$$

όπου $J(T_i)$ είναι ο Ιακωβιανός πίνακας της T_i. Θεωρούμε την αλλαγή μεταβλητών στον $\mathbb{R}^n$,

$$\Theta(x) = \sum_{i=1}^m c_i B_i^* T_i(B_i x),$$

της οποίας ο ιακωβιανός πίνακας δίνεται από την

$$J(\Theta(x)) = \sum_{i=1}^{m} c_i B_i^* J(T_i(B_i x)) B_i. \tag{6.3.6}$$

Κάθε T_i, ως βέλτιστη απεικόνιση μεταφοράς, έχει την μορφή $T_i = \nabla\varphi_i$, για κάποια κυρτή συνάρτηση φ_i, συνεπώς $J(T_i) = D^2\varphi_i$ και η αλλαγή μεταβλητών Θ είναι μονότονη. Κάτω από κατάλληλες υποθέσεις κανονικότητας και με την χρήση ενός προσεγγιστικού επιχειρήματος, μπορούμε να χρησιμοποιήσουμε την θεωρία κανονικότητας του Caffarelli (βλέπε [59]) από όπου έπεται ότι η Θ είναι αυστηρώς μονότονη C^1 αλλαγή μεταβλητών, η οποία είναι καλά ορισμένη από την $\bigcap_{i=1}^m B_i^{-1}(\mathbb{R}^{n_i})$ στον $\mathbb{R}^n$ (βλέπε [9]). Από τον ορισμό του D (βλέπε (6.3.4)) έχουμε

$$\det(J(\Theta(x)) = \det\Big(\sum_{i=1}^{m} c_i B_i^* J(T_i(B_i x)) B_i\Big) \tag{6.3.7}$$

$$\geqslant D \prod_{i=1}^{m} (\det J(T_i(B_i x)))^{c_i}. \tag{6.3.8}$$

Συνεπώς, έχουμε

$$\begin{aligned} K(h_1,\dots,h_m) &= \int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m h_i^{c_i}(y_i) \ : \ y = \sum_{i=1}^m c_i B_i^* y_i\right\} dy \\ &= \int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m h_i^{c_i}(y_i) \ : \sum_{i=1}^m c_i B_i^* y_i = \Theta(x)\right\} \det J(\Theta(x)) dx \\ &\geqslant D \int_{\mathbb{R}^n} \sup\left\{\prod_{i=1}^m h_i^{c_i}(y_i) \ : \sum_{i=1}^m c_i B_i^* y_i = \Theta(x)\right\} \prod_{i=1}^m (\det(J(T_i(B_i x)))^{c_i} dx, \end{aligned}$$

όπου στην τελευταία ανισότητα χρησιμοποιήσαμε την (6.3.7). Όμως, αν κάθε y_i είναι ίσο με το $T_i(B_i x)$, τότε $\sum_{j=i}^m c_i B_i^* y_i = \Theta(x)$, συνεπώς η τελευταία ποσότητα είναι μεγαλύτερη ή ίση από

$$\begin{aligned} D \int_{\mathbb{R}^n} \prod_{i=1}^m h_i(T_i(B_i x))^{c_i} \prod_{i=1}^m (\det J(T_i(B_i x)))^{c_i} dx \\ &= D \int_{\mathbb{R}^n} \prod_{i=1}^m \Big(h_i \circ T_i(B_i x) \det J(T_i(B_i x))\Big)^{c_i} dx_j \\ &= D \int_{\mathbb{R}^n} \prod_{i=1}^m f_i(B_i x)^{c_i} dx \\ &= D \cdot I(f_1,\dots,f_m), \end{aligned}$$

όπου στην τελευταία ισότητα χρησιμοποιήσαμε τις εξισώσεις Monge-Ampère (6.3.5). Η απόδειξη της πρότασης είναι πλήρης. □

Απόδειξη του θεωρήματος 6.3.5. Είναι σαφές ότι $F \geqslant F_g$ και $E \leqslant E_g$. Συνεπώς, συνδυάζοντας τα Λήμματα 6.3.6, 6.3.7 και την Πρόταση 6.3.8, έχουμε

$$\begin{aligned} F_g = \frac{1}{\sqrt{D}} \leqslant F &\leqslant \sup\left\{ I(f_1,\ldots,f_m) \;:\; \int_{\mathbb{R}^{n_i}} f_i = 1 \right\} \\ &\leqslant \frac{1}{D} \inf\left\{ K(h_1,\ldots,h_m) \;:\; \int_{\mathbb{R}^{n_i}} h_i = 1 \right\} \\ &\leqslant E \cdot \frac{1}{D} \leqslant E_g \frac{1}{D} \leqslant \frac{1}{\sqrt{D}}. \end{aligned}$$

Έχουμε ισότητα παντού, άρα $F = F_g = \frac{1}{\sqrt{D}}$ και $E = E_g = \sqrt{D}$. Η απόδειξη του θεωρήματος είναι πλήρης. □

6.4 Ανισότητες Sobolev

Σε αυτήν την ενότητα θα μελετήσουμε μια οικογένεια ανισοτήτων Sobolev οι οποίες βρίσκονται στο σύνορο της γεωμετρίας με την συναρτησιακή ανάλυση. Για κάθε ακέραιο $n \geqslant 1$ και για κάθε πραγματικό αριθμό $p \geqslant 1$ θεωρούμε τον *χώρο Sobolev*

$$W^{1,p}(\mathbb{R}^n) = \left\{ f \in L^p(\mathbb{R}^n) : \nabla f \in L^p(\mathbb{R}^n) \right\}.$$

Επιπλέον, αν $p \in [1, n)$ τότε ορίζουμε

$$p^* = \frac{np}{n-p}. \tag{6.4.1}$$

Το θεώρημα εμφύτευσης Sobolev $W^{1,p}(\mathbb{R}^n) \subset L^{p^*}(\mathbb{R}^n)$ ισχυρίζεται ότι υπάρχει σταθερά $S_n(p) > 0$ τέτοια ώστε

$$\|f\|_{L^{p^*}} \leqslant S_n(p) \Big(\int_{\mathbb{R}^n} |\nabla f|^p \Big)^{\frac{1}{p}}. \tag{6.4.2}$$

για κάθε $f \in W^{1,p}(\mathbb{R}^n)$. Χωρίς βλάβη της γενικότητας μπορούμε να υποθέσουμε ότι η $S_n(p)$ είναι η βέλτιστη σταθερά με την οποία ικανοποιείται η ανισότητα (6.4.2). Η βέλτιστη σταθερά $S_n(p)$ για $p > 1$ υπολογίστηκε για πρώτη φορά στην δεκαετία του 1960, σε μια μη δημοσιευμένη δουλειά του Rodemich, και έπειτα ανεξάρτητα από τον Aubin (βλέπε [8]) και τον Talenti (βλέπε [61]). Για $p = 1$ ήταν γνωστό πολύ καιρό πριν ότι η βέλτιστη μορφή της (6.4.2) είναι ισοδύναμη με την κλασική Ευκλείδεια ισοπεριμετρική ανισότητα.

Θα πάρουμε τις βέλτιστες σταθερές στις ανισότητες Sobolev ως εφαρμογή των όσων έχουμε αναφέρει σε προηγούμενα κεφάλαια. Η απόδειξη οφείλεται στους Cordero-Erausquin, Nazaret και Villani (βλέπε [18]). Σε σύγκριση με άλλες αποδείξεις, αυτή που θα παρουσιάσουμε είναι πιο στοιχειώδης, εφαρμόζεται για κάθε νόρμα στον $\mathbb{R}^n$, και δεν εμπλέκει μεθόδους συμμετρικοποιήσεων ή μερικές διαφορικές εξισώσεις Euler-Lagrange για σχετικά προβλήματα λογισμού μεταβολών. Επιπλέον, μέσω της απόδειξης αυτής εμφανίζεται απροσδόκητα ένα δυϊκό πρόβλημα, όπως έγινε και με την απόδειξη του Barthe για την ανισότητα Brascamp-Lieb.

Ο χώρος στον οποίο θα δουλέψουμε είναι ο $E = (\mathbb{R}^n, \|\cdot\|)$, όπου $\|\cdot\|$ είναι τυχούσα νόρμα στον $\mathbb{R}^n$. Συνεπώς, ο δυϊκός χώρος θα είναι ο $E^* = (\mathbb{R}^n, \|\cdot\|_*)$, όπου για $X \in E^*$ έχουμε,

$$\|X\|_* = \sup_{\|Y\|\leqslant 1} \langle X, Y\rangle \text{ και } \langle X, Y\rangle = \sum X_i Y_i.$$

Ο δυϊσμός μπορεί να εκφραστεί και μέσω της *ανισότητας Young*:

$$\langle X, Y\rangle \leqslant \frac{\lambda^{-p}}{p}\|X\|_*^p + \frac{\lambda^q}{q}\|Y\|^q \tag{6.4.3}$$

όπου $\lambda > 0$ και $q = \frac{p}{p-1}$ είναι ο συζυγής εκθέτης του p. Για $X : \mathbb{R}^n \to E^*$ στον L^p και $Y : \mathbb{R}^n \to E$ στον L^q, ολοκληρώνοντας την (6.4.3) και βελτιστοποιώντας ως προς λ, παίρνουμε την ανισότητα Hölder στην μορφή,

$$\int \langle X, Y\rangle \leqslant \Big(\int \|X\|_*^p\Big)^{\frac{1}{p}} \Big(\int \|Y\|^q\Big)^{\frac{1}{q}}. \tag{6.4.4}$$

Η ανισότητα αυτή εκφράζει το γεγονός πως ο δυϊκός χώρος του $L^p(\mathbb{R}^n, E)$ ταυτίζεται με τον $L^q(\mathbb{R}^n, E^*)$. Η τυχούσα νόρμα $\|\cdot\|$ είναι Lipschitz, συνεπώς είναι διαφορίσιμη σχεδόν παντού. Άρα, αν $x \in \mathbb{R}^n \setminus \{0\}$ είναι ένα σημείο διαφορισιμότητας της νόρμας, τότε η κλίση της νόρμας στο x είναι το μοναδικά ορισμένο διάνυσμα $x^* = \nabla(\|\cdot\|)(x)$ που ικανοποιεί τις

$$\|x^*\|_* = 1, \text{ και } \langle x, x^*\rangle = \|x\| = \sup_{\|y\|_*=1} \langle x, y\rangle. \tag{6.4.5}$$

Η απόδειξη που θα παρουσιάσουμε κάνει χρήση δύο βασικών ανισοτήτων: της ανισότητας αριθμητικού-γεωμετρικού μέσου και της ανισότητας Young (6.4.3) ή ισοδύναμα της ανισότητας Hölder στην (6.4.4). Το γεγονός ότι γνωρίζουμε τις περιπτώσεις ισότητας σε αυτές τις δύο ανισότητες θα μας δώσει ακριβώς τις περιπτώσεις ισότητας στις ανισότητες Sobolev.

Για $1 \leqslant p < n$ ορίζουμε την συνάρτηση h_p ως εξής:

$$\left\{ \begin{array}{ll} h_p(x) = \frac{1}{(\sigma_p + \|x\|^q)^{\frac{n-p}{p}}}, & \text{αν } p > 1 \\ h_1(x) = \frac{1_B(x)}{|B|^{\frac{n-1}{n}}} & \end{array} \right\} \tag{6.4.6}$$

όπου $q = \frac{p}{p-1}$ είναι ο συζυγής εκθέτης του p, η σταθερά $\sigma_p > 0$ καθορίζεται από την συνθήκη

$$\|h_p\|_{L^{p^*}} = 1 \tag{6.4.7}$$

και B είναι η μοναδιαία μπάλα στον $(\mathbb{R}^n, \|\cdot\|)$,

$$B = \{x \in \mathbb{R}^n : \|x\| \leqslant 1\}.$$

Οι συναρτήσεις h_p είναι αυτές που, όπως θα δούμε, βελτιστοποιούν την ανισότητα Sobolev. Παρατηρούμε ότι η τιμή της σταθεράς σ_p για την οποία ικανοποιείται η (6.4.7) δεν εξαρτάται από την επιλογή της νόρμας, και ότι η h_p δεν βρίσκεται απαραίτητα στον L^p αλλά αυτό δεν έχει ιδιαίτερη σημασία. Είναι φυσικό να αναζητήσουμε τις βέλτιστες συναρτήσεις για την ανισότητα Sobolev στον ομογενή χώρο Sobolev,

$$W^{1,p}(\mathbb{R}^n) = \{f \in L^{p^*}(\mathbb{R}^n) : \nabla f \in L^p(\mathbb{R}^n)\}.$$

Αυτός ο χώρος συμπίπτει με τον χώρο των συναρτήσεων f των οποίων η κλίση (με την έννοια των κατανομών) είναι στοιχείο του L^p και έχουν την ιδιότητα ότι το σύνολο $\{|f| \geqslant \alpha\}$ έχει πεπερασμένο μέτρο για κάθε $\alpha > 0$. Ο χώρος αυτός είναι ομογενής με την ίδια έννοια με την οποία η ανισότητα (6.4.2) είναι ομογενής κάτω από τον μετασχηματισμό $f \to f_\lambda \equiv f(\frac{\cdot}{\lambda})$. Είναι προτιμότερο να χρησιμοποιήσουμε τον χώρο αυτόν για την μελέτη της ανισότητας (6.4.2) από τον $W^{1,p}$: πράγματι για $p > 1$ οι βέλτιστες συναρτήσεις πάντα θα υπάρχουν στον $W^{1,p}$, όμως για $p \geqslant \sqrt{n}$ δεν θα ανήκουν στον $W^{1,p}$.

Θεώρημα 6.4.1 (βέλτιστες ανισότητες Sobolev). *Έστω $p \in (1, n)$. Αν $f, g \in L^{p^*}(\mathbb{R}^n)$ είναι δύο συναρτήσεις τέτοιες ώστε $\|f\|_{L^{p^*}} = \|g\|_{L^{p^*}}$ και $\nabla f \in L^p(\mathbb{R}^n)$, τότε*

$$\frac{\int |g|^{p^*(1-\frac{1}{n})}}{\left(\int |y|^q |g(y)|^{p^*} dy\right)^{\frac{1}{q}}} \leqslant \frac{p(n-1)}{n(n-p)} \|\nabla f\|_{L^p}. \tag{6.4.8}$$

Ισότητα έχουμε όταν $f = g = h_p$.

Άμεσες συνέπειες είναι οι εξής:

(i) *Η αρχή δυϊσμού,*

$$\sup_{\|g\|_{L^{p^*}}=1} \frac{\int |g|^{p^*(1-\frac{1}{n})}}{\left(\int |y|^q |g(y)|^{p^*} dy\right)^{\frac{1}{q}}} = \frac{p(n-1)}{n(n-p)} \inf_{\|f\|_{L^{p^*}}=1} \|\nabla f\|_{L^p} \tag{6.4.9}$$

(ii) *Η βέλτιστη ανισότητα Sobolev: αν $f \neq 0$ με $f \in L^{p^*}(\mathbb{R}^n)$, τότε*

$$\frac{\|\nabla f\|_{L^p}}{\|f\|_{L^{p^*}}} \geqslant \|\nabla h_p\|_{L^p} \tag{6.4.10}$$

(iii) *Η εμφύτευση Sobolev* $W^{1,p}(\mathbb{R}^n) \subset L^{p^*}(\mathbb{R}^n)$, *βλέπε* (6.4.2).

Παρατηρήσεις 6.4.2. (i) Οι βέλτιστες συναρτήσεις για $p = 1$ δεν υπάρχουν στον $W^{1,1}(\mathbb{R}^n)$. Πρέπει να αναζητηθούν στον χώρο των συναρτήσεων με φραγμένη κύμανση.

(ii) Η ανισότητα (6.4.8) παρουσιάζει ενδιαφέρον μόνο όταν $\int \|y\|^q |g(y)|^{p^*} < +\infty$. Τότε, η (6.4.8) αναγκάζει την g να ανήκει στον $L^{p^*(1-\frac{1}{n})}(\mathbb{R}^n)$.

(iii) Η κρίσιμη ιδιότητα της h_p είναι ότι σχεδόν για κάθε x έχουμε ισότητα στην ανισότητα Young (6.4.3) όταν $X = -\nabla h_p(x),\ Y = h_p^{\frac{p^*}{q}}(x)x$ και

$$\lambda = \lambda_p = \Big(\frac{n-p}{p-1}\Big)^{\frac{1}{q}}.$$

Πράγματι, μετά από υπολογισμούς και χρησιμοποιώντας την (6.4.5) οδηγούμαστε στην ισότητα

$$\Big(\frac{n-p}{p-1}\Big)\frac{\|x\|^q}{(\sigma_p + \|x\|^q)^n} = \frac{1}{p\lambda_p^p}\Big(\frac{n-p}{p-1}\Big)^p \frac{\|x\|^q}{(\sigma_p + \|x\|^q)^n} + \frac{\lambda_p^q}{q}\frac{\|x\|^q}{(\sigma_p + \|x\|^q)^n}.$$

Από αυτήν την ισότητα (ή με απευθείας υπολογισμό) βλέπουμε ότι η ίδια επιλογή για τα X και Y μας δίνει ισότητα και στην ανισότητα Hölder (6.4.4):

$$-\int \nabla h_p(x)\cdot (h_p^{\frac{p^*}{q}}(x)x)dx = \|\nabla h_p\|_{L^p}\Big(\int \|x\|^q h_p^{p^*}(x)dx\Big)^{\frac{1}{q}}. \tag{6.4.11}$$

Πριν προχωρήσουμε στην απόδειξη του Θεωρήματος 6.4.1, θα κάνουμε μία μικρή αναφορά στην σύνδεση των εξισώσεων τύπου Laplace με τις εξισώσεις Monge-Ampére και στην διαφορισιμότητα των κυρτών συναρτήσεων.

Οι εξισώσεις Monge-Ampère μπορούν να συνδεθούν με τις εξισώσεις τύπου Laplace, πιο συγκεκριμένα με τις γραμμικές ελλειπτικές εξισώσεις δεύτερης τάξης, της μορφής

$$\sum_{i,j}\alpha_{ij}\frac{\partial^2\varphi}{\partial x_i\partial x_j} + \sum_i \alpha_i\frac{\partial\varphi}{\partial x_i} + c\varphi = h,$$

όπου $(\alpha_{ij}(x))_{1\leqslant i,j\leqslant n}$ είναι μία θετικά ορισμένη συνάρτηση με τιμές τετραγωνικούς πίνακες $n\times n$, $(\alpha_i(x))_{1\leqslant i,j\leqslant n}$ μία συνάρτηση με τιμές διανύσματα και c, h πραγματικές συναρτήσεις. Η εξίσωση Laplace είναι η ειδική περίπτωση

$$\Delta\varphi = \sum_{i=1}^n \frac{\partial^2\varphi}{\partial x_i^2} = h.$$

Για να δούμε την σύνδεση με την εξίσωση Monge-Ampère

$$\det D^2\varphi(x) = F(x, \varphi(x), \nabla\varphi(x)), \tag{6.4.12}$$

αρκεί να σκεφτούμε ότι η $\det D^2\varphi$ είναι το γινόμενο των ιδιοτιμών της Hessian $D^2\varphi$, ενώ η $\Delta\varphi$ είναι το άθροισμα αυτών των ιδιοτιμών. Συνεπώς, τις εξισώσεις Laplace μπορούμε να τις βλέπουμε σαν γραμμικές εκδοχές των εξισώσεων Monge-Ampère. Επιπλέον, λαμβάνοντας υπόψιν την ανισότητα αριθμητικού-γεωμετρικού μέσου, έχουμε

$$(\det D^2\varphi)^{\frac{1}{n}} \leqslant \frac{\Delta\varphi}{n}. \tag{6.4.13}$$

Έστω τώρα φ μία κυρτή συνάρτηση σε έναν μετρικό χώρο X. Τότε, από το θεώρημα του Aleksandrov (βλέπε [4]), η φ είναι αυτομάτως δύο φορές διαφορίσιμη σχεδόν παντού στο $\mathrm{Int}(\mathrm{Dom}\varphi)$. Δηλαδή, σχεδόν για κάθε $x \in \mathrm{Int}(\mathrm{Dom}\varphi))$, έχουμε

$$\varphi(x+h) = \varphi(x) + \nabla\varphi(x)\cdot h + \langle D_A^2\varphi(x)h, h\rangle + o(|h|^2),$$

όπου $D_A^2\varphi$ είναι το απολύτως συνεχές κομμάτι της distributional Hessian $D_{\mathcal{D}'}^2\varphi$, η οποία είναι η γραμμική μορφή που ορίζεται στο $D(\Omega)$ από την ταυτότητα,

$$\langle D_{\mathcal{D}'}^2\varphi, \zeta\rangle = \int_\Omega \zeta D^2\varphi,$$

όπου $\Omega = \mathrm{Int}(\mathrm{Dom}\varphi)$ και είναι $D(\Omega)$ ο χώρος των C^∞ συναρτήσεων με συμπαγή φορέα στο Ω.

Ας προχωρήσουμε τώρα στην απόδειξη του Θεωρήματος 6.4.1.

Απόδειξη. Τα (i) και (ii) έπονται από την (6.4.8) ενώ η (iii) έπεται από την (ii), διότι κάθε συνάρτηση $f \in W^{1,p}$ μπορεί να προσεγγιστεί από συναρτήσεις $f_k \in W^{1,p}\cap L^{p^*}$ έτσι ώστε η $\|\nabla f_k\|_{L^p}$ να συγκλίνει στην $\|\nabla f\|_{L^p}$. Άρα αρκεί να δείξουμε την ανισότητα (6.4.8) για τυχούσες f και g. Επιπλέον, λόγω της ταυτότητας $|\nabla f| = |\nabla|f||$ αρκεί να περιοριστούμε στην περίπτωση όπου η f είναι μη αρνητική, και με ένα επιχείρημα πυκνότητας είναι αρκετό να επιλέξουμε τις f και g ομαλές και με συμπαγείς φορείς. Τέλος, επειδή η ανισότητα είναι ομογενής, μπορούμε να υποθέσουμε ότι $\|f\|_{L^{p^*}} = \|g\|_{L^{p^*}} = 1$.

Εισάγουμε τις πυκνότητες πιθανότητας

$$F(x) = f^{p^*}(x),\ G(y) = g^{p^*}(y)$$

στον $\mathbb{R}^n$. Από το Θεώρημα 5.2.3 γνωρίζουμε ότι η κλίση μίας κυρτής συνάρτησης φ (μοναδικά καθορισμένης σχεδόν παντού) ικανοποιεί την

$$\nabla\varphi\#(Fdx) = Gdy$$

και, επιπλέον, $\mathrm{supp}(Gdy) = \overline{\nabla\varphi(\mathrm{supp}(Fdx))}$. Αρχικά, θα αποδείξουμε ότι

$$\int G^{1-\frac{1}{n}} \leqslant \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta\varphi, \tag{6.4.14}$$

όπου $\Delta\varphi(x) = \mathrm{tr}(D^2\varphi(x))$ είναι το απόλυτα συνεχές κομμάτι της Laplacian $\Delta_{\mathcal{D}'}\varphi$ της φ.

Χρησιμοποιώντας τον ορισμό του προωθητικού μέτρου και την εξίσωση Monge-Ampére

$$F(x) = G(\nabla\varphi(x))\det D^2\varphi(x),$$

έχουμε

$$\begin{aligned}\int G^{1-\frac{1}{n}}dy &= \int G(y)G(y)^{-\frac{1}{n}}dy = \int G(\nabla\varphi(x))\det D_A^2\varphi(x)G(\nabla\varphi(x))^{-\frac{1}{n}}dx \\ &= \int F(x)G(\nabla\varphi(x))^{-\frac{1}{n}}dx = \int F(x)F(x)^{-\frac{1}{n}}(\det D_A^2\varphi(x))^{\frac{1}{n}}dx \\ &\leqslant \frac{1}{n}\int F(x)^{1-\frac{1}{n}}\Delta_A\varphi.\end{aligned} \tag{6.4.15}$$

όπου στην τελευταία ανισότητα χρησιμοποιήσαμε την (6.4.13). Τώρα, αφού η G έχει συμπαγή φορέα, έπεται ότι η $\nabla\varphi$ είναι φραγμένη και η φ μπορεί να επεκταθεί σε μία κυρτή συνάρτηση σε ολόκληρο τον $\mathbb{R}^n$. Επιπλέον, αφού η F είναι ομαλή και έχει συμπαγή φορέα, μπορούμε να γράψουμε

$$\frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_A\varphi \leqslant \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta_{\mathcal{D}'}\varphi = -\frac{1}{n}\int \nabla(F^{1-\frac{1}{n}})\nabla\varphi, \tag{6.4.16}$$

όπου στην πρώτη ανισότητα χρησιμοποιήσαμε την $\Delta_A\varphi \leqslant \Delta_{\mathcal{D}'}\varphi$.

Χρησιμοποιώντας τώρα τον αρχικό συμβολισμό $F = f^{p^*}$ και $G = g^{p^*}$ και συνδυάζοντας τις (6.4.15) και (6.4.16), έχουμε

$$\begin{aligned}\int g^{p^*\frac{n-1}{n}} &= \int g^{\frac{p(n-1)}{n-p}} \\ &\leqslant -\frac{1}{n}\int \langle\nabla(F^{1-\frac{1}{n}}),\nabla\varphi\rangle \\ &= -\frac{1}{n}\int \frac{n-1}{n}F^{-\frac{1}{n}}\langle\nabla F,\nabla\varphi\rangle \\ &= -\frac{n-1}{n^2}\int f^{-\frac{p*}{n}}p^*f^{p^*-1}\langle\nabla f,\nabla\varphi\rangle \\ &= -\frac{n-1}{n-p}\int f^{\frac{n(p-1)}{n-p}}\langle\nabla f,\nabla\varphi\rangle \\ &= -\frac{p(n-1)}{n(n-p)}\int f^{\frac{n(p-1)}{n-p}}\langle\nabla f,\nabla\varphi\rangle \\ &= -\frac{p(n-1)}{n(n-p)}\int f^{\frac{p^*}{q}}\langle\nabla f,\nabla\varphi\rangle,\end{aligned} \tag{6.4.17}$$

όπου έχουμε χρησιμοποιήσει τους ορισμούς των p^* και q που δόθηκαν παραπάνω.

Από την ανισότητα Hölder (βλέπε (6.4.4)) έχουμε

$$-\int f^{\frac{p^*}{q}} \langle \nabla f, \nabla\varphi\rangle \leqslant \|\nabla f\|_{L^p}\Big(\int f^{p^*}|\nabla\varphi|^q\Big)^{\frac{1}{q}}. \tag{6.4.18}$$

Όμως, από τον ορισμό του προωθητικού μέτρου, έχουμε

$$\int f^{p^*}|\nabla\varphi|^q = \int |y|^q g^{p^*}(y)dy.$$

Συνεπώς, συνδυάζοντας τις (6.4.17) και (6.4.18) παίρνουμε την ανισότητα (6.4.8).

Ας επιλέξουμε $f = g = h_p$ τότε παρατηρούμε ότι έχουμε ισότητα σε όλα τα βήματα της απόδειξης, άρα και στην (6.4.8). Βέβαια, η h_p δεν έχει συμπαγή φορέα, όμως στην συγκεκριμένη περίπτωση η απεικόνιση Brenier (βλέπε Παρατήρηση 5.2.4) ανάγεται στην ταυτοτική απεικόνιση $\nabla\varphi = x$ η οποία οδηγεί σε ισότητα στην (6.4.8) και στην (6.4.16) (ολοκληρώνοντας κατά μέρη). Τότε, παρατηρούμε ότι υπάρχει ισότητα και στην (6.4.18), το οποίο ολοκληρώνει την απόδειξη. □

Παρατήρηση 6.4.3. Η επιλογή $f = g = h_p$ γίνεται πιο εύκολα αντιληπτή αν παρατηρήσουμε τις περιπτώσεις ισότητας στην ανισότητα Hölder. Ισότητα στην (6.4.18) συνεπάγεται ότι $\|\nabla f(x)\|^p = kf^{p^*}(x)\|\nabla\varphi(x)\|^q$ σχεδόν για κάθε $x \in \mathbb{R}^n$. Αν τώρα υποθέσουμε ότι $\nabla\varphi(x) = x$, και ψάξουμε για μία ακτινικά συμμετρική συνάρτηση που να δίνει ισότητα στην (6.4.18), θα καταλήξουμε στην h_p.

Μελετώντας κανείς τις περιπτώσεις ισότητας στις παραπάνω ανισότητες Sobolev μπορεί να αποδείξει το εξής θεώρημα.

Θεώρημα 6.4.4 (περιπτώσεις ισότητας στην ανισότητα Sobolev)**.** *Μία συνάρτηση* $f \in L^{p^*}(\mathbb{R}^n)$ *είναι βέλτιστη για την ανισότητα Sobolev* (6.4.10) *αν και μόνο αν υπάρχουν* $C \in \mathbb{R}$, $\lambda \neq 0$ *και* $x_0 \in \mathbb{R}^n$ *ώστε*

$$f(x) = Ch_p(\lambda(x - x_0)).$$

Το θεώρημα αυτό είναι επίσης ανεξάρτητο της νόρμας που επιλέγουμε στον $\mathbb{R}^n$. Η απόδειξη, που βασίζεται στην στρατηγική που ακολουθήθηκε και στο προηγούμενο θεώρημα είναι αρκετά τεχνική (βλέπε [18]), όμως πιο απλή από την κλασική η οποία βασίζεται στην ανισότητα αναδιάταξης ($\|\nabla f^*\|_{L^p} \leqslant \|\nabla f\|_{L^p}$, όπου f^* είναι η μονότονη, ακτινικά συμμετρική αναδιάταξη της f (η ανισότητα αυτή είναι γνωστή ως *αρχή Pólya-Szegö*) και την αναγωγή του προβλήματος στην μία διάσταση.

Κλείνουμε αυτήν την ενότητα με την μορφή που παίρνει η ανισότητα Sobolev όταν $p = 1$. Έχουμε λοιπόν το εξής θεώρημα:

Θεώρημα 6.4.5. *Αν f είναι μία ομαλή συνάρτηση με συμπαγή φορέα, τότε*

$$\frac{\|\nabla f\|_{L^1}}{\|f\|_{L^{\frac{n}{n-1}}}} \geqslant n|B|^{\frac{1}{n}}.$$

Αυτή η ανισότητα επεκτείνεται σε συναρτήσεις με φραγμένη κύμανση. Ισότητα έχουμε όταν $f = h_1$.

Απόδειξη. Χωρίς βλάβη της γενικότητας αποδεικνύουμε το θεώρημα μόνο στην περίπτωση όπου η f είναι μη αρνητική συνάρτηση τέτοια ώστε $\|f\|_{L^{\frac{n}{n-1}}} = 1$. Εισάγουμε την απεικόνιση Brenier $\nabla\varphi$ η οποία προωθεί το μέτρο $F(x)dx = f^{\frac{n}{n-1}}(x)dx$ στο μέτρο $G(y)dy = h_1^{\frac{n}{n-1}}(y)dy$. Ακολουθώντας τον συλλογισμό της απόδειξης του Θεωρήματος 6.4.1, βλέπουμε ότι η σχέση (6.4.14),

$$\int G^{1-\frac{1}{n}} \leqslant \frac{1}{n}\int F^{1-\frac{1}{n}}\Delta\varphi$$

παίρνει την εξής μορφή:

$$|B|^{\frac{1}{n}} \leqslant \frac{1}{n}\int f\Delta\varphi \leqslant -\frac{1}{n}\int \nabla f\cdot\nabla\varphi,$$

όπου B είναι η μοναδιαία μπάλα στον $\mathbb{R}^n$. Πράγματι, χρησιμοποιώντας τον ορισμό της h_1

$$h_1(x) = \frac{1_B(x)}{|B|^{\frac{n-1}{n}}},$$

γράφουμε το αριστερό μέλος της (6.4.14) στην μορφή

$$\int G^{1-\frac{1}{n}} = \int h_1 = \int_B |B|^{\frac{1-n}{n}}dx = |B|^{\frac{1-n}{n}}\cdot|B| = |B|^{\frac{1}{n}},$$

ενώ το δεξιό γράφεται στην μορφή

$$\frac{1}{n}\int F^{1-\frac{1}{n}}\Delta\varphi = \frac{1}{n}\int f\Delta\varphi \leqslant -\frac{1}{n}\int \nabla f\nabla\varphi.$$

Από τον ορισμό της h_1, σχεδόν για κάθε $x\in \mathrm{supp}(f)$, έχουμε $\nabla\varphi(x)\in B$. Ειδικότερα, $-\nabla f\cdot\nabla\varphi \leqslant \|\nabla f\|_*$ συνεπώς,

$$n|B|^{\frac{1}{n}} \leqslant \int \|\nabla f\|_* = \|\nabla f\|_{L^1}. \tag{6.4.19}$$

Με ένα προσεγγιστικό επιχείρημα μπορούμε να εκφράσουμε αυτήν την ανισότητα ως ισοπεριμετρική ανισότητα: αν A κλειστό υποσύνολο του $\mathbb{R}^n$ έχουμε

$$\partial(A) \geqslant n|B|^{\frac{1}{n}}\cdot|A|^{\frac{n-1}{n}}, \tag{6.4.20}$$

όπου $\partial(A)$ το μέτρο επιφάνειας του συνόλου A ως προς την μετρική $\|\cdot\|$ (η οποία δεν είναι απαραίτητα η Ευκλείδεια) που ορίζεται από

$$\partial(A) = \liminf_{\varepsilon\to 0} \frac{|A+\varepsilon B| - |A|}{\varepsilon}.$$

Παρατηρούμε ότι $A+\varepsilon B$ είναι η ε-περιοχή του A ως προς την μετρική $\|\cdot\|$. Τέλος, στην (6.4.20) έχουμε ισότητα όταν το A είναι αφφινική εικόνα του B. Συνεπώς η ανισότητα (6.4.20) πρέπει να είναι βέλτιστη, και το ίδιο πρέπει να ισχύει για την (6.4.19). $\square$

Βιβλιογραφία

[1] Σ. Νεγρεπόντης, Θ. Ζαχαριάδης, Ν. Καλαμίδας και Β. Φαρμάκη, *Γενική Τοπολογία και Συναρτησιακή Ανάλυση.*

[2] Alberti G. and Ambrosio L., *A geometrical approach to monotone functions in* $\mathbb{R}^n$. Math.Z. 230,2(1999), 259-316.

[3] Aleksandrov A.. *Existence and uniqueqness of a convex surface with given integral curvature.* C.R. (Dokl.) Acad. Sci. URSS 35 (1942), 131-134.

[4] Aleksandrov A., *Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected to it.* Ucen. Zap. Leningrad. Gos. Univ. 37 (1939), 3-35.

[5] Ambrosio L., *Lecture notes on optimal transport problems. In mathematical aspects of evolving interfaces.* (Funchal, 2000), vol. 1812 of Lecture notes in Math. Springer, Berlin, 2003, pp. 1-52.

[6] Ambrosio L. and Pratelli A., *Existence and stability results in the* L^1 *theory of optimal transportation.* Lecture Notes in Mathematics Volume 1813, 2003, pp 123-160.

[7] Appell P., *Mémoire sur les déblais et les remblais des systèmes continus ou discontinus.* Mémoires présentés par divers Savants á l Académie des Sciences de l Institut de France, Paris 29 (1887), 1-208.

[8] Aubin T., *Problémes isopérimétriques et espaces de Sobolev.* J. Differential Geometry 11, 4 (1976), 573-598.

[9] Barthe F., *On a reverse form of the Brascamp-Lieb inequality.* Invent. Math. 134,2 (1998), 335-361.

[10] Barthe F., *Inégalités fonctionnelles et géométriques obtenues par transport de measures.* PhD thesis, Univ. Marne-la-Vallée,1997.

[11] Bernard P. and Buffoni B., *Optimal mass transportation and Mather theory.* J. Eur. Math. Soc. (JEMS) 9, 1 (2007), 85-121.

[12] Billingsley P., *Convergence of probability measures*, second ed. John Wiley& Sons Inc.,New York, 1999. A Wiley-Interscience Publication.

[13] Bobylev A. and Toscani G., *On the generalizztion of the Boltzmann H-theorem for a spatially homogeneous Maxwell gas.* J. Math. Phys. 33, 7 (1992), 2578-2586.

[14] Brenier Y., *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math. 44, 4(1991), 375-417.

[15] Brenier Y., *Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris, SérieI, 305 (1987), 805-808.

[16] Burton G., Douglas R.J., *Rearrangements and polar factorization of countably degenerate functions.* Proc. Roy. Soc. Edinburg sect.A 128, 4 (1998), 671-681.

[17] Caffarelli L.A., Feldman M. and McCann R.J., *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs.* L. Amer. Math. Soc. 15, 1 (2002), 1-26.

[18] Cordero-Erausquin, Nazaret D., Villani C., *A mass transportation approach to optimal Sobolev ang Gagliardo-Nirenberg inequalities.* To appear in Adv. Math.

[19] Cullen M. and Purser R.J., *Properties of the Lagrangian semi-geostrophic equations.* J. Atmospheric Sci. 46, 17 (1989), 2684-2697.

[20] Cullen M.J.P. and W. Gangbo, *A variational approach for the 2-dimensianal semigeostrophic shallow water equations.* Arch. Ration. Mech. Anal. 156, 3 (2001), 241-273.

[21] Cullen M. and Feldman M., *Lagrangian solutions of semigeostrophic equations in physical space.* SIAM J. Math. Anal. 37,5 (2006), 1371-1395.

[22] Cullen M.J.P., *A mathematical theory of large-scale atmosphere/ocean flow.* World Scientific, 2006.

[23] Darboux G., *Prix Bordin (géométrie)* C.R. Acad. Sci. 101 (1885), 1312-1316.

[24] Dudley R., *Probabilities and metrics: Convergence of laws on metric spaces, with a viewto statistical testing.* Aarhus Universitet, 1976.

[25] Dudley R., *Real analysis and probability.* Vol.74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[26] Eliassen A., *The quasi-static equations of motion.* Geofys. Publ. 17, 3 (1948).

[27] Evans L.C. and Gariepy R.F., *Measure theory and fine properties of functions.* CRC Press, Boca Ratou, FL, 1992.

[28] Evans L.C., *Partial differential equations and Monge-Kantorovich mass transfer.* In Current developments in mathematics, 1997 (Cambridge, MA). Int. Press, Boston, MA, 1999,pp.65-126.

[29] Evans L.C. and Gangbo W., *Differential equations methods for the Monge-Kantorovich mass transfer problem.* Mem. Amer. Math. Soc. 137, no. 653, 1999.

[30] Graf S., Mauldin R.D., *A classification of disintegrations of measures. In measure and measurable dynamics.* (Rochester, NY, 1987), no. 94 in Contemp. Math. Amer. Math. Soc., Providence, RI, 1989, pp. 147-158.

[31] Hoskins B.J., *Atmospheric frontogenesis models: some solutions.* Q.J.R. Met. Soc. 97 (1971), 139-153.

[32] Hoskins B.J., *The geostrophic momentum approximation and the semi-geostrophicequations.* J. Atmosph. Sciences 32 (1975), 233-242.

[33] Hoskins B.J., *The mathematical theory of frontogenesis.* Ann. Rev. of Fluid Mech. 14 (1982), 131-151.

[34] Jordan R., Kinderlehrer D. and Otto F., *The variational formulation of the Fokker-Planck equation.* SIAM J. Math. Anal., 29 (1998), 1-17.

[35] Kantorovich L.V., *Mathematical methods in the organization and planning of production.* Leningrad Univ., 1939. English translation in Management Science 6, 4 (1960), 363-422.

[36] Kantorovich L.V., *On an effective method of solving certain classes of extremal problems.* Dokl. Akad. Nauk. USSR 38 (1940), 212-215.

[37] Kantorovich L.V., *On the translocation of masses*. Dokl. Akad. Nauk. USSR 37 (1942), 199-201. English translation in J. Math. Sci. 133, 4 (2006), 1381-1382.

[38] Kantorovich L.V., *On a problem of Monge*. Uspekhi Mat. Nauk. 3 (1948), 225-226. English transkation in J. Math. Sci. 133, 4 (2006), 1383.

[39] Kantorovich L.V., *The best use of economic resources*. Oxford Pergamon Press 1965.

[40] Kantorovich L.M., *On the translocation of masses*, C.R.(Docl.) Acad. Sci. URSS 37 (1942), 199-201.

[41] Kantorovich L.M., *On a problem of Monge* (in Russian). Uspekhi Mat. Nauk.3 (1948), 225-226.

[42] L. Leindler, *On a certain converse of Hölderis inequality II*, Acta. Sci. Math. Szeged **33** (1972), 217-223.

[43] Lieb E. H., *Gaussians kernels have only Gaussians maximizers*. Invent.Math. 102, 1 (1990), 179-208.

[44] Mather J.N., *Minimal measures*. Comment. Math. Helv. 64, 3 (1989), 375-394.

[45] Mather J.N., *Action minimizing invariant measures for positive definite Lagrangian systems*. Math. Z. 207,2 (1991), 169-207.

[46] McCann R.J., *Existence and uniqueness of monotone measure-preserving maps*. Duke Math. J. 80, 2 (1995), 309-323.

[47] McCann R.J.,Gangbo W. , *The geometry of optimal transportation*. Acta Math. 177, 2 (1996), 113-161.

[48] McCann R.J., *A convexity principle for interacting gases*. Adv. Math. 128, 1 (1997), 153-179.

[49] McCann R.J., *Exact solutions to the transportation problem on the line*. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Scxi. 455, 1984 (1999), 1341-1380.

[50] Monge G.. *Mémoire sur la théorie des déblais et des remblais*. In Histoire de l Académie Royale des Sciences de Paris (1781), pp.666-704.

[51] Murata H., Tanaka H., *An inequality for certain functional of multidimensional propability distributions*. Hirishima Math. L. 4 (1974), 75-81.

[52] Otto F., *The geometry of dissipative evolution equations: the porous medium equation*. Comm. Partial Differential Equations, 26 (2001), 101-174.

[53] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. Szeged **34** (1973), 335-343.

[54] Pulvirenti A. and Toscani G., *The theory of the nonlinear Boltzmann equation for Maxwell molecules in Fourier representation*. Ann. Mat. pura ed appl. 171, 4 (1996), 181-204.

[55] Rachev S. and Ruschendorf L.. *Mass transportation problems*. Vol.I:Theory, Voll.II:Applications. Probability and its applications. Springer-Verlag, New York, 1998.

[56] Rockafellar R.T.. *Characterization of the subdifferentials of convex functions*. Pacific J. Math 17 (1966), 497-510.

[57] Rockafellar R.T., *Convex analysis*. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[58] Rudin W., *Real and complex analysis*, Mathematics series, McGraw-Hill International Editions, New York (1987)

[59] Villani C., *Topics in Optimal Transportation*, Graduate Texts in Mathematics **58**, Amer. Math. Soc. (2003).

[60] Villani C., *Optimal Transport, Old and New*, A series of comprehensive studies in mathematics **338**, Springer (2006)

[61] Talenti G., *Best constants in Sobolev inequality*. Ann. Mat. Pura Appl. (IV) 110 (1976), 353-372.

[62] Tanaka H., *An inequality for a functional of probability distributions and its application to Kac's one-dimensional model of a Maxwellian gas*. Z. Wahrscheinlichkeitstheorie und Vrw. Gebiete 27 (1973), 47-52.

[63] Tanaka H., *Probabilistic treatment of the Boltzman equation of Maxwellian molecules*. Z. Wahrsch. Verw. Gebiete 46, 1 (1978-1979), 67-105.

[64] Trudinger N.S. and Wang X.J., *On the Monge mass transfer problem*. Calc. Var. Partial Differential equations 13 (2001), 19-31.

[65] Sudakov V.N., *Geometric problems in the theory of infinite-dimensional propability distributions*. Proc. Steclov Inst. Math. 141 (1979), 1-178.