An Invitation to Statistics in Wasserstein Space

Wang Chao 2020212429

Shanghai University of Finance and Economics

May 7, 2022

Outline

Content

1. 4.6 Convergence Rates and a Central Limit Theorem on the Real Line
2. 4.7 Convergence of the Empirical Measure and Optimality
3. 5.1 A Steepest Descent Algorithm for the Computation of Frechet Means

Convergence Rates

$\Lambda = T_{\#}\lambda$

Lemma 4.6.1 (Number of Points Grows Linearly)

Let $N_i^n = \Pi_i^n(\mathcal{X})$ denote the total number of observed points.If $\tau_n/\log n \to \infty$,then there exists a constant $C_\Pi > 0$,depending only on the distribution of Π,such that almost surely

$$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^n}{\tau_n} \ge C_\Pi$$

Hint: The Chernoff bound

$$\mathbb{P}(X - \mathbb{E}(X) > t) \le \inf_{\lambda>0} e^{-\lambda t}\mathbb{E}(e^{\lambda(X-\mathbb{E}(X))})$$

$$\mathbb{P}(X - EX > t) = \mathbb{P}\left(e^{\lambda(X-EX)} > e^{\lambda t}\right)$$

$$\le \inf_{\lambda>0} e^{-\lambda t} \int_{e^{\lambda(X-EX)} > e^{\lambda t}} e^{\lambda(X-EX)}\, dP$$

The sketch of proof

N poisson point process.

$N_1^{(n)}, N_2^{(n)}, \cdots N_n^{(n)} \sim N$

$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\lambda_n} \geq 1, \quad a.s.$

$c > 1, \quad t > 0$

$P(N > \lambda/c) \leq \inf_{t>0} \frac{E e^{-Nt}}{e^{-\lambda t/c}} = \inf_{t>0} \exp[\lambda(e^{-t} + \frac{t}{c} - 1)]$

$t = \log c$

$\sum P(N > \lambda/c) \leq \exp(-\lambda_n d) < \infty \quad d = c^{-1}[c - 1 - \log c] > 0$

$P\left(\frac{\min N_i^n}{\lambda_n} \geq 1, \quad\right) = 1 \quad \Leftarrow B.C$

$\frac{\lambda_n}{\log n} \to \infty \cdot \underbrace{\sum \exp(-\lambda_n d)}_{< \infty}$

$\liminf_{n\to\infty} \frac{\min_{1\le i\le n} N_i^{(n)}}{\lambda_n} \geq 1, \quad a.s$

$\sqrt{n}(F - F_n) \to$

$\| F - \hat{F}_n \| \to$

As in the consistency proof,the idea is to write

$$F - \hat{F}_n = (F - F_n) + (F_n - \tilde{F}_n) + (\tilde{F}_n - \hat{F}_n)$$

The standard $O_{\mathbb{P}}$ terminology:
if X_n and Y_n are random variables,then $X_n = O_{\mathbb{P}}(Y_n)$ means that the sequence (X_n/Y_n) is bounded in probability,which by the definition is the condition

$$\forall \epsilon > 0, \exists M : \sup_n \mathbb{P}(|X_n/Y_n| > M) < \epsilon.$$

$X_n = o_{\mathbb{P}}(Y_n)$ means that $X_n/Y_n \to 0$ in probability.

$$\limsup_n \mathbb{P}(|X_n/Y_n| > M) = 0$$

$o(Y_n)$. a.s.

$O(Y_n)$. a.s

Convergence Rates

$T_n = n$.

Theorem 4.6.3 (Convergence Rates on $\mathbb{R}$)

Lemma 4.6.1

Suppose in addition to Assumptions 3 that $d = 1$, $\tau_n / \log n \to \infty$ and that Π is either a Poisson process or a binomial process. Then

$$W_2(\hat{\lambda}_n, \lambda) \le O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\sqrt[4]{\tau_n}}\right) + O_P(\sigma_n), \qquad \sigma_n = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^{(n)},$$

$T_n = n^2$.

$O_P\left(\frac{1}{\sqrt{n}}\right)$

where all the constants in the O_P terms are explicit.

$\sqrt{n}\,(\hat{\lambda}_n - \lambda) \to G_P$

$W(\hat{\lambda}_n, \lambda) = O_P\left(\frac{1}{\sqrt{n}}\right)$

The sketch of proof

$\mathcal{X}=\mathbb{R}$, $W_2(\gamma,\theta)=\|F_\theta^{-1}-F_\gamma^{-1}\|_{L_2(0,1)}$,

$\Lambda_1, \Lambda_2, \cdots \Lambda_n$

$F_{\lambda_n}^{-1} = \frac{1}{n}\sum_{i=1}^{n} F_{\Lambda_i}^{-1}$.

$\sqrt{n}\left(F_{\lambda_n}^{-1}-F_{\lambda}^{-1}\right)=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}F_{\Lambda_i}^{-1}-F_{\lambda}^{-1}\right) \longrightarrow GP$

GP: X_t $(t\in T)$

$\forall\ t_1 \cdots t_n \in T$

$(X_{t_1},\cdots X_{t_n}) \sim$ Multi-normal

$T = L_2(0,1)$

$W_2(\lambda_n,\lambda)=\|F_{\lambda_n}^{-1}-F_{\lambda}^{-1}\| = O_p(n^{-1/2})$ #

$F_{\hat\lambda_n}^{-1} - F_{\lambda}^{-1}$

$F^{-1}_{\hat\lambda_n} - F^{-1}_{\lambda_n}$

$(X_{f_1},\cdots X_{f_n})$, $f_1\cdots f_n \in L_2(0,1)$.

$\|F_{\lambda_n}^{-1}-F_{\hat\lambda_n}^{-1}\| = \left\|\frac{1}{n}\sum_{i=1}^{n}F_{\Lambda_i}^{-1}-\frac{1}{n}\sum_{i=1}^{n}F_{\hat\Lambda_i}^{-1}\right\| \le \frac{1}{n}\sum_{i=1}^{n}\|F_{\Lambda_i}^{-1}-F_{\hat\Lambda_i}^{-1}\| = \frac{1}{n}\sum_{i=1}^{n}W_2(\Lambda_i,\hat\Lambda_i)$

$\le \frac{1}{n}\sum_{i=1}^{n}W_2(\Lambda_i,\tilde\Lambda_i) + \frac{1}{n}\sum_{i=1}^{n}W_2(\tilde\Lambda_i,\hat\Lambda_i) \le \sqrt{C_{4,k}}\ \tau_n$

Central Limit Theorem

Theorem 4.6.5 (Asymptotic Normality)

In addition to the conditions of Theorem 4.6.3, assume that $\tau_n/n^2 \to \infty$, $\sigma_n = o_P(n^{-1/2})$ and λ possesses an invertible distribution function F_λ on K. Then

$\lambda_n \longrightarrow \hat{\lambda}_n$

$$\sqrt{n}(t_\lambda^{\hat{\lambda}_n} - i) \to Z \quad \text{weakly in} \quad L_2(\lambda)$$

for a zero-mean Gaussian process Z with the same covariance operator of T (the latter viewed as a random element in $L_2(\lambda)$, namely with covariance kernel

$$\kappa(x, y) = \mathrm{cov}(T(x), T(y)).$$

If the density f_λ exists and is (piecewise) continuous and bounded below on K, then the weak convergence also holds in $L_2(K)$.

The sketch of proof

$G_n = \sqrt{n}(P_n(f) - P(f))$
$f \in \mathcal{F}$.
$T(f_1, f_2) = Cov(f_1, f_2)$
$G_n \to G_P$ in $\mathcal{F}$
$T(f_1, f_2) = Cov(f_1, f_2)$
$\sqrt{n}(\bar{x} - \mu) \to N(0, 1)$

$X_n \to x \Rightarrow f(x_n) \to f(x)$.
$f(x)$, a.s. continuous

① $\sqrt{n}(F_{\hat{x}_n}^{-1} - F_{x_n}^{-1}) = o_p(1)$
② slutsky Thm.
$\sqrt{n}(F_{\hat{\lambda}_n}^{-1} - F_{\lambda}^{-1})$
$= \underline{\sqrt{n}(F_{\lambda_n}^{-1} - F_{\lambda}^{-1})} + \underbrace{\sqrt{n}(F_{\hat{\lambda}_n}^{-1} - F_{\lambda_n}^{-1})}_{o_p(1)}$
$\to G_P$ in $L_2(0,1)$
$\sqrt{n}(F_{\lambda}^{\hat{\lambda}_n} - 1) = \sqrt{n}(F_{\hat{\lambda}_n}^{-1} \circ F_{\lambda} - F_{\lambda}^{-1} \circ F_{\lambda})$
$= \underline{\sqrt{n}(F_{\hat{\lambda}}^{-1} - F_{\lambda}^{-1})} \circ F_{\lambda} \to G_P \circ F_{\lambda}$
$G_P \to G_P \circ F_{\lambda}$

③ continuous mapping Thm
$G_P \to G_P \circ F_{\lambda}$
$f(G_P) = G_P \circ F_{\lambda}$.

Content

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} F_{\Lambda_i}^{-1}\circ F_\lambda - \underline{F_\lambda^{-1}\circ F_\lambda}\right)$$

$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \underbrace{t_\lambda^{\Lambda_i}}_{T_i} - i\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} T_i - i\right) \to Z$$

Convergence of the Empirical Measure and Optimality

π: binomial $Z_n = n$.

$O_p(\frac{1}{\sqrt{n}})$

$\mu_n = n^{-1} \sum_{i=1}^{n} \delta(x_i)$

Lemma 4.7.1

Let $\mu \in P(\mathcal{X})$ be any measure. Then

$$\mathbb{E} W_p(\mu, \mu_n) = \begin{cases} = \infty & \mu \notin \mathcal{W}_p(\mathcal{X}) \\ \to 0 & \mu \in \mathcal{W}_p(\mathcal{X}) \end{cases}$$

Proof

We already know that $Y_n = W_p^p(\mu, \mu_n) \to 0$ almostly sure if and only if $\mu \in \mathcal{W}_p(\mathcal{X})$ in Proposition 2.2.6.

$\in \mathbb{R}$

$Y_n \to 0$ a.s

$\lim_{n\to\infty} E\, Y_n = E \lim_{n\to\infty} Y_n = 0$

$$0 \le Y_n \le \int_{\mathcal{X}^2} ||x - y||^p d\mu \otimes \mu_n = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{X}} ||x - X_i||^p d\mu := Z_n$$

Let $V = \int_{\mathcal{X}} ||x - X_1||^p d\mu$

$Z_1 \sim \cdots Z_n \sim V$.

$E Z_n \to E Z$

Continued

Z_n $\qquad\qquad\qquad\qquad$ $\mu \in W_p(\mathcal{X})$

$$\mathbb{E}[V] = \int_{\mathcal{X}^2} \|x-y\|^p d\mu \otimes \mu \le 2^p \int_{\mathcal{X}^2} \|x\|^p + \|y\|^p d\mu \otimes \mu < \infty$$

$Z_n \xrightarrow{a.s.} \le V$

It follows that $\mathbb{E}[Z_n] \to \mathbb{E}[V]$. By the General Dominated Convergence Theorem ,$\mathbb{E}[Y_n] = 0$. Jensen's inequality yields

$p \ge 1$

$$\mathbb{E}W_p(\mu, \mu_n) \le [\mathbb{E}W_p^p(\mu, \mu_n)]^{1/p} = [\mathbb{E}Y_n]^{1/p} \to 0$$

Lemma 4.7.3 ($\sqrt{n}$ Lower Bound)

Let $\mu \in P(\mathcal{X})$ be nondegenerate. Then there exists s constant $c(\mu) > 0$ such that for all $p \geq 1$ and all n

$$\mathbb{E} W_p(\mu_n, \mu) \geq \frac{c(\mu)}{\sqrt{n}}$$

proof

Let $X \sim \mu$, and let $a \neq b$ be two points in the support μ. Consider $f(x) = \min\{1, ||x - a||\}$.

The Kantorovich–Rubinstein Theorem:

$$W_1(\mu,\nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X}^2} d(x, y) d\pi(x, y) = \sup_{||\varphi||_{Lip} \leq 1} \left| \int_{\mathcal{X}} \varphi d\mu - \int_{\mathcal{X}} \varphi d\nu \right|$$

φ

$\Rightarrow \left| \int_{\mathcal{X}} \varphi d\mu - \int_{\mathcal{X}} \varphi d\nu \right|$

Using the Kantorovich–Rubinstein Theorem gives

Continued

$X_n := n^{-1/2}\left(\sum_{i=1}^{n} f(x_i) - \mathbb{E}f(x)\right) \to N(0, \operatorname{var} f(x))$ $=: X$

$$\sqrt{n}\mathbb{E}W_p(\mu_n,\mu) \geq \sqrt{n}\mathbb{E}W_1(\mu_n,\mu) \geq \mathbb{E}|n^{-1/2}(\sum_{i=1}^{n} f(X_i) - \mathbb{E}f(X))|$$

$\mathbb{E}|X_n| \to \mathbb{E}|X|$

$|X_n| \xrightarrow{d} |X|$

$$\to \sqrt{\frac{2Varf(X)}{\pi}}$$

Lemma 4.7.4 (Separated Support)

Suppose that there exist Borel sets $A, B \subset \mathcal{X}$ such that $\mu(A \cup B) = 1$,

$A \cap B \neq \phi$

$$\mu(A)\mu(B) > 0 \quad and \quad d_{\min} = \inf_{x\in A, y\in B} ||x - y|| > 0$$

Then,for any $p \geq 1$ there exists $c_p(\mu) > 0$ such that

$$\mathbb{E}W_p(\mu_n,\mu) \geq c_p(\mu)n^{-1/2p}$$

$n^{-\frac{1}{2p}}$ $\quad \frac{1}{\sqrt{n}}$

On the real line, it is easy to obtain a sufficient condition for the optimal rate $n^{-1/2}$ to be attained for W_1:

$$\mathbb{E}W_1(\mu_n,\mu) = \int_{\mathbb{R}} \mathbb{E}|F_n(t) - F(t)|dt \le n^{-1/2} \int_{\mathbb{R}} \sqrt{F(t)(1-F(t))}dt$$

∞

so that $W_1(\mu_n,\mu)$ is of the optimal order $n^{-1/2}$ if

$$J_1(\mu) := \int_{\mathbb{R}} \sqrt{F(t)(1-F(t))}dt < \infty$$

For any $\delta > 0$, we have for $X \sim \mu$ that
$\mathbb{E}(|X|^{2+\delta}) < \infty \Rightarrow J_1(\mu) < \infty \Rightarrow \mathbb{E}|X|^2 < \infty$

$X \in L_2(\mu)$

Convergence of the Empirical Measure and Optimality

$\mu = f + g$

Let f denote the density of the absolutely continuous part of μ.

Theorem 4.7.5 (Rate of Convergence of Empirical Measures)

Let $p \geq 1$ and $\mu \in \mathcal{W}_p(\mathbb{R})$. The condition

$$J_p(\mu) = \int_{\mathbb{R}} \frac{[F(t)(1 - F(t))]^{p/2}}{f(t)^{p-1}} dt < \infty$$

is necessary and sufficient for $\mathbb{E}W_p(\mu_n, \mu) = O(n^{-1/2})$

Convergence of the Empirical Measure and Optimality

Proposition 4.7.7

Let $\mu \in \mathcal{W}_1(\mathbb{R}^d)$ have an absolutely continuous part with respect to Lebesgue measure, and let v_n be any discrete measure supported on n points (or less). Then there exists a constant $C(\mu) > 0$ such that

$$W_p(\mu, v_n) \geq W_1(\mu, v_n) \geq C(\mu) n^{-1/d}.$$

Proof

$\underline{f}$: density of $\mu_c \leftarrow \mu$

$\sigma > 0$. M. $\underline{\geq \sigma} = \mu_c(\{x : f(x) \leq M\})$

$x_1 \cdots x_n$ support points of ν_n

$\forall\, \varepsilon > 0$, $\mu_{c,M}$.

$$\underline{\mu_{c,M}}\Big(\bigcup_{i=1}^{n} B_\varepsilon(x_i)\Big) \leq M \sum_{i=1}^{n} \mathrm{Leb}(B_\varepsilon(x_i)) = M \cdot n\varepsilon^d \underbrace{\mathrm{Leb}_d(B_1(0))}_{C_d}$$

$W_1(\nu_n, \mu) \geq \sigma \varepsilon$.

$= \sigma (\sigma / M C_d)^{1/d} \cdot n^{-1/d}$ if $\varepsilon^d = \sigma (n M C_d)^{-1}$

$\nexists$ $\mu_{c,M}(\quad) \leq \sigma$

if μ, supported on compact set

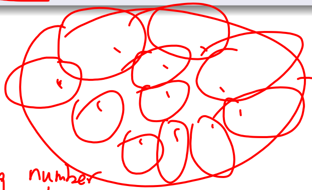
$\Rightarrow N(\mu, \varepsilon, 0) \leq K \varepsilon^{-d}$

Theorem 4.7.8

If for some $d > 2p$, $N(\mu, \epsilon, \epsilon^{dp/(d-2p)}) \leq K\epsilon^{-d}$, then $\mathbb{E}W_p^{(\mu_n, \mu)} \leq C_p n^{-1/d}$.
where

$d > 2$, $\mathbb{E}\, W_1(\mu_n, \mu) \leq n^{-1/d}$ $\qquad n^{-1/d}$

$N(\mu, \epsilon, \tau)$ = *minimal number of balls whose union has* μ *mass* $\geq 1 - \tau$



$X \subset \bigcup_{i=1}^{n} B_{\varepsilon}(x_i)$

$\mathbb{P}(X \subset \bigcup_{i=1}^{n} B_{\varepsilon}(x_i)) = 1$

$\mathbb{P}(X \qquad) > 1 - \tau$

covering number $N(X, \varepsilon, d)$

Content

A Steepest Descent Algorithm

$\mathbb{R}^d$

We assume that N is a fixed integer and consider a fixed collection $\mu^1, \mu^2, ..., \mu^N \in \mathcal{W}_2(\mathbb{R}^d)$ with μ^1 absolutely continuous with bounded density,which has unique Frechet mean by Proposition 3.1.8. Theorem 3.1.14 shows that if γ is absolutely continuous then the associated Frechet functional

$\bar{\mu}$

$\bar{\mu} = \underset{\mu \in \mathcal{W}_2(\mathbb{R}^d)}{\operatorname{argmin}} F(\gamma)$

$$F(\gamma) = \frac{1}{2N} \sum_{i=1}^{N} W_2^2(\mu^i, \gamma)$$

$\gamma_1 \to \gamma_2$

has Frechet derivative

$$F'(\gamma) = -\frac{1}{N} \sum_{i=1}^{N} \log_\gamma(\mu^i) = -\frac{1}{N} \sum_{i=1}^{N} (t_\gamma^{\mu^i} - i)$$

$\frac{1}{N} \sum^{i} (t_\gamma^{\mu^i} - i)$

A Steepest Descent Algorithm

We use $\gamma_j \in \mathcal{W}_2(\mathbb{R}^d)$ to represent our estimate of the Frechet mean at step j.If we introduce a step size $\tau_j > 0$,and to follow the steepest descent of F given by the negative of the gradient:

$$\gamma_{j+1} = [i - \tau_j F'(\gamma_j)]_{\#\gamma_j} = [i + \tau_j \frac{1}{N}\sum_{i=1}^{N}(t_{\gamma_j}^{\mu^i} - i)]_{\#\gamma_j}$$

$$x_{i+1} = x_i - \tau_i F'(x_i)$$

$$\gamma_{j+1} = [i - F'(\gamma_j)]_{\#\gamma_j}$$

$$= [\frac{1}{N}\sum_{i=1}^{N} t_{\gamma_j}^{\mu^i}]_{\#\gamma_j} = T_j \# \gamma_j$$

A Steepest Descent Algorithm

$\gamma_1 \quad \gamma_2 \quad \cdots$

Lemma 5.1.1 (Regularity of the Iterates)

If γ_0 is absolutely continuous and $\tau = \tau_0 \in [0,1]$,then $\gamma_1 = \exp(-\tau_0 F'(\gamma_0))$ is also absolutely continuous.

Lemma 5.1.2 (Optimal Stepsize)

If $\gamma_0 \in \mathcal{W}(\mathbb{R}^d)$ is absolutely continuous,then

$\in L_2(\gamma_0)$

$$F(\gamma_1) - F(\gamma_0) \leq -\|F'(\gamma_0)\|^2[\tau - \frac{\tau^2}{2}]$$

and the bound on the right-hand side of the last display is minimised when $\tau = 1$.

Proof of Lemma 5.1.2

Let $S_i = t_{\gamma_0}^{\mu^i}$ and $W_i = S_i - i$. Then

$$2NF(\gamma_0) = \sum_{i=1}^{N} W_2^2(\gamma_0, \mu^i) = \sum_{i=1}^{N} \underbrace{\int_{\mathbb{R}^d} \|S_i - i\|^2 d\gamma_0}_{L_2(\gamma_0)} = \sum_{i=1}^{N} \|W_i\|^2.$$

We invoke the inequality (2.3)

$$W_p^p(t_{\#_\mu}, s_{\#_\mu}) \le \int_{\mathcal{X}} \|t(x) - s(x)\|^p d\mu(x) \qquad \gamma_0$$

$\gamma_1 = \boxed{}\#\gamma_0 \qquad \mu^i = t_{\gamma_0}^{\mu^i}\#\gamma_0$

to get

$$W_2^2(\gamma_1, \mu^i) \le \int_{\mathbb{R}^d} \|(1-\tau)i + \frac{\tau}{N}\sum_{j=1}^{N} S_j - S_i\|_{\mathbb{R}^d}^2 d\gamma_0 = \| - W_i + \frac{\tau}{N}\sum_{j=1}^{N} W_j\|^2$$

A Steepest Descent Algorithm

$\langle -W_i + \frac{\tau}{N}\sum_{j=1}^{N} W_j, -W_i + \frac{\tau}{N}\sum_{j=1}^{N} W_j \rangle$

Proof of Lemma 5.1.2

$2NF(\gamma_1) \le \sum_{i=1}^{N} \| -W_i + \frac{\tau}{N}\sum_{j=1}^{N} W_j \|^2$

$$2NF(\gamma_1) \le \sum_{i=1}^{N} ||W_i||^2 - 2\frac{\tau}{N}\sum_{i,j=1}^{N}\langle W_i, W_j\rangle + N\tau^2 ||\sum_{j=1}^{N}\frac{1}{N}W_j||^2$$

$$= 2NF(\gamma_0) - 2N\tau||\sum_{i=1}^{N}\frac{1}{N}W_i||^2 + N\tau^2||\sum_{i=1}^{N}\frac{1}{N}W_i||^2,$$

Recall that $F'(\gamma_0) = -\frac{1}{N}\sum_{i=1}^{N}(t_{\gamma_0}^{\mu_i} - i) = -\frac{1}{N}\sum_{i=1}^{N} W_i$ yields

$= \zeta_i - i = W_i$

$$F(\gamma_1) - F(\gamma_0) \le -||F'(\gamma_0)||^2[\tau - \frac{\tau^2}{2}]$$

and it is obvious that $\tau - \frac{\tau^2}{2}$ is maximised at $\tau = 1$.

A Steepest Descent Algorithm

$F'(\gamma_0) - F(\gamma_p) \geq \cancel{\phi}\, \frac{1}{2} ||F'(\gamma_0)||^2$

A first step in the convergence analysis is that the sequence $F(\gamma_j)$ is nonincreasing and that for any integer k,

$\frac{1}{2}\sum_{j=0}^{k} ||F'(\gamma_j)||^2 \leq$

$$\frac{1}{2}\sum_{j=0}^{k} ||F'(\gamma_j)||^2 \leq \sum_{j=0}^{k} [F'(\gamma_j) - F'(\gamma_{j+1})] = F'(\gamma_0) - F'(\gamma_{k+1}) \leq F(\gamma_0).$$

which implies $||F'(\gamma_j)||^2$ must vanish as $j \to \infty$.

A Steepest Descent Algorithm

The resulting iteration is summarised as Algorithm 1.

Algorithm 1 Steepest descent via Procrustes analysis

(A) Set a tolerance threshold $\varepsilon > 0$.
(B) For $j = 0$, let γ_j be an arbitrary absolutely continuous measure. γ_0
(C) For $i = 1, \ldots, N$ solve the (pairwise) Monge problem and find the optimal transport map $\mathbf{t}_{\gamma_j}^{\mu^i}$ from γ_j to μ^i.
(D) Define the map $T_j = N^{-1} \sum_{i=1}^{N} \mathbf{t}_{\gamma_j}^{\mu^i}$.
(E) Set $\gamma_{j+1} = T_j \# \gamma_j$, i.e. push-forward γ_j via T_j to obtain γ_{j+1}.
(F) If $\|F'(\gamma_{j+1})\| < \varepsilon$, stop, and output γ_{j+1} as the approximation of $\bar{\mu}$ and $\mathbf{t}_{\gamma_{j+1}}^{\mu^i}$ as the approximation of $\mathbf{t}_{\bar{\mu}}^{\mu^i}$, $i = 1, \ldots, N$. Otherwise, return to step (C).