An Invitation to Statistics in Wasserstein Space

Shen Ziliang 2021213218

Shanghai University of Finance and Economics

January 23, 2022

Outline

Content

Regularity

theorem1.6.6 (Regularity in $\mathbb{R}$)

Let $\mu, v \in P(\mathbb{R})$ possess distribution functions F and G of class $C_k, k \geq 1$. Suppose further that supp v is an interval I (possibly unbounded) and that $G^{'} > 0$ on the interior of I. Then the optimal map is of classC_k as well. If $F, G \in C^0$ are merely continuous, then so is the optimal map.

Definition ($\alpha - Hölder$)

an open $\Omega \subseteq \mathbb{R}^d, k \geq 0$ and $\alpha \in (0, 1]$. We say that $f \in C^{k,\alpha}(\omega)$ if all the partial derivatives of order k of f are locally $\alpha - Hölder$ on Ω. For example, if $k = 1$, this means that for any $x \in \Omega$, there exists a constant L and an open ball B containing x such that

$$\|\nabla f(y) - \nabla f(z)\| \leq L\|y - z\|^{\alpha}, y, z \in B$$

Regularity

theorem1.6.7 (Regularity of Transport Maps)

Fix open sets $\Omega_1, \Omega_2 \subseteq \mathbb{R}^d$, with Ω_2 convex, and absolutely continuous measures $\mu, v \in P(\mathbb{R}^d)$ with finite second moments and bounded, strictly positive densities f, g, respectively, such that $\mu(\Omega_1) = 1 = v(\Omega_2)$. Let ϕ be such that $\nabla\phi_{\#}\mu = v$.

1. If Ω_1 and Ω_2 are bounded and f, g are bounded below, then ϕ is strictly convex and of class $C^{1,\alpha}(\Omega_1)$ for some $\alpha > 0$.
2. If $\Omega_1 = \Omega_2 = \mathbb{R}^d$ and $f, g \in C^{0,\alpha}$, then $\phi \in C^{0,\alpha}(\Omega_1)$.

if in addition $f, g \in C^{k,\alpha}$, then $\phi \in C^{k+2,\alpha}(\Omega_1)$ In other words, the optimal map $T = \nabla\phi \in C^{k+1,\alpha}(\Omega_1)$ is one derivative smoother than the densities, so has the same smoothness as the measures μ, v

Content

1. Regularity of the Transport Maps
2. Stability of solutions

Weak Convergence

Lemma 1.7.1 (Portmanteau)

Let $\mathscr{X}$ be a complete separable metric space and let μ ,$\mu_n \in P(\mathscr{X})$ Then the following are equivalent:

- $\mu_n \to \mu$ weakly ;
- $F_n(x) \to F(x)$ for any continuity point x of F. Here $\mathscr{X} = \mathbb{R}^d$,$F_n(x)$ is the distribution function of μ_n and F is that of μ ;
- for any open $G \subseteq \mathscr{X}$, liminf$\mu_n(G) \geq \mu(G)$;
- for any closed $F \subseteq \mathscr{X}$, liminf$\mu_n(F) \leq \mu(F)$;
- $\int h \mathrm{d}\mu_n \to \int h \mathrm{d}\mu$ for any bounded measurable h whose set of discontinuity points is a -null set.

Stability of Optimal Transference Plans

Theorem 1.7.2 (Weak Convergence and Optimal Plans)

Let μ_n and v_n converge weakly to μ and v , respectively, in $P(\mathscr{X})$ and let $c : \mathscr{X}^2 \to \mathbb{R}_+$ be continuous. If $\pi_n \in \mathit{\Pi}(\mu_{n,n})$ are optimal transference plans and

$$limsup_{n\to\infty} \int_{\mathscr{X}^2} c(x,y)\mathrm{d}\pi_n(x,y) < \infty \tag{1}$$

then (π_n) is a tight sequence and each of its weak limits $\pi \in \Pi(\mu, v)$ is optimal.

One can even let c vary with n under some conditions.

Prohorov's Theorem

1. If $\mathit{\Pi}$ is tight, then it is relatively compact;
2. Suppose S is separable and complete. if $\mathit{\Pi}$ is relatively compact , then it is tight.

Cyclical Monotonicity

Definition 1.7.3 (Cyclically Monotone Sets and Measures)

A set $\Gamma \subseteq \mathscr{X}^2$ is cyclically monotone if for any n and any $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ with $y_0 = y_n$

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i-1}) \tag{2}$$

A probability measure π on $\mathscr{X}^2$ is cyclically monotone if there exists a monotone Borel set Γ such that $\pi(\Gamma) = 1$.

Cyclical Monotonicity

Proposition 1.7.4 (Optimal Plans Are Cyclically Monotone)

$\mu, v \in P(\mathscr{X})$ and suppose that the cost function c is nonnegative and continuous. Assume that the optimal $\pi \in \Pi(\mu, v)$ has a finite total cost. Then supp π is cyclically monotone. In particular, π is cyclically monotone.

Proposition 1.7.5 (Cyclically Monotone Plans Are Optimal)

Let $\mu, v \in P(\mathscr{X}), c : \mathscr{X}^2 \to \mathbb{R}_+$ continuous and $\pi \in \Pi(\mu, v)$ is a cyclically monotone measure with $c(\pi)$ finite. Then π is optimal in $\Pi(\mu, v)$.

Theorem 1.7.6 (Rockafellar)

A nonempty $\Gamma \subseteq \mathscr{X}^2$ is quadratic cyclically monotone if and only if it is included in the graph of the subdifferential of a lower semicontinuous convex function that is not identically infinite

Uniform Convergence of Optimal Maps

Theorem 1.7.7 (Uniform Convergence of Optimal Maps)

Let μ_n, μ be absolutely continuous measures with finite second moments on an open convex set $U \subseteq \mathbb{R}^d$ such that $\mu_n \to \mu$ weakly, and let $v_n \to v$ weakly with $_n, v \in P(\mathbb{R}^d)$ with finite second moments. If T_n and T are continuous on U and $C(T_n)$ is bounded uniformly in n, then

$$sup_{x\in\Omega}\|T_n(x) - T(x)\| \to 0, n \to \infty$$

for any compact $\Omega \subseteq U$

Stability of Transport Maps

Definition (monotone)

we call a set-valued function (or multifunction) $u : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ **monotone** if whenever $y_i \in u(x_i), i = 1, 2,$

$$\langle y_2 - y_1, x_2 - x_1 \rangle \geq 0$$

If $d = 1$, this simply means that u is a nondecreasing (set-valued) function. Next, u is said to be **maximally monotone** if no points can be added to its graph while preserving monotonicity

$$\langle y^{'} - y, x^{'} - x \rangle \geq 0, whenever\, y \in u(x) \Longrightarrow y^{'} \in u(x^{'})$$

Continuity of Optimal Maps

Proposition 1.7.8 (Continuity at Singletons)

Let $x \in \mathbb{R}^d$such that $u(x) = y$ is a singleton. Then u is nonempty on some neighbourhood of x and it is continuous at x: if $x_n \to x$ and $y_n \in u(x_n)$, then $y_n \to y$.

Notice that this result implies that if a convex function ϕ is differentiable on some open set $E \subseteq \mathbb{R}^d$, then it is continuously differentiable there. If $f : \mathbb{R}^d \to \mathbb{R} \bigcup \infty$is any function, one can define its subgradient at x locally as

$$\partial f(x) = \{f(z) \geq f(x) + \langle y, z - x \rangle + o(\|z - x\|)\}$$

$$= \{y : liminf_{z \to x} \frac{f(z) - f(x) + \langle y, z - x \rangle}{\|z - x\|} \geq 0\}$$

Stability of Transport Maps

When f is convex, one can remove the $o(\|z - x\|)$ term and the inequality holds for all z, i.e., globally and not locally. Since monotonicity is related to convexity, it should not be surprising that monotonicity is in some sense a local property. Suppose that $u(x_0) = y_0$ is a singleton and that for some $y^* \in \mathbb{R}^d$, $\langle y - y^*, x - x_0 \rangle \geq 0, \forall x \in \mathbb{R}^d, y \in u(x)$

Definition (Lebesgue point)

Let $B_r(x_0) = \{x : \|x - x_0\| < r\}$ for $r \geq 0$ and $x_0 \in \mathbb{R}^d$. The interior of a set $G \subseteq \mathbb{R}^d$ is denoted by *intG* and the closure by $\overline{G}$. If G is measurable, then *LebG* denotes the Lebesgue measure of G. Finally, *convG* denotes the convex hull of G. A point x_0 is a Lebesgue point (or of Lebesgue density) of a measurable set $G \subseteq \mathbb{R}^d$,if for any $\epsilon > 0$ there exists $t_\epsilon > 0$ such that A point x_0 is a Lebesgue point (or of Lebesgue density) of a measurable set $G \subseteq \mathbb{R}^d$ if for any $\epsilon > 0$,there exists $t_\epsilon > 0$ such that

$$\frac{Leb(B_t(x_0) \bigcap G)}{Leb(B_t(x_0))} > 1 - \epsilon, 0 < t < t_\epsilon$$

Stability of Transport Maps

We denote the set of points of Lebesgue density of G by G^{den}. Here are some facts about G^{den}: clearly, $intG \subseteq G^{den} \subseteq G$. By the Hahn–Banach theorem,$G^{den} \subseteq int(conv(G))$: indeed, if x is not in $int(convG)$, then there is a separating hyperplane between x and $convG \supseteq G$, so the fraction above is at most $1/2$ for all $t > 0$. By the Hahn–Banach theorem, $G^{den} \subseteq int(conv(G))$: indeed, if x is not in $int(convG)$, then there is a separating hyperplane between x and $convG \supseteq G$, so the fraction above is at most $1/2$ for all $t > 0$.

Stability of Transport Maps

Lemma 1.7.9 (Density Points and Distance)

Let x_0 be a point of Lebesgue density of a measurable set $G \subseteq \mathbb{R}^d$. Then

$$\delta(z) = \delta_G(z) = inf_{x \in G}\|z - x\| = o(\|z - x_0\|), z \to x_0$$

Corollary of Lemma 1.7.9

For almost all $x \in G, \delta(z) = o(\|z - x\|)$ as $z \to x$.

This can be seen in other ways: since δ is Lipschitz, it is differentiable almost everywhere. If $z \in G$ and δ is differentiable at x, then $\nabla\delta(x)$ must be 0 (because δ is minimised there), and then $\delta(z) = o(\|z - x\|)$. We just showed that δ is differentiable with vanishing derivative at all Lebesgue points of x. The converse is not true: $G = \{\pm 1/n\}_{n=1}^{\infty}$ has no Lebesgue points, but $\delta(y) \leq 4y^2$ as $y \to 0$.

Stability of Transport Maps

Lemma 1.7.10 (Local Monotonicity)

Let $x_0 \in \mathbb{R}^d$ such that $u(x_0) = y_0$ and x_0 is a Lebesgue point of a set G satisfying

$$\langle y - y^*, x - x_0 \rangle \geq 0, \forall x \in G, \forall y \in u(x)$$

Then $y^* = y_0$. In particular, the result is true if the inequality holds on $G = O \backslash \mathscr{N}$ with $\emptyset \neq O$ open and $\mathscr{N}$ Lebesgue negligible.

Stability of Transport Maps

Assumptions 1

Let$\mu_n, \mu, v_n, v \in P(\mathbb{R}^d)$ with optimal couplings (with respect to quadratic cost) $\pi_n \in \Pi(\mu_{n,n}), \pi \in \Pi(\mu,)$and convex potentials ϕ_n and ϕ, respectively, such that

- (convergence) $\mu_n \to \mu$ and $_n \to$ weakly;
- (finiteness) the optimal couplings $\pi_n \in \Pi(\mu_{n,n})$ satisfy

$$limsup_{n\to\infty} \int_{\mathscr{X}^2} \frac{1}{2}\|x - y\|^2 \mathrm{d}\pi_n(x, y) < \infty$$

- (unique limit) the optimal $\pi \in \Pi(\mu,)$is unique.

We further denote the subgradients $\partial\phi_n$ and $\partial\phi$ by u_n and u, respectively.

Uniform Convergence and Pointwise Convergence a.s of Optimal Maps

Proposition 1.7.11 (Uniform Convergence of Optimal Maps)

Let **Assumptions 1** hold true and denote $E = supp\mu$ and E^{den} the set of its Lebesgue points. Let Ω be a compact subset of E^{den} on which u is univalued (i.e., $u(x)$ is a singleton for all $x \in \Omega$). Then u_n converges to u uniformly on Ω : $u_n(x)$ is nonempty for all $x \in \Omega$ and all $n > N_\Omega$, and

$$sup_{x\in\Omega} sup_{y\in u_n(x)} \|y - u_n(x)\| \to 0, n \to \infty$$

In particular, if u is univalued throughout $int(E)$ (so that $\phi \in C^1$ there), then uniform convergence holds for any compact $\Omega \subset int(E)$.

Corollary 1.7.12 (Pointwise Convergence μ-Almost Surely)

If in addition μ is absolutely continuous, then $u_n(x) \to u(x)$ μ-almost surely.