

Continuous Kantorovich Formulation — Exercises

Exercise 1 (Couplings in basic cases). (a) Let α, β be finite *positive* measures with *different* total masses. What is $\Pi(\alpha, \beta)$, the set of couplings?

- (b) (i) What are the couplings between δ_x and $\mathcal{U}[0, 1]$?
(ii) What are the couplings between $\frac{1}{2}(\delta_x + \delta_y)$ and $\mathcal{U}[0, 1]$?

Exercise 2 (Gaussian couplings in \mathbb{R} (zero means)). Let $\alpha = \mathcal{N}(0, \sigma_\alpha^2)$ and $\beta = \mathcal{N}(0, \sigma_\beta^2)$ with $\sigma_\alpha, \sigma_\beta \geq 0$. Consider *Gaussian* couplings $\pi = \mathcal{N}(0, \Sigma)$ on $\mathbb{R} \times \mathbb{R}$ with

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & c \\ c & \sigma_\beta^2 \end{pmatrix} \succeq 0.$$

- (a) Express the quadratic OT cost $\mathbb{E}[(X - Y)^2]$ in terms of c .
(b) Determine the constraint on c implied by $\Sigma \succeq 0$.
(c) Solve the Kantorovich problem *restricted to Gaussian couplings* by minimizing over feasible c .
(d) What is the rank of Σ as a function of $(\sigma_\alpha, \sigma_\beta, c)$, and what is the support of π in each case? Compare the optimal Gaussian coupling with the conclusion of Brenier's theorem.

Exercise 3 (Uniform \rightarrow two Diracs; two Diracs \rightarrow uniform). Fix two distinct points $x \neq y$ in \mathbb{R} and let

$$\alpha = \mathcal{U}[0, 1], \quad \beta = \frac{1}{2}\delta_x + \frac{1}{2}\delta_y.$$

Consider the quadratic cost $c(u, v) = |u - v|^2$ (the answers are the same for any $|u - v|^p$, $p \geq 1$, in 1D).

- (a) Find the W_p -optimal Monge map $T : [0, 1] \rightarrow \mathbb{R}$ that pushes α to β .
(b) Is there a Monge map $S : \mathbb{R} \rightarrow [0, 1]$ such that $S_\# \beta = \alpha$?
(c) Describe an optimal Kantorovich coupling $\pi \in \Pi(\alpha, \beta)$ between α and β .

Exercise 4 (Bottleneck cost: metric property and closed-form OT value). Let (X, \mathcal{B}) be a measurable space and let $m : X \rightarrow (0, \infty)$ be measurable (strictly positive everywhere). Define

$$c(x, y) := \begin{cases} 0, & x = y, \\ m(x) + m(y), & x \neq y. \end{cases}$$

- (a) Show that $d(x, y) := c(x, y)$ is a metric on X .
(b) For probability measures α, β on (X, \mathcal{B}) with $\int m d\alpha + \int m d\beta < \infty$, compute the Kantorovich value

$$W_c(\alpha, \beta) := \inf_{\pi \in \Pi(\alpha, \beta)} \int c(x, y) d\pi(x, y)$$

in closed form.

- (c) Give an explicit optimal plan in terms of the maximal common submeasure $\mu := \alpha \wedge \beta$.

Optional – More difficult exercises

Exercise 5 (Gaussian couplings in \mathbb{R}^d (zero means)). Let $\alpha = \mathcal{N}(0, \Sigma_\alpha)$ and $\beta = \mathcal{N}(0, \Sigma_\beta)$ with $\Sigma_\alpha, \Sigma_\beta \in \mathbb{S}_{++}^d$. Any Gaussian coupling on $\mathbb{R}^d \times \mathbb{R}^d$ has block covariance

$$\Sigma = \begin{pmatrix} \Sigma_\alpha & K \\ K^\top & \Sigma_\beta \end{pmatrix} \succeq 0.$$

- (a) Express the quadratic OT cost $\mathbb{E}\|X - Y\|^2$ in terms of K .
- (b) Determine the constraint on K implied by $\Sigma \succeq 0$ (give equivalent forms).
- (c) Solve the Kantorovich problem *restricted to Gaussian couplings* by minimizing the cost over feasible K .
- (d) What is the rank of Σ and the support of the coupling π for (i) strict feasibility and (ii) an optimizer from part (c)? Compare the optimal Gaussian coupling with Brenier's theorem.

Solution. (a) **Cost as a function of K .**

$$\mathbb{E}\|X - Y\|^2 = \mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2 - 2\mathbb{E}\langle X, Y \rangle = \text{tr}(\Sigma_\alpha) + \text{tr}(\Sigma_\beta) - 2\text{tr}(K).$$

(b) **Feasible K from $\Sigma \succeq 0$.** Since $\Sigma_\alpha \succ 0$, Schur complement gives

$$\Sigma \succeq 0 \iff \Sigma_\beta - K^\top \Sigma_\alpha^{-1} K \succeq 0.$$

Equivalently, with $M := \Sigma_\alpha^{-1/2} K \Sigma_\beta^{-1/2}$, this is $I - M^\top M \succeq 0$, i.e. $\|M\|_{\text{op}} \leq 1$. Thus the feasible K are exactly

$$K = \Sigma_\alpha^{1/2} M \Sigma_\beta^{1/2} \quad \text{with} \quad \|M\|_{\text{op}} \leq 1.$$

Another equivalent form is $K \Sigma_\beta^{-1} K^\top \preceq \Sigma_\alpha$.

(c) **Minimization over feasible K .** Using the parametrization in (b),

$$\text{tr}(K) = \text{tr}(M S), \quad S := \Sigma_\beta^{1/2} \Sigma_\alpha^{1/2}.$$

By von Neumann's trace inequality,

$$\max_{\|M\|_{\text{op}} \leq 1} \text{tr}(M S) = \text{tr}((S^\top S)^{1/2}) = \text{tr}\left((\Sigma_\beta^{1/2} \Sigma_\alpha \Sigma_\beta^{1/2})^{1/2}\right),$$

achieved for $M = UV^\top$ when $S = U \text{diag}(s) V^\top$ is an SVD. Therefore an optimal cross-covariance is

$$K^* = \Sigma_\alpha^{1/2} (\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2})^{1/2} \Sigma_\alpha^{1/2}.$$

Plugging into (a) yields the Gaussian-restricted Kantorovich value

$$\mathbb{W}_2^2(\alpha, \beta) = \text{tr}(\Sigma_\alpha) + \text{tr}(\Sigma_\beta) - 2\text{tr}\left((\Sigma_\beta^{1/2} \Sigma_\alpha \Sigma_\beta^{1/2})^{1/2}\right).$$

This is attained by the deterministic linear map

$$T(x) = Ax, \quad A = \Sigma_\alpha^{-1/2} (\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2})^{1/2} \Sigma_\alpha^{-1/2},$$

for which $\text{Cov}(X, T(X)) = \Sigma_\alpha A^\top = \Sigma_\alpha A = K^*$.

(d) **Rank of Σ , support of π , and Brenier.** Write the joint covariance as

$$\Sigma = \begin{pmatrix} \Sigma_\alpha & K \\ K^\top & \Sigma_\beta \end{pmatrix}.$$

- *Strict feasibility:* if $\Sigma_\beta - K^\top \Sigma_\alpha^{-1} K \succ 0$ (equivalently $\|M\|_{\text{op}} < 1$), then $\Sigma \succ 0$ and $\text{rank}(\Sigma) = 2d$. The Gaussian π has a smooth density on \mathbb{R}^{2d} (full support).
- *At the optimizer $K = K^*$:* the Schur complement vanishes,

$$\Sigma_\beta - (K^*)^\top \Sigma_\alpha^{-1} K^* = 0,$$

so $\text{rank}(\Sigma) = d$. In fact (X, Y) satisfies $Y = AX$ a.s., hence

$$\text{supp}(\pi) = \{(x, Ax) : x \in \mathbb{R}^d\},$$

a d -dimensional linear subspace (the graph of T).

- *Brenier comparison:* since α is absolutely continuous, Brenier's theorem asserts that the *globally* W_2 -optimal coupling is induced by a map $T = \nabla \phi$. For Gaussians, T is linear, symmetric positive semidefinite, and equals the A above. Thus the Gaussian-restricted optimizer coincides with the Brenier map, and the joint law is concentrated on its graph.

Exercise 6 (A non-Monge optimal plan in \mathbb{R}^2). Let α be the uniform probability measure on the vertical segment

$$S_0 := \{(0, s) : s \in [0, 1]\}$$

(with respect to 1D length, normalized), and let

$$\beta := \frac{1}{2} \mathcal{U}(S_-) + \frac{1}{2} \mathcal{U}(S_+), \quad S_\pm := \{(\pm 1, t) : t \in [0, 1]\}.$$

We consider the quadratic cost $c(x, y) = \|x - y\|^2$ on $\mathbb{R}^2 \times \mathbb{R}^2$.

- (a) (**Lower bound**) Show that for any coupling $\pi \in \Pi(\alpha, \beta)$,

$$\int \|x - y\|^2 d\pi(x, y) \geq 1.$$

- (b) (**An optimal coupling**) Define π^* by the disintegration: for $X = (0, s) \sim \alpha$,

$$\pi^*(\cdot \mid X = (0, s)) = \frac{1}{2} \delta_{(-1, s)} + \frac{1}{2} \delta_{(1, s)}.$$

Check that $\pi^* \in \Pi(\alpha, \beta)$ and compute its cost. Conclude that the Kantorovich value equals 1.

- (c) (**No Monge minimizer**) Let $T : S_0 \rightarrow S_- \cup S_+$ be a measurable map with $T_\# \alpha = \beta$. Show that

$$\int \|x - T(x)\|^2 d\alpha(x) > 1,$$

hence no Monge map attains the Kantorovich minimum.

- (d) (**Structure of the optimizer and comparison with Brenier**) Describe $\text{supp}(\pi^*)$ and explain why it is not contained in the graph of any map $x \mapsto T(x)$. Comment on why this does not contradict Brenier's theorem.

Solution. (a) **Lower bound.** Write $x = (0, s)$ and $y = (\xi, t)$ with $\xi \in \{-1, +1\}$ and $s, t \in [0, 1]$. Then

$$\|x - y\|^2 = (\xi - 0)^2 + (t - s)^2 = 1 + (t - s)^2 \geq 1.$$

Integrating over any coupling π gives $\int \|x - y\|^2 d\pi \geq 1$.

- (b) **Optimal coupling π^* .** By construction, for any Borel $J \subset [0, 1]$,

$$\pi^*(S_- \times (\{-1\} \times J)) = \int_{s \in J} \frac{1}{2} ds = \frac{1}{2} |J|,$$

and similarly for S_+ . Hence $(\text{proj}_Y)_\# \pi^* = \beta$, and trivially $(\text{proj}_X)_\# \pi^* = \alpha$. Moreover, under π^* we have $t = s$ a.s., so

$$\int \|x - y\|^2 d\pi^* = \int (1 + (t - s)^2) d\pi^* = \int 1 d\pi^* = 1.$$

Combined with (a), this proves that π^* is optimal and the Kantorovich value equals 1.

(c) No Monge minimizer. Suppose $T_{\#}\alpha = \beta$ with $T(0, s) = (\xi(s), u(s))$, where $\xi(s) \in \{-1, +1\}$ and $u(s) \in [0, 1]$. Because β assigns mass $\frac{1}{2}$ to each vertical segment with *uniform* t -coordinate, we must have

$$\lambda(\{s : \xi(s) = -1\}) = \lambda(\{s : \xi(s) = +1\}) = \frac{1}{2},$$

and the pushforward of Lebesgue on $A_- := \{s : \xi(s) = -1\}$ by u is uniform on $[0, 1]$ (same for A_+). Now compute the cost:

$$\int \|x - T(x)\|^2 d\alpha(x) = \int_0^1 (1 + (u(s) - s)^2) ds = 1 + \int_0^1 (u(s) - s)^2 ds.$$

If this equalled 1, then $(u(s) - s)^2 = 0$ a.e., hence $u(s) = s$ a.e. But then the t -marginal on the left segment is the image of Lebesgue by $s \mapsto s$ restricted to A_- , i.e. Lebesgue on A_- , which is *not* uniform on $[0, 1]$ unless $A_- = [0, 1]$ (impossible since $\lambda(A_-) = \frac{1}{2}$). Therefore $\int_0^1 (u(s) - s)^2 ds > 0$, and so every Monge map has cost > 1 . Consequently, no Monge map is optimal.

(d) Support and Brenier. The optimizer π^* is supported on two graphs:

$$\text{supp}(\pi^*) = \{(0, s), (-1, s) : s \in [0, 1]\} \cup \{(0, s), (+1, s) : s \in [0, 1]\}.$$

It is not contained in the graph of any single-valued map T , since for every $x = (0, s)$ the conditional law assigns positive mass to two distinct points $(-1, s)$ and $(+1, s)$. This does not contradict Brenier's theorem: that theorem requires the source to be absolutely continuous with respect to *Lebesgue measure on \mathbb{R}^2* . Here α is supported on a 1D set (a segment), so the assumption fails and an optimal plan need not be induced by a map.

Remark (approximating the optimum with maps). Although no map attains cost 1, one can build maps with cost arbitrarily close to 1 by partitioning $[0, 1]$ into many small pieces, sending half of them to S_- and half to S_+ , and rearranging $u(s)$ inside each piece to match the uniform t -marginal; the added vertical cost can be made as small as $O(n^{-2})$ with n pieces.