

Sesión II: Transporte Óptimo

Daniel López-Castaño
Johns Hopkins University

Geometría en Ciencias de Datos
Universidad Nacional de Colombia
Agosto 18-21

Un t3pico que ha sido parte de la far3ndula matem3tica desde siempre



Monge



Kantorovich

Nobel



Wasserstein



Brenier



Caffarelli



McCann



Otto



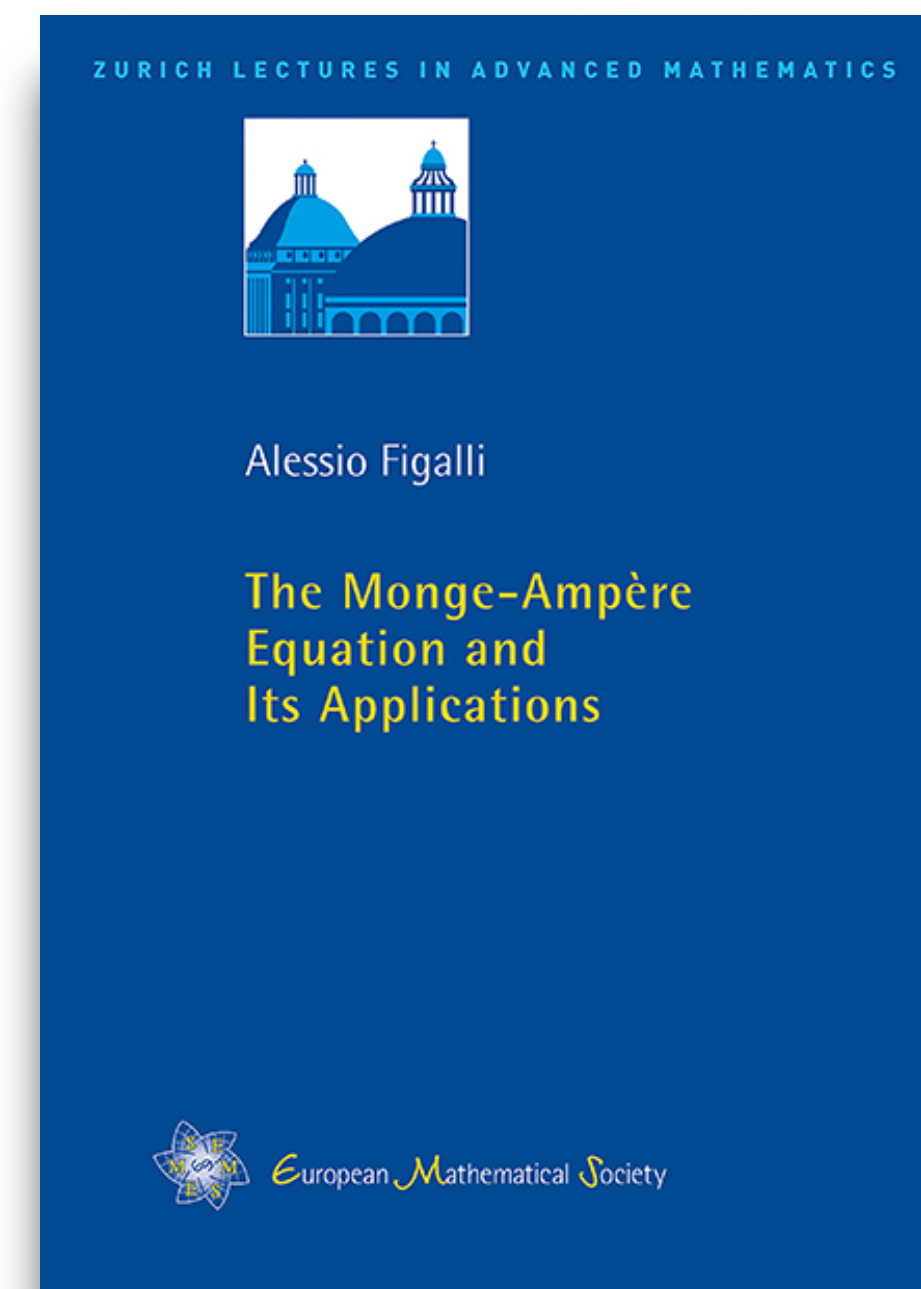
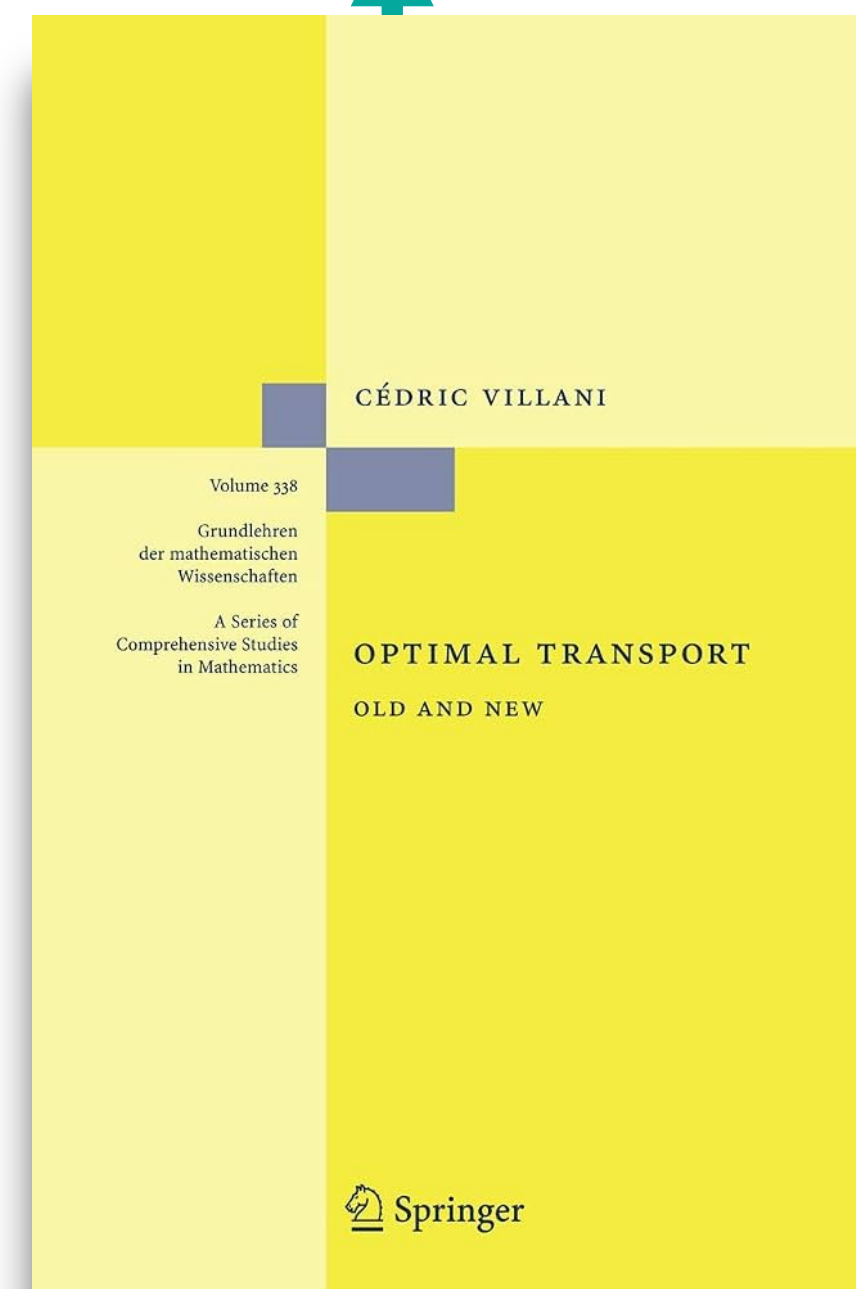
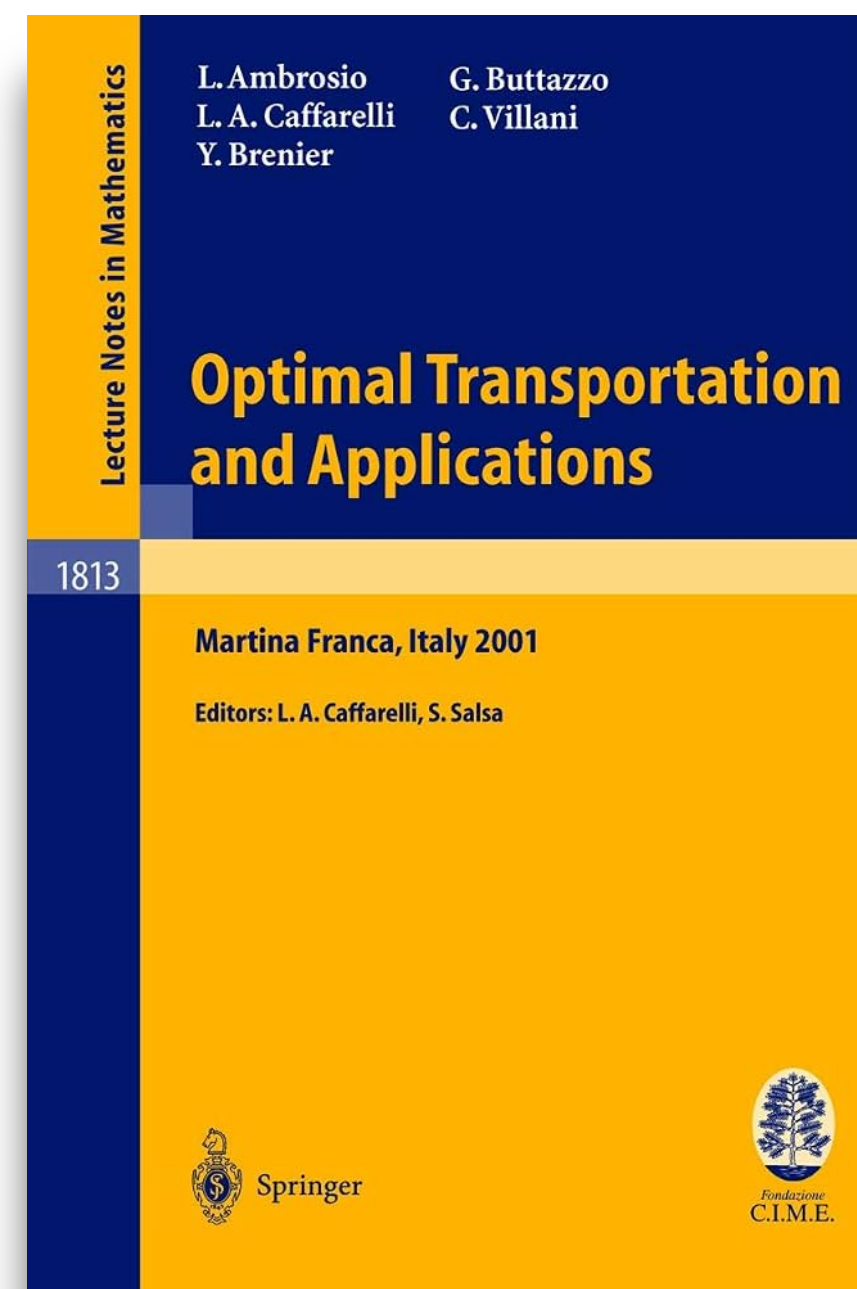
Villani

Fields



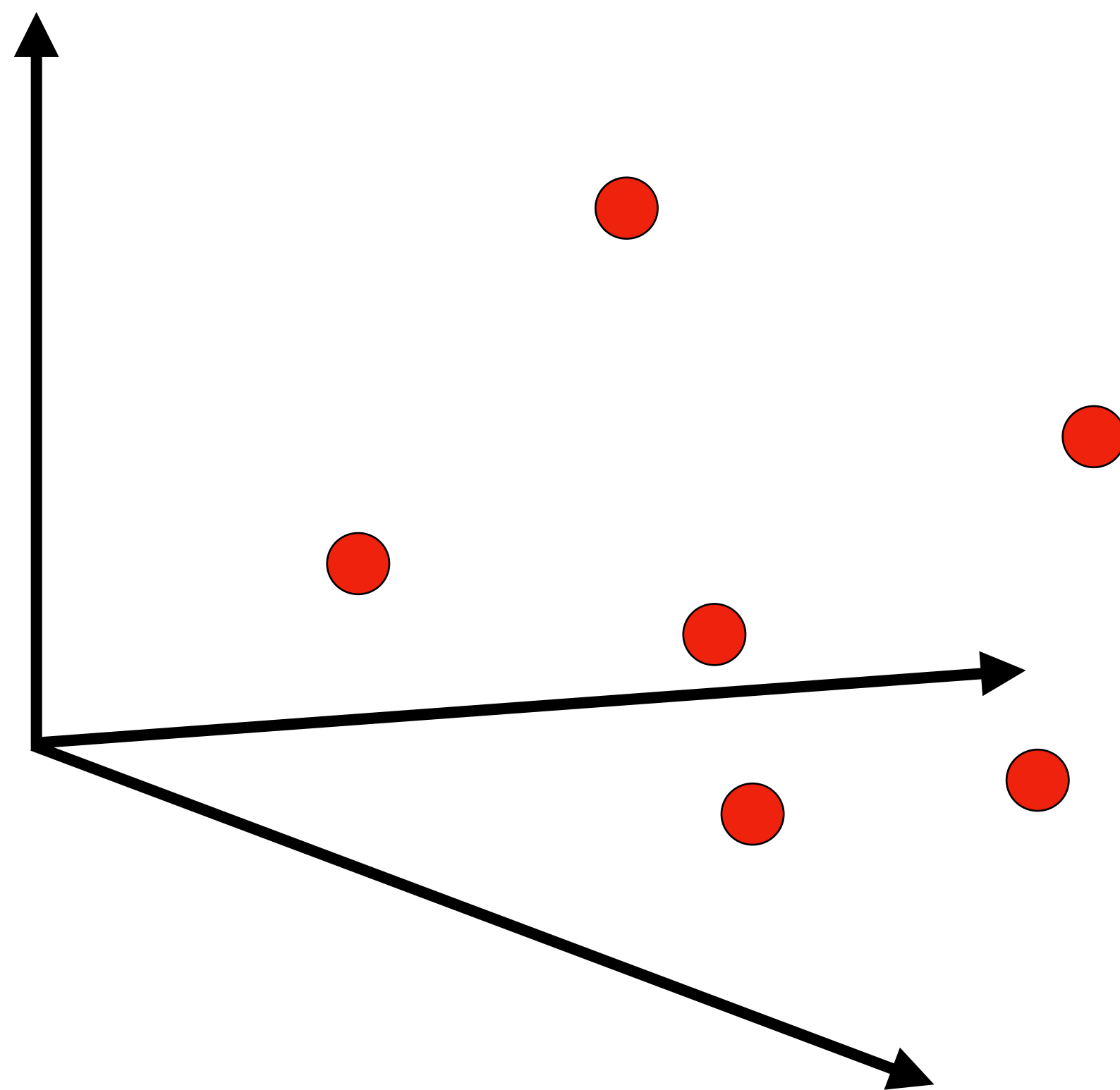
Figalli

Fields

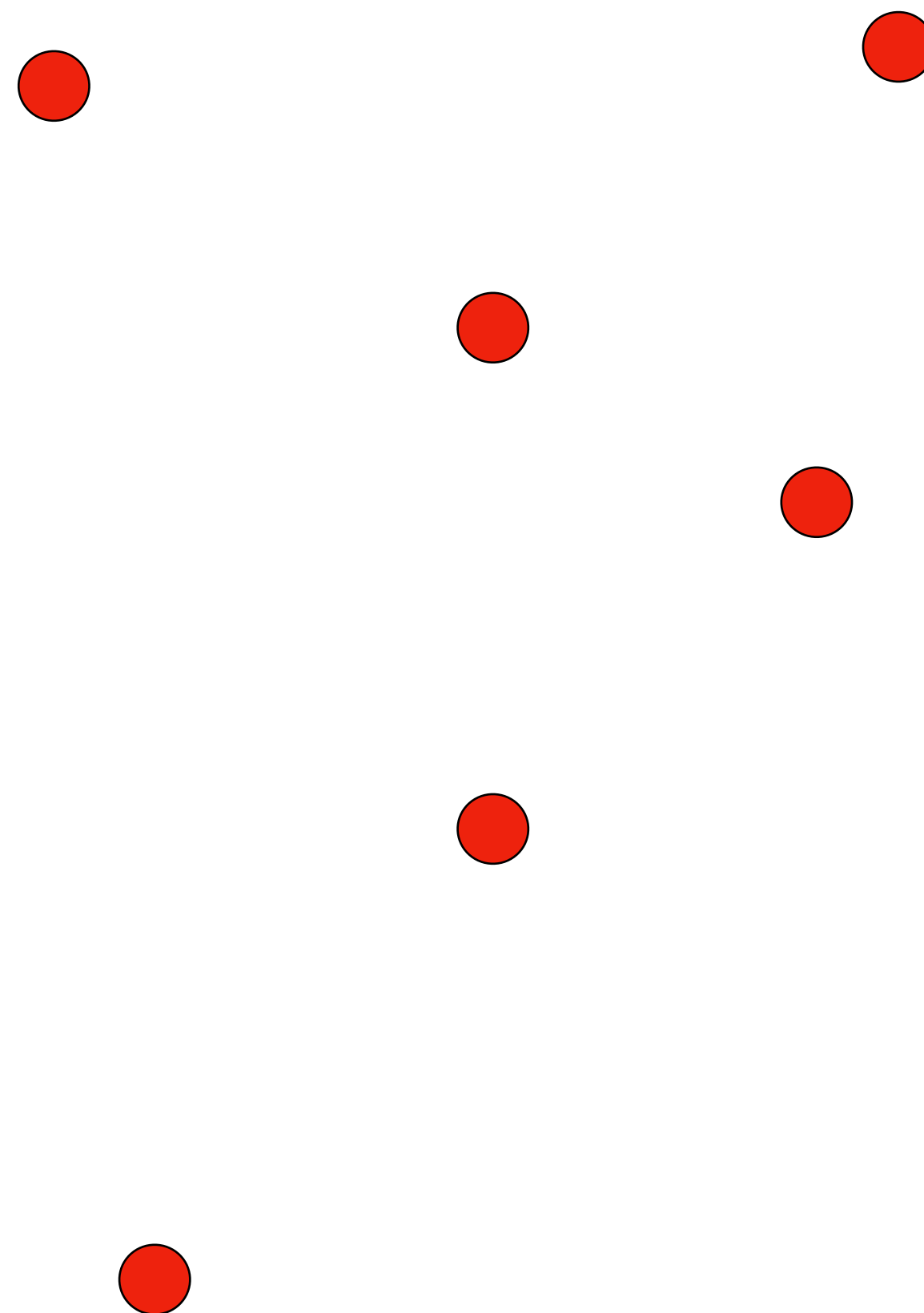
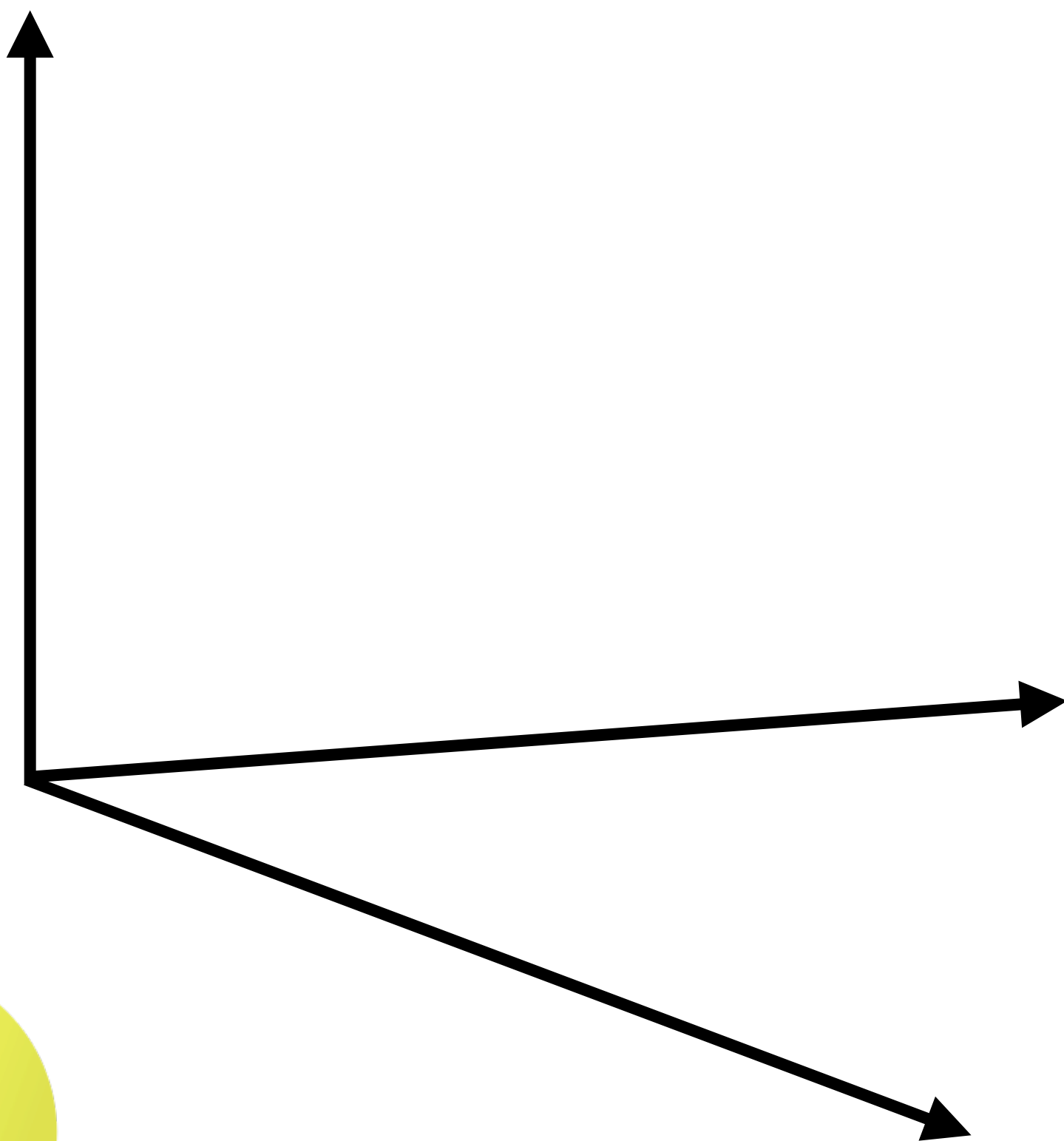


¿Qué es transporte óptimo?

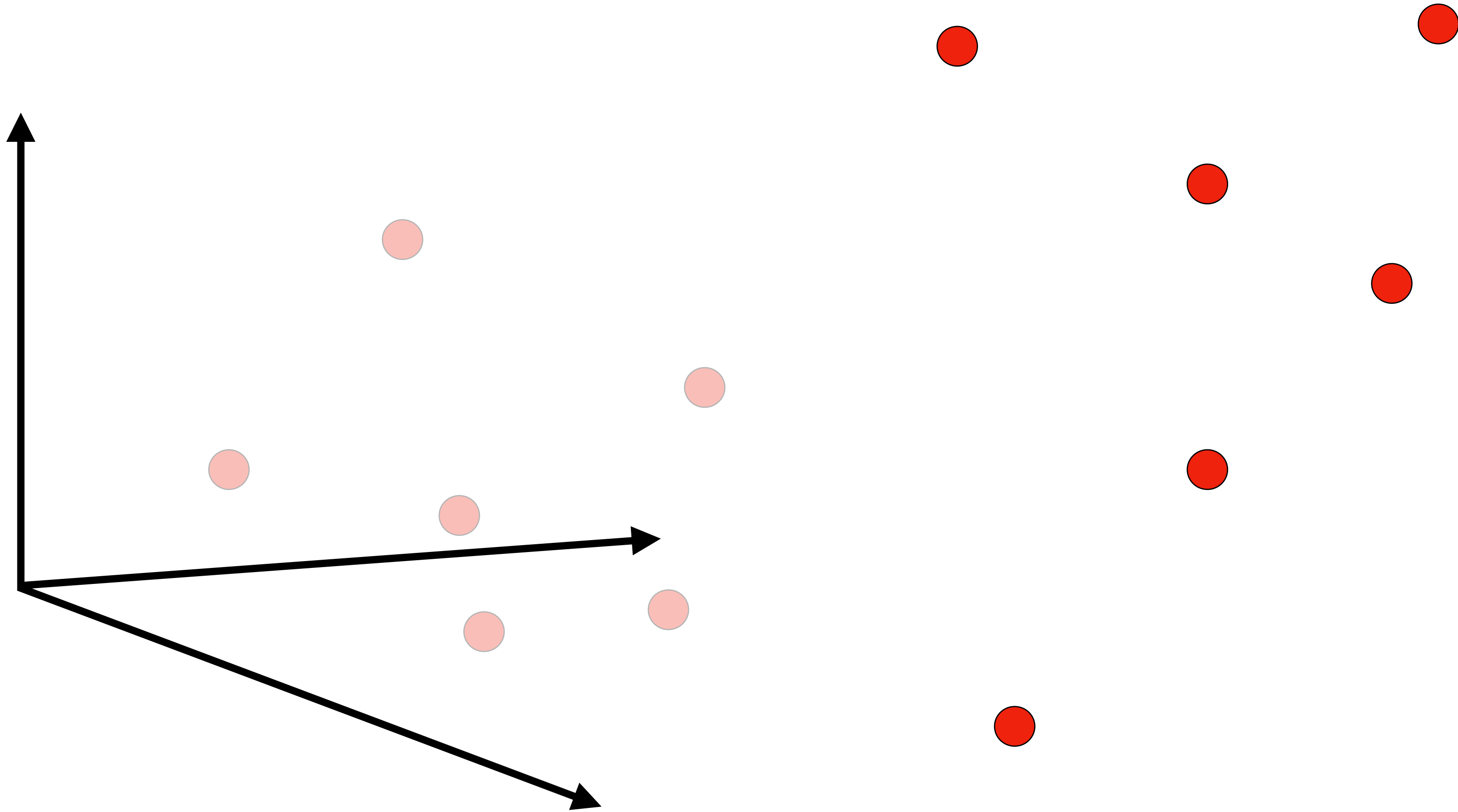
**Una manera geométrica
de comparar medidas de
probabilidad**



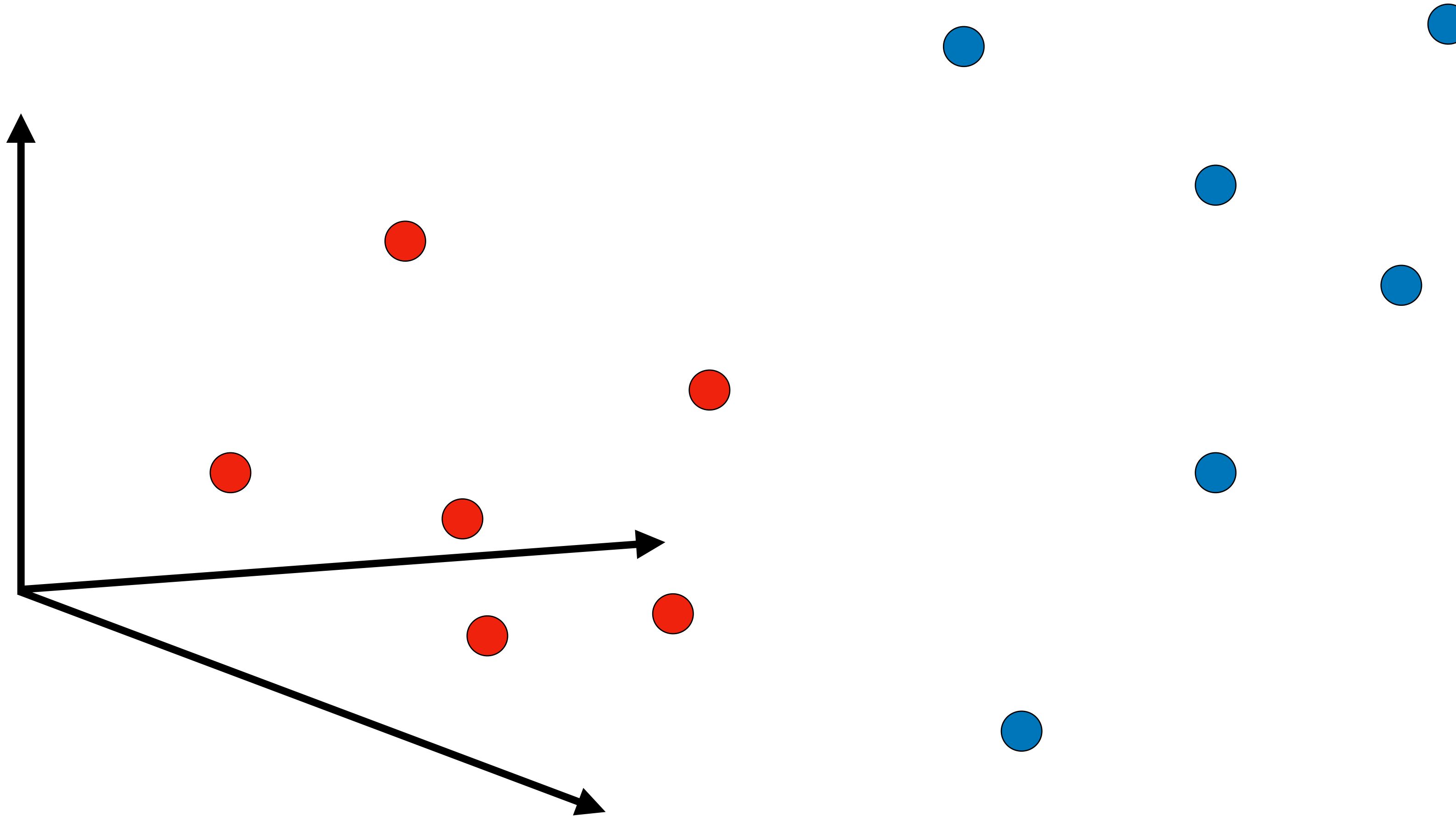




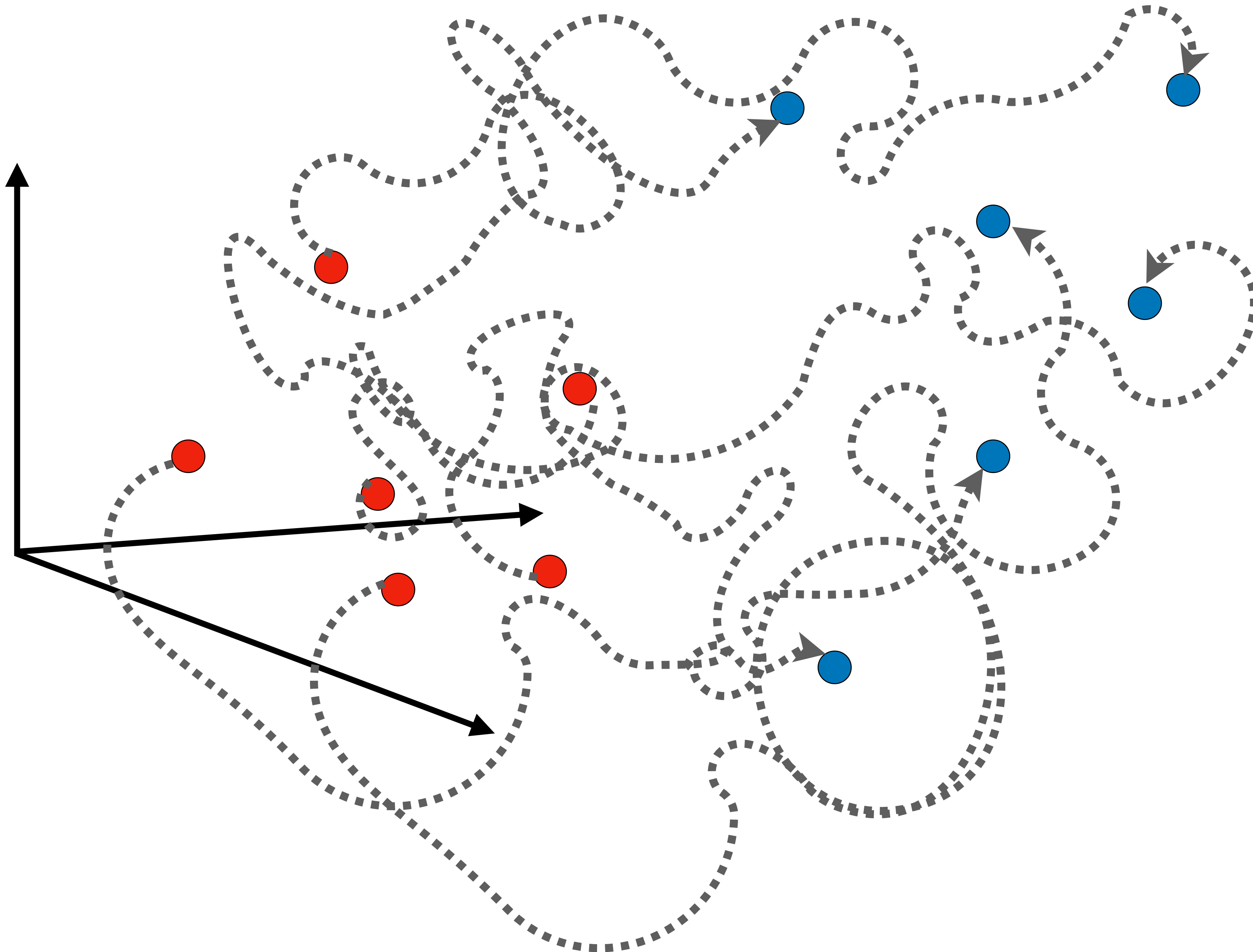
¿Quién fue a dónde?



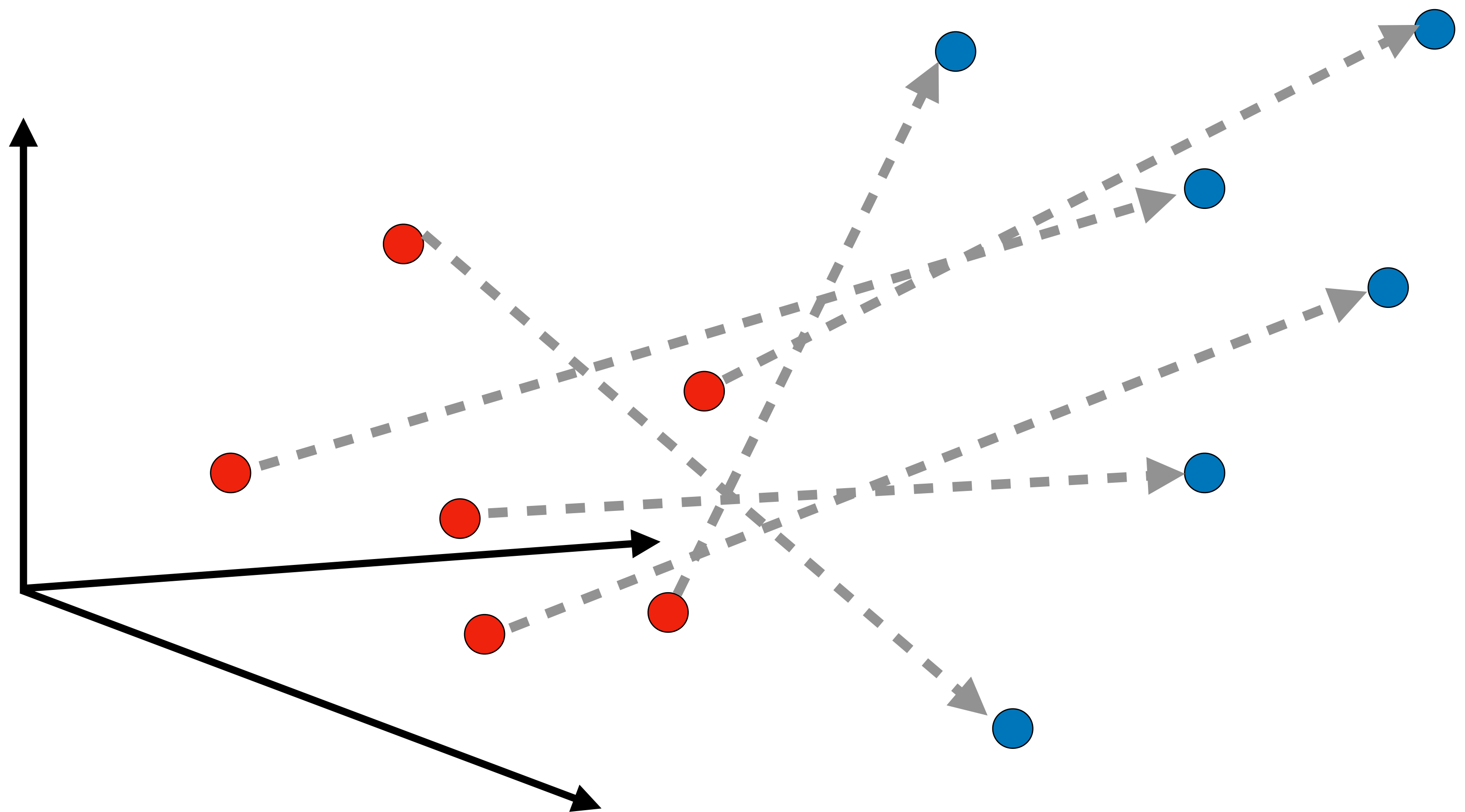
¿Quién fue a dónde?



¿Quién fue a dónde?

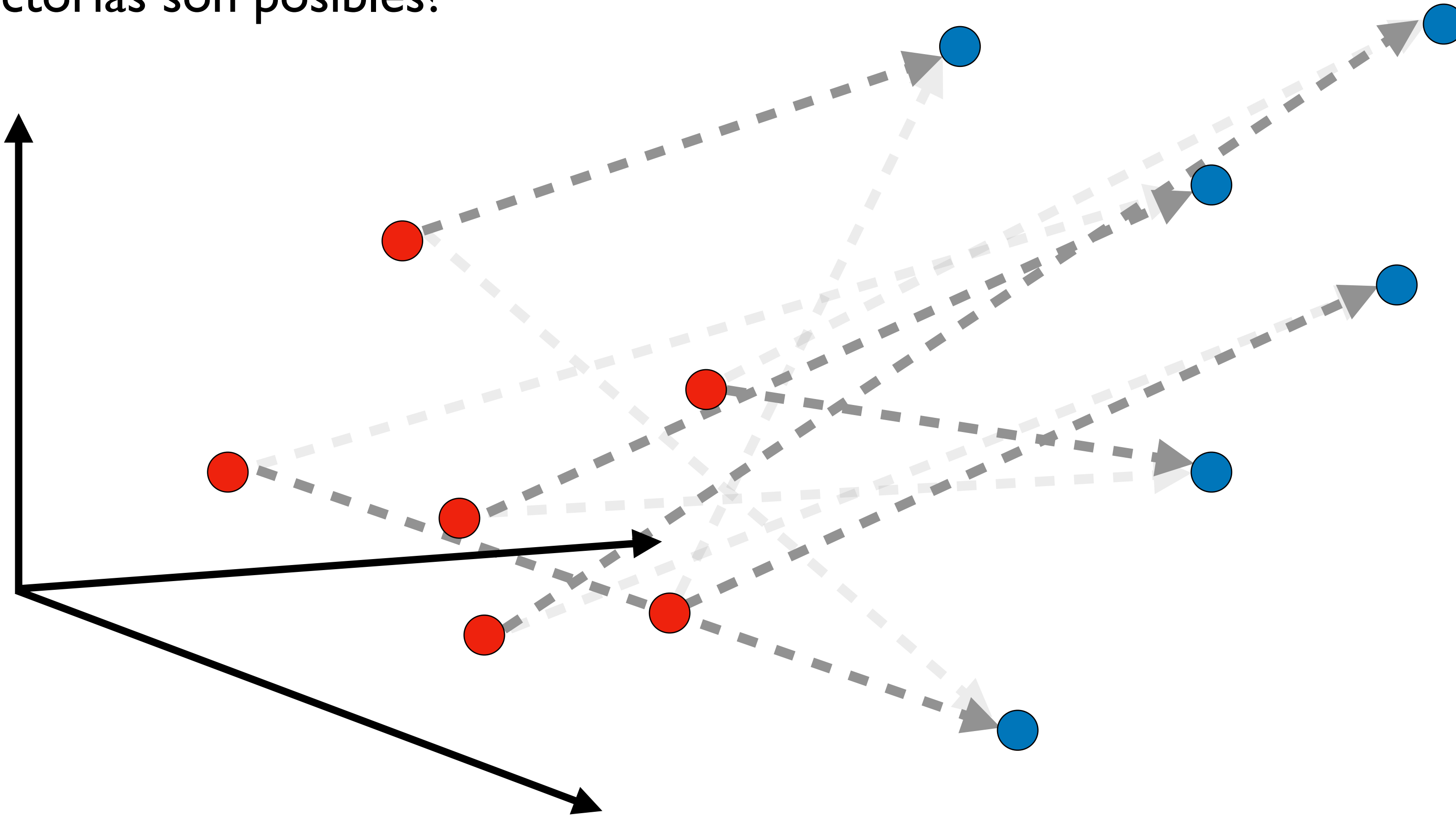


Asumamos que se desplazaron en linea recta



Asumamos que se desplazaron en linea recta

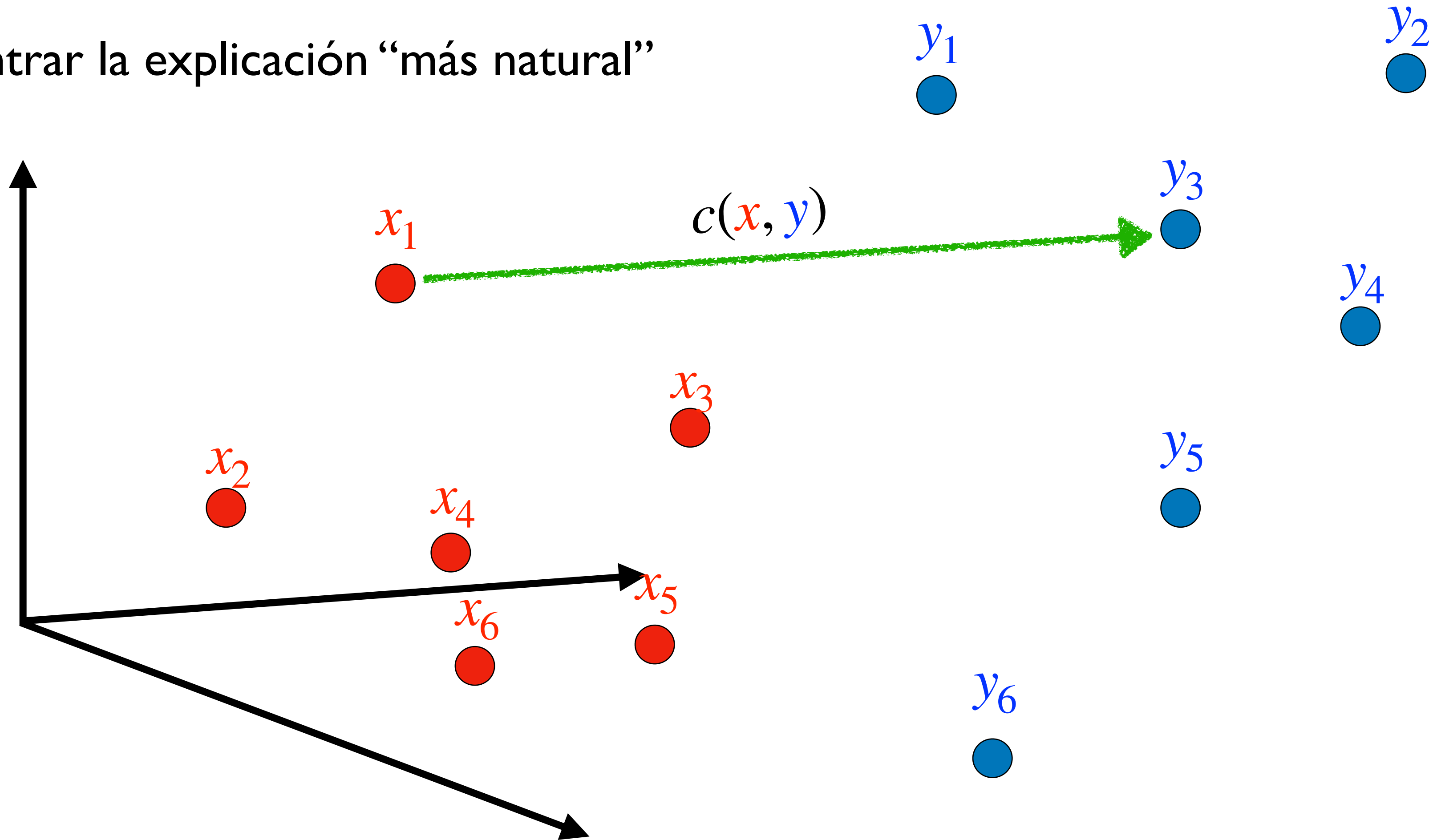
¿Cuántas trayectorias son posibles?



Hay $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ permutaciones posibles

Transporte óptimo v.0.0 aka problema de asignación lineal

Misión: Encontrar la explicación “más natural”

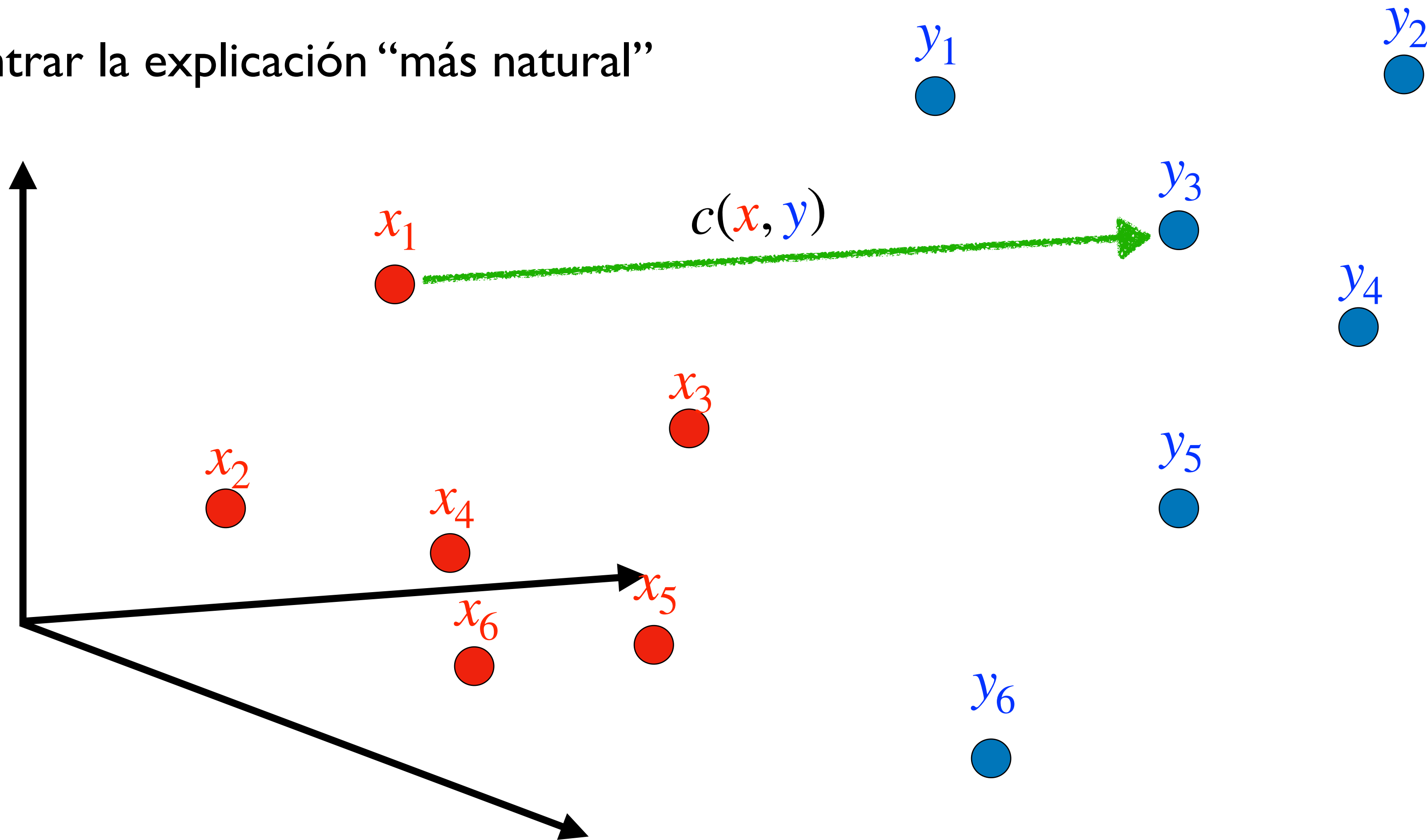


“natural” va a estar codificado con costos asociados a pares de puntos, y buscamos encontrar la mejor permutación

$$\min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^n c(x_i, y_{\sigma(i)})$$

Transporte óptimo v.0.0 aka problema de asignación lineal

Misión: Encontrar la explicación “más natural”



para resolver el problema de asignación lineal podemos correr el *algoritmo húngaro* o cualquier variante

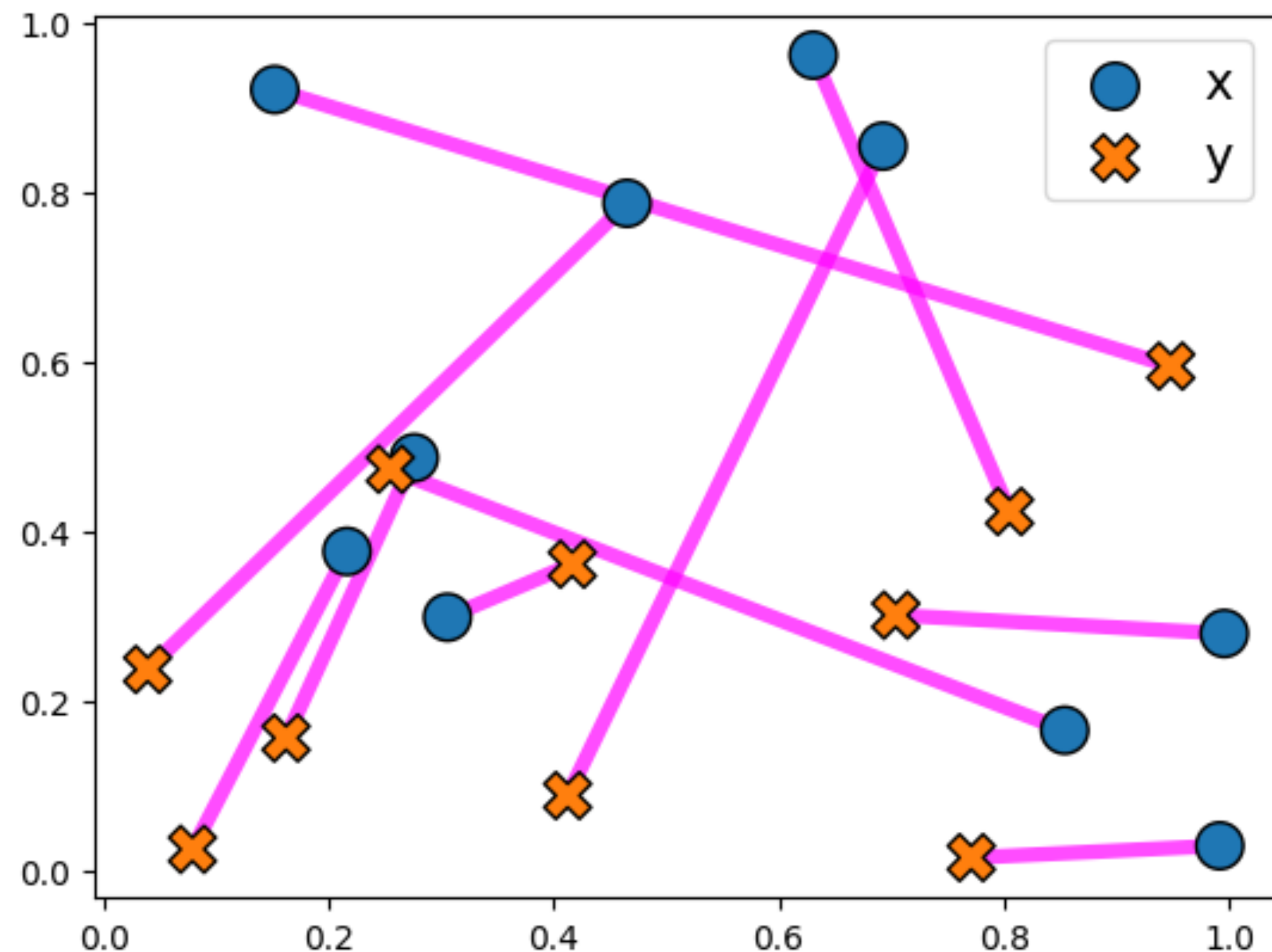
pero...

pero...

las correspondencias óptimas no son convenientes para ML

En efecto, computar correspondencias óptimas genera muchos **problemas**:

- ✱ **Mal escalamiento**, generalmente complejidad $O(n^3)$ para n puntos. [Colab 1]
- ✱ Son **inestables** y difíciles de paralelizar de manera eficiente en GPUs.

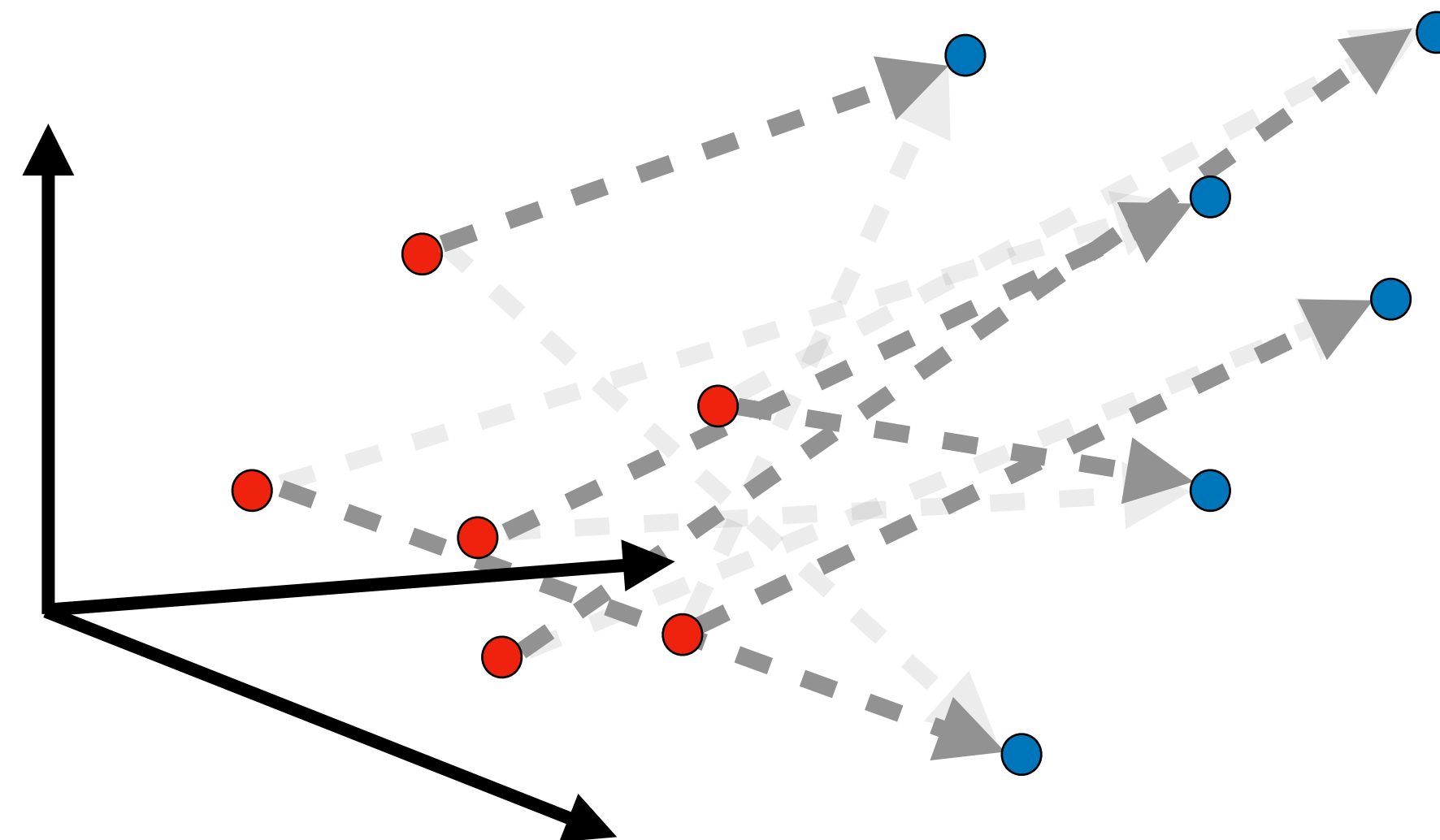


[Colab 2]

Las correspondencias óptimas nos dan más preguntas que respuestas:

... y lo que es más problemático: son frágiles, rígidas y les falta generalización

- ✱ ¿Qué pasa si los puntos x y y difieren en cantidad ($n \neq m$)? ¿y si tienen pesos?
- ✱ ¿Qué pasa si la función de costo no es obvia? (e.g. x y y viven en diferentes espacios)
- ✱ ¿Qué pasa si necesitamos generalizar y empezar a mover nuevos puntos fuera de la muestra?
- ✱ ¿Qué pasa si queremos estocasticidad? En vez de líneas rectas queremos desplazamientos estocásticos



✱ ¿Qué pasa si los puntos x y y difieren en cantidad ($n \neq m$)? ¿y si tienen pesos?

Optimal couplings
Kantorovich relaxation

✱ ¿Qué pasa si la función de costo no es obvia? (e.g. x y y viven en diferentes espacios)

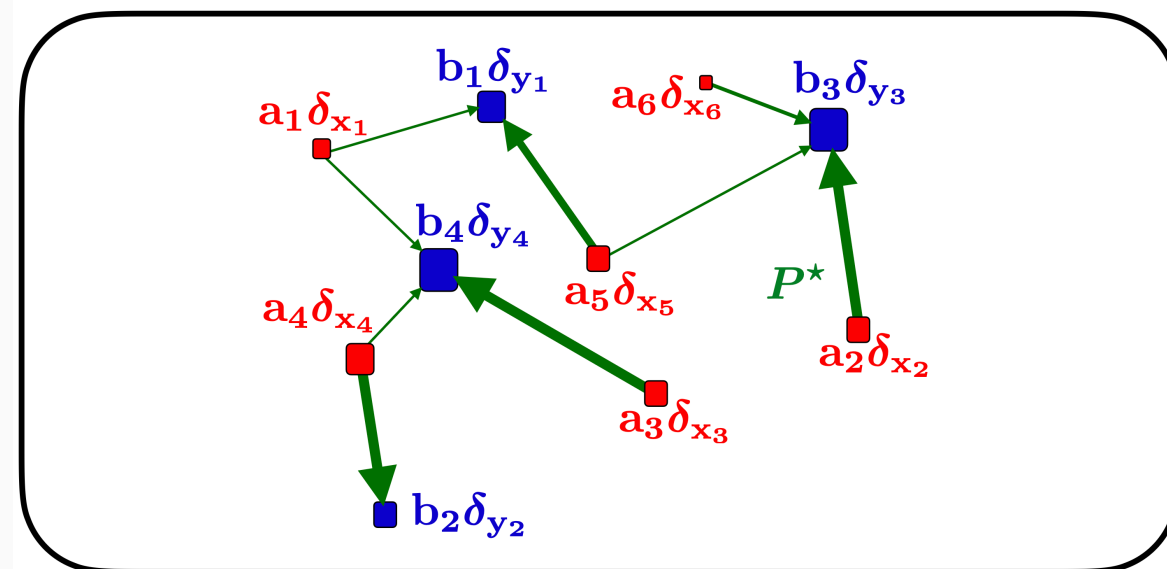
Gromov-Wasserstein
distances

✱ ¿Qué pasa si necesitamos generalizar y empezar a mover nuevos puntos fuera de la muestra?

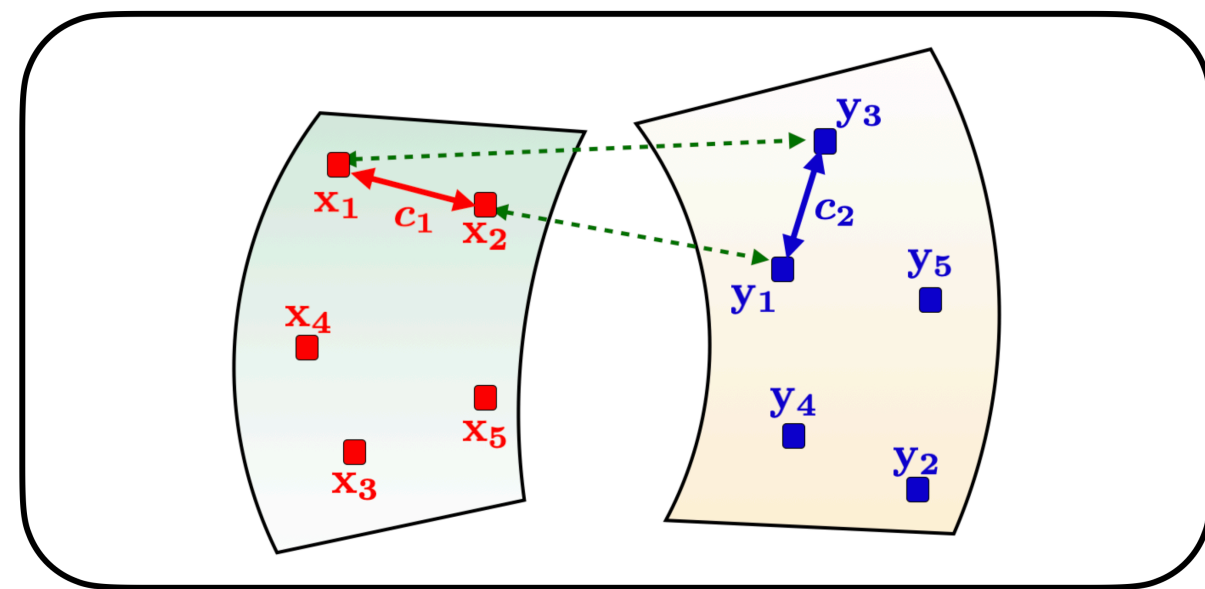
Monge problem
Transport map
Wasserstein distance

✱ ¿Qué pasa si queremos estocasticidad? En vez de líneas rectas queremos desplazamientos esto

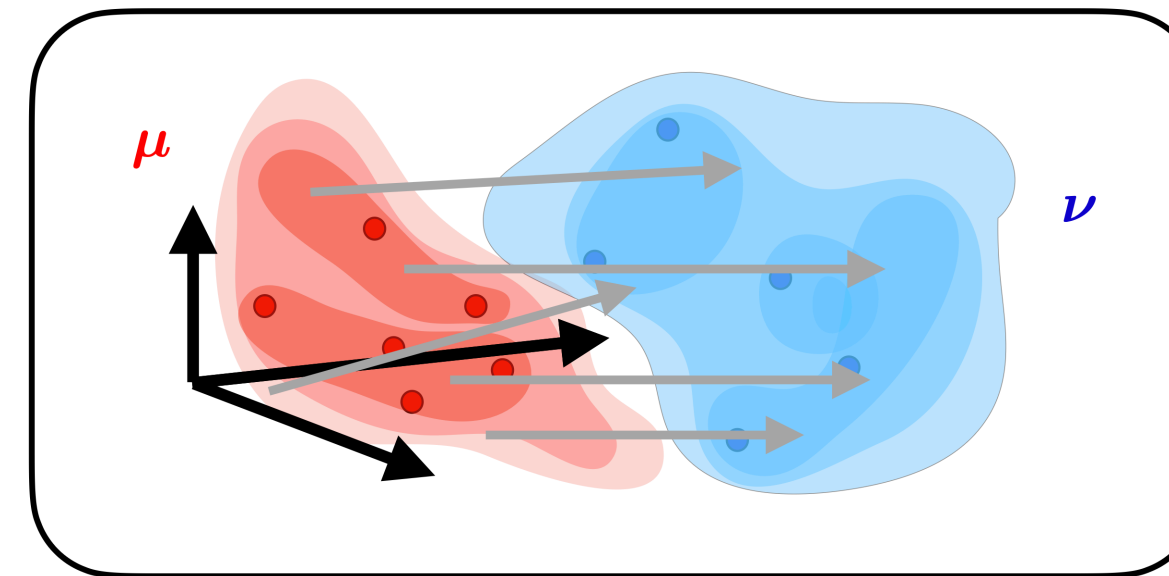
Dynamic OT
Benamou-Brenier



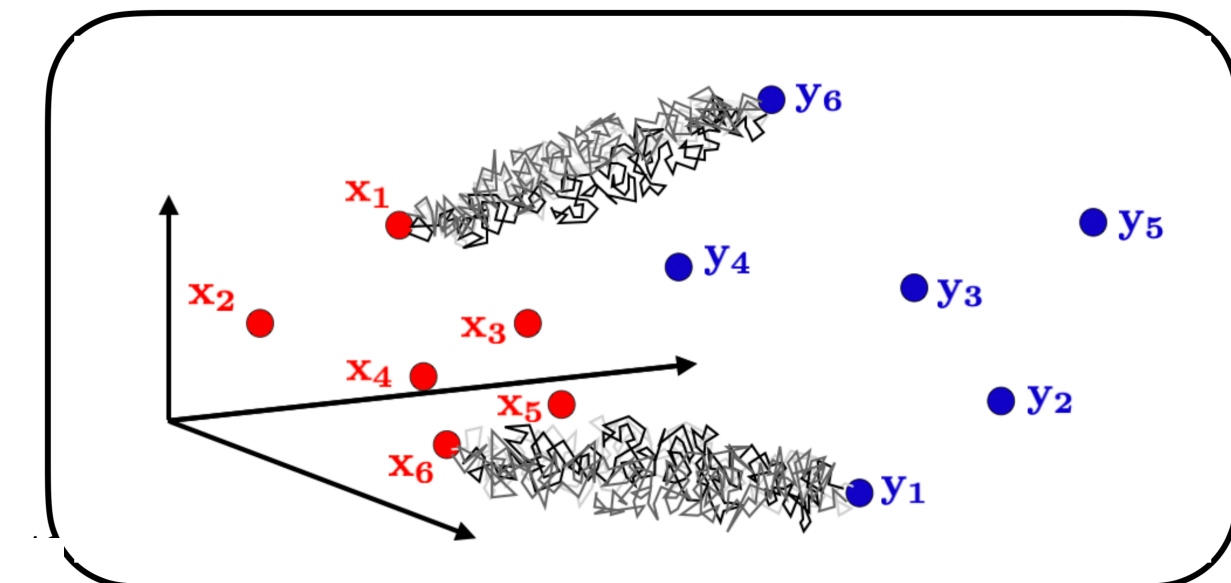
Optimal couplings
Kantorovich relaxation



Gromov-Wasserstein
distances

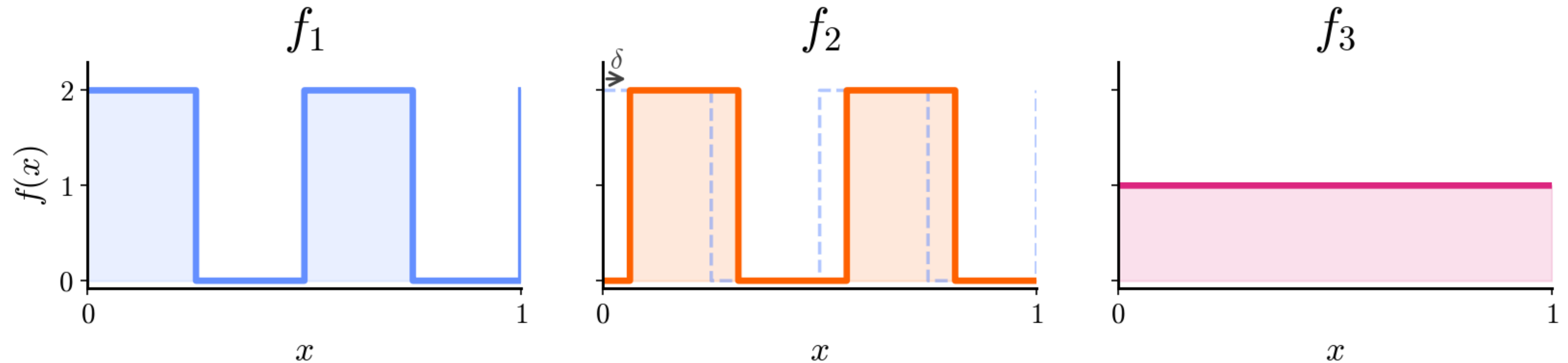


Monge problem
Transport map
Wasserstein distance



Dynamic OT
Benamou-Brenier

Transporte Óptimo



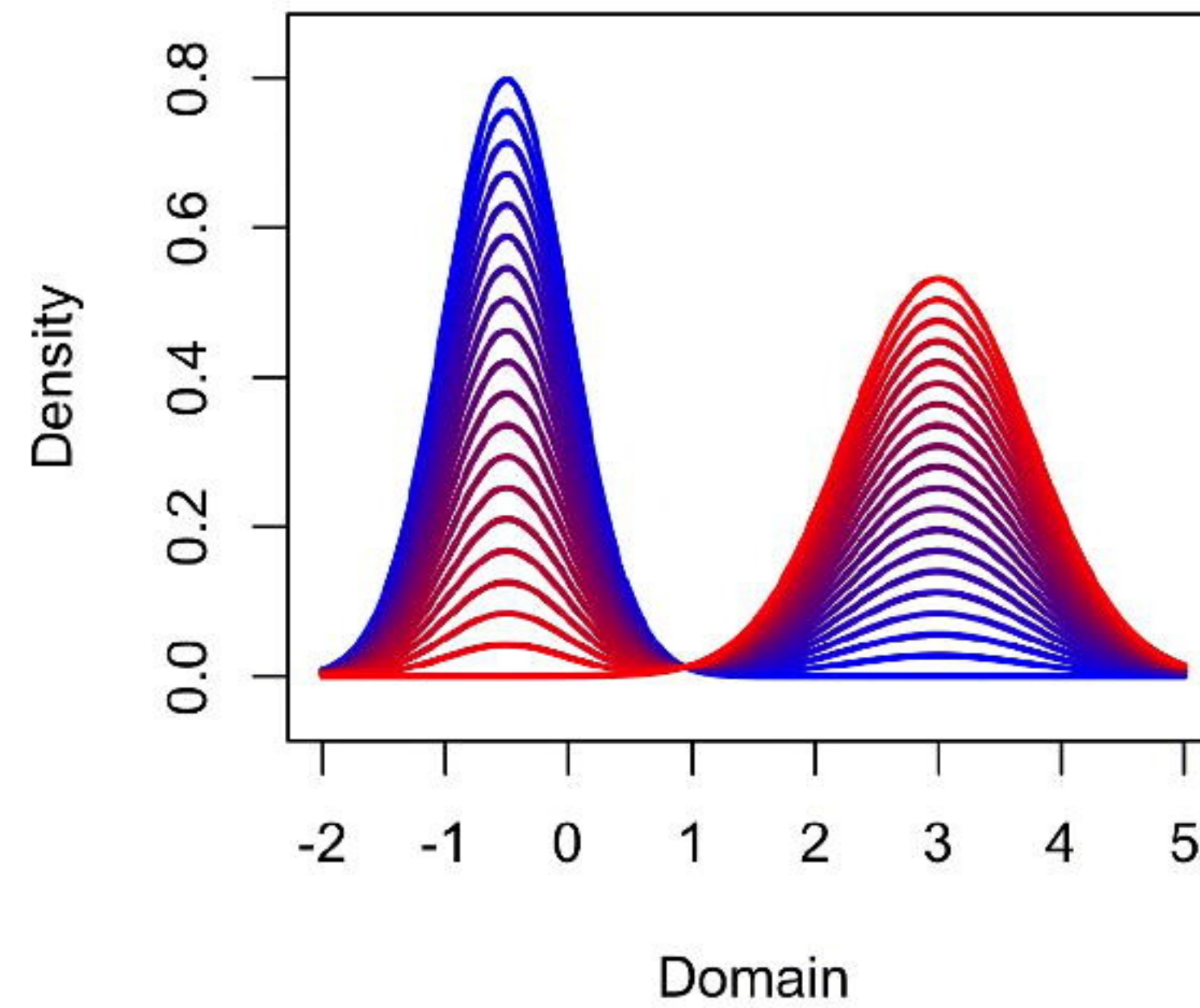
f_2 “debería ser” más cercano a f_1 que a f_3 , pero en distancia L_2 ...

$$\|f_1 - f_2\|_2 = \|f_1 - f_3\|_2 = \|f_2 - f_3\|_2 = 1$$

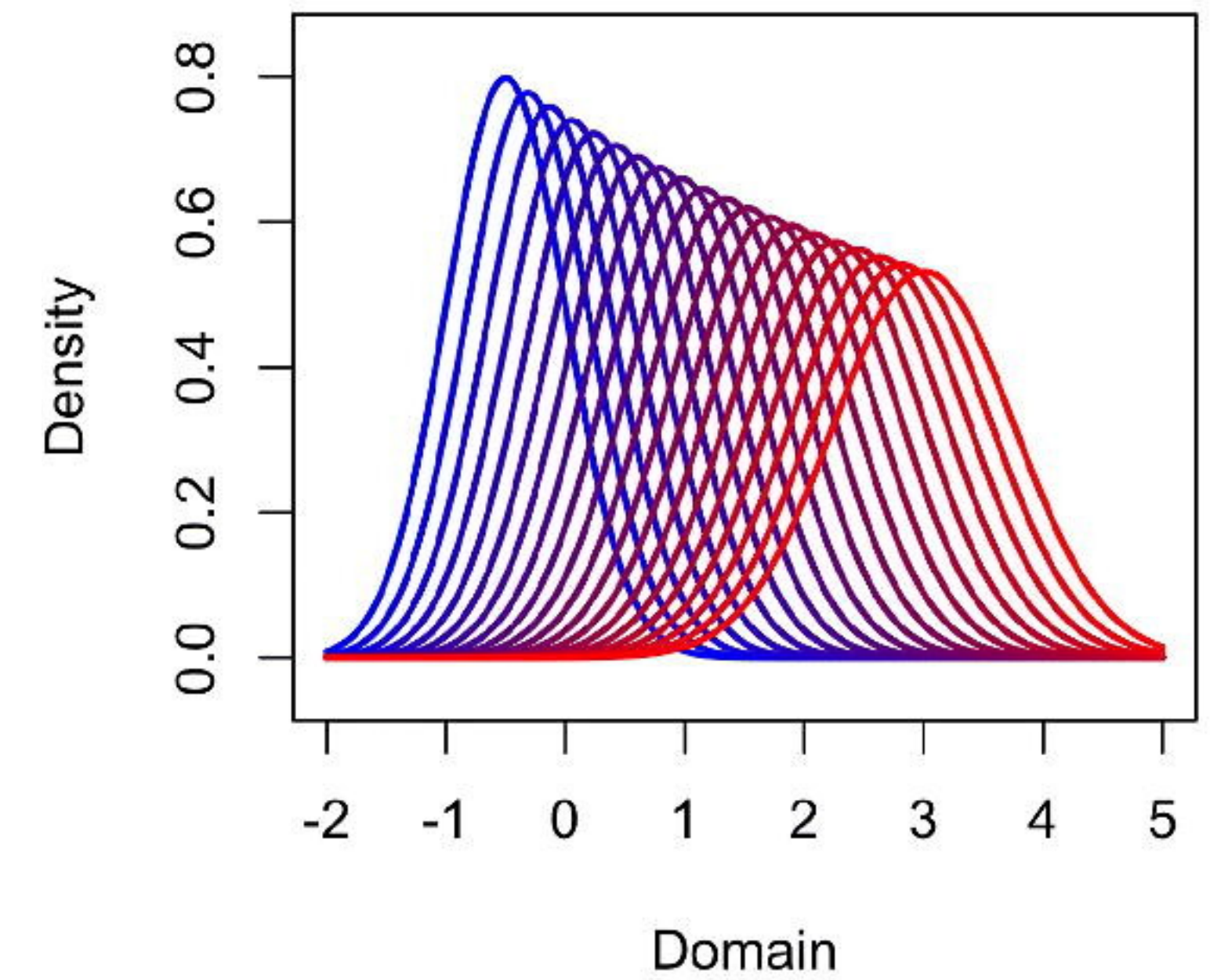
Moraleja: nociones comunes de distancia entre funciones **no dotan** a la variable independiente de una interpretación espacial



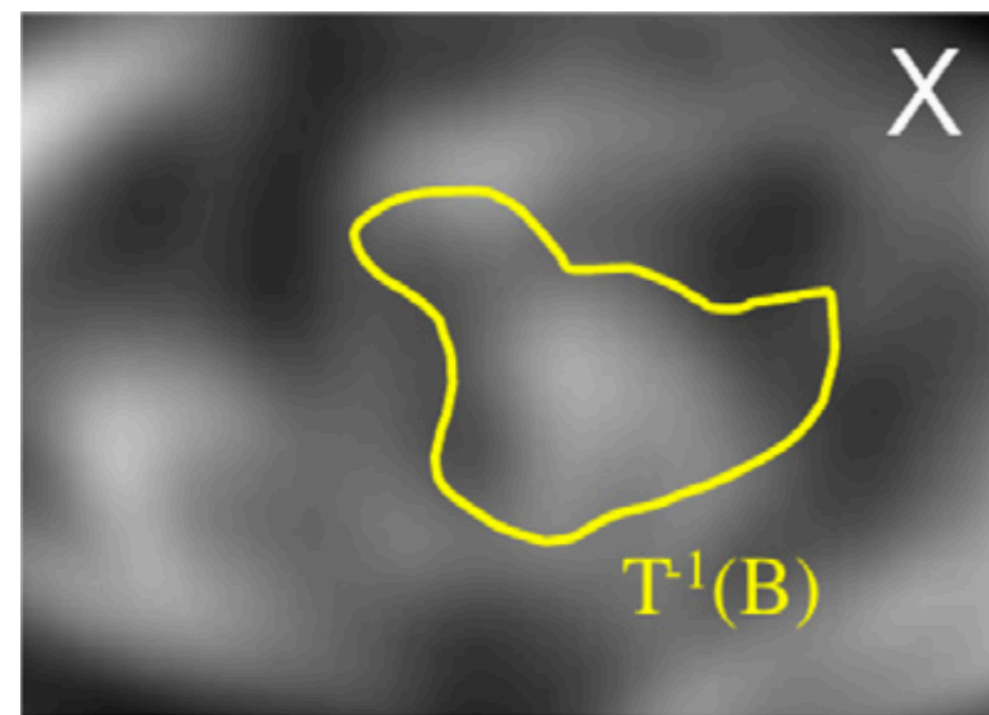
Interpolación
Lineal



Interpolación de
McCann



Problema de Monge

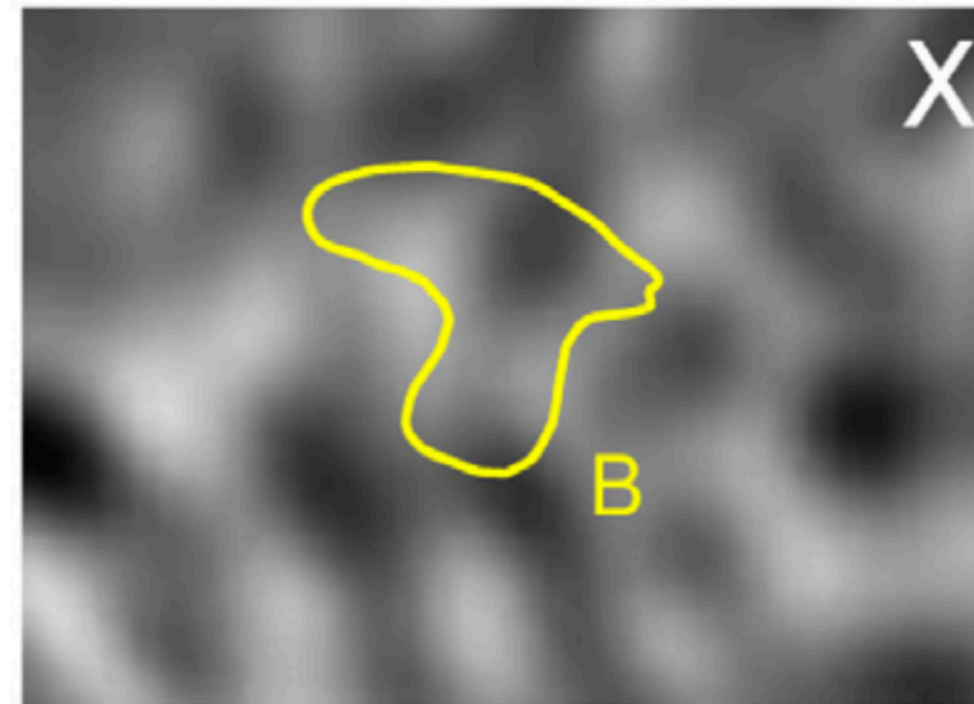


u

$$T : X \rightarrow X$$



minimizando
el trabajo

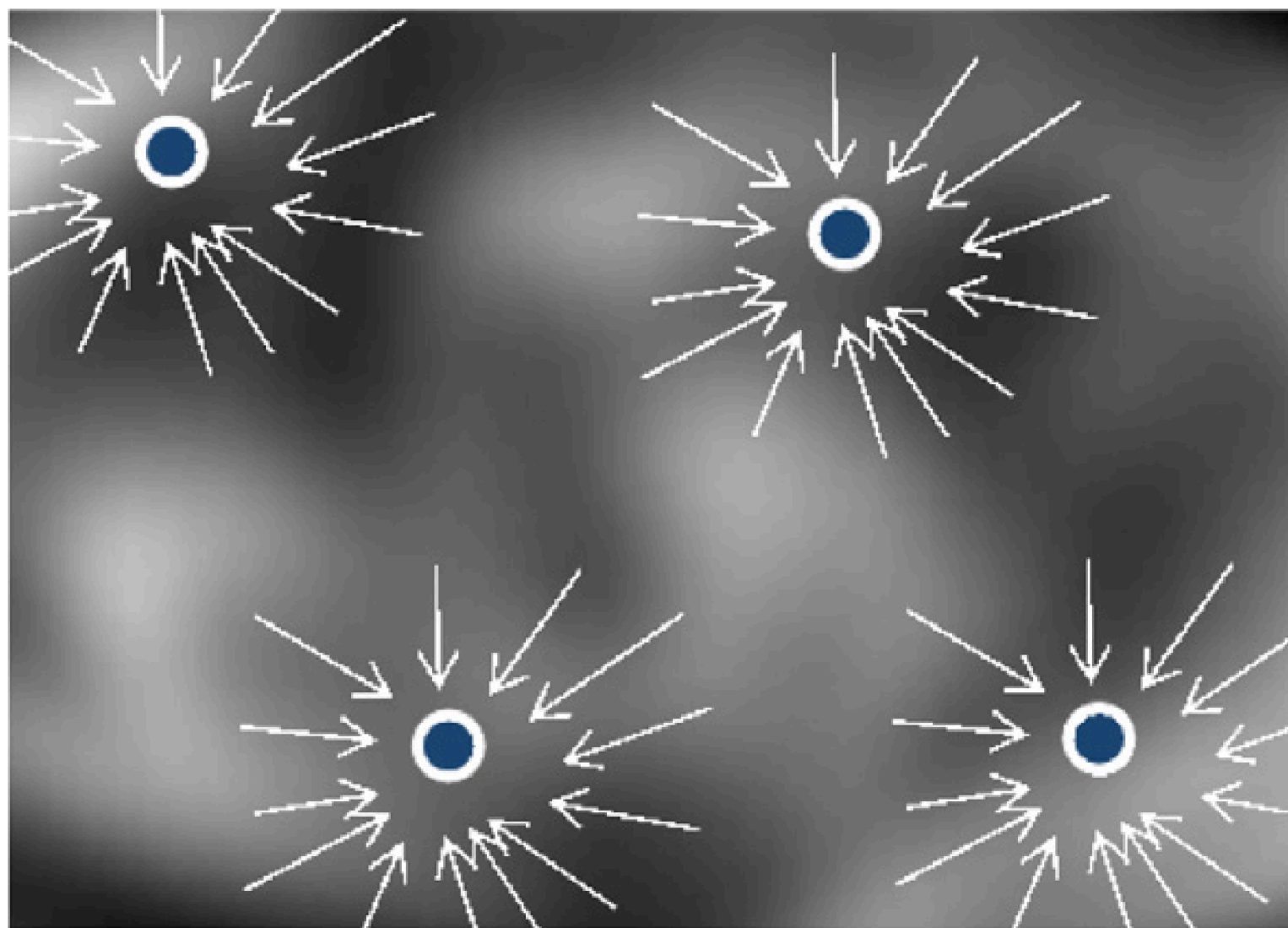


v

$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) u(x) dx$$

sujeto a

$$\forall B \subset X, \int_{T^{-1}(B)} u(x) dx = \int_B v(y) dy$$



¡necesitamos medida!

$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) d\mu \quad \text{sujeto a} \quad T\#\mu = \nu$$

donde

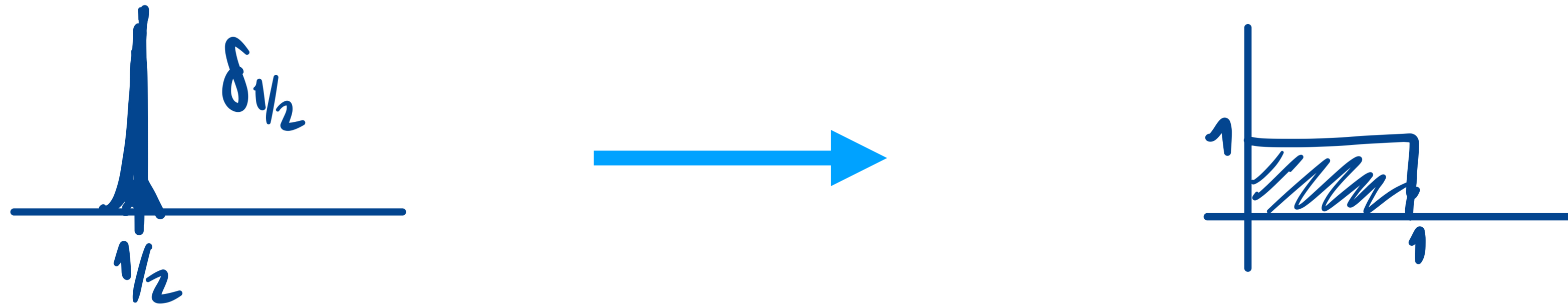
$$T\#\mu(B) = \mu(T^{-1}(B)) \quad \forall B \in \mathcal{B}$$

$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) d\mu \quad \text{sujeto a} \quad T\#\mu = \nu$$

Desafortunadamente, el problema de Monge es un problema de optimización horrible

Dificultad I: el conjunto de restricciones puede ser vacío

Dadas μ, ν , no necesariamente existe T tal que $\nu = T\#\mu$

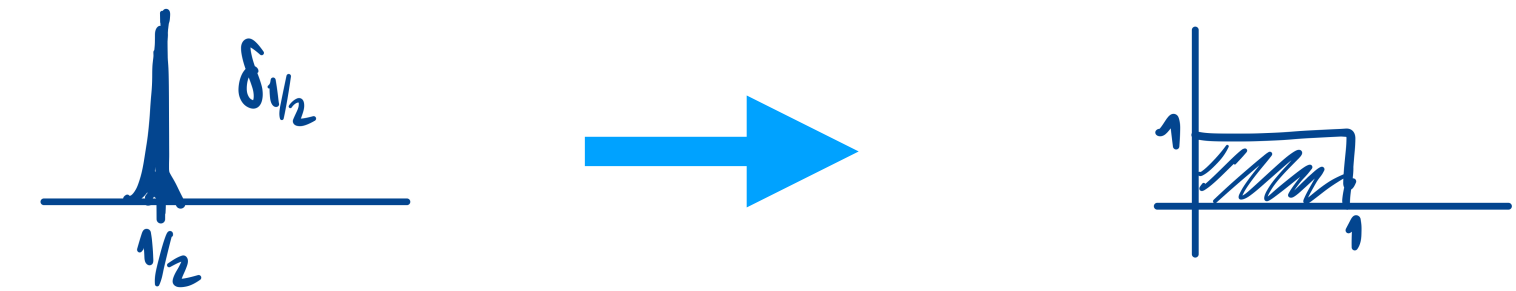


$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) d\mu \quad \text{sujeto a} \quad T\#\mu = \nu$$

Desafortunadamente, el problema de Monge es un problema de optimización horrible

Dificultad I: el conjunto de restricciones puede ser vacío

Dadas μ, ν , no necesariamente existe T tal que $\nu = T\#\mu$



Dificultad II: las soluciones pueden no ser únicas

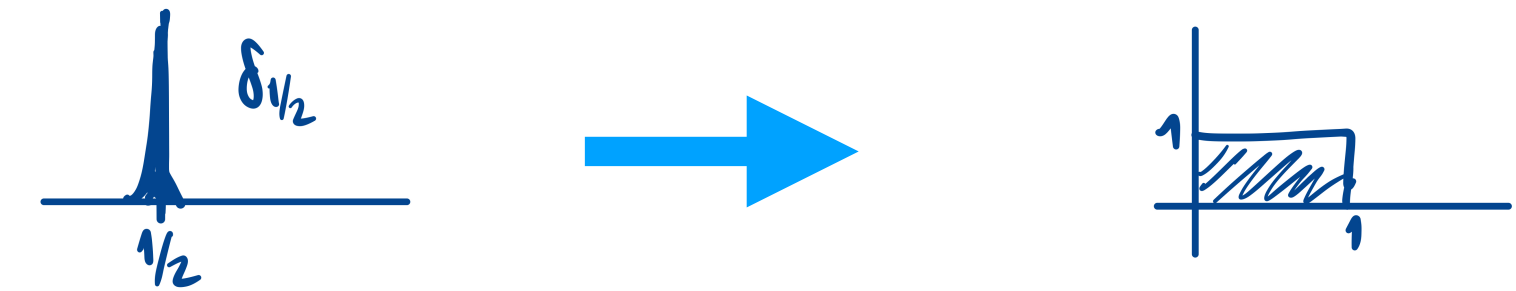


$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) d\mu \quad \text{sujeto a} \quad T\#\mu = \nu$$

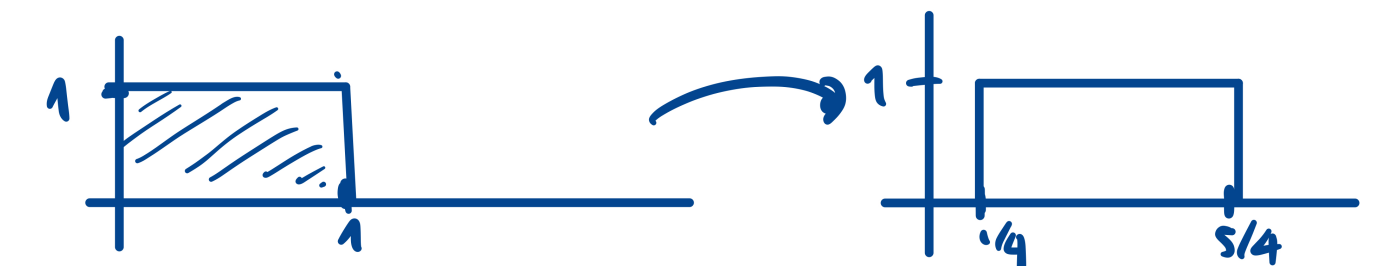
Desafortunadamente, el problema de Monge es un problema de optimización horrible

Dificultad I: el conjunto de restricciones puede ser vacío

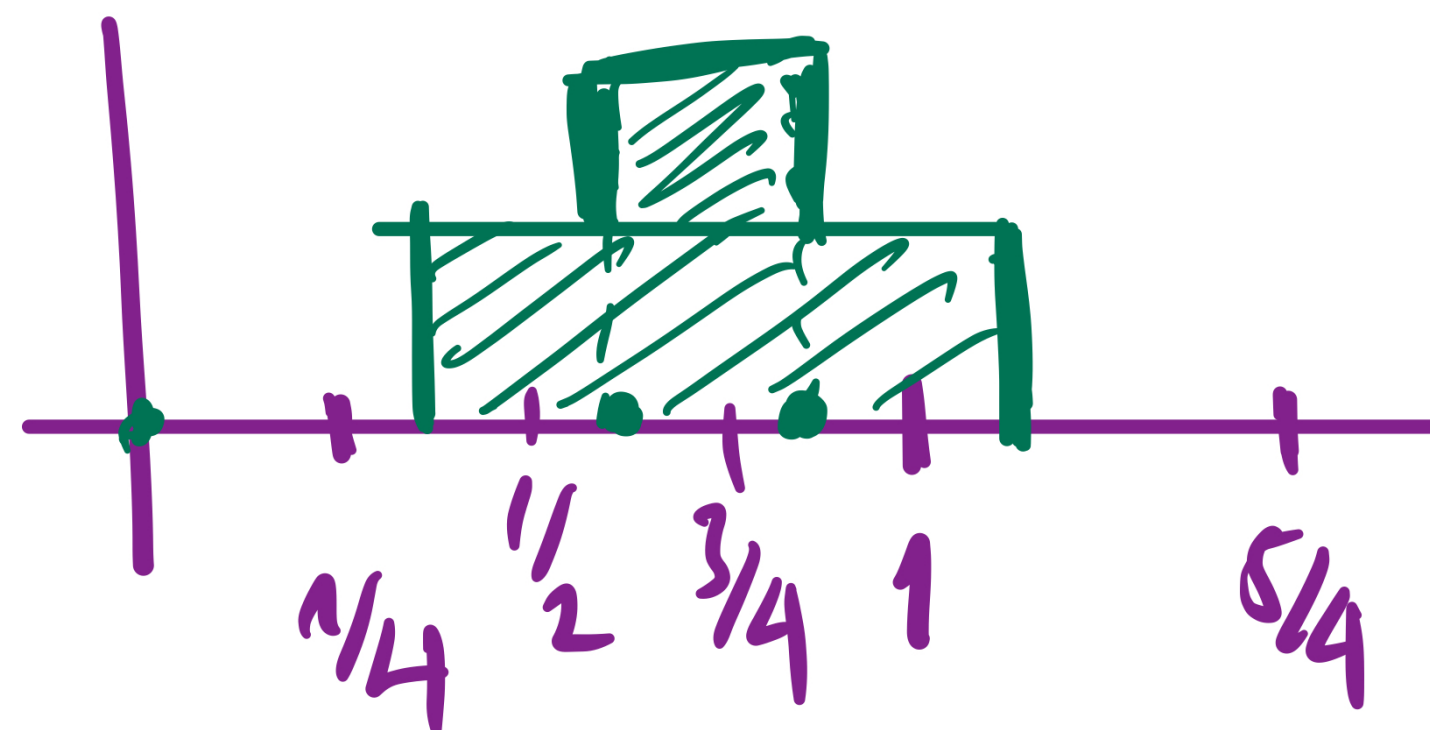
Dadas μ, ν , no necesariamente existe T tal que $\nu = T\#\mu$



Dificultad II: las soluciones pueden no ser únicas



Dificultad III: el conjunto de restricciones no es convexo

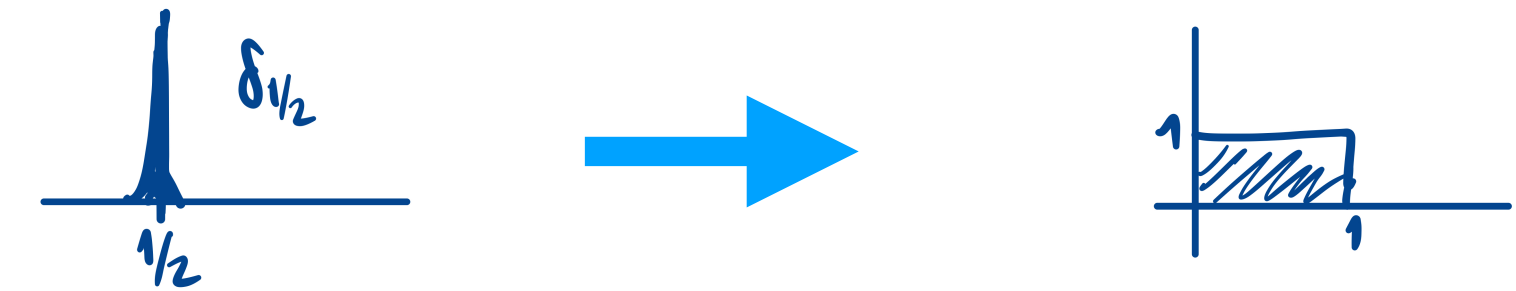


$$\inf_{T: X \rightarrow X} \int_X c(x, T(x)) d\mu \quad \text{sujeto a} \quad T\#\mu = \nu$$

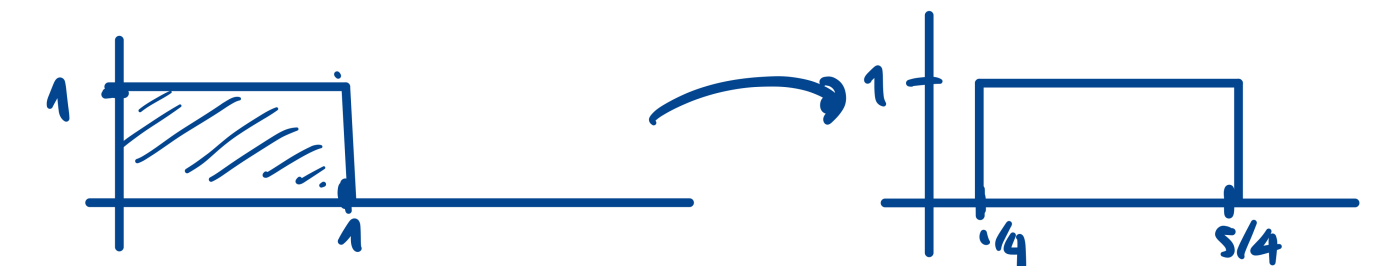
Desafortunadamente, el problema de Monge es un problema de optimización horrible

Dificultad I: el conjunto de restricciones puede ser vacío

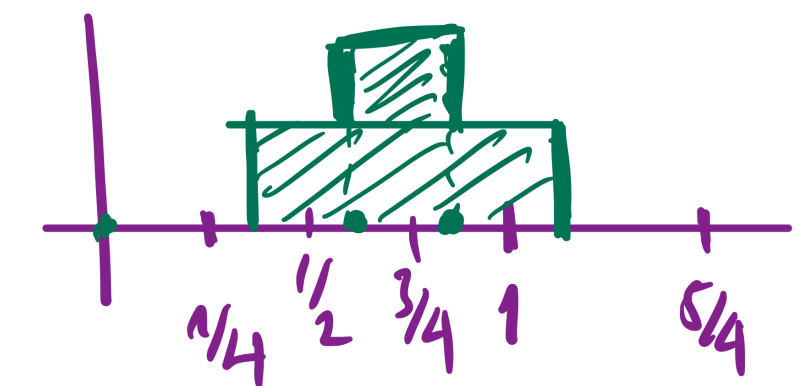
Dadas μ, ν , no necesariamente existe T tal que $\nu = T\#\mu$



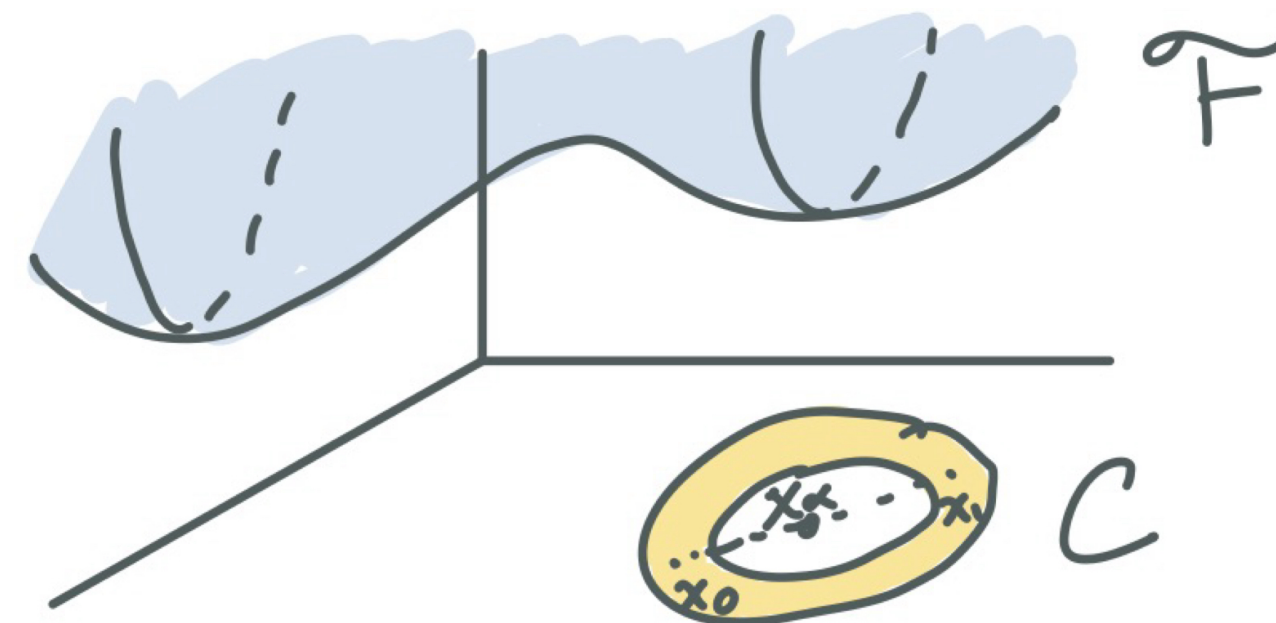
Dificultad II: las soluciones pueden no ser únicas



Dificultad III: el conjunto de restricciones no es convexo

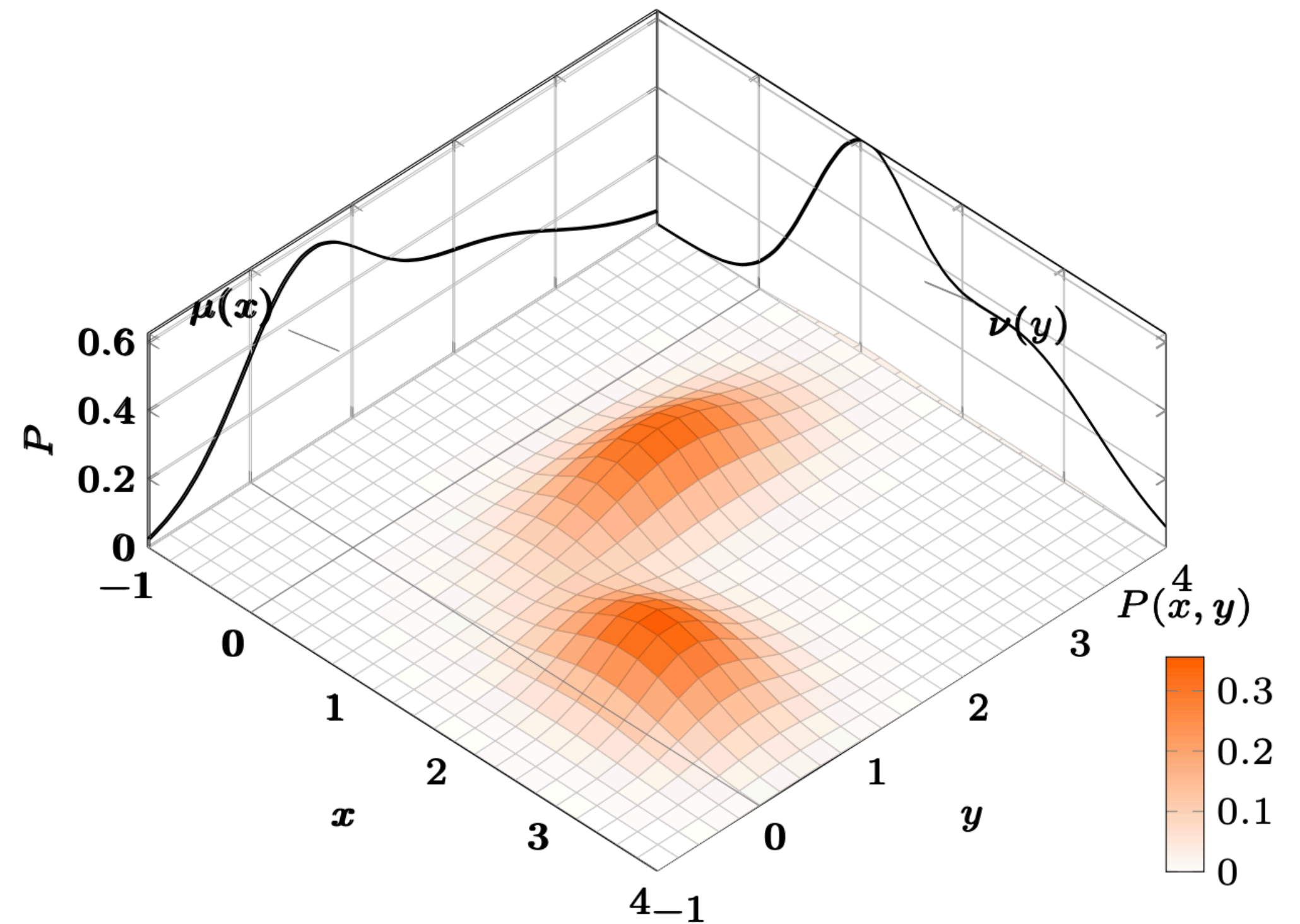
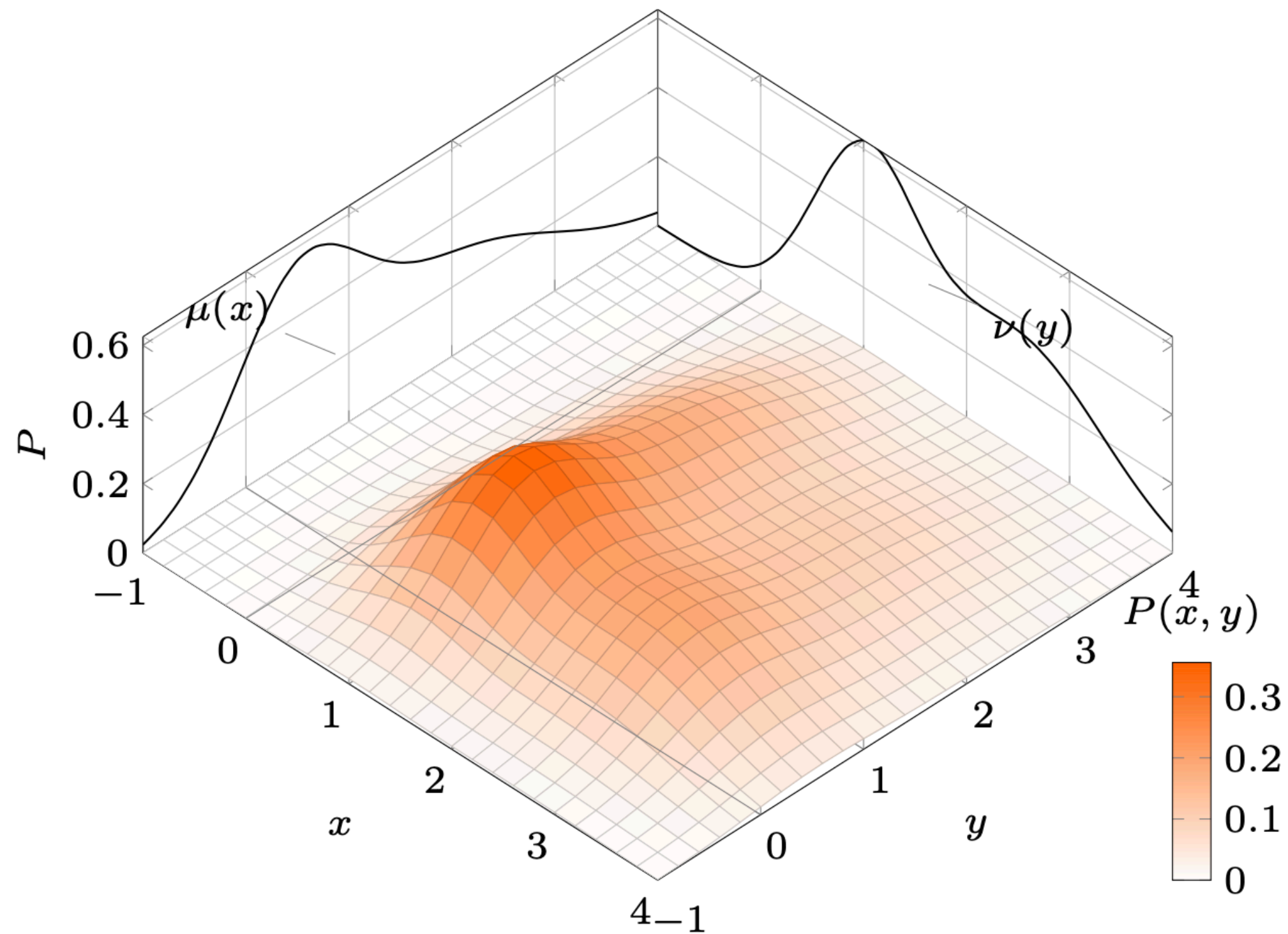
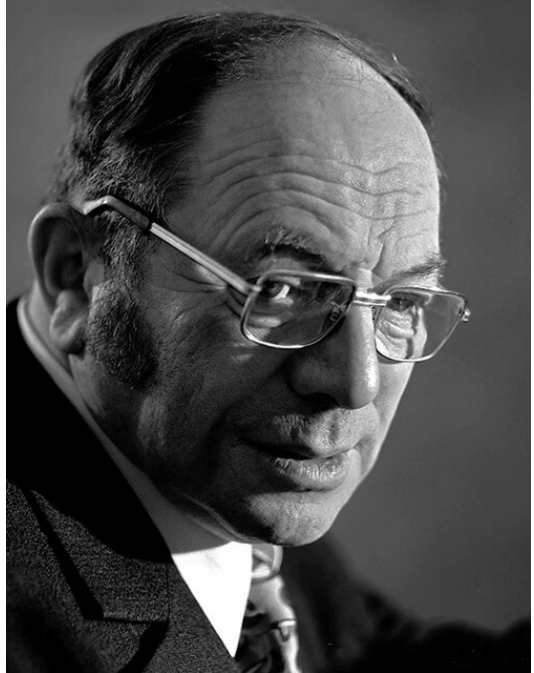


En **optimización**, queremos que nuestro conjunto de restricciones sea **convexo** puesto que la estrategia natural es “adivinar” un valor inicial, **perturbarlo** y ver si la función objetivo **decrece**



**Estamos jodidos
o qué hacemos? (;**

$$\Pi(\mu, \nu) \stackrel{\text{def}}{=} \{P \in \mathcal{P}(\Omega \times \Omega) \mid \forall A, B \subset \Omega, \\ P(A \times \Omega) = \mu(A), P(\Omega \times B) = \nu(B)\}$$



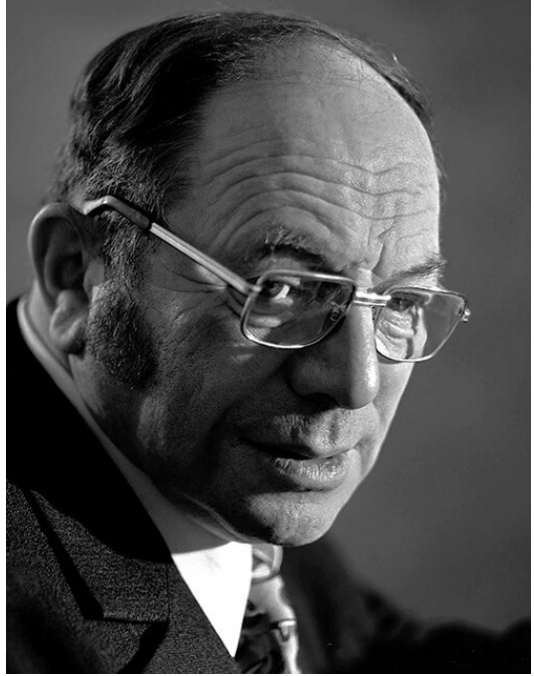
Relajación de Kantorovich

$$\inf_{P \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) P(dx, dy)$$

Relajación de Kantorovich

$$\inf_{P \in \Pi(\mu, \nu)} \mathcal{K}(P) = \inf_{P \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) P(dx, dy)$$

Bueno, y por qué es mejor? (:



~~Dificultad I:~~ el conjunto de restricciones puede ser vacío

Conjunto de restricciones es no vacío

$$\mu \otimes \nu \in \Pi(\mu, \nu)$$

~~Dificultad II:~~ el conjunto de restricciones no es convexo

Conjunto de restricciones es convexo

En efecto, dados $P_0, P_1 \in \Pi(\mu, \nu)$ definamos $P_\alpha := \alpha P_0 + (1-\alpha) P_1$.

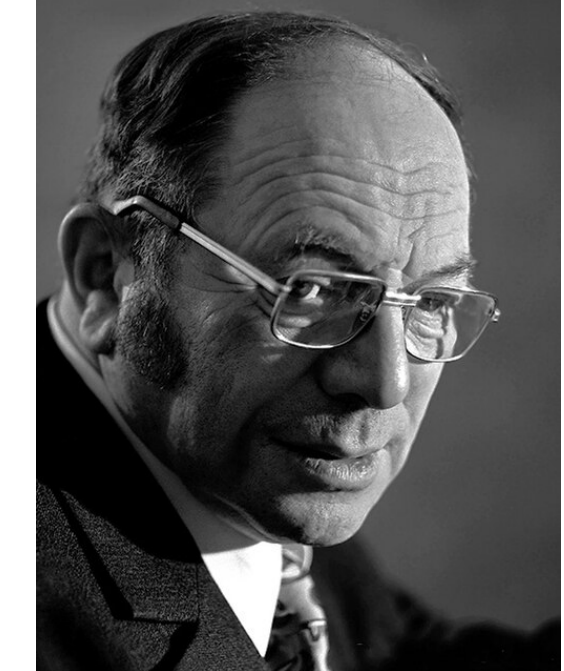
$$\begin{aligned} \text{Entonces, } \pi_x \# P_\alpha(A) &= P_\alpha(\pi_x^{-1}(A)) = \alpha P_0(\pi_x^{-1}(A)) + (1-\alpha) P_1(\pi_x^{-1}(A)) \\ &= \alpha \pi_x \# P_0(A) + (1-\alpha) \pi_x \# P_1(A) = \alpha \mu(A) + (1-\alpha) \mu(A) = \mu(A) \end{aligned}$$

$$\text{Análogamente } \pi_y \# P_\alpha(B) = \nu(B)$$

Relajación de Kantorovich

$$\inf_{P \in \Pi(\mu, \nu)} \mathcal{K}(P) = \inf_{P \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) P(dx, dy)$$

Bueno, y por qué es mejor? (:



~~Dificultad I:~~ el conjunto de restricciones puede ser vacío

Conjunto de restricciones es no vacío

$$\mu \otimes \nu \in \Pi(\mu, \nu)$$

~~Dificultad II:~~ el conjunto de restricciones no es convexo

Conjunto de restricciones es convexo

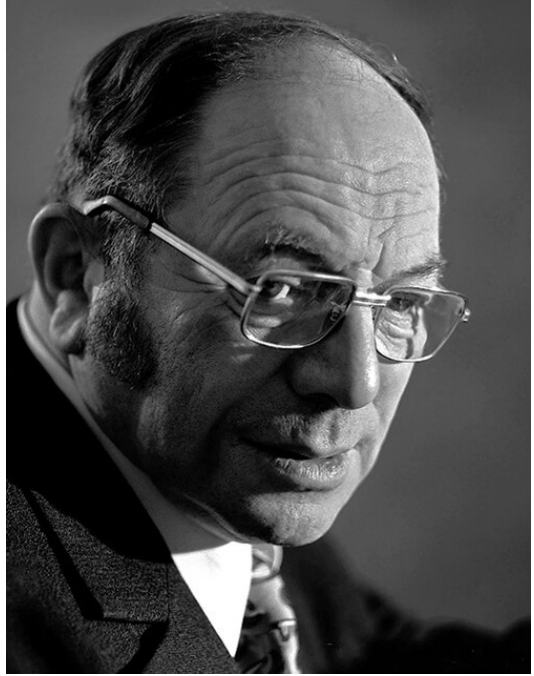
Funcion objetivo es convexa

$$\begin{aligned} \mathcal{K}(\alpha P_0 + (1-\alpha)P_1) &= \int_{X \times X} c(x, y) (\alpha P_0 + (1-\alpha)P_1)(dx, dy) \\ &= \alpha \int_{X \times X} c(x, y) P_0(x, y) + (1-\alpha) \int_{X \times X} c(x, y) P_1(x, y) \\ &= \alpha \mathcal{K}(P_0) + (1-\alpha) \mathcal{K}(P_1) \end{aligned}$$

Relajación de Kantorovich

$$\inf_{P \in \Pi(\mu, \nu)} \mathcal{K}(P) = \inf_{P \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) P(dx, dy)$$

Bueno, y por qué es mejor? (:



~~Dificultad I:~~ el conjunto de restricciones puede ser vacío

Conjunto de restricciones es no vacío

$$\mu \otimes \nu \in \Pi(\mu, \nu)$$

~~Dificultad II:~~ el conjunto de restricciones no es convexo

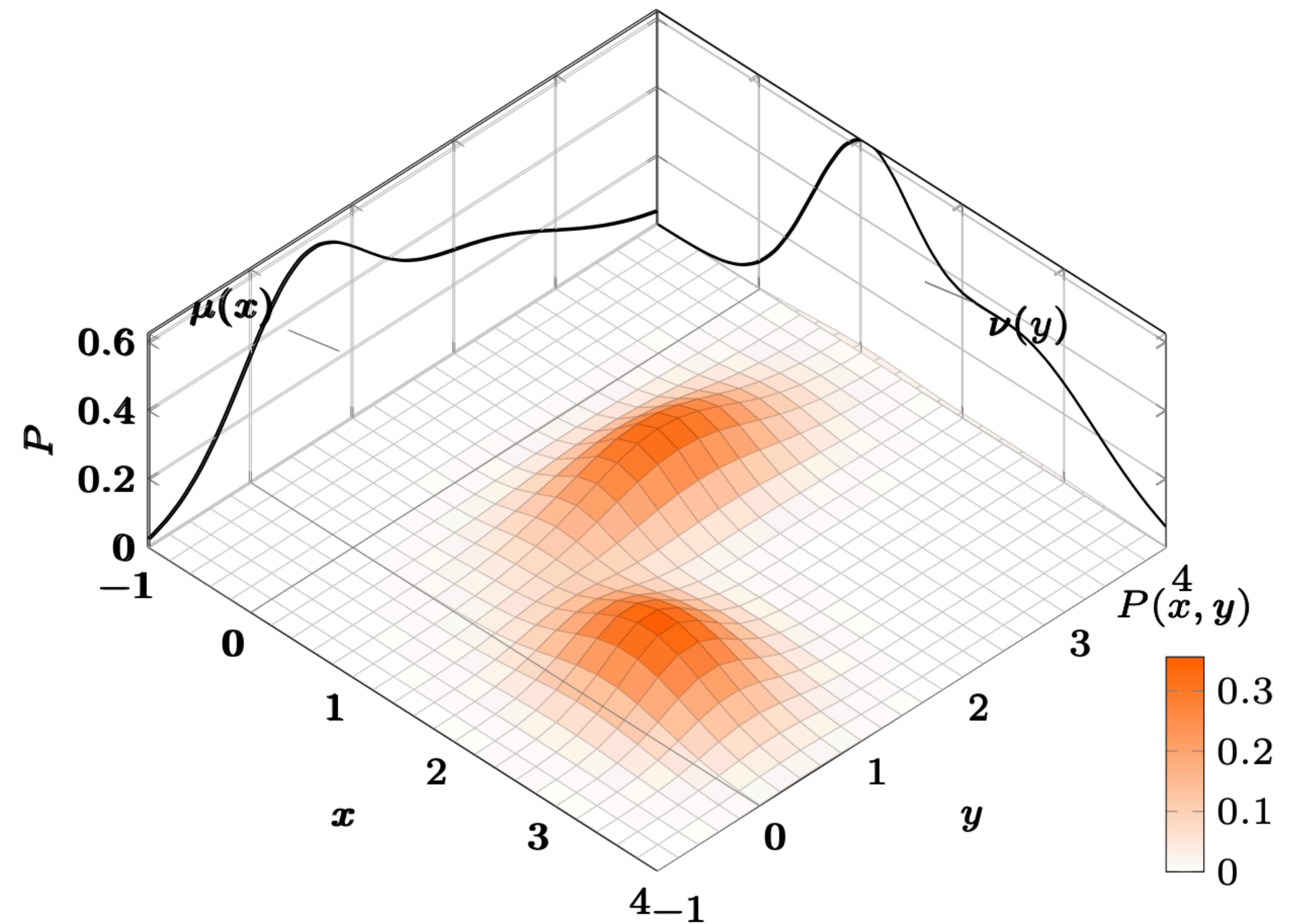
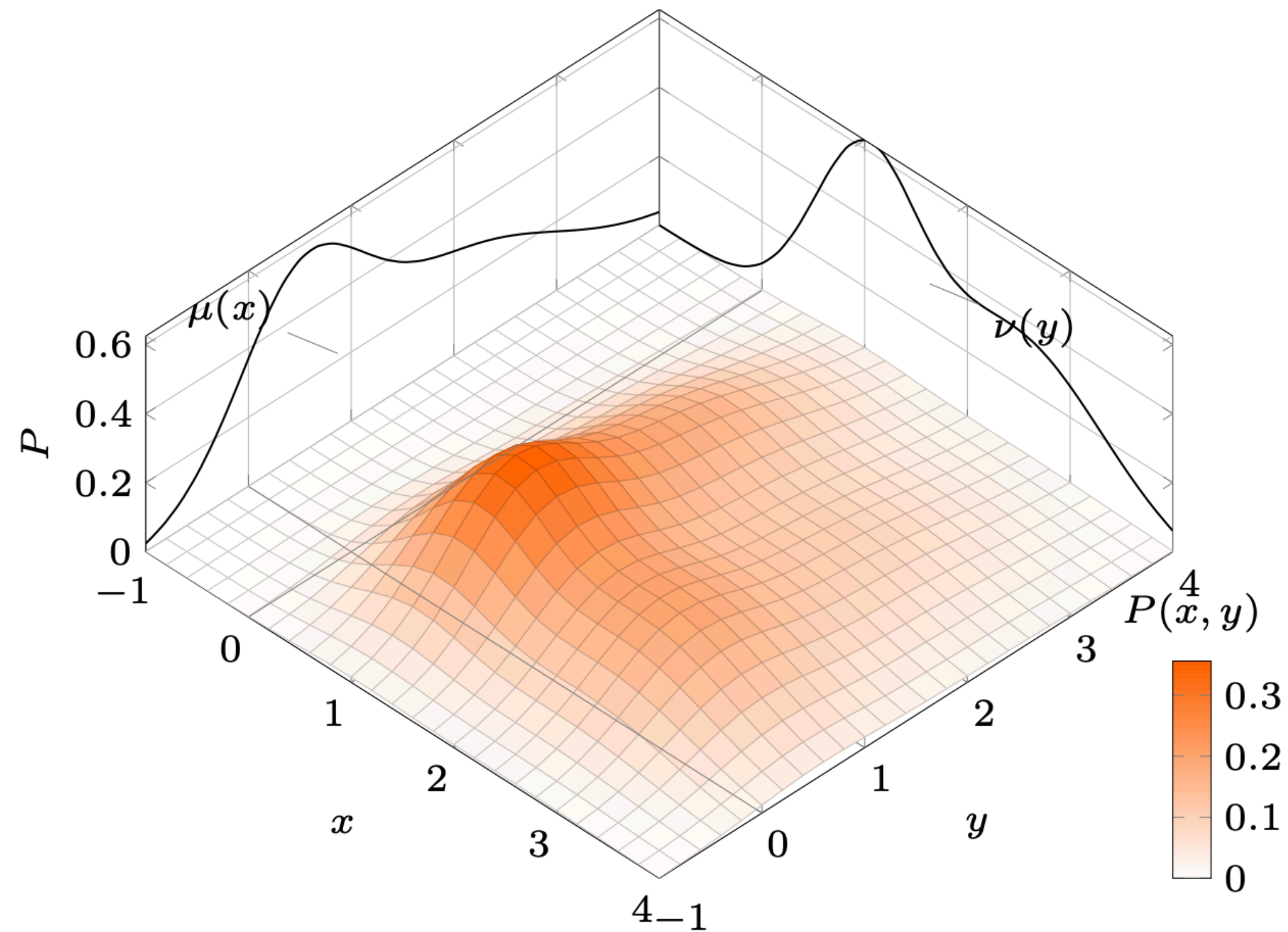
Conjunto de restricciones es convexo

Funcion objetivo es convexa

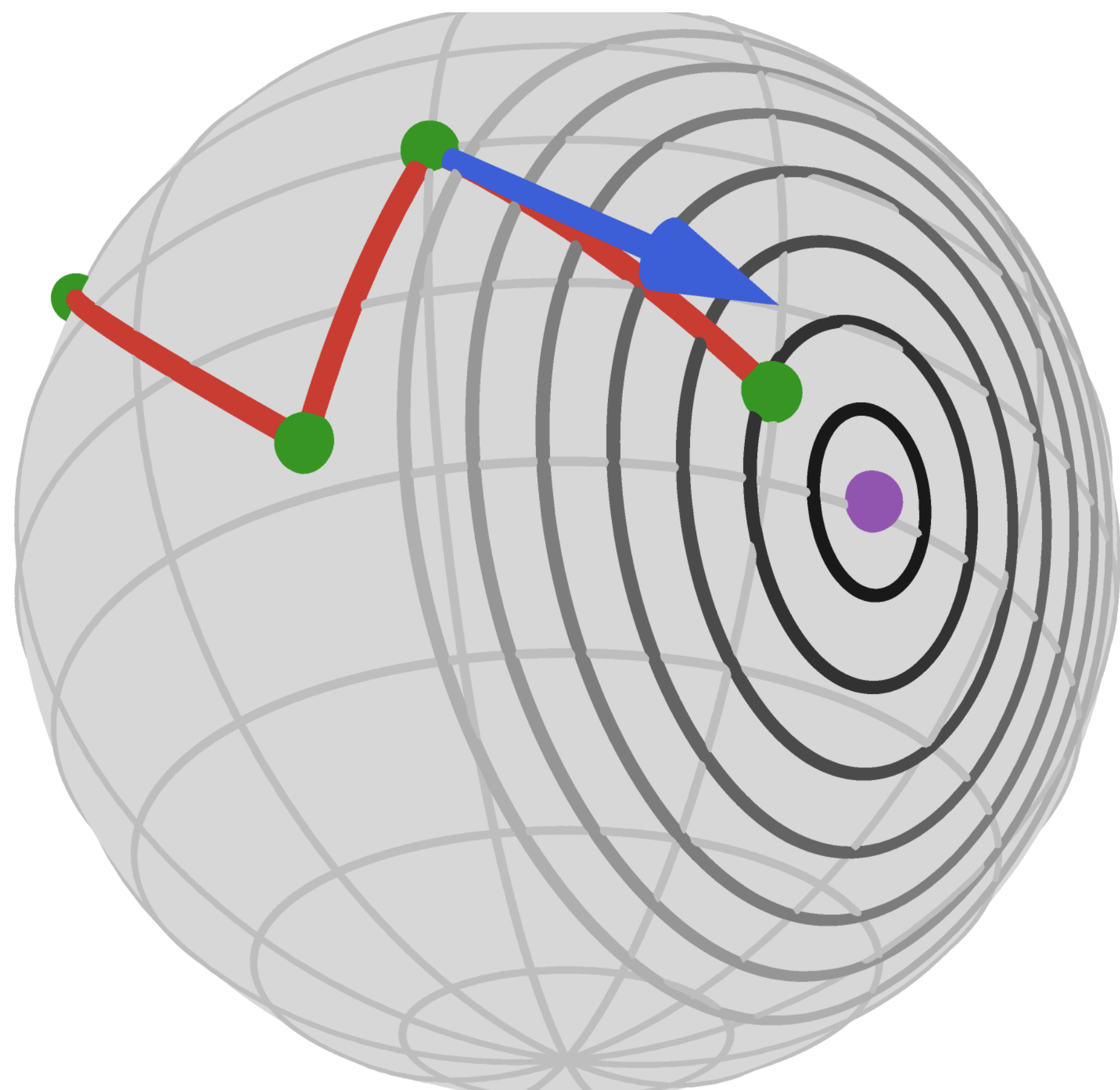
Problema de Kantorovich tiene una versión dual

Uno puede probar que siempre existe una solución vía cálculo de variaciones

$$\Pi(\mu, \nu) \stackrel{\text{def}}{=} \{P \in \mathcal{P}(\Omega \times \Omega) \mid \forall A, B \subset \Omega, \\ P(A \times \Omega) = \mu(A), P(\Omega \times B) = \nu(B)\}$$



$$W_2(\mu, \nu) \stackrel{\text{def}}{=} \left(\inf_{P \in \Pi(\mu, \nu)} \iint d(x, y)^2 P(dx, dy) \right)^{1/2}$$



Gracias!

