

## TEORÍA DE TRANSPORTE ÓPTIMO

### SUGERENCIAS PARA LOS EJERCICIOS

La numeración es la de las notas y de las guías (Ejercicio  $n.k$  = ejercicio  $k$  del Capítulo  $n$ ); las referencias remiten a las notas (<https://github.com/jfbonder/transporte-optimo>). Las sugerencias indican una idea o el resultado que conviene usar, no una solución completa. Sólo figuran los ejercicios que tienen sugerencia.

### CAPÍTULO 1. INTRODUCCIÓN

**Ejercicio 1.1** (Push-forward). (a)  $T^{-1}$  conmuta con uniones numerables y complementos, y toda  $\sigma$ -álgebra que haga medible a  $T$  está contenida en  $\Sigma'$  por definición de medibilidad. (b) Es  $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$ . (c)  $T_{\#}\mu = \sum_i m_i \delta_{T(x_i)}$ ; si  $T$  no es inyectiva, agrupar los  $x_i$  con la misma imagen. (d) Basta que  $\mu$  y  $\nu$  sean difusas (sin átomos), por ejemplo  $\mu = \nu = \mathcal{L}^1|_{[0,1]}$  y  $T(x) = x/2$ : ambos lados de la igualdad son 0.

**Ejercicio 1.2** (Caracterización integral del push-forward). Seguir la sugerencia del enunciado. Para el recíproco tomar  $\phi = \mathbf{1}_B$ : la igualdad dice exactamente  $\mu(T^{-1}(B)) = \nu(B)$ . Para el paso a  $L^1(\nu)$  separar  $\phi = \phi^+ - \phi^-$  y observar que  $\phi \circ T \in L^1(\mu)$  porque  $\int |\phi \circ T| d\mu = \int |\phi| d\nu$  por el paso anterior.

**Ejercicio 1.3** (Ejemplos de push-forward). Para medidas discretas,  $T_{\#}\delta_x = \delta_{T(x)}$ . Para densidades, usar  $T_{\#}\mu(B) = \mu(T^{-1}(B))$  con  $B = (-\infty, t]$  y derivar la función de distribución: en (a) se obtiene  $\frac{1}{2}\mathcal{L}^1|_{[1,3]}$ ; en (b),  $T^{-1}((-\infty, t]) = [-\sqrt{t}, \sqrt{t}]$  da la densidad  $\frac{1}{2\sqrt{t}}\mathbf{1}_{(0,1]}$  en ambos casos (¡las dos medidas tienen la misma imagen!); en (c) sumar las contribuciones de cada intervalo  $[k, k+1)$ : la primera imagen es  $\mathcal{L}^1|_{[0,1]}$  y la segunda tiene densidad  $\frac{2}{3}(\mathbf{1}_{[0,1)} + \mathbf{1}_{[0,1/2)})$ .

**Ejercicio 1.4** (Cambio de variables: contraejemplos). (a) En  $x^2$  falla la inyectividad (dos pre-ímagenes); en la carpa falla  $C^1$  en  $x = 1/2$  y también la inyectividad. Comparar la densidad verdadera del Ejercicio 1.3 con la que daría la fórmula. (b) Cubrir  $X$  con abiertos donde  $T$  sea inyectiva y aplicar la fórmula en cada uno; la imagen es la suma. (c) Si  $F(t) = \mathcal{L}^1(\{T \leq t\}) = \mathcal{L}^1([0, T^{-1}(t)]) = T^{-1}(t)$  y debe ser  $F(t) = t$ , entonces  $T^{-1} = \text{id}$ .

**Ejercicio 1.5** (Soporte del push-forward). (a) El complemento de  $\overline{T(\text{sop } \mu)}$  es un abierto  $U$  con  $T^{-1}(U) \cap \text{sop } \mu = \emptyset$ , luego  $T_{\#}\mu(U) = 0$ . Para la igualdad con soporte compacto: si  $y = T(x)$  con  $x \in \text{sop } \mu$ , todo entorno  $V$  de  $y$  cumple  $T^{-1}(V) \ni x$  abierto, luego de medida positiva. (b)  $\mu = \sum 2^{-k}\delta_k$  en  $\mathbb{R}$  y  $T(x) = 1/(1+x)$ :  $T(\text{sop } \mu)$  no contiene a 0, que sí está en el soporte de la imagen. (c)  $[0, 1]$ ,  $[0, \infty)$  y  $[0, 1]$ : el soporte es cerrado y los racionales tienen medida nula.

**Ejercicio 1.6** (Planes inducidos por mapas). (a)  $\pi(A \times Y) = \mu(\{x : x \in A, T(x) \in Y\}) = \mu(A)$  y  $\pi(X \times B) = \mu(T^{-1}(B)) = \nu(B)$ . (b) La medida producto  $\mu \otimes \nu$  siempre está en  $\Pi(\mu, \nu)$ . (c) Si  $\pi$  está concentrado en el gráfico de  $T$ , para  $\phi \in C_b(X \times Y)$  vale  $\int \phi d\pi = \int \phi(x, T(x)) d\pi = \int \phi(x, T(x)) d\mu$  porque la primera marginal de  $\pi$  es  $\mu$ ; el recíproco es la definición de  $(\text{id}, T)_{\#}\mu$ .

**Ejercicio 1.7** (Un plan que no proviene de un mapa). Un plan inducido por  $T$  carga sólo dos puntos del cuadrado, uno en cada “columna”  $\{0\} \times Y$ ,  $\{1\} \times Y$ ;  $\pi$  carga cuatro. En general  $\Pi(\mu, \nu)$  es el segmento  $\pi_t = \frac{t}{2}(\delta_{(0,0)} + \delta_{(1,1)}) + \frac{1-t}{2}(\delta_{(0,1)} + \delta_{(1,0)})$ ,  $t \in [0, 1]$ , y sólo los extremos  $t = 0, 1$  provienen de un mapa.

**Ejercicio 1.8** (El conjunto de mapas de transporte no es convexo). (a) Para  $T_1$ , calcular  $(T_1)_{\#}\mu$  por intervalos:  $[0, 1/4]$  va a  $[1, 5/4]$  y  $[1/4, 1]$  queda fijo. (b) La combinación  $T_\alpha$  es  $x+\alpha$  en  $[0, 1/4]$  y  $x + (1-\alpha)/4$  en  $[1/4, 1]$ ; su imagen es la unión de dos intervalos que se superponen o dejan

un hueco, por lo que la densidad de  $(T_\alpha)_\# \mu$  no es  $\mathbf{1}_{[1/4, 5/4]}$ . (c) Las restricciones de marginales son lineales en  $\pi$  y  $\pi \mapsto \int c d\pi$  es lineal.

**Ejercicio 1.9** (Caracterización de los planes por integrales). Si  $\pi \in \Pi(\mu, \nu)$ , aplicar el Ejercicio 1.2 a las proyecciones  $p_X, p_Y$  (que cumplen  $(p_X)_\# \pi = \mu, (p_Y)_\# \pi = \nu$ ) para ver que  $\phi \circ p_X \in L^1(\pi)$  y  $\int \phi(x) d\pi = \int \phi d\mu$ . Para el recíproco, tomar  $\phi = \mathbf{1}_A, \psi = 0$  y luego  $\phi = 0, \psi = \mathbf{1}_B$ .

**Ejercicio 1.10** (Mínimos de funcionales lineales y puntos extremales). (a)  $\ell$  es continua y  $K$  compacto. (b) El conjunto  $F$  de minimizadores es  $K \cap \ell^{-1}(\min_K \ell)$ , cerrado y convexo; si  $x = \frac{1}{2}(x_1 + x_2) \in F$  con  $x_i \in K$ , entonces  $\ell(x_1) + \ell(x_2) = 2 \min_K \ell$  obliga  $\ell(x_1) = \ell(x_2) = \min_K \ell$ . Un compacto convexo no vacío tiene puntos extremales (Krein–Milman aplicado a  $F$ ), y un extremal de una cara extremal  $F$  es extremal en  $K$ .

**Ejercicio 1.11** (Matrices biestocásticas). (a) Las dos condiciones de fila y columna dejan un solo grado de libertad; los extremales son  $t = 0, 1$ . (b)  $\mathcal{B}_n$  es cerrado y acotado, definido por igualdades y desigualdades lineales. Una matriz de permutación tiene entradas en  $\{0, 1\}$ : si es punto medio de  $\pi^1, \pi^2 \in \mathcal{B}_n$ , cada entrada 0 o 1 obliga a la misma entrada en ambas. (c) En la fila  $i_0$  hay otra entrada en  $(0, 1)$  (la fila suma 1), en su columna otra, etc.; como hay finitas entradas, el camino se cierra en un ciclo par. Sumar y restar  $\varepsilon$  alternadamente a lo largo del ciclo conserva las sumas de filas y columnas. (d) Por (c) los extremales tienen entradas en  $\{0, 1\}$ , es decir, son permutaciones; la segunda afirmación es Krein–Milman en dimensión finita (Carathéodory).

## CAPÍTULO 2. EXISTENCIA DE PLANES ÓPTIMOS

**Ejercicio 2.1** (Convergencia débil: ejemplos). Probar contra  $\phi \in C_b(\mathbb{R})$  y usar continuidad de  $\phi$  en el punto límite. (a)  $\delta_0$ ; (c)  $\delta_0$ ; (d)  $\mathcal{L}^1_{[0,1]}$  (comparar las integrales, la diferencia es  $O(1/n)\|\phi\|_\infty$ ); (e)  $\delta_0$  (teorema del valor medio); (f)  $\mathcal{L}^1_{[0,1]}$  (sumas de Riemann). En (b): si  $\delta_{n_k} \rightharpoonup \mu$ , tomar  $\phi$  con  $\phi(x) = 1$  para  $x \leq R$  y soporte compacto:  $\int \phi d\mu = 0$  para todo  $R$ , luego  $\mu = 0$ , que no es de probabilidad; la masa se escapa al infinito.

**Ejercicio 2.2** (Continuidad del push-forward). (a) Si  $\phi \in C_b(Y)$  entonces  $\phi \circ T \in C_b(X)$  y  $\int \phi d(T_\# \mu_n) = \int \phi \circ T d\mu_n$ . (b) Aplicar (a) a la medida empírica. (c)  $T = \mathbf{1}_{(0,\infty)}$  en  $\mathbb{R}$  y  $\mu_n = \delta_{1/n} \rightharpoonup \delta_0$ :  $T_\# \mu_n = \delta_1$  pero  $T_\# \delta_0 = \delta_0$ .

**Ejercicio 2.3** (Los mapas no son compactos). (a)  $f_n(x) = nx - \lfloor nx \rfloor$  (o cualquier reescalado de período  $1/n$  de una  $f$  que preserve  $\mathcal{L}^1_{[0,1]}$ ). Que preserve la medida se ve intervalo por intervalo; la convergencia débil a la media  $\frac{1}{2}$  es el lema de Riemann–Lebesgue para funciones periódicas: primero para  $g$  indicadora de un intervalo, luego por densidad. (b) Para  $\phi(x)\psi(y)$  el mismo argumento da  $\int \phi \int \psi$ ; concluir por densidad de las combinaciones de productos en  $C([0, 1]^2)$ . El producto  $\mathcal{L}^1 \otimes \mathcal{L}^1$  no está concentrado en ningún gráfico. Para el problema de Monge, el ínfimo sobre mapas puede no alcanzarse porque el límite de mapas no es un mapa. (c) No. Si  $f \in C^1$  preserva  $\mathcal{L}^1_{[0,1]}$ , entonces  $f'$  no se anula: cerca de un punto crítico  $x_0$  la imagen de un entorno de  $x_0$  tiene longitud mucho menor que el entorno, y la medida imagen concentraría masa cerca de  $f(x_0)$  (fórmula del Ejercicio 1.4(b)). Luego  $f$  es estrictamente monótona y, por el Ejercicio 1.4(c),  $f = \text{id}$  o  $f(x) = 1 - x$ ; ninguna sucesión con esos valores converge débilmente a  $\frac{1}{2}$ .

**Ejercicio 2.4** (La variación total no sirve).  $\pi_n$  está concentrada en el gráfico de  $T_n$  y  $\pi_m$  en el de  $T_m$ ; los dos gráficos son disjuntos (son conjuntos de medida nula distintos), así que el conjunto  $A = \text{graf}(T_n)$  da  $\pi_n(A) - \pi_m(A) = 1$ . Un espacio compacto para una métrica tiene sucesiones de Cauchy; una sucesión con distancias mutuas 1 no las tiene. Débilmente,  $\pi_n \rightharpoonup (\text{id}, \text{id})_\# \mu$  porque  $T_n \rightarrow \text{id}$  uniformemente salvo un conjunto de medida  $1/n$ .

**Ejercicio 2.5** (Regularización de Moreau–Yosida). (a)  $g_k(x) \leq g(x)$  tomando  $y = x$ ;  $g_k \geq \inf g$  es inmediato; la desigualdad triangular  $d(x, y) \leq d(x, x') + d(x', y)$  da  $g_k(x) \leq g_k(x') + k d(x, x')$ . (b)  $g_k$  es creciente en  $k$  y  $\leq g(x)$ . Si  $\lim_k g_k(x) < g(x) - \delta$ , elegir  $y_k$  casi óptimos:  $k d(x, y_k)$

acotado obliga  $y_k \rightarrow x$ , y la semicontinuidad inferior da  $\liminf g(y_k) \geq g(x)$ , contradicción. (c)  $\min\{g_k, k\}$  es Lipschitz, acotada y sigue creciendo hacia  $g$ .

**Ejercicio 2.6** (Integrales de funciones semicontinuas). (a) La desigualdad  $\geq$  es inmediata. Para  $\leq$ : las funciones  $g_k = \min\{f_k, k\}$  del Ejercicio 2.5 son continuas acotadas,  $\leq f$ , y crecen a  $f$ ; aplicar convergencia monótona. (b) Para cada  $g \in C_b$  con  $g \leq c$ ,  $\int g d\pi = \lim \int g d\pi_n \leq \liminf \int c d\pi_n$ ; tomar supremo en  $g$ . (c)  $c = \mathbf{1}_{\{x=y\}}$  con  $\pi_n = (\text{id}, T_n)_\# \mathcal{L}^1|_{[0,1]}$ ,  $T_n(x) = x + 1/n \bmod 1$  (Ejercicio 2.4):  $\int c d\pi_n = 0$  para todo  $n$ , pero  $\pi_n \rightarrow (\text{id}, \text{id})_\# \mathcal{L}^1$ , que tiene costo 1.

**Ejercicio 2.7** (Estabilidad de los planes óptimos). (a)  $\Pi(\mu, \nu)$  es débilmente compacto (Corolario 2.3.1). Si  $\gamma_{n_k} \rightarrow \gamma$ , para cualquier  $\eta \in \Pi(\mu, \nu)$ :  $\int c d\gamma = \lim \int c d\gamma_{n_k}$  (continuidad de  $c$  y convergencia débil)  $= \lim \int c_{n_k} d\gamma_{n_k}$  (convergencia uniforme)  $\leq \lim \int c_{n_k} d\eta = \int c d\eta$ . (b) La convergencia uniforme en compactos es directa en ambos casos. Para  $\varepsilon \rightarrow \infty$ , restar la constante  $\varepsilon$  (no cambia los planes óptimos) y multiplicar por  $2\varepsilon$ :  $2\varepsilon(\sqrt{\varepsilon^2 + |x-y|^2} - \varepsilon) = \frac{2\varepsilon|x-y|^2}{\sqrt{\varepsilon^2 + |x-y|^2} + \varepsilon} \rightarrow |x-y|^2$  uniformemente en compactos, así que los planes óptimos convergen a planes óptimos para el costo cuadrático.

**Ejercicio 2.8** (La convergencia débil es metrizable). Con  $d$  así definida,  $d(\mu_n, \mu) \rightarrow 0$  equivale a  $\int g_k d\mu_n \rightarrow \int g_k d\mu$  para todo  $k$ . Falta ver que probar contra una familia numerable adecuada equivale a la convergencia débil: usar que en un polaco toda sucesión débilmente convergente es ajustada (tight), aproximar  $\phi \in C_b$  por funciones Lipschitz acotadas (Ejercicio 2.5) y a éstas por los  $g_k$  uniformemente sobre un compacto que contiene casi toda la masa.

**Ejercicio 2.9** (Asignación y Kantorovich). Con masas uniformes el plan óptimo es un vértice de  $\mathcal{B}_n/n$ , es decir, una permutación (Ejercicios 1.11 y 3.5). La cota  $n+m-1$  es la del Ejercicio 3.4. Para el ejemplo con 5 entradas, elegir masas “genéricas” (por ejemplo  $a = (0,5,0,3,0,2)$ ,  $b = (0,4,0,35,0,25)$ ) y un costo con estructura monótona; para menos entradas, hacer coincidir masas, por ejemplo  $a = b$  con  $c_{ii} = 0$ .

**Ejercicio 2.10** (El segmento  $\Pi(\mu, \nu)$ ). El segmento es el del Ejercicio 1.7. El costo es afín en  $t$  (Ejercicio 1.8(c)), así que el mínimo se alcanza en un extremo, salvo que la función sea constante. Para  $c = \mathbf{1}_{\{x \neq y\}}$  y  $c = |x-y|^2$  el costo es  $(1-t) \cdot \text{cte}$ ; para  $-|x-y|^2$  es  $t \cdot \text{cte}$  negativa; ninguno es constante. Un costo de la forma  $\phi(x) + \psi(y)$  hace óptimos a todos los planes.

**Ejercicio 2.11** (Birkhoff). Que el óptimo del LP sea una permutación es el Ejercicio 3.5(a). La existencia de  $\sigma$  con  $M_{i\sigma(i)} > 0$  es el teorema de Hall aplicado al grafo de entradas positivas (o, directamente, la asignación de costo mínimo para  $-\log M$  tiene costo finito). Restar  $tM(\sigma)$  mantiene filas y columnas iguales y anula al menos una entrada más, así que el proceso termina en a lo sumo  $(n-1)^2 + 1$  pasos; en la práctica aparecen pocas permutaciones más que las cinco de partida.

**Ejercicio 2.12** (Costos cóncavos). Para  $p > 1$  el costo  $|x-y|^p$  es estrictamente convexo y el plan óptimo es el monótono, que casi no deja masa en la diagonal; para  $p < 1$  es cóncavo y conviene dejar quieta la masa común (Ejercicio 8.5); en  $p = 1$  todos los planes “monótonos a trozos” son óptimos y el solver devuelve uno cualquiera (Proposición 8.4.3). Medir la masa en la diagonal discreta como  $\sum_i \pi_{ii}$  sobre una grilla común.

**Ejercicio 2.13** (Convergencia débil de planes). Es la verificación numérica del Ejercicio 2.3(b): aproximar la integral con una grilla fina. Para la traslación módulo 1,  $f_n \rightarrow \text{id}$  salvo en un conjunto de medida  $1/n$ , así que  $\pi_n \rightarrow (\text{id}, \text{id})_\# \mathcal{L}^1$ , que sí está inducido por un mapa (la identidad); compárese con el Ejercicio 2.4.

### CAPÍTULO 3. UN BREVE REPASO DE DUALIDAD EN PROGRAMACIÓN LINEAL

**Ejercicio 3.1** (Duales de distintas formas). (a) Escribir  $Ax \geq b$  como  $Ax - s = b$ ,  $s \geq 0$ , y  $x = x^+ - x^-$ ; el dual de la forma estándar es  $\max\{b^T y : A^T y \leq c\}$ . Al deshacer las sustituciones aparecen las condiciones  $y \geq 0$  (por las holguras) o  $A^T y = c$  (por la ausencia de signo). (b)

Poner el dual en forma estándar y dualizar. (c)  $\min\{x : x \leq -1, x \geq 0\}$  tiene primal no factible y dual no acotado; para ambos no factibles hace falta al menos dos variables o restricciones contradictorias en ambos lados, por ejemplo  $\min\{-x_1 : x_1 - x_2 = 1, -x_1 + x_2 = 1, x \geq 0\}$ . La dualidad fuerte supone que uno de los dos es factible.

**Ejercicio 3.2** (Holgura complementaria). (a)  $c^T x - b^T y = c^T x - (Ax)^T y = x^T (c - A^T y)$ , suma de términos no negativos; es cero (óptimos, por dualidad fuerte) si y sólo si cada término lo es. (b) Las variables son  $\pi_{ij} \geq 0$ , las restricciones  $\sum_j \pi_{ij} = a_i$ ,  $\sum_i \pi_{ij} = b_j$ , y las variables duales  $\varphi_i, \psi_j$ ; la holgura de la variable  $\pi_{ij}$  es  $c_{ij} - \varphi_i - \psi_j$ .

**Ejercicio 3.3** (Un problema  $2 \times 2$  a mano). (a)  $\pi_{11} = t$  determina todo:  $\pi = \begin{pmatrix} t & 0,4-t \\ 0,5-t & t+0,1 \end{pmatrix}$  con  $t \in [0, 0,4]$ ; el costo es afín en  $t$ . (b) Normalizar  $\varphi_1 = 0$  y saturar las restricciones sobre el soporte del plan óptimo hallado en (a). (c) Las restricciones sólo involucran  $\varphi_i + \psi_j$ , invariantes por  $(\varphi + s, \psi - s)$ ; que no hay más libertad se ve porque el grafo de soporte es conexo (Ejercicio 5.2).

**Ejercicio 3.4** (Vértices del politopo de transporte). (a) Hay  $NM$  variables y  $N+M$  restricciones de igualdad, de las cuales una es redundante ( $\sum a_i = \sum b_j$ ); el interior relativo es no vacío porque  $a \otimes b$  tiene todas sus entradas positivas. (b) Si el grafo de entradas positivas contiene un ciclo, perturbar  $\pm \varepsilon$  alternadamente a lo largo del ciclo como en el Ejercicio 1.11(c). Un grafo sin ciclos con  $N+M$  vértices tiene a lo sumo  $N+M-1$  aristas. (c) Un programa lineal con óptimo finito lo alcanza en un vértice.

**Ejercicio 3.5** (El problema de asignación). (a) El mínimo de un funcional lineal sobre  $\mathcal{B}_n/N$  se alcanza en un extremal (Ejercicio 1.10), que es una permutación por Birkhoff. (b) Es la holgura complementaria del Ejercicio 3.2 para el plan  $\pi_{ij} = \frac{1}{N} \delta_{j\sigma(i)}$ . (c) Si  $i < k$  y  $\sigma(i) > \sigma(k)$ , intercambiar:  $(x_i - y_{\sigma(k)})^2 + (x_k - y_{\sigma(i)})^2 \leq (x_i - y_{\sigma(i)})^2 + (x_k - y_{\sigma(k)})^2$  equivale a  $(x_k - x_i)(y_{\sigma(i)} - y_{\sigma(k)}) \geq 0$ . Repetir hasta eliminar todos los cruces.

## CAPÍTULO 4. DUALIDAD

**Ejercicio 4.1** (Desigualdades “casi en todo punto” y planes). (a) El conjunto donde falla la desigualdad está contenido en  $(N \times Y) \cup (X \times M)$ , y  $\pi(N \times Y) = \mu(N) = 0$ ,  $\pi(X \times M) = \nu(M) = 0$ . (b) La desigualdad  $1 \leq \mathbf{1}_{\{x \neq y\}}$  falla exactamente en la diagonal, que tiene medida  $\mu \otimes \nu$  nula pero medida  $\pi$  igual a 1. (c) Con la definición c.t.p. el par  $(\frac{1}{2}, \frac{1}{2})$  sería admisible con valor dual 1, pero el plan diagonal tiene costo 0: la dualidad débil fallaría.

**Ejercicio 4.2** (Dualidad a mano). (a)  $-x + y \leq |x - y|$  siempre; sobre  $\Pi(\mu, \nu)$  se tiene  $y \geq x$  c.t.p., así que  $\int |x - y| d\pi = \int (y - x) d\pi = \int y d\nu - \int x d\mu = 1$  para *todo* plan: todos son óptimos. (b) Buscar  $\psi(y) = \alpha y + \beta$  y  $\varphi = \psi^c$ ; con  $T(x) = x + 1$  la saturación  $\varphi(x) + \psi(x + 1) = 1$  y  $\varphi(x) = \inf_y (|x - y|^2 - \psi(y))$  dan  $\alpha = 2$ . Para el costo cuadrático el plan óptimo es único (Capítulo 8). (c) En (a) el costo es una diferencia de funciones de  $x$  y de  $y$  por separado, y el problema es trivial.

**Ejercicio 4.3** (Funciones semicontinuas y medidas signadas). Si  $\mu \geq 0$  es el Ejercicio 2.6(a). Si no, hay  $A$  boreliano con  $\mu(A) < 0$  (descomposición de Hahn); por regularidad se puede tomar  $A$  compacto y aproximar  $-k\mathbf{1}_A$  por funciones continuas  $g_k \leq \min f$  que valen  $-k$  en  $A$  y son  $\geq -k$  fuera, con soporte en un entorno pequeño de  $A$ ; entonces  $\int g_k d\mu \approx -k\mu(A) \rightarrow +\infty$ .

**Ejercicio 4.4** (Dualidad para costos semicontinuos inferiores). (a) Compacidad de  $\Pi(\mu, \nu)$ ; para  $k \geq k_0$ ,  $\int c_{k_0} d\pi_k \leq \int c_k d\pi_k = \min_{\Pi} \int c_k$ , y la continuidad de  $c_{k_0}$  permite pasar al límite en  $k$ . Después  $k_0 \rightarrow \infty$  por convergencia monótona. (b) Si  $\varphi \oplus \psi \leq c_k \leq c$  entonces el par es admisible para  $c$ . La cadena de desigualdades es la del enunciado; el paso del medio es el teorema para costos continuos. (c) Sin compacidad hay que trabajar con costos  $c \geq 0$  semicontinuos y medidas ajustadas; el resultado sigue valiendo pero la demostración pasa por aproximar  $\mu, \nu$  por medidas de soporte compacto.

## CAPÍTULO 5. POTENCIALES DE KANTOROVICH EN EL CASO DISCRETO

**Ejercicio 5.1** (Un problema  $3 \times 3$  con potenciales no únicos). (a)  $\pi_{11} = \pi_{12} = \pi_{23} = \frac{1}{4}$ ,  $\pi_{33} = \frac{1}{2}$ ; costo  $\frac{1}{2}$ . (b) Las cuatro saturaciones dan  $\psi_1 = 0$ ,  $\psi_2 = 1$ ,  $\varphi_2 = 1 - \psi_3$ ,  $\varphi_3 = -\psi_3$ ; llamar  $t = \psi_3$ . De las cinco restricciones restantes, la que manda es  $\varphi_2 + \psi_2 \leq c_{22} = 0$ , o sea  $t \geq 2$ , y  $\varphi_1 + \psi_3 \leq c_{13} = 4$  da  $t \leq 4$ . (c) Los coeficientes de  $t$  son  $-a_2 - a_3 + b_3 = 0$ . (d)  $G_\pi$  tiene dos componentes:  $\{1\} \cup \{1, 2\}$  y  $\{2, 3\} \cup \{3\}$ ; la constante aditiva de cada componente es libre, y sólo una de las dos se fija con la normalización.

**Ejercicio 5.2** (Unicidad de los potenciales y conexión del soporte). (a) Es la holgura complementaria. (b) Fijar  $\varphi_{i_0}$ ; recorriendo las aristas de  $G_\pi$  (cada una da una ecuación  $\varphi_i + \psi_j = c_{ij}$ ) se determinan todos los valores, y la conexión garantiza que se llega a todos los vértices. (c) Cada componente aporta a lo sumo una constante libre; la dimensión puede ser menor si alguna restricción de admisibilidad entre componentes es una igualdad. Para el ejemplo con dimensión exactamente  $k$ , tomar  $k$  problemas “desacoplados” (soportes muy separados) y sumarlos.

**Ejercicio 5.3** (Costo 0–1 y variación total). (a) Para cualquier plan, la masa que sale de  $x_i$  es al menos  $(a_i - b_i)^+$ ; el plan que deja quieta  $\min\{a_i, b_i\}$  y reparte el resto de cualquier manera alcanza la cota. (b)  $\varphi_i = \mathbf{1}_{\{a_i > b_i\}}$ ,  $\psi_i = -\mathbf{1}_{\{a_i > b_i\}}$ : las restricciones  $\varphi_i + \psi_j \leq \mathbf{1}_{\{i \neq j\}}$  se verifican caso por caso y el valor dual es  $\sum (a_i - b_i)^+$ . (c) La cota inferior es  $\pi(\{x \neq y\}) \geq \sup_A (\mu(A) - \nu(A))$  porque  $\mu(A) - \nu(A) = \pi(A \times A^c) - \pi(A^c \times A) \leq \pi(\{x \neq y\})$ ; para la igualdad, descomponer  $\mu - \nu = (\mu - \nu)^+ - (\mu - \nu)^-$  (Hahn), dejar quieta la parte común  $\mu \wedge \nu$  y acoplar de cualquier modo las dos partes de la descomposición.

**Ejercicio 5.4** (La cara óptima). (a)  $\Pi^0 = \Pi(a, b) \cap \{\langle c, \pi \rangle = \min\}$  es la intersección de un politopo con un hiperplano de soporte, luego una cara; un extremal de una cara es extremal del politopo (Ejercicio 1.10(b)). (b) Para todo plan,  $\int |x - y| d\pi \geq \int (y - x) d\pi = \bar{y} - \bar{x} = 1$ , con igualdad si y sólo si  $\pi$  está concentrado en  $\{y \geq x\}$ , es decir,  $j \geq i - 1$ . Los extremales óptimos son las permutaciones con  $\sigma(i) \geq i - 1$ : la identidad ( $x_i \mapsto y_i$ ), la que manda  $x_0 \mapsto y_4$  y baja los demás un lugar, etc. (c) Sobre  $\{j \geq i - 1\}$  la saturación dice  $\varphi_i + \psi_j = y_j - x_i$ ; el grafo es conexo, así que  $\varphi_i = -x_i + s$ ,  $\psi_j = y_j - s$ , y estos pares son admisibles también fuera (allí  $|x_i - y_j| = x_i - y_j \geq y_j - x_i$ ).

## CAPÍTULO 6. POTENCIALES DE KANTOROVICH

**Ejercicio 6.1** (Propiedades de la  $c$ -transformada). (a)  $\chi^{c\bar{c}}(x) = \inf_y \{c(x, y) - \inf_{x'} [c(x', y) - \chi(x')]\} \geq \inf_y \{c(x, y) - c(x, y) + \chi(x)\}$  tomando  $x' = x$ . Aplicar esto y la monotonía de (c) a  $\chi^c$ :  $\chi^{c\bar{c}c} \leq \chi^c$  por  $\chi^{c\bar{c}} \geq \chi$ , y  $\geq$  por (a) aplicado a  $\chi^c$  con los papeles de  $X$  e  $Y$  intercambiados. (b) Si  $\varphi = \zeta^{\bar{c}}$  entonces  $\varphi^{c\bar{c}} = \zeta^{c\bar{c}c} = \zeta^{\bar{c}}$  por (a). (c) Directo de la definición. (d)  $\inf_y \inf_y \{c(x, y) - \zeta_i(y)\} = \inf_y \{c(x, y) - \sup_i \zeta_i(y)\}$ .

**Ejercicio 6.2** (Equicontinuidad del ínfimo). Para cada  $\alpha$ ,  $f(x, \alpha) \leq f(x', \alpha) + \omega(d(x, x'))$ ; tomar ínfimo en  $\alpha$  a ambos lados y luego intercambiar  $x$  y  $x'$ . En el Lema 6.2.1 se aplica con  $A = Y$  y  $f(x, y) = c(x, y) - \psi(y)$ : el módulo de continuidad de  $\psi^{\bar{c}}$  es el de  $c$  en la variable  $x$ , independientemente de  $\psi$ .

**Ejercicio 6.3** (Costos que son distancias). (a) Si  $\varphi = \zeta^{\bar{c}}$ , entonces  $\varphi(x) \leq d(x, y) - \zeta(y) \leq d(x, x') + d(x', y) - \zeta(y)$  y tomando ínfimo en  $y$ ,  $\varphi(x) \leq \varphi(x') + d(x, x')$ . Recíprocamente, si  $\varphi$  es 1-Lipschitz,  $\varphi^c(y) = \inf_x \{d(x, y) - \varphi(x)\} = -\varphi(y)$  (la desigualdad  $\geq$  es la condición de Lipschitz, y  $x = y$  da la igualdad), y entonces  $\varphi^{c\bar{c}} = (-\varphi)^{\bar{c}} = \varphi$  por el mismo cálculo. (b) Sustituir  $\psi = \varphi^c = -\varphi$  en el funcional dual. (c)  $t \mapsto t^\alpha$  es cóncava con  $0^\alpha = 0$ , luego subaditiva:  $(s+t)^\alpha \leq s^\alpha + t^\alpha$ ; las funciones 1-Lipschitz para esta distancia son exactamente las de seminorma de Hölder  $[u]_{0, \alpha} \leq 1$ .

**Ejercicio 6.4** ( $c$ -transformadas para el costo cuadrático). (a) Minimizar  $|x - y|^2 - \lambda|x|^2$  en  $x$ : si  $\lambda \geq 1$  el ínfimo es  $-\infty$  (tomar  $|x| \rightarrow \infty$ ); si  $\lambda < 1$  el mínimo está en  $x = y/(1 - \lambda)$ . (b)  $\chi^c(y) = -a \cdot y - |a|^2/4 - b$  (mínimo en  $x = y + a/2$ ) y  $\chi^{c\bar{c}} = \chi$ : las afines son  $c$ -cóncavas. (c)

$\varphi = \zeta^c$  equivale a  $|x|^2 - \varphi(x) = \sup_y \{2x \cdot y - (|y|^2 - \zeta(y))\}$ , un supremo de afines en  $x$ , es decir, una función convexa sci; recíprocamente toda convexa sci es su propia biconjugada. Desarrollando el cuadrado,  $\varphi^c(y) = |y|^2 + \inf_x \{u(x) - 2x \cdot y\} = |y|^2 - \bar{u}(2y)$ .

**Ejercicio 6.5** (La mejora de potenciales converge en el caso discreto). (a) Es el Ejercicio 6.1(a):  $\varphi^{c\bar{c}}$  es  $c$ -cóncava y  $(\varphi^{c\bar{c}})^{c\bar{c}} = \varphi^{c\bar{c}}$ . (b) En el  $2 \times 2$ , partir de  $\varphi = (0, 0)$ : se obtiene  $\psi = \varphi^c = (\min_i c_{i1}, \min_i c_{i2})$  y  $\varphi^{c\bar{c}}$ , un par admisible y ajustado cuyo valor dual se compara con el óptimo calculado en el Ejercicio 3.3; en general son distintos. La iteración se detiene porque la  $c$ -transformada discreta es un mínimo exacto, sin “memoria” de las masas  $a, b$ .

**Ejercicio 6.6** (El  $2 \times 2$  y su degeneración). El vértice óptimo cambia cuando los dos planes extremales tienen el mismo costo: escribir el costo de cada extremo del segmento  $\Pi(a, b)$  como función de  $s$  e igualar. En ese valor de  $s$  el costo es constante sobre el segmento y las dos saturaciones “cruzadas” pueden cumplirse a la vez, lo que deja un grado de libertad extra en los potenciales.

**Ejercicio 6.7** (Potenciales únicos y grafo conexo). Con las masas del Ejercicio 5.1 el plan óptimo tiene sólo cuatro entradas positivas y el grafo es desconexo. Perturbar  $a$  de modo que el plan óptimo necesite cinco entradas ( $n + m - 1$ , Ejercicio 3.4): con cinco aristas y seis vértices el grafo bipartito es un árbol, luego conexo, y los potenciales quedan determinados salvo la constante.

**Ejercicio 6.8** (Holgura complementaria y saturación espuria). La saturación sin masa ocurre cuando el plan óptimo no es único o cuando el vértice óptimo es degenerado (menos de  $n + m - 1$  entradas positivas pero  $n + m - 1$  restricciones duales activas). Con puntos aleatorios es raro; en una grilla regular abundan las igualdades de costos y aparece con frecuencia.

**Ejercicio 6.9** (Mejora de potenciales). Si  $\varphi$  ya es un potencial óptimo,  $\varphi^{c\bar{c}} = \varphi$  y  $\varphi^c$  es su compañero óptimo: la brecha se cierra en un paso. Para  $\varphi$  genérico, comparar  $\mathcal{D}(\varphi^{c\bar{c}}, \varphi^c)$  con el óptimo del programa lineal; el histograma muestra que la mejora suele ser grande pero rara vez completa (Ejercicio 6.5).

**Ejercicio 6.10** (Costo 0–1). Sobre soportes disjuntos el costo  $\mathbf{1}_{\{x \neq y\}}$  vale 1 en todo par, así que todo plan tiene costo 1; coincide con  $\|\mu - \nu\|_{TV} = 1$  para medidas mutuamente singulares (Ejercicio 5.3(c)).

## CAPÍTULO 7. ALGUNAS INTERPRETACIONES Y EXTENSIONES

**Ejercicio 7.1** (Acoplamientos y covarianza). (a) Desarrollar  $\mathbb{E}|X - Y|^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 - 2\mathbb{E}XY$  y escribir cada término con medias y varianzas. (b) Cauchy–Schwarz para  $\text{Cov}(X, Y) = \mathbb{E}[(X - m_\mu)(Y - m_\nu)]$ ; sustituir en (a). (c) Igualdad en Cauchy–Schwarz si y sólo si  $Y - m_\nu = a(X - m_\mu)$  c.s., y  $\text{Cov} > 0$  obliga  $a > 0$ ; entonces  $\nu$  es la imagen de  $\mu$  por  $x \mapsto ax + b$ .

**Ejercicio 7.2** (Cotas de Fréchet–Hoeffding). (a)  $H(x, y) \leq \pi((-\infty, x] \times \mathbb{R}) = F_\mu(x)$  y lo mismo con  $y$ ; para la cota inferior,  $H(x, y) = F_\mu(x) + F_\nu(y) - \pi((-\infty, x] \times \mathbb{R} \cup \mathbb{R} \times (-\infty, y]) \geq F_\mu(x) + F_\nu(y) - 1$ . (b) La cota superior es la función de distribución de  $(F_\mu^{[-1]}(U), F_\nu^{[-1]}(U))$  con  $U$  uniforme en  $(0, 1)$  (Ejercicio 8.2(a)), y la inferior la de  $(F_\mu^{[-1]}(U), F_\nu^{[-1]}(1 - U))$ .

**Ejercicio 7.3** (Matching estable: un ejemplo). (a) El cruzado,  $\{(x_1, y_2), (x_2, y_1)\}$ , tiene excedente 6; el codicioso empareja primero  $(x_1, y_1)$  y obtiene 5. (b) Con  $\varphi_1 = s$  se obtiene  $\psi_2 = 3 - s$ ,  $\psi_1 \in [4 - s, 5 - s]$ ,  $\varphi_2 = 3 - \psi_1$ . (c) El segmento está parametrizado por  $\psi_1$  (o por  $s$ , que es la constante aditiva); en un extremo la restricción  $\varphi_1 + \psi_1 \geq 4$  está saturada y el par no emparejado  $(x_1, y_1)$  es indiferente, en el otro lo está  $(x_2, y_2)$ . Al mover el parámetro, el excedente se traslada de los trabajadores a las empresas.

**Ejercicio 7.4** (Un criterio de optimalidad multimarginal). (a) Integrar  $x_1 + \dots + x_N = v$  contra  $\gamma$  y usar las marginales. (b) Para cualquier  $\tilde{\gamma} \in \Pi(\mu_1, \dots, \mu_N)$ , Jensen da  $\int h(\sum x_i) d\tilde{\gamma} \geq h(\int \sum x_i d\tilde{\gamma}) = h(v) = \int h(\sum x_i) d\gamma$ . (c) Desarrollando,  $\sum_{i \neq j} |x_i - x_j|^2 = 2N \sum_i |x_i|^2 - 2|\sum_i x_i|^2$ ;

el primer término tiene integral fija sobre  $\Pi(\mu_1, \dots, \mu_N)$ , así que minimizar  $c$  equivale a minimizar  $\int |\sum_i x_i|^2 d\gamma$ . (d) En el primer caso sí: el plan antimonótono, concentrado en  $\{(0, 1), (1, 0)\}$ , tiene suma constante 1. En el segundo no:  $x_1 + x_2 = x_1$  no puede ser constante bajo una marginal uniforme.

**Ejercicio 7.5** (Discretización y estabilidad). (a) La compacidad de  $X$  y la convergencia débil dan  $\int d(x, x')^2 d\alpha_n \rightarrow 0$  (el costo se anula sólo en la diagonal y  $(\text{id}, \text{id})_{\#}\mu$  es el único plan de costo cero entre  $\mu$  y  $\mu$ ; usar el Ejercicio 2.7 con  $\mu_n \rightarrow \mu$ ). Pegando  $\alpha_n, \gamma$  y  $\beta_n$  se obtiene un plan en  $X \times X \times Y \times Y$ ; su proyección a las coordenadas extremas está en  $\Pi(\mu_n, \nu_n)$  y difiere de  $\gamma$  en dos “desplazamientos” pequeños en  $L^2$ , lo que basta para la convergencia débil (probar contra funciones Lipschitz). (b) Combinar (a) con la semicontinuidad inferior del costo y la compacidad de los planes. (c) Aplicar (b) con  $\mu_n, \nu_n$  empíricas, que convergen débilmente por la ley de los grandes números.

## CAPÍTULO 8. TRANSPORTE ÓPTIMO EN DIMENSIÓN UNO

**Ejercicio 8.1** (Mapas monótonos explícitos).  $T_{\text{mon}} = F_{\nu}^{[-1]} \circ F_{\mu}$ . (a)  $T(x) = -\log(1 - x)$ ; plan inducido por un mapa. (b)  $T = 0$  en  $[0, 1/3]$  y  $T = 1$  en  $[1/3, 1]$ : sí es un mapa. (c)  $\mu$  tiene átomos y  $\nu$  no:  $\gamma_{\text{mon}}$  reparte la masa de 0 uniformemente en  $[0, 1/2]$  y la de 1 en  $[1/2, 1]$ , y no hay ningún mapa (un mapa mandarí cada átomo a un punto). (d)  $T(x) = m_1 + \frac{\sigma_1}{\sigma_0}(x - m_0)$ , que es el mapa de Brenier gaussiano en dimensión uno.

**Ejercicio 8.2** (Propiedades de la pseudoinversa). (a) Monotonía y continuidad por la izquierda salen de que  $F_{\mu}$  es no decreciente y continua por la derecha; la desigualdad de Galois es la definición de ínfimo. (b) Por (a),  $(F_{\mu}^{[-1]})_{\#}\mathcal{L}^1|_{(0,1)}$  tiene función de distribución  $\mathcal{L}^1(\{t : F_{\mu}^{[-1]}(t) \leq x\}) = \mathcal{L}^1((0, F_{\mu}(x)]) = F_{\mu}(x)$ , luego es  $\mu$ ; aplicar el Ejercicio 1.2. (c) Calcular  $F_{\mu}(x) = \mathcal{L}^1(\{G \leq x\}) = \sup\{t : G(t) \leq x\}$  y verificar que su pseudoinversa devuelve  $G$  usando la continuidad por la izquierda. (d) (b) da la inyectividad y (c) la sobreyectividad.

**Ejercicio 8.3** (Dominancia estocástica). (a)  $\Rightarrow$  (d):  $F_{\mu} \geq F_{\nu}$  equivale a  $F_{\mu}^{[-1]} \leq F_{\nu}^{[-1]}$ , y  $T_{\text{mon}}(x) = F_{\nu}^{[-1]}(F_{\mu}(x)) \geq F_{\mu}^{[-1]}(F_{\mu}(x)) = x$ . (d)  $\Rightarrow$  (c)  $\Rightarrow$  (b) son inmediatas. (b)  $\Rightarrow$  (a): si  $\gamma$  vive en  $\{y \geq x\}$  entonces  $\{y \leq t\} \subset \{x \leq t\}$   $\gamma$ -c.t.p., luego  $F_{\nu}(t) = \gamma(\{y \leq t\}) \leq \gamma(\{x \leq t\}) = F_{\mu}(t)$ . Con átomos,  $T_{\text{mon}}$  puede no transportar  $\mu$  en  $\nu$  y hay que reemplazar (c) y (d) por el plan comonótono; sin soporte pleno las igualdades  $F_{\mu}^{[-1]} \circ F_{\mu} = \text{id}$  valen sólo  $\mu$ -c.t.p.

**Ejercicio 8.4** (Costos con  $\partial^2 c / \partial x \partial y > 0$ ). (a) Copiar la demostración de la Proposición 8.3.5: la condición de dos puntos  $c(x, y) + c(x', y') \leq c(x, y') + c(x', y)$  se escribe como  $\int_x^{x'} \int_y^{y'} \partial_{xy}^2 c \leq 0$  (con el signo adecuado), lo que con  $x < x'$  y derivada cruzada positiva obliga  $y \geq y'$ . (b) Aplicar el Teorema 8.4.1 a  $\tilde{c}(x, y) = c(x, -y)$ , que tiene  $\partial_{xy}^2 \tilde{c} < 0$ , entre  $\mu$  y  $R_{\#}\nu$ , y componer con  $R$ . (c)  $\partial_{xy}^2(-|x - y|^2) = 2 > 0$  y  $\partial_{xy}^2(xy) = 1 > 0$ .

**Ejercicio 8.5** (Costos cóncavos: el plan monótono no es óptimo). (a)  $\gamma_{\text{mon}}$  empareja  $0 \mapsto 1$  y  $1 \mapsto 2$ , costo 1; el otro plan empareja  $0 \mapsto 2$  y  $1 \mapsto 1$ , costo  $\frac{1}{2}\sqrt{2} < 1$ . (b) Falla la convexidad estricta de  $h(z) = |z|^{1/2}$ ; la condición de dos puntos  $c(0, 1) + c(1, 2) \leq c(0, 2) + c(1, 1)$  es  $2 \leq \sqrt{2}$ , falsa. (c) Si un plan óptimo mueve una parte de la masa común  $x \mapsto y \neq x$  y otra parte  $x' \mapsto x$ , la desigualdad triangular estricta de  $h$  (subaditividad estricta por concavidad) muestra que mandar  $x' \mapsto y$  directamente y dejar  $x$  quieto cuesta menos; ver Santambrogio, Sec. 3.3.2.

**Ejercicio 8.6** (Densidades sin soporte pleno). (a) Enumerar los racionales  $q_k$  de  $[0, 1]$  y poner alrededor de cada uno un conjunto de Cantor de medida positiva  $K_k$  contenido en el intervalo de centro  $q_k$  y radio  $2^{-k}$ , eligiendo sus medidas de modo que la masa total que cae dentro de cada intervalo diádico sea menor que la mitad de su longitud (los  $K_k$  son cerrados sin interior, así que se puede controlar). Entonces  $A = \bigcup K_k$  tiene medida positiva en todo intervalo, y su complemento también. (b)  $F_{\mu}(x) = \mathcal{L}^1(A \cap [0, x]) / \mathcal{L}^1(A)$  es continua (sin átomos) y estrictamente creciente ( $A$  carga todo intervalo);  $T_{\text{mon}} = F_{\mu}$  es Lipschitz porque su derivada  $\mathbf{1}_A / \mathcal{L}^1(A)$  está

acotada, pero la inversa no lo es:  $F'_\mu = 0$  en un conjunto de medida positiva de todo intervalo, así que  $F_\mu^{-1}$  tiene pendientes arbitrariamente grandes.

**Ejercicio 8.7** (Un costo singular en la diagonal). (a) Basta exhibir un plan de costo finito (por ejemplo, para  $\mu$  uniforme,  $y = x + \frac{1}{2}$  mód 1) y usar la semicontinuidad inferior de  $c$  y la compacidad de  $\Pi(\mu, \mu)$ . (b) La monotonía cíclica de dos puntos para  $c = 1/|x - y|$  dice que, si  $(x, y)$  y  $(x', y')$  están en el soporte con  $x < x'$  y ambos pares del mismo lado de la diagonal, entonces  $y \leq y'$  (el costo es convexo en  $y - x$  a cada lado de la diagonal); de ahí las dos gráficas crecientes. Que no se superponen sale de comparar el costo de cruzar la diagonal con el de no hacerlo. El mapa resultante no es creciente globalmente: salta de una rama a la otra. (c)  $T(x) = x + \frac{1}{2}$  para  $x < \frac{1}{2}$  y  $T(x) = x - \frac{1}{2}$  para  $x \geq \frac{1}{2}$ .

**Ejercicio 8.8** (Mapas monótonos explícitos). Discretizar  $\mu$  y  $\nu$  con muchos puntos y comparar  $F_\nu^{[-1]} \circ F_\mu$  con las fórmulas del Ejercicio 8.1. En el caso (c) el solver devuelve un plan que reparte cada átomo en muchos puntos: contar cuántas entradas positivas hay por fila.

**Ejercicio 8.9** (Dominancia estocástica). Para gaussianas,  $T_{\text{mon}}(x) = m + \sigma x$  y la condición  $T_{\text{mon}} \geq \text{id}$  es  $m + (\sigma - 1)x \geq 0$  para todo  $x$ , que sólo vale si  $\sigma = 1$  y  $m \geq 0$ : la región es la semirrecta  $\{\sigma = 1, m \geq 0\}$ . Numéricamente se ve que la condición falla en las colas apenas  $\sigma \neq 1$ .

**Ejercicio 8.10** (Cotas de Fréchet–Hoeffding). Construir  $H$  como suma acumulada doble de la matriz del plan; el comonótono y el antimonótono se obtienen ordenando los átomos (Ejercicio 7.2(b)) y deben coincidir con las cotas en todos los nodos de la grilla.

**Ejercicio 8.11** (Costos cóncavos). Con las medidas del Ejercicio 8.5, el costo del plan monótono es 1 y el del cruzado es  $\frac{1}{2}2^p$ ; el monótono deja de ser óptimo exactamente cuando  $2^{p-1} < 1$ , es decir, para  $p < 1$ . Con densidades continuas, medir  $\sum_i \pi_{ii}$  sobre una grilla común.

**Ejercicio 8.12** (La fórmula del área y  $p > 1$ ). Para  $W_1$  usar la fórmula del Ejercicio 10.5 con las funciones de distribución empíricas (escalonadas). Para  $p = 2$  la fórmula correcta es  $W_2^2 = \int_0^1 |F_\mu^{[-1]} - F_\nu^{[-1]}|^2 dt$ , que integra en la variable de los cuantiles, no en  $x$ ; las dos integrales sólo coinciden para  $p = 1$ , donde el área entre las gráficas se puede medir tanto horizontal como verticalmente.

## CAPÍTULO 9. EL TEOREMA DE FACTORIZACIÓN POLAR DE BRENIER

**Ejercicio 9.1** (Traslaciones, dilataciones y gradientes). (a)  $T = \nabla u$  con  $u(x) = \frac{1}{2}|x|^2 + v \cdot x$ ,  $u = \frac{\lambda}{2}|x|^2$  y  $u = \frac{1}{2}x \cdot Ax + b \cdot x$ , convexas en los casos indicados; aplicar el Corolario 9.1.8. Para  $\lambda < 0$ , la condición de dos puntos  $(x - x') \cdot (\lambda x - \lambda x') = \lambda|x - x'|^2 \geq 0$  falla en cuanto el soporte tenga dos puntos. (b) La monotonía cíclica de dos puntos da  $(x - x') \cdot (-x + x') \geq 0$ , es decir  $|x - x'|^2 \leq 0$ , para todo par de puntos del soporte. (c)  $\nabla \bar{u} \circ \nabla u = \text{id}$  es la relación entre subdiferenciales de  $u$  y  $\bar{u}$  (Apéndice D); entonces  $(\nabla \bar{u})_\# \nu = \mu$  y  $\nabla \bar{u}$  es gradiente de una convexa.

**Ejercicio 9.2** (El problema de maximización).  $\int |x - y|^2 d\pi = \int |x|^2 d\mu + \int |y|^2 d\nu - 2 \int x \cdot y d\pi$ , así que maximizar equivale a minimizar  $\int x \cdot y d\pi$ , que es el problema de Kantorovich para  $\tilde{c}(x, y) = |x + y|^2$  (misma reducción), es decir, el costo cuadrático entre  $\mu$  y  $R_\# \nu$ . El Teorema 9.1.3 da un único mapa óptimo  $\nabla u$  para ese problema; el mapa buscado es  $T = -\nabla u = \nabla(-u)$ .

**Ejercicio 9.3** (Un mapa radial). (a)  $\mu(B(0, r)) = r^2$  y  $\nu(B(0, \rho)) = \frac{\rho^2}{2} - \frac{\rho^4}{16}$ ; igualar y resolver la cuadrática en  $\rho^2$  eligiendo la raíz que da  $\phi(0) = 0$ . (b)  $u$  es radial con  $u'(r) = \phi(r)$  creciente, luego convexa;  $\nabla u(x) = \phi(|x|)\frac{x}{|x|}$ , y  $T_\# \mu = \nu$  por construcción (basta comprobarlo en bolas centradas). (c) La medida de  $|x|$  bajo  $\mu$  es  $\mathcal{L}^1|_{[0,1]} \cdot 2r$ ; un mapa radial  $T(x) = \phi(|x|)x/|x|$  es gradiente de la convexa  $\int_0^{|x|} \phi$  si y sólo si  $\phi$  es creciente, y transporta  $\mu$  en  $\nu$  si y sólo si  $\phi$  transporta las leyes de los radios: es el mapa monótono unidimensional del Capítulo 8.

**Ejercicio 9.4** (Imágenes afines de una medida invariante por rotaciones). (a)  $L_B \circ L_A^{-1}(x) = b + BA^{-1}(x - a)$  es afín con matriz  $BA^{-1}$ , que es simétrica (y entonces óptima por el Ejercicio 9.1) si y sólo si  $A$  y  $B$  conmutan. (b)  $T$  transporta  $\mu$  en  $\nu$  si y sólo si  $(MA)_{\#}\rho = B_{\#}\rho$ ; por la invariancia de  $\rho$  esto vale si  $MA = BQ$  con  $Q$  ortogonal, y por la factorización polar toda matriz inversible se escribe así. Elevando al cuadrado  $MA = BQ$ :  $(MA)(MA)^T = MA^2M = B^2$ , de donde  $A^{-1}(AB^2A)^{1/2}A^{-1}$  es la única solución simétrica definida positiva. (c) Es la fórmula del mapa óptimo entre gaussianas  $N(a, A^2)$  y  $N(b, B^2)$ .

**Ejercicio 9.5** (Una reflexión que no es óptima). (a) Para  $(x, Sx)$  y  $(x', Sx')$  en el soporte, la condición de dos puntos es  $(x - x') \cdot (Sx - Sx') \geq 0$ , es decir  $|\hat{x} - \hat{x}'|^2 - (x_d - x'_d)^2 \geq 0$ . Entonces  $x_d$  es función 1-Lipschitz de  $\hat{x}$  sobre el soporte, y se extiende a todo  $\mathbb{R}^{d-1}$  (Kirschbraun, o directamente por la fórmula de McShane). (b) Un gráfico Lipschitz tiene medida nula (Fubini). (c) La uniforme sobre el segmento  $\{(s, s) : s \in [0, 1]\}$  en  $\mathbb{R}^2$ : como  $x \cdot Sx = 0$  para todo  $x$  del soporte, y también  $x \cdot y = 0$  para todo  $x \in \text{sop } \mu, y \in \text{sop } S_{\#}\mu$ , el costo  $\int |x - y|^2 d\pi = \int |x|^2 d\mu + \int |y|^2 d\nu - 2 \int x \cdot y d\pi$  es el mismo para todos los planes, y  $S$  es óptimo. En general, cualquier  $\mu$  concentrada en un gráfico 1-Lipschitz sirve, pero la optimalidad de  $S$  hay que verificarla con un criterio suficiente (por ejemplo, Knott-Smith, Ejercicio 9.6(a)), porque el recíproco del Teorema 8.3.2 es la Observación que sigue a ese teorema, que las notas no demuestran.

**Ejercicio 9.6** (Knott-Smith y un ejemplo de no unicidad). (a) Por Fenchel-Young,  $u(x) + \bar{u}(y) \geq x \cdot y$  para todo par, así que  $\varphi = \frac{1}{2}|x|^2 - u, \psi = \frac{1}{2}|y|^2 - \bar{u}$  es admisible para el costo  $\frac{1}{2}|x - y|^2$  y satura sobre el soporte de  $\pi$ : dualidad débil. (b) Con  $x = s(1, 1) + r(1, -1)$ ,  $x \cdot y - |x_1 - x_2| = s(y_1 + y_2) + r(y_1 - y_2) - 2|r|$ , acotado superiormente si y sólo si  $y_1 + y_2 = 0$  y  $|y_1 - y_2| \leq 2$ , con supremo 0. Los cuatro pares posibles cumplen  $x \cdot y = 0 = u(x) + \bar{u}(y)$ . (c) Son las mismas medidas del ejemplo de la Sección 12.3: la no unicidad del plan y la del baricentro tienen el mismo origen.

**Ejercicio 9.7** (La convexidad como condición de segundo orden). (a)  $\rho_t$  resuelve la ecuación de continuidad con velocidad  $v$ ; como  $\nabla \cdot (fv) = 0$ , la función constante  $\rho_t \equiv f$  también la resuelve, y la unicidad (para  $v$  suave) da  $\rho_t = f$ . (b) Desarrollar  $|T_t - x|^2$  y usar (a) para cambiar variables  $y = Y_t(x)$ ; derivar dos veces bajo la integral con  $\frac{d}{dt}Y_{-t} = -v(Y_{-t})$ . La integración por partes usa  $\int \nabla g \cdot v f = - \int g \nabla \cdot (fv) = 0$  para  $g = \nabla u \cdot v$ . (c) Si  $D^2u \geq 0$  el integrando es no negativo. Para el recíproco, tomar  $v$  soportado en una bola pequeña alrededor de  $x_0$ , de divergencia nula respecto de  $f$  (por ejemplo  $v = f^{-1} \nabla^\perp \eta$  en  $d = 2$ , o un rotor en general) y aproximadamente paralelo al autovector negativo.

**Ejercicio 9.8** (Costos cóncavos de la distancia). (a)  $c$  es Lipschitz en cada variable con constante 1 (la derivada de  $\arctan$  está acotada por 1), y las  $c$ -transformadas heredan el módulo de continuidad de  $c$  (Ejercicio 6.2). (b) Sobre los soportes,  $c$  es  $C^1$  fuera de la diagonal, que no toca  $\text{sop } \mu \times \text{sop } \nu$ ; la relación  $\nabla \varphi(x) = \nabla_x c(x, y)$  en el soporte del plan óptimo determina  $y$  a partir de  $x$  porque  $h(r) = \arctan r$  es estrictamente creciente con  $h'$  inyectiva en  $(0, \infty)$  (invertir  $\nabla_x c = h'(|x - y|) \frac{x - y}{|x - y|}$ ). (c) En la diagonal  $c$  no es diferenciable y la inversión de  $\nabla_x c$  falla; si los soportes se solapan, la masa común conviene dejarla quieta (Ejercicio 8.5(c)) y la estructura de mapa se pierde en general.

## CAPÍTULO 10. DISTANCIAS DE WASSERSTEIN

**Ejercicio 10.1** (Momentos).  $W_p^p(\mu, \delta_0) = \int |x|^p d\mu$  (el único plan es el producto). Las cuatro primeras tienen todos los momentos; la de Cauchy no tiene momento de orden 1; la última tiene  $\sum \frac{6}{\pi^2} n^{p-2} < \infty$  sólo para  $p < 1$ , así que no está en  $\mathcal{P}_1$ . La inclusión  $\mathcal{P}_p \subset \mathcal{P}_q$  es Hölder (o  $|x|^q \leq 1 + |x|^p$ ); es estricta por medidas como  $\sum c n^{-1-p} \delta_n$ .

**Ejercicio 10.2** (Escala, traslaciones y medias). (a) Si  $\pi \in \Pi(\mu, \nu)$  entonces  $(D_\lambda \times D_\lambda)_{\#}\pi \in \Pi(D_{\lambda\#}\mu, D_{\lambda\#}\nu)$  y el costo se multiplica por  $\lambda^p$ ; la biyección entre planes da la igualdad. Igual para isometrías. (b)  $|m_\mu - m_\nu| = |\int (x - y) d\pi| \leq \int |x - y| d\pi \leq (\int |x - y|^p d\pi)^{1/p}$ . (c) La traslación

da un plan de costo  $|v|$ , y (b) da la cota inferior. Para  $p \neq 2$  no hay descomposición exacta: con  $\mu = \delta_0$  y  $\nu = \frac{1}{2}(\delta_0 + \delta_2)$  se tiene  $W_1(\mu, \nu) = 1$ , mientras que  $|m_\mu - m_\nu| = 1$  y  $W_1(\tilde{\mu}, \tilde{\nu}) = 1$ .

**Ejercicio 10.3** (Masa que se escapa). (a) Contra  $\phi \in C_b$ ,  $|\int \phi d\mu_n - \phi(0)| \leq \frac{2}{n}\|\phi\|_\infty$ ; el único plan hacia  $\delta_0$  es el producto. (b)  $a_n = n^{1/r}$  con  $q < r < p$ . Deja de converger el momento de orden  $p$ :  $\int |x|^p d\mu_n = n^{p/r-1} \rightarrow \infty$ . (c)  $\mu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_{n^{1/p}}$  está en la bola, converge débilmente a  $\delta_0$ , pero  $W_p(\mu_n, \delta_0) = 1$ : si una subsucesión convergiera en  $W_p$ , su límite sería  $\delta_0$  (la convergencia en  $W_p$  implica la débil), contradicción.

**Ejercicio 10.4** (Convergencia contra funciones Lipschitz). La cota de  $\int |x| d\mu_n$  da ajuste:  $\mu_n(\{|x| > R\}) \leq C/R$ . Dada  $g \in C_b$  y  $\varepsilon > 0$ , elegir  $R$  con masa  $< \varepsilon$  fuera de  $B_R$  para todas las  $\mu_n$  y  $\mu$ , y una Lipschitz acotada  $g_k$  con  $|g - g_k| < \varepsilon$  en  $B_R$  y  $\|g_k\|_\infty \leq \|g\|_\infty$ ; entonces  $|\int g d(\mu_n - \mu)| \leq |\int g_k d(\mu_n - \mu)| + 4\varepsilon\|g\|_\infty + 2\varepsilon$ . La recíproca es falsa: en el Ejercicio 10.3,  $\mu_n \rightarrow \delta_0$  pero  $\int |x| d\mu_n \not\rightarrow 0$ .

**Ejercicio 10.5** (La fórmula del área vía Kantorovich–Rubinstein). (a)  $\int \varphi d\mu = \int \varphi'(x) \mu((x, \infty)) dx$  por Fubini (escribir  $\varphi(y) = \varphi(-\infty) + \int_{-\infty}^y \varphi'$ ) y  $\mu((x, \infty)) = 1 - F_\mu(x)$ . Entonces  $|\int \varphi d(\mu - \nu)| \leq \int |F_\mu - F_\nu|$  porque  $|\varphi'| \leq 1$ . (b) Con  $\varphi' = -\text{sgn}(F_\mu - F_\nu)$  la desigualdad es igualdad;  $\varphi$  es 1-Lipschitz aunque no sea  $C^1$  (la fórmula de (a) vale para Lipschitz por aproximación). (c)  $\mu = \delta_0$ ,  $\nu = \frac{1}{2}(\delta_0 + \delta_2)$ :  $W_p^p = \frac{1}{2}2^p$  pero  $\int |F_\mu - F_\nu|^p = 2 \cdot 2^{-p}$ ; sólo coinciden en  $p = 1$ . La fórmula correcta para  $p > 1$  es en la variable de los cuantiles (Proposición 8.4.5).

**Ejercicio 10.6** (Transporte en el rectángulo). (a) Cada densidad integra  $\frac{2}{9} \cdot 3 + \frac{1}{9} \cdot 3 = 1$  (áreas 3 y 3). (b)  $\int (-x) d(\mu_0 - \mu_1) = \int x (\rho_1 - \rho_0)$ :  $\rho_1 - \rho_0$  vale  $-\frac{1}{9}$  en  $\{x \leq 1\}$ , 0 en la franja central y  $\frac{1}{9}$  en  $\{x \geq 2\}$ ; el cálculo da  $\frac{1}{9}(\frac{5}{2} \cdot 2 - \frac{1}{2} \cdot 2) = \frac{4}{9}$ . (c) En cada horizontal  $y$  fija, la distribución de  $\rho_1(\cdot, y)$  está a la derecha de la de  $\rho_0(\cdot, y)$  (comparar las funciones de distribución en  $x$ ), así que el plan monótono unidimensional en cada horizontal mueve masa sólo hacia la derecha:  $u(x) - u(x') = x' - x = |x - x'|$  en su soporte. Saturación en el soporte más admisibilidad de  $u$  dan optimalidad (dualidad débil). (d) El plan no es único (en la franja central se puede reorganizar la masa entre horizontales sin cambiar el costo, siempre que se mueva hacia la derecha); el potencial sí está determinado, salvo constante, sobre el soporte de las medidas donde hay transporte, pero no en la franja central donde nada se mueve.

**Ejercicio 10.7** (Del disco a una densidad radial, con costo  $|x - y|$ ). (a)  $-|x| + |y| \leq |x - y|$  es la desigualdad triangular; hay igualdad si y sólo si  $y$  está en la semirrecta desde  $x$  alejándose del origen. (b) El mapa radial cumple  $|T(x)| = \phi(|x|) \geq |x|$  y  $T(x)$  está en el rayo de  $x$ : satura la desigualdad, y  $u$  es admisible; concluir por dualidad débil. El costo es  $\int (|T(x)| - |x|) d\mu$ . (c) Cualquier plan concentrado en  $\{y = tx, t \geq 1\}$  satura la misma desigualdad y por lo tanto es óptimo; hay muchos porque el problema de emparejar radios con radios de forma creciente tiene una única solución, pero el de emparejar “en cada rayo” se puede hacer de varias formas si se permite redistribuir el ángulo.

**Ejercicio 10.8** (La distancia SNCF). (a) La desigualdad triangular  $d(x, z) \leq d(x, y) + d(y, z)$  es  $P(x) + P(z) \leq P(x) + 2P(y) + P(z)$ . (b) Un plan que deja quieta  $\mu \wedge \nu$  y mueve el resto paga exactamente  $\int P d(\mu - \nu)^+ + \int P d(\mu - \nu)^- = \int P d|\mu - \nu|$ . Para la cota inferior,  $\varphi = P$  sobre el conjunto  $A$  que concentra a  $(\mu - \nu)^+$  y  $\varphi = -P$  fuera:  $|\varphi(x) - \varphi(y)| \leq P(x) + P(y) = d(x, y)$  para  $x \neq y$ , y  $\int \varphi d(\mu - \nu) = \int P d|\mu - \nu|$ .

**Ejercicio 10.9** (Distancia a un conjunto). (a) Para todo  $\nu \in \mathcal{P}(\Sigma)$  y todo plan,  $|x - y| \geq d(x, \Sigma)$  porque  $y \in \Sigma$ : cota inferior. La proyección  $p_\Sigma$  da un plan con costo exactamente  $\int d(x, \Sigma) d\mu$ . (b) Idéntico con potencias. (c) Con  $\Sigma$  finito,  $\nu \in \mathcal{P}(\Sigma)$  equivale a  $\#\text{sop } \nu \leq N$ . Para la condición de baricentro, derivar  $\sum_i \int_{V_i} |x - x_i|^2 d\mu$  respecto de  $x_i$ : los términos de borde de las celdas se cancelan porque el integrando es continuo al cruzar de una celda a la vecina, y queda  $\int_{V_i} (x - x_i) d\mu = 0$ .

**Ejercicio 10.10** (La distancia  $W_\infty$ ). (a)  $\pi \mapsto \pi\text{-ess sup } |x - y|$  es semicontinua inferiormente para la convergencia débil (si  $\pi_n \rightarrow \pi$  y  $\pi(\{|x - y| > s\}) > 0$ , entonces  $\pi_n(\{|x - y| > s\}) > 0$  para  $n$  grande, por ser un abierto), y  $\Pi(\mu, \nu)$  es compacto. La desigualdad triangular sale del

lema de pegado, como para  $W_p$ . (b)  $W_p \leq W_\infty$  es inmediato y  $W_p$  crece con  $p$  (Jensen), así que el límite existe. Para la desigualdad inversa, fijar  $s < W_\infty$ : la función  $\pi \mapsto \pi(\{|x - y| > s\})$  es semicontinua inferiormente y estrictamente positiva sobre el compacto  $\Pi(\mu, \nu)$ , luego  $\geq \delta > 0$ ; entonces  $W_p^p \geq s^p \delta$  y  $\lim_p W_p \geq s$ . (c) Por (a) y la compacidad. (d) Si  $x \in \text{sop } \mu$ , todo entorno de  $x$  tiene masa que debe ir a puntos de  $\text{sop } \nu$  a distancia  $\leq W_\infty$ . Ejemplo estricto:  $\mu = \frac{1}{2}(\delta_0 + \delta_1)$ ,  $\nu = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_1$ : los soportes coinciden ( $d_H = 0$ ) pero  $W_\infty = 1$ .

**Ejercicio 10.11** (La interpolación lineal no es geodésica). (a) Para  $s < t$ ,  $\nu_t$  tiene  $t - s$  más de masa en 1 que  $\nu_s$ ; el único plan que no es inmediato mueve masa  $t - s$  de 0 a 1, con costo cuadrático  $t - s$ . La longitud es  $\sup \sum \sqrt{t_{i+1} - t_i}$ , que para la partición uniforme vale  $\sqrt{n} \rightarrow \infty$ . (b) Dejar quieta la parte común  $\min\{1 - s, 1 - t\}\mu + \min\{s, t\}\nu$  y mover el resto, de masa  $|t - s|$ , con un plan óptimo entre  $\mu$  y  $\nu$  reescalado. Para discretas con soportes disjuntos, todo plan entre  $\nu_s$  y  $\nu_t$  debe mover al menos la masa  $|t - s|$  que cambia de un soporte al otro, a distancia  $\geq \text{dist}(\text{sop } \mu, \text{sop } \nu) > 0$ :  $W_2^2 \geq c|t - s|$  y la longitud explota como en (a). Para densidades suaves, el mapa  $T_h = \text{id} + h\xi$  con  $\nabla \cdot (f\xi) = f - g$  transporta  $\nu_s$  en una medida que difiere de  $\nu_{s+h}$  en  $O(h^2)$ , lo que da  $W_2(\nu_s, \nu_{s+h}) \leq Ch$ ; la existencia de  $\xi$  es un problema elíptico ( $\xi = f^{-1}\nabla w$  con  $\nabla \cdot \nabla w = f - g$  y condición de Neumann). (c)  $\nu_t$  es una mezcla y  $W_2^2(\cdot, \mu)$  es convexa en mezclas (Lema 12.3.1); a lo largo de la geodésica de desplazamiento no lo es (Ejercicio 12.9).

**Ejercicio 10.12** (Escala). Generar dos nubes de puntos, escalar y trasladar ambas con el mismo  $\lambda$  y  $v$ , y comparar `ot.emd2` (recordar que con `ot.dist` por defecto el costo ya es el cuadrado). La prueba a mano es el Ejercicio 10.2(a).

**Ejercicio 10.13** (Traslaciones). Escribir  $|x - y|^2 = |(x - m_\mu) - (y - m_\nu)|^2 + |m_\mu - m_\nu|^2 + 2(m_\mu - m_\nu) \cdot ((x - m_\mu) - (y - m_\nu))$  e integrar contra cualquier plan: el término cruzado se anula porque las medias de las medidas centradas son cero. Para  $\nu = (\tau_v)_\# \mu$  las centradas coinciden, así que  $W_2 = |v|$ . Para  $p \neq 2$  vale  $\leq |v|$  pero la igualdad depende de la cota de medias (Ejercicio 10.2(b)), que sí da igualdad también.

**Ejercicio 10.14** (La interpolación lineal tiene longitud infinita). Es el Ejercicio 10.11(a). Numéricamente, discretizar  $t$  en  $n$  pasos y sumar  $W_2(\nu_{t_i}, \nu_{t_{i+1}})$ : la suma crece como  $\sqrt{n}$ .

**Ejercicio 10.15** (Un tercer exponente). Tomar  $\mu_n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_{a_n}$  con  $a_n = n^{1/r}$ ,  $q < r < p$  (Ejercicio 10.3(b)); verificar numéricamente  $W_q^q = n^{q/r-1} \rightarrow 0$  y  $W_p^p = n^{p/r-1} \rightarrow \infty$ .

**Ejercicio 10.16** (Kantorovich–Rubinstein con el solver exacto). Los potenciales duales del programa lineal están determinados sólo salvo la constante  $(u + s, v - s)$  y, cuando el grafo de soporte del plan no es conexo, tienen más grados de libertad (Ejercicio 5.2); además POT devuelve un vértice dual cualquiera, no necesariamente el que proviene de una función 1-Lipschitz global. Sobre puntos comunes,  $v = -u$  vale para el par  $(\varphi, \varphi^c) = (\varphi, -\varphi)$  del Ejercicio 6.3, pero un par dual óptimo arbitrario puede no ser de esa forma.

## CAPÍTULO 11. LA DESIGUALDAD DE BRUNN–MINKOWSKI

**Ejercicio 11.1** (Casos elementales). (a) Para intervalos,  $|[a, b] + [c, d]| = (b - a) + (d - c)$ ; para compactos de  $\mathbb{R}$ , trasladar  $A$  y  $B$  para que  $\max A = 0 = \min B$ : entonces  $A \cup B \subset A + B$  y  $A \cap B = \{0\}$ . Para cajas,  $A + B$  es la caja de lados  $(b_i - a_i) + (d_i - c_i)$  y hay que probar  $\prod (\alpha_i + \beta_i)^{1/d} \geq \prod \alpha_i^{1/d} + \prod \beta_i^{1/d}$ : dividir por el miembro izquierdo y aplicar media aritmética–geométrica a  $\frac{\alpha_i}{\alpha_i + \beta_i}$  y a  $\frac{\beta_i}{\alpha_i + \beta_i}$ . (b)  $A + \lambda A = (1 + \lambda)A$  para  $A$  convexo; para compactos no convexos la igualdad puede fallar. (c)  $|A + B| = |A| + L|P(A)|$ , donde  $L$  es la longitud del segmento y  $P(A)$  la proyección del cubo sobre el hiperplano ortogonal al segmento (principio de Cavalieri); el término extra es positivo.

**Ejercicio 11.2** (La forma multiplicativa). Multiplicativa  $\Rightarrow$  aditiva:  $A'$  y  $B'$  tienen medida 1; escribir  $(1 - t)A + tB = \lambda((1 - t')A' + t'B')$  con  $\lambda = (1 - t)|A|^{1/d} + t|B|^{1/d}$  y  $t' = t|B|^{1/d}/\lambda$ , y aplicar la forma multiplicativa a  $A'$ ,  $B'$ :  $|(1 - t)A + tB| \geq \lambda^d$ . Aditiva  $\Rightarrow$  multiplicativa:  $|(1 - t)A + tB|^{1/d} \geq (1 - t)|A|^{1/d} + t|B|^{1/d} \geq |A|^{(1-t)/d}|B|^{t/d}$  por la desigualdad entre medias.

**Ejercicio 11.3** (Una identidad de la matriz de cofactores). (a)  $M_{ij}$  es, salvo signo, el determinante de la submatriz de  $D^2u$  que omite la fila  $i$  y la columna  $j$ ; al derivar respecto de  $x_i$  y sumar en  $i$ , cada término contiene derivadas terceras  $\partial_i \partial_k \partial_l u$  que aparecen dos veces con signos opuestos (intercambio de dos columnas). (b)  $\sum_{ij} \partial_i (M_{ij} \partial_j u) = \sum_{ij} M_{ij} \partial_{ij} u = \text{tr}(M D^2 u) = d \det D^2 u$  por (a). (c) Integrar (b) y usar la divergencia:  $\int_{\Omega} \det D^2 u = \frac{1}{d} \int_{\partial \Omega} (M \nabla u) \cdot n$ ; alternatively, cambio de variables  $y = \nabla u(x)$  directamente. La identidad de (b) es la que permite dar sentido débil a  $\det D^2 u$ .

**Ejercicio 11.4** (La desigualdad isoperimétrica). (a) Por Brunn–Minkowski,  $|A + \varepsilon B|^{1/d} \geq |A|^{1/d} + \varepsilon \omega_d^{1/d}$ ; elevar a la  $d$ , restar  $|A|$ , dividir por  $\varepsilon$  y hacer  $\varepsilon \rightarrow 0$ : queda  $d|A|^{(d-1)/d} \omega_d^{1/d}$  como término de primer orden. (b) Para la bola de radio  $r$ ,  $\text{Per} = d\omega_d r^{d-1}$  y el miembro derecho coincide. Para el cubo,  $|Q + \varepsilon B| = \ell^d + 2d\ell^{d-1}\varepsilon + O(\varepsilon^2)$ , y la desigualdad  $2d\ell^{d-1} \geq d\omega_d^{1/d} \ell^{d-1}$  dice  $2 \geq \omega_d^{1/d}$ . (c) Con  $d = 2$ ,  $\omega_2 = \pi$ :  $L \geq 2\sqrt{\pi \text{Área}}$ .

**Ejercicio 11.5** (Convexidad por desplazamiento de la entropía). (a) Cambio de variables con  $T_t$ , que es un difeomorfismo para  $t \in [0, 1]$  porque  $DT_t = (1-t)I + tD^2u$  es definida positiva. (b) Con (a),  $\int \rho_t \log \rho_t dy = \int \rho_0 \log \rho_t(T_t) dx = \int \rho_0 \log \frac{\rho_0}{\det DT_t} dx$ . Diagonalizar  $M$ :  $\log \det((1-t)I + tM) = \sum_i \log((1-t) + t\lambda_i)$ , suma de cóncavas. (c)  $\int \rho_t^m dy = \int \rho_0^m (\det DT_t)^{1-m} dx$ ; si  $g$  es cóncava y positiva,  $g^\alpha$  es convexa para  $\alpha \leq 0$  (composición de una convexa decreciente con una cóncava) y cóncava para  $0 \leq \alpha \leq 1$ ; aplicarlo a  $g(t) = (\det DT_t)^{1/d}$  y  $\alpha = d(1-m)$ :  $\alpha \leq 0$  es  $m \geq 1$ , y  $0 \leq \alpha \leq 1$  es  $1 - \frac{1}{d} \leq m \leq 1$ .

**Ejercicio 11.6** (Una desigualdad de tipo Sobolev).  $f(T(x)) - f(x) = \int_0^1 \nabla f(T_t(x)) \cdot (T(x) - x) dt$  e integrar contra  $\mu$ . Hölder con exponentes  $p$  (para  $\nabla f \circ T_t$ , cambiando variables a  $\rho_t$ ),  $q$  (para  $|T - x|$ , que da  $W_q$ ) y el tercero absorbido por la cota  $L^r$  de  $\rho_t$ :  $\int |\nabla f(T_t)|^p \rho_0 dx = \int |\nabla f|^p \rho_t dy \leq \|\nabla f\|_{L^{p,r'}} \|\rho_t\|_{L^r}$ , y ajustar los exponentes hasta llegar a la relación del enunciado. La cota de  $\|\rho_t\|_{L^r}$  es el Ejercicio 11.5(c) con  $m = r$ :  $\int \rho_t^r$  es convexa en  $t$ , luego acotada por el máximo en los extremos.

## CAPÍTULO 12. BARICENTROS DE WASSERSTEIN

**Ejercicio 12.1** (Traslaciones y escala). (a)  $\int (Ax + b) d\mu = A m_\mu + b$ . (b) Por el Lema 10.1.4,  $W_2^2(\tau_{v\#}\alpha, \beta) = W_2^2(\alpha, \beta) + |v|^2 + 2v \cdot (m_\alpha - m_\beta)$  (o directamente, componiendo planes con la traslación); sumar con pesos  $\lambda_i$  y minimizar en la posición de la media: el mínimo se alcanza trasladando por  $\bar{v}$ , y el término extra es  $\sum \lambda_i |v_i - \bar{v}|^2$  (varianza ponderada de los  $v_i$ ). (c) Tomar  $v_i = -m_{\mu_i}$  para centrar; para la dilatación,  $W_2(D_{\lambda\#}\alpha, D_{\lambda\#}\beta) = \lambda W_2(\alpha, \beta)$  (Ejercicio 10.2) y  $D_\lambda$  es una biyección de  $\mathcal{P}_2$ .

**Ejercicio 12.2** (Dos medidas: el valor mínimo). Con  $a = W_2(\mu_1, \nu)$ ,  $b = W_2(\nu, \mu_2)$  y  $a+b \geq W = W_2(\mu_1, \mu_2)$ , el mínimo de  $(1-t)a^2 + tb^2$  sobre  $a+b \geq W$  se alcanza en  $a = tW$ ,  $b = (1-t)W$  (multiplicadores de Lagrange, o completar cuadrados) y vale  $t(1-t)W^2$ . El punto  $\mu_t$  de la geodésica cumple exactamente  $W_2(\mu_1, \mu_t) = tW$  y  $W_2(\mu_t, \mu_2) = (1-t)W$ .

**Ejercicio 12.3** (Existencia para el problema multimarginal). No vacío: el producto  $\mu_1 \otimes \cdots \otimes \mu_N$ . Ajuste:  $\gamma(K_1 \times \cdots \times K_N)^c \leq \sum_i \mu_i(K_i^c)$ . Cerrado: las condiciones de marginal se escriben con integrales de funciones continuas acotadas  $\phi(x_i)$ , que pasan al límite débil. Semicontinuidad inferior:  $c$  es continua y acotada inferiormente (Teorema 2.3.6).

**Ejercicio 12.4** (Pegado de  $N$  planes). Medibilidad:  $y \mapsto \int \Phi d\pi_1^y \otimes \cdots \otimes d\pi_N^y$  es medible para  $\Phi$  producto de indicadoras (Fubini iterado con la medibilidad de  $y \mapsto \pi_i^y(A)$  del teorema de desintegración) y luego por clases monótonas. Marginales: tomar  $\Phi(y, x_1, \dots, x_N) = \phi(y, x_i)$ ; las otras integrales valen 1 y queda  $\int (\int \phi(y, x_i) d\pi_i^y) d\mu = \int \phi d\pi_i$  por la desintegración. La marginal de  $\gamma$  en  $x_i$  es la segunda marginal de  $\pi_i$ , es decir  $\mu_i$ .

**Ejercicio 12.5** (Detalles de la Proposición 12.6.1). (a) Es el Ejercicio 8.2(a); las propiedades se conservan por combinaciones con coeficientes positivos. (b) Ejercicio 8.2(c);  $\int x^2 d\bar{\mu} = \int_0^1 G^2 dt \leq$

$\sum \lambda_i \int_0^1 (F_{\mu_i}^{[-1]})^2 dt$  por convexidad. (c)  $F_{N(m, \sigma^2)}^{[-1]}(t) = m + \sigma \Phi^{-1}(t)$ : la combinación es  $\sum \lambda_i m_i + (\sum \lambda_i \sigma_i) \Phi^{-1}(t)$ . Para uniformes en  $[a_i, b_i]$ ,  $F^{[-1]}(t) = a_i + (b_i - a_i)t$ , afín en  $t$ , y la combinación es afín: uniforme en  $[\sum \lambda_i a_i, \sum \lambda_i b_i]$ .

**Ejercicio 12.6** (Gaussianas con covarianzas que conmutan). (a) Si las  $\Sigma_i$  conmutan, también lo hacen sus raíces y  $\bar{\Sigma}$ ; entonces  $(\bar{\Sigma}^{1/2} \Sigma_i \bar{\Sigma}^{1/2})^{1/2} = \bar{\Sigma}^{1/2} \Sigma_i^{1/2}$  y la ecuación de punto fijo se reduce a  $\bar{\Sigma} = \bar{\Sigma}^{1/2} \sum \lambda_i \Sigma_i^{1/2}$ , que se cumple por definición de  $\bar{\Sigma}$ . (b)  $T_i$  es lineal con matriz simétrica definida positiva (todas conmutan), luego gradiente de una cuadrática convexa;  $T_i$  manda  $N(0, \bar{\Sigma})$  en  $N(0, \Sigma_i^{1/2} \bar{\Sigma}^{-1/2} \Sigma_i \bar{\Sigma}^{-1/2} \Sigma_i^{1/2}) = N(0, \Sigma_i)$ ; y  $\sum \lambda_i T_i = (\sum \lambda_i \Sigma_i^{1/2}) \bar{\Sigma}^{-1/2} = \text{id}$ . (c) En  $d = 1$  todo conmuta y  $\bar{\sigma} = \sum \lambda_i \sigma_i$ .

**Ejercicio 12.7** (El criterio multimarginal). (a) La marginal  $i$ -ésima de  $\gamma$  es  $(T_i)_\# \mu = \mu_i$ ;  $A \circ (T_1, \dots, T_N) = \sum \lambda_i T_i = \text{id}$ , luego  $A_\# \gamma = \mu$ . (b)  $\int c d\gamma = \int \sum_i \lambda_i |T_i(x) - \sum_k \lambda_k T_k(x)|^2 d\mu = \sum_i \lambda_i \int |T_i(x) - x|^2 d\mu = \sum_i \lambda_i W_2^2(\mu, \mu_i)$  porque cada  $T_i$  es óptimo (Brenier). Por el Teorema 12.5.1, el mínimo multimarginal es  $\min \mathcal{F} = \mathcal{F}(\mu)$  (la Proposición 12.4.1 dice que  $\mu$  es baricentro), así que  $\gamma$  es óptimo. (c) En el gráfico de  $x \mapsto (T_i(x))_i$  la combinación  $\sum \lambda_i x_i$  no es constante sino igual a  $x$ : el plan multimarginal “recuerda” el baricentro.

**Ejercicio 12.8** (Cuantización y baricentros de Diracs). (a)  $W_2^2(\delta_{x_i}, \nu) = \int |x_i - y|^2 d\nu$ , así que  $\mathcal{F}(\nu) = \int \sum \lambda_i |x_i - y|^2 d\nu = \int (|y - \bar{x}|^2 + \sum \lambda_i |x_i - \bar{x}|^2) d\nu$ , mínimo exactamente en  $\nu = \delta_{\bar{x}}$ . (b) Descomponer  $\int |x - y_k|^2 d\mu|_k = \text{Var}(\mu|_k) + |m_k - y_k|^2$  con  $m_k$  la media de la celda; sumar con pesos  $b_k$ . La condición de optimalidad del Ejercicio 10.9(c) es la igualdad  $y_k = m_k$  en cada celda de Voronoi.

**Ejercicio 12.9** ( $W_2^2(\mu, \cdot)$  no es convexo a lo largo de geodésicas). (a) Los dos planes extremales entre  $\nu_0$  y  $\nu_1$  (horizontal y vertical) tienen el mismo costo 4, así que ambos son óptimos y hay al menos dos geodésicas (y todas sus mezclas);  $\nu_t$  es la del plan horizontal. (b) Los dos emparejamientos posibles entre  $\mu$  y  $\nu_t$  cuestan  $1 + 4t^2$  y  $1 + 4(1-t)^2$ . (c) A lo largo de una mezcla  $(1-s)\nu_0 + s\nu_1$  la función  $W_2^2(\mu, \cdot)$  sí es convexa; a lo largo de la geodésica de desplazamiento sube hasta 2 en el medio. La unicidad del baricentro (Teorema 12.3.2) usa la convexidad en mezclas más la hipótesis de densidad, no la geodésica.

**Ejercicio 12.10** (Rigidez en dimensión uno). Con masas iguales, cada  $\mu_i$  tiene cuantiles  $x_{(1)}^{(i)} \leq \dots \leq x_{(n)}^{(i)}$ ; el plan comonótono empareja los  $k$ -ésimos y da  $B_\# \gamma = \frac{1}{n} \sum_k \delta_{\sum_i \lambda_i x_{(k)}^{(i)}}$ , que es exactamente el promedio de cuantiles (Proposición 12.6.1). La optimalidad sale del Ejercicio 7.4 aplicado a pares, o comparando con el valor del LP para  $n \leq 4$  (el LP tiene  $n^N$  variables).

**Ejercicio 12.11** (Traslaciones). Es el Ejercicio 12.1(b); numéricamente, comparar el baricentro de las medidas trasladadas con el trasladado del baricentro y el valor de  $\mathcal{F}$  antes y después.

**Ejercicio 12.12** (Convexidad del funcional). La primera curva es convexa por el Lema 12.3.1; para la segunda, calcular  $\nu_s$  con la fórmula de la geodésica entre gaussianas (o con `ot` sobre muestras) y observar que  $s \mapsto \mathcal{F}(\nu_s)$  puede tener un máximo interior, como en el Ejercicio 12.9. Elegir  $\nu_0, \nu_1$  con covarianzas que no conmuten para que el efecto sea visible.

**Ejercicio 12.13** (Baricentros de formas). Discretizar cada figura como medida uniforme sobre los puntos de una grilla que caen dentro; `ot.lp.barycenter` resuelve el programa lineal multimarginal sobre la grilla soporte común (Teorema 12.5.1). Con  $n$  puntos por figura el LP tiene  $n^N$  variables: probar  $n \approx 100$  para  $N = 2$  y ver cómo crece el tiempo.

## CAPÍTULO 13. TRANSPORTE ENTRÓPICO Y EL ALGORITMO DE SINKHORN

**Ejercicio 13.1** (Propiedades de la entropía relativa). (a)  $\log \frac{\pi_{ij}}{a_i b_j} = \log \pi_{ij} - \log a_i - \log b_j$ ; sumar usando las marginales. (b) Sea  $A = \{(i, j) : \pi_{ij} > a_i b_j\}$ ,  $t = \pi(A)$ ,  $p = (a \otimes b)(A)$ : por convexidad de  $(s, r) \mapsto s \log \frac{s}{r}$  (o por la desigualdad de log-suma),  $\text{KL}(\pi | a \otimes b) \geq t \log \frac{t}{p} + (1-t) \log \frac{1-t}{1-p}$ , y

$\|\pi - a \otimes b\|_1 = 2(t-p)$ . La función  $g(t) = t \log \frac{t}{p} + (1-t) \log \frac{1-t}{1-p} - 2(t-p)^2$  cumple  $g(p) = g'(p) = 0$  y  $g''(t) = \frac{1}{t(1-t)} - 4 \geq 0$ . (c) De la Proposición 13.4.1,  $\varepsilon \text{KL}(\pi^\varepsilon | a \otimes b) \leq \langle C, a \otimes b \rangle - \langle C, \pi^\varepsilon \rangle \leq 2\|C\|_\infty$ ; combinar con (b). (d)  $\text{KL} = \sum_i \frac{1}{n} \log \frac{1/n}{1/n^2} = \log n$ .

**Ejercicio 13.2** (El valor entrópico como función de  $\varepsilon$ ). (a)  $\text{OT}_\varepsilon = \min_\pi \{\langle C, \pi \rangle + \varepsilon \text{KL}(\pi)\}$  es un mínimo de funciones afines crecientes de  $\varepsilon$  (porque  $\text{KL} \geq 0$ ): cóncava y no decreciente. La derivada es el teorema de la envolvente (Danskin): en los puntos de derivabilidad,  $\frac{d}{d\varepsilon} \text{OT}_\varepsilon = \text{KL}(\pi^\varepsilon)$  porque el minimizador es único. (b) La derivada de una cóncava es no creciente; y  $\langle C, \pi^\varepsilon \rangle = \text{OT}_\varepsilon - \varepsilon \text{KL}(\pi^\varepsilon)$  es no decreciente porque para  $\varepsilon < \varepsilon'$  la optimalidad de cada plan para su  $\varepsilon$  da  $\langle C, \pi^\varepsilon \rangle + \varepsilon \text{KL}(\pi^\varepsilon) \leq \langle C, \pi^{\varepsilon'} \rangle + \varepsilon \text{KL}(\pi^{\varepsilon'})$  y la simétrica; restar. (c) Proposición 13.4.1 y Teorema 13.4.2; la monotonía es (a).

**Ejercicio 13.3** (El soft-min). (a) Sacar  $e^{s/\varepsilon}$  del logaritmo; la monotonía es la del logaritmo y la exponencial. (b)  $(\varphi)_{ij}^{c,\varepsilon} = -\varepsilon \log \sum_i a_i e^{(\varphi_i - c_{ij})/\varepsilon}$  es menos un log-suma-exp, que es convexa y  $C^\infty$ ; su gradiente es el vector de pesos normalizados  $(a_i e^{(\varphi_i - c_{ij})/\varepsilon} / Z_j)_i$ , que suma 1. (c) Un gradiente con norma  $\ell^1$  igual a 1 da la constante de Lipschitz 1 para  $\|\cdot\|_\infty$ . Para  $\varepsilon \rightarrow \infty$ , desarrollar  $e^{x/\varepsilon} = 1 + x/\varepsilon + O(\varepsilon^{-2})$  dentro del logaritmo y usar  $\sum a_i = 1$ .

**Ejercicio 13.4** (Unicidad de los potenciales entrópicos). (a) Ambos planes tienen la forma  $a_i b_j e^{(\cdot)/\varepsilon}$  y están en  $\Pi(a, b)$ ; el Teorema 13.2.1 dice que esa forma caracteriza al único minimizador. Tomar logaritmos entrada a entrada. (b) De  $\varphi_i + \psi_j = \varphi'_i + \psi'_j$  para *todo*  $(i, j)$  (no sólo en un soporte), fijar  $j$ :  $\varphi_i - \varphi'_i$  no depende de  $i$ . Para  $\varepsilon = 0$  la igualdad sólo vale en el soporte del plan, que puede ser desconexo; el plan entrópico tiene todas sus entradas positivas. (c) Por simetría,  $(\psi^\varepsilon, \varphi^\varepsilon)$  es también un par de potenciales; por (b),  $\psi^\varepsilon = \varphi^\varepsilon + s$ , y con la normalización  $\sum \varphi = \sum \psi$  se obtiene  $s = 0$ .

**Ejercicio 13.5** (Los potenciales entrópicos convergen a potenciales de Kantorovich). (a) Las  $c$ -transformadas de funciones  $\varphi$  con  $\varphi_1 = 0$  están acotadas por  $\|C\|_\infty$  (hasta la constante que fija  $\varphi_1 = 0$ ):  $\varphi_j^c \in [\min_i c_{ij} - \max \varphi, c_{1j}]$  y, como  $\varphi = (\psi)^{c,\varepsilon}$  también, las cotas se cierran. El Lema 13.2.4 controla la diferencia entre la versión suavizada y la exacta. (b) Pasar al límite en  $\psi^\varepsilon = (\varphi^\varepsilon)^{c,\varepsilon}$  usando el Lema 13.2.4 y la continuidad de la  $c$ -transformada; la saturación sobre  $\text{sop } \pi^0$  sale de  $\pi^\varepsilon \rightarrow \pi^0$  y de  $\pi_{ij}^\varepsilon = a_i b_j e^{(\varphi_i + \psi_j - c_{ij})/\varepsilon}$ : donde  $\pi_{ij}^0 > 0$  el exponente debe tender a 0. Un par ajustado y saturado sobre un plan óptimo es óptimo (holgura complementaria). (c) En el  $3 \times 3$  el límite existe y selecciona un punto interior del segmento, determinado por un problema de “segunda orden” (Cominetti–San Martín); no es en general un extremo.

**Ejercicio 13.6** (Sinkhorn en un paso). (a)  $a_i b_j e^{(\varphi_i + \psi_j - f_i - g_j)/\varepsilon} = a_i b_j$  con  $\varphi = f$ ,  $\psi = g$ , que ya está en  $\Pi(a, b)$ ; el Teorema 13.2.1 da  $\pi^\varepsilon = a \otimes b$ . (b) Si  $\pi^\varepsilon = a \otimes b$ , tomar logaritmos en la fórmula de  $\pi^\varepsilon$ :  $c_{ij} = \varphi_i + \psi_j$ . Rango uno es  $\pi_{ij} = u_i v_j$ ; normalizando marginales,  $\pi = a \otimes b$ . (c) Con  $\pi_{11} = p$  las otras entradas quedan determinadas, y la fórmula del plan da  $\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = e^{(c_{12}+c_{21}-c_{11}-c_{22})/\varepsilon} =: \kappa$ : una ecuación cuadrática en  $p$ . Si  $\varepsilon \rightarrow 0$ ,  $\kappa \rightarrow 0$  o  $\infty$  y  $p$  tiende a un extremo del segmento (el vértice óptimo); si  $\varepsilon \rightarrow \infty$ ,  $\kappa \rightarrow 1$  y  $p \rightarrow a_1 b_1$ .

**Ejercicio 13.7** (Baricentros entrópicos). (a)  $\text{OT}_\varepsilon(a, b) = \min_\pi \Phi(a, \pi)$  con  $\Phi$  conjuntamente convexa (demostración de la Proposición 13.5.2:  $(p, q) \mapsto p \log \frac{p}{q}$  es la perspectiva de  $p \log p$ , Ejercicio 13.5.3). Si  $\text{OT}_\varepsilon(\frac{a+a'}{2}, b) = \frac{1}{2}(\text{OT}_\varepsilon(a, b) + \text{OT}_\varepsilon(a', b))$ , la igualdad en la convexidad de  $\Phi$  fuerza  $\pi^\varepsilon(a)_{ij}/a_i = \pi^\varepsilon(a')_{ij}/a'_i$  para todo  $(i, j)$  con  $b_j > 0$ ; sumando en  $j$  da  $a = a'$ . (b) Para  $a$  fijo, minimizar  $\text{KL}(\tilde{\pi} | \pi^k)$  sujeto a  $\tilde{\pi} \mathbf{1} = a$  da  $\tilde{\pi} = \text{diag}(a/m^k) \pi^k$  (multiplicadores de Lagrange fila a fila) con valor  $\text{KL}(a | m^k)$ ; luego minimizar  $\sum \lambda_k \text{KL}(a | m^k)$  en  $a \in \Sigma_n$ : la condición de primer orden es  $\log a_i = \sum \lambda_k \log m_i^k + \text{cte}$ , la media geométrica. La aritmética aparecería si se minimizara  $\sum \lambda_k \text{KL}(m^k | a)$ , con los argumentos intercambiados. (c) Unicidad por (a). Simetría: el problema es invariante por las reflexiones respecto de ambos ejes y el minimizador es único, luego invariante. Para  $\varepsilon \rightarrow 0$  los baricentros entrópicos convergen (compacidad) a un baricentro exacto simétrico, que es la combinación uniforme de los cuatro candidatos.

**Ejercicio 13.8** (Convergencia de valores). Calcular  $\text{OT}_\varepsilon$  como  $\langle C, \pi^\varepsilon \rangle + \varepsilon \text{KL}(\pi^\varepsilon | a \otimes b)$  (no con el valor que devuelve `ot.sinkhorn2`) y usar `ot.emd2` para  $\text{OT}$ . Si el plan óptimo del LP es único, la diferencia es  $\varepsilon \text{KL}(\pi^0) + o(\varepsilon)$ , es decir, lineal con constante  $\text{KL}(\pi^0 | a \otimes b) \leq \log(1/\min a_i)$ ; en escala log-log la pendiente tiende a 1.

**Ejercicio 13.9** (Sinkhorn y la mejora de potenciales). Sin regularización, la iteración  $\varphi \mapsto \varphi^{\text{c}\bar{c}}$  se detiene en un paso (Ejercicio 6.5) y no llega al óptimo; Sinkhorn con  $\varepsilon$  pequeño converge (Teorema 13.3.2), aunque cada vez más lento, hacia potenciales entrópicos que aproximan potenciales de Kantorovich (Ejercicio 13.5).

**Ejercicio 13.10** (Costo de cada iteración). Una iteración son dos productos matriz–vector con  $K^\varepsilon \in \mathbb{R}^{n \times m}$ :  $O(nm)$ . El solver exacto (red de transporte) escala peor, típicamente entre  $O(n^2)$  y  $O(n^3)$  en la práctica; medir con `time.perf_counter` y graficar en log-log.

**Ejercicio 13.11** (Un baricentro con  $\varepsilon \rightarrow 0$ ). Usar la versión en el dominio logarítmico para  $\varepsilon$  pequeño (la versión  $(u, v)$  produce `NaN`). La distancia al baricentro exacto decrece con  $\varepsilon$  pero el número de iteraciones para la misma tolerancia crece aproximadamente como  $1/\varepsilon$ .

**Ejercicio 13.12** (Unicidad del baricentro entrópico). Es el Ejercicio 13.7(c): la salida del algoritmo debe ser simétrica respecto de ambos ejes y repartir la masa entre los cuatro candidatos (con algo de difusión entrópica); comparar con el LP exacto, que devuelve un vértice, es decir, uno solo de los baricentros.

## APÉNDICE A. COMPLEMENTOS DE TEORÍA DE LA MEDIDA

**Ejercicio A.1** (Desintegraciones explícitas). (a)  $\pi_x = \nu$  para todo  $x$ . (b) La marginal es  $2x \mathcal{L}^1|_{(0,1)}$  y  $\pi_x = \frac{1}{x} \mathcal{L}^1|_{(0,x)}$  (uniforme en  $(0, x)$ ). (c)  $\pi_x = \delta_{x^2}$ . (d)  $\pi_x = \frac{1}{2}(\delta_x + \delta_{1-x})$ : en cada  $x$  la masa se divide en dos destinos. Verificar en cada caso  $\int \Phi d\pi = \int (\int \Phi(x, y) d\pi_x(y)) d\mu(x)$  para  $\Phi = \mathbf{1}_{A \times B}$ .

**Ejercicio A.2** (Independencia y planes inducidos). (a) Por unicidad de la desintegración:  $\nu$  (constante en  $x$ ) desintegra al producto. (b) Si  $\pi = (\text{id}, T)_\# \mu$ , entonces  $\{\delta_{T(x)}\}$  desintegra a  $\pi$ ; recíprocamente, si  $\pi_x = \delta_{T(x)}$ , la función  $T$  es medible (es  $x \mapsto \int y d\pi_x$ ) y  $\pi = (\text{id}, T)_\# \mu$ . (c) Para cada  $x$ ,  $\int |x - y|^2 d\pi_x(y) = |x - T(x)|^2 + \text{Var}(\pi_x) \geq |x - T(x)|^2$ , con igualdad si y sólo si  $\pi_x$  es una Dirac. No demuestra que los óptimos sean mapas porque en general  $T_\# \mu \neq \nu$ : promediar destinos cambia la marginal de llegada.

**Ejercicio A.3** (Asociatividad del pegado). (a) Ambas construcciones dan la medida cuya desintegración respecto de  $x_2$  es  $\pi_{12}^{x_2} \otimes \delta_{x_2} \otimes \pi_{23}^{x_2}$  pegada con la de  $\pi_{34}$  respecto de  $x_3$ ; probar la igualdad sobre funciones producto  $\phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\phi_4(x_4)$  escribiendo ambas como integrales iteradas, y concluir por clases monótonas. (b) Proyectar la medida pegada a  $(x_1, x_4)$ : es un plan entre  $\mu_1$  y  $\mu_4$ , y  $|x_1 - x_4| \leq |x_1 - x_2| + |x_2 - x_3| + |x_3 - x_4|$  más la desigualdad de Minkowski en  $L^p$  de la medida pegada. (c) Las desintegraciones son Diracs y el pegado compone los mapas.

**Ejercicio A.4** (Desintegración y transporte condicional). (a) Para  $\phi(x, y)$ ,  $\int \phi \circ T d\mu = \int (\int \phi(T^1(x), T_x^2(y)) d\mu_x(y)) d\mu^1(x)$ , y luego  $T_\# \mu^1 = \nu^1$  permite reescribirlo como  $\int \phi d\nu$ . (b)  $T$  es un mapa admisible; su costo es  $\int |x - T^1(x)|^2 d\mu^1 + \int (\int |y - T_x^2(y)|^2 d\mu_x(y)) d\mu^1$ , y cada término es el mínimo del problema correspondiente. Es estricta cuando el mapa óptimo no es triangular, por ejemplo entre dos gaussianas con covarianzas no diagonales (el mapa de Brenier es lineal simétrico, el de Knothe es triangular). (c) Con productos, el plan producto de los dos planes óptimos es admisible y da  $\leq$ ; para  $\geq$ , todo plan  $\gamma \in \Pi(\mu, \nu)$  tiene proyecciones en  $\Pi(\mu^1, \nu^1)$  y en  $\Pi(\alpha, \beta)$ , y el costo se separa en dos sumandos.

## APÉNDICE B. EL TEOREMA DE FENCHEL–ROCKAFELLAR

**Ejercicio B.1** (Supremos de funciones convexas y semicontinuas). El epigrafo de un supremo es la intersección de los epigrafos, y la intersección de convexos (resp. cerrados) es convexa (resp.

cerrada).  $F^*(y) = \sup_x \{\langle y, x \rangle - F(x)\}$  es un supremo de funciones afines continuas en  $y$  (para la topología débil-\*, porque cada  $y \mapsto \langle y, x \rangle$  lo es).

**Ejercicio B.2** (Ejemplos de transformadas). (a) Maximizar  $xy - x^p/p$  en  $x \geq 0$ :  $x = y^{1/(p-1)}$  para  $y \geq 0$ , valor  $y^{p'}/p'$ ; para  $y < 0$  el máximo es 0 en  $x = 0$ . (b)  $f_\infty^*(y) = \max\{y, 0\}$ , que es el límite de  $y^{p'}/p'$  cuando  $p' \rightarrow 1$  (para  $y \geq 0$ ). (c)  $|x|^* = \chi_{[-1,1]}$ ;  $(e^x)^*(y) = y \log y - y$  para  $y > 0$ , 0 en  $y = 0$ ,  $+\infty$  para  $y < 0$ ;  $(-\log x)^*(y) = -1 - \log(-y)$  para  $y < 0$  y  $+\infty$  si no;  $(x \log x)^*(y) = e^{y-1}$ . Para  $f^{**} = f$  usar el Ejercicio B.4: todas son convexas y semicontinuas inferiormente.

**Ejercicio B.3** (Mínimos locales y globales). Si  $x_0$  es mínimo local y  $f(x_1) < f(x_0)$ , los puntos  $(1-t)x_0 + tx_1$  con  $t$  pequeño tienen valor  $\leq (1-t)f(x_0) + tf(x_1) < f(x_0)$  y están tan cerca de  $x_0$  como se quiera. El conjunto de minimizadores es un conjunto de nivel de una convexa. Ejemplos:  $f(x, y) = x^2$  (minimizadores: el eje  $y$ );  $f(x) = e^x$  en  $\mathbb{R}$ .

**Ejercicio B.4** (Biconjugada). (a)  $F^{**}(x) = \sup_y \{\langle y, x \rangle - F^*(y)\}$  y  $F^*(y) \geq \langle y, x \rangle - F(x)$  dan  $F^{**} \leq F$ ;  $F^{**}$  es convexa sci por el Ejercicio B.1; si  $G \leq F$  es convexa sci entonces  $G^* \geq F^*$  y  $G = G^{**} \leq F^{**}$  (usando (b) para  $G$ ). (b) Si  $F^{**}(x_0) < F(x_0)$ , un hiperplano que separe estrictamente  $(x_0, F^{**}(x_0))$  del epigrafo de  $F$  da una función afín  $\ell \leq F$  con  $\ell(x_0) > F^{**}(x_0)$ , pero toda afín continua por debajo de  $F$  está por debajo de  $F^{**}$  (es de la forma  $\langle y, \cdot \rangle - \beta$  con  $\beta \geq F^*(y)$ ). El caso de hiperplanos “verticales” se trata como en el Teorema B.1.6, usando que  $F$  es propia. (c) La envolvente convexa:  $F^{**}(x) = (|x| - 1)^+$ .

**Ejercicio B.5** (El caso de dimensión finita). (a) La primera igualdad es la definición; la segunda es la caracterización de un convexo cerrado como intersección de los semiespacios que lo contienen (Hahn–Banach en dimensión finita). (b)  $\inf_x \{\chi_K(x) + c \cdot x\} = \min_K c \cdot x$  es el primal; el dual de Fenchel–Rockafellar es  $\sup_y \{-\chi_K^*(-y) - G^*(y)\}$  con  $G^* = \chi_{\{c\}}$ , lo que deja  $-\chi_K^*(-c) = \inf_K c \cdot x$  trivialmente; para obtener el dual de programación lineal hay que separar las restricciones:  $F = \chi_{\{x \geq 0\}}$  y  $G = \chi_{\{Ax=b\}} + c \cdot x$ , y calcular  $\chi_{\{Ax=b\}}^*(y)$  en términos de  $A^T$ . La hipótesis de continuidad en un punto del dominio común es la factibilidad del primal.

## APÉNDICE C. FACTORIZACIÓN POLAR DE MATRICES INVERSIBLES

**Ejercicio C.1** (Factorizaciones polares explícitas). (a)  $AA^T = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ ; su raíz se calcula diagonalizando (autovalores  $\frac{3 \pm \sqrt{5}}{2}$ ) y  $U = S^{-1}A$ . Para  $\lambda R_\theta$ :  $S = \lambda I$ ,  $U = R_\theta$ . (b)  $S = (AA^T)^{1/2}$  es la única raíz definida positiva y  $U = S^{-1}A$  es ortogonal porque  $UU^T = S^{-1}AA^TS^{-1} = I$ ; la otra factorización se obtiene aplicando la primera a  $A^T$ :  $U' = U$  y  $S' = U^TSU$ . (c)  $AA^T = S^2$  y  $A^TA = U^TS^2U$ ; son iguales si y sólo si  $S^2$  conmuta con  $U$ , si y sólo si  $S$  conmuta con  $U$  (la raíz cuadrada es función de  $S^2$ ).

**Ejercicio C.2** (Matrices singulares). (a)  $A_k = A + \frac{1}{k}I$  es inversible para  $k$  grande (salvo finitos  $k$ );  $A_k = S_k U_k$  con  $S_k \rightarrow (AA^T)^{1/2}$  por continuidad de la raíz (Ejercicio C.3) y  $U_k$  en el compacto  $O(d)$  tiene una subsucesión convergente. (b) Si  $A = 0$ , cualquier  $U$  sirve. Si  $A$  tiene rango  $d-1$ ,  $S$  tiene núcleo de dimensión 1 generado por  $e$ , y  $U$  está determinada sobre  $e^\perp$  (por  $Se = 0$  y  $SU = A$ ) pero libre sobre  $e$ : dos opciones,  $Ue = \pm e'$  con  $e'$  el vector unitario ortogonal a  $U(e^\perp)$ .

**Ejercicio C.3** (Raíz cuadrada y monotonía). (a) Sobre un compacto de matrices, la raíz es límite uniforme de polinomios en  $M$  (Weierstrass para  $\sqrt{t}$  en  $[0, R]$ ), y los polinomios son continuos. (b) Si  $M \preceq N$  y  $M^{1/2} \not\preceq N^{1/2}$ : probar primero que  $A \preceq B$  con  $A, B \succeq 0$  implica  $\|A^{1/2}x\| \leq \|B^{1/2}x\|$  y usar el argumento estándar de Löwner (la función  $\sqrt{t}$  es operador-monótona); o bien, en  $2 \times 2$ , reducir a matrices que conmutan. Para el contraejemplo del cuadrado:  $M = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ ,  $N = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ :  $N - M \succeq 0$  pero  $N^2 - M^2 = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}$  tiene determinante negativo. (c) La fórmula es una composición de inversiones, productos y raíces cuadradas de matrices definidas positivas, todas continuas por (a).

## APÉNDICE D. SUBDIFERENCIAL DE FUNCIONES CONVEXAS

**Ejercicio D.1** (Un ejemplo completo). Para  $u = |x|$ :  $\partial u(x) = \{\operatorname{sgn} x\}$  si  $x \neq 0$  y  $[-1, 1]$  en  $0$ ;  $\bar{u} = \chi_{[-1,1]}$ ;  $\partial \bar{u}(y) = \{0\}$  en  $(-1, 1)$ ,  $[0, \infty)$  en  $y = 1$  y  $(-\infty, 0]$  en  $y = -1$ . Para  $\max\{x^2, 2|x|\}$ , los puntos de no diferenciabilidad son  $0$  y  $\pm 2$ ; su conjugada se calcula por tramos y tiene tramos afines exactamente donde  $u$  tiene esquinas. Para  $|x| + x^2$ ,  $\partial u(0) = [-1, 1]$  y  $\bar{u}(y) = \frac{1}{4}(|y| - 1)_+^2$ , que es  $C^1$  con un tramo plano en  $[-1, 1]$ : la esquina de  $u$  se convierte en una zona plana de  $\bar{u}$ .

**Ejercicio D.2** (Subdiferencial y conjugación). (a)  $y \in \partial u(x)$  equivale a  $u(x) + \bar{u}(y) = x \cdot y$  (Proposición D.3.1); como  $\bar{\bar{u}} = u$ , la misma igualdad leída al revés dice  $x \in \partial \bar{u}(y)$ . (b) Igual, pero la vuelta usa  $\bar{\bar{u}} = u$ , que requiere  $u$  convexa sci (Ejercicio B.4). (c) Es (a) y (b) juntos; en Brenier,  $\nabla u \# \mu = \nu$  y  $\nabla \bar{u} \# \nu = \mu$  con  $\nabla \bar{u} \circ \nabla u = \operatorname{id}$  c.t.p.

**Ejercicio D.3** (Monotonía y monotonía cíclica). (a) Sumar las desigualdades  $u(x_{\sigma(i)}) \geq u(x_i) + y_i \cdot (x_{\sigma(i)} - x_i)$ : los términos  $u$  se cancelan. Para  $c = |x - y|^2$ ,  $\sum |x_i - y_i|^2 \leq \sum |x_{\sigma(i)} - y_i|^2$  equivale a  $\sum x_i \cdot y_i \geq \sum x_{\sigma(i)} \cdot y_i$ . (b)  $u$  es un supremo de afines, luego convexa; la monotonía cíclica dice que toda cadena cerrada que vuelve a  $x_0$  da suma  $\leq 0$ , así que  $u(x_0) = 0$  y  $u$  es finita. Para  $(x, y) \in \Gamma$ , agregar  $(x, y)$  al final de una cadena muestra  $u(x') \geq u(x) + y \cdot (x' - x)$  para todo  $x'$ . (c) El soporte de un plan óptimo es  $c$ -cíclicamente monótono (Teorema 8.3.2); por (a) esto es la monotonía cíclica clásica, y (b) da la  $u$  convexa.

**Ejercicio D.4** (Diferenciabilidad). (a) Seguir la sugerencia: la cota de Lipschitz local acota  $|y_k|$ ; si  $y_k \rightarrow y'$ , pasar al límite en  $u(z) \geq u(x + h_k) + y_k \cdot (z - x - h_k)$  da  $y' \in \partial u(x)$ , y la desigualdad con  $\delta$  da  $y' \cdot h_k / |h_k| \geq y \cdot h_k / |h_k| + \delta$  a lo largo de la subsucesión, luego  $y' \neq y$ . (b)  $f(x) = \sqrt{|x|}$ : la única pendiente  $p$  con  $\sqrt{|y|} \geq py$  para todo  $y$  es  $p = 0$ , pero  $f$  no es derivable en  $0$ . (c) Una convexa es localmente Lipschitz, luego derivable c.t.p. (Rademacher), y donde es derivable el subdiferencial es unitario. Ejemplo denso:  $u(x) = \sum_k 2^{-k} |x - q_k|$  con  $(q_k)$  los racionales.