Optimális transzport és Wasserstein-terek

Szakdolgozat

Írta: Szögi Roland

Matematika BSc, matematikus szakirány

Témavezető:

Titkos Tamás
Rényi Alfréd Matematikai Kutatóintézet

Eötvös Loránd Tudományegyetem
Természettudományi Kar
2022

Tartalomjegyzék

Köszönetnyilvánítás

Szeretném megköszönni a témavezetőmnek, Titkos Tamásnak a szakdolgozat megírásához nyújtott segítségét, az iránymutatását és a javaslatait. Továbbá szeretnék köszönetet mondani a családomnak és a barátaimnak a támogatásukért.

1. Bevezető

Az optimális transzport és a Wasserstein-terek elmélete a matematika egy nagyon széles körben alkalmazott területe. Többek között használják a logisztikában, a közgazdaságtanban, a képfeldolgozásban, az alakfelismerésben, a természetes nyelvfeldolgozásban és az idegtudományban is. Az, hogy ilyen sok területen alkalmazzák, annak köszönhető, hogy sok probléma lefordítható úgy, hogy a vizsgált objektumok valószínűségi eloszlások legyenek. Erre látunk is majd egy példát a 9. fejezetben.

Az optimális transzport probléma röviden arról szól, hogy két adott valószínűségi mérték és egy rögzített költségfüggvény mellett keressük a minimális összköltségű transzport leképezést vagy transzport tervet, ami az egyik mértéket a másikba transzformálja. Ez a minimális költség alkalmas arra, hogy egyfajta távolságként tudjuk használni egy rögzített metrikus tér kellőképpen koncentrált valószínűségi mértékei között. Abban a speciális esetben, amikor a szóban forgó mértékek egy (X, ϱ) szeparábilis és teljes metrikus tér véges p-edik momentummal rendelkező valószínűségi mértékei ($p \geq 1$ rögzített valós szám), a költségfüggvény pedig a távolság p-edik hatványa ϱ^p, akkor az optimális transzport tervek egy valódi metrikát határoznak meg, ezt nevezik Wasserstein-távolságnak.

A szakdolgozatban bemutatjuk a transzport probléma Monge- és Kantorovics-féle megfogalmazását, és bebizonyítjuk a téma legfontosabb alaptételeit. A dolgozatot a Wasserstein-távolság bevezetésével, és a Wasserstein-tér néhány szép tulajdonságának bemutatásával zárjuk.

2. Jelölések

- $\mathbb{N}$: a természetes számok halmaza
- $\mathbb{R}$: a valós számok halmaza
- id: az identikus leképezés
- $B_r(x)$: az x körüli r sugarú nyílt gömb
- $\chi(A)$: az A halmaz karakterisztikus függvénye
- $\overline{A}$: az A halmaz komplementere
- λ: a Lebesgue-mérték
- δ_x: az $x \in X$ pontra koncentrált Dirac-mérték
- $\mathrm{supp}(\mu)$: a μ mérték tartója
- $\mu \times \nu$: a μ és ν mértékek szorzata
- $\mu|_A$: a μ mérték megszorítása az A halmazra
- $\mu_n \rightharpoonup \mu$: a μ_n mértékek sorozata gyengén konvergál a μ mértékhez
- $\mu < \nu$: a μ és ν mértékekre teljesül, hogy a $\nu - \mu \not\equiv 0$ pozitív mérték
- $\mathcal{B}(X)$: az X tér Borel-halmazainak σ-algebrája
- $\mathcal{M}_+(X)$: az X tér Borel-mértékeinek halmaza
- $\mathcal{P}(X)$: az X tér Borel valószínűségi mértékeinek halmaza
- $\mathcal{P}_p(X)$: az X téren értelmezett véges p-edik momentumú valószínűségi mértékek halmaza
- $C(X)$: az X téren értélmezett folytonos valós értékű függvények halmaza
- $C_b(X)$: az X téren értelmezett folytonos és korlátos valós értékű függvények halmaza

- $L^1(\mu)$: a μ mérték szerint integrálható függvények halmaza
- π_x: a $\pi_x : X \times Y \to X, \pi(x, y) = x$ természetes vetítés
- π_y: a $\pi_y : X \times Y \to Y, \pi(x, y) = y$ természetes vetítés
- $N(\mu, \sigma^2)$: a μ várható értékű σ^2 szórásnégyzetű normális eloszlás

3. Az optimális transzport probléma Monge-féle megfogalmazása

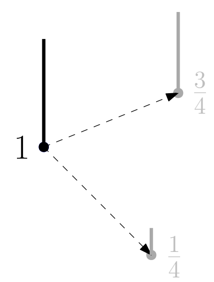
Szemléltetésként először két transzport problémát vizsgálunk meg, majd áttérünk az optimális transzport probléma Monge-féle megfogalmazására. A Monge-probléma megfogalmazása, valamint a hozzá kapcsolódó definíciók és állítások megtalálhatóak a [1] könyv 1.1 fejezetében, illetve a [2] könyv 1.1 fejezetében.

3.1. Bányák és gyárak

Tekintsük a kövtkező feladatot: adva van k darab bánya (ezeket jelölje $x_1, x_2, ..., x_k$) és l dab gyár ($y_1, y_2, ..., y_l$). A bányákban összesen 1 tonna vasérc van, és a gyáraknak összesen pontosan 1 tonna vasércre van szükségük. Jelölje $n(x_i)$ azt, hogy az i-edik bányában hány tonna vasérc van ($n(x_i) \in [0, 1]$), $m(y_j)$ pedig azt, hogy a j-edik gyárnak hány tonna vasércre van szüksége. Minden (x_i, y_j) párra $c(x_i, y_j)$ jelöli az x_i bányából az y_j gyárba törénő szállítás tonnánkénti költségét. Mindegyik x_i bányára igaz, hogy az ott bányászott $n(x_i)$ tonna vasércet nem lehet megosztani a gyárak között, tehát mind az $n(x_i)$ tonnát a gyárak egyikébe kell szállítani. A legolcsóbb összköltségű szállítást keressük.

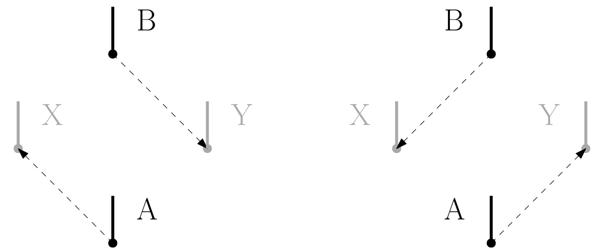
Néhány észrevétel a feladattal kapcsolatban:

- Előfordulhat, hogy a feladatnak nincs megoldása. Például, ha egy bánya van 1 tonna vasérccel, és két gyár $\frac{1}{4}$ és $\frac{3}{4}$ tonna igénnyel, akkor nem létezik a feladat feltételeinek megfelelő szállítás.



1. ábra

- Ha létezik megfelelő szállítás, akkor előfordulhat, hogy a legolcsóbb összköltségű szállítás nem egyértelmű. Például ha két bányánk van fél-fél tonna érccel, két gyár fél-fél tonna igénnyel, és bármely bányából bármely gyárba ugyanannyiba kerül a szállítás, akkor mindegy, hogy melyik bányából melyik gyárba szállítjuk az ércet.



2. ábra

- Mivel véges sok bánya és véges sok gyár esetén csak véges sok lehetséges szállítás van, ezért ha van megfelelő szállítás, akkor van optimális is.

A továbbiakban feltesszük, hogy létezik megfelelő szállítás.
Legyen $X = \{x_1, x_2, ..., x_k\}$ és $Y = \{y_1, y_2, ..., y_l\}$. Ekkor minden szállítást meg tudunk adni egy $T : X \to Y$ leképezéssel azáltal, hogy minden bánya esetén megmondjuk, hogy onnan a vasércet melyik gyárba szállítjuk. Ezeket a leképezéseket fogjuk majd transzport leképezéseknek hívni. Ekkor a T szállítás összköltsége:

$$C(T) := \sum_{i=1}^{k} c(x_i, T(x_i)) \cdot m(x_i).$$

Az a feltétel, hogy minden gyárba pontosan annyi ércet szállítunk, amennyire szükség van, a következőképpen adható meg:

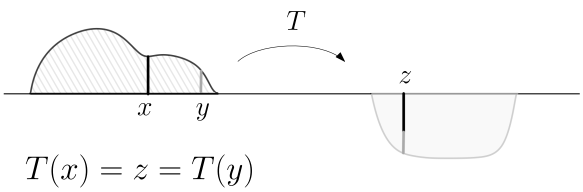
$$m(y_j) = \sum_{T(x_i)=y_j} n(x_i).$$

Ha a bányák és a gyárak száma megegyezik, továbbá minden bányában ugyanannyi érc van, és minden gyárnak ugyanannyi ércre van szüksége, akkor a fenti feladat egy minimális súlyú teljes párosítás keresése.

3.2. Földkupac és gödör

Most tekintsünk egy másik feladatot: adott egy földkupac és egy vele azonos térfogatú gödör. A gödröt szeretnénk feltölteni a földkupaccal a lehető legkevesebb munkával. Ezt a problémát a következőképpen modellezhetjük:

A földkupacnak és a gödörnek egy-egy Borel valószínűségi mértéket feleltetünk meg, ezek legyenek μ és ν. Szemléletesen, tetszőleges $A, B \in \mathcal{B}(\mathbb{R}^2)$ esetén $\mu(A)$ jelöli az A halmaz feletti föld mennyiségét, $\nu(B)$ pedig azt, hogy a gödör B halmaz alatti részének betöméséhez mennyi földre van szükség. Azt is megkötjük, hogy a sík egy x pontja feletti földoszlopot nem oszthatjuk meg különböző $y \neq z$ pontok között. Ekkor minden földpakolást le tudunk írni egy $T : \mathbb{R}^2 \to \mathbb{R}^2$ mérhető függvénnyel.

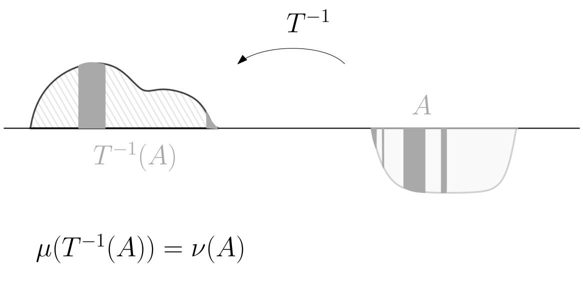


3. ábra

Tetszőleges $A \in \mathcal{B}(\mathbb{R}^2)$ halmaz esetén az A alatti rész térfogata meg kell, hogy egyezzen a $T^{-1}(A)$ feletti rész térfogatával, ahol $T^{-1}(A) = \{x \in \mathbb{R}^2 \,|\, T(x) \in A\}$. Formulával:

$$\mu(T^{-1}(A)) = \nu(A) \qquad \forall A \in \mathcal{B}\left(\mathbb{R}^2\right).$$

Ekkor ν-t a μ mérték T-szerinti előretoltjának, T-t pedig a μ és ν mértékek közötti transzport leképezésnek nevezzük.



4. ábra

A T transzport leképezés c-szerinti költsége:

$$C(T) := \int_{\mathbb{R}^2} c(x, T(x))\, d\mu(x).$$

A diszkrét esettel ellentétben itt nem egyértelmű, hogy ha létezik transzport leképezés, akkor létezik-e minimális költségű, hiszen itt a megfelelő szállítások halmaza nem feltétlenül véges.

3.3. A transzport probléma Monge-féle megfogalmazása

A bevezető példák után rátérünk az optimális transzport probléma Monge-féle megfogalmazására.

Megjegyezzük, hogy az (X, ϱ) metrikus térre sokszor csak X-ként fogunk hivatkozni. Az X tér Borel-halmazainak σ-algebráját $\mathcal{B}(X)$ jelöli, az X tér Borel valószínűségi mértékeinek halmazát pedig $\mathcal{P}(X)$.

3.1. Definíció. *Legyen X és Y két szeparábilis és teljes metrikus tér, $T : X \to Y$ Borel-mérhető függvény, $\mu \in \mathcal{P}(X)$. Ekkor a μ mérték T szerinti előretoltja az a $\nu \in \mathcal{P}(Y)$ mérték, amire*

$$\nu(E) = \mu(T^{-1}(E)) \qquad \forall E \in \mathcal{B}(Y).$$

Ennek a feltételnek a teljesülését így jelöljük: $\nu = T_{\#}\mu$, a T függvényt pedig transzport leképezésnek nevezzük.

3.2. Állítás. *A μ és ν mértékekre $\nu = T_{\#}\mu$ pontosan akkor teljesül, ha*

$$\int_Y \phi(y)\, d\nu(y) = \int_X \big(\phi \circ T\big)(x)\, d\mu(x) \tag{1}$$

igaz minden $\phi : Y \to \mathbb{R} \cup \{\pm\infty\}$ Borel-mérhető függvényre.

Bizonyítás. Először a $\Longleftarrow$ irányt bizonyítjuk. Legyen $A \in \mathcal{B}(Y)$ tetszőleges. Ekkor a $\phi = \chi(A)$ választással kapjuk, hogy

$$\int_Y \big(\chi(A)\big)(y)\, d\nu(y) = \int_X \big(\chi(A) \circ T\big)(x)\, d\mu(x).$$

Felhasználva, hogy

$$\nu(A) = \int_Y \big(\chi(A)\big)(y)\, d\nu(y), \quad \mu(T^{-1}(A)) = \int_X \big(\chi(A) \circ T\big)(x)\, d\mu(x),$$

rögtön adódik, hogy ekkor $\nu(A) = \mu(T^{-1}(A))$. Most belátjuk a $\Longrightarrow$ irányt:
Ha $\phi = \chi(A)$ valamilyen $A \in \mathcal{B}(Y)$ halmazra, és $\nu = T_{\#}\mu$, akkor az fentiek alapján teljesül az (1) egyenlőség. Ekkor nyilvánvalóan az egyszerű függvényekre is teljesül, amiből az integrál definícióját felhasználva rögtön következik, hogy tetszőleges ϕ mérhető függvényre

$$\int_Y \phi(y)\, d\nu(y) = \int_X \big(\phi \circ T\big)(x)\, d\mu(x).$$

□

Most már megfoglamazhatjuk az általános problémát, ami az optimális transzport probléma Monge-féle megfogalmazása. Legyen $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ egy alulról korlátos Borel-mérhető függvény, a μ és ν pedig Borel valószínűségi mértékek az X, illetve Y téren. Keressük a következő értéket:

$$\inf \left\{ \int_X c(x, T(x))\, d\mu(x) \,\middle|\, T : X \to Y, T_{\#}\mu = \nu \right\}. \tag{MP}$$

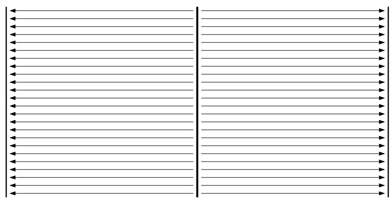
A fenti problémát a továbbiakban Monge-problémának nevezzük. Ha az (MP)-ben szereplő infimum felvétetik egy $\widehat{T}$ függvényre, akkor $\widehat{T}$-ot optimális transzport leképezésnek nevezzük.
Korábban már láttuk, hogy:

- Nem mindig létezik optimális transzport leképezés, hiszen előfodulhat, hogy transzport leképezés sincs.
- Az optimális transzport leképezés nem feltétlenül egyértelmű.
- Ha X és Y is véges, akkor csak véges sok transzport leképezés létezhet, ezért ha létezik transzport leképezés, akkor létezik optimális is.

Ha X és Y nem véges, akkor az is előfordulhat, hogy létezik transzport leképezés, viszont nincs optimális. Legyen $X = Y = \mathbb{R}^2$, $c(x, y) = |x - y|^2$. Legyen μ az a mérték, amit úgy kapunk, hogy 1 súlyt oszlatunk el egyenletesen a $\{0\} \times [0, 1]$ halmazon, a ν pedig az, amikor a $\{-1, 1\} \times [0, 1]$ halmazon oszlatunk el egyenletesen 1 súlyt. Nyilvánvaló, hogy 1-nél kisebb költségű transzport leképezés nem létezhet, hiszen a $\{0\} \times [0, 1]$ szakasz minden pontjától a $\{-1\} \times [0, 1] \cup \{1\} \times [0, 1]$ halmaz minden pontja legalább 1 távolságra van, és így

$$\int_{\{0\}\times[0,1]} |x - T(x)|^2 \, \mathrm{d}\mu(x) \geq \int_{\{0\}\times[0,1]} 1 \, \mathrm{d}\mu(x) = 1$$

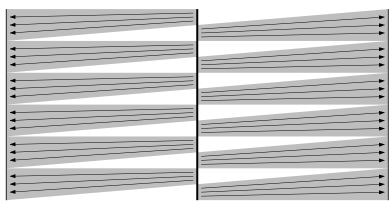
teljesül minden T transzport tervre. Az is érezhető, hogy az optimális az lenne, ha a középen lévő súlyt elfelezhetnénk, és a felét balra, felét pedig jobbra mozgatnánk vízszintesen. Ez azonban nem egy Monge-féle transzport leképezés.



5. ábra

Belátjuk, hogy a transzport leképezések költségének infimuma 1.

Osszuk fel a $\{0\} \times [0, 1]$ szakaszt $2n$ darab egyenlő hosszú szakaszra, a $\{-1\} \times [0, 1]$ és $\{1\} \times [0, 1]$ szakaszokat pedig n darab egyenlő hosszú szakaszra. A T_n képezze a $\{0\} \times [\frac{i-1}{2n}, \frac{i}{2n}]$ szakaszt az $\{1\} \times [\frac{i-1}{2n}, \frac{i+1}{2n}]$ szakaszra, ha i páratlan, a $\{-1\} \times [\frac{i-2}{2n}, \frac{i}{2n}]$ szakaszra, ha i páros, mindezt úgy, hogy az $\frac{1}{2n}$ hosszú szakaszokat a kétszeresükre nyújtjuk.



6. ábra

Nyilvánvalóan T_n egy transzport leképezés minden $n \in \mathbb{N}$-re, továbbá minden $x \in \{0\} \times [0,1]$ pontra $c(x, T_n(x)) \leq \sqrt{1 + \frac{1}{4n^2}}$, és így a T_n transzport leképezés költégére teljesül, hogy $C(T_n) \leq \sqrt{1 + \frac{1}{4n^2}}$, így az infimum valóban 1 lesz.

Belátjuk, hogy nincs optimális transzport leképezés. Indirekt tegyük fel, hogy a $T : \{0\} \times [0,1] \to \{-1,1\} \times [0,1]$ függvény optimális transzport leképezés, vagyis $C(T) = 1$. Ekkor majdnem minden $x \in \{0\} \times [0,1]$ pontra teljesülnie kell a $c(x, T(x)) = 1$ feltételnek, vagyis $T(0,y) = (1,y)$ vagy $T(0,y) = (-1,y)$ teljesül majdnem minden $y \in [0,1]$ esetén. Legyen $A = \{y \,|\, T(0,y) = (1,y)\}$, $B = \{y \,|\, T(0,y) = (-1,y)\}$. Ekkor $\mu(\{0\} \times A) > 0$ és $\mu(\{0\} \times B) > 0$ közül legalább az egyik teljesül, hiszen $\mu(\{0\} \times \overline{A \cup B}) = 0$. Feltehető, hogy $\mu(\{0\} \times A) > 0$. Ekkor:

$$T^{-1}(\{1\} \times A) = \{0\} \times A,\ \mu(\{0\} \times A) = 2 \cdot \nu(\{1\} \times A),$$

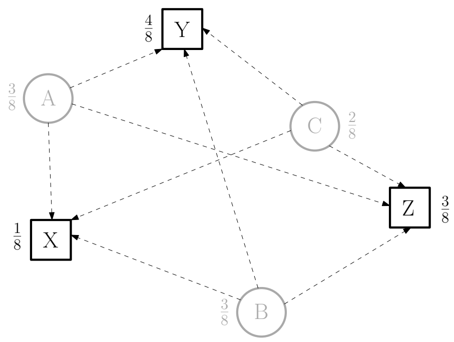
így a $\nu = T_{\#}\mu$ feltétel nem teljesül, ez pedig ellentmondás.

4. Az optimális transzport probléma Kantorovics-féle megfogalmazása

Először a 3.1 és 3.2 alfejezetekben látott példák egy relaxációját vizsgáljuk meg, majd rátérünk az optimális transzport probléma Kantorovics-féle megfogalmazására. Végül belátjuk, hogy a Kantorovics-féle megfogalmazás a Monge-féle megfogalmazás általánosítása. A fejezetben szereplő definíciók és állítások a [1] könyv 1.1, illetve a [2] könyv 1.1 fejezetében találhatóak meg.

4.1. A korábbi példák Kantorovics-féle átfogalmazása

Tekintsük ismét a bányás-gyáros feladatot azzal a különbséggel, hogy most megengedjük, hogy egy bányából több gyárba is lehessen szállítani. Ekkor minden szállítást le tudunk írni egy táblázattal, amelynek i-edik sorának j-edik oszlopában az x_i bányából az y_j gyárba szállított vasérc mennyisége áll.



7. ábra

Például a 7. ábra esetében, amikor három bányánk (A, B, C) és három gyárunk (X, Y, Z) van az ábrán látható ércmennyiségekkel és igényekkel, akkor egy szállítást ír le a következő táblázat:

	X	Y	Z
A	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
B	0	$\frac{3}{8}$	0
C	0	0	$\frac{2}{8}$

A táblázat egy sora tehát azt írja le, hogy az adott bányából melyik gyárba mennyi ércet szállítunk. A táblázat egy oszlopa pedig azt mutatja meg, hogy az adott gyárba melyik bányából mennyi érc érkezik. Ezek alapján egy sorösszeg az adott bányából szállított összes érc mennyiségével egyezik meg, ami tovább egyenlő a bányában lévő összes érc mennyiségével. Hasonlóan egy oszlopösszeg az adott gyárba érkező összes érc mennyiségével egyenlő, ami pontosan az adott gyár igényével egyezik meg. A táblázat egy valószínűségi eloszlást ír le, aminek az egyik marginálisa a bányákban lévő érc mennyiségének eloszlása, a másik marginálisa pedig a gyárak igényeinek eloszlása.

Hasonlóan, ha a földpakolási feladatnál sem követeljük meg, hogy a földoszlopokat ugyanoda kell vinni, egy szállítást le tudunk írni egy $\mathbb{R}^2 \times \mathbb{R}^2$-en értelmezett π valószínűségi mértékkel, ahol tetszőleges $A, B \in \mathcal{B}(\mathbb{R}^2)$ halmazok esetén $\pi(A \times B)$ jelöli azt, hogy az A halmaz feletti földmennyiségből mennyit rakunk a gödör B halmaz alatti részébe. Azt, hogy az A feletti összes földet elraktuk valahova, illetve, hogy a B alatti gödröt teljesen feltöltöttük, így formalizálhatjuk:

$$\pi(A \times \mathbb{R}^2) = \mu(A), \quad \pi(\mathbb{R}^2 \times B) = \nu(B).$$

Ez pontosan azt jelenti, hogy a π mérték marginálisai a μ és ν mértékek. Ezeket a mértékeket transzport terveknek nevezzük. A π transzport terv költsége a c költségfüggvényre vonatkozóan:

$$C(\pi) := \int_{\mathbb{R}^2 \times \mathbb{R}^2} c(x, y)\pi(dx, dy).$$

A cél a minimális költségű transzport terv megtalálása.

4.2. A transzport probléma Kantorovics-féle megfogalmazása

4.1. Definíció. *Legyen X és Y két szeparábilis és teljes metrikus tér, $\mu \in \mathcal{P}(X)$ és $\nu \in \mathcal{P}(Y)$ pedig két Borel valószínűségi mérték. Az $X \times Y$ téren értelmezett π Borel valószínűségi mérték egy transzport terv a μ és ν mértékek között, ha első marginálisa μ, második marginálisa ν, azaz:*

$$\pi(A \times Y) = \mu(A), \quad \pi(X \times B) = \nu(B) \qquad \forall A \in \mathcal{B}(X), B \in \mathcal{B}(Y).$$

A μ és ν mértékek közötti transzport tervek halmazát $\Pi(\mu,\nu)$-vel jelöljük. Megjegyezzük, hogy a $\Pi(\mu,\nu)$ halmaz soha nem üres, mert a $\mu\times\nu$ szorzatmérték mindig egy transzport terv μ és ν között.

Az optimális transzport probléma Kantorovics-féle megfogalmazása: Legyen X és Y két szeparábilis és teljes metrikus tér, a $c : X \times Y \to \mathbb{R}\cup\{+\infty\}$ egy alulról korlátos Borel-mérhető függvény. Legyen $\mu \in \mathcal{P}(X)$ és $\nu \in \mathcal{P}(Y)$ két valószínűségi mérték. Keressük a következő értéket:

$$\inf\left\{\int_{X\times Y} c(x,y)\,d\gamma(x,y)\,|\,\gamma\in\Pi(\mu,\nu)\right\}. \tag{KP}$$

Azokat a transzport terveket, amelyek minimalizálják a fenti kifejezést, optimális transzport terveknek nevezzük.

4.2. Állítás. *A $T : X \to Y$ leképezés esetén jelölje $(id,T) : X \to X\times Y$ azt a függvényt, amire $(id,T)(x) = \big(x,T(x)\big)$ minden $x\in X$ pont esetén. Amennyiben a $T : X\to Y$ Borel-mérhető függvényre teljesül, hogy $\nu = T_\#\mu$, akkor ebből következik, hogy $(id,T)_\#\mu \in \Pi(\mu,\nu)$. Vagyis ha $T : X \to Y$ egy transzport leképezés, akkor $\gamma_T := (id,T)_\#\mu$ egy transzport terv, továbbá ekkor a T transzport leképezés költsége megegyezik a γ_T transzport terv költségével. Formulával:*

$$\int_X c\big(x,T(x)\big)\,d\mu(x) = \int_{X\times Y} c(x,y)\,d\gamma_T(x,y).$$

Ezek alapján az optimális transzport probléma Kantorovics-féle megfogalmazása valóban a Monge-féle megfogalmazás egy relaxációja.

Bizonyítás. Minden $A\in\mathcal{B}(X)$ halmazra $(\mathrm{id},T)^{-1}(A\times Y) = A$, és így igaz a következő: $(id,T)_\#\mu(A\times Y) = \mu\big((\mathrm{id},T)^{-1}(A\times Y)\big) = \mu(A)$. Felhasználva, hogy $(\mathrm{id},T)^{-1}(X\times B) = T^{-1}(B)$, azt kapjuk, hogy

$$(id,T)_\#\mu(X\times B) = \mu\big((\mathrm{id},T)^{-1}(X\times B)\big) = \mu\big(T^{-1}(B)\big) = \nu(B),$$

vagyis ha $\nu = T_\#\mu$, akkor $(id,T)_\#\mu\in\Pi(\mu,\nu)$. A 3.2 állítás alapján abból, hogy $\gamma_T = (\mathrm{id},T)_\#\mu$, következik, hogy tetszőleges $\phi : X\times Y \to \mathbb{R}\cup\{\pm\infty\}$ Borel-mérhető függvényre

$$\int_{X\times Y}\phi(x,y)\,d\gamma_T(x,y) = \int_X \big(\phi\circ(\mathrm{id},T)\big)(x)\,d\mu(x).$$

Ezt a $\phi = c$ függvényre alkalmazva kapjuk, hogy

$$\int_{X\times Y} c(x,y)\, d\gamma_T(x,y) = \int_X c\big(x, T(x)\big)\, d\mu(x).$$

$\square$

Megjegyezzük, hogy egy optimális transzport leképezésből képzett transzport terv nem feltétlenül optimális. Például, ha a metrikus terünk $(\mathbb{R}, |\cdot|)$, a két mértékünk pedig $\frac{2}{3}\cdot\delta_0 + \frac{1}{3}\cdot\delta_1$ és $\frac{1}{3}\cdot\delta_0 + \frac{2}{3}\cdot\delta_1$, akkor egyetlen T transzport leképezés van, a $T(0) = 1$, $T(1) = 0$ függvény. A T-hez tartozó transzport tervnél olcsóbb, ha a 0 pontból viszünk át $\frac{1}{3}$ súlyt az 1 pontba.

5. Az optimális transzport terv létezése

Ebben a fejezetben egy, az optimális transzport terv létezésére vonatkozó tételt látunk be a [1] könyv 1.1, illetve a [2] könyv 1.1 fejezetében ismertetett módon.

Emlékeztetőül, egy X metrikus téren értelmezett $f : X \to \mathbb{R} \cup \{+\infty\}$ függvényt alulról félig folytonosnak nevezünk, ha minden $x_n \to x$ sorozatra teljesül, hogy $f(x) \leq \liminf_{n\to\infty} f(x_n)$. Tetszőleges I indexhalmaz esetén teljesül, hogy ha $\{f_\alpha \,|\, \alpha \in I\}$ az X-ből $\mathbb{R}$-be képező alulról félig folytonos függvények családja, akkor $f = \sup_\alpha f_\alpha$ is alulról félig folytonos ($f(x) := \sup_\alpha f_\alpha(x)\ \forall x \in X$). Továbbá egy $f : X \to \mathbb{R}\cup\{+\infty\}$ alulról korlátos függvény pontosan akkor alulról félig folytonos, ha léteznek olyan f_k k-Lipschitz függvények, amikre minden $x \in X$ esetén $f_k(x)$ monoton növő módon konvergál $f(x)$-hez.

Azt mondjuk, hogy az (X, ϱ) metrikus téren értelmezett $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ Borel valószínűségi mértékek sorozata gyengén konvergál a μ mértékhez, ha minden X-en értelmezett folytonos és korlátos valós értékű φ függvény esetén $\int_X \varphi\, d\mu_n \to \int_X \varphi\, d\mu$ teljesül. A gyenge konvergenciát így jelöljük: $\mu_n \rightharpoonup \mu$. Ismert tény, hogy ha X egy szeparábilis és teljes metrikus tér, akkor a $\mathcal{P}(X)$-en értelmezett gyenge konvergencia metrizálható (például a Lévy-Prohorov metrika metrizálja). Ekkor a gyenge topológiában a kompaktság és a sorozatkompatság ekvivalens fogalmak. Ezek alapján valószínűsági mértékek egy $\mathcal{S} \subseteq \mathcal{P}(X)$ családját gyengén kompaktnak nevezzük, ha bármely $\mathcal{S}$-beli sorozatnak van valószínűségi mértékhez gyengén konvergáló részsorozata. Ez pontosan azt jelenti, hogy $\mathcal{S}$ relatív kompakt a gyenge topológiában, azaz a lezártja kompakt. A valószínűsági mértékek egy $\mathcal{S} \subseteq \mathcal{P}(X)$ családját feszesnek nevezzük, ha minden $\varepsilon > 0$ esetén létezik olyan $K_\varepsilon \subseteq X$ kompakt halmaz, amire $\mu(X \setminus K_\varepsilon) \leq \varepsilon$ teljesül minden $\mu \in \mathcal{S}$ esetén. A Prohorov-tétel alapján szeparábilis és teljes metrikus téren értelmezett valószínűségi mértékek egy családja pontosan akkor feszes, ha gyengén kompakt. Fel fogjuk még használni, hogy minden metrikus téren értelmezett Borel valószínűségi mérték reguláris.

A Weierstrass-tétel azon változatára lesz szükségünk, miszerint egy X kompakt metrikus téren értelmezett alulról félig folytonos $f : X \to \mathbb{R}\cup\{+\infty\}$ függvény esetén létezik olyan $\bar{x} \in X$, amire $f(\bar{x}) = \min\{f(x) : x \in X\}$.

Szükségünk lesz még a monoton konvergencia tételre, ami azt állítja, hogy ha a $f_i : X \to \mathbb{R}_+$ függvények egy $(X, \mathcal{M}, \mu)$ mértéktéren értelmezett nemnegatív mérhető függvények, amelyekre $0 \leq f_1 \leq f_2, ...$, akkor tetszőleges $E \in \mathcal{M}$ halmazra teljesül,hogy

$$\int_E \lim_{n\to\infty} f_n(x)\, d\mu(x) = \lim_{n\to\infty} \int_E f_n(x)\, d\mu(x).$$

A dominált konvergencia tétel azt állítja, hogy ha az $f_i : X \to \mathbb{R}$ függvények egy $(X, \mathcal{M}, \mu)$ mértéktéren értelmezett mérhető függvények, amelyekhez létezik olyan $g \geq 0$ integrálható függvény, amire teljesül, hogy $|f_n| \leq g$ minden $n \in \mathbb{N}$ esetén, és az $(f_n)_{n\in\mathbb{N}}$ sorozat pontonként konvergál az f függvényhez, akkor:

$$\int_X f(x)\, d\mu(x) = \lim_{n\to\infty} \int_X f_n(x)\, d\mu(x).$$

Felhasználjuk még, hogy minden $f : X \to [0, \infty]$ mérhető függvényhez léteznek olyan $0 \leq e_1 \leq e_2 \leq ...$ egyszerű függvények, amelyekre az $(e_n)_{n\in\mathbb{N}}$ függvények monoton növő módon tartanak f-hez.

5.1. Lemma. *Legyen X és Y két metrikus tér, és legyen γ egy Borel valószínűségi mérték az $X \times Y$ téren, aminek első marginálisa μ, második marginális pedig a ν mérték. Ekkor minden $\varphi : X \to [0, \infty]$ Borel-mérhető függvényre*

$$\int_X \varphi(x)\, d\mu(x) = \int_{X\times Y} \varphi(x)\, d\gamma(x, y).$$

Hasonlóan minden $\psi : Y \to [0, \infty]$ Borel-mérhető függvényre

$$\int_Y \psi(y)\, d\nu(y) = \int_{X\times Y} \psi(y)\, d\gamma(x, y).$$

Bizonyítás. Mivel γ első marginális a μ mérték, ezért tetszőleges $A \in \mathcal{B}(X)$ halmazra:

$$\int_X \big(\chi(A)\big)(x)\, d\mu(x) = \mu(A) = \gamma(A \times Y) = \int_{X\times Y} \big(\chi(A)\big)(x)\, d\gamma(x, y).$$

Nyilvánvalóan ekkor minden $e : X \to \mathbb{R}$ egyszerű függvényre is teljesül, hogy

$$\int_X e(x)\, d\mu(x) = \int_{X\times Y} e(x)\, d\gamma(x, y).$$

Tudjuk, hogy minden $\varphi : X \to [0,\infty]$ Borel-mérhető függvényhez léteznek olyan $0 \leq e_1 \leq e_2 ...$ egyszerű függvények, amikre az $(e_n)_{n\in\mathbb{N}}$ sorozat monoton növő módon tart a φ függvényhez. Ekkor a monoton konvergencia tétel alapján

$$\int_X \varphi(x)\, d\mu(x) = \lim_{n\to\infty} \int_X e_n(x)\, d\mu(x)$$
$$\int_{X\times Y} \varphi(x)\, d\gamma(x,y) = \lim_{n\to\infty} \int_{X\times Y} e_n(x)\, d\gamma(x,y).$$

Felhasználva, hogy egyszerű függvényekre már tudjuk, hogy igaz az állítást, ebből már következik, hogy

$$\int_X \varphi(x)\, d\mu(x) = \int_{X\times Y} \varphi(x)\, d\gamma(x,y).$$

Hasonlóan látható be a másik állítás is. □

A φ és ψ függvények nemnegativitása helyett elég, ha alulról korlátosak, hiszen ekkor egy elég nagy szám hozzáadásával nemnegatív függvényt kapunk, amire alkalmazhatjuk a lemmát.

5.2. Lemma. *Ha $f : X \to [0,\infty]$ alulról félig folytonos függvény az X metrikus téren, akkor a $\gamma \to \int_X f(x)\, d\gamma(x)$ funkcionál is alulról félig folytonos lesz az X-en értelmezett valószínűségi mértékek halmazán a mértékek gyenge konvergenciájára nézve.*

Bizonyítás. Mivel f alulról félig folytonos, ezért léteznek olyan f_n folytonos és alulról korlátos függvények, amikre az $(f_n)_{n\in\mathbb{N}}$ sorozat monoton növő módon tart az f-hez. A monoton konvergencia tétel alapján:

$$\int_X f(x)\, d\gamma(x) = \int_X \lim_{n\to\infty} f_n(x)\, d\gamma(x) = \lim_{n\to\infty} \int_X f_n(x)\, d\gamma(x) = \sup_n \left\{ \int_X f_n(x)\, d\gamma(x) \right\}.$$

Mivel az f_n függvények folytonosak, a $\psi_n : \mathcal{P}(X) \to \mathbb{R}$, $\psi_n(\gamma) := \int_X f_n(x)\, d\gamma(x)$ funkcionál folytonos a gyenge konvergenciára nézve minden $n \in \mathbb{N}$ esetén, és így a $\gamma \to \int_X f(x)\, d\gamma(x)$ folytonos funcionálok szuprémuma, ebből adódóan pedig alulról félig folytonos. □

5.1. Tétel. *Legyen X és Y két szeparábilis és teljes metrikus tér, a μ és a ν két Borel valószínűségi mérték az X, illetve Y téren, és legyen a $c : X \times Y \to [0, \infty]$ alulról félig folytonos. Ekkor a Kantorovics-problémában szereplő infimum felvétetik, vagyis létezik optimális transzport terv.*

Bizonyítás. Először belátjuk, hogy $\Pi(\mu, \nu)$ feszes. Legyen $\varepsilon > 0$ rögzített, továbbá legyen $K_X \subseteq X$ és $K_Y \subseteq Y$ két olyan kompakt halmaz, amikre

$$\mu(X \setminus K_X) < \frac{\varepsilon}{2}, \quad \nu(Y \setminus K_Y) < \frac{\varepsilon}{2}.$$

Ilyen K_X, K_Y halmazok mindig léteznek a Prohorov-tétel alapján, hiszen egy egyetlen mértékből álló halmaz mindig gyengén kompakt, és így feszes is. A $K_X \times K_Y$ halmaz kompakt $X \times Y$-ban, és minden $\gamma \in \Pi(\mu, \nu)$ mértékre:

$$\begin{aligned}\gamma\big((X \times Y) \setminus (K_X \times K_Y)\big) &\leq \gamma\big((X \setminus K_X) \times Y\big) + \gamma\big(X \times (Y \setminus K_Y)\big) \\ &= \mu(X \setminus K_X) + \nu(Y \setminus K_Y) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

Tehát a $\Pi(\mu, \nu) \subseteq \mathcal{P}(X \times Y)$ feszes, és így a Prohorov-tétel alapján gyengén kompakt, azaz minden $\Pi(\mu, \nu)$-beli $(\gamma_n)_{n \in \mathbb{N}}$ sorozatnak van valószínűségi mértékhez gyengén konvergáló részsorozata. Belátjuk, hogy a konvergens részsorozat határértéke mindig $\Pi(\mu, \nu)$-beli, vagyis $\Pi(\mu, \nu)$ kompakt a gyenge topológiában. Ehhez elég azt belátni, hogy ha egy $(\gamma_n)_{n \in \mathbb{N}}$ $\Pi(\mu, \nu)$-beli sorozat gyengén konvergál egy γ valószínűségi mértékhez, akkor szükségképpen $\gamma \in \Pi(\mu, \nu)$. Legyen $(\gamma_n)_{n \in \mathbb{N}}$ egy $\Pi(\mu, \nu)$-beli sorozat, amire $\gamma_n \rightharpoonup \gamma$. Legyen $\varphi \in C_b(X)$ tetszőleges függvény. A $\phi : X \times Y \to \mathbb{R}$ $\phi(x, y) := \varphi(x)$ egy folytonos és korlátos függvény, így a gyenge konvergencia definíciója alapján

$$\int\limits_{X \times Y} \varphi(x)\, d\gamma(x, y) = \lim_{n \to \infty} \int\limits_{X \times Y} \varphi(x)\, d\gamma_n(x, y) = \lim_{n \to \infty} \int\limits_{X} \varphi(x)\, d\mu(x) = \int\limits_{X} \varphi(x)\, d\mu(x).$$

Itt a második egyenlőségnél a 5.1 lemmát használtuk. A fenti egyenlőség teljesül minden $\varphi \in C_b(X)$ függvényre. Belátjuk, hogy ebből következik, hogy γ első marginálisa μ. Ehhez először azt látjuk be, hogy minden $F \in \mathcal{B}(X)$ zárt halmazra $\mu(F) = \gamma(F \times Y)$. Legyen $f_n(x) = \max\{1 - n \cdot d(x, F), 0\}$, ahol $d(x, F)$ jelöli az x

pont távolságát az F halmaztól. Az f_n függvények folytonosak, és monoton csökkenő módon tartanak az F karakterisztikus függvényéhez, ha $n \to \infty$. A $g \equiv 1$ függvényre $|f_n| \le g$, és a g függvény μ és γ szerinti integrálja is véges (hiszen μ és γ valószínűségi mértékek), így a dominált konvergencia tétel alapján

$$\mu(F) = \int_X \big(\chi(F)\big)(x)\, d\mu(x) = \lim_{n\to\infty} \int_X f_n(x)\, d\mu(x)$$
$$\gamma(F \times Y) = \int_{X\times Y} \big(\chi(F)\big)(x)\, d\gamma(x,y) = \lim_{n\to\infty} \int_{X\times Y} f_n(x)\, d\gamma(x,y).$$

Mivel az f_n függvények folytonosak, minden $n \in \mathbb{N}$ esetén

$$\int_X f_n(x)\, d\mu(x) = \int_{X\times Y} f_n(x)\, d\gamma(x,y), \tag{2}$$

és így minden $F \in \mathcal{B}(X)$ zárt halmazra

$$\mu(F) = \gamma(F \times Y). \tag{3}$$

Legyen $\mu'(A) := \gamma(A \times Y)$ minden $A \in \mathcal{B}(X)$ esetén. Nyilvánvalóan ez is egy Borel valószínűségi mérték X-en, és az előzőek alapján minden $F \subseteq X$ zárt halmazra $\mu(F) = \mu'(F)$. Mivel a μ és μ' mértékek Borel valószínűségi mértékek az X metrikus téren, ezért regulárisak, vagyis tetszőleges X-beli Borel-halmazt tudunk belülről közelíteni zártakkal. Emiatt tetszőleges $A \in \mathcal{B}(X)$ Borel-halmazra

$$\mu(A) = \mu'(A) = \gamma(A \times Y),$$

vagyis a γ mérték első marginálisa a μ mérték. Hasonlóan látható be, hogy γ második marginálisa pedig ν. Ezzel beláttuk, hogy $\Pi(\mu,\nu)$ kompakt. A lemma alapján a $\gamma \to \int_{X\times Y} c(x,y)\, d\gamma(x,y)$ funkcionál alulról félig folytonos a gyenge konvergenciára nézve, a $\Pi(\mu,\nu)$ kompakt a gyenge topológiában, így a Weierstrass-tétel alapján a funkcionál felveszi a minimumát. □

Megjegyezzük, hogy a c nemnegativitása helyett elég csak azt feltennünk, hogy alulról korlátos. A bizonyítás ugyanúgy működik. Legyen K alsó korlátja c-nek. Ekkor a c_n-eket választhatjuk olyanoknak, hogy K azoknak is alsó korlátja legyen (minden c_n helyett vehetjük a $\max\{c_n, K\}$ függvényt). A $c-K$ és a c_n-K függvények mindegyike pozitív, így ezekre alkalmazhatjuk a monoton konvergencia tételt:

$$\begin{aligned}\int_{X\times Y} \big(c(x,y)-K\big)\, d\gamma(x,y) &= \lim_{n\to\infty} \int_{X\times Y} \big(c_n(x,y)-K\big)\, d\gamma(x,y)\\ \Longrightarrow \int_{X\times Y} c(x,y)\, d\gamma(x,y) - K &= \lim_{n\to\infty} \int_{X\times Y} c_n(x,y)\, d\gamma(x,y) - K\\ \Longrightarrow \int_{X\times Y} c(x,y)\, d\gamma(x,y) &= \lim_{n\to\infty} \int_{X\times Y} c_n(x,y)\, d\gamma(x,y).\end{aligned}$$

Itt felhasználtuk, hogy a γ valószínűségi mérték és így $\int_{X\times Y} K\, d\gamma(x,y) = K$. Felhasználtuk még azt is, hogy abból, hogy γ valószínűségi mérték, az is következik, hogy a negatív részek integrálja véges, és így az integrálok értelmesek.

6. Az optimális transzport alaptétele

Ebben a fejezetben megfogalmazzuk, és belátjuk az optimális transzport alaptételét, ami (bizonyos feltételek mellett) karakterizálja az optimális transzport terveket. A tételeket az [1] könyv 1.2, illetve a [2] könyv 1.6 fejezetében leírt módon bizonyítjuk.

6.1. Definíció. *Legyen $c : X \times Y \to \mathbb{R}$ egy tetszőleges Borel-mérhető függvény. A $\Gamma \subseteq X \times Y$ halmaz c-ciklikusan monoton, ha*

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

teljesül tetszőleges $N \in \mathbb{N}$ és $(x_i, y_i) \in \Gamma$, $i = 1, 2, ...N$ pontok esetén az $\{1, 2, ..., N\}$ halmaz minden σ permutációjára.

6.2. Definíció. *Legyen $c : X \times Y \to \mathbb{R}$ egy tetszőleges Borel-mérhető függvény, a $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ pedig egy tetszőleges függvény. A ψ függvény c_+-transzformáltja az a $\psi^{c+} : X \to \mathbb{R} \cup \{\pm\infty\}$ függvény, amire*

$$\psi^{c+}(x) := \inf_{y \in Y} \{c(x, y) - \psi(y)\}.$$

Egy $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$ függvény c_+-transzformáltja az a $\varphi^{c+} : Y \to \mathbb{R} \cup \{\pm\infty\}$ függvény, amire

$$\varphi^{c+}(y) := \inf_{x \in X} \{c(x, y) - \varphi(x)\}.$$

6.3. Definíció. *Legyen $c : X \times Y \to \mathbb{R}$ egy tetszőleges Borel-mérhető függvény. A $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$ függvény c-konkáv, ha létezik olyan $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ függvény, amire $\varphi = \psi^{c+}$. Hasonlóan, a $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ függvény c-konkáv, ha létezik olyan $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$, amire $\psi = \varphi^{c+}$.*

6.4. Állítás. *A $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ függvény pontosan akkor c-konkáv, ha*

$$\varphi^{c+c+} = \varphi.$$

Bizonyítás. Könnyen belátható, hogy minden $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ függvényre:

$$\psi^{c+c+c+} = \inf_{\tilde{y} \in Y} \sup_{\tilde{x} \in X} \inf_{y \in Y} \{c(x, \tilde{y}) - c(\tilde{x}, \tilde{y}) + c(\tilde{x}, y) - \psi(y)\}.$$

Ebből $\tilde{x} = x$ választással kapjuk, hogy $\psi^{c+c+c+} \geq \psi^{c+}$, az $y = \tilde{y}$ választással pedig azt, hogy $\psi^{c+c+c+} \leq \psi^{c+}$. Tehát minden $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ függvényre teljesül, hogy $\psi^{c+} = \psi^{c+c+c+}$, ebből pedig már nyilvánvaló, hogy φ pontosan akkor c-konkáv, ha $\varphi^{c+c+} = \varphi$. $\square$

6.5. Definíció. *Legyen $c : X \times Y \to \mathbb{R}$ egy tetszőleges Borel-mérhető függvény. A $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ c-konkáv függvény c-szuperdifferenciálja a következő halmaz:*

$$\partial^{c+}\varphi := \{(x, y) \in X \times Y \mid \varphi(x) + \varphi^{c+}(y) = c(x, y)\}.$$

A c-szuperdifferenciál az $x \in X$ pontban:

$$\partial^{c+}\varphi(x) := \{y \in Y \mid (x, y) \in \partial^{c+}\varphi\}.$$

Most megadjuk a c-szuperdifferenciál egy ekvivalens karakterizációját: az $y \in Y$ pontra $y \in \partial^{c+}\varphi(x)$ pontosan akkor teljesül, ha:

$$\varphi(x) = c(x, y) - \varphi^{c+}(y), \quad \varphi(z) \leq c(z, y) - \varphi^{c+}(y) \qquad \forall z \in X. \tag{4}$$

Ugyanis ha $y \in \partial^{c+}\varphi(x)$, akkor a szuperdifferenciál definíciója alapján

$$\varphi(x) = c(x, y) - \varphi^{c+}(y),$$

a φ^{c+} definíciója alapján pedig:

$$\varphi^{c+}(y) = \inf_{x \in X}\{c(x, y) - \varphi(x)\} \leq c(z, y) - \varphi(z) \qquad \forall z \in X.$$

Ebből átrendezéssel kapjuk, hogy:

$$\varphi(z) \leq c(z, y) - \varphi^{c+}(y) \qquad \forall z \in X.$$

A (4) ekvivalens a következővel:

$$\varphi(x) - c(x, y) \geq \varphi(z) - c(z, y) \qquad \forall z \in X, \tag{5}$$

ugyanis (4)-ből rögtön következik, hogy:

$$\varphi(x) - c(x, y) = -\varphi^{c+}(y) \geq \varphi(z) - c(z, y) \qquad \forall z \in X,$$

és (5)-ből következik, hogy:

$$\begin{aligned}
\varphi^{c+}(y) &= \inf_{z\in X}\{c(z,y)-\varphi(z)\} = c(x,y)-\varphi(x),\\
\varphi(z) &\le c(z,y)+\varphi(x)-c(x,y)\\
&= c(z,y)-(c(x,y)-\varphi(x))\\
&= c(z,y)-\varphi^{c+}(y) \qquad\qquad \forall z\in X.
\end{aligned}$$

Vagyis igaz, hogy az $y \in Y$ pontra $y \in \partial^{c+}\varphi(x)$ pontosan akkor teljesül, ha:

$$\varphi(x)-c(x,y) \ge \varphi(z)-c(z,y) \qquad \forall z \in X.$$

6.6. Állítás. *Egy φ c-konkáv függvény c-szuperdifferenciálja egy c-ciklikusan monoton halmaz.*

Bizonyítás. Ha $(x_i, y_i) \in \partial^{c+}\varphi$, akkor:

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

az $1, 2, ..., n$ halmaz minden σ permutációjára. Az első egyenlőség a szuperdfifferenciál definíciójából következik, az utolsó egyenlőtlenség pedig a c_+-transzformált definíciójából, hiszen

$$\begin{aligned}
\varphi(x)+\varphi^{c+}(y) &= \varphi(x)+\inf_{\tilde{x}\in X}(c(\tilde{x},y)-\varphi(\tilde{x})\\
&\le \varphi(x)+c(x,y)-\varphi(x)\\
&= c(x,y).
\end{aligned}$$

□

Az előkészületek után kimondjuk és belátjuk az optimális transzport alaptételét.

6.7. Tétel. *Legyen $c : X \times Y \to \mathbb{R}$ folytonos, alulról korlátos függvény, legyen $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, továbbá teljesüljön, hogy*

$$c(x,y) \le a(x)+b(y)$$

valamilyen $a \in L^1(\mu)$, $b \in L^1(\nu)$ függvényekre. Ekkor ekvivalensek a következők:
i) a γ transzport terv optimális
ii) $\mathrm{supp}(\gamma)$ *c-ciklikusan monoton*
iii) létezik olyan φ c-konkáv függvény, amire $\max\{\varphi, 0\} \in L^1(\mu)$ *és* $\mathrm{supp}(\gamma) \subseteq \partial^{c+}\varphi$.

Bizonyítás. A $c(x,y) \le a(x) + b(y)$ egyenlőtlenség miatt minden $\tilde{\gamma} \in \Pi(\mu,\nu)$ transzport tervre:

$$\int\limits_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y) \le \int\limits_{X\times Y} (a(x)+b(y))\,d\tilde{\gamma}(x,y) = \int\limits_X a(x)\,d\mu(x) + \int\limits_Y b(y)\,d\nu(y) < \infty.$$

Az $\int_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y)$ integrál létezik, mert a c alulról korlátos, a fentiek alapján az integrál véges, azaz c integrálható. Ebből következik, hogy $\max\{c,0\}$ is integrálható, a c alulról korlátosságából pedig az is következik, hogy $c \in L^1(\tilde{\gamma})$ minden $\tilde{\gamma}$ transzport tervre.

(i) $\Longrightarrow$ (ii):

Indirekt tegyük fel, hogy a γ optimális transzport terv, de $\mathrm{supp}(\gamma)$ nem c-ciklikusan monoton. Ez azt jelenti, hogy létezik olyan n természetes szám, $(x_i,y_i) \in \mathrm{supp}(\gamma)$, $i = 1,2,...,n$ pontok és az $\{1,2,....,n\}$ halmaznak egy olyan σ permutációja, amire:

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

Válasszuk úgy az $\varepsilon > 0$ számot, hogy $\varepsilon < \frac{1}{2n}\big(\sum_{i=1}^{n} c(x_i,y_i) - \sum_{i=1}^{2n} c(x_i,y_{\sigma(i)})\big)$ teljesüljön. Ekkor a c folytonossága miatt létezik olyan $r > 0$ szám, hogy minden $i = 1,...,n$ értékre:

$$\begin{aligned} c(x,y) &> c(x_i,y_i) - \varepsilon \qquad \forall (x,y) \in B_r(x_i) \times B_r(y_i) \\ c(x,y) &< c(x_i,y_i) + \varepsilon \qquad \forall (x,y) \in B_r(x_i) \times B_r(y_{\sigma(i)}), \end{aligned}$$

ahol $B_r(x)$ jelöli az x körüli r sugarú nyílt gömböt. Legyen $V_i := B_r(x_i) \times B_r(y_i)$. Ekkor $\gamma(V_i) > 0$, mert $(x_i,y_i) \in \mathrm{supp}(\gamma)$. Definiáljuk a következő mértékeket:

$$\gamma_i := \frac{1}{\gamma(V_i)}\gamma|_{V(i)}, \quad \mu_i := (\pi_x)_{\#}\gamma_i, \nu_i := (\pi_y)_{\#}\gamma_i, \quad \tilde{\gamma}_i := \mu_i \times \nu_{\sigma(i)}.$$

Legyen

$$0 < \varepsilon_0 < \frac{1}{n}\min_i \gamma(V_i), \quad \tilde{\gamma} := \gamma - \varepsilon_0 \cdot \sum_{i=1}^{n} \gamma_i + \varepsilon_0 \cdot \sum_{i=1}^{n} \tilde{\gamma}_i.$$

Belátjuk, hogy $\tilde{\gamma} \in \Pi(\mu,\nu)$, és hogy

$$\int\limits_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y) < \int\limits_{X\times Y} c(x,y)\,d\gamma(x,y),$$

ez pedig ellentmond a feltevésnek, miszerint γ optimális transzport terv μ és ν között. Először azt látjuk be, hogy $\tilde{\gamma}$ pozitív mérték. Az nyilvánvaló, hogy előjeles mérték, azt kell bizonyítani, hogy pozitív is. Ehhez elég a $\gamma - \varepsilon_0 \cdot \sum_{i=1}^n \gamma_i$ mértékről belátni, hogy pozitív, ehhez pedig elég, ha minden i-re $(i = 1, ..., n)$ teljesül, hogy $\varepsilon_0 \cdot \gamma_i < \frac{1}{n} \cdot \gamma$. Az ε_0 definíciója alapján:

$$\varepsilon_0 < \frac{1}{n} \min_i \gamma(V_i) \leq \frac{1}{n}\gamma(V_i) \implies \frac{\varepsilon_0}{\gamma(V_i)} < \frac{1}{n}.$$

Abból, hogy $\frac{\varepsilon_0}{\gamma(V_i)} < \frac{1}{n}$ és $\gamma|_{V_i} \leq \gamma$ következik, hogy:

$$\varepsilon_0 \cdot \gamma_i = \frac{\varepsilon_0}{\gamma(V_i)} \cdot \gamma|_{V_i} < \frac{1}{n} \cdot \gamma.$$

Most belátjuk, hogy $\tilde{\gamma}$ marginálisai μ és ν. Valóban,

$$\begin{aligned}(\pi_x)_\#\tilde{\gamma} &= (\pi_x)_\#(\gamma - \varepsilon_0 \cdot \sum_{i=1}^n \gamma_i + \varepsilon_0 \cdot \sum_{i=1}^n \tilde{\gamma}_i) \\ &= (\pi_x)_\#\gamma - \varepsilon_0 \cdot \sum_{i=1}^n (\pi_x)_\#\gamma_i + \varepsilon_0 \cdot \sum_{i=1}^n (\pi_x)_\#\tilde{\gamma}_i \\ &= \mu - \varepsilon_0 \cdot \sum_{i=1}^n \mu_i + \varepsilon_0 \cdot \sum_{i=1}^n \mu_i \\ &= \mu\end{aligned}$$

és

$$\begin{aligned}(\pi_y)_\#\tilde{\gamma} &= (\pi_y)_\#(\gamma - \varepsilon_0 \cdot \sum_{i=1}^n \gamma_i + \varepsilon_0 \cdot \sum_{i=1}^n \tilde{\gamma}_i) \\ &= (\pi_y)_\#\gamma - \varepsilon_0 \cdot \sum_{i=1}^n (\pi_y)_\#\gamma_i + \varepsilon_0 \cdot \sum_{i=1}^n (\pi_y)_\#\tilde{\gamma}_i \\ &= \nu - \varepsilon_0 \cdot \sum_{i=1}^n \nu_i + \varepsilon_0 \cdot \sum_{i=1}^n \nu_{\sigma(i)} \\ &= \nu.\end{aligned}$$

Végül belátjuk, hogy fennáll az

$$\int\limits_{X\times Y} c(x,y)\, d\tilde{\gamma}(x,y) < \int\limits_{X\times Y} c(x,y)\, d\gamma(x,y)$$

egyenlőtlenség. Valóban,

$$
\begin{aligned}
\int\limits_{X\times Y} c(x,y)\,d\gamma(x,y) - \int\limits_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y)& \\
&= \varepsilon_0 \cdot \sum_{i=1}^{n} \int\limits_{X\times Y} c(x,y)\,d\gamma_i(x,y) - \varepsilon_0 \cdot \sum_{i=1}^{n} \int\limits_{X\times Y} c(x,y)\,d\tilde{\gamma_i}(x,y) \\
&\geq \varepsilon_0 \cdot \sum_{i=1}^{n} (c(x_i,y_i) - \varepsilon) - \varepsilon_0 \cdot \sum_{i=1}^{n} (c(x_i, y_{\sigma(i)} + \varepsilon) \\
&= \varepsilon_0 \cdot (\sum_{i=1}^{n} c(x_i,y_i) - \sum_{i=1}^{n} c(x_i, y_{\sigma(i)} - 2\cdot n\cdot\varepsilon) \\
&> 0.
\end{aligned}
$$

Itt felhasználtuk, hogy a γ_i mértékek a $B_r(x_i) \times B_r(y_i)$ halmazokra vannak koncentrálva, a $\tilde{\gamma_i}$ mértékek pedig a $B_r(x_i) \times B_r(y_{\sigma(i)})$ halmazokra, továbbá hogy

$$
\begin{aligned}
c(x,y) &> c(x_i,y_i) - \varepsilon \qquad \forall (x,y) \in B_r(x_i) \times B_r(y_i) \\
c(x,y) &< c(x_i,y_i) + \varepsilon \qquad \forall (x,y) \in B_r(x_i) \times B_r(y_{\sigma(i)}).
\end{aligned}
$$

Azt is felhasználtuk, hogy a γ_i, illetve $\tilde{\gamma_i}$ mértékek valószínűségi mértékek, és hogy az ε-t úgy választottuk, hogy

$$
\varepsilon < \frac{1}{2n}\left(\sum_{i=1}^{n} c(x_i,y_i) - \sum_{i=1}^{2n} c(x_i, y_{\sigma(i)})\right)
$$

teljesüljön.

(ii) $\Longrightarrow$ (iii):

Azt kell belátnunk, hogy ha $\Gamma \subseteq X\times Y$ egy ciklikusan monoton halmaz, akkor létezik olyan c-konkáv φ függvény, amire teljesül, hogy $\Gamma \subseteq \partial^{c+}\varphi$ és $\max\{\varphi, 0\} \in L^1(\mu)$. Rögzítsünk egy $(\bar{x},\bar{y}) \in \Gamma$ pontot. Mivel olyan φ függvényt akarunk, ami c-konkáv, és a c-szuperdifferenciálja tartalmazza Γ-t, tetszőleges $N \in \mathbb{N}$ természetes szám és

$(x_i, y_i) \in \Gamma$, $i = 1, 2, ..., N$ pontok esetén teljesülni kell, hogy:

$$\begin{aligned}
\varphi(x) &\leq c(x, y_1) - \varphi^{c+}(y_1) = c(x, y_1) - c(x_1, y_1) + \varphi(x_1) \\
&\leq \big(c(x, y_1) - c(x_1, y_1)\big) + c(x_1, y_2) - \varphi^{c+}(y_2) \\
&= \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + \varphi(x_2) \\
&\leq ... \\
&\leq \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + ... + \big(c(x_N, \bar{y} - c(\bar{x}, \bar{y})\big) + \varphi(\bar{x}).
\end{aligned}$$

Itt felhasználtuk, hogy definíció alapján $\varphi(x) + \varphi^{c+}(y) \leq c(x, y)$, továbbá azt is, hogy ha φ olyan függvény, amire $\Gamma \subseteq \partial^{c+}\varphi$, akkor minden $(x, y) \in \Gamma \subseteq \partial^{c+}\varphi$ pontpárra $\varphi(x) + \varphi^{c+}(x) = c(x, y)$. Ezek alapján definiáljuk $\varphi(x)$-et a fenti kifejezés infimumaként az $N \in \mathbb{N}$ és $\{(x_i, y_i) \,|\, i = 1, ..., N\} \subset \Gamma$ feltételek mellett. Ha egy konstanst hozzáadunk φ-hez, az nem rontja el a fenti tulajdonság teljesülését, ezért a $\varphi(\bar{x})$-et elhagyhatjuk. Belátjuk, hogy az így definiált φ jó lesz. Tehát φ definíciója a következő:

$$\begin{aligned}
\varphi(x) := \inf\Big\{ \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + ... + \\
+ \big(c(x_N, \bar{y}) - c(\bar{x}, \bar{y})\big) \,|\, N \in \mathbb{N}, (x_i, y_i) \in \Gamma, i = 1, ..., N\Big\}.
\end{aligned}$$

Belátjuk, hogy a φ függvény c-konkáv. Ehhez azt kell belátnunk, hogy

$$\varphi(x) = \inf_y \{(x, y) - \psi(y)\}$$

valamilyen ψ függvényre. Írjuk fel $\varphi(x)$-et a következő módon:

$$\begin{aligned}
\varphi(x) &= \inf\Big\{ \big(c(x, y_1) - c(x_1, y_1)\big) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + ...+ \\
&\qquad + \big(c(x_N, \bar{y} - c(\bar{x}, \bar{y})\big) \,|\, N \in \mathbb{N}, (x_i, y_i) \in \Gamma, i = 1, ..., N\Big\} \\
&= \inf_{y_1}\Big\{ c(x, y_1) + \inf\big\{ - c(x_1, y_1) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + ...+ \\
&\qquad + \big(c(x_N, \bar{y} - c(\bar{x}, \bar{y})\big) \,|\, N \in \mathbb{N}, (x_i, y_i) \in \Gamma, i = 1, ..., N\big\}\Big\} \\
&= \inf_{y}\Big\{ c(x, y) + \inf\big\{ - c(x_1, y) + \big(c(x_1, y_2) - c(x_2, y_2)\big) + ...+ \\
&\qquad + \big(c(x_N, \bar{y} - c(\bar{x}, \bar{y})\big) \,|\, N \in \mathbb{N}, (x_i, y_i) \in \Gamma, i = 1, ..., N, y_1 = y\big\}\Big\}.
\end{aligned}$$

Ezek alapján legyen:

$$\psi(y) = -\inf\Big\{ -c(x_1,y) + \big(c(x_1,y_2) - c(x_2,y_2)\big) + ... + \big(c(x_N, \bar{y} - c(\bar{x},\bar{y})\big) \mid$$
$$N \in \mathbb{N}, (x_i,y_i) \in \Gamma, i = 1,...,N, y_1 = y\Big\}.$$

A ψ-t így választva rögtön adódik, hogy a φ c-konkáv. Az $N = 1$ és $(x_1, y_1) = (\bar{x}, \bar{y})$ választással kapjuk, hogy:

$$\varphi(x) \leq c(x,\bar{y}) - c(\bar{x},\bar{y}) < a(x) + b(\bar{y}) - c(\bar{x},\bar{y}).$$

Ebből következik, hogy:

$$\begin{aligned}\max\{\varphi(x), 0\} &\leq \max\{a(x) + b(\bar{y}) - c(\bar{x},\bar{y}), 0\} \\ &\leq |a(x) + b(\bar{y}) - c(\bar{x},\bar{y})| \\ &\leq |a(x)| + |b(\bar{y})| + |c(\bar{x},\bar{y})|.\end{aligned}$$

Ebből, az $a \in L^1(\mu)$ feltételből, és abból, hogy μ valószínűségi mérték, az következik, hogy $\max\{\varphi, 0\} \in L^1(\mu)$. Már csak azt kell belátnunk, hogy $\Gamma \subseteq \partial^{c+}\varphi$. Ehhez legyen $(\tilde{x},\tilde{y}) \in \Gamma$, és legyen $(x_1,y_1) = (\tilde{x},\tilde{y})$. Ekkor a φ definíciója alapján:

$$\varphi(x) \leq c(x,\tilde{y}) - c(\tilde{x},\tilde{y}) + \inf\Big\{\big(c(\tilde{x},y_2) - c(x_2,y_2)\big) + ... + \big(c(x_N,\bar{y}) - c(\bar{x},\bar{y})\big) \mid$$
$$N \in \mathbb{N}, (x_i,y_i) \in \Gamma, i = 2,...,N\Big\}$$
$$= c(x,\tilde{y}) - c(\tilde{x},\tilde{y}) + \varphi(\tilde{x}).$$

A c-szuperdifferenciál (5) egyenlettel adott karakterizációjából és a fenti egyenlőtlenségből következik, hogy $(\tilde{x},\tilde{y}) \in \partial^{c+}\varphi$.

(iii) $\Longrightarrow$ (i):

Most feltesszük, hogy létezik olyan φ c-konkáv függvény, amire $\max\{\varphi, 0\} \in L^1(\mu)$ és $\mathrm{supp}(\gamma) \subseteq \partial^{c+}\varphi$. Be kell látnunk, hogy ekkor γ optimális. Legyen $\tilde{\gamma} \in \Pi(\mu,\nu)$ egy tetszőleges transzport terv. Azt kell belátnunk, hogy

$$\int\limits_{X\times Y} c(x,y)\,d\gamma(x,y) \leq \int\limits_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y).$$

Tudjuk, hogy:

$$\begin{aligned}\varphi(x) + \varphi^{c+}(y) &= c(x,y) \qquad \forall (x,y) \in \mathrm{supp}(\gamma),\\ \varphi(x) + \varphi^{c+}(y) &\leq c(x,y) \qquad \forall (x,y) \in X \times Y.\end{aligned}$$

Ebből pedig már következik, hogy:

$$\begin{aligned}\int_{X\times Y} c(x,y)\,d\gamma(x,y) &= \int_{X\times Y} \big(\varphi(x)+\varphi^{c+}(y)\big)\,d\gamma(x,y)\\ &= \int_X \varphi(x)\,d\mu(x) + \int_Y \varphi^{c+}(y)\,d\nu(y)\\ &= \int_{X\times Y} \big(\varphi(x)+\varphi^{c+}(y)\big)\,d\tilde{\gamma}(x,y)\\ &\le \int_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y).\end{aligned}$$

Ellenőriznünk kell még, hogy a fenti egyenletekben szereplő integrálok léteznek. Az $\int_{X\times Y} \varphi(x)\,d\gamma(x,y)$ integrál létezéséhez azt kell belátnunk, hogy nem lehet φ pozitív és negatív részének is a γ mérték szerinti integrálja ∞. Azt, hogy a

$$\int_{X\times Y} c(x,y)\,d\gamma(x,y) \quad \text{és} \quad \int_{X\times Y} c(x,y)\,d\tilde{\gamma}(x,y)$$

integrálok léteznek és végesek, tudjuk a bizonyítás elején lévő megjegyzésből. A $\max\{\varphi,0\} \in L^1(\mu)$ feltételből következik, hogy létezik az

$$\int_X \varphi(x)\,d\mu(x)$$

integrál, hiszen a feltétel alapján a pozitív rész integrálja véges. Ebből következik, hogy léteznek az

$$\int_{X\times Y} \varphi(x)\,d\gamma(x,y) \quad \text{és} \quad \int_{X\times Y} \varphi(x)\,d\tilde{\gamma}(x,y)$$

integrálok (mert ezek ugyanakkor léteznek, mint $\int_X \varphi(x)\,d\mu(x)$, és az értékük megegyezik vele). Mivel az

$$\int_{X\times Y} c(x,y)\,d\gamma(x,y) \quad \text{és} \quad \int_{X\times Y} \varphi(x)\,d\gamma(x,y)$$

integrálok léteznek, továbbá $\int_{X\times Y} c(x,y)\,d\gamma(x,y) < \infty$,

$$\int_{X\times Y} \varphi^{c+}(y)\,d\gamma(x,y)$$

is létezik, hiszen $\varphi(x)+\varphi^{c+}(y) = c(x,y)$, ha $(x,y) \in \operatorname{supp}(\gamma)$, és

$$\int_{X\times Y} \varphi^{c+}(y)\,d\gamma(x,y) = \int_{X\times Y} c(x,y)\,d\gamma(x,y) - \int_{X\times Y} \varphi(x)\,d\gamma(x,y).$$

Az $\int_{X\times Y} \varphi^{c+}(y)\, d\gamma(x,y)$ létezéséből következik, hogy $\int_Y \varphi^{c+}(y)\, d\nu(y)$ létezik, abból pedig az, hogy $\int_{X\times Y} \varphi^{c+}(y)\, d\tilde{\gamma}(x,y)$ is létezik. Ahhoz, hogy

$$\int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^{c+}(y)\, d\nu(y)$$

értelmes legyen, elég belátni, hogy $\int_Y \varphi^{c+}(y)\, d\nu(y)$ értéke nem lehet $+\infty$. Ez a következőképpen látható be:

$$\varphi^{c+}(y) = \inf_{x\in X}\{c(x,y) - \varphi(x)\} \leq c(\tilde{x},y) - \varphi(\tilde{x}) \qquad \forall \tilde{x} \in X.$$

Az egyenlőtlenség mindkét oldalát integráljuk a γ mérték szerint:

$$\begin{aligned}\int\limits_{X\times Y} \varphi^{c+}(y)\, d\gamma(x,y) &\leq \int\limits_{X\times Y} c(\tilde{x},y)\, d\gamma(x,y) - \int\limits_{X\times Y} \varphi(\tilde{x})\, d\gamma(x,y)\\ &= \int\limits_{X\times Y} c(\tilde{x},y)\, d\gamma(x,y) - \varphi(\tilde{x}) < \infty,\end{aligned}$$

hiszen $\int\limits_{X\times Y} c(\tilde{x},y)\, d\gamma(x,y) < \infty$.

□

Az előző tétel fontos következményei (a tétel feltételeinek teljesülése mellett):

- Az hogy egy transzport terv optimális-e, csak a tartójától függ. ((i)$\Longleftrightarrow$(ii))
- Ha γ egy optimális transzport terv a marginálisai között, és egy $\tilde{\gamma}$ valószínűségi mértékre teljesül, hogy $\mathrm{supp}(\tilde{\gamma}) \subseteq \mathrm{supp}(\gamma)$, akkor $\tilde{\gamma}$ is optimáis transzport terv a saját marginálisai között. ((i)$\Longleftrightarrow$(ii))
- Tegyük fel, hogy $T : X \to Y$ egy olyan leképezés, amire $T(x) \in \partial^{c+}\varphi(x)$ minden x-re egy rögzített φ c-konkáv függvényre, amire $\max\{\varphi, 0\} \in L^1(\mu)$. Ekkor az optimális transzport alaptétele alapján minden olyan $\mu \in \mathcal{P}(X)$ mértékre, amire teljesül, hogy $c(x,y) \leq a(x) + b(y)$ valamilyen $a \in L^1(\mu)$ és $b \in L^1(T_{\#}\mu)$ függvényekre, T optimális lesz a μ és a $T_{\#}\mu$ között.

Az optimális transzport alaptételéből tudjuk, hogy ha a $\mu \in \mathcal{P}(X)$ és $\nu \in \mathcal{P}(Y)$ mértékekre teljesülnek a tétel feltételei, akkor minden γ optimális transzport tervhez létezik olyan φ c-konkáv függvény, amire $\mathrm{supp}(\gamma) \subseteq \partial^{c+}\varphi$. Ennél egy erősebb állítás is igaz: Ha egy γ optimális transzport tervre $\mathrm{supp}(\gamma) \subseteq \partial^{c+}\varphi$, akkor $\mathrm{supp}(\gamma') \subseteq \partial^{c+}\varphi$ minden optimális γ' transzporttervre. A $\max\{\varphi, 0\} \in L^1(\mu)$ feltételből következik, hogy $\max\{\varphi^{c+}, 0\} \in L^1(\nu)$, hiszen:

$$\varphi^{c+}(y) = \inf_{x \in X} \{c(x,y) - \varphi(x)\} \leq c(\bar{x}, y) - \varphi(\bar{x}) \leq a(\bar{x}) + b(y) - \varphi(\bar{x})$$

valamilyen rögzített $\bar{x} \in X$-re. Ebből az egyenlőtlenségből és a $b \in L^1(\nu)$ feltételből következik, hogy $\max\{\varphi^{c+}, 0\} \in L^1(\nu)$. Ezt felhasználva az állítást az alábbi módon láthatjuk be:

$$\begin{aligned}
\int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^{c+}(y)\, d\nu(y) &= \int_{X \times Y} \big(\varphi(x) + \varphi^{c+}(y)\big)\, d\gamma'(x,y) \\
&\leq \int_{X \times Y} c(x,y)\, d\gamma'(x,y) \\
&= \int_{X \times Y} c(x,y)\, d\gamma(x,y) \\
&= \int_{X \times Y} \big(\varphi(x) + \varphi^{c+}(y)\big)\, d\gamma(x,y) \\
&= \int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^{c+}(y)\, d\nu(y).
\end{aligned}$$

Itt a

$$\int_{X \times Y} c(x,y)\, d\gamma'(x,y) = \int_{X \times Y} c(x,y)\, d\gamma(x,y)$$

egyenlőség azért teljesül, mert γ és γ' is optimális, a

$$\int_{X \times Y} c(x,y)\, d\gamma(x,y) = \int_{X \times Y} \big(\varphi(x) + \varphi^{c+}(y)\big)\, d\gamma(x,y)$$

egyenlőség pedig azért, mert $\mathrm{supp}(\gamma) \subseteq \partial^{c+}\varphi$. Az egyenlőtlenség két végén álló két kifejezés megegyezik, és mindkettő véges, hiszen $\int_{X \times Y} c(x,y)\, d\gamma(x,y) < \infty$. Emiatt a közbülső egyenlőtlenségnek egyenlőséggel kell teljesülnie, ami azt jelenti, hogy $(x,y) \in \partial^{c+}\varphi$ teljesül γ' majdnem minden (x,y) pontra. Ebből és a c folytonosságából következik, hogy $\mathrm{supp}(\gamma') \subseteq \partial^{c+}\varphi$. (Az, hogy az integrálok léteznek, ugyanúgy látható be, mint korábban.)

7. Dualitás

Ebben a fejezetben megfogalmazzuk a Kantorovics-probléma duálisát, és belátjuk a dualitás tételt az [1] könyv 1.3, és a [2] könyv 1.2 fejezetei alapján.
A duális probléma a következő. Legyen $\mu \in \mathcal{P}(X)$ és $\nu \in \mathcal{P}(Y)$ két valószínűségi Borel-mérték. Keressük a következő értéket:

$$\sup\left\{ \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) \,\middle|\, \varphi \in L^1(\mu), \psi \in L^1(\nu), \right. \\ \left. \varphi(x) + \psi(y) \le c(x,y) \quad \forall (x,y) \in X \times Y \right\}. \tag{DP}$$

Mutatunk egy heurisztikus érvelést, ami megmutatja, hogyan kapjuk meg a Kantorovics-problémából a duálist:

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X\times Y} c(x,y)\,d\gamma(x,y) = \inf_{\gamma \in \mathcal{M}_+(X\times Y)} \left\{ \int_{X\times Y} c(x,y)\,d\gamma(x,y) + \chi(\gamma) \right\},$$

ahol $\chi(\gamma) = 0$, ha $\gamma \in \Pi(\mu,\nu)$, különben $\chi(\gamma) = \infty$. Ez a $\chi(\gamma)$ a következőképp adható meg:

$$\chi(\gamma) = \sup_{\varphi,\psi} \left\{ \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \int_{X\times Y} \big(\varphi(x) + \psi(y)\big)\,d\gamma(x,y) \right\},$$

ahol $\varphi \in C_b(X)$, és $\psi \in C_b(Y)$. Tetszőleges $(\varphi,\psi) \in C_b(X) \times C_b(Y)$ függvénypárra teljesül a következő, ha $\gamma \in \Pi(\mu,\nu)$:

$$\int_{X\times Y} \big(\varphi(x) + \psi(y)\big)\,d\gamma(x,y) = \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y),$$

így a szuprémum értéke 0 lesz. Ha $\gamma \notin \Pi(\mu,\nu)$, akkor léteznek olyan $\varphi \in C_b(X)$, $\psi \in C_b(Y)$ függvények, amikre

$$\int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \int_{X\times Y} \big(\varphi(x) + \psi(y)\big)\,d\gamma(x,y) \neq 0,$$

hiszen különben minden $\varphi \in C_b(X)$ függvényre teljesülne, hogy

$$\int_X \varphi(x)\,d\mu(x) = \int_{X\times Y} \varphi(x)\,d\gamma(x,y),$$

így γ első marginálisa a μ mérték lenne, és hasonlóan, minden $\psi \in C_b(Y)$ függvényre teljesülne, hogy

$$\int_Y \psi(y)\, d\nu(y) = \int_{X\times Y} \psi(y)\, d\gamma(x,y),$$

amiből következik, hogy a γ mérték második marginálisa a ν mérték. (Ennek a bizonyítását láttuk a 5.1 tétel bizonyításában.) Vegyünk egy (φ, ψ) függvénypárt, amire (7) teljesül. A φ és ψ függvények mindegyikét ugyanazzal a konstanssal megszorozva a (7) egyenlet bal oldala tetszőlegesen naggyá tehető, így a szuprémum ∞ lesz. Ezek alapján:

$$\begin{aligned}
&\inf_{\gamma\in\Pi(\nu,\mu)} \int_{X\times Y} c(x,y)\, d\gamma(x,y) = \\
&= \inf_{\gamma\in\mathcal{M}_+(X\times Y)} \sup_{\varphi,\psi} \Bigg\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) + \int_X \varphi(x)\, d\mu(x) + \\
&\qquad + \int_Y \psi(y)\, d\nu(y) - \int_{X\times Y} \big(\varphi(x)+\psi(y)\big)\, d\gamma(x,y) \Bigg\}.
\end{aligned}$$

Ha az infimumot és a szuprémumot fel tudnánk cserélni, ezt kapnánk:

$$\begin{aligned}
&\inf_{\gamma\in\Pi(\nu,\mu)} \int_{X\times Y} c(x,y)\, d\gamma(x,y) \\
&= \inf_{\gamma\in\mathcal{M}_+(X\times Y)} \sup_{\varphi,\psi} \Bigg\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) + \int_X \varphi(x)\, d\mu(x) + \\
&\qquad + \int_Y \psi(y)\, d\nu(y) - \int_{X\times Y} \big(\varphi(x)+\psi(y)\big)\, d\gamma(x,y) \Bigg\} \\
&= \sup_{\varphi,\psi} \inf_{\gamma\in\mathcal{M}_+(X\times Y)} \Bigg\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) + \int_X \varphi(x)\, d\mu(x) + \\
&\qquad + \int_Y \psi(y)\, d\nu(y) - \int_{X\times Y} \big(\varphi(x)+\psi(y)\big)\, d\gamma(x,y) \Bigg\} \\
&= \sup_{\varphi,\psi} \Bigg\{ \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) + \\
&\qquad + \inf_{\gamma\in\mathcal{M}_+(X\times Y)} \left\{ \int_{X\times Y} \big(c(x,y)-\varphi(x)-\psi(y)\big)\, d\gamma(x,y) \right\} \Bigg\},
\end{aligned}$$

ahol a φ és ψ függvényekre az a megkötés, hogy $\varphi \in C_b(X)$ és $\psi \in C_b(Y)$. Most vizsgáljuk meg ezt a kifejezést:

$$\inf_{\gamma \in \mathcal{M}_+(X \times Y)} \left\{ \int_{X \times Y} \big(c(x,y) - \varphi(x) - \psi(y)\big)\, d\gamma(x,y) \right\}.$$

Ha minden (x,y) pontra teljesül, hogy $\varphi(x) + \psi(y) \leq c(x,y)$, akkor minden γ mértékre egy nemnegatív függvényt integrálunk, így az infimum értéke 0 lesz, hiszen ha $\gamma \equiv 0$, akkor az integrál értéke 0. Ha viszont van olyan (x,y) pont, amire $\varphi(x) + \psi(y) > c(x,y)$, akkor a $\gamma := K \cdot \delta_{(x,y)}$ választással, és a K konstanst elég nagyra választva, a

$$\int_{X \times Y} \big(c(x,y) - \varphi(x) - \psi(y)\big)\, d\gamma(x,y)$$

integrál tetszőlegesen nagy abszolútértékű negatív számot felvehet, így ekkor a kifejezés értéke $-\infty$. Azt kaptuk, hogy

$$\inf_{\gamma \in \Pi(\nu,\mu)} \int c(x,y)\, d\gamma(x,y) = \sup_{\varphi,\psi} \left\{ \int \varphi(x)\, d\mu(x) + \int \psi(y)\, d\nu(y) \right\},$$

ahol $\varphi \in C_b(X)$, $\psi \in C_b(Y)$ és $\varphi(x) + \psi(y) \leq c(x,y) \quad \forall (x,y) \in X \times Y$.

7.1. Tétel (Dualitás)**.** *Legyen $\mu \in \mathcal{P}(X)$ és $\nu \in \mathcal{P}(Y)$ két valószínűségi Borel-mérték, $c : X \times Y \to \mathbb{R}$ egy folytonos és alulról korlátos függvény. Ha léteznek olyan $a \in L^1(\mu)$ és $b \in L^1(\nu)$ függvények, amikre:*

$$c(x,y) \leq a(x) + b(y) \qquad \forall (x,y) \in X \times Y,$$

akkor a Kantorovics-probléma minimuma megegyezik a duális probléma szuprémumával. Ekkor a duális probléma szuprémuma egyben maximum is, és létezik (φ, φ^{c+}) alakú maximalizáló függvénypár, ahol φ egy c-konkáv függvény.

Bizonyítás. Legyen $\gamma \in \Pi(\mu,\nu)$ egy transzport terv, és legyenek a $\varphi \in L^1(\mu)$ és $\psi \in L^1(\nu)$ olyan függvények, amikre $\varphi(x) + \psi(y) \leq c(x,y) \quad \forall (x,y) \in X \times Y$. Ekkor:

$$\begin{aligned} \int_{X \times Y} c(x,y)\, d\gamma(x,y) &\geq \int_{X \times Y} \big(\varphi(x) + \psi(y)\big)\, d\gamma(x,y) \\ &= \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y). \end{aligned}$$

A bal oldalon γ-ra infimumot véve, a bal oldalon φ-re és ψ-re szuprémumot véve kapjuk, hogy (KP) $\geq$ (DP). A (KP) $\leq$ (DP) egyenlőtlenség belátásához vegyünk egy $\gamma \in \Pi(\mu, \nu)$ optimális transzport tervet. Ekkor az optimális transzport alaptétele alapján létezik olyan φ c-konkáv függvény, amire $\max\{\varphi, 0\} \in L^1(\mu)$ és $\operatorname{supp}(\gamma) \subseteq \partial^{c+}\varphi$. Az optimális transzport alaptételének bizonyítása után láttuk, hogy $\max\{\varphi^{c+}, 0\} \in L^1(\nu)$.

$$\begin{aligned} \int_{X\times Y} c(x,y)\, d\gamma(x,y) &= \int_{X\times Y} \big(\varphi(x) + \varphi^{c+}(y)\big)\, d\gamma(x,y) \\ &= \int_X \varphi(x)\, d\mu(x) + \int_Y \varphi^{c+}(y)\, d\nu(y) \end{aligned} \tag{6}$$

Mivel $\int_{X\times Y} c(x,y)\, d\gamma(x,y) < \infty$, ezért $\varphi \in L^1(\mu)$ és $\varphi^{c+} \in L^1(\nu)$. (Az, hogy az integrálok léteznek, és nincs "$\infty - \infty$" alakú kifejezés a fenti képletben, hasonlóan látható be, mint az optimális transzport alaptételénél.) A φ^{c+} definíciója alapján $\varphi(x)+\varphi^{c+}(y) \leq c(x,y) \quad \forall (x,y) \in X\times Y$, vagyis φ és φ^{c+} kielégítik a duális probléma feltételeit, amiből a (6) egyenlőség alapján következik, hogy (KP) $\leq$ (DP). $\square$

Ha teljesülnek az optimális transzport alaptételének és a dualitás tételnek a feltételei, akkor minden φ c-konkáv függvényre, amire (φ, ψ) maximalizálja a duális problémát valamilyen ψ-re, és minden γ optimális transzport tervre teljesül, hogy:

$$\operatorname{supp}(\gamma) \subseteq \partial^{c+}\varphi.$$

Az optimális transzport alaptételének bizonyítása után láttuk, hogy létezik olyan $\varphi \in L^1(\mu)$ c-konkáv függvény, amire

$$\varphi^{c+} \in L^1(\nu), \quad \operatorname{supp}(\gamma) \subseteq \partial^{c+}\varphi$$

minden γ optimális transzport tervre. Vegyünk egy $(\tilde{\varphi}, \tilde{\psi})$ függvénypárt, ami szintén maximalizálja a duális problémát. A duális probléma feltételei alapján minden $(x,y) \in X\times Y$ pontpárra teljesül, hogy $\tilde{\varphi}(x)+\tilde{\psi}(y) \leq c(x,y)$. Ebből következik, hogy $\tilde{\psi} \leq \tilde{\varphi}^{c+}$, és így $(\tilde{\varphi}, \tilde{\varphi}^{c+})$ is maximalizálja a duális problémát. Azt kell belátnunk, hogy $\operatorname{supp}(\gamma) \subseteq \partial^{c+}\tilde{\varphi}$. Azt, hogy $\tilde{\varphi}^{c+} \in L^1(\nu)$, ugyanúgy láthatjuk be, mint a dualitás tétel bizonyatásában. Legyen γ egy optimális transzport terv, ekkor:

$$\begin{aligned}\int_X \tilde{\varphi}(x)\,d\mu(x) + \int_Y \tilde{\varphi}^{c+}(y)\,d\nu(y) &= \int_X \varphi(x)\,d\mu(x) + \int_Y \varphi^{c+}(y)\,d\nu(y)\\ &= \int_{X\times Y} \big(\varphi(x) + \varphi^{c+}(y)\big)\,d\gamma(x,y)\\ &= \int_{X\times Y} c(x,y)\,d\gamma(x,y)\\ &\geq \int_{X\times Y} \big(\tilde{\varphi}(x) + \tilde{\varphi}^{c+}(y)\big)\,d\gamma(x,y)\\ &= \int_X \tilde{\varphi}(x)\,d\mu(x) + \int_Y \tilde{\varphi}^{c+}(y)\,d\nu(y).\end{aligned}$$

Az egyenlőtlenség két végén álló két kifejezés megegyezik, és ezek végesek, hiszen tudjuk, hogy $\int_{X\times Y} c(x,y)\,d\gamma(x,y) < \infty$. Emiatt a közbülső egyenlőtlenségnek egyenlőséggel kell teljesülnie, ami azt jelenti, hogy $\operatorname{supp}(\gamma) \subseteq \partial^{c+}\tilde{\varphi}$.

7.2. Definíció. *Egy c-konkáv φ függvényt Kantorovics-potenciálnak nevezünk, ha (φ, φ^{c+}) maximalizálja a duális problémát.*

Megjegyezzük, hogy ha (φ, ψ) maximalizálja a duális problémát, akkor (φ, φ^{c+}) is, azaz (φ, φ^{c+}) ekkor Kantorovics-potenciál. Könnyen belátható, hogy a Kantorovics-potenciálok pontosan azok a c-konkáv függvények, amiknek a szuperdifferenciálja tartalmazza az összes optimális transzport terv tartóját. Ha (φ, φ^{c+}) Kantorovics-potenciál, akkor a $\partial^{c+}\varphi$ tartalmazza az összes optimális transzport terv tartóját. (Ezt a dualitás tétel bizonyítása után láttuk be.) Másfelől, ha $\partial^{c+}\varphi$ tartalmazza egy γ optimális transzport terv tartóját, akkor (φ, φ^{c+}) optimális, hiszen:

$$\int\limits_{X\times Y} c(x,y)\,d\gamma(x,y) = \int\limits_{X\times Y} (\varphi(x)+\varphi^{c+}(y))\,d\gamma(x,y) = \int\limits_X \varphi(x)\,d\mu(x)+\int\limits_Y \varphi^{c+}(y)\,d\nu(y).$$

8. A Wasserstein-tér definíciója

Ebben a fejezetben definiáljuk a mértékek Wasserstein-távolságát, és belátjuk, hogy ez valóban egy metrika a mértékek egy megfelelő halmazán. A fejezetben ismertetett definíciók és tételek megtalálhatóak a [1] könyv 2.1, illetve a [2] könyv 5.1 fejezetében, a diszintegrációs tétel pedig a [3] könyv 2. fejezetében.

8.1. Definíció. *Legyen (X, ϱ) egy szeparábilis és teljes metrikus tér, $p \geq 1$ rögzített szám. A véges p-edik momentumú valószínűségi mértékek halmazát jelölje $\mathcal{P}_p(X)$:*

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) \,\middle|\, \exists x_0 \in X : \int_X \varrho^p(x, x_0)\, d\mu(x) < \infty \right\}$$

A $\mu, \nu \in \mathcal{P}_p(X)$ mértékek p-Wasserstein-távolsága:

$$d_{\mathcal{W}_p}(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times X} \varrho^p(x, y)\, d\gamma(x, y) \right)^{\frac{1}{p}}.$$

A $(\mathcal{P}_p(X), d_{\mathcal{W}_p})$ párt az X feletti Wasserstein-térnek nevezzük, és a következőképpen jelöljük: $\mathcal{W}_p(X)$.

Be kell látnunk, hogy $d_{\mathcal{W}_p}$ metrika $\mathcal{P}_p(X)$-en, azaz $\mathcal{W}_p(X)$ egy metrikus tér. A fenti definíció alapján két mérték p-Wasserstein-távolsága megegyezik a két mérték közötti optimális transzport költségének p-edik gyökével a $c(x, y) = \varrho^p(x, y)$ költségfüggvény mellett.

Ha μ tetszőleges és $\nu = \delta_{x_0}$ valamilyen $x_0 \in X$ pontra, akkor egyetlen transzport terv van μ és ν között: a $\mu \times \delta_{x_0}$ szorzatmérték, így ez nyilván optimális. Ezek alapján:

$$d_{\mathcal{W}_p}(\mu, \delta_{x_0}) = \left(\int_X \varrho^p(x, x_0)\, d\mu(x) \right)^{\frac{1}{p}}.$$

Mielőtt bebizonyítanánk, hogy $d_{\mathcal{W}_p}$ metrika $\mathcal{P}_p(X)$-en, belátunk egy lemmát, amire szükségünk lesz a háromszög-egyenlőtlenség bizonyításához. Ehhez szükség van némi előkészületre.

Legyen μ egy Borel-mérték az X téren, az Y pedig legyen egy topologikus tér. A $\{\mu_y \,|\, y \in Y\}$ mértékcsaládot a μ mérték $f : X \to Y$ függvény szerinti diszintegrációjának nevezzük, ha minden y-ra a μ_y egy $f^{-1}(\{y\})$ halmazra koncentrált valószínűségi mérték, továbbá teljesül, hogy minden $\phi \in C(X)$ függvényre az $y \to \int_X \phi(x)\, d\mu_y(x)$ leképezés Borel-mérhető, és

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

ahol $\nu = f_{\#}\mu$.

Abban az esetben, amikor $X = Y \times Z$ és az f az Y-ra való vetítés, a μ_y mértékek az $Y \times Z$ téren értelmezett, az $\{y\} \times Z$ halmazokra koncentrált mértékek. Ekkor a $\{y\} \times Z$ és a Z azonsításával a μ_y mértékekre tekinthetünk úgy, mintha a Z téren lennének értelmezve, és teljesül, hogy

$$\int_{Y \times Z} \phi(y, z)\, d\mu(y, z) = \int_Y \int_Z \phi(y, z)\, d\mu_y(z)\, d\nu(y).$$

Szükségünk lesz a következő tételre, ami a diszintegráció létezéséről szól (2.4 tétel a [4] könyvben):Legyen X és Y két szeparábilis és teljes metrikus tér, a μ egy mérték X-en, az $f : X \to Y$ egy Borel-mérhető függvény, és legyen $\nu = f_{\#}\mu$. Ekkor létezik az X téren értelmezett valószínűségi mértékeknek egy olyan $\{\mu_y \,|\, y \in Y\}$ családja, amire:

(i) tetszőleges $A \in \mathcal{B}(X)$ esetén a $y \to \mu_y(A)$ függvény Borel-mérhető,

(ii) tetszőleges $A \in \mathcal{B}(X)$ esetén $\mu(A) = \int_Y \mu_y(A)\, d\nu(y)$

(iii) a μ_y az $f^{-1}(\{y\})$ halmazra van koncentrálva ν majdnem minden y-ra.

Továbbá, ha $\{\mu'_y \,|\, y \in Y\}$ az X téren értelmezett valószínűségi mértékeknek egy olyan családja, ami teljesíti a fenti feltételeket, akkor $\mu'_y = \mu_y$ teljesül μ majdnem minden y-ra.

8.1. Lemma. *Legyen X egy szeparábilis és teljes metrikus tér, és legyenek $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(X)$ valószínűségi mértékek X-en. Legyenek $\gamma^1 \in \Pi(\mu_1, \mu_2)$ és $\gamma^2 \in \Pi(\mu_2, \mu_3)$ transzport tervek. Ekkor létezik legalább egy olyan $\sigma \in \mathcal{P}(X \times X \times X)$,* *amire:*

$$(\pi_{x,y})_\# \sigma = \gamma^1$$
$$(\pi_{y,z})_\# \sigma = \gamma^2,$$

ahol $\pi_{x,y}$, illetve $\pi_{y,z}$ jelöli az első két, illetve második két változóra való vetítést.

Bizonyítás. Először vegyük a γ^1 mérték π_y vetítés szerinti diszintegrációját. Így kapunk egy $\{\gamma_y^1 \,|\, y \in Y\}$ mértékcsaládot, ahol $\gamma_y^1 \in \mathcal{P}(X)$ minden $y \in Y$ esetén. A γ_y^1 mértékekre definíció szerint teljesül, hogy:

$$\int_{X\times X} \phi(x,y)\, d\gamma^1(x,y) = \int_X \int_X \phi(x,y)\, d\gamma_y^1(x)\, d\mu_2(y)$$

minden nemnegatív ϕ Borel-mérhető függvényre. Vegyük γ^2-nek is a π_y szerinti diszintegrációját, ez legyen $\{\gamma_y^2 \,|\, y \in Y\}$. Ekkor:

$$\int_{X\times X} \phi(y,z)\, d\gamma^2(y,z) = \int_X \int_X \phi(y,z)\, d\gamma_y^2(z)\, d\mu_2(y).$$

Definiáljuk σ-t a következő módon:

$$\int_{X\times X\times X} \phi(x,y,z)\, d\sigma(x,y,z) := \int_X \int_{X\times X} \phi(x,y,z)\, d(\gamma_y^1 \times \gamma_y^2)(x,z)\, d\mu_2(y),$$

ahol $\gamma_y^1 \times \gamma_y^2$ jelöli a γ_y^1 és γ_y^2 mértékek szorzatát. Minden ϕ kétváltozós függvényre teljesül, hogy:

$$\begin{aligned}
\int_{X\times X\times X} \phi(x,y)\, d\sigma(x,y,z) &= \int_X \int_{X\times X} \phi(x,y)\, d(\gamma_y^1 \times \gamma_y^2)(x,z)\, d\mu_2(y) \\
&= \int_X \int_X \phi(x,y)\, d\gamma_y^2(x)\, d\mu_2(y) \\
&= \int_{X\times X} \phi(x,y)\, d\gamma^2(x,y).
\end{aligned}$$

Ezek alapján $(\pi_{x,y})_\# \sigma = \gamma^2$. Hasonlóan látható be, hogy $(\pi_{y,z})_\# \sigma = \gamma^1$. □

Az előkészületek után beláthatjuk, hogy a $\mathcal{W}_p(X)$ tér valóban egy metrikus tér.

8.2. Tétel. *A $d_{\mathcal{W}_p} : \mathcal{P}_p(X) \times \mathcal{P}_p(X) \to \mathbb{R}$ függvény metrika $\mathcal{P}_p(X)$-en.*

Bizonyítás. A Wasserstein-távolság definíciójából nyilvánvaló, hogy $d_{\mathcal{W}_p}$ szimmetrikus, továbbá az is, hogy $d_{\mathcal{W}_p} \geq 0$, és $d_{\mathcal{W}_p}(\mu, \mu) = 0$ minden $\mu \in \mathcal{P}_p(X)$ esetén. Belátjuk, hogy a $d_{\mathcal{W}_p}(\mu, \nu) = 0$ feltételből következik, hogy $\mu = \nu$. Az, hogy $d_{\mathcal{W}_p}(\mu, \nu) = 0$ azt jelenti, hogy létezik olyan $\gamma \in \Pi(\mu, \nu)$, amire

$$\int_{X \times X} \varrho(x, y)^p \, d\gamma(x, y) = 0.$$

Ekkor γ szükségszerűen az $x = y$ halmazra van koncentrálva. Ebből már következik, hogy $\mu = \nu$, hiszen tetszőleges ϕ függvényre:

$$\int_X \phi(x) \, d\mu(x) = \int_{X \times X} \phi(x) \, d\gamma(x, y) = \int_{X \times X} \phi(y) \, d\gamma(x, y) = \int_X \phi(y) \, d\nu(y).$$

Ezt a karakterisztikus függvényekre alkalmazva azonnal adódik, hogy $\mu = \nu$. Azt kell még belátnunk, hogy $d_{\mathcal{W}_p}$ teljesíti a háromszög-egyenlőtlenséget. Legyenek $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$ valószínűségi mértékek, $\gamma^1 \in \Pi(\mu_1, \mu_2)$ és $\gamma^2 \in \pi(\mu_2, \mu_3)$ optimális transzport tervek. A lemma alapján létezik olyan $\sigma \in \mathcal{P}(X \times X \times X)$ mérték, amire $(\pi_{x,y})_{\#}\sigma = \gamma^2$ és $(\pi_{y,z})_{\#}\sigma = \gamma^1$. Legyen $\gamma := (\pi_{x,z})_{\#}\sigma$. Ekkor $(\pi_x)_{\#}\gamma = (\pi_x)_{\#}\sigma = (\pi_x)_{\#}\gamma^1 = \mu_1$, és hasonlóan $(\pi_z)_{\#}\gamma = \mu_3$. Ez azt jelenti, hogy $\gamma \in \Pi(\mu_1, \mu_3)$, és:

$$\begin{aligned}
d_{\mathcal{W}_p}(\mu_1, \mu_3) &\leq \left(\int\limits_{X \times X} \varrho^p(x, z) \, d\gamma(x, z) \right)^{\frac{1}{p}} = \left(\int\limits_{X \times X \times X} \varrho^p(x, z) \, d\sigma(x, y, z) \right)^{\frac{1}{p}} \\
&\leq \left(\int\limits_{X \times X \times X} \big(\varrho(x, y) + \varrho(y + z)\big)^p \, d\sigma(x, y, z) \right)^{\frac{1}{p}} \\
&\leq \left(\int\limits_{X \times X \times X} \varrho^p(x, y) \, d\sigma(x, y, z) \right)^{\frac{1}{p}} + \left(\int\limits_{X \times X \times X} \varrho^p(y, z) \, d\sigma(x, y, z) \right)^{\frac{1}{p}} \\
&= \left(\int\limits_{X \times X} \varrho^p(x, y) \, d\gamma^1(x, y) \right)^{\frac{1}{p}} + \left(\int\limits_{X \times X} \varrho^p(y, z) \, d\gamma^2(y, z) \right)^{\frac{1}{p}} \\
&= d_{\mathcal{W}_p}(\mu_1, \mu_2) + d_{\mathcal{W}_p}(\mu_2, \mu_3).
\end{aligned}$$

A $d_{\mathcal{W}_p}$ értéke nem lehet $+\infty$, mert a háromszög-egyenlőtlenség alapján:

$$\begin{aligned} d_{\mathcal{W}_p}(\mu,\nu) &\le d_{\mathcal{W}_p}(\mu,\delta_{x_0}) + d_{\mathcal{W}_p}(\delta_{x_0},\nu) \\ &= \left(\int\limits_X \varrho^p(x,x_0)\,d\mu(x) \right)^{\frac{1}{p}} + \left(\int\limits_X \varrho^p(x,x_0)\,d\nu(x) \right)^{\frac{1}{p}} < \infty. \end{aligned}$$

$\square$

9. A Wasserstein-távolság és a Wasserstein-terek néhány fontos tulajdonsága

Ebben a fejezetben bemutatjuk a Wasserstein-távolság néhány hasznos tulajdonságát. Összehasonlításképp és az előnyöket kihangsúlyozandó minden esetben megemlítünk egy-egy olyan metrikát is, ami a vizsgált tulajdonsággal nem rendelkezik.

9.1. A Wasserstein-távolság érzékeny az alul fekvő távolságra

Sok esetben hasznos, ha a valószínűségi mértékeken értelmezett távolságból visszanyerhető az alulfekvő tér metrikája. Szemléltetésképp tekintsük azt az esetet, amikor a metrikus tér az $(\mathbb{R}, |\cdot|)$, a számegyenes Borel valószínűségi mértkein pedig a Kolmogorov-Smirnov metrikát vesszük.

A $\mu, \nu \in \mathcal{P}(\mathbb{R})$ valószínűségi mértékek Kolmogorov-Smirnov távolsága:

$$d_{KS}(\mu, \nu) = \sup_{t \in \mathbb{R}} |F_\mu(t) - F_\nu(t)|,$$

ahol F_μ a μ mérték eloszlásfüggvénye, F_ν pedig a ν mértéké. Ha $x \neq y$, akkor $d_{KS}(\delta_x, \delta_y) = 1$, vagyis a Kolmogorov-Smirnov távolság nem tartalmaz információt az alul fekvő térben lévő metrikáról. Viszont a Lévy-távolság már tartalmaz. A μ és ν mértékek Lévy-távolsága:

$$d_L(\mu, \nu) = \inf \left\{ \varepsilon > 0 \mid \forall x \in \mathbb{R} : F_\mu(x - \varepsilon) - \varepsilon \leq F_\nu(x) \leq F_\mu(x + \varepsilon) + \varepsilon \right\}.$$

A δ_x és δ_y mértékek Lévy-távolsága: $d_L(\delta_x, \delta_y) = \min\{|x - y|, 1\}$, vagyis a Lévy-metrikából ki tudjuk nyerni, hogy az x és y távolsága kisebb-e 1-nél, és ha igen, akkor a pontos értékét is megkapjuk.

Ezzel szemben a Wasserstein-távolságból az alaptér metrikája egyértelműen kiolvasható, ugyanis tetszőleges (X, ϱ) metrikus tér és $x, y \in X$ pontok esetén a δ_x és δ_y Dirac-mértékek között egyetlen transzport terv van, a szorzatmérték, ami éppen $\delta_{(x,y)}$, és így

$$d_{\mathcal{W}_p}(\delta_x, \delta_y) = \left(\int_{X \times X} \varrho^p(u, v) \, d\delta_{(\delta_x, \delta_y)}(u, v) \right)^{\frac{1}{p}} = \left(\varrho^p(x, y)\right)^{\frac{1}{p}} = \varrho(x, y).$$

Másképp megfogalmazva, a $\mathcal{W}_p(X)$ Wasserstein tér tartalmazza az alul fekvő (X, ϱ) tér egy izometrikus másolatát, nevezetesen:

$$\{\delta_x : x \in X\} \subseteq \mathcal{W}_p(X).$$

9.2. Approximálhatóság

Sokszor hasznos, ha abszolút folytonos eloszlásokat véges sok pontra koncentrált mértékekkel tudunk approximálni. A példa kedvéért tekintsük most a $([0,1], |\cdot|)$ metrikus teret, a $\mathcal{P}([0,1])$-et pedig lássuk el a totális variációs távolsággal, d_{TV}-vel:

$$d_{TV}(\mu, \nu) = \sup_{A \in \mathcal{B}([0,1])} |\mu(A) - \nu(A)|.$$

Legyen μ az egyenletes eloszlás a $[0,1]$ intervallumon, ν pedig az egyenletes eloszlás a $\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\}$ halmazon. Azt várnánk, hogy ha n elég nagy, akkor a ν mérték jó közelítése lesz μ-nek. Csakhogy az $\widehat{A} = [0,1] \setminus \{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\}$ egy Borel-halmaz, és így

$$d_{TV}(\mu, \nu) = \sup_{A \in \mathcal{B}([0,1])} |\mu(A) - \nu(A)| \geq |\mu(\widehat{A}) - \nu(\widehat{A})| = 1.$$

Ez nem csak azt jelenti, hogy a μ és ν mértékek nincsenek közel egymáshoz, hanem azt is, hogy a lehető legtávolabb vannak egymástól, hiszen μ és ν valószínűségi mértékek, és így

$$d_{TV}(\mu, \nu) = \sup_{A \in \mathcal{B}([0,1])} |\mu(A) - \nu(A)| \leq 1.$$

Hasonlóan látható be, hogy ha a μ abszolút folytonos a $\lambda|_{[0,1]}$ mértékre, és

$$\nu = \sum_{i=1}^{n} \lambda_i \cdot \delta_{x_i}$$

valamilyen $\lambda_i \geq 0$, $\sum_{i=1}^{n} \lambda_i = 1$ és $x_i \in [0,1]$ számokra, akkor $d_{TV}(\mu, \nu) = 1$. Ezzel szemben tetszőleges (X, ϱ) szeparábilis és teljes metrikus tér esetén a

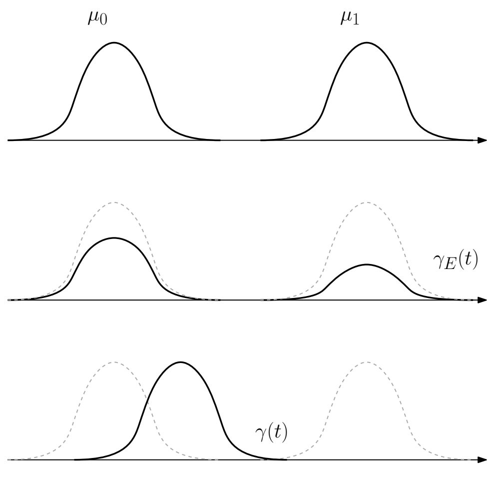
$$\left\{ \sum_{i=1}^{n} \lambda_i \cdot \delta_{x_i} : n \in \mathbb{N}, \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = 1, x_i \in X \right\}$$

halmaz sűrű $\mathcal{W}_p(X)$-ben, így minden valószínűségi mértéket tudunk véges sok pontra koncentrált mértékekkel approximálni.

9.3. Átlagolás

Tegyük fel, hogy van két hasonló objektumunk, mondjuk két eloszlásunk vagy két képünk, és szeretnénk ezek átlagát értelmezni úgy, hogy az átlag struktúrája minél jobban hasonlítson a kiindulási objektumok struktúrájára. Például ha van két normális eloszlásunk, $\mu_0 = N(0,1)$ és $\mu_1 = N(1,1)$, akkor ezek euklideszi átlaga, a $\frac{\mu_0+\mu_1}{2}$ mérték, nem egy normális eloszlás. Ha viszont vesszük a két eloszlást összekötő egyértelmű geodetikust a $\mathcal{W}_2(\mathbb{R})$ Wasserstein-térben, és ennek felezőpontjaként definiáljuk a két eloszlás átlagát, akkor az $N(0,1)$ és $N(1,1)$ eloszlások átlaga $N(\frac{1}{2},1)$ lesz, ami szintén normális eloszlás.

Hasonló a helyzet akkor is, amikor a két objektum közül az egyiket szeretnénk áttranszformálni a másikba: úgy szeretnénk ezt megtenni, hogy a folyamat közben minden időpillanatban a kiindulási és végső objektumokhoz hasonló struktúrájú objektumunk legyen. Például ha a $\mu_0 = N(0,1)$ eloszlást szeretnénk áttranszformálni a $\mu_1 = N(1,1)$ eloszlásba, akkor egy olyan γ görbét keresünk, amire $\gamma(0) = \mu_0$, $\gamma(1) = \mu_1$ és $\gamma(t)$ a $\gamma(0)$ és $\gamma(1)$ mértékekhez "hasonló" mérték minden $t \in [0,1]$ számra. A transzformációt elvégezhetnénk a $\gamma_E(t) = (1-t)\cdot\mu_0 + t\cdot\mu_1$ görbe alapján is, de ekkor a közbülső $\gamma(t)$ mértékek szerkezete jelentősen eltérne $\gamma(0)$ és $\gamma(1)$ szerkezetétől (lásd 8. ábra). Ehelyett megint vehetjük a $N(0,1)$ és $N(1,1)$ eloszlásokat összekötő geodetikust a $\mathcal{W}_2(\mathbb{R})$ Wasserstein-térben, ez a $\gamma(t) = N(t,1)$ görbe lesz.



8. ábra

Amikor két kép átlagát szeretnénk definiálni, nem pixelenként akarjuk őket kiátlagolni, mert ezzel az eljárással az eredmény általában egy értelmezhetetlen kép lenne. Ehelyett azt akarjuk, hogy az átlag egy értelmes kép legyen, ami tartalmazza mindkét kép "jellegzetességeit". Ha az egyik képet szeretnénk áttranszformálni a másikba, akkor azt úgy szeretnénk csinálni, hogy az első képből kiindulva a másodikra egyre jobban hasonlító képeken keresztül jussunk el a második képhez. Csinálhatnánk azt, hogy az első kép minden pixelére megnézzük, hogy a második képben milyen pixel felel meg neki, és minden pixelt külön-külön a megfelelő pixellé transzformálunk, de így a közbülső képek nem feltétlenül lennének értelemesek. Megpróbálhatjuk a transzformációt megint a Wasserstein-térben lévő geodetikus alapján definiálni, de ehhez először eloszlásokként kell elkódolni a képeket. A képekhez úgy szeretnénk eloszlásokat rendelni, hogy az eloszlások tartalmazzák a képekről azokat az információkat, amik számunkra fontosak.

Például tegyük fel, hogy fekete-fehér képek sötétségével szeretnénk foglalkozni. Minden fekete-fehér képhez egyszerűen hozzá lehet rendelni egy olyan eloszlást, ami azt írja le, mennyire sötét az adott kép. A fekete-fehér képeket a következőképpen kódolják el: minden pixelhez hozzárendelnek egy 0 és 255 közötti számot, ami azt jelzi, hogy az adott pixel színe hol helyezkedik el a fekete és a fehér közötti skálán (0: fekete, 255: fehér). Az egyszerűség kedvéért osszuk le a számokat 255-tel, hogy a pixelekhez rendelt értékek 0 és 1 között legyenek. Minden fekete-fehér képhez rendeljünk egy mértéket a következőképpen: a $t \in [0, 1]$ számra $n(t)$ súlyt rakunk, ahol $n(t)$ jelöli a kép azon pixeleinek a számát, amikhez a t számot rendeltük. Ezt a mértéket még osszuk le a pixelek számával, hogy valószínűségi mértéket kapjunk.

Ezek alapján egy csupa fekete pixelből álló képhez a δ_0 mérték tartozik, a csupa fehér pixelekből álló képhez pedig a δ_1. Ha a μ_0 és μ_1 mértékek átlagát a $\frac{\mu_0+\mu_1}{2}$ mértékként definiálnánk, akkor a fekete és fehér kép átlaga egy olyan kép lenne, amihez a $\frac{\delta_0+\delta_1}{2}$ mérték tartozik. Ez egy olyan képnek felel meg, amiben a pixelek fele fehér, fele pedig fekete. Ehelyett azt szeretnénk, hogy a két kép átlaga egy szürke kép legyen, amihez a $\delta_{\frac{1}{2}}$ mérték tartozik. Ez megegyezik a $\mathcal{W}_2([0, 1])$ Wasserstein-térben a δ_0 és δ_1 mértékeket összekötő egyetlen geodetikus, a $\gamma(t) = \delta_t$ görbe, felezőpontjával. Fontos megjegyezni, hogy itt tényleg felhasználjuk, hogy $p = 2$, azaz a $\mathcal{W}_2([0, 1])$

térben vagyunk, ahol egyértelmű a geodetikus. Ugyanis például a $\mathcal{W}_1([0,1])$ térben a $\gamma(t) = \delta_t$ mellett a $\gamma_E(t) = (1-t) \cdot \delta_0 + t \cdot \delta_1$ görbe is egy geodetikus δ_0 és δ_1 között. Tehát a $\mathcal{W}_1([0,1])$ térben az $\frac{1}{2}(\delta_0 + \delta_1)$ is egy geodetikus felezőpontja.

Ha a fekete képet a $\gamma(t) = (1-t) \cdot \delta_0 + t \cdot \delta_1$ görbe mentén transzformáljuk át a a fehér képpé, akkor a közbülső képek olyanok, amikben csak fekete és fehér pixelek vannak, és egyre több a fehér pixel. Ha viszont a $\mathcal{W}_2([0,1])$-beli $\gamma(t) = \delta_t$ geodetikus alapján végezzük a transzformációt, akkor végig egyszínű, egyre világosabb képeken keresztül jutunk el a fekete képből a fehérbe.

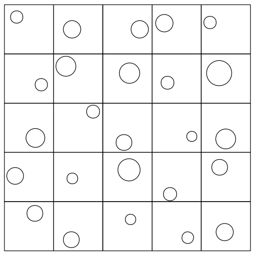
A két kép átlagához hasonlóan a két kép $1-t$ és t súlyokkal vett átlagát is definiálhatjuk (az egyértelműen létező) $\mathcal{W}_2([0,1])$-beli γ geodetikus segítségével: minden $t \in [0,1]$ számra vegyük a geodetikuson a $\gamma(t)$-nek megfelelő képet. Általában, ezzel a módszerrel definiálhatjuk két $\mathcal{W}_2([0,1])$-beli mérték súlyozott átlagát.

Tegyük fel, hogy szeretnénk több eloszlás átlagát definiálni a $\mathcal{W}_2(X)$ térben. A μ_0, μ_1, μ_2 átlagát megpróbálhatnánk úgy definiálni, hogy vesszük a μ_0-t és μ_1-et összekötő geodetikust, vesszük ennek a felezőpontját, azt összekötjük a μ_2-vel a geodetikus mentén, és vesszük ennek az $1:2$ arányú osztópontját. Azonban ez nem asszociatív, vagyis más eredményt kapunk, ha a mértékek sorrendjét felcseréljük.

A fenti megközelítés helyett induljunk ki az $\mathbb{R}^d$-beli súlypontok, vagy más szóval baricentrumok, egy másik karakterizációjából: az $x_i \in \mathbb{R}^d$ pontok $\lambda_i \geq 0$ súlyokkal vett baricentruma az az egyértelmű y pont, ami minimalizálja a $\sum_{i=1}^n \lambda_i |y - x_i|^2$ kifejezést. Ez motiválja a Wasserstein-baricentrum definícióját: a μ_i mértékek $\lambda_i \geq 0$ súlyokkal vett baricentruma a $\mathcal{W}_2(X)$ térben az a ν mérték, ami minimalizálja a következő kifejezést:

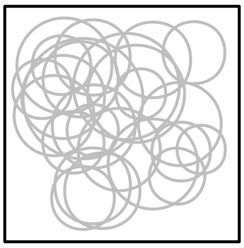
$$\sum_{i=1}^n \lambda_i \cdot d^2_{\mathcal{W}_p}(\nu, \mu_i).$$

A Wasserstein-baricentrumok fogalma használható több kép átlagának definiálására is. Tegyük fel, hogy van egy megfigyelésből álló 25 elemű mintánk, lásd 9. ábra.



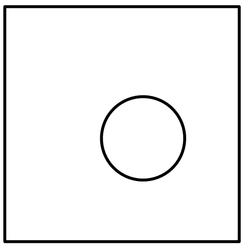
9. ábra

Azt látjuk, hogy a megfigyelések során véletlenszerűen elhelyezett, különböző sugarú körvonalakat kaptunk (reprezentálják ezeket a megfigyeléseket a $\mu_1, ..., \mu_{25}$ mértékek). Tehát az lenne egy jó átlagfogalom, ahol ennek a mintának az átlaga egy körhöz hasonlít. Ha a mértékeket csak simán az euklideszi módon kiátlagoljuk, azaz a $\frac{1}{25} \cdot \sum_{i=1}^{25} \mu_i$ mértéket vesszük, akkor a 10. ábrához hasonló képnek megfelelő mértéket kapunk.



10. ábra

Azonban ha ehelyett a μ_i mértékek $\mathcal{W}_2$-beli baricentrumát vesszük, akkor az eredmény egy kört ábrázoló képre fog hasonlítani.



11. ábra

Hivatkozások

[1] Luigi Ambrosio and Nicola Gigli. *A user's guide to optimal transport*, volume 2062. Springer, 2013.

[2] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87. Birkhäuser, Cham, 2015.

[3] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lecture 2: The Kantorovich Problem*, pages 13–22. Springer International Publishing, Cham, 2021.

[4] Daniele Semola Luigi Ambrosio, Elia Brué. *Lectures on Optimal Transport*. Springer, Cham, 2021.

NYILATKOZAT

Név: Szögi Roland

ELTE Természettudományi Kar, szak: Matematika Bsc

NEPTUN azonosító: XNLZX4

Szakdolgozat címe:
Optimális transzport és Wasserstein-terek

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2022. május 18.

Szögi Roland

a hallgató aláírása