Sinkhorn — Exercises

Exercise 1 (2×2 entropic OT with perturbation)**.** Let

$$C = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad a = \begin{pmatrix} p \\ 1-p \end{pmatrix}, \qquad b = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

with $p \in (0,1)$, and let $\varepsilon > 0$. Consider

$$\min_{P \in \mathbb{R}_+^{2\times 2}} \langle C, P \rangle \; + \; \varepsilon \sum_{i,j} P_{ij}(\log P_{ij} - 1) \quad \text{subject to} \quad P\mathbf{1} = a, \quad P^\top \mathbf{1} = b.$$

(1) Show that every admissible transport plan can be written as

$$P(x) = \begin{pmatrix} x & y(x) \\ z(x) & w(x) \end{pmatrix}$$

for a single scalar x, and give explicit formulas for $y(x)$, $z(x)$, $w(x)$, as well as the admissible interval $I(p)$ for x.

(2) Rewrite the objective as a scalar function $F_\varepsilon(x)$ and compute $F'_\varepsilon(x)$.

(3) Explain why the optimal $x_\varepsilon(p)$ is necessarily in the interior of $I(p)$ if $p > 0$. Optional: from the optimality condition $F'_\varepsilon(x) = 0$, give the expression of the optimal $x_\varepsilon(p)$.

(4) Optional: perform a expansion of $x_\varepsilon(p)$ as a function of $e^{-\varepsilon/2}$, assuming $p < 1/2$, shows $x_\varepsilon(p) \sim p + ce^{-2/\varepsilon}$ for some constant $c = c(p)$.

Exercise 2 (Dual of entropic-regularized linear program)**.** Consider the entropic-regularized linear program

$$\min_{P \geq 0,\ A(P)=b} \langle C, P \rangle + \varepsilon\, H(P), \qquad H(P) := \sum_i P_i(\log P_i - 1),$$

where P is a finite-dimensional nonnegative vector (think of $P \in \mathbb{R}_+^N$, or a matrix reshaped as a vector), $C \in \mathbb{R}^N$, $A : \mathbb{R}^N \to \mathbb{R}^m$ is linear, $b \in \mathbb{R}^m$, and $\varepsilon > 0$.

(1) Write the Lagrangian $\mathcal{L}(P, \lambda)$.

(2) Assuming we can swap $\inf_P$ and $\sup_\lambda$, write the dual problem.

(3) Give the primal–dual optimality relation between P and λ.

(4) Give the expression of $A(P)$ in the case of the marginal constraint of classical OT. Compute $A^\top$, and show that the dual you obtained matches the classical dual of entropic OT.

Exercise 3 (Entropic multimarginal OT with 3 marginals)**.** Let $a \in \mathbb{R}_+^{n_1}$, $b \in \mathbb{R}_+^{n_2}$, $c \in \mathbb{R}_+^{n_3}$ and $C \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. For $T \in \mathbb{R}_+^{n_1 \times n_2 \times n_3}$ define the three marginalization operators

$$\pi_1(T) := \Big(\sum_{j,k} T_{ijk}\Big)_{i=1}^{n_1}, \qquad \pi_2(T) := \Big(\sum_{i,k} T_{ijk}\Big)_{j=1}^{n_2}, \qquad \pi_3(T) := \Big(\sum_{i,j} T_{ijk}\Big)_{k=1}^{n_3}.$$

Consider the problem

$$\min_{T \geq 0} \langle C, T \rangle + \varepsilon \sum_{i,j,k} T_{ijk}(\log T_{ijk} - 1) \quad \text{s.t.} \quad \pi_1(T) = a,\ \pi_2(T) = b,\ \pi_3(T) = c.$$

(1) Write the Lagrangian using multipliers f, g, h for π_1, π_2, π_3.

(2) Assuming one can swap $\min_T$ and $\max_{f,g,h}$, write the dual.

(3) Write the Sinkhorn iterations in the log domain for the three potentials.

(4) *(Bonus)* Explain why the usual Hilbert-metric/Birkhoff argument for 2 marginals does not give a contraction here.

Exercise 4 (Gaussian example for Sinkhorn in dual form)**.** Let α and β be probability measures on $\mathbb{R}$ and let $c(x,y) = \frac{1}{2}(x-y)^2$. The entropic OT dual can be written

$$\max_{f,g} \int f(x)\, d\alpha(x) + \int g(y)\, d\beta(y) - \varepsilon \iint \exp\Big(\tfrac{f(x)+g(y)-c(x,y)}{\varepsilon}\Big)\, d\alpha(x)\, d\beta(y).$$

(1) Recall the Sinkhorn update on f (dual form).

(2) Assume $\alpha = \mathcal{N}(m_1, \sigma_1^2)$, $\beta = \mathcal{N}(m_2, \sigma_2^2)$, and $g(y)$ is quadratic, say $g(y) = \frac{1}{2}ay^2 + by + c$. Compute the integral on the right-hand side and show that the updated f is also quadratic.

(3) Give an intuitive explanation of why (for this quadratic/Gaussian setting) the optimal dual potentials remain quadratic, and what this implies for the optimal coupling $\pi^\star$.

Exercise 5 (Envelope theorem and gradients of the entropic cost)**.** Recall the envelope theorem in the finite-dimensional smooth case:

Let $V(\theta) := \min_z F(z,\theta)$ with F smooth, and for each θ assume there is a unique minimizer $z^\star(\theta)$. Then

$$\nabla_\theta V(\theta) = \nabla_\theta F\big(z^\star(\theta), \theta\big),$$

i.e. the derivative of V w.r.t. θ is the partial derivative of F at the optimizer (no derivative of $z^\star$ needed).

Let $a \in \mathbb{R}_+^m$, $b \in \mathbb{R}_+^n$ be histograms and $C \in \mathbb{R}^{m\times n}$ a cost matrix. The entropic OT cost is

$$\mathsf{W}_\varepsilon(a,b;C) := \min_{\substack{P\in\mathbb{R}_+^{m\times n}\\ P\mathbf{1}=a,\ P^\top\mathbf{1}=b}} \Big\{ \langle C, P\rangle + \varepsilon \sum_{i,j} P_{ij}(\log P_{ij} - 1) \Big\}.$$

(1) Using the envelope theorem on this *primal* problem, give the gradient $\nabla_C \mathsf{W}_\varepsilon(a,b;C)$.

(2) Assume the cost comes from positions: $x = (x_1, \dots, x_m) \subset \mathbb{R}^d$, $y = (y_1, \dots, y_n) \subset \mathbb{R}^d$, and

$$C_{ij} = \tfrac{1}{2}\|x_i - y_j\|^2.$$

Compute $\nabla_{x_1} \mathsf{W}_\varepsilon(a,b;C)$.

(3) Recall the dual of entropic OT

$$\mathsf{W}_\varepsilon(a,b;C) = \max_{f\in\mathbb{R}^m,\, g\in\mathbb{R}^n} \langle f, a\rangle + \langle g, b\rangle - \varepsilon \sum_{i,j} \exp\Big(\tfrac{f_i+g_j-C_{ij}}{\varepsilon}\Big).$$

Using again the envelope theorem, give $\nabla_a \mathsf{W}_\varepsilon(a,b;C)$.

Exercise 6 (Entropic OT, joint convexity, and semi-relaxed reformulation)**.** Let $a = (a_1, \dots, a_m) \in \Delta_m$ and $b = (b_1, \dots, b_n) \in \Delta_n$ be probability vectors and let $C \in \mathbb{R}^{m\times n}$ be a cost matrix. Consider the (balanced) entropic OT problem

$$W_\varepsilon(a,b) := \min_{\substack{P\in\mathbb{R}_+^{m\times n}\\ P\mathbf{1}_n=a,\ P^\top\mathbf{1}_m=b}} \Big\{ \langle C, P\rangle + \varepsilon\, H(P) \Big\}, \qquad H(P) := \sum_{i,j} P_{ij}(\log P_{ij} - 1).$$

It is well known (and you may take it for granted) that the dual of this problem is

$$W_\varepsilon(a,b) = \sup_{f\in\mathbb{R}^m,\, g\in\mathbb{R}^n} \left\{ \langle f, a\rangle + \langle g, b\rangle - \varepsilon \sum_{i=1}^m \sum_{j=1}^n \exp\Big(\tfrac{f_i+g_j-C_{ij}}{\varepsilon}\Big) \right\}.$$

(1) Using the dual expression above, show that *both* maps

$$a \longmapsto W_\varepsilon(a,b) \quad \text{and} \quad b \longmapsto W_\varepsilon(a,b)$$

are convex, and in fact that that $W_\varepsilon(a,b)$ is jointly convex in (a,b).

(2) Show that the outer problem

$$\min_{a\in\Delta_m} \; W_\varepsilon(a,b)$$

is equivalent to the semi-relaxed problem

$$\min_{\substack{P\geq 0 \\ P^\top \mathbf{1}_m = b}} \; \langle C, P\rangle + \varepsilon\, H(P).$$

(3) Solve this semi-relaxed problem in closed form.

Exercise 7 (Invariance and primal/dual recovery (continuous setting))**.** Let (X,μ) and (Y,ν) be measurable spaces and let α on X and β on Y be probability measures. For a measurable cost $c : X\times Y \to \mathbb{R}$ and $\varepsilon > 0$, consider the entropic optimal transport problem

$$\min_{\pi\in\Pi(\alpha,\beta)} \left\{ \int_{X\times Y} c(x,y)\, d\pi(x,y) + \varepsilon \int_{X\times Y} \left(\log \frac{d\pi}{d(\alpha\otimes\beta)}(x,y) - 1 \right) d\pi(x,y) \right\},$$

where $\Pi(\alpha,\beta)$ is the set of couplings of α and β and we assume $\pi \ll \alpha\otimes\beta$ for all admissible π.

(1) (**Invariance under potentials**) Let $u : X \to \mathbb{R}$ and $v : Y \to \mathbb{R}$ be integrable. Show that replacing the cost by

$$c'(x,y) = c(x,y) + u(x) + v(y)$$

does *not* change the optimal plan: the minimizer for c' is the same coupling as for c.

(2) (**Scaling of the cost**) For $\lambda > 0$, denote by $\pi_{c,\varepsilon}$ the optimal coupling for cost c and regularization ε, and by $\pi_{\lambda c,\varepsilon}$ the one for the scaled cost λc. Show that

$$\pi_{\lambda c,\varepsilon} = \pi_{c,\ \varepsilon/\lambda},$$

i.e. *scaling the cost by λ is the same as keeping the cost and scaling the regularization by* $1/\lambda$.

Optional – More difficult exercises

Exercise 8 (Unbalanced OT as constrained entropic problem)**.** Let $C \in \mathbb{R}^{m\times n}$, $a \in \mathbb{R}^m_+$, $b \in \mathbb{R}^n_+$, $\tau > 0$. Define, for $r \in \mathbb{R}^m_+$, $a \in \mathbb{R}^m_+$,

$$\mathrm{KL}(r \,|\, a) := \sum_{i=1}^{m} \Big(r_i \log \frac{r_i}{a_i} - r_i + a_i \Big),$$

and similarly for $\mathrm{KL}(s \,|\, b)$. Consider the unbalanced OT problem

$$\min_{P \in \mathbb{R}^{m\times n}_+} \langle C, P\rangle + \tau\, \mathrm{KL}(P\mathbf{1}_n \,|\, a) + \tau\, \mathrm{KL}(P^\top \mathbf{1}_m \,|\, b). \tag{UOT}$$

(1) Show that (UOT) is equivalent to

$$\min_{\substack{P\geq 0\\ P\mathbf{1}_n = r\\ P^\top \mathbf{1}_m = s}} \langle C, P\rangle + \tau\, \mathrm{KL}(r \,|\, a) + \tau\, \mathrm{KL}(s \,|\, b), \tag{$\star$}$$

and argue that if $(P^\star, r^\star, s^\star)$ solves $(\star)$, then $P^\star$ is an optimal *balanced* OT plan between the (relaxed) marginals $(r^\star, s^\star)$.

(2) Write the Lagrangian of $(\star)$ using dual vectors $f \in \mathbb{R}^m$, $g \in \mathbb{R}^n$ for the constraints $P\mathbf{1}_n = r$ and $P^\top \mathbf{1}_m = s$.

(3) Assuming we can swap $\inf_{P,r,s}$ and $\sup_{f,g}$, derive the dual involving only the two potentials f, g.

(4) Show that when $\tau \to +\infty$, the dual reduces to the usual OT dual.

Solution. **(1) Equivalence.** In (UOT) the terms $\mathrm{KL}(P\mathbf{1}_n|a)$ and $\mathrm{KL}(P^\top\mathbf{1}_m|b)$ depend on P only through its left and right marginals

$$r := P\mathbf{1}_n \in \mathbb{R}^m_+, \qquad s := P^\top \mathbf{1}_m \in \mathbb{R}^n_+.$$

Therefore we can make r, s explicit and write

$$\min_{P\geq 0}\big[\langle C, P\rangle + \tau\, \mathrm{KL}(P\mathbf{1}_n|a) + \tau\, \mathrm{KL}(P^\top \mathbf{1}_m|b)\big] = \min_{P\geq 0} \min_{\substack{r = P\mathbf{1}_n\\ s = P^\top \mathbf{1}_m}} \big[\langle C, P\rangle + \tau\, \mathrm{KL}(r|a) + \tau\, \mathrm{KL}(s|b)\big].$$

Swapping the two mins gives $(\star)$. Now fix any (r, s); among all $P \geq 0$ with $P\mathbf{1}_n = r$, $P^\top \mathbf{1}_m = s$ the term $\tau\, \mathrm{KL}(r|a) + \tau\, \mathrm{KL}(s|b)$ is constant, so minimizing over P with those marginals is exactly *classical* OT with cost C and marginals (r, s). Hence, at the optimum $(P^\star, r^\star, s^\star)$ of $(\star)$, $P^\star$ must solve the classical OT problem with marginals $(r^\star, s^\star)$.

(2) Lagrangian. Introduce $f \in \mathbb{R}^m$ and $g \in \mathbb{R}^n$. The Lagrangian is

$$\mathcal{L}(P, r, s; f, g) = \langle C, P\rangle + \tau\, \mathrm{KL}(r|a) + \tau\, \mathrm{KL}(s|b) + \langle f, P\mathbf{1}_n - r\rangle + \langle g, P^\top \mathbf{1}_m - s\rangle.$$

Expand the linear parts:

$$\mathcal{L} = \langle C, P\rangle + \langle f, P\mathbf{1}_n\rangle + \langle g, P^\top \mathbf{1}_m\rangle - \langle f, r\rangle - \langle g, s\rangle + \tau\, \mathrm{KL}(r|a) + \tau\, \mathrm{KL}(s|b).$$

Note that

$$\langle f, P\mathbf{1}_n\rangle + \langle g, P^\top \mathbf{1}_m\rangle = \sum_{i,j} (f_i + g_j) P_{ij}.$$

Hence

$$\mathcal{L} = \sum_{i,j} (C_{ij} + f_i + g_j) P_{ij} + \sum_i \big(\tau\, \mathrm{KL}(r_i|a_i) - f_i r_i\big) + \sum_j \big(\tau\, \mathrm{KL}(s_j|b_j) - g_j s_j\big).$$

(3) Dual. The dual function is

$$\varphi(f, g) = \inf_{P\geq 0, r\geq 0, s\geq 0} \mathcal{L}(P, r, s; f, g).$$

First minimize over $P \geq 0$:
$$\inf_{P\geq 0} \sum_{i,j} (C_{ij} + f_i + g_j) P_{ij}.$$
This is finite iff $C_{ij} + f_i + g_j \geq 0$ for all (i,j), otherwise the inf is $-\infty$. Under that condition, the inf is 0 (achieved at $P = 0$). So

$$\varphi(f,g) = \begin{cases} \inf_{r\geq 0} \sum_i \big(\tau \,\mathrm{KL}(r_i|a_i) - f_i r_i\big) + \inf_{s\geq 0} \sum_j \big(\tau \,\mathrm{KL}(s_j|b_j) - g_j s_j\big), & \text{if } C_{ij} + f_i + g_j \geq 0 \ \forall i,j, \\ -\infty, & \text{otherwise.} \end{cases}$$

The two infima decouple componentwise. For one coordinate, define for $r \geq 0$

$$\psi(r) := \tau\Big(r \log \frac{r}{a} - r + a\Big) - fr.$$

Derivative:
$$\psi'(r) = \tau \log \frac{r}{a} - f.$$
Setting to 0:
$$\tau \log \frac{r}{a} = f \quad \Longrightarrow \quad r = a e^{f/\tau}.$$
Plugging back:
$$\inf_{r\geq 0} \psi(r) = \tau\Big(a e^{f/\tau} \cdot \frac{f}{\tau} - a e^{f/\tau} + a\Big) - f a e^{f/\tau} = \tau a - \tau a e^{f/\tau}.$$
(Same cancellation as usual in KL-conjugate.) Thus

$$\inf_{r\geq 0} \big(\tau \,\mathrm{KL}(r|a) - fr\big) = \tau a - \tau a e^{f/\tau},$$

and similarly
$$\inf_{s\geq 0} \big(\tau \,\mathrm{KL}(s|b) - gs\big) = \tau b - \tau b e^{g/\tau},$$
where the products are componentwise. Summing all coordinates,

$$\varphi(f,g) = \begin{cases} \tau \sum_i a_i - \tau \sum_i a_i e^{f_i/\tau} + \tau \sum_j b_j - \tau \sum_j b_j e^{g_j/\tau}, & \text{if } C_{ij} + f_i + g_j \geq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Hence the dual problem is

$$\max_{f\in\mathbb{R}^m,\, g\in\mathbb{R}^n} \Big[\tau\langle \mathbf{1}, a\rangle - \tau \sum_i a_i e^{f_i/\tau} + \tau \langle \mathbf{1}, b\rangle - \tau \sum_j b_j e^{g_j/\tau}\Big] \quad \text{s.t.} \quad f_i + g_j \leq C_{ij} \quad \forall i,j.$$

(Equivalently, $C_{ij} + f_i + g_j \geq 0$ after a sign flip; the above is the standard OT-sign convention.)

(4) Limit $\tau \to +\infty$. Consider
$$-\tau a_i e^{f_i/\tau}, \qquad -\tau b_j e^{g_j/\tau}.$$
For fixed f_i, g_j, as $\tau \to +\infty$ we have $e^{f_i/\tau} = 1 + \frac{f_i}{\tau} + o(\tau^{-1})$, so

$$-\tau a_i e^{f_i/\tau} = -\tau a_i - a_i f_i + o(1),$$

and similarly for b_j. Thus

$$\tau \sum_i a_i - \tau \sum_i a_i e^{f_i/\tau} = -\sum_i a_i f_i + o(1), \quad \tau \sum_j b_j - \tau \sum_j b_j e^{g_j/\tau} = -\sum_j b_j g_j + o(1).$$

Therefore, up to an additive $o(1)$ independent of (f,g),

$$\varphi(f,g) \xrightarrow[\tau\to\infty]{} -\langle a, f\rangle - \langle b, g\rangle \quad \text{with} \quad f_i + g_j \leq C_{ij}.$$

Maximizing this limit subject to $f_i + g_j \leq C_{ij}$ is exactly the classical Kantorovich dual. So the unbalanced dual converges to the classical OT dual when $\tau \to +\infty$.

Exercise 9 (Unbalanced OT + entropic regularization)**.** Let $C \in \mathbb{R}^{m\times n}$, $a \in \mathbb{R}^m_+$, $b \in \mathbb{R}^n_+$, and let $\tau > 0$, $\varepsilon > 0$. Denote

$$H(P) := \sum_{i,j} P_{ij}(\log P_{ij} - 1), \qquad \mathrm{KL}(r \mid a) := \sum_i \big(r_i \log \tfrac{r_i}{a_i} - r_i + a_i\big).$$

Consider

$$\min_{P\ge 0} \langle C, P\rangle + \varepsilon\, H(P) + \tau\, \mathrm{KL}(P\mathbf{1} \mid a) + \tau\, \mathrm{KL}(P^\top \mathbf{1} \mid b).$$

(1) Compute the dual problem and show it involves only two dual potentials $f \in \mathbb{R}^m$, $g \in \mathbb{R}^n$.

(2) For fixed g, give the maximizer in f (unbalanced Sinkhorn update).

(3) Show that when $\tau \to +\infty$ the dual reduces to the classical entropic OT dual.

Solution. **(1) Dual.** Introduce $r = P\mathbf{1}$, $s = P^\top \mathbf{1}$ and write

$$\min_{\substack{P\ge 0\\ P\mathbf{1}=r\\ P^\top\mathbf{1}=s}} \langle C, P\rangle + \varepsilon H(P) + \tau \mathrm{KL}(r|a) + \tau \mathrm{KL}(s|b).$$

With Lagrange multipliers f, g,

$$\mathcal{L}(P, r, s; f, g) = \langle C, P\rangle + \varepsilon H(P) + \tau \mathrm{KL}(r|a) - \langle f, r\rangle + \tau \mathrm{KL}(s|b) - \langle g, s\rangle + \langle f, P\mathbf{1}\rangle + \langle g, P^\top \mathbf{1}\rangle.$$

Since $\langle f, P\mathbf{1}\rangle + \langle g, P^\top \mathbf{1}\rangle = \sum_{i,j}(f_i + g_j)P_{ij}$, we get

$$\mathcal{L} = \sum_{i,j}(C_{ij} + f_i + g_j)P_{ij} + \varepsilon H(P) + \tau \mathrm{KL}(r|a) - \langle f, r\rangle + \tau \mathrm{KL}(s|b) - \langle g, s\rangle.$$

Minimizing over $P \ge 0$ gives

$$\inf_{P\ge 0}\big\{\langle C + f + g, P\rangle + \varepsilon H(P)\big\} = -\varepsilon \sum_{i,j} \exp\Big(-\tfrac{C_{ij}+f_i+g_j}{\varepsilon}\Big).$$

For r and s,

$$\inf_{r\ge 0}\big(\tau \mathrm{KL}(r|a) - \langle f, r\rangle\big) = \tau\langle \mathbf{1}, a\rangle - \tau \sum_i a_i e^{f_i/\tau}, \quad \inf_{s\ge 0}\big(\tau \mathrm{KL}(s|b) - \langle g, s\rangle\big) = \tau\langle \mathbf{1}, b\rangle - \tau \sum_j b_j e^{g_j/\tau}.$$

Hence the dual is

$$\max_{f,g}\; -\varepsilon \sum_{i,j} e^{-(C_{ij}+f_i+g_j)/\varepsilon} + \tau \sum_i a_i - \tau \sum_i a_i e^{f_i/\tau} + \tau \sum_j b_j - \tau \sum_j b_j e^{g_j/\tau}.$$

(2) Update in f. Fix g. For each i,

$$\Phi_i(f_i) = -\varepsilon \sum_j e^{-(C_{ij}+f_i+g_j)/\varepsilon} + \tau a_i - \tau a_i e^{f_i/\tau}.$$

Setting $\Phi_i'(f_i) = 0$ gives

$$\sum_j e^{-(C_{ij}+f_i+g_j)/\varepsilon} = a_i e^{f_i/\tau}.$$

Let $K_{ij} = e^{-C_{ij}/\varepsilon}$, $v_j = e^{-g_j/\varepsilon}$, and $S_i = \sum_j K_{ij} v_j$. Then

$$e^{-f_i/\varepsilon} S_i = a_i e^{f_i/\tau} \quad\Longrightarrow\quad f_i = -\frac{\varepsilon\tau}{\varepsilon+\tau}\log\Big(\frac{a_i}{S_i}\Big).$$

(3) Limit $\tau \to +\infty$. As $\tau \to +\infty$,

$$\tau a_i - \tau a_i e^{f_i/\tau} \to -a_i f_i, \qquad \tau b_j - \tau b_j e^{g_j/\tau} \to -b_j g_j,$$

so the dual converges to

$$\max_{f,g}\; -\varepsilon \sum_{i,j} e^{-(C_{ij}+f_i+g_j)/\varepsilon} - \langle a, f\rangle - \langle b, g\rangle,$$

the classical entropic OT dual.

Exercise 10 (φ-divergence, dual form and unbalanced OT)**.** Let α and β be finite nonnegative measures on a measurable space X, and let $\varphi : \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ be convex, lsc, with $\varphi(1) = 0$. The φ-divergence of α w.r.t. β is

$$D_\varphi(\alpha \,|\, \beta) := \begin{cases} \displaystyle\int \varphi\Big(\frac{d\alpha}{d\beta}(x)\Big)\, d\beta(x), & \alpha \ll \beta, \\ +\infty, & \text{otherwise.} \end{cases}$$

(1) Recall the convex conjugate $\varphi^\star(t) := \sup_{u \geq 0}\big(ut - \varphi(u)\big)$.

(2) Show that

$$D_\varphi(\alpha \,|\, \beta) = \sup_f \left\{ \int f \, d\alpha - \int \varphi^\star(f)\, d\beta \right\}.$$

(3) Consider now the unbalanced OT problem

$$\min_{\pi \geq 0} \int c(x,y)\, d\pi(x,y) + D_\varphi(\pi_1 \,|\, \alpha) + D_\varphi(\pi_2 \,|\, \beta),$$

where π_1, π_2 are the first and second marginals of π. Using the dual formula in (2) for each divergence, write a dual problem involving only two test functions f, g on X.

Solution. **(1)** This is the usual Legendre–Fenchel transform.

(2) If $\alpha \ll \beta$ and $u = \frac{d\alpha}{d\beta}$, then pointwise

$$\varphi(u) = \sup_t (ut - \varphi^\star(t)).$$

Integrating and swapping sup and $\int$ gives

$$D_\varphi(\alpha \,|\, \beta) = \int \varphi(u)\, d\beta = \sup_f \int \big(uf - \varphi^\star(f)\big)\, d\beta = \sup_f \left\{ \int f\, d\alpha - \int \varphi^\star(f)\, d\beta \right\}.$$

If $\alpha \not\ll \beta$ the RHS is $+\infty$, matching the LHS.

(3) Apply (2) to each marginal term:

$$D_\varphi(\pi_1 \,|\, \alpha) = \sup_f \left\{ \int f(x)\, d\pi_1(x) - \int \varphi^\star(f(x))\, d\alpha(x) \right\},$$

$$D_\varphi(\pi_2 \,|\, \beta) = \sup_g \left\{ \int g(y)\, d\pi_2(y) - \int \varphi^\star(g(y))\, d\beta(y) \right\}.$$

Insert these in the primal:

$$\min_{\pi \geq 0} \sup_{f,g} \Big[\int c\, d\pi + \int f\, d\pi_1 - \int \varphi^\star(f)\, d\alpha + \int g\, d\pi_2 - \int \varphi^\star(g)\, d\beta \Big].$$

Use $\int f\, d\pi_1 = \int f(x)\, d\pi(x,y)$ and the same for g to get

$$\min_{\pi \geq 0} \sup_{f,g} \int \big(c(x,y) + f(x) + g(y)\big)\, d\pi(x,y) - \int \varphi^\star(f)\, d\alpha - \int \varphi^\star(g)\, d\beta.$$

Assuming we can interchange min and sup, we get the dual

$$\sup_{f,g} \left\{ -\int \varphi^\star(f)\, d\alpha - \int \varphi^\star(g)\, d\beta \;\middle|\; c(x,y) + f(x) + g(y) \geq 0 \text{ for all } x, y \right\}.$$

(Equivalently, $f(x) + g(y) \geq -c(x,y)$.) This is the general unbalanced OT dual under a generic φ-divergence, and for the special choice $\varphi(u) = u \log u - u + 1$ one recovers the KL-based dual seen earlier.

Exercise 11 (Barycentric projection from entropic OT)**.** Let $a = \sum_{i=1}^m a_i \delta_{x_i}$ and $b = \sum_{j=1}^n b_j \delta_{y_j}$ be two discrete measures in $\mathbb{R}^d$, and let P_ε be the optimal plan of entropic OT between (a, b) for some cost $C_{ij} = c(x_i, y_j)$. For x_i with $a_i > 0$ define the barycentric projection

$$T_\varepsilon(x_i) := \frac{1}{a_i} \sum_{j=1}^n (P_\varepsilon)_{ij}\, y_j.$$

(1) Suppose the unregularized OT problem (with the same cost) has a unique Monge map $T : \{x_i\} \to \{y_j\}$, and $P_\varepsilon \rightharpoonup P^\star$ where $P^\star = (\mathrm{Id}, T)_\# a$ is supported on the graph of T. Show that

$$T_\varepsilon(x_i) \longrightarrow T(x_i) \quad \text{for every } i.$$

(2) Assume now that $c(x, y) = \frac{1}{2}\|x - y\|^2$ and recall from the previous exercise (envelope theorem) that

$$\nabla_{x_i} \mathsf{W}_\varepsilon(a, b) = a_i x_i - \sum_{j=1}^n (P_\varepsilon)_{ij}\, y_j.$$

Express $T_\varepsilon(x_i)$ in terms of $\nabla_{x_i} \mathsf{W}_\varepsilon(a, b)$ and x_i, and comment on the fact that the barycentric projection is exactly the "OT force" at x_i rescaled by a_i.

Solution. **(1)** Since $P_\varepsilon \rightharpoonup P^\star$ and $P^\star$ is supported on $\{(x_i, T(x_i))\}$, we have for each i

$$\sum_j (P_\varepsilon)_{ij} \to \sum_j (P^\star)_{ij} = a_i, \qquad \sum_j (P_\varepsilon)_{ij} y_j \to \sum_j (P^\star)_{ij} y_j = a_i T(x_i).$$

Dividing by $a_i > 0$ gives $T_\varepsilon(x_i) \to T(x_i)$.

(2) From the envelope-theorem computation we have

$$\nabla_{x_i} \mathsf{W}_\varepsilon(a, b) = a_i x_i - \sum_j (P_\varepsilon)_{ij} y_j = a_i x_i - a_i T_\varepsilon(x_i),$$

hence

$$T_\varepsilon(x_i) = x_i - \frac{1}{a_i} \nabla_{x_i} \mathsf{W}_\varepsilon(a, b).$$

So the barycentric projection is obtained by moving x_i in the direction of *minus* the gradient of the entropic OT cost, with step size $1/a_i$. This makes explicit that T_ε is the OT "target point" encoded in the gradient of the cost.