MAT 1502 / Fields Academy: Assignment #2

Due: Tuesday 1 March — Note: No classes nor office hours Feb 22 and 24

NB: Recall your project topics and references are due by Tuesday Feb 15.

1. **Dependence of optimizer on transport distance exponent:** Fix a pair of compactly supported Borel probability measure $\mu^{\pm}$ given by spherically symmetric densities $d\mu^{\pm}(x) = f^{\pm}(|x|)d^n x$ on $\mathbf{R}^n$.

a) Among maps $G_{\#}\mu^+ = \mu^-$, find the map which minimizes the cost $c_p(x,y) := |x-y|^p$ for $p=2$.

b) Address the analogous question for each $p > 1$.

c) Find a pair of (not necessarily spherically symmetric) $L^1(\mathbf{R}^n)$ measures $\mu^{\pm} \in \mathcal{P}(\mathbf{R}^n)$ for which the c_p-optimal map $G_p \neq G_2$ for some $1 < p \neq 2$.

2 **Bi-twisted benefits:**

a) Given smooth compact domains $M^{\pm} \subset \mathbf{R}^n$, show a benefit $b \in C^1(M^+ \times M^-)$ satisfies the bi-twist condition (A1) if and only if for each $(x_0, y_0) \in M^+ \times M^-$, the cross-difference

$$\Delta_0(x,y) := b(x,y) + b(x_0,y_0) - b(x,y_0) - b(x_0,y)$$

has no critical points (x,y) apart from (x_0,y_0).

b) If instead $M^{\pm}$ are compact manifolds (with no boundaries), find a benefit $b \in C^1(M^+ \times M^-)$ satisfying the bi-twist condition or show no such function exists.

3 **An asymmetric dual:** Let $M^{\pm} \subset \mathbf{R}^n$ be smooth and compact domains and set $b(x,y) = x \cdot y$. Given measures $\mu^{\pm} \in \mathcal{P}(M^{\pm})$ with $L^1(\mathbf{R}^n)$ densities $d\mu^{\pm}(x) = f^{\pm}(x)d^n x$ define

$$J(u) := \int_{M^+} u(x)d\mu^+(x) + \int_{M^-} u^b(y)d\mu^-(y)$$

a) Show $J(u)$ is a convex function on $C(M^+)$.

b) Given $u, w \in C(M^+)$, show $s \in \mathbf{R} \mapsto J(u+sw)$ is differentiable and find a formula for its derivative

$$\frac{d}{ds}\bigg|_{s=0} J(u+sw).$$

c) Show $J(u)$ attains its minimum on $C(M^+)$.

d) Combine (b) with (c) to give a different proof that a convex gradient exists pushing μ^+ forward to μ^- than the one shown in class.

4. **The Ma-Trudinger-Wang regularity conditions:** Let $M_0^{\pm} \subset \mathbf{R}^n$ be non-empty convex domains and $M^{\pm}$ denote their closures. Let $h^{\pm} \in C^{\infty}(M^{\pm})$ both have Lipschitz constant less than $2^{-1/2}$ and set $b(x,y) = x \cdot y + h^+(x)h^-(y)$.

a) Show b is bi-twisted (A1) and non-degenerate (A2).

d) If $M^{\pm}$ are small enough balls, show (A4) $D_x b(x, M^-)$ and $D_y b(M^+, y)$ are convex.

c) If both $h^{\pm}$ are convex, show b satisfies (A3).

d) Find $h^{\pm}$ for which b violates (A3).

MAT1502 PS2

Jeffrey Liang

February 2022

1. (a) We claim the map that satisfies this is the following: $G(x) = (F_-^{-1} \circ F_+)(|x|) * \frac{x}{|x|}$ where $F_\pm$ are the cumulative distribution functions of $f^\pm$. As all the hypotheses of Theorem 1.1 in Guillen and McCann (2010) are satisfied, it suffices to show that G is the gradient of a convex function. It is easily verified using the Leibniz integral rule that the gradient of $F(x) = \int_0^{|x|} F_-^{-1}(F_+(r))dr$ is G. Therefore we need to show F is convex. Recall that a differentiable function F is convex iff $< \nabla F(x) - \nabla F(y), x - y > \geq 0 \, \forall x, y$. If we expand, we obtain $< \nabla F(x), x > + < \nabla F(y), y > - < \nabla F(x), y > - < \nabla F(y), x >= ||G(x)|| * ||x|| + ||G(y)|| * ||y|| - ||G(x)|| * ||y|| * cos\theta - ||G(y)|| * ||x|| * cos\theta$ since $G(x)$ points in the same direction as x. But $||G(z)||$ is a non decreasing function of $||z||$! Wlog $||x|| \geq ||y||$. Then $||G(x)|| * (||x|| - ||y|| cos\theta) + ||G(y)|| * (||y|| - ||x|| cos\theta)$ is clearly non negative as the first term is positive and large enough to cancel the second term, even if $cos\theta = 1$. QED.

(b) We claim the same map is optimal. $D_x c(x, y) = p|x - y|^{p-1} * \frac{x-y}{|x-y|}$ so for a fixed x, it is injective and it is uniformly bounded in x and y a.e. as $\mu^\pm$ are compactly supported. Then all hypotheses of Theorem 2.9 in Guillen and McCann (2010) are satisfied. Using the functional form of G, we obtain that the potential $u = \int_0^{|x|} p|(1 - \lambda(r))r|^{p-1}dr$, where $\lambda(r) := \frac{(F_-^{-1} \circ F_+)(r)}{r}$ and $G(x) = \lambda(|x|) * x$. Since $\mu^\pm$ are compactly supported, we can bound $|r - \lambda(r)r|$ by a constant. We have $\nabla u = p|(1 - \lambda(|x|))|^{p-1}|x|^{p-1} * \frac{x}{|x|}$ since we can bound this by a constant, u is Lipschitz so it follows from the theorem that G is optimal.

(c) Consider the case where μ^+ is the average of two Dirac masses supported at x, x' and similarly for μ^-. Unless $c(x, y) = c(x, y')$, the Kantorovich solution will only be supported on one of (x,y) or (x,y'). In the case p = 2: minimizing this cost is equivalent to maximizing $x \cdot y$ since $|x - y|^2 =< x - y, x - y >=< x, x > + < y, y > -2 < x, y >$. Now let us consider $n = 2$. (This example can be easily generalized to higher dimensions.) Let $x = (1, 0)$ and $x' = (2, 2)$. And $y = (100, 0), y' = (2, 2)$. For p = 2, the unique optimal map takes x to y' and x' to y with cost $1/2 * (|(98, -2)|^2 + |(1, 2)|^2) = 4,806.5$. Let us check the other possible mapping G' that takes x to y and x' to y'. In this case, we have cost $1/2 * (|(99, 0)|^2 + 0) = 4,900.5$. So G is in fact optimal. However, consider p close to 1. Since the sum of the norms of the transportation vectors of G' is less than for G ($99 < 100 < 98.02 + \sqrt{5}$). Therefore G' is instead optimal. Now of course these measures are not absolutely continuous but we can take a small ball around each of the points and consider the uniform measure on that ball. Since the cost is continuous, the balls can be made small enough so that all mass in a ball gets mapped to a single ball and these results still hold true.

2. (a) This theorem holds for any manifold without boundaries.

(=>) $D_x\Delta_0(x,y) = D_x b(x,y) - D_x b(x,y_0)$

This equals zero iff $y = y_0$ by the bi-twist condition. Similarly $D_y\Delta_0(x,y) = 0$ iff $x = x_0$. Since these must both be zero for (x,y) to be a critical point, the conclusion follows.

(<=) Suppose (A1) does not hold. Wlog $\exists x, y, y'$ s.t. $D_x b(x,y) = D_x b(x,y')$. Then if we set $x_0 := x$ and $y_0 := y$, (x_0, y') is a critical point by the above computation.

(b) No such function exists. The cross difference is a continuous function. Since $M^\pm$ are compact, it attains its maximum and minimum. b will have an extrema at some point other than (x_0, y_0) and this must be a critical point. Therefore by part (a), b is not bi-twisted.

3. (a)

$$
\begin{aligned}
J(tu+(1-t)v) &= \int (tu+(1-t)v)(x)d\mu^+(x) + \int (tu+(1-t)v)^b(y)d\mu^-(y) \\
&= t\int u(x)d\mu^+(x) + (1-t)\int v(x)d\mu^+(x) \\
&+ \int \sup_{x\in M^+}(<x,y> - tu(x) - (1-t)v(x))d\mu^-(y) \\
&\leq t\int u(x)d\mu^+(x) + (1-t)\int v(x)d\mu^+(x) \\
&+ t\int \sup_{x\in M^+}(<x,y> - u(x))d\mu^-(y) + (1-t)\int \sup_{x\in M^+}(<x,y> - v(x))d\mu^-(y) \\
&= tJ(u) + (1-t)J(v)
\end{aligned}
$$

where the inequality holds because the convex conjugate is a convex function on $C(M^+)$ as a supremum over affine functions.

(b) The derivative of the first term is easily determined to be $\int w(x)d\mu^+(x)$. As for the second term, we note that the supremum in the definition of the convex conjugate is attained because the objective is a continuous function on a compact set. The convex conjugate is convex in s as a supremum of affine functions. So it is differentiable a.e. Then by the envelope theorem, the derivative of $(u+sw)^b$ with respect to s is $-w(x^\star)$ where $x^\star$ is the maximizing argument. For any s and y, the derivative is bounded in magnitude by $\max_{x\in M^+}|w(x)|$ an integrable function over M^-. Therefore the Dominated Convergence Theorem applies and we may differentiate under the integral sign. Thus the formula we are looking for is: $\int w(x)d\mu^+(x) - \int w(x^\star(y))d\mu^-(y)$.

(c) We can take a similar approach as in the Proof of Theorem 2.9 in Guillen and McCann (2010). Since $u^{bb} \leq u$, a minimizer must be c-convex, in particular convex. We claim that the set of c-convex functions normalized to be zero at a fixed point x_0 is relatively compact. Since adding constants to a function u doesn't change the value of the functional $J(u)$, a minimizer resides in this subset if one exists. Arzela-Ascoli requires that this subset is equicontinuous and pointwise bounded. Lemma 3.1 in Guillen and McCann (2010) shows that the gradient can be uniformly bounded. This then implies both equicontinuity and pointwise boundedness. Finally we note that J is continuous as $||u-u'|| < \epsilon \ (=>) \ |J(u) - J(u')| < 2\epsilon$. Therefore, a minimizer exists.

(d) Since J is minimized at a point $u^\star$, it's variational derivative must be 0 for all w. So we have $0 = \int w(x)d\mu^+(x) - \int w(x^\star(y))d\mu^-(y)$ for all $w \in C(M^+)$. u^b is convex so it is differentiable a.e. and its gradient is precisely $x^\star$. The gradient of u^{bb} and the gradient of u^b are inverses. Therefore if we let $w(x) = g\circ\nabla u^{bb}$, ($\nabla u^{bb}$ is continuous a.e.) where g is an arbitrary continuous function on M^-, we have $\int g\circ\nabla u^{bb}(x)d\mu^+(x) = \int g\circ\nabla u^{bb}(\nabla u^b(y))d\mu^-(y) = \int g(y)d\mu^-(y)$ since these gradients are inverses. But this means that ∇u^{bb}, the gradient of a convex function, pushes forward μ^+ to μ^-.

4. (a)

$$D_x b(x,y) = y + \nabla h^+(x)h^-(y) = y' + \nabla h^+(x)h^-(y') = D_x b(x,y')$$
$$y - y' = (h^-(y') - h^-(y))\nabla h^+(x)$$

but $||(h^-(y') - h^-(y))\nabla h^+(x)|| \leq 2^{-1/2}||y - y'|| * 2^{-1/2}$ from the Lipschitz constants of $h^{\pm}$. So this cannot hold unless $||y - y'|| = 0$, i.e. $y = y'$. The analogous argument applies for the symmetric case. So b is bi-twisted.

For showing non degeneracy, we compute the matrix of interest:

$$\frac{\partial b}{\partial x_i \partial y_j} = \frac{\partial}{\partial y_j}(y_i + \frac{\partial h^+}{\partial x_i}h^-(y)) = I + \nabla h^+(\nabla h^-)^T$$

By the matrix determinant lemma, the determinant of this matrix is $1 + (\nabla h^+)^T(\nabla h^-)$ which is greater than or equal to 1/2 because of the Lipschitz bounds and the Cauchy-Schwarz inequality. In particular, it is non-zero. Therefore, b is non-degenerate.

(b) Polyak (2001) showed that the image of a small enough ball under a diffeomorphism from R^n to itself is convex. By part (a), $D_x b(x,y)$ is manifestly a diffeomorphism from M^- with its image. Therefore, the result follows.

(c) We compute:

$$\begin{aligned}
\frac{\partial^4}{\partial s^2 \partial t^2}|_{s=t=0} b(x(s), y(t)) &= \frac{\partial^4}{\partial s^2 \partial t^2}|_{s=t=0}(x(s)y(t) + h^+(x(s))h^-(y(t))) \\
&= \ddot{x}(0)\ddot{y}(0) + (\dot{x}(0)\nabla^2 h^+(x(0))\dot{x}(0) + \nabla h^+(x(0)) \cdot \ddot{x}(0)) \\
&\quad (\dot{y}(0)\nabla^2 h^-(y(0))\dot{y}(0) + \nabla h^-(y(0)) \cdot \ddot{y}(0)) \qquad (1)
\end{aligned}$$

Wlog we also have that $(x(0), y(t))$ is a b-segment, i.e.

$$\begin{aligned}
\frac{\partial^2}{\partial t^2}(y(t) + \nabla h^+(x(0)) * h^-(y(t))) &= 0 \\
\ddot{y}_t + \nabla h^+(x(0))\dot{y}(t)\nabla^2 h^-(y(t))\dot{y}(t) + \nabla h^+(x(0))\dot{y}(t) \cdot \ddot{y}(t) &= 0 \\
\nabla h^+(x(0))(\dot{y}(t)\nabla^2 h^-(y(t))\dot{y}(t) + \nabla h^-(y(t)) \cdot \ddot{y}(t)) &= -\ddot{y}(t) \qquad (2) \\
a * \nabla h^+(x(0)) &= -\ddot{y}(0) \qquad (3)
\end{aligned}$$

where $a = (\dot{y}(0)\nabla^2 h^-(y(0))\dot{y}(0) + \nabla h^-(y(0)) \cdot \ddot{y}(0))$. We need to show $a \geq 0$. Recall we also have the following:

$$\begin{aligned}
\frac{\partial^2}{\partial s \partial t}|_{s=t=0} b(x(s), y(t)) &= 0 \qquad (4) \\
\dot{x}(0)\dot{y}(0) + \nabla h^+(x(0)) \cdot \dot{x}(0) * \nabla h^-(y(0)) \cdot \dot{y}(0) &= 0 \qquad (5)
\end{aligned}$$

But $||\nabla h^+(x(0)) \cdot \dot{x}(0) * \nabla h^-(y(0)) \cdot \dot{y}(0)|| \leq 1/2 * ||\dot{x}(0)|| * ||\dot{y}(0)||$ by the Lipschitz bound so it must be that $|cos\theta| \leq 1/2$, where θ is the angle between $\dot{x}(0)$ and $\dot{y}(0)$. There is a gap in my argument here because I need to show that $a \geq 0$.

Finally, substituting equation 3 in to equation 1, we obtain:

$$\begin{aligned}\frac{\partial^4}{\partial s^2 \partial t^2}|_{s=t=0} b(x(s), y(t)) &= \ddot{x}(0)\ddot{y}(0) + a(\dot{x}(0)\nabla^2 h^+(x(0))\dot{x}(0)) + \ddot{x}(0) \cdot (-\ddot{y}(0)) \\ &= a\dot{x}(0)\nabla^2 h^+(x(0))\dot{x}(0)\end{aligned}$$

If $a \geq 0$, this quantity is non negative since h^+ is convex.

(d) Suppose $M^{\pm}$ are the closed unit ball around the origin in $R^n, n > 1$. Define $h^+(x) = \frac{1}{4\sqrt{2}} x_1^2$ and $h^-(y) = -\frac{1}{4\sqrt{2}} y_2^2$. These indeed have Lipschitz constant less than $2^{-1/2}$ since their derivative is bounded below that. Further let $x(s) = s * e_1$ and $y(t) = t * e_2$, where e_i is the ith element of the standard basis on R^n. $(x(0), y(t))$ is also a b-segment since $\ddot{y}(0) = 0$ and $\nabla h^+(x(0)) = 0$. We also have Equation 4 holding because in Equation $5, \dot{x}(0)$ and $\dot{y}(0)$ are orthogonal and $\nabla h^{\pm}(0) = 0$. However, while $\dot{x}(0)\nabla^2 h^+(x(0))\dot{x}(0) = \frac{1}{2\sqrt{2}} > 0$, $a = -\frac{1}{2\sqrt{2}} < 0$, so (A3) does not hold for b with these $h^{\pm}$.

References

[1] N. Guillen and R. McCann, "Five lectures on optimal transportation: Geometry, regularity and applications," 2010. http://www.math.toronto.edu/mccann/papers/FiveLectures.pdf

[2] B. T. Polyak, "Convexity of Nonlinear Image of a Small Ball with Applications to Optimization," Set-valued analysis, vol. 9, no. 1, pp. 159–168, 2001, doi: 10.1023/A:1011287523150.