Victor M. Panaretos
Yoav Zemel

An Invitation to Statistics in Wasserstein Space

Supplementary Proofs

September 2, 2019

Springer

Contents

Introduction

This document contains proofs omitted from the book. To avoid confusion, results stated in this document follow the numbering scheme Lemma 2, while those from the book follow the scheme Proposition 1.5.2.

Chapter 1
Proofs for Chapter 1

Proof (Proof of Remark 1.3.2). Suppose that μ is supported on n points and ν on m points. Then there cannot exist a transport map from μ to ν if $m > n$ and there cannot be a transport map from ν to μ if $n > m$. Consequently, if one is interested in solving both Monge problems, the only possible case is when $n = m$. If we now assume that the weights are ordered:

$$\mu = \sum_{i=1}^{n} a_i \delta\{x_i\}, \qquad \nu = \sum_{i=1}^{n} b_i \delta\{y_i\}, \qquad 0 \le a_1 \le \cdots \le a_n; \quad 0 \le b_1 \le \cdots \le b_n,$$

then transport maps exist if and only if $a_i = b_i$, $i = 1, \dots, n$. One can then split the problem into smaller uniform problems: suppose for example that $n = 7$ and $a_5 < a_6 = a_7$. Then x_7 and x_6 can only be sent to y_7 or y_6, and this creates a uniform discrete problem of size 2. Arguing inductively, we see that the only interesting discrete case for the Monge problem is the uniform one (with the same number of points).

Proof (Proof of Proposition 1.5.2). It is well known that $F^{-1}(u) \le x$ if and only if $u \le F(x)$. Let $A = \{u : G^{-1}(u) > F^{-1}(u)\} \subseteq (0,1)$ and notice that for $u \in A$, $F^{-1}(u) \le x < G^{-1}(u)$ if and only if $G(x) < u \le F(x)$. A similar equivalence holds when $u \in B = (0,1) \setminus A$. It follows from Fubini's theorem that

$$\int_A |G^{-1}(u) - F^{-1}(u)|\,\mathrm{d}u = \int_A \left(\int_{F^{-1}(u)}^{G^{-1}(u)} 1\,\mathrm{d}x \right) \mathrm{d}u = \int_{\mathbb{R}} \left(\int_{G(x)}^{F(x)} 1_A(u) \mathbf{1}\{F(x) \ge G(x)\}\,\mathrm{d}u \right) \mathrm{d}x;$$

$$\int_B |G^{-1}(u) - F^{-1}(u)|\,\mathrm{d}u = \int_B \left(\int_{G^{-1}(u)}^{F^{-1}(u)} 1\,\mathrm{d}x \right) \mathrm{d}u = \int_{\mathbb{R}} \left(\int_{F(x)}^{G(x)} 1_B(u) \mathbf{1}\{G(x) \ge F(x)\}\,\mathrm{d}u \right) \mathrm{d}x.$$

Since $1_A(u) + 1_B(u) = 1$, summing up these equalities yields the result.

Proof (Proof of Theorem 1.6.6). The optimal map is $G^{-1} \circ F$ by Theorem 1.5.1, and the discussion in the preceding paragraph proves the result when $k = 1$, since we have a composition of C^1 functions. When $k = 2$, we let $H = G^{-1}$ and use the formula $H'(t) = 1/G'(H(t))$ for all $t \in (0,1)$. Then both G' and H are C^1, so that H' is C^1, and consequently H is C^2. By induction we see that if G is C^k, then so is H. If in addition F is C^k, then $T = G^{-1} \circ F$ is C^k.

For the case $k = 0$, observe that G is strictly increasing, because $\operatorname{supp}\nu$ is an interval. Since G is assumed continuous, so is $H = G^{-1}$, so that $T = H \circ F$ must be continuous too.

Proof (Proof of Lemma 1.7.9). For any $1 > \varepsilon > 0$ there exists $0 < t_\varepsilon$ such that for $t < t_\varepsilon$,

$$\frac{\mathrm{Leb}(B_t(x_0) \cap G)}{\mathrm{Leb}(B_t(x_0))} > 1 - \varepsilon^d.$$

Fix z such that $t = t(z) = \|z - x_0\| < t_\varepsilon$. The intersection of $B_t(x_0)$ with $B_{2\varepsilon t}(z)$ includes a ball of radius εt centred at $y = x_0 + (1-\varepsilon)(z - x_0)$, so that

$$\frac{\mathrm{Leb}(B_t(x_0) \cap B_{2\varepsilon t}(z))}{\mathrm{Leb}(B_t(x_0))} \geq \frac{\mathrm{Leb}(B_{\varepsilon t}(y))}{\mathrm{Leb}(B_t(x_0))} = \varepsilon^d.$$

It follows that $G \cap B_{2\varepsilon t}(z)$ is nonempty. In other words: for any $\varepsilon > 0$ there exists t_ε such that if $\|z - x_0\| < t_\varepsilon$, then there exists $x \in G$ with $\|z - x\| \leq 2\varepsilon t(z) = 2\varepsilon\|z - x_0\|$. This means precisely that $\delta(z) = o(\|z - x_0\|)$ as $z \to x_0$.

Proof (Proof of Lemma 1.7.10). Set $z_t = x_0 + t(y^* - y_0)$ for $t > 0$ small. It is possible that $z_t \notin G$; but Lemma 1.7.9 guarantees existence of $x_t \in G$ with $\|x_t - z_t\|/t \to 0$. By Proposition 1.7.8 $u(x_t)$ is nonempty for t small enough. For $y_t \in u(x_t)$,

$$\begin{aligned} 0 \leq \langle y_t - y^*, x_t - x_0 \rangle &= \langle y_t - y^*, x_t - z_t \rangle + \langle y_t - y^*, z_t - x_0 \rangle \\ &= \langle y_t - y^*, x_t - z_t \rangle + t\,\langle y_t - y_0, y^* - y_0 \rangle - t\|y^* - y_0\|^2. \end{aligned}$$

It now follows from the Cauchy–Schwarz inequality that

$$\|y^* - y_0\|^2 \leq \|y_t - y_0\|\|y^* - y_0\| + t^{-1}\|x_t - z_t\|\left(\|y_t - y_0\| + \|y^* - y_0\|\right).$$

As $t \searrow 0$ the right-hand side vanishes, since $y_t \to y_0$ (Proposition 1.7.8) and $\|x_t - z_t\|/t \to 0$. It follows that $y^* = y_0$.

We now prove Proposition 1.7.11, by establishing the two properties:

- if a sequence in the graph of u_n converges, then the limit is in the graph of u (Lemma 2);
- sequences in the graph of u_n are bounded if the domain is bounded (Proposition 4).

Each property will in turn be proven using an intermediate lemma.

Lemma 1 (points in the limit graph are limit points) *Let $x_0 \in \operatorname{supp}\mu$ be such that $u(x_0) = \{y_0\}$ is a singleton. Then there exists a sequence $(x_n, y_n) \in u_n$ that converges to (x_0, y_0).*

Proof. This is essentially the same argument as used in the proof of Theorem 1.7.2. Invoking the continuity of u at x_0 (Proposition 1.7.8), for any k there exists $\delta = \delta_k > 0$ such that if $x \in B_\delta(x_0) = \{x : \|x - x_0\| < \delta\}$ then $u(x)$ is nonempty and if $y \in u(x)$, then $\|y - y_0\| < 1/k$. Assume without loss of generality that $\delta_k \to 0$, and set $B_k = B_{\delta_k}(x_0)$, $V_k = B_{1/k}(y_0)$. Then $u(B_k) \subseteq V_k$, so

$$\pi(B_k \times V_k) = \pi\{(x,y) : x \in B_k, y \in u(x) \cap V_k\} = \pi\{(x,y) : x \in B_k, y \in u(x)\} = \mu(B_k) > 0,$$

because B_k is a neighbourhood of $x_0 \in \operatorname{supp}(\mu)$. Since $B_k \times V_k$ is open, we have by the portmanteau lemma 1.7.1 that $\pi_n(B_k \times V_k) > 0$ for $n \geq N_k$. But π_n is concentrated on the graph of u_n, so when $n \geq N_k$ there exist $(x_n, y_n) \in u_n \cap [B_k \times V_k]$, so that $\|x_n - x_0\| < \delta_k$ and $\|y_n - y_0\| < 1/k$. This completes the proof.

Lemma 2 (limit points are in the limit graph) *Let x_0 be a Lebesgue point of $E = \operatorname{supp}\mu$ (for example $x_0 \in \operatorname{int}E$) such that $u(x_0) = \{y_0\}$ is a singleton. If a subsequence $(x_{n_k}, y_{n_k}) \in u_{n_k}$ converges to (x_0, y^*), then $y^* = y_0$.*

Proof. The set $\mathscr{N} \subseteq \mathbb{R}^d$ of points where u contains more than one element has Lebesgue measure zero. Moreover, there exists a neighbourhood V of x_0 on which u is nonempty (Proposition 1.7.8). It follows that x_0 is a Lebesgue point of $G = (E \cap V) \setminus \mathscr{N}$, and $u(x)$ has one and only element for every $x \in G$. Let us fix $(x,y) \in u$ with $x \in G$. Application of Lemma 1 to the sequence $\{u_{n_k}\}_{k=1}^\infty$ at x yields sequences $x'_{n_k} \to x$ and $y'_{n_k} \to y$ with $(x'_{n_k}, y'_{n_k}) \in u_{n_k}$. Consequently,

$$\langle y - y^*, x - x_0 \rangle = \lim_{l \to \infty} \left\langle y'_{n_k} - y_{n_k}, x'_{n_k} - x_{n_k} \right\rangle \geq 0, \qquad \forall x \in G\ \forall y \in u(x).$$

It now follows from Lemma 1.7.10 that $y^* = y_0$.

We now know that if $u_n(x)$ converges and $u(x) = \{y\}$, then $u_n(x) \to y$. It therefore suffices to show that $u_n(x)$ remains in a bounded set. To this end we shall use another result about monotone functions: if x is in the convex hull of $x_1, \dots, x_m$, $y_i \in u(x_i)$, and $y \in u(x)$, then $\|y\|$ can be bounded in terms of $\|y_i\|$ and the distance of x from the boundary of $\operatorname{conv}(x_1, \dots, x_m)$. It will be convenient to introduce the ℓ_∞ balls $B_\varepsilon^\infty(x_0) = \{x : \|x - x_0\|_\infty < \varepsilon\}$ and their closures $\overline{B}_\varepsilon^\infty(x_0)$, because unlike the ℓ_2 balls, ℓ_∞ balls are polytopes and equal the convex hull of their finitely many vertices. (For that purpose, we could have also chosen ℓ_1 balls.)

We will need the following easy result about ℓ_∞ balls: let $Z = \{z_1, \dots, z_m\}$, $m = 2^d$ be a collection of vectors with the following property: for each collection $(e_1, \dots, e_d) \in \{\pm 1\}^d$ there exists a vector $y \in Z$ such that $|y_j| > 1$ and $y_j e_j > 0$ for all $j = 1, \dots, d$. Then $\operatorname{conv}Z \supseteq \overline{B}_1^\infty(0)$. In geometric terms this means that if we have 2^d points that are "more extreme" than the vertices of the unit ℓ_∞ ball around zero, then the convex hull of these points includes this ℓ_∞ ball.

The proof of this result is a straightforward consequence of the Hahn–Banach theorem. We show that $e = (e_1, \dots, e_d)$ cannot be separated from Z with a hyperplane for any $e_j \in \{\pm 1\}$. Indeed, let $x \in \mathbb{R}^d \setminus \{0\}$ be any vector and set $J = \{j : e_j x_j > 0\}$. Pick $w, y \in Z$ such that $w_j e_j > 0$ if and only if $y_j e_j < 0$ if and only if

$j \in J$. Since $|w_j| > 1$ and $|y_j| > 1$ this gives $x_j y_j < x_j e_j < x_j w_j$ whenever $x_j \neq 0$ and since $x \neq 0$,

$$\langle x, y\rangle < \langle x, e\rangle < \langle x, w\rangle .$$

Lemma 3 (continuity of convex hulls) *Let $Z = \{z_i\}_{i\in I} \subseteq \mathbb{R}^d$ be an arbitrary collection of points and let $\tilde{Z} = \{\tilde{z}_i\}_{i\in I}$ be another collection such that $\|\tilde{z}_i - z_i\|_\infty \leq \varepsilon$ for all $i \in I$. If* $\mathrm{conv}Z \supseteq B^\infty_\rho(x_0)$*, then* $\mathrm{conv}\tilde{Z} \supseteq B^\infty_{\rho-\varepsilon}(x_0)$.

Proof. Without loss of generality $\varepsilon < \rho$. Fix $\varepsilon < \rho' < \rho$. Each vertex of $\overline{B}^\infty_{\rho'}(x_0)$ takes the form

$$y = x_0 + \rho'(e_1, \ldots, e_d), \qquad e_d \in \{\pm 1\},$$

and can be written as a (finite) convex combination $y = \sum a_i z_i$ with $z_i \in Z$. If we define $\tilde{y} = \sum a_i \tilde{z}_i \in \mathrm{conv}\tilde{Z}$, then $\|\tilde{y} - y\|_\infty \leq \varepsilon$. It follows that $\tilde{y}$ is "more extreme" than the vertex

$$x = x_0 + (\rho' - \varepsilon)(e_1, \ldots, e_d)$$

of the ℓ_∞-ball $B^\infty_{\rho'-\varepsilon}(x_0)$, in the sense that $y_j - x_0$ has a larger absolute value than $x_j - x_0$ but the same sign for all $j = 1, \ldots, d$. For each of the 2^d vertices we can find a corresponding $\tilde{y}$, and consequently $\mathrm{conv}\tilde{Z} \supseteq B^\infty_{\rho'-\varepsilon}(x_0)$ by the discussion before the lemma. Since $\rho' < \rho$ was arbitrary this completes the proof.

Proposition 4 (boundedness) *Let $\Omega \subseteq \mathrm{int}(\mathrm{conv}(\mathrm{supp}(\mu)))$ be compact. Then there exist $N(\Omega)$ and a constant $R(\Omega)$ such that for all $n > N(\Omega)$, $u_n(x)$ is nonempty for all $x \in \Omega$ and $\sup_{x\in\Omega} \sup_{y\in u_n(x)} \|y\| \leq R(\Omega)$ is bounded uniformly.*

Proof. If we set $E = \mathrm{supp}(\mu)$ and $F = \mathrm{conv}(E)$, then Ω is a compact subset of the open set $\mathrm{int}F$. Consequently, there exists $\delta = \delta(\Omega) > 0$ such that $\overline{B}^\infty_{3\delta}(\Omega) \subseteq \mathrm{int}F$. We may construct a finite collection $\{\omega_j\} \subseteq \Omega$ such that the union of $B^\infty_\delta(\omega_j)$ includes Ω. Since each vertex of $\cup_j \overline{B}^\infty_{3\delta}(\omega_j)$ is in F, it can be written as a convex combination of elements of E. Consequently, there exists a finite set $Z = \{z_1, \ldots, z_m\} \subseteq E$ with $\mathrm{conv}Z \supseteq B^\infty_{3\delta}(\omega_j)$ for any j.

The ball $B_i = B^\infty_\delta(z_i)$ is an open neighbourhood of an element of $\mathrm{supp}\mu$ and therefore has positive measure, say $2\varepsilon_i > 0$. By the portmanteau lemma 1.7.1 $\mu_n(B_i) > \varepsilon_i$ for all n large and all $i = 1, \ldots, m$. We can set $\varepsilon = \min_i \varepsilon_i > 0$ and invoke the tightness of $\{\nu_n\}$ to find a compact set K_ε with $\inf_n \nu_n(K_\varepsilon) > 1 - \varepsilon$. A simple calculation shows that this construction guarantees the existence of $x_{ni} \in B_i$ and $y_{ni} \in u_n(x_{ni})$ such that $y_{ni} \in K_\varepsilon$. Setting

$$\tilde{Z} = X_n = \{x_{n1}, \ldots, x_{nm}\},$$

noticing that by definition $\|x_{ni} - z_i\|_\infty \leq \delta$ and applying Lemma 3, we obtain

$$\mathrm{conv}X_n = \mathrm{conv}(\{x_{n1}, \ldots, x_{nm}\}) \supseteq B^\infty_{3\delta-\delta}(\omega_j) = B^\infty_{2\delta}(\omega_j) \qquad \text{for all } j.$$

Recall that $B^\infty_\delta(\omega_j)$ cover Ω. From this it follows that $\mathrm{conv}X_n \supseteq B^\infty_\delta(\Omega) \supseteq B_\delta(\Omega)$ (since $\|x\| \geq \|x\|_\infty$, ℓ_2-balls are always included in ℓ_∞-balls of the same radius).

We are now in a position to employ the property of monotonicity mentioned above. From [2, Lemma 1.2(4)] we conclude that for any $\omega\in\Omega$ and any $y_0\in u_n(\omega)$,

$$\|y_0\|\le\frac{[\sup_{x,z\in X_n}\|x-z\|][\max_{x\in X_n}\inf_{y\in u_n(x)}\|y\|]}{d(\omega,\mathbb{R}^d\setminus\operatorname{conv}X_n)}\le\frac{1}{\delta}\left[\sup_{k,l}\|x_{nk}-x_{nl}\|\right]\left[\max_i\inf_{y\in u_n(x_{ni})}\|y\|\right].$$

To bound the infimum at the right-hand side, we can take y to be y_{ni}, which all lie in the compact set K_ε. To bound the supremum independently of n, we use the approximation $\|x_{nk}-z_k\|\le\sqrt{d}\|x_{nk}-z_k\|_\infty\le\sqrt{d}\delta$, so that $\|x_{nk}-x_{nl}\|\le 2\sqrt{d}\delta+\|z_k-z_l\|$. Hence

$$\forall n>N(\delta)\,\forall\omega\in\Omega\quad\forall y_0\in u_n(\omega):\qquad\|y_0\|\le\frac{1}{\delta}\left(2\sqrt{d}\delta+\max_{k,l}\|z_k-z_l\|\right)\sup_{y\in K_\varepsilon}\|y\|.$$

Recall that δ depends only on Ω, ε and Z only on δ, and K_ε only on ε, so $N(\delta)=N(\delta(\Omega))$ and the bound at the right-hand side does not depend on n.

Finally, the fact that u_n is not empty on Ω is a consequence of the almost convexity of $\operatorname{dom}u$ ([2, Corollary 1.3(2)]).

Proof (Proof of Proposition 1.7.11). After all the hard work, the proof is now straightforward.

There exists N_Ω such that for all $n>N_\Omega$, $u_n(x)$ is nonempty and (Proposition 4)

$$\sup_{x\in\Omega}\sup_{y\in u_n(x)}\|y\|\le C_{\Omega,d}<\infty,\qquad n>N_\Omega,$$

where $C_{\Omega,d}$ is a constant that depends only on Ω (and the dimension d).

If uniform convergence did not hold, then one could find $\varepsilon>0$ and subsequences $(x_{n_k},y_{n_k})\in u_{n_k}$ with $x_{n_k}\in\Omega$ and

$$\|y_{n_k}-u(x_{n_k})\|>\varepsilon,\qquad k=1,2,\dots.$$

Since the x_{n_k}'s are bounded (in Ω) and the y_{n_k}'s are bounded too, they have subsequences that converge to $x\in\Omega$ and some y, that must equal $u(x)$ by Lemma 2. Using again the continuity of u at x (Proposition 1.7.8), we get (up to subsequences)

$$\varepsilon<\|y_{n_k}-u(x_{n_k})\|\le\|y_{n_k}-y\|+\|y-u(x)\|+\|u(x)-u(x_{n_k})\|\to 0,\qquad k\to\infty,$$

a contradiction.

Chapter 2
Proofs for Chapter 2

Proof (Proof of Proposition 2.2.3). Suppose that (2.6) holds. If $\mu_n \in \mathscr{K}$, then there exists a measure μ_0 such that $\mu_n \to \mu_0$ weakly (up to a subsequence), and as (2.4) holds for that subsequence, it converges in the Wasserstein space.

Conversely, if (2.6) does not hold, then we can find a sequence $\mu_n \in \mathscr{K}$ such that for some $\varepsilon > 0$,

$$\int_{\{x:\|x\|>n\}} \|x\|^p \, \mathrm{d}\mu_n(x) > \varepsilon, \qquad n = 1, 2, \dots.$$

Obviously no subsequence of μ_n can converge in the Wasserstein space, in view of (2.4). Thus $\overline{\mathscr{K}}$ is not compact in $\mathscr{W}_p$.

Now the existence of g clearly implies (2.6). Conversely let $R_1 < R_2 < \dots$ such that $\sup_\mu \int_{\|x\|>R_k} \|x\|^p \, \mathrm{d}\mu(x) < k^{-3}$ and $R_k \to \infty$, set $g(R_{k+1}) = k$ and extend it by linear interpolation. Then $\int \|x\|^p g(x) \, \mathrm{d}\mu(x) < R_1^p + \sum k^{-2} < \infty$ for all $\mu \in \mathscr{K}$.

Proof (Proof of Theorem 2.2.7). The first collection is dense by Proposition 2.2.6, and the second collection is larger than the first. Let $\mu = n^{-1} \sum_{i=1}^n \delta\{x_i\}$ be a finitely supported measure with rational weights (with x_i possibly not distinct) and $\varepsilon > 0$. Pick $a_i \in A$ with $\|a_i - x_i\| < \varepsilon$ and set $\nu = n^{-1} \sum_{i=1}^n \delta\{a_i\}$. Then $W_p(\mu, \nu) \le \varepsilon$, and so the third set is also dense. Finally, for any $\sigma > 0$ define μ_σ as the convolution of μ with a uniform measure on a ball of size σ, i.e. with density

$$g(x) = \frac{\sigma^{-d}}{c_d} \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{\|x - x_i\| \le \sigma\}, \qquad c_d = \mathrm{Leb}\left\{x \in \mathbb{R}^d : \|x\| \le 1\right\}.$$

Then μ_σ is absolutely continuous and compactly supported with

$$\mathscr{W}_p^p(\mu_\sigma, \mu) \le \frac{\sigma^{-d}}{c_d} \frac{1}{n} \sum_{i=1}^n \int_{B_\sigma(x_i)} \|x - x_i\|^p \, \mathrm{d}x \le \sigma^p.$$

It follows that the fourth collection is dense too. For the fifth, use the same convolution with a Gaussian measure instead of a uniform one.

Proof (Proof of Proposition 2.2.8). Let (μ_n) be a Cauchy sequence in $\mathscr{W}_p(\mathscr{X})$. It follows from (2.1) that $W_1(\mu,\nu) \le W_p(\mu,\nu)$ for any $\mu,\nu \in P(\mathscr{X})$. Thus (μ_n) is a Cauchy sequence in $\mathscr{W}_1(\mathscr{X})$. In that space the Kantorovich–Rubinstein theorem (1.11) states that

$$W_1(\mu,\nu) = \sup_{\|\varphi\|_{\mathrm{Lip}} \le 1} \left| \int_{\mathscr{X}} \varphi \,\mathrm{d}\mu - \int_{\mathscr{X}} \varphi \,\mathrm{d}\nu \right|, \qquad \|\varphi\|_{\mathrm{Lip}} = \sup_{x \ne y} \frac{|\varphi(x) - \varphi(y)|}{\|x-y\|}.$$

In particular $W_1(\mu,\nu)$ is larger than the *bounded Lipschitz norm*

$$W_1(\mu,\nu) \ge \|\mu - \nu\|_{\mathrm{BL}} = \sup_{\|\varphi\|_{\mathrm{BL}} \le 1} \left| \int_{\mathscr{X}} \varphi \,\mathrm{d}\mu - \int_{\mathscr{X}} \varphi \,\mathrm{d}\nu \right|, \qquad \|\varphi\|_{\mathrm{BL}} = \|\varphi\|_{\mathrm{Lip}} + \|\varphi\|_\infty,$$

which metrises weak convergence in $P(\mathscr{X})$ [4, Theorem 11.3.3]. Thus (μ_n) is a Cauchy sequence with $\|\cdot\|_{\mathrm{BL}}$. Since $P(\mathscr{X})$ is complete with this norm [4, Corollary 11.5.5], (μ_n) converges weakly to $\mu \in P(\mathscr{X})$. If we now fix N, then the lower semicontinuity of the Wasserstein distance (2.5) gives

$$W_p^p(\mu_N,\mu) \le \liminf_{k\to\infty} W_p^p(\mu_N,\mu_k).$$

Since the sequence (μ_n) is Cauchy, the right-hand side vanishes as $N \to \infty$. Thus $W_p(\mu_N,\mu) \to 0$ and completeness is established.

Proof (Proof of Proposition 2.2.9). By Theorem 2.2.7 there exists a compactly supported measure ν with $W_p(\mu,\nu) < \varepsilon/2$, so that $\overline{B}_{\varepsilon/2}(\nu) \subseteq \overline{B}_\varepsilon(\mu)$. We can consequently assume without loss of generality that there exists a compact $K \subset \mathscr{X}$ with $\mu(K) = 1$.

Pick a sequence $x_n \in \mathscr{X}$ of elements that has no partial limits and that are of distance at least $\delta > 0$ from K (i.e. such that $d(x_n,K) = \inf_{x\in K} \|x - x_n\| \ge \delta$ for all n, for instance $\|x_n\| \to \infty$), assume without loss of generality that $\varepsilon < \delta$ and set

$$\mu_n = (1-\alpha_n)\mu + \alpha_n \delta\{x_n\}, \qquad \alpha_n = \varepsilon^p / W_p^p(\mu, \delta\{x_n\}).$$

Then μ_n is a probability measure because

$$W_p^p(\mu,\delta\{x_n\}) = \int_K \|x - x_n\|^p \,\mathrm{d}\mu(x) \ge \delta^p \ge \varepsilon^p,$$

so $\alpha_n \in [0,1]$ for all n. To bound $W_p(\mu_n,\mu)$ observe that we may leave the common $(1-\alpha_n)$ mass in place, so that

$$W_p^p(\mu_n,\mu) \le \alpha_n W_p^p(\mu,\delta\{x_n\}) = \varepsilon^p \qquad \Longrightarrow \quad \mu_n \in \overline{B}_\varepsilon(\mu).$$

We need to show that no subsequence of μ_n can converge in the Wasserstein space. By extracting a subsequence, we may assume that $\alpha_n \to \alpha \in [0,1]$. If (a subsequence of) μ_n converges in the Wasserstein space (or even weakly), then the limit must be $(1-\alpha)\mu + \alpha\delta\{x\}$ with x a limit of (x_n). By the hypothesis on the sequence (x_n), this can only happen if $\alpha = 0$. To finish the proof we only need to show that $W_p(\mu_n, \mu)$ is bounded away from zero.

Clearly $W_p^p(\mu_n, \mu) \geq \alpha_n d^p(x_n, K)$; let us show that this is bounded below. Indeed, let $d_K = \sup_{x,y \in K} \|x - y\|$ be the diameter of K and observe that

$$W_p^p(\mu, \delta\{x_n\}) = \int_K \|x - x_n\|^p \, \mathrm{d}\mu(x) \leq [d(x_n, K) + d_K]^p \leq d^p(x_n, K) \left[1 + \frac{d_K}{\delta}\right]^p,$$

so that

$$\alpha_n d^p(x_n, K) = \frac{\varepsilon^p d^p(x_n, K)}{W_p^p(\mu, \delta\{x_n\})} \geq \frac{\varepsilon^p \delta^p}{(\delta + d_K)^p} > 0.$$

Thus $\alpha = 0$ is impossible too and no subsequence of (μ_n) converges.

Proof (Proof of Lemma 2.3.3). Since G_j is continuous, classical arguments on quantile functions yield $G_j(G_j^{-1}(v)) = v$ for all $v \in (0,1)$, and the same holds for F_j. If μ and ν have the same copula then

$$G(G_1^{-1}(v_1), \ldots, G_d^{-1}(v_d)) = C(v_1, \ldots, v_d) = F(F_1^{-1}(v_1), \ldots, F_d^{-1}(v_d)).$$

If we now change variables and set $v_j = F(x_j)$, then $F(x_1, \ldots, x_d) = G(G_1^{-1}(F_1(x_1)), \ldots, G_d^{-1}(F_d((x_d)))$ for all x_j in the range of F_j^{-1}. Defining now $T_j = G_j^{-1} \circ F_j$, it follows that $\nu = (T_1, \ldots, T_d)\#\mu$, and this map is optimal, hence equals $\mathbf{t}_\mu^\nu$, because the T_j's are nondecreasing.

Conversely, $\mathbf{t}_\mu^\nu$ of the form (2.8) ensures that T_j is nondecreasing, since optimality will be violated otherwise. The push forward constraint of $\mathbf{t}_\mu^\nu$ means that T_j must push the j-th marginal of μ to that of ν; as we have seen in Section 1.5, this entails $T_j = G_j^{-1} \circ F_j$. Consequently for all $v_j \in (0,1)$,

$$C_\nu(v_1, \ldots, v_d) = G(G_1^{-1}(v_1), \ldots, G_d^{-1}(v_d)) = F(F_1^{-1}(v_1), \ldots, F_d^{-1}(v_d)) = C_\mu(v_1, \ldots, v_d).$$

Proof (Proof of Lemma 2.4.2). The lower semicontinuity (2.5) implies that $\mathscr{W}_p$ is a closed set in the weak topology. Since the Wasserstein topology is finer than the induced weak topology, any weak Borel subset of $\mathscr{W}_p$ is Borel in the Wasserstein topology, yielding one implication. For the converse, consider first a Wasserstein closed ball $\mathscr{B} = \{\nu : W_p(\mu, \nu) \leq \varepsilon\}$. If $\nu_n \to \nu$ and ν_n in $\mathscr{B}$, then the lower semicontinuity (2.5) yields that $\nu \in \mathscr{B}$. Thus B is weakly closed, and belongs to the weak Borel σ algebra. Any open Wasserstein ball can be written as a countable union of closed Wasserstein balls, thus also included in the weak Borel σ-algebra. Since $\mathscr{W}_p$ is separable, any open set can be written as a countable union of open balls, so it belongs to the weak Borel σ-algebra. This completes the proof.

Proof (Proof of Theorem 2.4.4). We repeat the proof of the Riesz–Fischer theorem of completeness of L_p spaces. Let f_n be a Cauchy sequence in $\mathscr{L}_p(\mu)$. For each k let n_k be such that $\|f_n - f_m\|_{\mathscr{L}_p(\mu)} < 1/k^2$ if $m, n \geq n_k$. Define $f : \mathscr{X} \to \mathscr{X}$ and $g : \mathscr{X} \to \mathbb{R} \cup \{\infty\}$ by

$$f = f_1 + \sum_{k=1}^{\infty} f_{n_{k+1}} - f_{n_k}, \qquad g(x) = \|f_1(x)\|_{\mathscr{X}} + \sum_{k=1}^{\infty} \|f_{n_{k+1}}(x) - f_{n_k}(x)\|_{\mathscr{X}}.$$

Then $\|f(x)\|_{\mathscr{X}} \le g(x)$ for all $x \in \mathscr{X}$ and

$$\|f\|_{\mathscr{L}_p(\mu)} \le \|g\|_{L_p(\mu)} \le \|f_1\|_{\mathscr{L}_p(\mu)} + \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

This means that for μ-almost every x, $g(x) < \infty$. Since $\mathscr{X}$ is complete, at each such point $f(x)$ is defined and belongs to $\mathscr{X}$. Clearly $\|f(x) - f_{n_k}(x)\|_{\mathscr{X}} \le g(x)$ and $f_{n_k}(x) \to f(x)$ as $k \to \infty$, μ-almost surely. By the dominated convergence theorem $f_{n_k} \to f$ in $\mathscr{L}_p(\mu)$, and since $\{f_n\}$ is Cauchy it follows that $f_n \to f$.

Proof (Proof of Lemma 2.4.6). Suppose that $\Lambda_n \to \Lambda$ in $\mathscr{W}_p(\mathscr{X})$ and fix $\varepsilon > 0$. Define the sets

$$B_n = \{x : \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^{\Lambda}\| \ge \varepsilon\},$$

so that

$$\|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^{\Lambda}\|_{\mathscr{L}_p(\mu)}^p = \int_{\mathscr{X}} \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^{\Lambda}\|^p \,\mathrm{d}\mu \le \varepsilon^p + \int_{B_n} \|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^{\Lambda}\|^p \,\mathrm{d}\mu.$$

Since $\|a-b\|^p \le 2^p\|a\|^p + 2^p\|b\|^p$, the last integral is no larger than

$$2^p \int_{B_n} \|\mathbf{t}_\mu^{\Lambda_n}\|^p \,\mathrm{d}\mu + 2^p \int_{B_n} \|\mathbf{t}_\mu^{\Lambda}\|^p \,\mathrm{d}\mu = 2^p \int_{(\mathbf{t}_\mu^{\Lambda_n})^{-1}(B_n)} \|x\|^p \,\mathrm{d}\Lambda_n(x) + 2^p \int_{(\mathbf{t}_\mu^{\Lambda})^{-1}(B_n)} \|x\|^p \,\mathrm{d}\Lambda(x).$$

Since (Λ_n) and Λ are tight in the Wasserstein space, they must satisfy the absolute uniform continuity (2.7). Let $\delta = \delta_\varepsilon$ as in (2.7). Invoking Villani [9, Corollary 5.23], we see that $\mu(B_n) < \delta$ for all $n > N = N_\varepsilon$. By the measure preserving property of the optimal maps, the last two integrals are taken on sets of measures at most δ. Consequently, for all $n > N_\varepsilon$,

$$\|\mathbf{t}_\mu^{\Lambda_n} - \mathbf{t}_\mu^{\Lambda}\|_{\mathscr{L}_p(\mu)} \le \varepsilon^p + 2^{p+1}\varepsilon,$$

and this completes the proof upon letting $\varepsilon \to 0$.

Proof (Proof of Lemma 2.4.8). If f is a limit of simple functions f_n, and $\mathscr{N}$ is the set on which $f_n(\omega)$ does not converge to f, then $f(\Omega \setminus \mathscr{N})$ is included in the closure of the union of $f_n(\Omega \setminus \mathscr{N})$. This is a countable union of finite sets; hence $f(\Omega \setminus \mathscr{N})$ is separable.

Conversely, let (b_j) be dense in $f(\Omega \setminus \mathscr{N})$. For each n, $f(\Omega \setminus \mathscr{N})$ is included in the countable union $\cup_k B_{1/n}(b_k)$. By the monotone convergence theorem, there exists a finite $M = M(n)$ such that the probability that f is in the first M balls is at least $1 - 1/n$. If we make these balls disjoint ($C_1 = B_{1/n}(b_1); C_{k+1} = B_{1/n}(b_{k+1}) \setminus \cup_{j=1}^k B_{1/n}(b_j)$) and let

$$f_n(\omega) = \sum_{k=1}^{M(n)} b_k \mathbf{1}\{f(\omega) \in C_k\},$$

then f_n is a simple function and $\mathbb{P}(\|f_n - f\| \geq 1/n) < 1/n$, so that $\|f_n - f\| \to 0$ in probability. Consequently, there exists a subsequence f_{n_k} that converges to f almost surely on $\Omega \setminus \mathscr{N}$. Finally, define $g_n(\omega) = f_n(\omega)\mathbf{1}\{\|f_n(\omega)\| \leq 2\|f(\omega)\|\}$. The sequence (g_{n_k}) satisfies the desired properties.

Proof (Proof of Proposition 2.4.9). First we remark that projections are continuous: if $\mathbf{t}$ and $\mathbf{s}$ are random functions in $\mathscr{L}_2(\theta_0)$, then $\langle \mathbf{t}, \mathbf{s} \rangle$ is a random function in $L_2(\theta_0)$. Thus, the integral on the left-hand side of (2.9) is a random variable, and so the expectation is taken on $\mathbb{R}$. In the middle integral, the expectation is the Bochner expectation of the random element $\left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle$ in $L_2(\theta_0)$. The expectation on the right-hand side of (2.9) is the Bochner expectation of the random element $\mathbf{t}_{\theta_0}^{\Lambda}$ in $\mathscr{L}_2(\theta_0)$.

Suppose initially that Λ is a simple function, that is

$$\Lambda(\omega) = \sum_{j=1}^{n} \lambda_j \mathbf{1}\{\omega \in \Omega_j\}, \qquad \lambda_j \in \mathscr{W}_2(\mathscr{X}); \qquad \Omega = \bigcup_{j=1}^{n} \Omega_j.$$

If we let $\alpha_j = \mathbb{P}\Omega_j$, then equation (2.9) states that

$$\sum_{j=1}^{n} \alpha_j \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \sum_{j=1}^{n} \alpha_j \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 = \int_{\mathscr{X}} \left\langle \sum_{j=1}^{n} \alpha_j \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0,$$

which is true by linearity and by finiteness of each of the summands:

$$\int_{\mathscr{X}} \left| \left\langle \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right| \mathrm{d}\theta_0 \leq \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\lambda_j} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0} \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0} = W_2(\theta_0, \lambda_j) W_2(\theta_0, \theta) < \infty.$$

Now suppose that Λ is measurable and $\mathbb{E}W_2(\Lambda, \delta_0) < \infty$. Since $\mathscr{X}$ is separable, the Wasserstein space $\mathscr{W}_2(\mathscr{X})$ is separable too, so $\Lambda(\Omega)$ is separable. But it has been shown that $\Lambda \mapsto \mathbf{t}_{\theta_0}^{\Lambda}$ is continuous from $\mathscr{W}_2(\mathscr{X})$ to $\mathscr{L}_2(\theta_0)$ (Lemma 2.4.6). Consequently, $\mathbf{t}_{\theta_0}^{\Lambda}(\Omega)$ is separable, and by Lemma 2.4.8 there exists a sequence of simple functions $\mathbf{t}_n(\omega)$ that converge to $\mathbf{t}_{\theta_0}^{\Lambda}(\omega)$ for almost every ω and $\|\mathbf{t}_n\|_{\mathscr{L}_2(\theta_0)} \leq 2\|\mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)}$. We may assume without loss of generality that $\mathbf{t}_n$ are optimal maps: indeed, define the simple random measures $\Lambda_n = \mathbf{t}_n \# \theta_0$. Then

$$\|\mathbf{t}_{\theta_0}^{\Lambda_n}\|_{\mathscr{L}_2(\theta_0)} = W_2(\Lambda_n, \delta_0) = \|\mathbf{t}_n\|_{\mathscr{L}_2(\theta_0)},$$

and $\Lambda_n(\omega) \to \Lambda(\omega)$ by Lemma 2.4.5, so $\mathbf{t}_{\theta_0}^{\Lambda_n(\omega)} \to \mathbf{t}_{\theta_0}^{\Lambda(\omega)}$ almost surely by Lemma 2.4.6. Thus, $\mathbf{t}_n$ can be replaced by $\mathbf{t}_{\theta_0}^{\Lambda_n}$.

As (2.9) has been established for Λ_n, it suffices to show that each expression of (2.9) equals the limit as $n \to \infty$ of the same expression with Λ replaced by Λ_n.

We begin with the right-hand side. Since for all $\omega \in \Omega$

$$\sup_n \|\mathbf{t}_{\theta_0}^{\Lambda_n(\omega)}\|_{\mathscr{L}_2(\theta_0)} \leq 2\|\mathbf{t}_{\theta_0}^{\Lambda(\omega)}\|_{\mathscr{L}_2(\theta_0)} = 2W_2(\Lambda(\omega), \delta_0),$$

and the latter is integrable, it follows from the dominated convergence theorem that $\mathbb{E}\|\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)} \to 0$ and the definition of the Bochner integral implies that

$$\mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} \to \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} \qquad \text{in } \mathscr{L}_2(\theta_0).$$

Consequently

$$\int_{\mathscr{X}} \left| \left\langle \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right| \mathrm{d}\theta_0 \leq \sqrt{\int_{\mathscr{X}} \left\| \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda} \right\|^2 \mathrm{d}\theta_0} \sqrt{\int_{\mathscr{X}} \left\| \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\|^2 \mathrm{d}\theta_0}$$

vanishes as $n \to \infty$, since $\|\mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbb{E}\mathbf{t}_{\theta_0}^{\Lambda}\|_{\mathscr{L}_2(\theta_0)} \to 0$.

Next we deal with the middle integral of (2.9). We have by continuity of the projections that almost surely

$$\left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \to \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \qquad \text{in } L_2(\theta_0), \qquad n \to \infty,$$

and as before

$$\sup_n \left\| \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right\|_{L_2(\theta)} \leq W_2(\theta_0, \theta) \sup_n \left\| \|\mathbf{t}_{\theta_0}^{\Lambda_n}\|_{\mathscr{X}} \right\|_{L_2(\theta)} \leq 2W_2(\theta_0, \theta) \left\| \mathbf{t}_{\theta_0}^{\Lambda} \right\|_{\mathscr{L}_2(\theta_0)}$$

so again the dominated convergence theorem gives

$$\mathbb{E} \left\| \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle - \left\langle \mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \right\|_{L_2(\theta_0)} \to 0, \qquad n \to \infty.$$

Of course the same holds if we subtract the identity from $\mathbf{t}_{\theta_0}^{\Lambda_n}$ and $\mathbf{t}_{\theta_0}^{\Lambda}$. The definition of the Bochner integral means that

$$\mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \to \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \qquad \text{in } L_2(\theta),$$

which of course implies

$$\int_{\mathscr{X}} \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0 \to \int_{\mathscr{X}} \mathbb{E} \left\langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0, \qquad n \to \infty.$$

Lastly we treat the left-hand side of (2.9). Define the random variables

$$Y_n = \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda_n}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0, \qquad Y = \int_{\mathscr{X}} \left\langle \mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i} \right\rangle \mathrm{d}\theta_0.$$

Then again

$$\sup_n |Y_n| \leq 2W_2(\theta_0, \theta) \left\| \mathbf{t}_{\theta_0}^{\Lambda} \right\|_{\mathscr{L}_2(\theta_0)}$$

and

$$|Y_n - Y| \leq W_2(\theta_0, \theta) \|\mathbf{t}_{\theta_0}^{\Lambda_n} - \mathbf{t}_{\theta_0}^{\Lambda}\|_{L_2(\theta_0)} \to 0, \qquad n \to \infty,$$

so the dominated convergence theorem applies and $\mathbb{E}Y_n \to \mathbb{E}Y$.

Chapter 3
Proofs for Chapter 3

Proof (Proof of Theorem 3.1.5). **Step 1:** tightness of $(\overline{\mu}_k)$. The entire collection $\mathscr{K} = \{\mu_k^i\}$ is tight, since all the sequences converge in distribution (Theorem 2.2.1). For any $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset \mathscr{X}$ such that $\mu(K_\varepsilon) \geq 1 - \varepsilon/N$ for all $\mu \in \mathscr{K}$.

Let π_k be any multicoupling of $(\mu_k^1, \dots, \mu_k^N)$. Then the marginal constraints of π_k imply that $\pi_k(K_\varepsilon^N) \geq 1 - \varepsilon$. By Proposition 3.1.2, $\overline{\mu}_k$ must take the form $M\#\pi_k$ for some multicoupling π_k. Since M is continuous, $K'_\varepsilon = M(K_\varepsilon^N)$ is compact and by definition

$$\overline{\mu}_k(K'_\varepsilon) = \pi_k(M^{-1}(M(K_\varepsilon^N))) \geq \pi_k(K_\varepsilon^N) \geq 1 - \varepsilon,$$

for all k. This proves tightness, and we may assume that $\overline{\mu}_k \to \overline{\mu}$ weakly.

Step 2: $W_2(\overline{\mu}_k, \overline{\mu}) \to 0$. We shall show that (π_k) stays in a compact subset of $\mathscr{W}(\mathscr{X}^N)$, where $\mathscr{X}^N$ is endowed with the Hilbert norm $\|x\|^2 = \sum \|x_i\|^2$. If $\|x\|^2 > R$, then $\|x_i\|^2 > R/N$ for some i, so

$$\int_{\|x\| > R'} \|x\|^2 \, \mathrm{d}\pi_k(x) \leq \sum_{i,j=1}^N \int_{\|x_j\| > R'/\sqrt{N}} \|x_i\|^2 \, \mathrm{d}\pi_k(x).$$

We need to show that each summand vanishes, uniformly over k, as $R' \to \infty$; equivalently, when $R = R'/\sqrt{N} \to \infty$. When $i = j$,

$$\sup_k \int_{\|x_i\| > R} \|x_i\|^2 \, \mathrm{d}\pi_k(x) = \sup_k \int_{\|x_i\| > R} \|x_i\|^2 \, \mathrm{d}\mu_k^i(x_i) \to 0, \qquad R \to \infty,$$

since μ_k^i converge in $\mathscr{W}_2(\mathscr{X})$ (Theorem 2.2.1). For $i \neq j$ we use the formula

$$\int_{\|x_j\| > R} \|x_i\|^2 \, \mathrm{d}\pi_k(x) = \int_0^\infty \pi_k(\{x : \|x_i\|^2 \mathbf{1}\{\|x_j\| > R\} > t\}) \, \mathrm{d}t$$

The integrand is smaller than $\sup_k \pi_k(\{x : \|x_j\| > R\}) = \sup_k \mu_k^j(\{\|x_j\| > R\}) = f_j(R) \to 0$, since (μ_k^j) is tight.

Moreover, for any random variable X

$$\mathbb{E}X^2\mathbf{1}\{X>A\}=A^2\mathbb{P}(X>A)+\int_{A^2}^{\infty}\mathbb{P}(X^2>t)\,\mathrm{d}t\geq\int_{A^2}^{\infty}\mathbb{P}(X^2>t)\,\mathrm{d}t,$$

so that the last displayed integral is no larger than

$$\leq\int_0^{1/\sqrt{f_j(R)}}+\int_{1/\sqrt{f_j(R)}}^{\infty}\leq\sqrt{f_j(R)}+\sup_k\int_{\|x_i\|>1/\sqrt[4]{f_j(R)}}\|x_i\|^2\,\mathrm{d}\mu_k^i(x_i),$$

which vanishes as $R\to\infty$ because $f_j(R)\to 0$ and μ_k^i converge in $\mathscr{W}_2(\mathscr{X})$.

This shows that the entire collection of multicouplings stays in a compact set of $\mathscr{W}_2(\mathscr{X}^N)$, and since

$$\sup_k\int_{\|M(x)\|>R}\|M(x)\|^2\,\mathrm{d}\pi_k(x)=\sup_k\int_{\|\overline{x}\|>R}\|\overline{x}\|^2\,\mathrm{d}\pi_k(x)$$

vanishes as $R\to\infty$ and $\overline{\mu}_k=M\#\pi_k$, we deduce that $\overline{\mu}_k$ stay in a compact set of $\mathscr{W}_2(\mathscr{X})$.

Step 3: a moment bound for $\overline{\mu}_k$. Let $R^i=\int_{\mathscr{X}}\|x\|^2\,\mathrm{d}\mu^i(x)$ denote the second moment of μ^i. Since the second moments can be interpreted as a (squared) Wasserstein distance to the Dirac mass at 0, the second moment of μ_k^i converges to R^i and so for k large it is smaller than R^i+1. By Corollary 3.1.3, for k large

$$\int_{\mathscr{X}}\|x\|^2\,\mathrm{d}\overline{\mu}_k(x)\leq\frac{1}{N}\sum_{i=1}^{N}R^i+1\leq\max(R^1,\dots,R^N)+1:=R+1.$$

Consequently, the Fréchet means $\overline{\mu}_k$ can be found (for k large) in the Wasserstein ball

$$B=\{\mu\in\mathscr{W}_2(\mathscr{X}):W^2(\mu,\delta_0)\leq R+1\},$$

with δ_0 a Dirac measure at the origin. Let F_k denote the Fréchet functional corresponding to $(\mu_k^1,\dots,\mu_k^N)$. If $\mu,\nu\in B$ then, since $\mu_k^i\in B$ for k large,

$$|F_k(\mu)-F_k(\nu)|\leq\frac{1}{2N}\sum_{i=1}^{N}[W_2(\mu,\mu_k^i)+W_2(\nu,\mu_k^i)]W_2(\mu,\nu)\leq 2\sqrt{R+1}\,W_2(\mu,\nu).$$

In other words, all the F_k's are uniformly Lipschitz on B. Suppose now that $\overline{\mu}_k\to\overline{\mu}$ in $\mathscr{W}_2(\mathscr{X})$. Let $\mu\in B$, $\varepsilon>0$ and k_0 such that $W_2(\overline{\mu}_k,\overline{\mu})<\varepsilon/(2\sqrt{R+1})$ for all $k\geq k_0$. Since $F_k\to F$ pointwise we may assume that $|F(\mu)-F_k(\mu)|<\varepsilon$ when $k\geq k_0$ and the same holds for $\mu=\overline{\mu}$. Then for all $k\geq k_0$

$$\varepsilon+F(\mu)\geq F_k(\mu)\geq F_k(\overline{\mu}_k)\geq F_k(\overline{\mu})-\varepsilon\geq F(\overline{\mu})-2\varepsilon.$$

Since $\varepsilon>0$ and $\mu\in B$ are arbitrary, we conclude that $\overline{\mu}$ minimises F over B and hence over the entire Wasserstein space $\mathscr{W}_2(\mathscr{X})$.

Proof (Proof of Lemma 3.1.11). Agueh & Carlier [1, Proposition 3.8] show that there exist convex lower semicontinuous potentials ψ_i^* on $\mathbb{R}^{d_1}$ and φ_i^* on $\mathbb{R}^{d_2}$ whose gradients push μ forward to μ^i and ν to ν^i respectively, and such that

$$\frac{1}{N}\sum_{i=1}^{N}\psi_i(x) \le \frac{1}{2}\|x\|^2, \quad x \in \mathbb{R}^{d_1}; \qquad \frac{1}{N}\sum_{i=1}^{N}\varphi_i(y) \le \frac{1}{2}\|y\|^2, \quad y \in \mathbb{R}^{d_2},$$

with equality μ- and ν-almost surely respectively. Define the convex function $\phi_i : \mathbb{R}^{d_1+d_2} \to \mathbb{R}\cup\{\infty\}$ by $\phi_i(x,y) = \psi_i(x) + \varphi_i(y)$. Then the gradient of

$$\phi_i^*(x,y) = \sup_{u,v}\langle x,u\rangle + \langle y,v\rangle - \psi_i(u) - \varphi_i(v) = \psi_i^*(x) + \varphi_i^*(y)$$

pushes $\mu^i \otimes \nu^i$ forward to $\mu \otimes \nu$ and

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i(x,y) \le \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 = \frac{1}{2}\|(x,y)\|^2, \quad (x,y) \in \mathbb{R}^{d_1+d_2},$$

with equality $\mu\otimes\nu$-almost surely. By Proposition 3.1.10, $\mu\otimes\nu$ is the Fréchet mean of $(\mu^i\otimes\nu^i)$.

Proof (Proof of Lemma 3.1.12). If $x \mapsto \phi(x)$ is convex, then $x \mapsto \phi(U^{-1}x)$ is convex with gradient $U\nabla\phi(U^{-1}x)$ at (almost all) x and conjugate $x \mapsto \phi^*(U^{-1}x)$. If ϕ_i are convex potentials with $\nabla\phi_i^*\#\mu = \mu^i$, then $\nabla[(\phi_i\circ U^{-1}]^*) = \nabla(\phi_i^*\circ U^{-1})$ pushes $U\#\mu$ forward to $U\#\mu^i$ and by [1, Proposition 3.8]

$$\frac{1}{N}\sum_{i=1}^{N}(\varphi_i\circ U^{-1})(Ux) = \frac{1}{N}\sum_{i=1}^{N}\phi_i(x) \le \frac{1}{2}\|x\|^2 = \frac{1}{2}\|Ux\|^2$$

with equality for μ-almost any x. A change of variables $y = Ux$ shows that the set of points y such that $\sum(\varphi_i\circ U^{-1})(y) < N\|y\|^2/2$ is $(U\#\mu)$-negligible, completing the proof.

Proof (Proof of Proposition 3.2.4). Let $\mathrm{proj}_K : \mathscr{X} \to K$ denote the projection onto the set K, which is well-defined since K is closed and convex, and of course satisfies

$$\|x-y\| \ge \|x - \mathrm{proj}_K(y)\|, \qquad x \in K, \quad y \in \mathscr{X},$$

with equality if and only if $y \in K$. Let $\pi \in \Pi(\Lambda,\gamma)$ be optimal. By the hypothesis $\Lambda(K) = 1$, so that the above inequality holds for Λ-almost every x and all y, hence π-almost surely. Define the projection $\gamma_K = \mathrm{proj}_K\#\gamma$ of γ onto K, and recall that $\mathbf{i}$ denotes the identity mapping on $\mathscr{X}$. Then $(\mathbf{i}\times\mathrm{proj}_K)\#\pi \in \Pi(\Lambda,\gamma_K)$ and

$$W_2^2(\Lambda,\gamma) = \int_{K\times\mathscr{X}}\|x-y\|^2\,\mathrm{d}\pi(x,y) \ge \int_{K\times\mathscr{X}}\|x-\mathrm{proj}_K(y)\|^2\,\mathrm{d}\pi(x,y) \ge W_2^2(\Lambda,\gamma_K).$$

Taking expectations gives $F(\gamma) \ge F(\mathrm{proj}_K\#\gamma)$. Again equality holds if and only if $\gamma(K) = 1$, in which case $\mathrm{proj}_K\#\gamma = \gamma$, which completes the proof.

Proof (Proof of Corollary 3.2.10). We shall prove the assertion under the assumption that there exists a divergent function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $r^2g(r)$ is convex and

$$\text{with probability one} \int_{\mathbb{R}^d} \|x\|^2 g(\|x\|)\, \mathrm{d}\Lambda(x) \leq M.$$

Since $g \to \infty$, this implies that Λ takes values in a compact subset $\mathscr{K}$ of $\mathscr{W}_2(\mathbb{R}^d)$ (Proposition 2.2.3). Conversely, most divergent g satisfy the convexity condition (at least for r large, and they can be modified for r small to preserve convexity throughout $\mathbb{R}_+$). Examples of g include r^ε or $\log(r+1)$ or $\log\log(r+10)$ or $\log\log\log(r+100)$ and so on, so this assumption is not far from requiring the law of Λ to be supported on a compact subset of $\mathscr{W}_2(\mathbb{R}^d)$.

It follows from convexity that (see the proof of Corollary 3.1.3) for all n,

$$\int_{\mathbb{R}^d} \|x\|^2 g(\|x\|)\, \mathrm{d}\overline{\mu}_n(x) \leq \sum_{i=1}^n \int_{\mathbb{R}^d} \|x\|^2 g(\|x\|)\, \mathrm{d}\Lambda_i(x) \leq M.$$

Since balls in $\mathbb{R}^d$ are compact, this implies that (μ_n) is tight, and $g \to \infty$ implies that it is tight also in $\mathscr{W}_2(\mathbb{R}^d)$. Now conclude as in Step 3 of the proof of Theorem 3.1.5 (see page 17 in this document): by Corollary 3.1.3

$$W_2^2(\overline{\mu}_n, \delta_0) = \int_{\mathbb{R}^d} \|x\|^2\, \mathrm{d}\overline{\mu}_n(x) \leq \sup_{\Lambda} \int_{\mathbb{R}^d} \|x\|^2\, \mathrm{d}\Lambda(x) := R < \infty$$

since Λ stays in the compact set $\mathscr{K}$. The empirical Fréchet functionals F_n are uniformly $2\sqrt{R}$-Lipschitz on the ball $B = \{\mu : W_2^2(\mu, \delta_0) \leq R\}$ and, consequently, they converge to F uniformly on $\mathscr{K} \cup \{\lambda\}$, where λ is the (assumed unique) Fréchet mean of Λ. Any limit of $\overline{\mu}_n$ must therefore be a minimiser, and the proof is complete.

Proof (Proof of Theorem 3.2.11). Let X be a random element in a separable Hilbert space $\mathscr{X}$ with $\mathbb{E}\|X\|^2 < \infty$. Let (e_j) be an orthonormal basis of $\mathscr{X}$ and define

$$x_0 = \sum_{j=1}^\infty e_j \mathbb{E}\left\langle X, e_j \right\rangle.$$

Then $x_0 \in \mathscr{X}$ because

$$\sum_{j=1}^\infty [\mathbb{E}\left\langle X, e_j \right\rangle]^2 \leq \sum_{j=1}^\infty \mathbb{E}[\left\langle X, e_j \right\rangle]^2 = \mathbb{E} \sum_{j=1}^\infty [\left\langle X, e_j \right\rangle]^2 = \mathbb{E}\|X\|^2 < \infty.$$

By definition $\mathbb{E}\left\langle X - x_0, e_j \right\rangle = 0$ for all j, so $\mathbb{E}\langle X, x\rangle = \langle x_0, x\rangle$ for all $x \in \mathscr{X}$, and

$$\mathbb{E}\|X - x\|^2 - \mathbb{E}\|X - x_0\|^2 = \mathbb{E}\langle 2X - x - x_0, x_0 - x\rangle = \|x_0 - x\|^2.$$

Therefore the Fréchet mean of F_Λ^{-1} in $L_2(0,1)$ is the unique element g satisfying

$$\mathbb{E}\int_0^1 F_\Lambda^{-1}(t) f(t)\, \mathrm{d}t = \mathbb{E}\left\langle F_\Lambda^{-1}, f\right\rangle = \langle g, f\rangle = \int_0^1 g(t) f(t)\, \mathrm{d}t, \qquad f \in L_2(0,1).$$

We need to show that g can be identified with a quantile function, that is, a nondecreasing left-continuous function, and that $g(t) = \mathbb{E}F_\Lambda^{-1}(t)$ for all $t \in (0,1)$. Although F_Λ^{-1} is an element of L_2 where pointwise evaluations are undefined, the left-continuity allows us to define them in a measurable way. Taking F_Λ^{-1} to be the unique left-continuous element in the equivalence class, observe that for any $t \in (0,1)$, the quantity

$$\lim_{m\to\infty} m\int_{t-1/m}^{t} F_\Lambda^{-1}(u)\,\mathrm{d}u = \lim_{m\to\infty} m\left\langle F_\Lambda^{-1}, 1[t-1/m,t]\right\rangle$$

is a random variable (measurable from $(\Omega,\mathscr{F},\mathbb{P})$ to $\mathbb{R}$), and the limit exists and equals $F_\Lambda^{-1}(t)$ by left-continuity. Now observe that since F_Λ^{-1} is nondecreasing,

$$\|F_\Lambda^{-1}\|^2 = \int_0^1 |F_\Lambda^{-1}(s)|^2\,\mathrm{d}s \geq \begin{cases} \int_t^1 |F_\Lambda^{-1}(t)|^2\,\mathrm{d}s & F_\Lambda^{-1}(t) \geq 0 \\ \int_0^t |F_\Lambda^{-1}(t)|^2\,\mathrm{d}s & F_\Lambda^{-1}(t) \leq 0 \end{cases} \geq \min(t,1-t)|F_\Lambda^{-1}(t)|^2,$$

so $\mathbb{E}|F_\Lambda^{-1}(t)|^2 < \infty$ for all t. From this we obtain that F_Λ^{-1} is locally bounded in expectation: for all $0 < s < t < 1$

$$\mathbb{E}\sup_{x\in[s,t]} |F_\Lambda^{-1}(x)|^2 = \mathbb{E}[\max(|F_\Lambda^{-1}(s)|,|F_\Lambda^{-1}(t)|)]^2 < \infty.$$

Since $m1[t-1/m,t]$ integrates to unity, we have for all $m > 2/t$,

$$|m\left\langle F_\Lambda^{-1}, 1[t-1/m,t]\right\rangle| \leq \sup_{x\in[t-1/m,t]} |F_\Lambda^{-1}(x)| \leq \sup_{x\in[t/2,t]} |F_\Lambda^{-1}(x)|$$

and this is integrable, so the dominated convergence theorem gives

$$\mathbb{E}F_\Lambda^{-1}(t) = \lim_{m\to\infty} m\mathbb{E}\left\langle F_\Lambda^{-1}, 1[t-1/m,t]\right\rangle = \lim_{m\to\infty} m\left\langle g, 1[t-1/m,t]\right\rangle.$$

Moreover Fubini's theorem yields

$$\mathbb{E}\int_a^b F_\Lambda^{-1}(t)\,\mathrm{d}t = \int_a^b \mathbb{E}F_\Lambda^{-1}(t)\,\mathrm{d}t$$

for all $0 < a < b < 1$. (The quantity $F_\Lambda^{-1}(t)$ is jointly measurable from $\Omega \times [a,b]$ because $F_\Lambda^{-1}(t) \leq x$ if and only if $t \leq F_\Lambda(x)$, and the set of Λ such that $F_\Lambda(x) \leq \alpha$ is closed (Remark 3.2.5, so $F_\Lambda(x)$ is measurable for all x.) Since linear combinations of indicators of intervals are dense in $L_2(0,1)$, we get $g(t) = \mathbb{E}F_\Lambda^{-1}(t)$ for all $t \in (0,1)$.

Using this, one can easily deduce the desired properties. If $s < t$, then $F_\Lambda^{-1}(s) \leq F_\Lambda^{-1}(t)$, so $\mathbb{E}F_\Lambda^{-1}(s) \leq \mathbb{E}F_\Lambda^{-1}(t)$. This implies that the sequence is nondecreasing in m. To prove left-continuity, fix $t \in (0,1)$ and $\varepsilon > 0$. Pick m such that

$$m\int_{t-1/m}^{t} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u = m\left\langle \mathbb{E}F_\Lambda^{-1}, 1[t-1/m,t]\right\rangle \geq \mathbb{E}F_\Lambda^{-1}(t) - \varepsilon/2,$$

and then $\delta < 1/m$ such that

$$m\int_{t-1/m}^{t-\delta} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge m\int_{t-1/m}^{t} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u - \varepsilon/2 \ge \mathbb{E}F_\Lambda^{-1}(t) - \varepsilon,$$

which exists because the integral converges. If $s \in (t-\delta, t)$ then using the monotonicity of $\mathbb{E}F_\Lambda^{-1}$ again and noticing that $s - 1/k > t - \delta$ for k large, we obtain

$$\mathbb{E}F_\Lambda^{-1}(s) = \lim_{k\to\infty} k\int_{s-1/k}^{s} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge \mathbb{E}F_\Lambda^{-1}(t-\delta) \ge m\int_{t-1/m}^{t-\delta} \mathbb{E}F_\Lambda^{-1}(u)\,\mathrm{d}u \ge \mathbb{E}F_\Lambda^{-1}(t) - \varepsilon,$$

This proves that $\mathbb{E}F_\Lambda^{-1}$ can be viewed as a quantile function.

Proof (Proof of Proposition 3.2.12). We repeat the calculations of Ambrosio et al. [3, Theorem 10.2.2 and Proposition 10.2.6] for the particular case $p = 2$. Define a three-coupling $\mu = (\mathbf{i}, \mathbf{t}_{\theta_0}^{\Lambda}, \mathbf{t}_{\theta_0}^{\theta})\#\theta_0 \in P(\mathscr{X}^3)$ and notice that its relevant projections are optimal couplings of (θ_0, Λ) and (θ_0, θ) but not necessarily of (Λ, θ). By definition

$$\int_{\mathscr{X}} \langle \mathbf{t}_{\theta_0}^{\Lambda} - \mathbf{i}, \mathbf{t}_{\theta_0}^{\theta} - \mathbf{i}\rangle \,\mathrm{d}\theta_0 = \int_{\mathscr{X}^3} \langle x_2 - x_1, x_3 - x_1\rangle \,\mathrm{d}\mu; \quad W_2^2(\theta_0, \Lambda) = \int_{\mathscr{X}^3} \|x_2 - x_1\|^2 \,\mathrm{d}\mu;$$
$$\delta^2 = W_2^2(\theta_0, \theta) = \int_{\mathscr{X}^3} \|x_1 - x_3\|^2 \,\mathrm{d}\mu; \quad W_2^2(\theta\ ,\Lambda) \le \int_{\mathscr{X}^3} \|x_2 - x_3\|^2 \,\mathrm{d}\mu.$$

Integrating the equality

$$\frac{1}{2}\|x_2 - x_3\|^2 - \frac{1}{2}\|x_2 - x_1\|^2 + \langle x_2 - x_1, x_3 - x_1\rangle = \frac{1}{2}\|x_1 - x_3\|^2 \tag{3.1}$$

with respect to μ yields the second inequality of the proposition. If the measures are compatible then the relevant marginal of μ is optimal for (Λ, θ), and the inequality holds as equality.

For the other inequality, let β be another three-coupling that optimally couples (θ_0, θ) and (Λ, θ). Then

$$W_2^2(\theta, \Lambda) = \int_{\mathscr{X}^3} \|x_2 - x_3\|^2 \,\mathrm{d}\beta \quad \text{and} \quad W_2^2(\theta_0, \Lambda) \le \int_{\mathscr{X}^3} \|x_1 - x_2\|^2 \,\mathrm{d}\beta.$$

Integration of (3.1) with respect to β yields

$$\frac{1}{2}W_2^2(\theta, \Lambda) - \frac{1}{2}W_2^2(\theta_0, \Lambda) \ge \frac{1}{2}\delta^2 - \int_{\mathscr{X}^3} \langle x_2 - x_1, x_3 - x_1\rangle \,\mathrm{d}\beta.$$

All that remains is to bound the last displayed integral by a constant times δ, when the integral is taken with respect to either β or μ. To this end, we apply the Cauchy–Schwarz inequality

$$\left|\int_{\mathscr{X}^3}\langle x_2-x_1,x_3-x_1\rangle\,\mathrm{d}\mu\right|\le\sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\mu}\sqrt{\int_{\mathscr{X}^3}\|x_3-x_1\|^2\,\mathrm{d}\mu}=\delta W_2(\theta_0,\Lambda),$$
$$\left|\int_{\mathscr{X}^3}\langle x_2-x_1,x_3-x_1\rangle\,\mathrm{d}\beta\right|\le\sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\beta}\sqrt{\int_{\mathscr{X}^3}\|x_3-x_1\|^2\,\mathrm{d}\beta}$$

where the last displayed square root again equals δ, and

$$\sqrt{\int_{\mathscr{X}^3}\|x_2-x_1\|^2\,\mathrm{d}\beta}\le\sqrt{\int_{\mathscr{X}^3}2\|x_1\|^2\,\mathrm{d}\beta+\int_{\mathscr{X}^3}2\|x_2\|^2\,\mathrm{d}\beta}=\sqrt{2W_2^2(\theta_0,\delta_0)+2W_2^2(\Lambda,\delta_0)}.$$

This completes the proof.

Proof. Proof of Proposition 3.2.14] Suppose that $F'(\gamma)\neq 0$ and define $\mathbf{t}=\mathbb{E}\mathbf{t}_\gamma^\Lambda$ and $W=\mathbf{t}-\mathbf{i}$. If we show that actually $\mathbf{t}=\mathbf{t}_\gamma^{\mathbf{t}\#\gamma}$, i.e. that $\mathbf{t}$ is optimal, then the result will follow immediately. Indeed, if we set $\nu_s=[\mathbf{i}+s(W-\mathbf{i})]\#\gamma$, then $W_2(\nu_s,\gamma)=s\|W\|_{\mathscr{L}_2(\gamma)}$ for $s\in[0,1]$ and by Theorem 3.2.13,

$$0=\lim_{s\to0^+}\frac{F(\nu_s)-F(\gamma)+\int_{\mathscr{X}}\langle W(x),sW(x)\rangle\,\mathrm{d}\gamma(x)}{s\|W\|_{\mathscr{L}_2(\gamma)}}=\lim_{s\to0^+}\frac{F(\nu_s)-F(\gamma)}{s\|W\|_{\mathscr{L}_2(\gamma)}}+\|W\|_{\mathscr{L}_2(\gamma)}.$$

Since $\|W\|_{\mathscr{L}_2(\gamma)}>0$, this means that $F(\nu_s)-F(\gamma)$ is negative when s is small, and therefore γ cannot be the Fréchet mean.

Let us now show the optimality of $\mathbf{t}$. If Λ is a simple random measure, then the result follows immediately. Otherwise, there exists a sequence of simple optimal maps $\mathbf{t}_n$ that converge to $\mathbf{t}$ in $\mathscr{L}_2(\gamma)$ (see the proof of Proposition 2.4.9). Let us show that $\mathbf{t}$ is monotone. There exists a set B with $\gamma(B)=1$ such that

$$\langle\mathbf{t}_n(y)-\mathbf{t}_n(x),y-x\rangle\ge0,\qquad x,y\in B\quad n=1,2,\dots.$$

Fix an integer k, let $R=R_k$ such that $\gamma[B_R(0)]\ge1-1/k$ and define $D_k\subseteq\mathscr{X}^2$ by

$$D_k=\{(x,y):x,y\in B\cap B_R(0),\quad\langle\mathbf{t}(y)-\mathbf{t}(x),y-x\rangle<-2/k\}.$$

If $(x,y)\in D_k$ then either

$$\|\mathbf{t}_n(x)-\mathbf{t}(x)\|\ge\frac{1/k}{\|x-y\|}\ge\frac{1/k}{2R}$$

or that same lower bound holds for $\|\mathbf{t}_n(y)-\mathbf{t}(y)\|$. By Markov's inequality and since $\|\mathbf{t}_n-\mathbf{t}\|_{\mathscr{L}_2(\gamma)}\to0$, when $n\ge N_k$ is large enough this happens with γ measure at most $1/k$. Define

$$B_k=B\cap B_R(0)\cap\{x:\|\mathbf{t}_n(x)-\mathbf{t}(x)\|\le1/(2Rk)\},\qquad n=N_k.$$

Then $\gamma(B_k)\ge1-2/k$ and

$$\langle\mathbf{t}(y)-\mathbf{t}(x),y-x\rangle\ge-\frac{2}{k},\qquad x,y\in B_k.$$

If we now set

$$B' = \cap_{j=1}^{\infty} \cup_{k=j}^{\infty} B_k,$$

then $\gamma(B') = 1$ and $\langle \mathbf{t}(y) - \mathbf{t}(x), y - x \rangle \geq 0$ for all $x, y \in B'$. Similarly one shows that $\mathbf{t}$ is cyclically monotone, that is, the measure $\pi = (\mathbf{i}, \mathbf{t})\#\gamma$ is cyclically monotone, hence optimal by Proposition 1.7.5.

Chapter 4
Proofs for Chapter 4

Proof (Proof of Theorem 4.2.4). We need to show the following claims:

1. Λ is measurable (so that the expectations in the statement of the theorem are well-defined).
2. $\mathbb{E}\phi(x) = \langle \mathbb{E}\mathbf{t}(x), x\rangle /2 = \|x\|^2/2$ for all x.
3. For any $\theta \in \mathscr{W}_2(\mathscr{X})$ Fubini's identity holds:
$$\mathbb{E}\int_{\mathscr{X}} \phi(x)\,\mathrm{d}\theta(x) = \int_{\mathscr{X}} \mathbb{E}\phi(x)\,\mathrm{d}\theta(x), \qquad \theta \in \mathscr{W}_2.$$
4. The dual integrals
$$\int_{\mathscr{X}} \left(\frac{1}{2}\|y\|^2 - \phi^*(y)\right) \mathrm{d}\Lambda(y)$$
are measurable real-valued quantities with a well-defined expectation.

For all $p, q \geq 1$ we have

$$W_p^p(\mathbf{t}\#\theta, \mathbf{s}\#\theta) \leq \int_{\mathscr{X}} \|\mathbf{t}(x) - \mathbf{s}(x)\|^p\,\mathrm{d}\theta(x) \leq \|\mathbf{t}-\mathbf{s}\|_{G_q}^p \int_{\mathscr{X}} (1+\|x\|^q)^p\,\mathrm{d}\theta(x).$$

Thus if $\theta \in \mathscr{W}_{pq}(\mathscr{X})$ then the map $\mathbf{t} \mapsto \mathbf{t}\#\theta$ from G_q to $\mathscr{W}_p$ is Lipschitz continuous. When $q = 1$ we obtain in particular

$$W_p(\mathbf{t}\#\theta, \mathbf{s}\#\theta) \leq \|\mathbf{t}-\mathbf{s}\|_{G_1} \left(\int_{\mathscr{X}} (1+\|x\|)^p\,\mathrm{d}\theta(x)\right)^{1/p}.$$

For $p = 2$, this establishes claim 1. Moreover, the Fréchet functional is seen to be finite, as it is finite for λ; choose $\theta = \lambda$ and $\mathbf{s} = \mathbf{i}$.

Since $\mathbf{t}$ is continuous, ϕ can be recovered as the line integral

$$\phi(x) = \int_0^1 \langle \mathbf{t}(sx), x\rangle\,\mathrm{d}s,$$

and we argue that $\phi \in G_2(\mathscr{X})$. The condition $\mathbf{t} \in G_1(\mathscr{X}, \mathscr{X})$ implies that $\mathbf{t}$ is bounded on bounded subsets of $\mathscr{X}$, namely $M_R = \sup_{\|x\| \le R} \|\mathbf{t}(x)\| < \infty$ for all $R \ge 0$. From the Cauchy–Schwarz inequality we get

$$|\phi(x) - \phi(y)| \le \int_0^1 \|\mathbf{t}(sy)\| \|y - x\| \,\mathrm{d}s + \int_0^1 \|\mathbf{t}(sy) - \mathbf{t}(sx)\| \|x\| \,\mathrm{d}s.$$

As $y \to x$, we may assume $\|y\| \le \|x\| + 1$ and so the first integral is smaller than $\|x - y\| M_{\|y\|} \le \|x - y\| M_{1+\|x\|}$ and thus vanishes. The second integrand converges to 0 as $y \to x$ and is bounded by $2\|x\| M_{1+\|x\|}$ so the corresponding integral vanishes by the dominated convergence theorem. This shows that ϕ is continuous. By definition $\|\mathbf{t}(sx)\| \le \|\mathbf{t}\|_{G_1}(1 + \|sx\|)$ so that

$$|\phi(x)| \le \|\mathbf{t}\|_{G_1}(1 + \|x\|)\|x\| \le 2\|\mathbf{t}\|_{G_1}(1 + \|x\|^2), \qquad x \in \mathscr{X}.$$

Thus $\phi \in G_2(\mathscr{X})$ and $\|\phi\|_{G_2} \le 2\|\mathbf{t}\|_{G_1}$; in particular ϕ is a (measurable) random element in $G_2(\mathscr{X})$. Since $\theta \in \mathscr{W}_2$, this guarantees that the integral of $|\phi| \mathrm{d}(\theta)$ is finite. By linearity, the Fubini equality holds if $\mathbf{t}$ is simple (the random element takes finitely many values in $G_1(\mathscr{X}, \mathscr{X})$). Otherwise, since it is Bochner measurable, we can find a sequence $\mathbf{t}_n$ of simple random elements such that $\|\mathbf{t}_n - \mathbf{t}\|_{G_1} \to 0$ almost surely and in expectation. Denote the corresponding potentials $\phi_n \in G_2(\mathscr{X})$ and observe that

$$\left| \int_{\mathscr{X}} \phi(x) \,\mathrm{d}\theta(x) - \int_{\mathscr{X}} \phi_n(x) \,\mathrm{d}\theta(x) \right| \le \int_{\mathscr{X}} \|\phi - \phi_n\|_{G_2}(1 + \|x\|^2) \,\mathrm{d}\theta(x) \le 2\|\mathbf{t} - \mathbf{t}_n\|_{G_1} \int_{\mathscr{X}} (1 + \|x\|^2) \,\mathrm{d}\theta(x)$$

vanishes as $n \to \infty$, since $\theta \in \mathscr{W}_2(\mathscr{X})$. This shows in particular that the quantity $\int_{\mathscr{X}} \phi \,\mathrm{d}\theta$ is a (measurable) random variable and its expectation is the limit of that of $\int_{\mathscr{X}} \phi_n \,\mathrm{d}\theta$. Similarly, for all x we have $|\phi(x) - \phi_n(x)| \le 2(1 + \|x\|^2)\|\mathbf{t} - \mathbf{t}_n\|_{G_1}$ which again implies that $\phi(x)$ is a (measurable) random variable with expectation

$$\lim_{n\to\infty} \mathbb{E}\phi_n(x) = \lim_{n\to\infty} \mathbb{E} \int_0^1 \langle \mathbf{t}_n(sx), x\rangle \,\mathrm{d}s = \lim_{n\to\infty} \int_0^1 \langle \mathbb{E}\mathbf{t}_n(sx), x\rangle \,\mathrm{d}s = \int_0^1 \langle \mathbb{E}\mathbf{t}(sx), x\rangle \,\mathrm{d}s,$$

establishing the second claim (the second equality is by linearity, since $\mathbf{t}_n$ is simple). As $\mathbb{E}\mathbf{t}$ is the identity, we obtain $\mathbb{E}\phi(x) = \|x\|^2/2$ for all $x \in \mathscr{X}$, that is, (i). Next, we have $\mathbb{E}\phi_n(x) \to \mathbb{E}\phi(x)$ and $|\mathbb{E}\phi_n(x)| \le 2\mathbb{E}\|\mathbf{t}_n\|_{G_1}(1 + \|x\|^2)$ and for n large this is smaller than $2(1 + \|x\|^2)(1 + \mathbb{E}\|\mathbf{t}\|_{G_1}) < \infty$. Therefore $\int_{\mathscr{X}} \mathbb{E}\phi_n(x) \,\mathrm{d}\theta(x) \to \int_{\mathscr{X}} \mathbb{E}\phi(x) \,\mathrm{d}\theta(x)$ by dominated convergence, establishing the third claim. The fourth claim follows from this, as the dual integral is the difference between the squared Wasserstein distance and the the integral with respect to θ, and both are measurable and integrable because the Fréchet functional is finite.

Proof (of the statement in Remark 4.2.5). Suppose that $\|T(x)\| = \|x\|^2$ for all x and let $\lambda \in \mathscr{W}_2(\mathscr{X}) \setminus \mathscr{W}_4(\mathscr{X})$. Then for $X \sim \lambda$ we have $\mathbb{E}\|T(X)\|^2 = \mathbb{E}\|X\|^4 < \infty$, so $T\#\lambda$ is not in $\mathscr{W}_2(\mathscr{X})$. In a similar way, one shows that if T grows faster than linearly, then for any $p \geq 1$ there exists a probability measure $\lambda \in \mathscr{W}_p(\mathscr{X})$ that T "throws away" from the space in that $T\#\lambda \notin \mathscr{W}_p(\mathscr{X})$.

Moreover, the linear growth condition appears to be the weakest condition compatible with the mean identity assumption. Indeed, if $T_1, \ldots, T_n : \mathbb{R} \to \mathbb{R}$ ($n \geq 2$) are nondecreasing with pointwise average equal to the identity, then for $x \geq 0$,

$$0 \leq T_n(x) - T_n(0) = nx - [T_1(x) - T_1(0)] - \sum_{i=2}^{n-1} [T_i(x) - T_i(0)] \leq nx - [T_1(x) - T_1(0)],$$

with a similar inequality for negative x. This means that $|T_1(x) - T_1(0)| \leq n|x|$, so $\|T_1\|_{G_1} \leq n$, and the same holds for each of the T_i's.

We now prove the equivalence between the weak and vague topology mentioned in page 93 in the book.

Lemma 5 *Let $\mu_n, \mu \in P(\mathscr{X})$.*

(a) If $\mu_n \to \mu$ vaguely and $\mathscr{X} = \mathbb{R}^d$, then $\mu_n \to \mu$ weakly.
(b) If $\mathscr{X}$ is an infinite-dimensional Hilbert space, then $\mu_n \to \mu$ vaguely. In other words, as μ_n and μ are arbitrary, the vague topology is trivial.

Proof. (a) We first show that the sequence $\{\mu_n\}$ is tight. Let K_1 be compact such that $\mu(K_1) > 1 - \varepsilon$ and let $K_2 = \{x : \|x - K_1\| \leq \varepsilon\}$ be the closed ε-enlargement of K_1. Then K_2 is compact, and there exists a continuous function $0 \leq f \leq 1$ that equals 1 on K_1 and 0 outside K_2. Hence

$$\mu_n(K_2) \geq \int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu \geq \mu(K_1)$$

is larger than $1 - 2\varepsilon$ for n sufficiently large. Since K_2 is compact, tightness is established. Any weak limit μ^* must equal μ on compacta, and, consequently, on countable unions of compacta. On $\mathbb{R}^d$ this entails that μ^* and μ are equal on the π-system of all closed sets. Since the collection of subsets on which μ^* and μ agree is a λ-system, $\mu^* = \mu$ by a monotone class argument.

(b) Since compact sets in $\mathscr{X}$ have empty interior, any continuous compactly supported (or C_0) function $f : \mathscr{X} \to \mathbb{R}$ is identically zero: let $x \in \mathscr{X}$ and $\varepsilon > 0$. Let $K \subseteq \mathscr{X}$ be compact such that $|f| \leq \varepsilon$ outside K, and $\delta > 0$ such that if $\|y - x\| < \delta$ then $|f(y) - f(x)| < \varepsilon$. It is possible to choose such y that is not in K, so that $|f(x)| \leq |f(y)| + \varepsilon \leq 2\varepsilon$. Since x and ε are arbitrary we obtain $f \equiv 0$. This means that the criterion for μ_n to converge vaguely to μ is that

$$0 = \lim_{n\to\infty} \int_{\mathscr{X}} 0 \, \mathrm{d}\mu_n = \lim_{n\to\infty} \int_{\mathscr{X}} 0 \, \mathrm{d}\mu = 0,$$

and is always satisfied.

Proof (Proof of Lemma 4.4.2). Denote the total number of points by $N_i = \widetilde{\Pi}_i(\mathbb{R}^d)$, suppose that it is nonzero and let $\Psi(A) = \int_A \psi(x) \, \mathrm{d}x$ be the probability measure corresponding to the density ψ. For every $y \in K$ define $\tilde{\mu}_y = \delta\{y\} * \psi_\sigma$ and its restricted renormalised version $\mu_y = (1/\tilde{\mu}_y(K))\tilde{\mu}_y|_K$. Then $\widehat{\Lambda}_i = (1/N_i)\sum_{j=1}^{N_i} \mu_{x_j}$ with $N_i \geq 1$ and $x_j \in K$ (because $\Lambda_i(K) = 1$).

A coupling (certainly not optimal, unless $N_i = 1$) of $\widehat{\Lambda}_i$ and $\widetilde{\Lambda}_i = \widetilde{\Pi}_i/N_i$ can be constructed by sending the $1/N_i$ mass of μ_{x_j} to x_j. This gives

$$W_2^2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) \leq \frac{1}{N_i}\sum_{j=1}^{N_i} W_2^2(\mu_{x_j}, \delta\{x_j\}) = \frac{1}{N_i}\sum_{j=1}^{N_i} \frac{1}{\tilde{\mu}_{x_j}(K)}\int_K \|x - x_j\|^2 \psi_\sigma(x - x_j)\,\mathrm{d}x.$$

A change of variables shows that each of the last displayed integrals is bounded by σ^2, since ψ was assumed to have unit variance and so ψ_σ has variance σ^2. The proof will be complete if we can find a lower bound for $\tilde{\mu}_y(K)$ that is uniform in σ and in $y \in K$. Clearly

$$\tilde{\mu}_y(K) = \int_K \psi_\sigma(x - y)\,\mathrm{d}x = \int_{(K-y)/\sigma} \psi(x)\,\mathrm{d}x = \Psi\left(\frac{K - y}{\sigma}\right).$$

Let us first eliminate σ. The set $K_y = K - y$ is a convex set that includes the origin; it follows that $K_y \subseteq (1+\varepsilon)K_y$ for all $\varepsilon > 0$. Consequently $K_y/\sigma \supseteq K_y$ as long as $\sigma \leq 1$. Recalling that $\psi(x) = \psi_1(\|x\|)$ with ψ_1 nonincreasing and strictly positive, we find

$$\Psi\left(\frac{K - y}{\sigma}\right) \geq \Psi(K - y) = \int_{K-y} \psi(x)\,\mathrm{d}x \geq \int_{K-y} \psi_1(d_K)\,\mathrm{d}x = \psi_1(d_K)\mathrm{Leb}K > 0.$$

We have again used the notation $d_K = \sup\{\|x - y\| : x, y \in K\}$ for the finite diameter of the compact set K.

If we now define $C_{\psi,K} = [\psi_1(d_K)\mathrm{Leb}K]^{-1} < \infty$, then putting everything together gives

$$W_2^2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) \leq \frac{1}{N_i}\sum_{j=1}^{N_i} W_2^2(\mu_{x_j}, \delta\{x_j\}) \leq C_{\psi,K}\sigma^2 \qquad \text{if } \sigma \leq 1.$$

Finally, if $N_i = 0$, then by construction $W_2(\widehat{\Lambda}_i, \widetilde{\Lambda}_i) = 0$.

Here is an example: suppose that $K = [0,\infty)^2$ and $y = 0$. Then $\tilde{\mu}_y(K) = 1/4$ for all $\sigma > 0$. But actually $W_2^2(\mu_y, \delta_y) = W_2^2(\tilde{\mu}_y, \delta_y) = \sigma^2$ by the isotropy of ψ: this can be seen by "folding" each quadrant onto the positive quadrant in $\mathbb{R}^2$. If now K is $[0,1]^2$, then after this folding there is still mass in the positive quadrant outside of K, so in fact $W_2^2(\mu_y, \delta_y) < \sigma^2$.

Proof (Proof of Theorem 4.4.1). Let us first show the convergence in probability of $\widehat{\Lambda}_i$ to Λ_i. Let $N_i^{(n)} = \Pi_i^{(n)}(\mathscr{X}) = \widetilde{\Pi}_i^{(n)}(\mathscr{X})$ denote the total number of observed points. We may assume without loss of generality that τ_n are integers: otherwise, we replace τ_n by $\lfloor \tau_n \rfloor$ and Λ_i by $(\tau_n/\lfloor \tau_n \rfloor)\Lambda_i$ which converges to Λ_i because the fraction $\tau_n/\lfloor \tau_n \rfloor \to 1$.

Then, $\Pi_i^{(n)}$ has the same distribution as the superposition of τ_n independent copies of Π, say $\{P_j^{(n)}\}$, and $\widetilde{\Pi}_i^{(n)}$ has the same distribution as a superposition of

$\{\widetilde{P}_j^{(n)}\}$, independent copies of $T_i\#\Pi$ which have mean measure Λ_i. Consequently, (e.g., Karr [6, Proposition 4.8])

$$\frac{1}{\tau_n}\widetilde{\Pi}_i^{(n)} \overset{d}{=} \frac{1}{\tau_n}\sum_{j=1}^{\tau_n}\widetilde{P}_j^{(n)} \xrightarrow{n} \Lambda_i, \qquad \text{in probability},$$

with '$\xrightarrow{n}$' denoting weak convergence of measures. (The convergence will be almost surely if $\widetilde{P}_j^{(n+1)} = \widetilde{P}_j^{(n)}$; but otherwise the convergence is only in probability unless further conditions are imposed).

The proof of this result is just a conditional version of the empirical measure setting in Proposition 2.2.6 with n replaced by τ_n: for any continuous bounded $f : K \to \mathbb{R}$,

$$\int_K f \,\mathrm{d}\frac{1}{\tau_n}\widetilde{\Pi}_i^{(n)} \to \int_K f \,\mathrm{d}\Lambda_i, \qquad \text{in probability},$$

and one then finds a countable collection (f_j) that suffices to conclude the weak convergence. In particular when $f \equiv 1$ we obtain $N_i^{(n)}/\tau_n \xrightarrow{p} 1$ and conclude from Slutsky's theorem that

$$\widetilde{\Pi}_i^{(n)}/N_i^{(n)} \xrightarrow{n} \Lambda_i \qquad \text{in probability}. \tag{4.1}$$

The weak convergence is equivalent to Wasserstein convergence, since K is compact (Corollary 2.2.2). Finally, by Lemma 4.4.2 and the triangle inequality

$$W_2(\widehat{\Lambda}_i,\Lambda_i) \le W_2\left(\Lambda_i, \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}\right) + W_2\left(\frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}, \widehat{\Lambda}_i\right) \le W_2\left(\Lambda_i, \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}\right) + \sqrt{C_{\psi,K}}\,\sigma_i^{(n)} \to 0,$$

because $\sigma_i^{(n)} \to 0$ as $n \to \infty$. This proves claim (1) in probability.

Let us now prove claim (2). Define on $\mathscr{W}_2(K)$ the functionals

$$\begin{aligned}
F(\gamma) &= \frac{1}{2}\mathbb{E}W_2^2(\Lambda,\gamma);\\
F_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\Lambda_i,\gamma);\\
\tilde{F}_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\widetilde{\Lambda}_i,\gamma), \qquad \widetilde{\Lambda}_i = \frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}} \qquad \text{or } \lambda^{(0)} \text{ if } N_i^{(n)} = 0;\\
\widehat{F}_n(\gamma) &= \frac{1}{2n}\sum_{i=1}^n W_2^2(\widehat{\Lambda}_i,\gamma), \qquad \widehat{\Lambda}_i = \lambda^{(0)} \text{ if } N_i^{(n)} = 0.
\end{aligned}$$

Assumptions 3 imply that λ is the unique minimiser of F, and we wish to show that any sequences of minimisers $\widehat{\lambda}_n$ of $\widehat{F}_n$ must converge to λ. To this end we shall bound the differences between any two consecutive functionals uniformly in γ. This is possible because all the relevant measures lie in a bounded set of the Wasserstein space $\mathscr{W}_2(\mathbb{R}^d)$. Indeed, if μ, ν and ρ are probability measures on K, then

$$W_2(\mu,\nu) \le \sqrt{\sup_{\pi\in P(K^2)} \int_{K^2} \|x-y\|^2 \, \mathrm{d}\pi(x,y)} \le \sqrt{\sup_{x,y\in K} \|x-y\|^2} = d_K < \infty; \tag{4.2}$$

$$|W_2^2(\mu,\rho) - W_2^2(\nu,\rho)| = |W_2(\mu,\rho)+W_2(\nu,\rho)||W_2(\mu,\rho)-W_2(\nu,\rho)| \le 2d_K W_2(\mu,\nu), \tag{4.3}$$

so that

$$\sup_{\gamma\in\mathscr{W}_2(K)} |\widehat{F}_n(\gamma) - \tilde{F}_n(\gamma)| \le \frac{d_K}{n}\sum_{i=1}^n W_2\left(\widehat{\Lambda}_i, \widetilde{\Lambda}_i\right) \le d_K\sqrt{C_{\psi,K}}\frac{1}{n}\sum_{i=1}^n \sigma_i^{(n)}$$

by Lemma 4.4.2. The right-hand side vanishes by our assumptions.

Similarly,

$$\sup_{\gamma\in\mathscr{W}_2(K)} |\tilde{F}_n(\gamma) - F_n(\gamma)| \le \frac{1}{n}\sum_{i=1}^n W_2\left(\Lambda_i, \widetilde{\Lambda}_i\right) = \frac{1}{n}\sum_{i=1}^n X_{ni} = \overline{X}_n.$$

Now X_{ni} is a function of T_i and $\Pi_i^{(n)}$, so by construction $(X_{ni})_{i=1}^n$ are independent and identically distributed. Therefore $\mathbb{E}\overline{X}_n = \mathbb{E}X_{n1}$. Since $X_{ni} \in [0,d_K]$ by (4.2)[1] and $X_{ni} \to 0$ in probability by (4.1), we have $\mathbb{E}\overline{X}_n \to 0$ by the bounded convergence theorem. In general L_1 convergence does not imply almost sure convergence, but here we deal with averages so the latter can be established. The centred versions $Y_{ni} = X_{ni} - \mathbb{E}X_{ni}$ are again bounded, and repeating the proof of the fourth moment law of large numbers (Durrett [5, Theorem 2.3.5]), we have

$$\mathbb{P}\left(\left(\overline{X}_n - \mathbb{E}\overline{X}_n\right)^4 > \varepsilon\right) = \mathbb{P}(\overline{Y_n}^4 > \varepsilon) \le \frac{n\mathbb{E}\left[Y_{n1}^4\right] + 3n(n-1)\mathbb{E}\left[Y_{n1}^2\right]}{\varepsilon^4 n^4} \le \frac{3\max(d_K^4, d_K^2)}{\varepsilon^4 n^2}.$$

Put $\varepsilon = n^{-1/5}$ and apply the Borel–Cantelli lemma while observing that $\mathbb{E}\overline{X}_n \to 0$ to conclude $|\overline{X}_n| \le |\overline{X}_n - \mathbb{E}\overline{X}_n| + |\mathbb{E}\overline{X}_n| \to 0$ almost surely.

Uniform convergence of F_n to F comes from a combination of the uniform Lipschitz bound (4.3), the strong law of large numbers and compactness of $\mathscr{W}_2(K)$ (Corollary 2.2.5). For each $\gamma \in \mathscr{W}_2$,

$$F_n(\gamma) \xrightarrow{a.s.} F(\gamma),$$

Fix $\varepsilon > 0$, invoke the total boundedness of $\mathscr{W}_2(K)$ to find a finite ε-cover $\gamma_1,\dots,\gamma_m$, $m = m(\varepsilon)$. By virtue of (4.3), F_n and F are uniformly d_K-Lipschitz. For any $\gamma \in \mathscr{W}_2(K)$ choose j such that $W_2(\gamma,\gamma_j) < \varepsilon$. Then

[1] One may need to increase d_K in case of empty point processes depending on the choice of $\lambda^{(0)}$, but it will in any case be finite.

$$\begin{aligned}|F_n(\gamma)-F(\gamma)| &\le |F_n(\gamma)-F_n(\gamma_j)|+|F_n(\gamma_j)-F(\gamma_j)|+|F(\gamma_j)-F(\gamma)|\\ &\le d_K W_2(\gamma,\gamma_j)+|F_n(\gamma_j)-F(\gamma_j)|+d_K W_2(\gamma,\gamma_j)\\ &\le 2d_K\varepsilon+|F_n(\gamma_j)-F(\gamma_j)|.\end{aligned}$$

Thus almost surely

$$\limsup_{n\to\infty}\sup_{\gamma\in\mathscr{W}_2(K)}|F_n(\gamma)-F(\gamma)|\le 2d_K\varepsilon.$$

Since $\varepsilon>0$ is arbitrary, we conclude that

$$\sup_{\gamma\in\mathscr{W}_2(K)}|\widehat{F}_n(\gamma)-F(\gamma)|\to 0,\qquad \text{almost surely.}$$

Convergence of minimisers is now standard. If a subsequence of $\widehat{\lambda}_n$ converges to μ, then the uniform convergence of $\widehat{F}_n$ to F and the continuity of F imply that $\widehat{F}_{n_k}(\widehat{\lambda}_{n_k})\to F(\mu)$. The definition of $\widehat{\lambda}_n$ gives $\widehat{F}_{n_k}(\widehat{\lambda}_{n_k})\le\widehat{F}_{n_k}(\lambda)\to F(\lambda)$. Consequently, $F(\mu)\le F(\lambda)$ and it must be that $\mu=\lambda$ because λ is the unique minimiser of F. Since $\widehat{\lambda}_n$ is a sequence in the compact set $\mathscr{W}_2(K)$, this means that $W_2(\widehat{\lambda}_n,\lambda)\to 0$ almost surely.

Lastly, we prove convergence almost surely in (1) under the more stringent assumptions on τ_n and on Π, mentioned in the end of the theorem's statement. Let us begin by showing that for all $a=(a_1,\dots,a_d)\in\mathbb{R}^d$,

$$\mathbb{P}\left(\frac{\widetilde{\Pi}_i^{(n)}((-\infty,a])}{\tau_n}-\Lambda_i((-\infty,a])\to 0\right)=1,\quad (-\infty,a]=(-\infty,a_1]\times\cdots\times(-\infty,a_d].$$

To simplify we shall write a instead of $(-\infty,a]$ henceforth. Recall that $\widetilde{P}_j^{(n)}$ are generic point processes, distributed as $T_i\#\Pi$ and independent across j. We may assume that they are constructed as $T_i\#P_j^{(n)}$ with $P_j^{(n)}$ distributed as Π.

Define the random variables

$$X_{nj}=\widetilde{P}_j^{(n)}(a)-\Lambda_i(a),\quad j=1,\dots,\tau_n;\qquad S_n=\sum_{j=1}^{\tau_n}X_{nj}.$$

The idea is now to use the fourth-moment law of large numbers (Durrett [5, Theorem 2.3.5]) conditional on Λ_i. Here is an informal argument. Since $\Lambda_i=T_i\#\lambda$, conditioning on Λ_i is equivalent to conditioning on T_i. The random variables X_{nj} have conditional mean zero by construction; and since T_i and $\{\Pi_i^{(n)}\}$ are independent, X_{nj} are also conditionally independent across j. It follows that

$$\mathbb{E}[S_n^4|T_i]=\sum_{j=1}^{\tau_n}\mathbb{E}[X_{nj}^4|T_i]+\sum_{j<l}\mathbb{E}[X_{nj}^2X_{nl}^2|T_i]=\tau_n\mathbb{E}[X_{11}^4|T_i]+3\tau_n(\tau_n-1)\mathbb{E}[X_{11}X_{12}|T_i].$$

To see this formally, we set $k = \tau_n$ and define $\Phi : (M(U))^k \times C_b(U,K) \to \mathbb{R}_+$ by

$$\Phi(p_1,\dots,p_k,f) = \left[\sum_{j=1}^{k} f\#p_j(a) - f\#\lambda(a)\right]^4, \quad f \in C_b(U,K); \quad p_j \in M(U).$$

(Recall that $M(U)$ is the collection of finite Borel measures on U endowed with the topology of weak convergence.) Then $S_n^4 = \Phi(P_1^{(n)},\dots,P_k^{(n)},T_i)$ and we claim that Φ is continuous (hence measurable). Indeed, if V_n are random vectors that converge weakly to V and f_n are continuous functions that converge uniformly to f, then $f_n(V_n) \to f(V)$ weakly by the continuous mapping theorem and Slutsky's theorem. Another application of Slutsky's theorem then shows that Φ is continuous. Finally, Φ is integrable because $0 \le f\#\lambda(a) \le 1$ and $\mathbb{E}[T_i\#P_j^{(n)}(a)]^4 \le \mathbb{E}[\Pi(\mathbb{R}^d)]^4 < \infty$ by the hypothesis.

Since $\{P_j^{(n)}\}$ and T_i are independent, one can evaluate the conditional expectation $\mathbb{E}[S_n^4|T_i]$ by taking the expectation with respect to P. That is, if we define $g : C_b(U,K) \to \mathbb{R}_+$ by

$$g(f) = \mathbb{E}_P\left[\Phi(P_1^{(n)},\dots,P_k^{(n)},f)\right] = \int_{[M(U)]^k} \Phi(p_1,\dots,p_k,f)\,\mathrm{d}(p_1,\dots,p_k), \quad f \in C_b(U,K),$$

then [5, Lemma 6.2.1] gives $\mathbb{E}[S_n^4|T_i] = g(T_i)$.

The same idea shows that for each j,

$$\mathbb{E}[X_{nj}|T_i] = \int_{M(U)} T_i\#p_j(a)\,\mathrm{d}p_j - T_i\#\lambda(a) = \lambda(T_i^{-1}(a)) - \lambda(T_i^{-1}(a)) = 0.$$

This provides the formal justification for the expression for $\mathbb{E}[S_n^4|T_i]$. If we now take the expectation with respect to T_i and apply Markov's inequality, we obtain

$$P\left[\left(\frac{S_n}{\tau_n}\right)^4 > \varepsilon\right] \le \frac{\mathbb{E}[S_n^4]}{\varepsilon^4\tau_n^4} = \frac{\tau_n\mathbb{E}[X_{11}^4] + 3\tau_n(\tau_n-1)\mathbb{E}[X_{11}^2X_{12}^2]}{\varepsilon^4\tau_n^4}.$$

Since the expectations are finite, the right-hand side is bounded by a constant times τ_n^{-2}, which is a convergent sum by the hypothesis. The result now follows from the Borel–Cantelli lemma.

Now that we have convergence for a fixed $a \in \mathbb{R}^d$, we use a standard approximation by rationals to obtain the convergence for *all* $a \in \mathbb{R}^d$. Indeed, we have

$$\mathbb{P}\left(\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \to 0 \text{ for any } a \in \mathbb{Q}^d\right) = 1.$$

If $a \in \mathbb{R}^d$ is arbitrary, then we can find rational sequences $a^k \nearrow a \swarrow b^k$ that converge monotonically coordinatewise to a. We can then use the approximations

$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \le \frac{\widetilde{\Pi}_i^{(n)}(b^k)}{\tau_n} - \Lambda_i(b^k) + \Lambda_i(b^k) - \Lambda_i(a);$$
$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \ge \frac{\widetilde{\Pi}_i^{(n)}(a^k)}{\tau_n} - \Lambda_i(a^k) + \Lambda_i(a^k) - \Lambda_i(a).$$

The resulting errors

$$\Lambda_i(b^k) - \Lambda_i(a) = \Lambda_i((-\infty, b^k] \setminus (-\infty, a]) \quad \text{and} \quad \Lambda_i(a^k) - \Lambda_i(a) = -\Lambda_i((-\infty, a] \setminus (-\infty, a^k])$$

both vanish as $k \to \infty$: the first set converges monotonically to the empty set; the second one does not converge to empty set but rather to $(-\infty, a] \setminus (-\infty, a)$, which is a union of d rays of dimension $d-1$. When a is a continuity point of Λ_i, this is still a Λ_i-null set. We may therefore conclude that with probability one

$$\frac{\widetilde{\Pi}_i^{(n)}(a)}{\tau_n} - \Lambda_i(a) \to 0, \qquad \text{for all } a \in \mathbb{R}^d \text{ continuity point of } \Lambda_i.$$

Taking $a = \infty$, we see that $\tau_n / N_i^{(n)} \to 1$ almost surely, so that

$$\frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}} \to \Lambda_i \text{ weakly.}$$

Since all these measures are concentrated on the compact set $K \subset \mathbb{R}^d$, the convergence holds in Wasserstein distance too. Finally,

$$W_2(\widehat{\Lambda}_i, \Lambda_i) \le W_2(\widehat{\Lambda}_i, \widetilde{\Pi}_i^{(n)}/N_i^{(n)}) + W_2(\widetilde{\Pi}_i^{(n)}/N_i^{(N)}, \Lambda_i) \to 0, \qquad n \to \infty,$$

by Lemma 4.4.2 if $\sigma_i^{(n)} \to 0$.

Extensions of Theorem 4.4.3 to the boundary. Stronger statements can be made when we can control the behaviour at the boundary of K. For example, when $\mathscr{X} = \mathbb{R}$, $K = [a,b]$ and the construction guarantees that $\widehat{T_i^{-1}}(a) = a$ and $\widehat{T_i^{-1}(b)} = b$, because in the one-dimensional case we *do* know that the Fréchet mean $\widehat{\lambda}_n$ is strictly positive on K. Consequently, the convergence in Theorem 4.4.3 actually holds on the whole of K. This can also be seen in elementary ways by properties of nondecreasing functions on the real line (Panaretos & Zemel [8]).

The interpretation of this property when $d = 1$ in terms of the set-valued framework is more propitious for extensions to multivariate setups. Let u be the set-valued function represented by T_i^{-1}. If $x = b \in \partial K$, then $u(x)$ is a subset of the ray $[b, \infty)$ (because u is nondecreasing and $u(z) \to b$ as $z \nearrow b$). In other words, there is a unique $y \in K$ that can be an element of $u(x)$, namely $y = b$. The same thing happens at $x = a$, which is the only other point of the boundary of K.

Now suppose that $\mathscr{X} = \mathbb{R}^d$ and u is as above. Assume that for each $x \in \partial K$, $u(x) \cap K$ contains exactly one element y. Let x_n be a sequence in U that converges to $x \in \partial K$. If $y_n \in u(x_n)$ and $y_n \to y$, then it is not difficult to see that $y \in u(x)$

(this property is called upper semicontinuity of set-valued functions and proven in Alberti & Ambrosio [2, Corollary 1.3]). Since y_n must be in K, it follows that they must converge to y. The same convergence holds when $y_n \in u_n(x_n)$, where u_n is represented by $\widehat{T_i^{-1}}$. In other words, we have extended the uniform convergence on compact subsets of U to uniform convergence on U itself.

Finally, for Corollary 4.4.4 we have assumed that $T_i(x) \in U$ for all $x \in U$. Let us see two sufficient conditions for this to be a consequence rather than an assumption: one in terms of T_i, the other in terms of the geometry of K. What we do know is that $T_i(x) \in K$ for all $x \in U$ and it is of interest to see whether this property suffices. Suppose that $y = T_i(x) \in \partial K$ for some $x \in \mathrm{int}K$. By the Hahn–Banach theorem there exists $\alpha \in \mathbb{R}^d \setminus \{0\}$ with $\langle y, \alpha\rangle \geq \sup \langle K, \alpha\rangle$. Let $x' = x + t\alpha$ for $t > 0$ small enough such that $x' \in U$. Then $y' = T_i(x') \in K$, so that

$$0 \leq \left\langle y' - y, x' - x\right\rangle = t\left\langle y' - y, \alpha\right\rangle.$$

One way to obtain a contradiction is to assume that T_i is strictly monotone on U; and this happens when the convex potential of T_i is strictly convex on U. Conditions for this are given in Theorem 1.6.7.

Another way is to assume that α separates y from K strictly, in the sense that

$$\langle y, \alpha\rangle > \left\langle y', \alpha\right\rangle, \qquad y' \in K \setminus \{y\}.$$

When such a strict separator exists (and $y \in K$), we say that y is an *exposed* point of K. When this is the case, the inequality $0 \leq t\langle y' - y, \alpha\rangle$ entails $y' = y$, because $t > 0$. This is a contradiction to the injectivity of T_i. Hence when *any* boundary point of K is exposed, T_i must map U into U. Examples for such K include the unit ball or any ellipsoid in $\mathbb{R}^d$ and more generally, when it can be written as $\partial K = \{x : \varphi_K(x) = 0\}$, for some strictly convex function φ_K. Indeed, if α creates a supporting hyperplane to K at y and $\langle \alpha, y\rangle = \langle \alpha, y'\rangle$ for $y \neq y'$, then as φ_K is strictly convex on the line segment $[y, y']$, it is impossible that $y' \in K$ without the hyperplane intersecting the interior of K. Although this condition excludes some interesting cases, perhaps most prominently polyhedral sets such as $K = [0,1]^d$, such sets can be approximated by convex sets that do satisfy it (Krantz [7, Proposition 1.12]).

Proof (Proof of Lemma 4.6.1). If Π is a binomial process ($\widetilde{\Lambda}_i$ is the empirical measure), then $N_i^{(n)} = \tau_n$ for all i and all n and there is nothing to prove.

Let us begin with the Poisson case, in which case the argument is more transparent. In this case $N_1^{(n)}, \ldots, N_n^{(n)}$ are independent Poisson random variables with parameter τ_n. We can then use a Chernoff bound as follows: if N has a Poisson(τ) distribution, then for any $c > 1$ and any $t \geq 0$,

$$\mathbb{P}(N \leq \tau/c) = \mathbb{P}(e^{-Nt} \geq e^{-\tau t/c}) \leq \frac{\mathbb{E}e^{-Nt}}{e^{-\tau t/c}} = \exp\left[\tau\left(e^{-t} + \frac{t}{c} - 1\right)\right].$$

The bound is optimised when $t = \log c$, yielding

$$\mathbb{P}(N \leq \tau/c) = \exp -\tau\alpha, \qquad \alpha = \alpha(c) = c^{-1}[c - 1 - \log c] > 0.$$

Since $\tau_n \to \infty$, this in particular shows that the probability that $N_i^{(n)}/\tau_n < 1/c$ vanishes as $n \to \infty$.

By Bonferroni's inequality, and since $N_i^{(n)}$ have the same distribution,

$$\mathbb{P}\left(\min_{1\leq i\leq n} N_i^{(n)} \leq \frac{\tau_n}{c}\right) \leq n\mathbb{P}\left(N_1^{(n)} \leq \frac{\tau_n}{c}\right) \leq n\exp[-\alpha(c)\tau_n].$$

If $\tau_n/\log_n \to \infty$ then for n large, the expression in the exponent is smaller than $-3\log n$. Summation over n of the probability on the left-hand side is therefore convergent, and the Borel–Cantelli lemma gives

$$\liminf_{n\to\infty} \frac{\min_{1\leq i\leq n} N_i^{(n)}}{\tau_n} \geq 1 \qquad \text{almost surely.}$$

One then shows the reverse inequalities by analogous calculations.

When Π is no longer Poisson, we replace the above argument with a Chernoff bound on binomial distributions, using a very crude bound.

Denote by p the probability that Π has no points. If τ is an integer, then $N_i^{(n)}$ is a sum of independent integer-valued random variables X_i. Since X_i is always an integer, we have the lower bound $X_i \geq \mathbf{1}\{X_i \geq 1\}$. Thus $N_i^{(n)}$ is stochastically larger than a random variable $N \sim B(\tau_n, 1-p)$. Set $q = 1-p$ and use the Chernoff bound as follows: for any $t \geq 0$

$$\mathbb{P}\left(N \leq \frac{\tau q}{c}\right) = \mathbb{P}\left(\exp(-Nt) \geq \exp\left(-t\frac{\tau q}{c}\right)\right) \leq \mathbb{E}\exp(-Nt)\exp\left(t\frac{\tau q}{c}\right) = \left[s^{q/c}\left(1 - q + \frac{q}{s}\right)\right]^{\tau},$$

where $s = e^t \geq 1$. The bound is optimised when $s = (c-q)/(1-q) > 1$, and we obtain

$$\mathbb{P}\left(N_i^{(n)} \leq \tau_n q/c\right) \leq \beta^{\tau_n}, \qquad \beta = \beta(q,c) = c\left((1-q)/(c-q)\right)^{1-q/c} < 1.$$

One then concludes as before that if $\tau_n/\log n \to \infty$, then almost surely

$$\liminf_{n\to\infty} \frac{\min_{1\leq i\leq n} N_i^{(n)}}{\tau_n} \geq 1 - p.$$

Finally, we treat the case where τ_n are not integers. We claim that in any case, the probability that $N_1^{(n)} = 0$ is p^{τ_n}. Indeed, recall that the Laplace functional of $\Pi_1^{(n)}$ is

$$f \mapsto \mathbb{E}e^{-\Pi_1^{(n)}f} = [L_\Pi(f)]^{\tau_n} = \left[\mathbb{E}e^{-\Pi f}\right]^{\tau_n}, \qquad f : \mathscr{X} \to \mathbb{R}_+.$$

By the bounded convergence, we may recover the zero probabilities by taking $f \equiv m$ to be a constant function:

$$\mathbb{P}(N_i^{(n)}=0)=\lim_{m\to\infty}\mathbb{E}e^{-mN_i^{(n)}}=\lim_{m\to\infty}[L_\Pi(m)]^{\tau_n}=\lim_{m\to\infty}[\mathbb{E}e^{-m\Pi(\mathscr{X})}]^{\tau_n}=p^{\tau_n}.$$

By infinite divisibility, $N_i^{(n)}$ has the same law as the sum of $\lfloor\tau_n\rfloor$ (the largest integer not larger than τ_n) independent integer valued random variables with zero probability $p'=p^{\tau_n/\lfloor\tau_n\rfloor}\le p$. The same argument then gives

$$\liminf_{n\to\infty}\frac{\min_{1\le i\le n}N_i^{(n)}}{\lfloor\tau_n\rfloor}\ge 1-p,$$

and as $\tau_n\to\infty$, we may replace $\lfloor\tau_n\rfloor$ by τ_n, which completes the proof.

Proof (Proof of Theorem 4.6.3). Recall from Section 1.5 that $W_2(\gamma,\theta)=\|F_\theta^{-1}-F_\gamma^{-1}\|_{L_2(0,1)}$ (quantile functions). The empirical Fréchet mean λ_n that minimises F_n is found by averaging the quantile functions of Λ_i (see Subsection 3.1.4), so that

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

By the central limit theorem in Hilbert spaces, the above expression converges weakly to a Gaussian limit GP with $\mathbb{E}\|GP\|^2<\infty$ as $n\to\infty$. In particular,

$$W_2(\lambda_n,\lambda)=\|F_{\lambda_n}^{-1}-F_\lambda^{-1}\|=O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Replacing λ_n with $\widehat{\lambda}_n$ results in an error of

$$\left\|F_{\lambda_n}^{-1}-F_{\widehat{\lambda}_n}^{-1}\right\|=\left\|\frac{1}{n}\sum_{i=1}^n F_{\Lambda_i}^{-1}-\frac{1}{n}\sum_{i=1}^n F_{\widehat{\Lambda}_i}^{-1}\right\|\le\frac{1}{n}\sum_{i=1}^n\left\|F_{\Lambda_i}^{-1}-F_{\widehat{\Lambda}_i}^{-1}\right\|=\frac{1}{n}\sum_{i=1}^n W_2(\Lambda_i,\widehat{\Lambda}_i).$$

Invoking the triangle inequality, we bound this by the sum of the amplitude term and the smoothing term:

$$\le\frac{1}{n}\sum_{i=1}^n W_2\left(\Lambda_i,\widetilde{\Lambda}_i\right)+\frac{1}{n}\sum_{i=1}^n W_2\left(\widetilde{\Lambda}_i,\widehat{\Lambda}_i\right)\le\frac{1}{n}\sum_{i=1}^n W_2\left(\Lambda_i,\widetilde{\Lambda}_i\right)+\sqrt{C_{\psi,K}}\sigma_n$$

by Lemma 4.4.2.

Define (as in the proof of Theorem 4.4.1) $X_{ni}=W_2(\Lambda_i,\widetilde{\Lambda}_i)$ and recall that

$$\widetilde{\Lambda}_i=\frac{\widetilde{\Pi}_i^{(n)}}{N_i^{(n)}}\quad\text{if } N_i^{(n)}>0,\qquad\text{and } \lambda^{(0)} \text{ otherwise.}$$

Set $S_{ni}=\mathbf{1}\{N_i^{(n)}>0\}$ and write

$$X_{ni}=W_2(\Lambda_i,\widetilde{\Lambda}_i)S_{ni}+W_2(\Lambda_i,\lambda^{(0)})(1-S_{ni})\le W_2(\Lambda_i,\widetilde{\Lambda}_i)S_{ni}+d_K(1-S_{ni}),$$

where d_K may need be increased if $\lambda^{(0)}(K) < 1$. The last term is zero for n large by Lemma 4.6.1 so converges at any rate: if $a_n \to \infty$ is any sequence, then

$$\mathbb{P}\left(a_n \sum_{i=1}^{n} 1 - S_{ni} > \varepsilon\right) = \mathbb{P}\left(a_n \sum_{i=1}^{n} \mathbf{1}\{N_i^{(n)} = 0\} > \varepsilon\right) \le \mathbb{P}\left(a_n \sum_{i=1}^{n} \mathbf{1}\{N_i^{(n)} = 0\} > 0\right) \to 0.$$

It remains to find the rate of the average of $X_{ni}S_{ni}$. As a first step, we replace probability calculations by expectations, using Markov's inequality:

$$\mathbb{P}\left(a_n \frac{1}{n} \sum_{i=1}^{n} X_{ni}S_{ni} > \varepsilon\right) \le \frac{a_n \mathbb{E}\sum_{i=1}^{n} X_{ni}S_{ni}}{n\varepsilon} = \frac{a_n \mathbb{E}X_{n1}S_{n1}}{\varepsilon}.$$

The idea is now to replace W_2 by W_1 (using (2.2)), which one can evaluate in terms of distribution functions by Corollary 1.5.3. Let us introduce $a = \inf K$ and $b = \sup K$, so that $K = [a,b]$ (since K is compact and convex). For any measure θ on $\mathbb{R}$ denote $\theta((-\infty,t])$ by $\theta(t)$. Then

$$\frac{\mathbb{E}X_{n1}^2 S_{n1}}{d_K} \le \mathbb{E}S_{n1} W_1\left(\Lambda_1, \frac{\widetilde{\Pi}_1^{(n)}}{N_1^{(n)}}\right) = \int_a^b \mathbb{E}\left|\Lambda_1(t) - \frac{\widetilde{\Pi}_1^{(n)}(t)}{N_1^{(n)}}\right| S_{n1}\,\mathrm{d}t = \int_a^b \mathbb{E}\left|B_t\right|\,\mathrm{d}t,$$

where B_t is defined by the above equation. Let us assume that Π is a Poisson process. Fix $t \in [a,b]$ and notice that conditional on Λ_1 and on the event $N_1^{(n)} = k$, $B_t = 0$ if $k = 0$ and otherwise follows a centred renormalised binomial distribution, of the form $B_t = B(k, \Lambda_1(t))/k - \Lambda_1(t)$. The variance of B_t is smaller than $1/(4k)$, and this does not depend on Λ_1. Thus $\mathbb{E}B_t^2 | N_1^{(n)} \le S_{n1}/(4N_1^{(n)})$.

The random variable $N_1^{(n)}$ follows a Poisson distribution with parameter $\tau = \tau_n$. Taking expectations and noticing that $1/k \le 2/(k+1)$, we find

$$\mathbb{E}\frac{S_{n1}}{N_1^{(n)}} = \sum_{k=1}^{\infty} \frac{1}{k} e^{-\tau} \frac{\tau^k}{k!} \le \sum_{k=1}^{\infty} 2e^{-\tau} \frac{\tau^k}{(k+1)!} = 2\tau^{-1} \sum_{k=1}^{\infty} e^{-\tau} \frac{\tau^{k+1}}{(k+1)!} \le \frac{2}{\tau},$$

so that $\mathbb{E}B_t^2 \le (2\tau_n)^{-1}$ and $\mathbb{E}|B_t| \le (2\tau_n)^{-1/2}$. Thus ($d_K = b - a$)

$$\mathbb{E}X_{n1}S_{n1} \le \sqrt{d_K \int_a^b (2\tau_n)^{-1/2}\,\mathrm{d}t} = d_K (2\tau_n)^{-1/4}.$$

If instead of a Poisson process Π is a binomial process, then $N_1^{(n)} = \tau_n$ and $\mathbb{E}B_t^2 \le (4\tau_n)^{-1}$ so the same result holds with an improved constant (and a shorter proof). We conclude that $W_2(\widehat{\lambda}_n, \lambda)$ is smaller than the sum of terms of orders $n^{-1/2}$, $d_K(2\tau_n)^{-1/4}$, $\sqrt{C_{\psi,K}}\sigma_n$ and a last one that is identically zero for n large.

Proof (Proof of Theorem 4.6.5). The hypotheses guarantee that $\sqrt{n}(F_{\widehat{\lambda}_n}^{-1} - F_{\lambda_n}^{-1})$ is $o_{\mathbb{P}}(1)$ and so by Slutsky's theorem

$$\sqrt{n}\left(F^{-1}_{\widehat{\lambda}_n}-F^{-1}_{\lambda}\right)=\sqrt{n}(F^{-1}_{\widehat{\lambda}_n}-F^{-1}_{\lambda})\to GP\quad\text{weakly in } L_2(0,1),$$

where GP is the Gaussian process defined in the proof of Theorem 4.6.3. By the continuous mapping theorem, in order to conclude the weak convergence

$$\sqrt{n}(\mathbf{t}_{\lambda}^{\widehat{\lambda}_n}-\mathbf{i})=\sqrt{n}\left(F^{-1}_{\widehat{\lambda}_n}\circ F_\lambda-F^{-1}_{\lambda}\circ F_\lambda\right)=\left[\sqrt{n}\left(F^{-1}_{\widehat{\lambda}_n}-F^{-1}_{\lambda}\right)\right]\circ F_\lambda\to GP\circ F_\lambda,$$

in $L_2(K)$, it suffices to show that $h\mapsto h\circ F_\lambda$ is continuous from $L_2(0,1)$ to $L_2(K)$. Once this is shown, we can also write $Z=GP\circ F_\lambda$ as the (weak) limit of the process

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n F^{-1}_{\Lambda_i}\circ F_\lambda-F^{-1}_{\lambda}\circ F_\lambda\right)=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{t}_{\lambda}^{\Lambda_i}-\mathbf{i}\right)=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n T_i-\mathbf{i}\right),$$

which again by the central limit theorem in $L_2(K)$ is a mean zero Gaussian process and has covariance kernel

$$\kappa(s,t)=\mathbb{E}[(T(s)-s)(T(t)-t)]=\mathrm{cov}(T(s),T(t)),\qquad s,t\in\mathrm{int}K.$$

To prove the purported continuity of $h\mapsto h\circ F_\lambda$, we first notice that this map is linear, so one needs only show continuity at 0. This is a straightforward consequence of the change of variables formula: let $[a,b]=K$ and notice that F_λ is strictly increasing and piecewise continuously differentiable on $[a,b]$ with derivative bounded below by $\delta>0$. Hence for all $p\ge 1$

$$\|h\circ F_\lambda\|^p_{L_p(K)}=\int_a^b|h^p(F_\lambda(s))|\,\mathrm{d}s=\int_{F_\lambda(a)}^{F_\lambda(b)}|h^p(t)|\frac{1}{F'_\lambda(F^{-1}_\lambda(t))}\,\mathrm{d}t\le\frac{1}{\delta}\|h\|^p_{L_p(0,1)},$$

so when $p=2$ this map is $\delta^{-1/2}$-Lipschitz.

The very same map $h\mapsto h\circ F_\lambda$ is continuous from $L_2(0,1)$ to $L_2(\lambda)$ under the assumptions of the theorem.

Proof (Proof of Lemma 4.7.1). If $\mu\in\mathscr{W}_p(\mathscr{X})$ then $Y_n=W_p^p(\mu,\mu_n)\to 0$ almost surely by Proposition 2.2.6. Furthermore

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

Observe that Z_n is a sample average from the random variable $V=\int_{\mathscr{X}}\|x-X_1\|^p\,\mathrm{d}\mu(x)$ with finite expectation:

$$\mathbb{E}V=\int_{\mathscr{X}^2}\|x-y\|^p\,\mathrm{d}\mu\otimes\mu(x,y)\le 2^p\int_{\mathscr{X}^2}[\|x\|^p+\|y\|^p]\,\mathrm{d}\mu\otimes\mu(x,y)=2^{p+1}W_p^p(\mu,\delta_0)<\infty.$$

By Fatou's lemma

$$\mathbb{E}V=\mathbb{E}\liminf(Z_n-Y_n)\le\liminf\mathbb{E}[Z_n-Y_n]=\mathbb{E}V-\limsup\mathbb{E}Y_n.$$

Since Y_n is nonnegative, this implies that $\mathbb{E}Y_n \to 0$. Jensen's inequality yields

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

as $n \to \infty$.

Proof (Proof of Remark 4.7.2). For $\mu = (\delta_0+\delta_1)/2$ the quantity $W_p^p(\mu,\mu_2)$ equals $1/2$ or zero, each with probability $1/2$, so $\mathbb{E}W_p(\mu,\mu_2) = 2^{-1-\frac{1}{p}}$. The quantity $W_p^p(\mu,\mu_3)$ equals $1/2$ with probability $1/4$ and $1/6$ with probability $3/4$, so

$$\mathbb{E}W_p(\mu,\mu_3) = 2^{-\frac{1}{p}}\left[\frac{1}{4}+\frac{3}{4}3^{-\frac{1}{p}}\right] > 2^{-\frac{1}{p}}\left[\frac{1}{4}+\frac{1}{4}\right] = \mathbb{E}W_p(\mu,\mu_2)$$

for all $p > 1$.

Proof (Proof of Lemma 4.7.4). Let $k \sim B(n, q = \mu(A))$ denote the number of points from the sample $(X_1,\dots,X_n)$ that fall in A. Then a mass of $|k/n - q|$ must travel between A and B, a distance of at least $d_{\min}$. Thus $W_p^p(\mu_n,\mu) \ge d_{\min}^p |k/n-q|$, and so

$$\mathbb{E}W_p(\mu_n,\mu) = d_{\min}\mathbb{E}\sqrt[p]{\frac{1}{n}\left|q-\frac{k}{n}\right|} = d_{\min}n^{-1/(2p)}\mathbb{E}\sqrt[p]{\sqrt{n}\left|q-\frac{k}{n}\right|} = d_{\min}n^{-1/(2p)}\mathbb{E}\sqrt[p]{|Z_n|}.$$

Since k follows a binomial distribution, $Z_n = \sqrt{n}(q-k/n)$ converges weakly to $Z \sim N(0, q(1-q))$ with variance $q(1-q) > 0$. The convergence holds also in $\mathscr{W}_2$ by Theorem 2.2.1 and it also follows that $\mathbb{E}|Z_n|^{1/p} \to \mathbb{E}|Z|^{1/p}$. Thus

$$n^{1/(2p)}\mathbb{E}W_p(\mu_n,\mu) \to d_{\min}[q(1-q)]^{1/p}c_p > 0, \qquad \left(c_p = \mathbb{E}|N(0,1)|^{1/p}\right).$$

Proof (Proof of Proposition 4.7.6). Let μ be uniform on $[-1,-\theta]\cup[\theta,1]$ for $\theta \in (0,1)$. Then as in Lemma 4.7.4 with $d_{\min} = 2\theta$ and $q = 1/2$ we have

$$\liminf_{n\to\infty} n^{1/(2p)}\mathbb{E}W_p(\mu_n,\mu) \ge (2\theta)4^{-1/p}c_p.$$

We shall construct a measure with strictly positive density everywhere by letting $\theta = \theta_n \to 0$ slowly and put very small density on $[-\theta,\theta]$. Without loss of generality $\varepsilon_n < \cdots < \varepsilon_1 = 1$. Define a symmetric distribution function F by $F(\varepsilon_n) = 1/2 + 1/n^2$ and $F(-\varepsilon_n) = 1/2 - 1/n^2$ for all n, as follows: let g be an interpolation of ε_n, that is, $g : [1,\infty) \to (0,1]$ and $g(n) = \varepsilon_n$. It is possible to choose g smooth and strictly decreasing. In order to have $F(g(x)) = 1/2 + x^{-2}$ set $F(t) = 1/2 + [g^{-1}(t)]^{-2}$ for $t \in (0,1)$. The derivative $f = F'$ on $(0,1)$ is smooth and strictly positive by the inverse function theorem. Multiply f by a constant to make it integrate to $1/2$, define $f(-t) = f(t)$ for $t \in (-1,0)$ and let F be the corresponding distribution function and μ the resulting measure.

The event that k sample points fall in $[-1,-\varepsilon_n]$, $n-k$ points in $[\varepsilon_n,1]$ and none in $(-\varepsilon_n,\varepsilon_n)$ has probability

$$\binom{n}{k}\left(\frac{1}{2}-\frac{1}{n^2}\right)^k\left(\frac{1}{2}-\frac{1}{n^2}\right)^{n-k},$$

and when it occurs a mass of at least $|1/2-k/n|$ must travel at least $2\varepsilon_n$. We obtain

$$\begin{aligned}\mathbb{E}W_p(\mu_n,\mu) &\geq \sum_{k=0}^{n}\left(\frac{1}{2}-\frac{1}{n^2}\right)^k\left(\frac{1}{2}-\frac{1}{n^2}\right)^{n-k}\binom{n}{k}2\varepsilon_n\sqrt[p]{\left|\frac{1}{2}-\frac{k}{n}\right|}\\ &=\left(1-\frac{2}{n^2}\right)^n 2\varepsilon_n \quad 2^{-n}\sum_{k=0}^{n}\binom{n}{k}\sqrt[p]{\frac{1}{n}\left|\frac{n}{2}-k\right|}\approx 2^{1-1/p}c_p n^{-1/(2p)}\varepsilon_n,\end{aligned}$$

completing the proof. For example, $\varepsilon_n = 1/\sqrt{\log n}$ gives $F(x)=1/2+C\exp(-2/x^2)$ for $x>0$ close to zero.

Chapter 5
Proofs for Chapter 5

Proof (Proof of Lemma 5.1.1). By Ambrosio et al. [3, Proposition 6.2.12], there exist γ_0-null sets $\mathscr{N}_1$ such that on $\mathbb{R}^d \setminus \mathscr{N}_1$, $\mathbf{t}_{\gamma_0}^{\mu^1}$ is differentiable, $\nabla \mathbf{t}_{\gamma_0}^{\mu^1} > 0$ (positive definite), and $\mathbf{t}_{\gamma_0}^{\mu^1}$ is strictly monotone:

$$\left\langle \mathbf{t}_{\gamma_0}^{\mu^1}(x) - \mathbf{t}_{\gamma_0}^{\mu^1}(x'), x - x' \right\rangle > 0, \qquad x, x' \notin \mathscr{N}_1, \quad x \neq x',$$

and with weak inequalities on $\mathscr{N}_2, \ldots, \mathscr{N}_N$. Since $\mathbf{t}_{\gamma_0}^{\gamma_1} = (1-\tau)\mathbf{i} + \tau N^{-1} \sum_{i=1}^N \mathbf{t}_{\gamma_0}^{\mu^i}$, it stays strictly monotone (hence injective) and $\nabla \mathbf{t}_{\gamma_0}^{\gamma_1} > 0$ outside $\mathscr{N} = \cup \mathscr{N}_i$, which is a γ_0-null set.

Let h_0 denote the density of γ_0 and set $\Sigma = \mathbb{R}^d \setminus \mathscr{N}$. Then $\mathbf{t}_{\gamma_0}^{\gamma_1}|_\Sigma$ is injective and $\{h_0 > 0\} \setminus \Sigma$ is Lebesgue negligible because

$$0 = \gamma_0(\mathscr{N}) = \gamma_0(\mathbb{R}^d \setminus \Sigma) = \int_{\mathbb{R}^d \setminus \Sigma} h_0(x)\, \mathrm{d}x = \int_{\{h_0 > 0\} \setminus \Sigma} h_0(x)\, \mathrm{d}x,$$

and the integrand is strictly positive. Since $|\det \nabla \mathbf{t}_{\gamma_0}^{\mu^i}| > 0$ on Σ we obtain that $\gamma_1 = \mathbf{t}_{\gamma_0}^{\mu^i} \# \gamma_0$ is absolutely continuous by [3, Lemma 5.5.3].

Proof (Proof of Proposition 5.3.7). As has been established in the discussion before Proposition 5.3.6, the limit γ must be absolutely continuous, so η is well-defined.

In view of Theorem 2.2.1, it suffices to show that if $h : (\mathbb{R}^d)^{N+1} \to \mathbb{R}$ is any continuous nonnegative function such that

$$|h(t_1, \ldots, t_N, y)| \leq \frac{2}{N} \sum_{i=1}^N \|t_i\|^2 + 2\|y\|^2,$$

then

$$\int_{\mathbb{R}^d} g_n \,\mathrm{d}\gamma_n = \int_{(\mathbb{R}^d)^{N+1}} h \,\mathrm{d}\eta_n \to \int_{(\mathbb{R}^d)^{N+1}} h \,\mathrm{d}\eta \int_{\mathbb{R}^d} g \,\mathrm{d}\gamma, \qquad g_n(x) = h(\mathbf{t}_{\gamma_j}^{\mu^1}(x), \ldots \mathbf{t}_{\gamma_j}^{\mu^n}(x), x),$$

and g defined analoguosly.

Define $g_n : \mathbb{R}^d \to \mathbb{R}$ by $g_n(x) = h(\mathbf{t}_{\gamma_j}^{\mu^1}, \dots \mathbf{t}_{\gamma_j}^{\mu^n}, x)$ and analogously define g. We wish to show that $\int g_n \, \mathrm{d}\gamma_n \to \int g \, \mathrm{d}\gamma$; specifying $h = \|y - \bar{t}\|^2$ yields the statement of the proposition.

We denote the integrands by g_n and g respectively and divide the proof into several steps. It is perhaps instructive to assume in first reading that g_n and g are bounded and continuous, in which case one can jump directly to Step 2. The first of these assumptions is satisfied when the μ^i have bounded supports, and the second can be obtained under the regularity conditions in Theorem 1.6.7.

Step 0: redefinition on null sets. At a given $x \in \mathbb{R}^d$, $g_n(x)$ can be undefined, either because some $\mathbf{t}_{\gamma_n}^{\mu^i}(x)$ is empty, or because it can be multivalued (see Subsection 1.7.2). Redefine $g_n(x)$ at such points by setting it to 0 in the former case and choosing an arbitrary representative otherwise. Apply the same procedure for g. Then g_n and g are finite, nonnegative functions (in the proper sense) throughout $\mathbb{R}^d$. We claim that this modification is inconsequential and does not affect $\int g \, \mathrm{d}\gamma$. Indeed, the set of ambiguity points is a γ-null set: this is a consequence of the absolute continuity of γ, together with Remark 2.3 and Corollary 1.3 in Alberti & Ambrosio [2] (see the paragraphs preceding Assumptions 1 for a more detailed discussion). Similarly, the value of the integral $\int g_n \, \mathrm{d}\gamma_n$ remains unaltered after this modification. Finally, by Proposition 1.7.8, the set of points where g is not continuous is a γ-null set, before and after the modification, because h is continuous.

Step 1: approximation by bounded functions. Since γ_n converge in the Wasserstein space, they satisfy the uniform integrability condition (2.4) by Theorem 2.2.1, and hence the uniform absolute continuity (2.7) that we repeat here for convenience:

$$\forall \varepsilon > 0 \exists \delta > 0 \forall n \geq 1 \forall A \subseteq \mathbb{R}^d \text{ Borel}: \quad \gamma_n(A) \leq \delta \quad \Longrightarrow \int_A \|x\|^2 \, \mathrm{d}\gamma_n(x) < \varepsilon. \tag{5.1}$$

The δ's can be chosen in such a way that (5.1) holds true for the finite collection $\{\mu^1, \dots, \mu^N\}$ as well. Fix $\varepsilon > 0$, set $\delta = \delta_\varepsilon$ as in (5.1), and invoke (2.4) to find an $R = R_\varepsilon \geq 1$ such that

$$\forall i \, \forall n: \quad \int_{\{\|x\|^2 > R\}} \|x\|^2 \, \mathrm{d}\gamma_n(x) + \int_{\{\|x\|^2 > R\}} \|x\|^2 \, \mathrm{d}\mu^i(x) < \frac{\delta}{2N}.$$

The bound (holding γ_n-almost surely)

$$g_n(x) \leq 2\|x\|^2 + \frac{2}{N} \sum_{i=1}^N \|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2$$

implies that the sets $A_n = \{x : g_n(x) \geq 4R\}$ satisfy

$$A_n \subseteq \{x : \|x\|^2 > R\} \cup \bigcup_{i=1}^N \{x : \|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2 > R\}.$$

To deal with the sets in the union observe that

$$\gamma_n(\{x : \|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2 > R\}) = \mu^i(\{x : \|x\|^2 > R\}) < \frac{\delta}{2N},$$

so that $\gamma_n(A_n) < \delta$. We use this in conjunction with (5.1) to bound

$$\begin{aligned}\int_{A_n} g_n(x)\,\mathrm{d}\gamma_n(x) &\le 2\int_{A_n} \|x\|^2\,\mathrm{d}\gamma_n(x) + \frac{2}{N}\sum_{i=1}^{N}\int_{A_n} \|\mathbf{t}_{\gamma_n}^{\mu^i}(x)\|^2\,\mathrm{d}\gamma_n(x)\\ &\le 2\varepsilon + \frac{2}{N}\sum_{i=1}^{N}\int_{\mathbf{t}_{\gamma_n}^{\mu^i}(A_n)} \|x\|^2\,\mathrm{d}\mu^i(x) \le 4\varepsilon,\end{aligned}$$

where we have used the measure-preservation property $\mu^i(\mathbf{t}_{\gamma_n}^{\mu^i}(A_n)) = \gamma_n(A_n) < \delta$.

Define the truncation $g_{n,R}(x) = \min(g_n(x), 4R)$. Then $0 \le g_n - g_{n,R} \le g_n \mathbf{1}\{g_n > 4R\}$, so

$$\int [g_n(x) - g_{n,R}(x)]\,\mathrm{d}\gamma_n(x) \le \int_{A_n} g_n(x)\,\mathrm{d}\gamma_n(x) \le 4\varepsilon, \qquad n = 1, 2, \ldots.$$

The analogous truncated function g_R satisfies

$$0 \le g_R(x) \le 4R \quad \forall x \in \mathbb{R}^d \quad \text{and} \quad \{x : g_R \text{ is continuous }\} \text{ is of } \gamma\text{-full measure.} \tag{5.2}$$

Step 2: convergence of g_n to g. Let $E = \operatorname{supp}\gamma$. The sets

$$\mathscr{N}^i = (E \setminus E^{\mathrm{den}}) \cup \{x : \mathbf{t}_{\gamma}^{\mu^i}(x) \text{ contains more than one element}\}, \qquad i = 1, \ldots, N,$$

are γ-negligible and so is their union $\mathscr{N}$. As $n \to \infty$, Proposition 1.7.11 implies pointwise convergence (in a set-valued sense) of $\mathbf{t}_{\gamma_n}^{\mu^i}(x)$ to $\mathbf{t}_{\gamma}^{\mu^i}(x)$ for any $i = 1, \ldots, N$ and any $x \in E \setminus \mathscr{N}$. Thus $g_n \to g$ pointwise on $x \in E \setminus \mathscr{N}$ (for whatever choice of representatives selected to define g_n); consequently, $g_{n,R} \to g_R$ on $E \setminus \mathscr{N}$.

If E were compact, we could strengthen this to uniform convergence by Egorov's theorem. In order to restrict the integrands to a bounded set we invoke the tightness of the sequence (γ_n) and introduce a compact set K_ε such that $\gamma_n(\mathbb{R}^d \setminus K_\varepsilon) < \varepsilon/R$ for all n. Clearly, $g_{n,R} \to g_R$ on $E' = K_\varepsilon \cap E \setminus \mathscr{N}$, and by Egorov's theorem (valid as $\mathrm{Leb}(E') \le \mathrm{Leb}(K_\varepsilon) < \infty$), there exists a Borel set $\Omega = \Omega_\varepsilon \subseteq E'$ on which the convergence is uniform, and $\mathrm{Leb}(E' \setminus \Omega) < \varepsilon/R$. Let us write

$$\int g_{n,R}\,\mathrm{d}\gamma_n - \int g_R\,\mathrm{d}\gamma = \int g_R\,\mathrm{d}(\gamma_n - \gamma) + \int_{\Omega} (g_{n,R} - g_R)\,\mathrm{d}\gamma_n + \int_{\mathbb{R}^d \setminus \Omega} (g_{n,R} - g_R)\,\mathrm{d}\gamma_n,$$

and bound each of the three integrals on the right-hand side as $n \to \infty$.

Step 3: bounding the first two integrals. The first integral vanishes as $n \to \infty$, by (5.2) and the portmanteau lemma 1.7.1, as the bounded function g_R is continuous besides a γ-null set. The second integral obviously tends to 0 as $n \to \infty$, since $g_{n,R}$ converge to g_R uniformly on Ω.

Step 4: bounding the third integral. The integrand is smaller than $8R$, so the integral is bounded by $8R\gamma_n(\mathbb{R}^d\setminus\Omega)$. The complement of $\Omega\subseteq E' = E\cap K_\varepsilon\setminus\mathscr{N}$ is included in the union $\mathscr{N}\cup(E'\setminus\Omega)\cup(\mathbb{R}^d\setminus E)\cup(\mathbb{R}^d\setminus K_\varepsilon)$, where the first set is Lebesgue-negligible and the second has Lebesgue measure smaller than ε/R. The hypothesis of the densities of γ_n implies that $\gamma_n(A)\le C\mathrm{Leb}(A)$ for any Borel set $A\subseteq\mathbb{R}^d$ and any $n\in\mathbb{N}$; it follows from this and $\gamma_n(\mathbb{R}^d\setminus K_\varepsilon)<\varepsilon/R$ that

$$\left|\int_{\mathbb{R}^d\setminus\Omega}(g_{n,R}-g_R)\,\mathrm{d}\gamma_n\right|\le 8R(C\varepsilon/R+\gamma_n(\mathbb{R}^d\setminus E)+\varepsilon/R)=8\left(R\gamma_n(\mathbb{R}^d\setminus E)+C\varepsilon+\varepsilon\right).$$

The weak (or even Wasserstein) convergence of γ_n to γ alone does not suffice for bounding the limit $\gamma_n(\mathbb{R}^d\setminus E)$, because E is closed and the portmanteau lemma gives the inequality in the wrong direction. Once again the uniform bound on the densities comes to our rescue. Write the open set $E_1=\mathbb{R}^d\setminus E$ as a countable union of closed sets A_k with[1] $\mathrm{Leb}(E_1\setminus A_k)<1/k$, and conclude that

$$\limsup_{n\to\infty}\gamma_n(E_1)\le\limsup_{n\to\infty}\gamma_n(A_k)+\limsup_{n\to\infty}\gamma_n(E_1\setminus A_k)\le\gamma(A_k)+\frac{C}{k}=\frac{C}{k},$$

where we have used the portmanteau lemma again, $A_k\cap\mathrm{supp}(\gamma)=\emptyset$ and $\gamma_n(A)\le C\mathrm{Leb}(A)$.

Step 5: concluding. By Steps 3 and 4, we have for all k

$$\limsup_{n\to\infty}\left|\int g_{n,R}\,\mathrm{d}\gamma_n-\int g_R\,\mathrm{d}\gamma\right|\le\limsup_{n\to\infty}\left|\int_{\mathbb{R}^d\setminus\Omega}(g_{n,R}-g_R)\,\mathrm{d}\gamma_n\right|\le\frac{8R_\varepsilon C}{k}+8(C+1)\varepsilon.$$

Letting $k\to\infty$, then incorporating the truncation error yields

$$\limsup_{n\to\infty}\left|\int g_n\,\mathrm{d}\gamma_n-\int g\,\mathrm{d}\gamma\right|\le 8(C+1)\varepsilon+8\varepsilon.$$

The proof is complete upon noticing that ε is arbitrary.

[1] This is possible even if E_1 is unbounded: let $E_1^m=E_1\cap[-m,m]^d$, find a closed set $A_k^m\subseteq E_1^m$ with $\mathrm{Leb}(E_1^m\setminus A_k^m)<2^{-m}/k$ and choose $A_k=\cup_m A_k^m$, which stays closed even though the union is countable.

References

1. M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *Society for Industrial and Applied Mathematics*, 43(2):904–924, 2011.
2. G. Alberti and L. Ambrosio. A geometrical approach to monotone functions in $\mathbb{R}^n$. *Math. Z.*, 230(2):259–316, 1999.
3. L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Springer, 2nd edition, 2008.
4. R. M. Dudley. *Real Analysis and Probability*, volume 74. Cambridge University Press, 2002.
5. R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.
6. A. Karr. *Point Processes and their Statistical Inference*, volume 7. CRC press, 1991.
7. S. G. Krantz. *Convex Analysis*. Textbooks in Mathematics. CRC Press, 2014.
8. V. M. Panaretos and Y. Zemel. Amplitude and phase variation of point processes. *The Annals of Statistics*, 44(2):771–812, 2016.
9. C. Villani. *Optimal Transport: Old and New*. Springer, 2008.