ON THE GROMOV–PROHOROV DISTANCE

SVANTE JANSON

Abstract. We survey some basic results on the Gromov–Prohorov distance between metric measure spaces. (We do not claim any new results.)

We give several different definitions and show the equivalence of them. We also show that convergence in the Gromov–Prohorov distance is equivalent to convergence in distribution of the array of distances between finite sets of random points.

1. Introduction

Gromov [5] introduced a notion of convergence for metric measure spaces $X = (X, d, \mu)$, where (X, d) is a complete and separable metric space, and μ is a finite Borel measure on X. We assume in the sequel that μ is a probability measure, i.e., $\mu(X) = 1$; the extension to arbitrary finite measures (as in [5]) is straightforward and left to the reader.

Gromov's convergence can be expressed in terms of a metric, known as the *Gromov–Prohorov* metric. In fact, there are several natural definitions that are either completely equivalent, or equivalent within (small) constant factors; these include Gromov's original definition of $\underline{\square}_a$ [5, $3\frac{1}{2}.B$], and the version d_{GP} by Villani [11, p. 762] and Greven, Pfaffelhuber and Winter [4] (Definitions 3.1 and 3.4 below, respectively).

Gromov [5] also considered a different notion of convergence, based on distances between random points in the space (Definition 4.1 below), and proved a convergence criterion [5, p. 131] relating this and convergence in his metric. In fact, these are equivalent (Theorem 4.2).

The purpose of the present note is to survey some different definitions and give proofs of the equivalence of them. The results all are known, and we try to give original references, but there might be unintentional omissions.

Remark 1.1. The Gromov–Prohorov distance d_{GP} is closely related to the *Gromov–Hausdorff distance* d_{GH} ([3, Chapter 7], [11, Chapter 27]) and the *Gromov–Hausdorff–Prohorov distance* d_{GHP} ([11, p. 762], [9, Section 6]). Informally, convergence in the Gromov–Prohorov distance means that there is "almost" a measure preserving isometry, but this may ignore parts of the spaces with zero or small measure; convergence in the Gromov–Hausdorff distance does not involve measures at all, and means that the spaces are almost isometric; convergence in the Gromov–Hausdorff–Prohorov distance combines both aspects. □

Date: 27 May, 2020; revised 2 June, 2020.

2010 *Mathematics Subject Classification.* 60B05; 60B10, 54E35.

Partly supported by the Knut and Alice Wallenberg Foundation.

Remark 1.2. We consider throughout only complete separable metric spaces. Several of the definitions and results extend to more general metric spaces, but there are also serious technical problems in this case, and we prefer to say no more about it. □

2. Preliminaries

2.1. **Some notation.** We denote Lebesgue measure on $[0,1]$ by λ, and let $[0,1]$ denote the measure space $([0,1],m)$.

If $x \in X$, where X is a metric space, and $r > 0$, then $B(x,r) := \{y \in X : d(y,x) \leqslant r\}$ is the closed ball with centre x and radius r.

If X is a metric space, then $\mathcal{P}(X)$ is the space of all (Borel) probability measures on X. We equip $\mathcal{P}(X)$ with the standard topology of weak convergence; see e.g. [1] or [2].

If $\mu \in \mathcal{P}(X)$, then $\xi \sim \mu$ means that ξ is a random variable in X with distribution μ. We use $\xrightarrow{\mathrm{d}}$ and $\xrightarrow{\mathrm{p}}$ for convergence in distribution and in probability, respectively, of random variables.

If $\mu \in \mathcal{P}(X)$, then $\operatorname{supp}\mu$ denotes the *support* of μ, i.e., the smallest closed subset of X with full measure. We have

$$\operatorname{supp}\mu = \bigl\{x \in X : \mu(B(x,r)) > 0 \quad \forall r > 0\bigr\}. \tag{2.1}$$

If X and Y are metric spaces, $\varphi : X \to Y$ is measurable, and $\mu \in \mathcal{P}(X)$, then $\varphi_\sharp(\mu) \in \mathcal{P}(Y)$ denotes the *push-forward* of μ, defined by

$$f_\sharp(A) = f\bigl(\varphi^{-1}(A)\bigr) \tag{2.2}$$

for any measurable $A \subseteq Y$. Equivalently, if $\xi \sim \mu$, then $\varphi_\sharp(\mu)$ is the distribution of $\varphi(\xi)$ (which is a random variable in Y).

A measurable map $\varphi : (X,\mu) \to (Y,\nu)$, where (X,μ) and (Y,ν) are probability spaces, is *measure preserving* if $\varphi_\sharp(\mu) = \nu$.

2.2. **The Prohorov distance.** Let $X = (X,d)$ be a complete separable metric space.

If B is a subset of X and $\varepsilon > 0$, let

$$B^\varepsilon := \{x : d(x,B) \leqslant \varepsilon\}. \tag{2.3}$$

The *Prohorov distance* $d_{\mathsf{P},a}(\mu,\mu')$ (where $a > 0$ is a parameter, usually chosen to be 1) between two probability measures μ and μ' in $\mathcal{P}(X)$ is defined as the infimum of $\varepsilon > 0$ such that, for every Borel set $B \subseteq X$,

$$\mu'(B) \leqslant \mu(B^\varepsilon) + a\varepsilon. \tag{2.4}$$

It is easily seen that this is symmetric in μ and μ', and that (2.4) (for every B) implies also

$$\mu(B) \leqslant \mu'(B^\varepsilon) + a\varepsilon. \tag{2.5}$$

Remark 2.1. Note that different choices of the parameter a yield distances that are equivalent within constant factors. (We use a only for greater flexibility and precision in the equivalences below.) In fact, $d_{\mathsf{P},a}$ equals $a^{-1}d_{\mathsf{P},1}$ evaluated in the metric space (X,ad) with a rescaled metric. □

Remark 2.2. The Prohorov distance has also a dual formulation: $d_{\mathsf{P},a}(X'X')$ equals the minimal $\varepsilon \geqslant 0$ such that there exist two random variables $\xi \sim \mu$ and $\xi' \sim \mu'$ in X such that

$$\mathbb{P}\big(d(\xi,\xi') > \varepsilon\big) \leqslant a\varepsilon. \tag{2.6}$$

See [10, Corollary 7.5.2]. □

Remark 2.3. The Prohorov distance is a metric on $\mathcal{P}(X)$ that generates the weak topology [1, Appendix III], [2, Theorem 8.3.2]. □

See further [1], [2], [10] and the survey [6].

3. The Gromov–Prohorov distance

We give in this section several definitions of a (pseudo)distance between two (complete, separable) metric measure spaces $X = (X, d, \mu)$ and $X' = (X', d', \mu')$. The definitions are all equivalent within constant factors, and we can choose any of them as the definition of the *Gromov–Prohorov* distance $d_{\mathsf{GP}}(X, X')$. (Our default choice is $d_{\mathsf{GP}} := d_{\mathsf{GP},1}$.)

The original definition by Gromov [5, Section $3\frac{1}{2}$.3] can be written as follows. Here $a > 0$ is an arbitrary parameter; the distances $\underline{\square}_a$ for different values of a are obviously equivalent, and usually we choose $a = 1$.

Definition 3.1. $\underline{\square}_a(X, X')$ *is the infimum of* $\varepsilon > 0$ *such that there exist measure preserving maps* $\varphi : [0, 1] \to X$ *and* $\varphi' : [0, 1] \to X'$ *and a set* $W_\varepsilon \subseteq [0, 1]$ *such that*

$$\lambda(W_\varepsilon) \leqslant a\varepsilon \tag{3.1}$$

$$\big|d\big(\varphi(x_1), \varphi(x_2)\big) - d'\big(\varphi'(x_1), \varphi'(x_2)\big)\big| \leqslant \varepsilon, \quad x_1, x_2 \in [0, 1] \setminus W_\varepsilon. \tag{3.2}$$

We give an alternative, equivalent, definition. Recall that a *coupling* of the measures μ on X and μ' on X' is a probability measure ν on $X \times X'$ such that the marginals are μ and μ'. Recall also that a *relation* between X and X' is any subset $R \subseteq X \times X'$.

Definition 3.2. $\underline{\square}_a(X, X')$ *is the infimum of* $\varepsilon > 0$ *such that there exist a Borel relation* $R \subseteq X \times X'$ *and a coupling* ν *of* μ *and* μ', *such that*

$$\nu(R) \geqslant 1 - a\varepsilon, \tag{3.3}$$

$$(x_1, x_1'), (x_2, x_2') \in R \implies \big|d(x_1, x_2) - d'(x_1', x_2')\big| \leqslant \varepsilon. \tag{3.4}$$

It is easily seen that we may require the relation R to be closed.

Proposition 3.3. *Definitions 3.1 and 3.2 agree.*

Proof. Given φ, φ' and W_ε as in Definition 3.1, define

$$R_0 := \big\{\big(\varphi(x), \varphi'(x)\big) : x \in [0, 1] \setminus W_\varepsilon\big\}. \tag{3.5}$$

Then (3.2) shows that (3.4) holds for R_0. Let $R := \overline{R_0}$; then (3.4) holds by continuity.

Furthermore, let $\Phi := (\varphi, \varphi') : [0, 1] \to X \times X'$ and let ν be the probability measure $\Phi_\sharp(\lambda)$ on $X \times X'$. Then ν is a coupling of μ and μ', and

$$\nu(R) = \lambda\big(\Phi^{-1}(R)\big) \geqslant \lambda\big([0, 1] \setminus W_\varepsilon\big) \geqslant 1 - a\varepsilon. \tag{3.6}$$

Hence, the conditions in Definition 3.2 hold.

Conversely, suppose that R and ν are as in Definition 3.2. Then ν is a probability measure on the Polish space $X \times X'$, and thus there exists a measure preserving map $\Phi : [0,1] \to (X \times X', \nu)$, see [7, Theorem 3.19 or Lemma 3.22]. Write $\Phi = (\varphi, \varphi')$. Then, φ and φ' are measure preserving maps $[0,1] \to X$ and $[0,1] \to X'$. Let $W_\varepsilon := [0,1] \setminus \Phi^{-1}(R)$. Then (3.2) holds by (3.4), and

$$\lambda(W_\varepsilon) = 1 - \lambda(\Phi^{-1}(R)) = 1 - \nu(R) \leqslant a\varepsilon. \tag{3.7}$$

Hence, φ, φ' and W_ε are as in Definition 3.1. □

Another metric was defined by Villani [11, p. 762] and Greven, Pfaffelhuber and Winter [4].

Definition 3.4. *$d_{\mathsf{GP},a}(X, X')$ equals the infimum of $\varepsilon > 0$ such that there exists a metric space Z with subspaces $Y, Y' \subseteq Z$ and isometries $\varphi : X \to Y$ and $\varphi' : X' \to Y'$ such that the Prohorov distance*

$$d_{\mathsf{P},a}\bigl(\varphi_\sharp(\mu), \varphi'_\sharp(\mu')\bigr) \leqslant \varepsilon. \tag{3.8}$$

In other words, $d_{\mathsf{GP},a}(X, X')$ is the infimum of the Prohorov distance between $\varphi_\sharp(\mu)$ and $\varphi'_\sharp(\mu')$ over all metric spaces Z and isometric embeddings $\varphi : X \to Z$ and $\varphi' : X' \to Z$.

Note that we may assume that the metric space Z in Definition 3.4 is complete and separable, since otherwise we may replace Z by first its completion and then the closure of $Y \cup Y'$ (or conversely).

Proposition 3.5 (Löhr [8])**.** *For any metric measure spaces X and X' and any $a > 0$,*

$$\boxminus_a(X, X') = 2d_{\mathsf{GP},2a}(X, X'). \tag{3.9}$$

Proof. We argue as for the corresponding result for the Gromov–Hausdorff–Prohorov distance in [9]; see also [3, Section 7.3].

Let $\varepsilon > \boxminus_a(X, X')$ and let R and ν be as in Definition 3.2. Let $Z := X \sqcup X'$ be the disjoint union of X and X', and define a metric δ on Z that equals d on X, d' on X', and, for $x \in X$ and $x' \in X'$,

$$\delta(x, x') := \inf\bigl(d(x, y) + \varepsilon/2 + d'(y', x') : (y, y') \in R\bigr). \tag{3.10}$$

It is easily verified that this really defines a metric, see e.g. [9, Proof of Proposition 6], and that $\delta(x, x') = \varepsilon/2$ when $(x, x') \in R$.

Regard X and X' as subspaces of Z, and let (ξ, ξ') be a random variable in $X \times X'$ with distribution ν. If $(\xi, \xi') \in R$, then $\delta(\xi, \xi') = \varepsilon/2$; hence,

$$\mathbb{P}\bigl(\delta(\xi, \xi') > \varepsilon/2\bigr) \leqslant \mathbb{P}\bigl((\xi, \xi') \notin R\bigr) = 1 - \nu(R) \leqslant a\varepsilon = 2a \cdot \varepsilon/2. \tag{3.11}$$

Hence, see Remark 2.2, $d_{\mathsf{P},2a}(\mu, \mu') \leqslant \varepsilon/2$. Consequently, by Definition 3.4,

$$d_{\mathsf{GP},2a}(X, X') \leqslant d_{\mathsf{P},2a}(\mu, \mu') \leqslant \varepsilon/2. \tag{3.12}$$

Conversely, suppose that $d_{\mathsf{GP},2a}(X, X') \leqslant \varepsilon$. Then there exist Y, Y' and φ, φ' as in Definition 3.4, with a replaced by $2a$. We may assume that $X = Y$ and $X' = Y'$. Thus,

$$d_{\mathsf{P},2a}(\mu, \mu') \leqslant \varepsilon. \tag{3.13}$$

By Remark 2.2 there exist random variables ξ and ξ' in Z such that

$$\mathbb{P}\bigl[d(\xi,\xi') > \varepsilon\bigr] \leqslant 2a\varepsilon. \tag{3.14}$$

Let $R := \bigl\{(x,x') \in X \times X' : d(x,x') \leqslant \varepsilon\bigr\}$. This is a closed relation, and it follows from (3.14) that if ν is the distribution of (ξ,ξ'), then

$$\nu(R) = \mathbb{P}\bigl[(\xi,\xi') \in R\bigr] = \mathbb{P}\bigl[d(\xi,\xi') \leqslant \varepsilon\bigr] \geqslant 1 - 2a\varepsilon. \tag{3.15}$$

Furthermore, (3.4) holds with ε replaced by 2ε. Hence, Definition 3.2 shows that $\underline{\square}_a(X,X') \leqslant 2\varepsilon$. ☐

Remark 3.6. Definitions 3.2 and 3.4 are analogues of similar definitions in [3] and [9] for the related Gromov–Hausdorff and Gromov–Hausdorff–Prohorov distances. In particular, they correspond to the definition and Proposition 6 in [9, Section 6.2] if we ignore the Hausdorff part; note that the only significant difference between the conditions in [9, Proposition 6] and in Definition 3.2 above is that in [9] (for d_{GHP}), the coupling R is supposed to be a *correspondence*, i.e., the projections of $R \to X$ and $R \to X'$ are onto. (In other words, every $x \in R$ is related to some $x' \in X'$, and conversely.) ☐

Remark 3.7. It is easy to see, perhaps simplest from Definition 3.2, that the triangle inequality holds for $\underline{\square}_a$ and $d_{\mathsf{GP},a}$; hence the distances above are pseudometrics. Note that $d_{\mathsf{GP}}(X,X') = 0$ may hold not only for isomorphic X and X' (in the obvious sense that there exists a measure preserving bijection). In fact, for any $X = (X,\mu)$, if we let $X' := \operatorname{supp}\mu$, then

$$d_{\mathsf{GP}}\bigl((X,\mu),(X',\mu)\bigr) = 0. \tag{3.16}$$

We will see in Theorem 3.8 that this is essentially the only way that d_{GP} fails to be a metric. ☐

We note two basic results by Gromov [5], to which we refer for proofs.

Theorem 3.8 (Gromov [5, Corollary in $3\frac{1}{2}.6$]). *If $X = (X,\mu)$ and (X',μ') are metric measure spaces, then $d_{\mathsf{GP}}(X,X') = 0$ if and only if $(\operatorname{supp}\mu,\mu)$ and $(\operatorname{supp}\mu',\mu')$ are isomorphic metric measure space.*

In other words, d_{GP} is a metric on the set $\mathcal{X}$ of equivalence classes (under isomorphism) of metric measure spaces (X,μ) with full support, $\operatorname{supp}\mu = X$. ☐

Theorem 3.9 (Gromov [5, Corollary in $3\frac{1}{2}.12$]). *The metric space $(\mathcal{X}, d_{\mathsf{GP}})$ is complete and separable.* ☐

4. Convergence

Gromov [5] considered also convergence of metric spaces in terms of arrays of distances between points in the following way (in our notation).

For an integer $\ell \geqslant 1$, let $\mathcal{M}_\ell$ be the space of real $\ell \times \ell$ matrices; note that $\mathcal{M}_\ell = \mathbb{R}^{\ell^2}$ is a complete separable metric space.

For a metric space $X = (X,d)$ and $\ell \geqslant 1$, let $\rho_\ell : X^\ell \to \mathcal{M}_\ell$ be the map given by the entries

$$\rho_\ell(x_1,\dots,x_\ell)_{ij} = \rho_\ell(x_1,\dots,x_\ell;X,d)_{ij} := d(x_i,x_j). \tag{4.1}$$

If $X = (X,d,\mu)$ is a metric measure space, define for $\ell \geqslant 1$, the measure

$$\tau_\ell(X) = \tau_\ell(X,d,\mu) := \rho_{\ell\sharp}(\mu^\ell) \in \mathcal{P}(\mathcal{M}_\ell), \tag{4.2}$$

the push-forward of the measure $\mu^\ell \in \mathcal{P}(X^\ell)$ along ρ_ℓ. In our setting with a probability measure μ, we can, equivalently, define τ_ℓ by letting $\xi_1, \dots, \xi_\ell$ be i.i.d. (independent, identically distributed) random points in X with $\xi_i \sim \mu$; then

$$\tau_\ell(X) := \mathcal{L}\bigl(\rho_\ell(\xi_1, \dots, \xi_\ell; X)\bigr), \tag{4.3}$$

the distribution of the random matrix $\rho_\ell(\xi_1, \dots, \xi_\ell) \in \mathcal{M}_\ell$.

We then define convergence of a sequence of metric measure space as follows. (All unspecified limits below are as $n \to \infty$.)

Definition 4.1. *Let $(X_n)_1^\infty$ and X be metric measure spaces. We say that $X_n \xrightarrow{\mathrm{G}} X$ if for every $\ell \geqslant 1$,*

$$\tau_\ell(X_n) \to \tau_\ell(X) \qquad \text{in } \mathcal{P}(\mathcal{M}_\ell). \tag{4.4}$$

By (4.3), the condition (4.4) can also be written

$$\rho_\ell(\xi_1^{(n)}, \dots, \xi_\ell^{(n)}; X_n) \xrightarrow{\mathrm{d}} \rho_\ell(\xi_1, \dots, \xi_\ell; X), \tag{4.5}$$

where $(\xi_i^{(n)})$ are i.i.d. random points in X_n with $\xi_i^{(n)} \sim \mu_n$.

In fact, as stated in the next theorem, convergence in this sense is equivalent to convergence in the Gromov–Prohorov distance.

Theorem 4.2 (Greven, Pfaffelhuber and Winter [4, Theorem 5]). *Let $(X_n)_1^\infty$ and X be metric measure spaces. Then $X_n \xrightarrow{\mathrm{G}} X$ if and only if $d_{\mathsf{GP}}(X_n, X) \to 0$.*

Remark 4.3. We use the notation $\xrightarrow{\mathrm{G}}$ in honour of Gromov, since the property (4.4) is studied in [5]; see e.g. [5, $3\frac{1}{2}$.14], which discusses the relation with convergence in the Gromov–Prohorov distance. However (as far as we know), Theorem 4.2 is not stated explicitly in [5]. (The easy implication $\Longleftarrow$ is implicit in [5, $3\frac{1}{2}$.6]; the converse is almost, but not quite, in [5, $3\frac{1}{2}$.14].) $\square$

Remark 4.4. Gromov [5, $3\frac{1}{2}$.5] proved the far from obvious fact that if X and Y are two metric measure spaces such that the measures have full support, then

$$\tau_\ell(X) = \tau_\ell(Y), \qquad \forall \ell \geqslant 1 \tag{4.6}$$

if and only if X and Y are isomorphic.

Equivalently, for any metric measure spaces X and Y, (4.6) holds if and only if $d_{\mathsf{GP}}(X, Y) = 0$. (Cf. Theorem 3.8, which is proved in [5] using this fact.) $\square$

Remark 4.5. As remarked by Gromov [5, $3\frac{1}{2}$.14], if we instead of (4.4) just assume that

$$\tau_\ell(X_n) \to \nu_\ell, \qquad \ell \geqslant 1, \tag{4.7}$$

for some probability measures $\nu_\ell \in \mathcal{P}(\mathcal{M}_\ell)$, then X_n does not have to converge, i.e., (4.7) does not imply that the limits $\nu_\ell = \tau_\ell(X)$ for some metric measure space X. For example [5, $3\frac{1}{2}$.14 and $3\frac{1}{2}$.18], if X_n is the unit sphere

S^n with normalized surface measure and, say, the intrinsic (Riemannian) metric d_n, and $(\xi_i^{(n)})_i$ are i.i.d. uniformly random points in X_n, then,

$$d_n\bigl(\xi_i^{(n)},\xi_j^{(n)}\bigr) \xrightarrow{\text{p}} \pi/2, \tag{4.8}$$

for any distinct i and j, and thus (4.7) holds with ν_ℓ the point mass at the matrix $\frac{\pi}{2}\bigl(\mathbf{1}\{i \neq j\}_{i,j=1}^{\ell}\bigr)$. However, there is no metric measure space (X,d,μ) with $\tau_\ell(X)=\nu_\ell$, which would mean that if ξ_1,ξ_2 are i.i.d. random points in X with $\xi_i\sim\mu$, then $d(\xi_1,\xi_2)=\pi/2$ a.s. (This would imply that for any $r<\pi/2$ and μ-a.e. $x_1\in X$, $\mu\bigl(B(x,r)\bigr)=0$, and thus $x\notin\operatorname{supp}\mu$; hence $\mu(\operatorname{supp}\mu)=0$, a contradiction.) ∎

Remark 4.6. We have (implicitly) assumed above that ℓ is a finite integer. However, we can also use the same definitions (4.1)–(4.3) for $\ell=\infty$, noting that $M_\infty=\mathbb{R}^{\infty^2}$ still is a Polish space, i.e., it can be regarded as a complete separable metric space. (The choice of metric is of no importance to us.)

It is easy to see that the condition (4.4), or equivalently (4.5), for every finite ℓ is equivalent to the same condition for $\ell=\infty$. Hence, $X_n \xrightarrow{\text{G}} X$ can also be defined by $\tau_\infty(X_n)\to\tau_\infty(X)$ in $\mathcal{P}(\mathcal{M}_\infty)$, or by

$$\rho_\infty(\xi_1^{(n)},\xi_2^{(n)},\dots;X_n) \xrightarrow{\text{d}} \rho_\infty(\xi_1,\xi_2,\dots;X). \tag{4.9}$$

∎

Proof of Theorem 4.2. Let $X_n=(X_n,d_n,\mu_n)$ and $X=(X,d,\mu)$. As above, let $(\xi_i^{(n)})_i$ be i.i.d. random points in X_n with $\xi_i^{(n)}\sim\mu_n$, and let $(\xi_i)_i$ be i.i.d. random points in X with $\xi_i\sim\mu$. (We may also write $\xi^{(n)}$ and ξ without index when the index does not matter.)

First, suppose that $d_{\mathsf{GP}}(X_n,X)\to 0$. By Proposition 3.5, then $\underline{\Box}_1(X_n,X)\to 0$, and thus there exists a sequence $\varepsilon_n\to 0$ such that $\underline{\Box}_1(X_n,X)<\varepsilon_n$ and hence, see Definition 3.2, there exists a coupling ν_n of μ_n and μ and a Borel relation $R_n\subseteq X_n\times X$ such that (3.3)–(3.4) hold for ν_n, R_n and ε_n (with d_n and d). We may assume that each pair $(\xi_i^{(n)},\xi_i)$ has the distribution ν on $X_n\times X$; thus

$$\mathbb{P}\bigl((\xi_i^{(n)},\xi_i)\in R_n\bigr)\geqslant 1-\varepsilon_n \tag{4.10}$$

by (3.3). Together with (3.4) and the definition (4.1), this implies

$$\begin{aligned}\mathbb{P}\Bigl(\bigl|\rho_\ell(\xi_1^{(n)},\dots,\xi_\ell^{(n)};X_n)-\rho_\ell(\xi_1,\dots,\xi_\ell;X)\bigr|\leqslant \ell^2\varepsilon_n\Bigr)&\\ \geqslant \mathbb{P}\bigl((\xi_i^{(n)},\xi_i)\in R_n, i=1,\dots,\ell\bigr)\geqslant 1-\ell\varepsilon_n.&\end{aligned} \tag{4.11}$$

This implies easily (4.5) for each ℓ, and thus $X_n\xrightarrow{\text{G}}X$.

Conversely, suppose that $X_n\xrightarrow{\text{G}}X$, so that (4.5) holds. Fix $r>0$, let $h(t):=(1-t/r)_+$ for $t\geqslant 0$, and define $g_n: X_n\to[0,\infty)$ by

$$g_n(x):=\mathbb{E}\,h\bigl(d_n(x,\xi^{(n)})\bigr),\qquad n\geqslant 1, \tag{4.12}$$

and similarly $g: X\to[0,\infty)$ by $g(x):=\mathbb{E}\,h\bigl(d(x,\xi)\bigr)$ Then $0\leqslant h\leqslant 1$ and $h\bigl(d_n(x,y)\bigr)=0$ unless $y\in B(x,r)$; hence

$$0\leqslant g_n(x)\leqslant \mu_n\bigl(B(x,r)\bigr). \tag{4.13}$$

For any $m \geqslant 1$, we have

$$g_n(x)^m = \mathbb{E} \prod_{i=1}^{m} h\bigl(d_n(x, \xi_i^{(n)})\bigr) \tag{4.14}$$

and thus, if we define $H : \mathcal{M}_{m+1} \to \mathbb{R}$ by $H\bigl((a_{ij})_{i,j}\bigr) := \prod_{i=1}^m h(a_{m+1,i})$,

$$\begin{aligned}\mathbb{E}\bigl[g_n(\xi^{(n)})^m\bigr] &= \mathbb{E} \prod_{i=1}^{m} h\bigl(d_n(\xi_{m+1}^{(n)}, \xi_i^{(n)})\bigr) \\ &= \mathbb{E}\, H\bigl(\rho_{m+1}(\xi_1^{(n)}, \dots, \xi_{m+1}^{(n)}; X_n)\bigr). \end{aligned} \tag{4.15}$$

Similarly,

$$\mathbb{E}\bigl[g(\xi)^m\bigr] = \mathbb{E}\, H\bigl(\rho_{m+1}(\xi_1, \dots, \xi_{m+1}; X)\bigr). \tag{4.16}$$

Note that H is a bounded continuous function on $\mathcal{M}_{m+1}$. Consequently, the assumption $X_n \xrightarrow{\mathrm{G}} X_\infty$ implies by (4.5) and (4.15)–(4.16)

$$\mathbb{E}[g_n(\xi^{(n)})^m] \to \mathbb{E}[g(\xi)^m], \qquad m \geqslant 1. \tag{4.17}$$

Thus, by the method of moments (recalling that $g(\xi)$ is bounded by (4.13), and thus its distribution is determined by its moments)

$$g_n(\xi^{(n)}) \xrightarrow{\mathrm{d}} g(\xi). \tag{4.18}$$

If $x \in \operatorname{supp} \mu$, then $\mathbb{P}\bigl(d(x,\xi) \leqslant r/2\bigr) = \mu\bigl(B(x, r/2)\bigr) > 0$ and thus $g(x) > 0$. Hence, $g(x) > 0$ μ-a.e., i.e.,

$$g(\xi) > 0 \qquad \text{a.s.} \tag{4.19}$$

Fix $\varepsilon > 0$. By (4.19), there exists $\kappa > 0$ such that $\mathbb{P}\bigl(g(\xi) \leqslant \kappa\bigr) < \varepsilon$. Then, by (4.18), there exists n_0 such that if $n \geqslant n_0$, then

$$\mathbb{P}\bigl(g_n(\xi^{(n)}) \leqslant \kappa\bigr) < \varepsilon. \tag{4.20}$$

Consider only $n \geqslant n_0$, and let

$$A_n := \bigl\{x \in X_n : \mu_n\bigl(B(x,r)\bigr) \geqslant \kappa\bigr\}. \tag{4.21}$$

By (4.13) and (4.20), we have

$$\mu_n(A_n) = \mathbb{P}\bigl(\mu_n(B(\xi^{(n)}, r)) \geqslant \kappa\bigr) \geqslant \mathbb{P}\bigl(g_n(\xi^{(n)}) \geqslant \kappa\bigr) > 1 - \varepsilon. \tag{4.22}$$

Pick recursively points $x_{n1}, x_{n2}, \dots, x_{iN}$ in A_n such that the balls $B_{ni} := B(x_{ni}, r)$ are disjoint, and stop when this is no longer possible. Since $\mu_n(B_{ni}) \geqslant \kappa$ for every i by the definition of A_n, this process has to stop at some $N = N_n \leqslant 1/\kappa$.

If $x \in A_n$, then $B(x,r)$ has to intersect some $B_{ni} = B(x_{ni}, r)$, and thus $x \in B(x_{ni}, 2r)$. Consequently, A_n is covered by the N balls $\widetilde{B}_{ni} := B(x_{ni}, 2r)$. Hence, by (4.22),

$$\mu_n\Bigl(\bigcup_{i=1}^{N} \widetilde{B}_{ni}\Bigr) \geqslant \mu_n(A_n) > 1 - \varepsilon. \tag{4.23}$$

Furthermore, since $X_n \xrightarrow{\mathrm{G}} X$, and thus by (4.5) $\rho_2(\xi_1^{(n)}, \xi_2^{(n)}; X_n) \xrightarrow{\mathrm{d}} \rho_2(\xi_1, \xi_2; X)$, we have $d_n(\xi_1^{(n)}, \xi_2^{(n)}) \xrightarrow{\mathrm{d}} d(\xi_1, \xi_2)$, and thus the sequence

$d_n(\xi_1^{(n)}, \xi_2^{(n)})$ of random variables is tight [7, Lemma 4.8]. Hence, there exists $D < \infty$ such that for all n,

$$\mathbb{P}\bigl(d_n(\xi_1^{(n)}, \xi_2^{(n)}) > D\bigr) < \kappa^2. \tag{4.24}$$

Suppose now that $x, y \in A_n$ and $d_n(x, y) > D + 2r$. If $\xi_1^{(n)} \in B(x, r)$ and $\xi_2^{(n)} \in B(y, r)$, then $d_n(\xi_1^{(n)}, \xi_2^{(n)}) \geqslant d_n(x, y) - 2r > D$; consequently, using the independence of $\xi_1^{(n)}$ and $\xi_2^{(n)}$ together with the definition (4.21) of A_n,

$$\mathbb{P}\bigl(d_n(\xi_1^{(n)}, \xi_2^{(n)}) > D\bigr) \geqslant \mathbb{P}\bigl(\xi_1^{(n)} \in B(x, r)\bigr)\, \mathbb{P}\bigl(\xi_2^{(n)} \in B(y, r)\bigr) \geqslant \kappa^2, \tag{4.25}$$

which contradicts (4.24). Consequently, $d_n(x, y) \leqslant D + 2r$ whenever $x, y \in A_n$, i.e.,

$$\operatorname{diam}(A_n) \leqslant D + 2r \tag{4.26}$$

and thus

$$\operatorname{diam}\Bigl(\bigcup_{i=1}^{N} \widetilde{B}_{ni}\Bigr) \leqslant D + 6r. \tag{4.27}$$

We have shown that for each positive ε and r, there exists n_0, N_0 $(= 1/\kappa^2)$ and D_1 $(= D + 6r)$ such that for each $n \geqslant n_0$, there exists a collection $\{\widetilde{B}_{ni}\}_{i=1}^{N_n}$ of subsets $\widetilde{B}_{ni} \subseteq X_n$ such that

$$N_n \leqslant N_0, \tag{4.28}$$

$$\operatorname{diam}\bigl(\widetilde{B}_{ni}\bigr) \leqslant 4r \tag{4.29}$$

$$\operatorname{diam}\Bigl(\bigcup_{i=1}^{N} \widetilde{B}_{ni}\Bigr) \leqslant D_1, \tag{4.30}$$

$$\mu_n\Bigl(X_n \setminus \bigcup_{i=1}^{N} \widetilde{B}_{ni}\Bigr) < \varepsilon. \tag{4.31}$$

Furthermore, by increasing N_0 and D_1 if necessary, this holds also for each $n < n_0$, as an easy consequence of the fact that each μ_n is a tight measure (as is every probability measure in a Polish space [1, Theorem 1.4]).

This shows that the sequence (X_n) satisfies condition III in the corollary on p. 131–132 in Gromov [5]; since we also assume (4.4), this corollary shows that X_n converges to some metric measure space Y in $\underline{\square}_1$, or equivalently in d_{GP}. Furthermore (as in the proof of this corollary in [5]), $d_{\mathsf{GP}}(X_n, Y) \to 0$ implies $\tau_\ell(X_n) \to \tau_\ell(Y)$ by the first part of the proof, and thus $\tau_\ell(Y) = \tau_\ell(X)$ for every $\ell \geqslant 1$. By Remark 4.4, this implies $d_{\mathsf{GP}}(X, Y) = 0$, and thus also $d_{\mathsf{GP}}(X_n, X) \to 0$. □

Acknowledgement

I thank Gabriel Berzunza Ojeda for help with references.

References

[1] Patrick Billingsley: *Convergence of Probability Measures.* Wiley, New York, 1968.

[2] V. I. Bogachev: *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[3] Dmitri Burago, Yuri Burago & Sergei Ivanov: *A Course in Metric Geometry.* American Mathematical Society, Providence, RI, 2001.

[4] Andreas Greven, Peter Pfaffelhuber & Anita Winter: Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Prob. Theo. Rel. Fields* **145** (2009), no. 1-2, 285–322.

[5] Misha Gromov: *Metric Structures for Riemannian and Non-Riemannian Spaces.* Birkhäuser, Boston, MA, 1999, 2001.

[6] Svante Janson: Probability distances. Preprint, 2020. `http://www2.math.uu.se/~svante/papers/sjN21.pdf`

[7] Olav Kallenberg: *Foundations of Modern Probability.* 2nd ed., Springer, New York, 2002.

[8] Wolfgang Löhr: Equivalence of Gromov–Prohorov– and Gromov's $\underline{\Box}_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* **18** (2013), no. 17, 10 pp.

[9] Grégory Miermont: Tessellations of random maps of arbitrary genus. *Ann. Sci. Éc. Norm. Supér. (4)* **42** (2009), no. 5, 725–781.

[10] Svetlozar T. Rachev, Lev B. Klebanov, Stoyan V. Stoyanov & Frank J. Fabozzi: *The Methods of Distances in the Theory of Probability and Statistics.* Springer, New York, 2013.

[11] Cédric Villani: *Optimal transport.* Springer-Verlag, Berlin, 2009.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

Email address: `svante.janson@math.uu.se`

URL: `http://www2.math.uu.se/∼svante/`