

¿QUÉ ES EL TRANSPORTE ÓPTIMO?

Jorge Salazar¹

Resumen. En este artículo haremos una breve introducción a la teoría del Transporte óptimo. El interés en el Transporte óptimo, es reciente, o mejor, renació en los años 40's con los trabajos de L. Kantorovich en la ex Unión Soviética y F. L. Hitchcock y T. Koopmans en los Estados Unidos. El Transporte Óptimo tomó impulso a partir de los años 80's cuando los métodos desarrollados en 3 líneas de investigación disímiles convergían en la versión moderna del viejo problema de Monge, descrito en "Mémoire sur la théorie des déblais e de remblais", publicado en 1781. El Transporte Óptimo es una teoría en pleno desarrollo y con un potencial enorme de aplicaciones..

Palabras Claves: Desintegración Dimensional de Medidas, Dimensión de Medidas, Función de Densidad de Medidas, medidas de Hausdorff.

Abstract. In this article, we make a brief introduction to the theory of Optimal Transportation. The interest in this subject is recent, or reborn in the 40's with the work of L. Kantorovich in the ex-Soviet Union and F. L. Hitchcock and T. Koopmans in the USA. Optimal Transportation took impulse from the 80's, when 3 very different areas of research came together around the modern formulation of the old Monge transport problem, expressed in "Memoire sur la theorie des deblais e de remblais" of 1781. Optimal Transportation is a theory in plain expansion and with a huge potential of applications.

Keywords and phrases. Optimal Transportation, marginal measures, couplings, cyclically monotone sets, C-Convexity, Wasserstein distance.

Recibido: Agosto 2015.

Aceptado: Agosto 2015.

1. INTRODUCCIÓN

1.1. **El problema de Monge (Gaspard Monge, 1746-1818).** En su "Mémoire sur la théorie des déblais e de remblais", 1781, Monge describe el problema de transportar el material extraído de varias minas a los sitios de construcción.

La producción de cada mina y la cantidad necesaria en cada construcción son conocidas y el total de la producción es igual al total del consumo. El problema consiste en determinar el destino al cual se debe enviar el material de cada mina, de forma a minimizar el costo del transporte. El costo en el problema de Monge era determinado por el producto de la distancia por la cantidad de material transportado.

1.2. Leonid V. Kantorovich, 1912-1986. Kantorovich trabajó en varias áreas de matemáticas y es mas conocido por la programación lineal, muy utilizada en economía. De hecho, Kantorovich obtuvo el premio Nobel en economía en 1975, compartido con T. Koopmans.

En los años 1940's, Kantorovich consideró un problema más amplio que el problema de Monge (y más natural), al permitir que la "masa" original se "disperse" y mostró que resolverlo es equivalente a resolver un problema "dual" en el cual el problema original de "minimizar el costo" se transforma en uno que consiste en "maximizar la ganancia."

Entre las nociones desarrolladas para resolver el problema del transporte y su dual están los conjuntos C-cíclicamente monótonos, funciones C-cóncavas/C-convexas, distancias entre medidas (planes de transferencia), entre otras. Aquí hablaremos de las dos primeras, dejando de fuera las distancias entre medidas.

1.3. Los años 1980's. En la historia de la matemática, existen varios ejemplos de investigaciones independientes, sobre temas completamente diferentes, que posteriormente resultan unificados en una única

Teoría matemática abaragente. Esto sucedió con el Transporte Óptimo durante los años 80's, con tres líneas de investigación en áreas distintas e incluso una de ellas fuera de la matemática.

En los Sistemas Dinámicos, John Mather consideró medidas estacionarias minimizantes de la acción de un Lagrangiano, las cuales son soluciones de un problema variacional y generalizan las curvas minimizantes.

Entre tanto, en la Mecánica de Fluidos Incompresibles, Yann Brenier define operadores de proyección sobre transformaciones conservando la medida, a través de un emparejamiento óptimo. Atrayendo de esta forma la atención de la comunidad que trabaja en las Ecuaciones Diferenciales Parciales.

Por otro lado Mike Cullen, trabajando con Ecuaciones semi-geostróficas en frentes atmosféricos, reinterpreta el cambio de variable de Hoskins en términos del emparejamiento óptimo y pone en evidencia el hecho que la minimalidad corresponde a una condición de estabilidad.

¹Jorge Salazar, Dirección Prometeo: Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias Naturales y Matemáticas, Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador E-mail address: jsalaza@espol.edu.ec

1.4. Nuevo milenio. Desde el inicio del siglo 21, el interés está centrado en la descripción cualitativa del Transporte Óptimo.

L. Caffarelli, C. Evans, W. Gangbo, R. McCann y otros trabajaron en describir mejor la estructura del transporte óptimo y encontraron otras aplicaciones. F. Otto introduce un formalismo ligado a la geometría diferencial. En la actualidad existe una interacción muy fructífera entre el transporte óptimo, la Geometría Diferencial, el Análisis Funcional y las Ecuaciones Diferenciales.

2. TRANSPORTE DE MEDIDA DE PROBABILIDAD

Con la palabra “transporte,” hay tendencia a pensar en el tránsito vehicular o en el transporte masivo. Más de estos temas no se ocupa el Transporte óptimo. El Transporte óptimo estudia el transporte o lo que también se puede llamar transferencia de “medidas.” Y que es una medida? y que tipo de transporte?

2.1. Medidas. (*Discusión heurística*) Todos tenemos idea de lo que es la medida de longitud, es decir la noción de longitud, no el resultado de la medición de la longitud de un objeto. Intuitivamente, una medida es un ente matemático que nos permite integrar (sumar) funciones.

Matemáticamente hablando, para poder integrar una función, debemos tener una “medida” y la función debe ser “integrable” con respecto a esta medida. Sin embargo, desde el punto de vista físico, podemos distinguir dos tipos de funciones “integrables.” Para explicar mejor este punto, tomemos dos ejemplos:

Supongamos que una función representa la temperatura en cada punto de un sólido. Si integramos esta función con respecto a la medida de volumen y dividimos para el volumen total, obtenemos la temperatura media.

Consideremos ahora el mismo sólido y una función que representa la densidad de masa. Si integramos esta función con respecto a la medida de volumen, obtenemos la masa total. Podemos también integrar esta función en dominios parciales (partes o “subconjuntos” del sólido) y obtendremos la masa de esa parte del sólido.

En el primer ejemplo el integral de la temperatura, solo tiene sentido “físico” cuando dividimos para el volumen. En el ejemplo de la densidad de masa, a cada parte del sólido podemos asociar su masa, obteniendo de este modo una nueva medida o “noción” de masa.

En general, integrando funciones (positivas), podemos obtener nuevas medidas (positivas). En la práctica, hacemos esto cuando éstas tienen sentido físico como la medida de masa del segundo ejemplo. Mas, matemáticamente, basta que sea una función positiva, “integrable.”

2.2. Transporte. En el mismo sólido, podemos determinar el calor “por unidad de volumen” en función de la temperatura, densidad de masa, calor específico (que también puede ser variable), etc. Al igual que la distribución de masa, también tiene sentido físico asociar a cada parte del sólido el calor contenido en esa parte en un instante dado, i.e. el integral del calor sobre cada parte del sólido. Esto resulta en lo que podemos denominar de distribución de calor.

A diferencia de la distribución de masa, que por tratarse de un sólido, podemos imaginar que es constante, la distribución del calor varía en el tiempo debido a la transferencia de calor por conducción o difusión. Si suponemos que el sólido está aislado y por lo tanto no hay intercambio de energía con el medio exterior, la cantidad total de energía calorífica contenida en el sólido es constante en el tiempo y lo único que varía es la distribución del calor

Figure 1.
16 bolas cada una con “peso” 1/16



2.2.1. Transporte / transferencia de distribuciones. Observando la distribución de calor en un instante dado y un cierto tiempo más tarde, vemos dos distribuciones diferentes, con el mismo total. El paso de una configuración a la otra es un buen ejemplo de lo que llamamos transporte o transferencia de distribuciones. Este es el tipo de “transporte” del que se ocupa el Transporte Óptimo.

2.3. Medidas/distribuciones de probabilidad. Una vez que las distribuciones tienen el mismo total, dividiendo por el total, podemos considerar que éste es 1. Una distribución

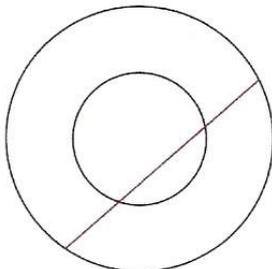
cuyo total es 1 se llama distribución de probabilidad.

Una medida o distribución de probabilidad es un sistema que asigna, a cada evento (=parte o subconjunto del conjunto todo), un “peso” o “proporción” llamada medida del evento, que es un número real entre 0 y 1 (no tiene unidades).

La unión de dos eventos disjuntos tiene su medida igual a la suma de las medidas de los dos eventos individuales. El conjunto todo tiene medida 1 y el conjunto vacío tiene medida 0.

En la figura 1, tenemos una urna con 16 bolas. Si asignamos el mismo “peso” a cada bola, entonces la medida de un conjunto unitario (con- teniendo una bola) tiene medida 1/16. Cada conjunto constituido por bolas de la urna es un evento. La medida de cada evento es el número de bolas dividido por 16.

Figure 2.
Cuerda escogida “al azar”



2.4. La paradoja de Bertrand. Para motivar la definición rigurosa de un espacio de probabilidad, vamos a relatar el ejemplo dado por Joseph Louis Francois Bertrand (1822–1900) en su libro “Calcul des probabilités” publicado en 1889.

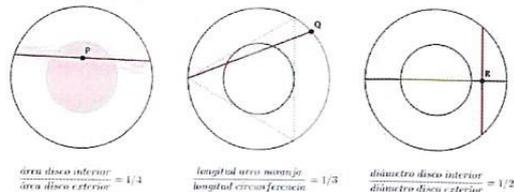
Consideremos una circunferencia de radio 2 cm y escojamos una cuerda de esta circunferencia al azar. Cuál es la probabilidad de que esta cuerda intersecte el círculo concéntrico de radio 1 cm? (Figura 2)

2.5. Tres interpretaciones/tres soluciones. La frase “la cuerda intersecta el círculo interior” no es un evento en el sentido de las probabilidades, en cuanto no se determine el espacio de probabilidad y la medida de probabilidad a la que nos estamos refiriendo.

2.5.1. Primera interpretación/solución. Todas las cuerdas (excepto los diámetros) están determinadas por su punto medio, de forma unívoca. Toda cuerda intersecta el disco de radio 1 si y solamente si el punto medio pertenece al disco interior (de radio 1). Luego, para escoger una cuerda basta escoger su punto medio y desde luego, podemos convenir que el punto es escogido en una región con probabilidad proporcional al área de esa región. En ese caso, la respuesta a la pregunta: Cual es la probabilidad de que una cuerda escogida al aleatoriamente intersecte el círculo concéntrico

de radio 1 cm? Es el cociente entre el área del disco de radio 1 dividido por el área del disco de radio 2. O sea 1/4. (Figura 3.)

2.5.2. Segunda interpretación/solución. Por simetría a, podemos considerar únicamente las cuerdas que nacen en un punto fijo. En ese caso, las cuerdas están determinadas por el segundo punto sobre la circunferencia. Las cuerdas que intersectan el disco de radio 1 son las que tienen el segundo punto en el tercio de circunferencia que queda frente al punto fijo. (Ver el esquema central de la figura 3). La medida natural asociada a los subconjuntos de la circunferencia es proporcional a la longitud de arco. En este caso, la respuesta es el cociente entre la longitud del arco “favorable” dividido para la longitud de la circunferencia. Es decir 1/3.



2.5.3. Tercera interpretación/solución. Por simetría a, podemos considerar únicamente las cuerdas “verticales.” Esta vez, las cuerdas están determinadas por el punto de intersección con el diámetro horizontal. (Ver el esquema derecho de la figura 3). La medida “uniforme” asociada a los subconjuntos del diámetro es proporcional a su longitud. Para que la cuerda intersecte el disco interior, el punto de intersección con el diámetro horizontal debe estar en el segmento central, que corresponde al diámetro del disco pequeño. La longitud de este segmento es igual a la mitad de la longitud del diámetro del disco exterior. Interpretando las cosas de esta forma, la respuesta es 1/2.

2.6. Espacio de probabilidad. El ejemplo de Bertrand nos muestra que debemos definir cuidadosamente lo que queremos decir cuando decimos “al azar” o “aleatoriamente.” La forma correcta es introduciendo un espacio de probabilidad, que es una estructura matemática precisa:

Empezamos con un conjunto Ω , no vacío, y una familia de subconjuntos de Ω con una cierta estructura, que interpretamos como “eventos” (todo lo que puede suceder). Esta familia de subconjuntos se la conoce con el nombre de σ -álgebra y la describimos a continuación.

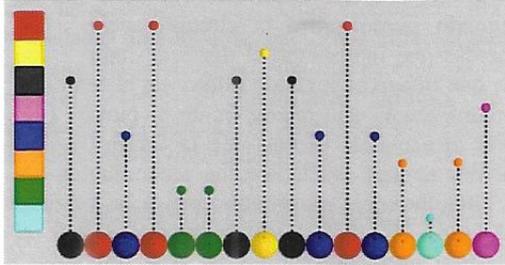
2.7. σ -álgebra. Denotemos 2^Ω la familia de todos los subconjuntos de Ω . Una σ -álgebra es una familia de subconjuntos de Ω , que vamos a denotar \mathcal{F} ($\mathcal{F} \subset 2^\Omega$), con las siguientes propiedades:

$$(1) \emptyset \in \mathcal{F}.$$

- (2) Si $A \in \mathcal{F}$, entonces $\Omega \setminus A \in \mathcal{F}$.
- (3) Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$.

Figure 4.

A cada bola (numerada), asociamos su color



Los elementos de \mathcal{F} se llaman “conjuntos medibles.”

Cuando Ω es un conjunto finito (como en el ejemplo de la urna, ver la figura 1), \mathcal{F} es usualmente la familia de todos los subconjuntos, i.e. 2^Ω . Como la urna de la figura 1 tiene 16 bolas (numeradas), la familia 2^Ω tiene 216 elementos!

2.8. Medida de probabilidad. Definición Una medida de probabilidades es una función $\mathcal{P}: \mathcal{F} \rightarrow [0,1]$ que verifica

- (1) $\mathcal{P}(\emptyset) = 0$ y $\mathcal{P}(\Omega) = 1$.
- (2) Si A_1, A_2, \dots son elementos de \mathcal{F} , disjuntos dos a dos, entonces

$$\mathcal{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathcal{P}(A_k)$$

2.9. Transformaciones y medida imagen. Si a las bolas de la urna anterior “asociamos” su color (ver la figura 4.), obtenemos un nuevo espacio (conjunto) de 8 elementos (colores). La proporción de bolas de cada color es una medida de probabilidad sobre el espacio de los 8 colores. A esta medida se la conoce como la medida imagen de la medida original, a través de la aplicación “color.” (Ver la figura 5.) También podemos decir que la medida original es “transportada” hacia la medida imagen.

2.9.1. Otro ejemplo discreto. La aplicación

$$f : (x, y) \in \{(0,0),(1,0),(0,1),(1,1)\} \rightarrow x+y \in \{0, 1, 2\}$$

Figure 5.

En la urna, c/bola tiene medida 1/16. En el conjunto de colores, c/color tiene medida de acuerdo al número de bolas.

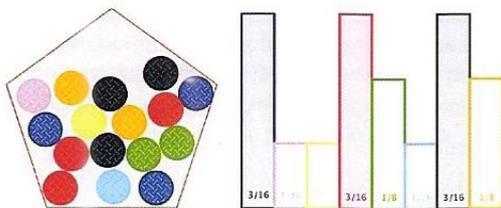
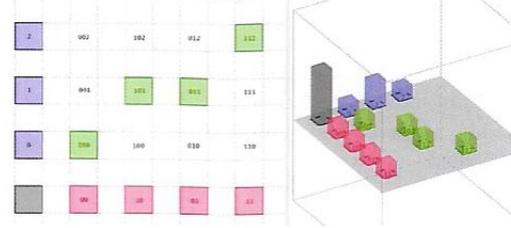


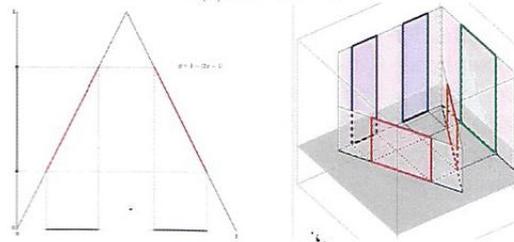
Figure 6.
 $f(x,y) = x + y$



Transporta la medida uniforme, i.e., la medida en que todos los elementos (o mejor dicho, los conjuntos unitarios) tienen medida 1/4, en la medida $v(0) = v(2) = 1/4$ y $v(1) = 1/2$. (Ver la figura 6.)

2.9.2. Ejemplo en el intervalo $[0, 1]$. La función $f(x) = 1 - |2x - 1|$, $x \in [0, 1]$, transporta la medida de longitud sobre $[0, 1]$ en sí misma. De hecho, medida imagen de cualquier intervalo contenido en el conjunto de llegada, llamémosle I , es igual a la suma de las longitudes de los dos intervalos de la preimagen. (Ver la figura 7.) Un cálculo elemental muestra que esta suma de longitudes de los intervalos de la preimagen es igual a la longitud del intervalo I . Luego la medida imagen del intervalo I es su longitud.

Figure 7.
 $f(x) = 1 - |2x - 1|$



2.9.3. Otro ejemplo en el intervalo $[0, 1]$. La función $g(x) = \sqrt{1-x}$, $x \in [0, 1]$, transporta la medida de longitud en la medida $2y dy$. (Ver la figura 7.)

Más explícitamente, la medida imagen de un intervalo $[a, b] \subset [0, 1]$ (contenido en el conjunto de llegada) es $b^2 - a^2$.

Efectivamente, la medida imagen es la longitud de la preimagen, que en este caso es el conjunto $\{x \in [0, 1]; \sqrt{1-x} \in [a, b]\} = [1-b^2, 1-a^2]$, Cuya longitud es la medida imagen del intervalo imagen. Note que para todo $0 \leq a < b \leq 1$,

$$\int_a^b 2y dy = b^2 - a^2$$

Por esta razón, decimos que la medida imagen en este ejemplo es dada por $2y dy$, i.e., la medida de longitud con “densidad” $2y$. Este punto puede ser aclarado viendo este proceso como un cambio de variable.

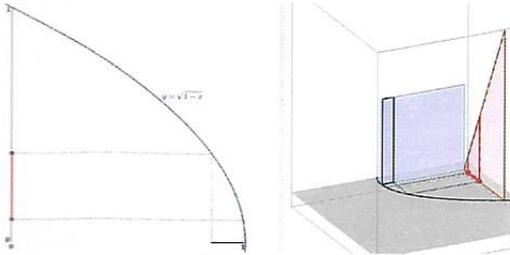
2.9.4. *Notación.* Dados dos espacios medibles, $(\mathcal{X}, \mathcal{F})$ y $(\mathcal{Y}, \mathcal{G})$, la medida imagen de una medida μ , definida sobre el espacio $(\mathcal{X}, \mathcal{F})$, con respecto a una transformación $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{Y}, \mathcal{G})$, \mathcal{F}/\mathcal{G} -medible, se denota $T\#\mu$ y se define por

$$(2.1) \quad T\#\mu(A) = \mu(T^{-1}(A)),$$

Para todo $A \in \mathcal{G}$. Como T es medible, $T^{-1}(A) \in \mathcal{F}$ y por tanto la expresión del lado derecho de (2.1) tiene sentido. El lector puede verificar que la fórmula (2.1) define ciertamente una medida.

Figure 8.

El área del triángulo rojo indica la medida del intervalo imagen (base del triángulo)



2.10. **Transporte de medidas / Transformaciones admisibles.** Supongamos ahora que tenemos dos espacios de probabilidad $(\mathcal{X}, \mathcal{F}, \mu)$ y $(\mathcal{Y}, \mathcal{G}, \nu)$ y que queremos encontrar una transformación $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{Y}, \mathcal{G})$ (\mathcal{F}/\mathcal{G} -medible) que transporte la medida μ en la medida ν , i.e. tal que

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

Las transformaciones con esta propiedad se llaman **TRANSFORMACIONES ADMISIBLES**.

2.10.1. *Ejemplo.* Consideremos los espacios

$$X = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$$

Con la medida uniforme $(\mu(\{(x, y)\}) = 1/4$, para todo $(x, y) \in X$ y $Y = \{0, 1, 2\}$

Con la medida $\nu(\{0\}) = \nu(\{2\}) = 1/4$ y $\nu(\{1\}) = 1/2$.

De las 81 (3^4) transformaciones de X en Y , 12 son admisibles. Las figuras 9 y 10 representan dos de ellas (además de la figura 6).

Observe el cambio en el punto de vista, en el párrafo anterior, la medida μ y la transformación T eran dadas y calculábamos la medida imagen $T\#\mu$. En este párrafo, las dos medidas $(\mu$ y $\nu)$ son conocidas y queremos estudiar todas las transformaciones que transportan μ en ν .

Este programa, i.e. Transportar una medida dada en otra igualmente fija (a través de una

aplicación), no siempre es posible. Por ejemplo una medida de Dirac solo se puede transportar en otra medida de Dirac.

Figure 9.
transformación admisible

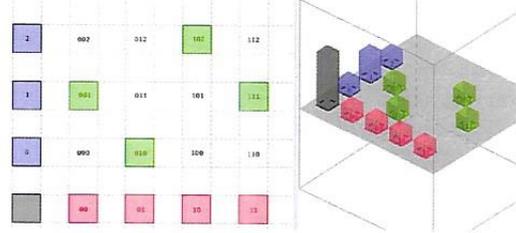
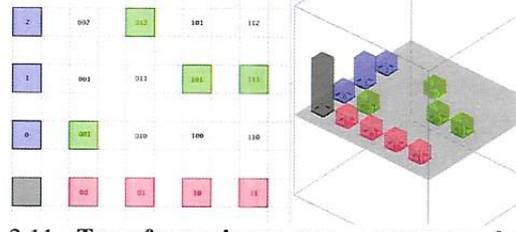


Figure 10.
transformación admisible 2



2.11. **Transformaciones que conservan la medida.** Un caso particular de esta situación es cuando consideramos un solo espacio de probabilidad, digamos $(\mathcal{X}, \mathcal{F}, \mu)$, y que queremos encontrar todas las transformaciones $T : (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{F})$ (\mathcal{F}/\mathcal{F} -medibles) que conservan la medida μ , i.e.

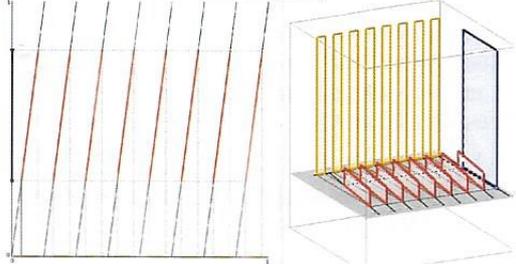
$$T\#\mu = \mu.$$

Las transformaciones que conservan la medida, con la operación de composición, forman un semigrupo con elemento neutro (la aplicación identidad). Las transformaciones invertibles que conservan la medida forman un grupo.

2.11.1. *Ejemplos.* La función $f(x) = 1 - |2x - 1|$ conserva la medida de longitud sobre el intervalo $[0, 1]$. (Ver la figura 7.) Otro ejemplo es

Figure 11.

$f_8(x) = 8x - [8x]$ conserva la medida de longitud sobre el intervalo $[0, 1]$



La función $f_n(x) = nx - [nx]$, siendo $n \in \mathbb{N}$ y $[y]$ el mayor entero menor o igual que y . (Ver la figura 11.)

3. PLANES DE TRANSFERENCIA

Como dijimos en la sección anterior, no siempre existen transformaciones que transportan una medida en otra (ambas fijas). Esto se arregla permitiendo que la “masa” de la primera medida se “disperse” en la segunda, como vamos a explicar a continuación.

3.1. Emparejamiento. Dados dos espacios de probabilidad $(\mathcal{X}, \mathcal{F}, \mu)$ y $(\mathcal{Y}, \mathcal{G}, \nu)$, queremos encontrar un espacio de probabilidad $(\Omega, \mathcal{E}, \mathcal{P})$ y dos variables aleatorias (aplicaciones medibles), digamos $X : (\Omega, \mathcal{E}) \rightarrow (\mathcal{X}, \mathcal{F})$ y $Y : (\Omega, \mathcal{E}) \rightarrow (\mathcal{Y}, \mathcal{G})$, tales que las leyes de X y Y sean μ y ν , respectivamente. i.e.,

$$(3.1) \quad X_{\#} P = \mu \text{ y } Y_{\#} P = \nu.$$

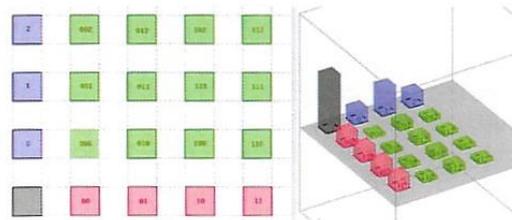
Un par (X, Y) , que verifica (3.1), se llama emparejamiento de μ y ν .

Note que (X, Y) es una variable aleatoria a valores en el espacio producto $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. Si (X, Y) es un emparejamiento de μ y ν , entonces la medida imagen de P por (X, Y) , i.e.,

$$\Pi = (X, Y)_{\#} P,$$

Figure 12.

Medida producto: cada punto tiene medida igual al producto de las medidas de sus componentes



Tiene a μ y ν como medidas marginales sobre $(\mathcal{X}, \mathcal{F})$ y $(\mathcal{Y}, \mathcal{G})$, respectivamente. Por abuso de lenguaje, a la medida Π también se la llama emparejamiento de μ y ν .

Recordemos que las medidas marginales son las medidas imagen de las proyecciones sobre los espacios factores,

$(x, y) \in \mathcal{X} \times \mathcal{Y} \rightarrow x \in \mathcal{X}$ y $(x, y) \in \mathcal{X} \times \mathcal{Y} \rightarrow y \in \mathcal{Y}$. Mas explícitamente, las medidas marginales son las medidas sobre

$(\mathcal{X}, \mathcal{F})$ y $(\mathcal{Y}, \mathcal{G})$, definidas (respectivamente) por:
 $A \in \mathcal{F} \rightarrow \Pi(A \times \mathcal{Y})$ y $B \in \mathcal{G} \rightarrow \Pi(\mathcal{X} \times B)$.

3.2. Medidas admisibles. Desde luego, en el párrafo anterior, siempre podemos escoger $\Omega =$

$\mathcal{X} \times \mathcal{Y}, \mathcal{E} = \mathcal{F} \otimes \mathcal{G}$ y las variables X y Y igual a las proyecciones sobre \mathcal{X} y \mathcal{Y} , respectivamente.

En ese caso el problema se reduce a encontrar todas las medidas de probabilidad sobre $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, tal que las proyecciones X y Y sigan las leyes μ y ν , respectivamente.

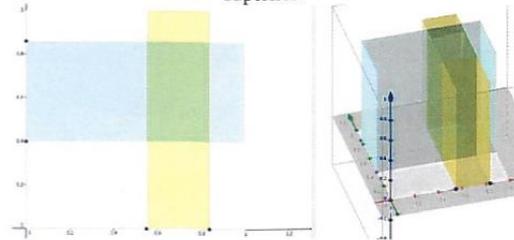
Una medida con estas características se conoce con el nombre de medida admisible o plan de transferencia entre μ y ν , o como dijimos en el párrafo anterior, también se la denomina emparejamiento de μ y ν .

3.2.1. Ejemplo: Medida Producto. Al contrario del problema de la sección anterior (i.e., encontrar transformaciones que transportan una medida en otra), encontrar un plan de transferencia siempre es posible.

De hecho, basta tomar la medida producto $\mu \times \nu$. (Ver la figura 12)

Figure 13.

La medida producto sobre $[0, 1] \times [0, 1]$ es la medida de superficie



En el caso del producto del espacio de probabilidad $[0, 1]$, con la medida de longitud, multiplicado por sí mismo, la medida producto es la medida de superficie sobre el cuadrado $[0, 1] \times [0, 1]$. (Ver la figura 12.)

3.2.2. Problema: Siendo $\mathcal{X} = \mathcal{Y} = [0, 1]$, el intervalo unidad con la medida de longitud, encontrar una medida sobre el cuadrado unidad $[0, 1] \times [0, 1]$ de forma que ambas medidas marginales sean la medida de longitud en $[0, 1]$.

Algunas soluciones “deterministas”: Interpretar, en este contexto, las transformaciones de $[0, 1]$ en $[0, 1]$ consideradas en las figuras 7 y 11. (Observe los gráficos 3D en la derecha de dichas figuras.)

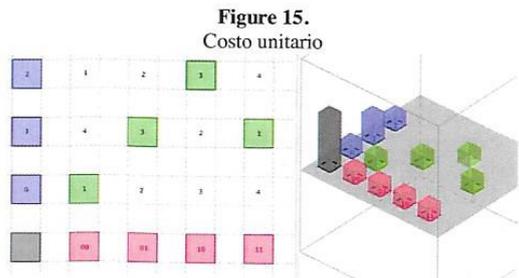
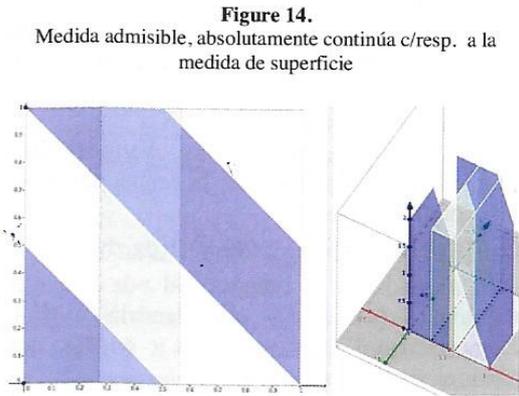
Una solución “difusa”: Desde luego, la medida producto (medida de superficie) cumple las condiciones. La figura 14 contiene un ejemplo de una solución absolutamente continua con respecto a la medida de superficie, pero diferente de ésta. Esta medida, llámémosla Π , es igual a la medida de superficie, multiplicada por la densidad $2\chi_D$, siendo D la región de color lila de la figura 14. En otras palabras,

$$\forall A \subset [0, 1] \times [0, 1] \text{ (medible), } \Pi(A) = 2 \times \text{área}(A \cap D).$$

El lector puede verificar que efectivamente Π es admisible.

4. COSTO DEL TRANSPORTE Y PROBLEMA DE MONGE

Aquí vamos a introducir un nuevo elemento: El costo del transporte.



4.1. Costo entre puntos. El costo asociado al transporte (“por unidad de masa”) de cada punto $x \in \mathcal{X}$ a cada punto $y \in \mathcal{Y}$, es una función (medible) $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$. (Otras hipótesis de regularidad serán impuestas sobre c , más adelante.)

Por ejemplo, en la figura 15, hemos asociado, a cada punto del espacio producto, un costo.

Existen $3^4 = 81$ transformaciones del espacio $\{00, 01, 10, 11\}$ en el espacio $\{0, 1, 2\}$. En el lado derecho de la figura 15 están representadas:

La medida uniforme,

$\mu(\{00\}) = \mu(\{10\}) = \mu(\{01\}) = \mu(\{11\}) = \frac{1}{4}$ y la medida

$$v(\{0\}) = v(\{2\}) = 1/4, \quad v(\{1\}) = \frac{1}{2}.$$

De las 81 transformaciones, 12 son admisibles con respecto a estas medidas. (Las transformaciones admisibles son aquellas en que 0 y 2 tienen exactamente una preimagen cada uno y 1 tiene exactamente dos preimágenes.)

Los cuadraditos verdes en la figura 15 representa el gráfico de una transformación admisible. El costo total del transporte es

$$C(T) = 1 \times \frac{1}{4} + 3 \times \frac{1}{4} + 3 \times \frac{1}{4} + 1 \times \frac{1}{4} = 2$$

4.2. El problema de Monge. Consiste en encontrar la transformación admisible que minimiza el costo total del transporte. Dada una transformación admisible T , el costo total del transporte se calcula integrando $C(x, T(x))$ con respecto a la medida μ . i.e.,

$$C(T) = \int_{\mathcal{X}} C(x, T(x)) \mu(dx)$$

(Pudiendo este valor ser infinito).

4.3. Costo total mínimo. Denotando \mathcal{A} el conjunto de las aplicaciones admisibles que transportan μ en ν , definimos

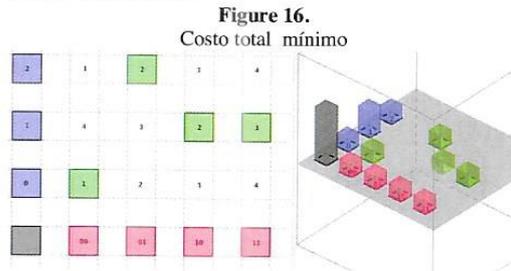
$$\mathcal{I} = \inf_{T \in \mathcal{A}} C(T)$$

(Ponemos $\mathcal{I} = \infty$, si $\mathcal{A} = \emptyset$.)

La transformación de la figura 16 tiene un costo total asociado de

$$C(T_0) = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 2 \times \frac{1}{4} + 1 \times \frac{1}{4} = \frac{3}{2}.$$

Note que este costo es el menor de entre los costos asociados a las otras 12 transformaciones admisibles. Por lo tanto, T_0 resuelve el problema de Monge. Note que esta transformación no es única, existe otra igualmente económica.



5. EL PROBLEMA DE MONGE-KANTOROVICH

En vez de minimizar el costo sobre todas las transformaciones admisibles, podemos minimizar el costo sobre todos los planes de transferencia (medidas admisibles sobre $\mathcal{X} \times \mathcal{Y}$).

El costo asociado a una medida Π es

$$C(\Pi) = \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) \Pi(dx, dy)$$

Y el objetivo es encontrar el plan de transferencia que minimiza este valor. Denotando Π el conjunto de los planes de transferencia entre μ y ν , definimos

$$\mathcal{J} = \inf_{\Pi \in \Pi} C(\Pi)$$

5.1. Planos de transferencia “deterministas”. Note que una transformación admisible $T: \mathcal{X} \rightarrow \mathcal{Y}$ corresponde a un plan de transferencia “concentrado” en el gráfico de T . Como los planes de transferencia permiten que la “masa” del conjunto de salida se disperse (i.e., partes de la producción de un sitio van para diferentes destinos), el problema de

Monge-Kantorovich permite mayor margen de maniobra, lo que puede resultar en una reducción en el costo del transporte. Resumiendo,

$$(5.1) \quad \mathcal{J} \leq \mathcal{I}.$$

(Un plan de transferencia dado por una función se llama plan de transferencia determinista.)

5.2. Monge-Kantorovich vs Monge. Existen condiciones bastante generales, bajo las cuales tenemos igualdad en (5.1). Sin embargo, las condiciones bajo las cuales existe solución del problema de Monge son muy restrictivas. En lo que resta de esta pequeña introducción al Transporte Óptimo, desarrollaremos la teoría al rededor del segundo problema, es decir el problema de Monge-Kantorovich.

En el ejemplo siguiente, $J = I$, sin embargo, únicamente el problema de Monge-Kantorovich tiene solución.

5.2.1. Ejemplo. Tomamos $\mathcal{X} [0, 1)$ y $\mathcal{Y} (-1, 1)$, ambos con la probabilidad "uniforme", i.e., la medida de longitud normalizada. Suponemos además que el costo del transporte es $C(x, y) = (x^2 - y^2)^2$.

Verificamos inmediatamente que $\mathcal{J} = 0$ y que la única solución al problema de Monge-Kantorovich es la medida "uniforme" concentrada en gráfico $x = |y|$.

Mas, cualquiera que sea $n \in \mathbb{N}$, la transformación.

$$T_n(x) = \begin{cases} 2x - \frac{2k}{2^n}, & \text{si } \frac{2k}{2^n} \leq x < \frac{2k+1}{2^n} \\ -2x + \frac{2k+1}{2^n}, & \text{si } \frac{2k+1}{2^n} \leq x < \frac{2k+2}{2^n} \end{cases} \quad 0 \leq k \leq 2^{n-1}$$

Es admisible y tiene un costo inferior a $1/n$. Por lo tanto, también $I = 0$.

5.2.2. Observación. La noción de transformación admisible no es simétrica, en cuanto que la de plan de transferencia si lo es. Note que en el ejemplo anterior la solución está concentrada en un gráfico, más la transformación va del segundo espacio en el primero. En conclusión, podemos decir que el problema de Monge-Kantorovich tiene más sentido.

6. EXISTENCIA DE PLANES DE TRANSFERENCIA ÓPTIMOS

Sean \mathcal{X} y \mathcal{Y} dos espacios métricos separables y completos. Sean μ y ν sendas medidas de probabilidad, borelianas, sobre estos espacios. i.e. μ y ν están definidas sobre los borelianos B_X y B_Y (respectivamente) y toman valores en el intervalo unidad $[0, 1]$.

Supongamos que el costo del transporte es dado por una función C :

$\mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+ \cup \{\infty\}$, semicontinua inferiormente. (Como $X \times Y$ es también un espacios métrico separable y completo, la función C es medible.)

Bajo estas condiciones, siempre existirá un plan de transferencia óptimo. i.e., un plan de transferencia, digamos Π_0 , tal que

$$\mathcal{J} = C(\Pi_0) = \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) \Pi(dx, dy).$$

De hecho, esta afirmación se basa en un argumento variacional relacionado con la topología de la convergencia débil en el espacio $P(X \times Y)$ (espacio de las medidas de probabilidad, Borelianas, sobre $X \times Y$).

Para aplicarlo, necesitamos verificar dos propiedades: La primera es la semicontinuidad inferior del funcional

$$C = \Pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) = \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) \Pi(dx, dy).$$

La segunda es la compacidad de Π en la topología de la convergencia débil. Demostraremos estos puntos en los dos párrafos siguientes.

6.1. Semicontinuidad inferior del funcional

C . Supongamos que $\{\Pi_k\}_{k \in \mathbb{N}} \subset \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ es una sucesión de medidas de probabilidad Borelianas, que converge débilmente para una medida $\Pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Entonces,

$$C(\Pi) \leq \liminf_{k \rightarrow \infty} C(\Pi_k)$$

6.1.1. Demostración. Como $C \geq 0$, existe una familia, no decreciente, de funciones continuas, digamos $\{C_l\}_{l \in \mathbb{N}}$ convergiendo puntualmente hacia C . Por el teorema de la convergencia monótona,

$$(6.1) \quad C(\Pi) = \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) \Pi(dx, dy) = \lim_{l \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} C_l(x, y) \Pi(dx, dy).$$

Por la convergencia debil, para todo $l \in \mathbb{N}$ fijo,

$$(6.2) \quad C(\Pi) = \int_{\mathcal{X} \times \mathcal{Y}} C_l(x, y) \Pi(dx, dy) = \lim_{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} C_l(x, y) \Pi_k(dx, dy)$$

Substituyendo la ecuación (6.2) en la ecuación (6.1), obtenemos

$$C(\Pi) = \lim_{l \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} C_l(x, y) \Pi_k(dx, dy)$$

Luego, como $C_l \leq C$, para todo $l \in \mathbb{N}$,

$$\liminf_{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{Y}} C_l(x, y) \Pi_k(dx, dy)$$

6.2. Compacidad. El conjunto $\Pi \subset P(X \times Y)$ es pre compacto en la topología de la convergencia débil.

6.2.1. Demostración. Según el teorema de Prokhorov, si Z es un espacio métrico, separable y completo y $\Psi \subset P(Z)$ es un subconjunto del espacio de medidas de probabilidad Borelianas, entonces Ψ es pre-

compacto si y solo si para todo $\epsilon > 0$, existe un conjunto compacto $K_\epsilon \subset Z$, tal que

$$\Pi(Z \setminus K_\epsilon) \leq \epsilon, \text{ para todo } \Pi \in \Psi.$$

Como $X \times Y$ es un espacio métrico, separable y completo (ya que cada uno de los factores lo es), para mostrar que Π es precompacto en la topología de la convergencia débil, fijamos $\epsilon > 0$, y escogemos un compacto $K \subset X$ tal que $\mu(K^c) \leq \epsilon/2$ y un compacto $L \subset Y$ tal que $\nu(L^c) \leq \epsilon/2$. Esto es posible gracias a la regularidad de las medidas de Borel. Ahora, note que

$$(X \times Y) \setminus (K_\epsilon \times L_\epsilon) \subseteq (X \setminus K_\epsilon) \cup (Y \setminus L_\epsilon).$$

Por lo tanto,

$$\Pi((X \times Y) \setminus (K_\epsilon \times L_\epsilon)) \leq \Pi(X \setminus K_\epsilon) + \Pi(Y \setminus L_\epsilon).$$

Pero como $\Pi \in \Pi$, las medidas marginales son μ y ν , respectivamente. Luego

$$\Pi(X \setminus K_\epsilon) = \mu(X \setminus K_\epsilon) \leq \epsilon/2 \text{ y } \Pi(Y \setminus L_\epsilon) = \nu(Y \setminus L_\epsilon) \leq \epsilon/2.$$

Por lo anterior,

$$(6.3) \quad \Pi((X \times Y) \setminus (K_\epsilon \times L_\epsilon)) \leq \epsilon, \text{ para todo } \Pi \in \Pi.$$

Sólo nos falta mostrar que Π es cerrado. Efectivamente, la propiedad

$$\int_{X \times Y} \varphi(x) \Pi(dx, dy) = \int_X \varphi(x) \mu(dx).$$

Para todo $\varphi: X \rightarrow \mathbb{R}$, continua y limitada, se conserva por pasaje al límite. Idem para ν . Entonces, cualquier medida adherente a Π tiene como medidas marginales a μ y ν y por lo tanto pertenece a Π .

6.3. Existencia de planes de transferencia óptimos. Gracias a las dos propiedades anteriores, ahora podemos mostrar que existe un plan de transferencia óptimo: Escogemos una sucesión $\{\Pi_k\}_{k \in \mathbb{N}} \subset \Pi$ (de planes de transferencia) tal que

$$\mathcal{J} = \lim_{k \rightarrow \infty} C(\Pi_k).$$

Por la compacidad de Π , tomando una subsucesión si es necesario, podemos asumir que $\{\Pi_k\}_{k \in \mathbb{N}}$ es débilmente convergente hacia un plan de transferencia que llamaremos $\Pi \in \Pi$. Por la semicontinuidad inferior del funcional C ,

$$C(\Pi) \leq \lim_{k \rightarrow \infty} C(\Pi_k) = \mathcal{J}.$$

6.3.1. Observación. Este teorema de existencia no garantiza que el costo mínimo sea finito! De hecho, puede suceder que todos los planes de transferencia tengan un costo infinito. Para eliminar esta situación, es usual suponer que C es $\mu \times \nu$ -integrable.

7. MONOTONÍA C-CÍCLICA Y PROBLEMA DUAL

Ahora vamos a introducir dos conceptos nuevos para después integrar- los en la caracterización de los de los planes de transferencia óptimos.

7.1. Conjuntos C-cíclicamente monótonos.

Decimos que un conjunto $\Gamma \subseteq X \times Y$ es C-cíclicamente monótono si para toda familia

$(x_1, y_1), \dots, (x_N, y_N)$ de elementos de Γ , tenemos (poniendo $y_{N+1} = y_1$)

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

Note que si redireccionamos parte del material de x_1 para y_2 , el material que iba para y_2 a partir de x_2 debería ser redireccionado para y_3 y así sucesivamente hasta completar el ciclo y que el nuevo plan sea admisible, (Ver la figura 17.)

Intuitivamente, si $\Gamma \subseteq X \times Y$ es C-cíclicamente monótono, entonces es imposible mejorar el costo del transporte con este método.

7.1.1. Ejemplos. Si $C(x, y) = c_1(x) + c_2(y)$, entonces todos los subconjuntos de $X \times Y$ son C-cíclicamente monótonos.

Mas generalmente, si $C(x, y) = h(c_1(x) + c_2(y))$, siendo h una función convexa sobre $(c_1 + c_2)^{-1}(X \times Y)$, entonces todos los subconjuntos de nivel

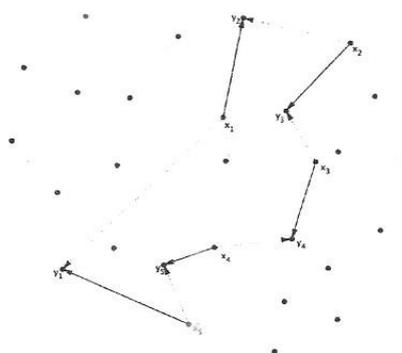
$$\{(x, y) \in X \times Y; c_1(x) + c_2(y) = a_0\}, \quad a_0 \in \mathbb{R}$$

Son C-cíclicamente monótonos.

De hecho, para todo $a_0 \in (c_1 + c_2)^{-1}(X \times Y)$, existe $m \in \mathbb{R}$ tal que para todo $a \in (c_1 + c_2)^{-1}(X \times Y)$,

$$h(a_0) + m \cdot (a - a_0) \leq h(a).$$

Figure 17.
Redistribución



Luego, para toda familia finita $\{(x_i, y_i); i = 1, \dots, N\} \subset (c_1 + c_2)^{-1}(a_0)$,

$$\begin{aligned} \sum_{i=1}^N C(x_i, y_{i+1}) &= \sum_{i=1}^N h(c_1(x_i) + c_2(y_{i+1})) \\ &\geq \sum_{i=1}^N h(a_0) + m(c_1(x_i) + c_2(y_{i+1}) - a_0) \\ &= \sum_{i=1}^N h(a_0) \\ &= \sum_{i=1}^N C(x_i, y_i) \end{aligned}$$

7.2. Problema dual. Supongamos que un transportista propone hacerse cargo del

transporte. Paga $\psi(x)$ por unidad de insumos en el sitio de producción x y los vende a $\phi(y)$ en el sitio de consumo y . En ese caso, para el consorcio productor/consumidor, el costo del transporte de x a y es $\phi(y) - \psi(x)$. Para que la propuesta del transportista sea "competitiva," debemos tener

$$(7.1) \quad \phi(y) - \psi(x) \leq C(x, y), \quad \forall (x, y) \in X \times Y.$$

Un par de "precios" (ψ, ϕ) , verificando (7.1) se llama competitivo o admisible.

El problema del transportista es maximizar la ganancia

$$(7.2) \quad G(\psi, \phi) = \int_Y \phi(y) \nu(dy) - \int_X \psi(x) \mu(dx)$$

Bajo la condición (7.1). Desde luego, para que los integrales en (7.2) tengan sentido, asumimos además que $\psi \in L^1(\mu)$ y $\phi \in L^1(\nu)$.

Este es el problema dual de Kantorovich. En él, no hay nada que haga referencia a los planes de transferencia admisibles y la desigualdad (7.1) debe verificarse en todos los puntos y no en casi todo punto, ya que como dijimos, no hay ninguna medida de referencia en $X \times Y$.

7.3. Nota. La discusión heurística previa a la condición (7.1) nos lleva a pensar que ψ y ϕ son funciones con valores reales. Sin embargo, para poder encontrar pares competitivos, es muchas veces necesario permitir que estas funciones tomen valores infinitos. Por otro lado, para que el lado izquierdo de (7.1) tenga siempre sentido, debemos exigir que $\psi > -\infty$ y que $\phi > +\infty$. i.e.

$$(7.3) \quad \psi : X \rightarrow \mathbb{R} \cup \{+\infty\} \text{ y } \phi : Y \rightarrow \mathbb{R} \cup \{-\infty\}.$$

7.3.1. Condiciones de los pares de precios admisibles/Notación. Denotaremos \mathcal{A} el conjunto de los pares (ψ, ϕ) que verifican (7.3) y (7.1).

Supondremos además que para todo $(\psi, \phi) \in \mathcal{A}$, ψ no es la constante

$+\infty$ y ϕ no es la constante $-\infty$. i.e., existe $x \in X$ tal que $\psi(x) < +\infty$

y existe $y \in Y$ tal que $\phi(y) < -\infty$.

7.3.2. Monotonía C -cíclica del conjunto de contacto. Para todo $(\psi, \phi) \in \mathcal{A}$, el conjunto de contacto,

$$(7.4) \quad \Gamma(\psi, \phi) = \{(x, y) \in X \times Y; \phi(y) - \psi(x) = C(x, y)\}, \text{ es } C\text{-cíclicamente monótono.}$$

7.3.3. Desigualdad costo total/ganancia total. Para cualquier plan de transferencia $\Pi \in \Pi$, para todo par $(\psi, \phi) \in \mathcal{A}$, $\psi \in L^1(\mu)$ y $\phi \in L^1(\nu)$, tenemos

$$\begin{aligned} \mathcal{G}(\psi, \phi) &= \int_Y \phi(y) \nu(dy) - \int_X \psi(x) \mu(dx) \\ &= \int_{X \times Y} (\phi(y) - \psi(x)) \Pi(dx, dy) \\ &= \int_{X \times Y} C(x, y) \Pi(dx, dy) \\ &= C(\Pi) \end{aligned}$$

Por lo tanto,

$$(7.5) \quad \mathcal{K} = \sup \{ \mathcal{G}(\psi, \phi); \psi \in L^1(\mu), \phi \in L^1(\nu), (\psi, \phi) \in \mathcal{A} \} \leq \mathcal{J}.$$

Por otro lado, si existe un plan de transferencia $\Pi \in \Pi$ y un par $(\psi, \phi) \in \mathcal{A}$, $\psi \in L^1(\mu)$ y $\phi \in L^1(\nu)$, tal que

$$\Pi(\phi - \psi = C) = 1,$$

Entonces $\mathcal{J} \leq C(\Pi) = G(\psi, \phi) \leq \mathcal{K}$. En consecuencia, $\mathcal{K} = \mathcal{J}$ y tanto Π como el par (ψ, ϕ) son óptimos con respecto a sus respectivos problemas.

7.3.4. Ejemplo. Sea $X = Y = [0, 1]$, con la medida de probabilidad

$(\mu = \nu)$ definida por la función de acumulación

$$F(a) = \frac{2}{\pi} \arcsin(a), \quad a \in [0, 1]$$

i.e

$$(7.6) \quad \mu(dx) = \frac{2}{\pi\sqrt{1-x^2}} dx$$

Supongamos que el costo del transporte es

$$C(x, y) = \frac{1}{x^2 + y^2}.$$

Para todo $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$(7.7) \quad 2 - y^2 - x^2 \leq \frac{1}{x^2 + y^2}$$

El conjunto de contacto es el cuarto de arco de circunferencia

$$\Gamma = \{(x, y) \in [0, 1] \times [0, 1]; x^2 + y^2 = 1\}.$$

Sobre Γ , C es constante, ya que $2 - y^2 - x^2 = \frac{1}{x^2 + y^2}$

La medida de longitud normalizada sobre Γ , llamémosla Π , es una medida admisible, pues las medidas marginales son ambas dadas por (7.3.4). (La medida de longitud normalizada de un subarco de Γ , cuyo ángulo central es $\theta \in [0, \frac{\pi}{2}]$, es $\frac{2\theta}{\pi}$).

Concluimos que Π es una medida admisible y el par $\psi(x) = x^2 - 1$ y $\phi(y) = 1 - y^2$ es competitivo $((\psi, \phi) \in \mathcal{A})$. Además,

$$\mathcal{J} \leq C(\Pi) = \int_{X \times Y} \frac{1}{x^2 + y^2} \Pi(dx, dy) = 1$$

$$\mathcal{K} \geq \int_{[0,1]} (1 - y^2) \frac{1}{\pi\sqrt{1-y^2}} dy - \int_{[0,1]} (x^2 - 1) \frac{1}{\pi\sqrt{1-x^2}} dy = 1$$

Luego, $\mathcal{K} = \mathcal{J} = 1$ y tanto Π como el par (ψ, ϕ) son óptimos con respecto a sus respectivos problemas.

Observación: El carácter óptimo del par (ψ, ϕ) no depende de las medidas marginales dadas. De hecho, cualquier medida de probabilidad P sobre Γ es un plan de transferencia admisible y óptimo, entre las medidas marginales de P , sean las que fueren, y (ψ, ϕ) es igualmente óptimo. Esto se debe simplemente a que el soporte de P está contenido en Γ que es igual al conjunto $\{(x, y); \phi(y) - \psi(x) = C(x, y)\}$.

Esta observación es de carácter general.

8. C-CONVEXIDAD/CONCAVIDAD

Al establecer un par de precios (ψ, ϕ) , es natural seleccionar para $\phi(y)$ el mayor valor posible, sin dejar de cumplir la condición (7.1). i.e.

$$(8.1) \phi(y) = \inf_{x \in X} \{\psi(x) + C(x, y)\}, \forall y \in Y.$$

De igual forma, sería de esperar que

$$(8.2) \psi(x) = \sup_{y \in Y} \{\phi(y) - C(x, y)\}, \forall x \in X.$$

Un par $(\psi, \phi) \in A$, que verifica (8.1) y (8.2) se llama ajustado.

8.1. Ajuste. Dado un par de precios $(\psi, \phi) \in A$, vamos a "ajustarlos," sin preocuparnos de momento con la integralidad, ni siquiera nos preocuparemos con la medibilidad (por ahora). Definamos

$$(8.3) \phi_1(y) = \inf_{x \in X} \{\psi(x) + C(x, y)\}, x \in X$$

Gracias a (7.1), necesariamente $\phi_1 \geq \phi$. En particular, $\phi_1 = -\infty$. Como para todo $(x, y) \in X \times Y$, $\phi_1(y) \leq \psi(x) + C(x, y)$, si $\phi_1 < +\infty$, podemos substraer $\psi(x)$ en ambos lados de la desigualdad y obtener (7.1). i.e. Si $\phi_1 < +\infty$, el par $(\psi, \phi_1) \in A$.

El conjunto de contacto del nuevo par $(\psi, \phi_1) \in A$ es mayor, i.e. $\Gamma(\psi, \phi_1) \supseteq \Gamma(\psi, \phi)$.

8.1.1. Contraejemplo. Hemos supuesto explícitamente que $\phi_1 < +\infty$ ya que la fórmula (8.3) no lo garantiza, como lo podemos verificar fácilmente tomando $X = Y = \{0, 1\}$, $C(0, y) = 0$, $C(1, y) = +\infty$, $\psi(0) = +\infty$, $\psi(1) = 0$ y $\phi(y) = 0$. En este caso, $\phi_1 \equiv +\infty$.

Proseguimos (asumiendo $\phi_1 < +\infty$) definiendo

$$\psi_1(x) = \sup_{y \in Y} \{\phi_1(y) - C(x, y)\}, y \in Y$$

Por (7.1), $\psi_1 \leq \psi = +\infty$. Como para todo $(x, y) \in X \times Y$,

$$\psi_1(x) \geq \phi(y) - C(x, y),$$

Si $\psi_1 > -\infty$, entonces $(\psi_1, \phi_1) \in A$ y el conjunto de contacto se amplía aún más, $\Gamma(\psi_1, \phi_1) \supseteq \Gamma(\psi, \phi_1)$.

Continuando, pongamos

$$\phi_2(y) = \inf_{x \in X} \{\psi_1(x) + C(x, y)\}.$$

Es fácil verificar que $\phi_2 = \phi_1$ y por lo tanto este proceso es estacionario a partir de este punto.

Esto sugiere que al considerar el problema dual de Kantorovich, podemos concentrarnos en pares ajustados de precios, e incluso obtener una de las funciones en términos de la otra (siempre y cuando podamos recuperar la medibilidad, pero de eso nos ocuparemos más tarde).

Sin embargo, si definimos ϕ a partir de (8.1), nada garantiza que ψ verifique (8.2). De hecho, la fórmula (8.2) es válida si y solamente si la función ψ tiene la forma (8.2) (pudiendo

ser otra la función en el lugar de ϕ). Esto nos sugiere la siguiente definición.

8.2. Transformada-C^Y. Estando fijo el costo del transporte C, y dada una función $\xi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$, definimos la transformada-C^Y de ξ de la siguiente manera.

$$(8.4) \xi^Y(x) = \sup_{y \in Y} \{\xi(y) - C(x, y)\}, \forall x \in X,$$

Nota: $\xi^Y : X \rightarrow \mathbb{R} \cup \{\pm\infty\}$.

8.2.1. Ejemplo. Siendo el costo

$$C(x, y) = \frac{1}{x+y}, (x, y) \in [0, 1] \times [0, 1],$$

La transformada-C^Y de $\xi = 1$ es

$$\psi(x) = \sup_{y \in Y} \left\{ \frac{1}{y} - \frac{1}{x+y} \right\} = \begin{cases} +\infty, & \text{si } x \neq 0 \\ 0, & \text{si } x = 0 \end{cases}$$

8.2.2 Ejemplo. Siendo

$$C(x, y) = \frac{1}{|x^2 - y^2 - x + y|}, (x, y) \in [0, 1] \times [0, 1],$$

La transformada-C^Y de $\xi = 0$ es

$$\psi(x) = \sup_{y \in Y} \left\{ -\frac{1}{|x^2 - y^2 - x + y|} \right\} = \begin{cases} -\frac{1}{x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8}, \\ -\frac{1}{\frac{1}{4} - x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8} \end{cases}$$

8.3. C-conconvexidad. Decimos que una función $\psi : X \rightarrow \mathbb{R} \cup \{+\infty\}$, no idénticamente igual a $+\infty$, es C-conconvexa si es la transformada-C^Y de alguna función $\xi : Y \rightarrow \mathbb{R} \cup \{-\infty\}$. i.e. ψ es C-conconvexa si

$$(8.5) \psi = \xi^Y,$$

$\Psi(x) > -\infty$, para todo $x \in X$ y $\psi(x) < +\infty$, para algún $x \in X$.

8.3.1. Ejemplo. Siguiendo las reglas de la teoría de la medida para las operaciones aritméticas con $\pm\infty$ (en particular, $+\infty \cdot 0 = 0$), la función ψ del ejemplo 8.2.1 se representa

$$\psi = +\infty \cdot \chi_{\{x=0\}}.$$

Esta función es el caso más extremo de función C-conconvexa, ya que es finita en apenas un punto.

8.3.2. Ejemplo. La función

$$\psi(x) = \begin{cases} -\frac{1}{x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8}, \\ -\frac{1}{\frac{1}{4} - x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8} \end{cases}$$

Es C-conconvexa respecto al costo $C(x, y) = \frac{1}{|x^2 - y^2 - x + y|}$ en $[0, 1] \times [0, 1]$.

(Ver el ejemplo 8.2.2.)

8.4. Transformada-C_X. La transformada-C_X de una función ζ :

$X \rightarrow \mathbb{R} \cup \{+\infty\}$ (con respecto a un costo C) se define como sigue.

$$(8.6) \zeta_X(y) = \inf_{x \in X} \{\zeta(x) + C(x, y)\}, \forall y \in Y.$$

Nota: $\zeta_X : Y \rightarrow \mathbb{R} \cup \{\pm\infty\}$.

8.4.1. *Ejemplo.* La transformada- $C_{\mathcal{X}}$ de la función $\psi = +\infty \cdot \chi_{\{x=0\}}$ del ejemplo 8.3.1, es

$$\varphi(y) = \inf_{x \in (0,1]} \left\{ +\infty + \frac{1}{x+y} \right\}^{\psi(0)+C(0,y)} = \frac{1}{y}$$

Recordemos que la función $\psi = +\infty \cdot \chi_{\{x=0\}}$ es C-convexa, ya que $\psi = \left(\frac{1}{4}\right)^y$ Y. (Ver el ejemplo 8.2.1.) Note que se tiene

$$\frac{1}{y} = \left(\left(\frac{1}{y} \right)^y \right)_x$$

8.4.2. *Ejemplo.* Calculemos ahora la transformada- $C_{\mathcal{X}}$ de la función ψ , del ejemplo 8.3.2. Poniendo

$$\phi_1(y) = \inf_{x(1-x) \geq \frac{1}{8}} \left\{ \frac{1}{|x^2 + y^2 - x + y|} - \frac{1}{x(1-x)} \right\}$$

Y

$$\phi_2(y) = \inf_{x(1-x) < \frac{1}{8}} \left\{ \frac{1}{|x^2 + y^2 - x + y|} - \frac{1}{x(1-x)} \right\}$$

La transformada- $C_{\mathcal{X}}$ de ψ es $\varphi(y) = \phi_1(y) \wedge \phi_2(y)$. Dejamos al lector verificar que

$$(8.7) \quad \phi^Y(x) = \begin{cases} -\frac{1}{x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8}, \\ -\frac{1}{\frac{1}{4}-x(1-x)}, & \text{si } x(1-x) < \frac{1}{8} \end{cases}$$

8.5. **C-concavidad.** Decimos que una función $\varphi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$, no idénticamente igual a $-\infty$, es C-cóncava si es la transformada- $C_{\mathcal{X}}$ de alguna función $\zeta: X \rightarrow \mathbb{R} \cup \{+\infty\}$. i.e. φ es C-cóncava si

$$(8.8) \quad \varphi = \zeta_{\mathcal{X}},$$

$\varphi(y) < +\infty$, para todo $y \in Y$ y $\varphi(y) > -\infty$, para algún $y \in Y$.

8.5.1. *Ejemplo.* La función $\frac{1}{y}$ es C-cóncava respecto a $C(x, y) = \frac{1}{x+y}$ en $[0, 1] \times (0, 1]$. (Ver el ejemplo 8.4.1.)

8.5.2. *Ejemplo.* La función φ , definida en (8.7), es C-cóncava. Además, el lector puede verificar que para todo $x \in \mathcal{X}$,

$$\phi^Y(x) = \begin{cases} -\frac{1}{x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8}, \\ -\frac{1}{\frac{1}{4}-x(1-x)}, & \text{si } x(1-x) < \frac{1}{8} \end{cases}$$

Que es la función ψ del ejemplo 8.3.2.

8.6. **Sumario de algunos hechos inmediatos.**

Un par $(\psi, \varphi) \in A$ es ajustado si y solamente si $\psi = \varphi^Y$ y $\varphi = \psi_{\mathcal{X}}$.

Siendo $\xi: \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ una función cualquiera, entonces ξ^Y es C-convexa si y solo si $\xi^Y > -\infty$ y $\xi^Y = +\infty$ (si y solo si $\xi^Y, \xi \in A$).

Siendo $\zeta: \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$ una función cualquiera, entonces $\zeta_{\mathcal{X}}$ es C-cóncava si y solo si $\zeta_{\mathcal{X}} < +\infty$ y $\zeta_{\mathcal{X}} = -\infty$ (si y solo si $(\zeta, \zeta_{\mathcal{X}}) \in A$).

A). En particular, para todo $\xi: Y \rightarrow \mathbb{R} \cup \{-\infty\}$ tal que ξ^Y es C-convexa,

$(\xi^Y)_{\mathcal{X}}$ es C-cóncava si y solamente si

$$((\xi^Y), (\xi^Y)_{\mathcal{X}}) \in A.$$

Por último, para todo $\zeta: X \rightarrow \mathbb{R} \cup \{+\infty\}$ tal que $\zeta_{\mathcal{X}}$ es C-cóncava, $(\zeta_{\mathcal{X}})^Y$ es C-convexa si y solamente si

$$(\zeta_{\mathcal{X}})^Y, \zeta_{\mathcal{X}} \in A.$$

8.7. **Lema.** Sea $\xi: \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ una función tal que ξ^Y es C-convexa y $(\xi^Y)_{\mathcal{X}}$ es C-cóncava, i.e. $((\xi^Y), (\xi^Y)_{\mathcal{X}}) \in A$. Entonces

$$(8.9) \quad \xi^Y = ((\xi^Y)_{\mathcal{X}})^Y$$

En otras palabras, $((\xi^Y), (\xi^Y)_{\mathcal{X}})$ es ajustado.

8.8. **Corolario.** Sea $\psi: X \rightarrow \mathbb{R} \cup \{+\infty\}$ una función tal que $(\psi, \psi_{\mathcal{X}}) \in A$. Entonces ψ es C-convexa si y solamente si

$$\psi = (\psi_{\mathcal{X}})^Y.$$

8.9. **C-subdiferencial.** Si $(\psi, \psi_{\mathcal{X}}) \in A$ y ψ es C-convexa, decimos que ψ y $\psi_{\mathcal{X}}$ son C-conjugadas y definimos el C-subdiferencial de ψ como siendo el conjunto de contacto $\Gamma(\psi, \psi_{\mathcal{X}})$ (ver (7.4)). i.e.

$$(8.10) \quad \partial C \psi = \{(x, y) \in X \times Y; \psi_{\mathcal{X}}(y) - \psi(x) = C(x, y)\}.$$

Para todo $x \in \mathcal{X}$ fijo, el C-subdiferencial de ψ en el punto x es

$$\partial C \psi(x) = \{y \in Y; (x, y) \in \partial C \psi\}.$$

Note que $y \in \partial C \psi(x)$ si y solamente si $\forall z \in \mathcal{X}$,

$$\psi(x) + C(x, y) \leq \psi(z) + C(z, y).$$

8.9.1. *Ejemplo.* Las funciones $\psi = +\infty \cdot \chi_{\{y=1\}}$ y $\varphi(y)=1$ son conjugadas. (Ver el ejemplo 8.4.1.) El subdiferencial de ψ es

$$\partial C \psi = \{(x, y) \in [0, 1] \times (0, 1]; \frac{1}{y} - \infty \cdot \chi_{\{x \neq 0\} = \frac{1}{x+y}}\} = \{(0, y); y \in (0, 1]\}$$

Por lo tanto,

$$\partial C \psi(x) = \begin{cases} (0, 1], & \text{si } x = 0, \\ \emptyset, & \text{si } x \neq 0 \end{cases}$$

8.9.2. *Ejemplo.* Para poner en evidencia la simetría que existe entre las funciones de los ejemplos 8.4.2 y 8.5.2, las escribiremos de la siguiente forma

$$\psi(x) = \begin{cases} -\frac{1}{x(1-x)}, & \text{si } x(1-x) \geq \frac{1}{8}, \\ -\frac{1}{1-4x(1-x)}, & \text{si } x(1-x) < \frac{1}{8} \end{cases}$$

$$\phi(x) = \begin{cases} -\frac{1}{y(1-y)}, & \text{si } y(1-y) \geq \frac{1}{8}, \\ -\frac{1}{1-4y(1-y)}, & \text{si } y(1-y) < \frac{1}{8} \end{cases}$$

Es decir, $\varphi(y) = \psi(y) - 4$. Estas dos funciones son conjugadas respecto al costo

$$C(x, y) = \frac{1}{|x^2 - y^2 - x + y|}, (x, y) \in [0, 1] \times [0, 1].$$

Dejamos al lector la tarea de encontrar el conjunto $\partial C \psi$.

8.10. C -superdiferencial. Sea $(\psi, \phi) \in A$ es un par ajustado, i.e. ψ y ϕ son C -conjugadas. El C -superdiferencial de ϕ es lo mismo que el C -subdiferencial de ψ . i.e.

$$(8.11) \quad \partial^C \phi = \{(x, y) \in \mathcal{X} \times \mathcal{Y}; \phi(y) - \psi(x) = C(x, y)\}.$$

Mas ahora centramos la atención en la función ϕ y consideramos ψ únicamente como resultado de la transformada-CX de ϕ ($\psi = \phi^y$). Así, para todo $y \in Y$ fijo, el C -superdiferencial de ϕ en el punto y es

$$\partial^C \phi(y) = \{x \in \mathcal{X}; (x, y) \in \partial C \phi\}$$

Note que $x \in \partial^C \phi(y)$ si y solamente si

$$\forall z \in \mathcal{Y}, \phi(y) - C(x, y) \geq \phi(z) - C(x, z).$$

8.10.1. *Ejemplo.* Como vimos en el ejemplo 8.9.1, las funciones $\psi = +\infty \cdot \chi_{\mathcal{X}^c}$ y $\phi(y) = 1$ son conjugadas. El superdiferencial de ϕ es $\partial C \phi = \{(0, y); y \in (0, 1]\}$.

Por lo tanto, para todo $y \in (0, 1]$, $\partial C \phi(y) = \{0\}$.

9. DESCRIPCIÓN DE LOS PLANES DE TRANSFERENCIA ÓPTIMOS

En 7.3.3, vimos que para todo $\Pi \in \Pi$ y para todo $(\psi, \phi) \in A$, tal que $\psi \in L^1(\mu)$ y $\phi \in L^1(\nu)$, tenemos (desigualdad (7.5))

$$\mathcal{K} \leq \mathcal{J}.$$

Después dimos un ejemplo explícito en el cual $\mathcal{K} = \mathcal{J}$. (Ver el Ejemplo 7.3.4.) La forma de establecer esta igualdad es exhibiendo un plan de transferencia $\Pi \in \Pi$ y un par de funciones integrables $(\psi, \phi) \in A$, tal que el conjunto de contacto $\Gamma(\psi, \phi)$ tenga Π -medida total. i.e.

$$\Pi(\Gamma(\psi, \phi)) = 1.$$

Integrando el costo C y la diferencia $\phi - \psi$ con respecto a Π , obtenemos

$C(\Pi) = G(\psi, \phi)$ y por lo tanto, se verifica la desigualdad inversa:

$$\mathcal{J} \leq C(\Pi) = G(\psi, \phi) \leq \mathcal{K}.$$

El teorema de la dualidad de Kantorovich dice que, bajo condiciones extremadamente generales, $\mathcal{K} = \mathcal{J}$ y que, en muchos casos, el problema dual es también resoluble, i.e., existe al menos un par $(\psi, \phi) \in A$, $\psi \in L^1(\mu)$, $\phi \in L^1(\nu)$, tal que $\mathcal{K} = G(\psi, \phi)$.

9.1. Teorema de la dualidad. Sean (\mathcal{X}, μ) y (\mathcal{Y}, ν) dos espacios métricos separables y completos, con sendas probabilidades borelianas μ y ν . Sea $C : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ una función semicontínua inferiormente. Entonces,

$$\mathcal{K} = \mathcal{J}$$

Denotando

$$L^1_+(\mu) = \{\psi \in L^1(\mu); \psi > -\infty, \psi \neq +\infty\}$$

$$L^1_+(\nu) = \{\phi \in L^1(\nu); \phi > -\infty, \phi \neq -\infty\},$$

Tenemos además las igualdades

$$\mathcal{K} = \sup_{\psi \in L^1_+(\mu)} G(\psi, \psi_x) = \sup_{\phi \in L^1_+(\nu)} G(\phi^y, \phi)$$

En los supremos anteriores, podríamos considerar únicamente funciones C -convexas y C -cóncavas, respectivamente, sin alterar el resultado.

Por otro lado, nada garantiza que el valor común sea finito.

9.2. Caracterización de los planes de transferencia óptimos. Si C es una función real ($C < +\infty$) y J es finito, entonces las siguientes afirmaciones son equivalentes:

(1) $\Pi \in \Pi$ es óptimo.

(2) Existe un conjunto medible $\Gamma \Pi \subseteq X \times Y$, C -cíclicamente monótono, tal que $\Pi(\Gamma \Pi) = 1$.

(3) Existe una función C -convexa, medible, denotémosla ψ , tal que $\psi_x(y) - \psi(x) = C(x, y)$ sobre un conjunto C -cíclicamente monótono, de medida Π -total. i.e. Existe un conjunto medible $\Gamma \Pi \subseteq \partial C \psi$, C -cíclicamente monótono, tal que $\Pi(\Gamma \Pi) = 1$. (Nada garantiza que el conjunto $\partial C \psi$, definido por esta igualdad, sea medible.)

(4) Existe un par de funciones medibles $(\psi, \phi) \in A$, tal que

$$\Pi(\Gamma(\psi, \phi)) = 1.$$

Además, bajo las condiciones de este teorema (i.e., C real y \mathcal{J} finito), existe un conjunto

medible $\Gamma_U \subseteq \mathcal{X} \times \mathcal{Y}$, C -cíclicamente monótono, tal que para todo $\Pi \in \Pi$, $\Pi(\Gamma_U) = 1$ si y solamente si Π es óptimo.

(Note que Γ_U es el mismo para todos los planes de transferencia.)

9.3. Existencia de soluciones del problema dual. Si $C < +\infty$ y

$$\int_{\mathcal{X} \times \mathcal{Y}} C(x, y) \mu \otimes \nu(dx, dy) < +\infty,$$

Entonces el problema dual también tiene solución. i.e., existe un par $(\psi, \phi) \in A$, $\psi \in L^1(\mu)$ y $\phi \in L^1(\nu)$, tal que

$$\mathcal{K} = \int_{\mathcal{Y}} \phi(y) \nu(dy) - \int_{\mathcal{X}} \psi(x) \mu(dx)$$

Podemos incluso exigir que ψ sea C -convexa y que $\phi = \psi_x$ sobre un conjunto de medida ν -total.

Bajo las condiciones de este teorema y asumiendo además que C es continua, existe un conjunto cerrado $\Gamma_U \subseteq \mathcal{X} \times \mathcal{Y}$, C -cíclicamente monótono, tal que

(1) Para todo $\Pi \in \Pi$, $\Pi(\Gamma_U) = 1$ si y solamente si Π es óptimo.

(2) Para toda $\psi \in L^1(\mu)$, C -convexa, el par (ψ, ψ_λ) es óptimo si y solamente si $\partial^C \psi \supseteq \Gamma_U$.

Las demostraciones de estos teoremas (9.1, 9.2 y 9.3) son en la mayor parte simples, mas contienen algunos puntos sutiles, que requieren del lector un buen dominio de la teoría de la medida. En este primer abordaje, preferimos no demostrarlos y pasar a discutir más bien las conclusiones.

10. CONCLUSIONES

La formulación moderna del problema del Transporte Óptimo, en espacios métricos y costos extremadamente generales, permiten que el ámbito de aplicaciones de esta teoría se amplíe a horizontes impensables en el pasado.

El desarrollo llevado a cabo entre los años 40's y 60's por Frank Lauren Hitchcock, L. Kantorovich, T. Koopmans, y otros, se centra en problemas económicos relacionados con la asignación óptima de recursos y la programación lineal. El desarrollo posterior del Transporte

Óptimo, en especial a partir de los años 80's se encuentra íntimamente ligado a reformulaciones y generalizaciones de principios matemáticos fundamentales de teorías como los Sistemas Dinámicos, la formulación matemática de la Mecánica de Fluidos Incompresibles, las Ecuaciones Diferenciales Parciales, permitiendo grandes avances en los resultados y un mejor entendimiento de los fenómenos envueltos.

Sin embargo, fuera de la Matemática y Física Matemática, y a excepción de la etapa pionera de mediados del siglo pasado, prácticamente no hay estudios de aplicación de esta teoría, que al momento cuenta con una gran cantidad de resultados cualitativos muy precisos sobre su estructura. El encuadramiento del Transporte Óptimo en variedades diferenciables abre aún más las posibilidades de aplicación.

Desde luego, también podemos considerar costos que provengan de integrales de tipo Lagrangiano, o de otro tipo, introduciendo la dimensión temporal. Aquí hemos omitido este aspecto, para mantener esta discusión a un nivel introductorio.

A pesar de estos avances formidables, en tantos frentes, cuando pensamos en aplicaciones, hay siempre situaciones que salen del marco teórico. Por ejemplo, si queremos aplicar el Transporte Óptimo en la distribución de energía eléctrica, vamos a tener que estudiar los efectos de la pérdida de energía en el transporte, algo que no está considerado en la teoría. Otro aspecto importante es la fluctuación en la demanda, lo que podría resultar en una dependencia temporal intrincada.

En resumen, fuera del círculo de especialistas, poco se conoce de los logros y potencialidades de esta fascinante teoría. Las aplicaciones del Transporte Óptimo, fuera del ámbito de la Matemática, están apenas comenzando, por lo que este es un campo muy prometedor para la investigación tanto en Matemática pura como aplicada.

REFERENCIAS BIBLIOGRÁFICAS

- [1]. L. Ambrosio, Lecture notes on optimal transport problems, in *Mathematical aspects of evolving interfaces*, CIME summer school held in Funchal (Portugal), Springer LMN, 2003.
- [2]. L. Ambrosio and N. Gigli, *A user's guide to optimal transport*, 2012.
- [3]. Feldman M Caffarelli, L. and R. McCann, Constructing optimal maps for monge's transport problem as a limit of strictly convex costs, *J. Amer. Math. Soc.* 15 (2002), 1–26 (electronic).
- [4]. L. Evans and W. Gangbo, Differential equations methods for the monge-kantorovich mass transfer problem, vol. 137, *Mem. Amer. Math. Soc.*, 1999.
- [5]. W. Gangbo, The monge mass transfer problem and its applications, in *monge ampère equation: applications to geometry and optimization*, *Contemp. Math.*, Amer. Math. Soc. 226 (1999), 79–104.
- [6]. W. Gangbo and R. McCann, The geometry of optimal transportation, *Acta Math.* 177 (1996), 113–161.
- [7]. N. Gigli, On the geometry of the space of probability measures in R^n endowed with the quadratic optimal transport distance.
- [8]. L. Kantorovich, On an effective method of solving certain classes of extremal problems, *Dokl. Akad. Nauk. USSR* 28 (1940), 212–215, English translation in *J. Math. Sci.* 133, 4 (2006).
- [9]. , On the translocation of masses, *Dokl. Akad. Nauk. USSR* 37 (1942), 199–201.
- [10]. G. Monge, Mémoire sur la théorie des déblais et des remblais, *Histoire de l'Académie Royale des Sciences de Paris* (1781), 666–704.
- [11]. A. Pratelli, On the equality between monge's infimum and kantorovich's mini-mum in optimal mass transportation, *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 43 (2007), 1–13.
- [12]. S. Rachev and L. Rüschendorf, *Mass transportation problems. vol. i, Probability and its Applications*, Springer-Verlag, New York, 1998.
- [13]. V. Sudakov, Geometric problems in the theory of infinite dimensional probability distributions, *Proceedings of Steklov Institute* 141 (1979), 1–178.
- [14]. H. Thorisson, *Coupling, stationarity, and regeneration*, Springer-Verlag, 2000.
- [15]. N. Trudinger and X.-J. Wang, On the monge mass transfer problem, *Calc. Var. Partial Differential Equations* 13 (2001), 19–31.
- [16]. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
- [17]. *Optimal transport, old and new*, Springer-Verlag, 2008.
- [18]. Q. Xia, Optimal paths related to transport problems, *Commun. Contemp. Math.* 5 (2003), 251–279.
- [19]. , Interior regularity of optimal transport paths, *Calc. Var. Partial Differential Equations* 20 (2004), 283–299.