REAL TREES

SVANTE JANSON

Abstract. We survey the definition and some elementary properties of real trees. The main purpose is to give a number of different definitions and show the equivalence between them. We discuss also the four-point inequality, and mention the connection to the theory of Gromov hyperbolic spaces.

1. Introduction

There are several different but equivalent definitions of real trees (also called $\mathbb{R}$-trees). We collect several of them as follows. We define below conditions (T1) and (T2a)–(T2j) on a metric space (T, d); we will show that assuming (T1), the conditions (T2a)–(T2j) are all equivalent. We then make the following definition. (Which we state already here, although it is not yet justified.)

Definition 1.1. A *real tree* (or $\mathbb{R}$-*tree*) is a metric space $T = (T, d)$ that satisfies condition (T1) and one (and thus all) of (T2a)–(T2j).

Remark 1.2. Some authors assume also that the metric space T is complete. We will not do so. See further Remark 6.6 below. □

1.1. **Some notation.** Throughout, $T = (T, d)$ is a metric space. We often write $d_{x,y}$ for $d(x, y)$.

$B(x, r) := \{y : d(x, y) < r\}$ denotes the open ball with centre $x \in T$ and radius $r > 0$.

If $\psi_1 : [0, a] \to T$ and $\psi_2 : [0, b] \to T$ are continuous maps with $\psi_1(a) = \psi_2(0)$, their *concatenation* $\psi_1 * \psi_2 : [0, a + b] \to T$ is defined by

$$\psi_1 * \psi_2(t) := \begin{cases} \psi_1(t), & 0 \leqslant t \leqslant a, \\ \psi_2(t - a), & a \leqslant t \leqslant a + b. \end{cases} \tag{1.1}$$

The concatenation is clearly a continuous map $[0, a + b] \to T$.

If $s, t \in \mathbb{R}$, then $s \wedge t := \min\{s, t\}$ and $s \vee t := \max\{s, t\}$. (These operations have priotiry over addition and subtraction.)

2. The conditions

In this section we state the conditions on a metric space $T = (T, d)$, beginning with the central (T1).

(T1) For any $x, y \in T$, there exists a unique isometric embedding $\varphi_{x,y}$ of the closed interval $[0, d_{x,y}] \subset \mathbb{R}$ into T such that $\varphi_{x,y}(0) = x$ and $\varphi_{x,y}(d_{x,y}) = y$.

Date: 8 June, 2020; revised 9 June 2020.

Partly supported by the Knut and Alice Wallenberg Foundation.

Assume that (T1) holds. We then denote the image $\varphi_{x,y}([0,d_{x,y}]) \subseteq T$ by $[x,y]$; thus $[x,y]$ is a connected compact subset of T, homeomorphic with $[0,1]$ if $x \neq y$. We similarly define $[x,y) = [x,y] \setminus \{y\}$, $(x,y] = [x,y] \setminus \{x\}$, $(x,y) = [x,y] \setminus \{x,y\}$. (If $x = y$, then $[x,x] = \{x\}$, and $[x,x) = (x,x] = (x,x) = \emptyset$.)

Obviously, $\varphi_{y,x}(t) = \varphi_{x,y}(d_{x,y} - t)$ and $[y,x] = [x,y]$.

Furthermore, still assuming (T1), let $x,y,z \in T$. Since $\varphi_{x,y}$ and $\varphi_{x,z}$ are isometries,

$$[x,y] \cap [x,z] = \bigl\{\varphi_{x,y}(t) : t \in [0, d_{x,y} \wedge d_{x,z}],\, \varphi_{x,y}(t) = \varphi_{x,z}(t)\bigr\}. \tag{2.1}$$

We define, noting that the maximum exists (i.e., the supremum is attained) by continuity,

$$\Delta(x,y,z) := \max\bigl\{t \in [0, d_{x,y} \wedge d_{x,z}] : \varphi_{x,y}(t) = \varphi_{x,z}(t)\bigr\}, \tag{2.2}$$

$$\gamma(x,y,z) := \varphi_{x,y}(\Delta(x,y,z)) = \varphi_{x,z}(\Delta(x,y,z)) \in [x,y] \cap [x,z]. \tag{2.3}$$

Further properties of these objects are given in Section 3.

We turn to the conditions (T2a)–(T2j). These are stated for a metric space $T = (T,d)$ such that (T1) holds, so we can use the notations just introduced.

(T2a) For any $x,y \in T$ and any $z \in (x,y)$, x and y are in different components of $T \setminus \{z\}$.
(T2b) For any $x,y,z \in T$, $[y,z] \subseteq [x,y] \cup [x,z]$.
(T2c) For any $x,y,z \in T$, $[x,y] \cap [x,z] \cap [y,z] \neq \emptyset$.
(T2d) For any $x,y,z \in T$, $\gamma(x,y,z) \in [y,z]$.
(T2e) For any injective continuous map $\psi : [0,1] \to T$,

$$d\bigl(\psi(0),\psi(t)\bigr) + d\bigl(\psi(t),\psi(1)\bigr) = d\bigl(\psi(0),\psi(1)\bigr), \qquad t \in [0,1]. \tag{2.4}$$

(T2f) For any injective continuous map $\psi : [0,1] \to T$, $\psi([0,1]) \subseteq [\psi(0),\psi(1)]$.
(T2g) For any injective continuous map $\psi : [0,1] \to T$, $\psi([0,1]) = [\psi(0),\psi(1)]$.
(T2h) Any injective continuous map $\psi : [0,1] \to T$ equals $\varphi_{x,y}$ up to parametrization, where $x = \psi(0)$ and $y = \psi(1)$; i.e., $\psi = \varphi_{x,y} \circ h$ for some strictly increasing homeomorphism $[0,1] \to [0,d_{x,y}]$.
(T2i) For any injective continuous map $\psi : [0,1] \to T$, $\psi([0,1]) \supseteq [\psi(0),\psi(1)]$.
(T2j) For any continuous map $\psi : [0,1] \to T$, $\psi([0,1]) \supseteq [\psi(0),\psi(1)]$.

As said in the introduction, we have the following equivalences.

Theorem 2.1. *Assume that $T = (T,d)$ is a metric space such that* (T1) *holds. Then* (T2a)–(T2j) *are all equivalent.*

The proof is given in Section 4.

Remark 2.2. Condition (T1) alone is not sufficient. Examples of spaces satisfying (T1) without being real trees are the Euclidean space $\mathbb{R}^d$, $d \geqslant 2$, and any convex subset of $\mathbb{R}^d$ of dimension $\geqslant 2$; for example the unit disc. □

3. Consequences of (T1)

In this section we assume (T1) (and sometimes further conditions), and show some lemmas used in the proof of Theorem 2.1.

Lemma 3.1. *Suppose that* (T1) *holds. Then T is connected, pathwise connected, and locally pathwise connected. Hence, if $V \subset T$ is an open subset of T, then V is a union of open (pathwise) connected components.*

Proof. T is obviously pathwise connected by (T1). Thus, T is connected.

Furthermore, T is locally pathwise connected, since every open ball $B(x,r)$ is pathwise connected. (Every $y \in B(x,r)$ is connected to x by the path $[x,y] \subseteq B(x,r)$.) □

In particular, if $z \in T$, then the components of $T \setminus \{z\}$ are open and pathwise connected. These components are called the *branches* at z.

Lemma 3.2. *Suppose that* (T1) *holds. If $x,y \in T$ and $z,w \in [x,y]$, then $[z,w] \subseteq [x,y]$, and, furthermore,*

$$\varphi_{z,w}(t) := \varphi_{x,y}(d_{x,z}+t), \qquad 0 \leqslant t \leqslant d_{z,w}. \tag{3.1}$$

Proof. By symmetry, we may assume $d_{x,z} \leqslant d_{x,w}$. Since $\varphi_{x,y}$ is an isometry with $\varphi_{x,y}(0) = x$, we have $z = \varphi_{x,y}(d_{x,z})$ and $w = \varphi_{x,y}(d_{x,w})$; furthermore, $d(z,w) = |d_{x,z} - d_{x,w}| = d_{x,w} - d_{x,z}$. Let

$$\varphi(t) := \varphi_{x,y}(d_{x,z}+t), \qquad 0 \leqslant t \leqslant d_{x,w} - d_{x,z} = d_{z,w}. \tag{3.2}$$

Then φ is an isometry and it follows that $\varphi = \varphi_{z,w}$. The result follows. □

Lemma 3.3. *Suppose that* (T1) *holds. Then, for any $x,y \in T$,*

$$[x,y] = \{z : d(x,z) + d(z,y) = d(x,y)\}. \tag{3.3}$$

Proof. If $z \in [x,y]$, then by definition $z = \varphi_{x,y}(s)$ for some $s \in [0, d_{x,y}]$. Since $\varphi_{x,y}$ is an isometry, we have $d(x,z) = d\big(\varphi_{x,y}(0), \varphi_{x,y}(s)\big) = s$ and $d(z,y) = d\big(\varphi_{x,y}(s), \varphi_{x,y}(d_{x,y})\big) = d(x,y) - s$. Hence,

$$d(x,z) + d(z,y) = s + \big(d(x,y) - s\big) = d(x,y). \tag{3.4}$$

Conversely, suppose that $z \in T$ with $d(x,z) + d(z,y) = d(x,y)$. Define $\varphi : [0, d_{x,y}] \to T$ as the concatenation $\varphi := \varphi_{x,z} * \varphi_{z,y}$, see (1.1). Then, φ is a continuous map $[0, d_{x,y}] \to T$ with $\varphi(0) = x$ and $\varphi(d_{x,y}) = y$. Furthermore, φ is an isometry on $[0, d_{x,z}]$ and on $[d_{x,z}, d_{x,y}]$. It follows that if $s \in [0, d_{x,z}]$ and $t \in [d_{x,z}, d_{x,y}]$, then, by the triangle inequality,

$$\begin{aligned} d\big(\varphi(s), \varphi(t)\big) &\leqslant d\big(\varphi(s), \varphi(d_{x,z})\big) + d\big(\varphi(d_{x,z}), \varphi(t)\big) \\ &= \big(d_{x,z} - s\big) + \big(t - d_{x,z}\big) = t - s. \end{aligned} \tag{3.5}$$

On the other hand, if we have strict inequality $d\big(\varphi(s), \varphi(t)\big) < t - s$ for some s,t with $0 \leqslant s \leqslant t \leqslant d_{x,y}$, then, similarly,

$$\begin{aligned} d(x,y) &\leqslant d\big(x, \varphi(s)\big) + d\big(\varphi(s), \varphi(t)\big) + d\big(\varphi(t), y\big) \\ &= s + d\big(\varphi(s), \varphi(t)\big) + d_{x,y} - t \\ &< s + (t-s) + (d_{x,y} - t) = d_{x,y} = d(x,y), \end{aligned} \tag{3.6}$$

a contradiction.

Consequently, φ is an isometry, and thus $\varphi = \varphi_{x,y}$, by the uniquness assumption in (T1). Hence, $z = \varphi(d_{x,z}) = \varphi_{x,y}(d_{x,z}) \in [x,y]$. □

Lemma 3.4. *Suppose that* (T1) *holds. If $x,y \in T$ and $z,w \in [x,y]$ with $z \in [x,w]$, then $w \in [z,y]$.*

Proof. Since $z \in [x,w]$, we have $d_{x,z} \leqslant d_{x,w}$. Hence, by Lemma 3.3,

$$d_{y,w} = d_{x,y} - d_{x,w} \leqslant d_{x,y} - d_{x,z} = d_{y,z}. \tag{3.7}$$

Hence, by Lemma 3.2,

$$w = \varphi_{y,x}(d_{y,w}) = \varphi_{y,z}(d_{y,w}) \in [y,z]. \tag{3.8}$$

□

Lemma 3.5. *Suppose that* (T1) *holds. Then, for any $x,y,z \in T$,*

$$\bigl\{t \in [0, d_{x,y} \wedge d_{x,z}] : \varphi_{x,y}(t) = \varphi_{x,z}(t)\bigr\} = [0, \Delta(x,y,z)] \tag{3.9}$$

and

$$[x,y] \cap [x,z] = [x, \gamma(x,y,z)] = \bigl\{\varphi_{x,y}(t) : t \in [0, \Delta(x,y,z)]\bigr\}. \tag{3.10}$$

Proof. Let

$$J := \bigl\{t \in [0, d_{x,y} \wedge d_{x,z}] : \varphi_{x,y}(t) = \varphi_{x,z}(t)\bigr\}. \tag{3.11}$$

Thus, the definition (2.2) says $\Delta(x,y,z) = \max J$. If $t \in J$, let $w := \varphi_{x,y}(t) \in [x,y] \cap [x,z]$. Then $[x,w] \subseteq [x,y] \cap [x,z]$ by Lemma 3.2. Hence, if $0 \leqslant s \leqslant t$, then

$$\varphi_{x,y}(s) = \varphi_{x,w}(s) = \varphi_{x,z}(s) \tag{3.12}$$

and consequently, $s \in J$. This shows that J is an interval, and (3.9) follows.

Finally, (3.10) follows from (3.9), using (2.1), (2.3) and (3.12). □

Lemma 3.6. *Suppose that* (T1) *and* (T2d) *hold. Then, for any $x,y,z \in T$,*

$$[x,y] \cap [x,z] \cap [y,z] = \{\gamma(x,y,z)\} \tag{3.13}$$

and

$$\Delta(x,y,z) = \tfrac{1}{2}\bigl(d(x,y) + d(x,z) - d(y,z)\bigr). \tag{3.14}$$

In particular, Δ is a continuous function $T^3 \to \mathbb{R}$, and $\gamma(x,y,z)$ is a symmetric function of x,y,z.

Proof. By the definition (2.3) and the assumption (T2d),

$$\gamma(x,y,z) \in [x,y] \cap [x,z] \cap [y,z]. \tag{3.15}$$

Let $w \in [x,y] \cap [x,z] \cap [y,z]$. Then, by Lemma 3.3,

$$d_{x,w} + d_{y,w} = d_{x,y}, \tag{3.16}$$

$$d_{x,w} + d_{z,w} = d_{x,z}, \tag{3.17}$$

$$d_{y,w} + d_{z,w} = d_{y,z}, \tag{3.18}$$

and consequently

$$2d_{x,w} = d_{x,y} + d_{x,z} - d_{y,z}. \tag{3.19}$$

Hence, $d_{x,w}$ is uniquely determined by x,y,z, and thus so is $w = \varphi_{x,y}(d_{x,w})$. Consequently, (3.15) implies (3.13). Furthermore, (2.3) implies $\Delta(x,y,z) = d(x,\gamma(x,y,z))$ and thus (3.14) follows from (3.19). □

Lemma 3.7. *Suppose that* (T1) *and* (T2d) *hold, and let $x,y,z \in T$.*

(i) *Then*

$$z \in [x,y] \iff \Delta(z,x,y) = 0. \tag{3.20}$$

(ii) *If* $z \notin [x,y]$ *and* $d(w,z) < \Delta(z,x,y)$*, then*

$$\Delta(x,y,w) = \Delta(x,y,z). \tag{3.21}$$

Proof. (i): Immediate by Lemma 3.3 and (3.14).

(ii): Let, using the symmetry of γ in Lemma 3.5, $u := \gamma(x,y,z) = \gamma(z,x,y)$ and $v := \gamma(x,z,w)$. Then $v \in [z,w]$ and thus, using the assumption,

$$d(z,v) \leqslant d(z,w) < \Delta(z,x,y) = d(z,u). \tag{3.22}$$

Since $u, v \in [x,z]$, this implies

$$d(x,u) = d(x,z) - d(u,z) < d(x,z) - d(v,z) = d(x,v). \tag{3.23}$$

Hence, if $d(x,u) < t < d(x,v)$, then, by (2.1),

$$\varphi_{x,w}(t) = \varphi_{x,z}(t) \neq \varphi_{x,y}(t) \tag{3.24}$$

On the other hand, $u \in [x,y]$ too, and thus

$$\varphi_{x,w}(d_{x,u}) = \varphi_{x,z}(d_{x,u}) = u = \varphi_{x,y}(d_{u,}). \tag{3.25}$$

Hence, $\Delta(x,y,w) = d_{x,u} = \Delta(x,y,z)$ by Lemma 3.5. □

4. Proof of Theorem 2.1

(T2a) $\Longrightarrow$ (T2j): If (T2a) holds and $\psi : [0,1] \to T$ is a continuous map, let $x := \psi(0)$ and $y := \psi(1)$. If $z \in [x,y]$ but $z \notin \psi([0,1])$, then ψ is a curve connecting x and y in $T \setminus \{z\}$, which contradicts (T2a).

(T2j) $\Longrightarrow$ (T2b): Define $\psi := \varphi_{y,x} * \varphi_{x,z}$, see (1.1). This is a continuous map $[0, d_{y,x} + d_{x,z}] \to T$ and thus $\psi_1(t) := \psi\big(t(d_{y,x} + d_{x,z})\big)$ is a continuous map $\psi_1 : [0,1] \to T$. We have $\psi_1([0,1]) = [y,x] \cup [x,z]$, $\psi_1(0) = y$ and $\psi_1(1) = z$. Hence (T2j) yields (T2b).

(T2b) $\Longrightarrow$ (T2c): Let $A := [y,z] \cap [x,y]$ and $B := [y,z] \cap [x,z]$. These are closed subsets of $[y,z]$ and $A \cup B = [y,z]$ by assumption; furthermore, A and B are both nonempty since $y \in A$ and $z \in B$. Since $[y,z]$ is homeomorphic to $[0,1]$, it is connected. Consequently, $A \cap B \neq \emptyset$.

(T2c) $\Longrightarrow$ (T2d): By (2.3), it remains to show that $\gamma(x,y,z) \in [y,z]$.

Let $w \in [x,y] \cap [x,z] \cap [y,z]$. By Lemma 3.5, then $w \in [x, \gamma(x,y,z)]$. Since $w, \gamma(x,y,z) \in [x,y]$, Lemma 3.4 shows that $\gamma(x,y,z) \in [w,y]$. Hence, using also $w \in [y,z]$ and Lemma 3.2,

$$\gamma(x,y,z) \in [w,y] = [y,w] \subseteq [y,z]. \tag{4.1}$$

(T2d) $\Longrightarrow$ (T2a): Let $z \in (x,y)$. Then $z = \varphi_{x,y}(d_{x,z})$ with $0 < d_{x,z} < d_{x,y}$. Partition $T \setminus \{z\} = U_1 \cup U_2 \cup U_3$ where

$$U_1 := \{w \in T \setminus \{z\} : \Delta(x,y,w) < d_{x,z}\}, \tag{4.2}$$

$$U_2 := \{w \in T \setminus \{z\} : \Delta(x,y,w) = d_{x,z}\}, \tag{4.3}$$

$$U_3 := \{w \in T \setminus \{z\} : \Delta(x,y,w) > d_{x,z}\}. \tag{4.4}$$

By Lemma 3.6, $\Delta(x,y,w)$ is a continuous function of w, and thus U_1 and U_3 are open subsets of $T \setminus \{z\}$ (and of T). Furthermore, U_2 is open by Lemma 3.7.

Hence, $T \setminus \{z\} = U_1 \cup U_2 \cup U_3$ is a partition into three disjoint open sets. Each connected component of $T \setminus \{z\}$ has to be a subset of one of these, and since $x \in U_1$ and $y \in U_3$, x and y are in different components. (Although not needed, it is easy to see that U_1 and U_3 are connected, while U_2 may be empty, connected or disconnected with any number of components, finite or infinite.)

(T2j) $\Longrightarrow$ (T2i): Trivial.

(T2i) $\Longrightarrow$ (T2f): Suppose that (T2f) fails, and let $\varphi : [0,1] \to T$ be an injective continuous map such that $\varphi(u) \notin [x,y]$ for some $u \in (0,1)$, where $x = \varphi(0)$, $y = \varphi(1)$. Since $[x,y]$ is a compact, and thus closed, subset of T, the set $U := \{t \in (0,1) : \varphi(t) \notin [x,y]\}$ is an open subset of $(0,1)$. Hence, the component of U containing u is an open interval $(a,b) \subseteq (0,1)$.

Let $z := \varphi(a)$ and $w := \varphi(b)$, and note that $z, w \in [x,y]$. Since $a < b$ and φ is injective, we have $z \neq w$. By Lemma 3.2, $[z,w] \subseteq [x,y]$. The map $\psi(t) := \varphi(a + (b-a)t)$ is an injective continuous map $[0,1] \to T$ such that $\psi(0) = z$, $\psi(1) = w$ and

$$\psi([0,1]) \cap [z,w] = \varphi([a,b]) \cap [z,w] \subseteq \varphi([a,b]) \cap [x,y] = \varphi(\{a,b\}) = \{z,w\}. \tag{4.5}$$

Hence, if $v \in (z,w)$, then $v \notin \psi([0,1])$. Thus, (T2i) does not hold.

(T2f) $\Longrightarrow$ (T2d): Let $x, y, z \in T$, and let $w := \gamma(x,y,z)$. By Lemmas 3.2 and 3.5,

$$[y,w] \cap [w,z] = [w,y] \cap [w,z] \subseteq [x,y] \cap [x,z] = [x,w] \tag{4.6}$$

and thus, since $w \in [x,y]$,

$$[y,w] \cap [w,z] \subseteq [x,w] \cap [w,y] = \{w\}. \tag{4.7}$$

Define $\psi : [0, d_{y,w} + d_{w,z}] \to T$ as the concatenation $\psi := \varphi_{y,w} * \varphi_{w,z}$, see (1.1). Then ψ is continuous, and it follows from (4.7) that ψ is injective. Consequently, (T2f) implies, after a change of variables,

$$w = \psi(d_{y,w}) \in [\psi(0), \psi(d_{y,w} + d_{w,z})] = [y,z]. \tag{4.8}$$

(T2f) $\Longleftrightarrow$ (T2e): An immediate consequence of Lemma 3.3.

(T2f) $\Longrightarrow$ (T2h): Suppose that $\psi : [0,1] \to T$ is an injective continuous map. Let $x := \psi(0)$ and $y := \psi(1)$. By (T2f), $\psi : [0,1] \to [x,y]$. Since $\varphi_{x,y}$ is a homeomorphism, $h := \varphi_{x,y}^{-1} \circ \psi : [0,1] \to [0, d_{x,y}]$ is an injective continuous map with $h(0) = 0$ and $h(1) = d_{x,y}$. In particular, the image $h([0,1])$ is connected, and it follows that $h([0,1]) = [0, d_{x,y}]$. Hence h is a continuous bijection, and thus a homeomorphism $[0,1] \to [0, d_{x,y}]$; furthermore, h has to be stricly increasing. Finally, the definition of h yields $\psi = \varphi_{x,y} \circ h$.

(T2h) $\Longrightarrow$ (T2g): Trivial.

(T2g) $\Longrightarrow$ (T2f): Trivial.

This completes the proof of Theorem 2.1. □

5. Subtrees

We have, as a simple consequence of the definition and Theorem 2.1 a simple result for subsets (with the induced metric) of a real tree.

Theorem 5.1. *A pathwise connected subspace of a real tree is a real tree.*

Proof. Let $T_1 \subseteq T$, where T is a real tree, and suppose that T_1 is pathwise connected. If $x, y \in T_1$, then there exists a continuous map $\psi : [0,1] \to T_1$ with $\psi(0) = x$ and $\psi(1) = y$. By (T2j) (for the real tree T), $\psi([0,1]) \supseteq [x,y] := \varphi_{x,y}\big([0, d_{x,y}]\big)$, where $\varphi_{x,y}$ is the mapping in (T1) for the real tree T. Hence, $\varphi_{x,y} : [0, d_{x,y}] \to T_1$, and thus (T1) holds for T_1 too; uniqueness follows because $\varphi_{x,y}$ obviously is unique in T_1 if it is unique in T.

Finally, (T2e) holds in T_1 since it holds in T. (In fact, we could here argue with any of (T2a)–(T2j).) □

We can strengthen Theorem 5.1.

Theorem 5.2. *A connected subspace of a real tree is a real tree.*

Proof. Let $T_1 \subseteq T$, where T is a real tree, and suppose that T_1 is connected. Let $x, y \in T_1$, and consider $[x,y]$ (in the real tree T). Let $z \in (x,y)$, and suppose that $z \notin T_1$. By (T2a) and Lemma 3.1, the components of $T \setminus z$ are disjoint open sets, with x and y in different sets. Let U be the component containing x, and V the union of all other components; then $T \setminus z = U \cup V$ where U and V are open disjoint subsets with $x \in U$ and $y \in V$. Consequently, $T_1 = (T_1 \cap U) \cup (T_1 \cap V)$, where $T_1 \cap U$ and $T_1 \cap V$ are two nonempty disjoint open subsets of T_1. This contradicts the assumption that T_1 is connected, and this contradiction shows that $(x,y) \subseteq T_1$.

Consequently, if $x, y \in T_1$, then $[x,y] \subseteq T_1$, which shows that T_1 is pathwise connected. Hence, T_1 is a real tree by Theorem 5.1. □

6. The four-point inequality

A different type of characterization of real trees is given by the following theorem, see e.g. [3] and the references there; the condition (6.1) is called the *four-point inequality* or *four-point condition.* (This characterization is less intuitive, but technically very useful.)

Theorem 6.1. *A metric space T is a real tree if and only if T is connected and for any four points $x, y, z, w \in X$*

$$d(x,y) + d(z,w) \leqslant \big(d(x,z) + d(y,w)\big) \vee \big(d(x,w) + d(y,z)\big). \tag{6.1}$$

Remark 6.2. It is easily verified that (6.1) is trivial if two or more of x, y, z, w coincide; hence it does not matter whether we require x, y, z, w to be distinct or not. □

Remark 6.3. By considering all permutations of x, y, z, w, it follows that (6.1) is equivalent to the condition that (for any x, y, z, w), among the three sums

$$d(x,y) + d(z,w), \qquad d(x,z) + d(y,w), \qquad d(x,w) + d(y,z), \tag{6.2}$$

two are equal and the third is equal or less than the other two. □

Theorem 6.1 is a simple corollary of the following more general result together with Theorem 5.1. For proofs see e.g. Dress [3, Theorem 8], Gromov [5, §6.1], or the references mentioned in [3]; see also [4].

Theorem 6.4. *Let X be a metric space. Then X can be isometrically embedded into a real tree if and only if the four-point inequality* (6.1) *holds for any four points $x, y, z, w \in X$.* □

Proof of Theorem 6.1 from Theorem 6.4. Suppose that T is connected and that (6.1) holds. Then, by Theorem 6.4, $T \subseteq \widehat{T}$ for some real tree $\widehat{T}$. Since T is connected, T is a real tree by Theorem 5.2. □

Among the consequences we mention the following.

Theorem 6.5. *If T is a real tree, then so is its completion $\tilde{T}$.*

Proof. By continuity, if the four-point inequality (6.1) holds in the dense subset T of $\tilde{T}$, then it holds in $\tilde{T}$. Furthermore, since T is connected, so is $\tilde{T}$. hence, $\tilde{T}$ is a real tree by Theorem 6.1. □

Remark 6.6. By Theorem 6.5, we may in many situations assume without loss of generality that real trees are complete, since we can replace an arbitrary real tree by its completion. □

The four-point inequality (6.1) can be rewritten in several ways. Define, for three points x, y, z in a general metric space (X, d), the *Gromov product*

$$(x, y)_z := \tfrac{1}{2}\bigl(d(z, x) + d(z, y) - d(x, y)\bigr). \tag{6.3}$$

Note that $(x, y)_z \geqslant 0$ by the triangle inequality, and that (6.3) meaures how far the triangle inequality is from being an equality. Note also that in a real tree, Lemma 3.6 shows that $(x, y)_z = \Delta(z, x, y)$, which equals the distance from z to $[x, y]$.

Lemma 6.7. *The four-point inequality* (6.1) *is equivalent to*

$$(x, y)_w \geqslant (x, z)_w \wedge (y, z)_w. \tag{6.4}$$

Proof. By the definition (6.3), the inequality (6.4) holds if and only if at least one of the following holds:

$$d_{x,w} + d_{y,w} - d_{x,y} \geqslant d_{x,w} + d_{z,w} - d_{x,z}, \tag{6.5}$$

$$d_{x,w} + d_{y,w} - d_{x,y} \geqslant d_{y,w} + d_{z,w} - d_{y,z}. \tag{6.6}$$

These are equivalent to, respectively,

$$d_{y,w} + d_{x,z} \geqslant d_{z,w} + d_{x,y}, \tag{6.7}$$

$$d_{x,w} + d_{y,z} \geqslant d_{z,w} + d_{x,y}, \tag{6.8}$$

and thus at least one of them holds if and only iff (6.1) holds. □

Appendix A. Gromov hyperbolic spaces

This appendix is a long remark on an approximate version of the four-point inequality, which is important in other contexts; the appendix can be skipped by those interested in trees only. $X = (X, d)$ is a metric space.

Gromov [5] defined *δ-hyperbolic metric spaces* as metric spaces where the four-point inequality holds up to an error δ. More precisely, he used the version (6.4), and the definition is as follows, for a given $\delta \geqslant 0$:

Definition A.1 (Gromov)**.** A metric space $X = (X, d)$ is δ-hyperbolic if, for all $x, y, z, w \in X$,

$$(x, y)_w \geqslant (x, z)_w \wedge (y, z)_w - \delta. \tag{A.1}$$

By Lemma 6.7, Theorem 6.4 can be formulated as follows.

Theorem A.2. *A metric space is* 0*-hyperbolic if and only if it can be isometrically embedded in a real tree.*

However, the main interest of Definition A.1 is for $\delta > 0$. The value of δ is often not important, and one says that a metric space X is *Gromov hyperbolic* if it is δ-hyperbolic for some $\delta < \infty$.

Remark A.3. Intuitively, at large distances, the δ in (A.1) is insignificant, and thus the large scale geometry of a Gromov hyperbolic space is "tree-like". □

There are several alternatives to Definition A.1, with conditions that are equivalent in the sense that if one holds, then so do the others, but with δ replaced by $C\delta$ for some (small) constant C. (The conditions are in general not equivalent with the same δ.) Some of these alternative conditions are given (or implicit) in the following lemmas. See further e.g. [5], [2] and [6].

Lemma A.4. *The inequality* (A.1) *holds if and only if*

$$d(x,y) + d(z,w) \leqslant \big(d(x,z) + d(y,w)\big) \vee \big(d(x,w) + d(y,z)\big) + 2\delta. \tag{A.2}$$

Proof. As the proof of Lemma 6.7, adding 2δ to the left-hand sides of (6.5)–(6.8). □

Lemma A.5 ([5])**.** *Let X be a metric space and let $o \in X$ be fixed. If* (A.1) *holds for $w = o$ and all $x, y, z \in X$, then it holds for all $x, y, z, w \in X$ with δ replaced by 2δ.*

Proof. We first show that the assumption implies that, for any x, y, z, w,

$$(x,y)_o + (z,w)_o \geqslant \big((x,z)_o + (y,w)_o\big) \wedge \big((x,w)_o + (y,z)_o\big) - 2\delta. \tag{A.3}$$

To see this, we first note that both sides are symmetric in x and y, and also in z and w; hence, by interchanging (x, y) and/or (z, w) if necessary, we may assume that $(x, z)_o$ is the largest of the four numbers $(x, z)_o, (x, w)_o, (y, z)_o, (y, w)_o$. In this case, the assumption implies

$$(x,y)_o \geqslant (x,z)_o \wedge (y,z)_o - \delta = (y,z)_o - \delta, \tag{A.4}$$

$$(z,w)_o \geqslant (x,z)_o \wedge (x,w)_o - \delta = (x,w)_o - \delta, \tag{A.5}$$

and thus

$$(x,y)_o + (z,w)_o \geqslant (y,z)_o + (x,w)_o - 2\delta, \tag{A.6}$$

verifying (A.3).

Next, by exanding all Gromov products in (A.3) according to the definition (6.3), we see that (A.3) does not depend on the choice of o; hence (A.3) holds for every $o, x, y, z, w \in X$.

In particular, we can choose $o = w$ in (A.3). Since $(w, v)_w = 0$ for every v by (6.3), we obtain

$$(x,y)_w \geqslant (x,z)_w \wedge (y,z)_w - 2\delta, \tag{A.7}$$

as asserted. □

A *geodesic* in a metric space X is an isometric curve, i.e., an isometric mapping of an interval $I \subseteq \mathbb{R}$ into X. If a geodesic φ is defined on an interval $I = [a, b]$ that is closed and finite, we say that the geodesic has the endpoints $\varphi(a)$ and $\varphi(b)$, and that it *joins* $\varphi(a)$ and $\varphi(b)$. (Then, the geodesic necessarily has length $d_{x,y}$, and we may choose $I = [0, d_{x,y}]$.)

Note that condition (T1) says that for every $x, y \in X$, there is a unique geodesic from x to y (provided we normalize $I = [0, d_{x,y}]$).

More generally, a metric space $X = (X, d)$ is *geodesic*, if every pair $x, y \in X$ is joined by a geodesic. In other words, we assume the existence part of (T1), but uniqueness is not assumed. In particular, every real tree is geodesic.

Gromov hyperbolicity is often studied for geodesic metric spaces. This can be done without real loss of generality by the following result by Bonk and Schramm [1], to which we refer for a proof.

Theorem A.6 (Bonk and Schramm [1, Theorem 4.1])**.** *Let $\delta \geqslant 0$. A metric space is δ-hyperbolic if and only if it can be isometrically embedded into a δ-hyperbolic complete geodesic metric space.* □

In particular, for $\delta = 0$, we recover Theorem 6.4 (together with Theorem 6.5).

For geodetic metric spaces, there are further conditions equivalent to Gromov hyperbolicity. We note first an extension of Lemma 3.3 to general geodesic spaces.

Lemma A.7. *Let (X, d) be a geodesic metric space, and let $x, y, z \in X$. Then $(x, y)_z = 0$ if and only if there exists a geodesic from x to y that contains z.*

Proof. As for Lemma 3.3. □

In a geodesic metric space, it is enough to assume the hyperbolicity condition (A.1) when the left-hand side is 0, i.e., by Lemma A.7, when w is on a geodesic joining x and y.

Lemma A.8. *Let (X, d) be a geodesic metric space. Suppose that, for every $x, y, z \in X$ and every w that lies on a geodesic from x to y, the inequality* (A.1) *holds, or equivalently*

$$(x, z)_w \wedge (y, z)_w \leqslant \delta. \tag{A.8}$$

Then X is 3δ-hyperbolic.

Proof. See [2, p. 286 and Exercise 8.4.5] or [6, Proof of Theorem 2.34]. □

Remark A.9. In particular, using Theorem A.2, a geodesic metric space is a real tree if and only if (A.8) holds with $\delta = 0$. In fact, it is easily seen that this condition implies that geodesics are unique, so (T1) holds. (Given two geodesics from x to y, let z and w be points on them with $d(z, x) = d(w, x)$, and conclude $z = w$ from (A.8).) Then, by Lemma A.7 again, (A.8) with $\delta = 0$ when $w \in [x, y]$ is equivalent to $[x, y] \subseteq [x, z] \cup [y, z]$, which is (T2b). □

The condition in Lemma A.8 can be regarded as a condition on the geometry of the triangle xyz. One can pursue this point of view further.

Definition A.10. Let (X, d) be a geodesic metric space. A *triangle xyz* in X is a set of three *vertices* $x, y, z \in X$ together with three *sides* xy, xz and yz that are geodesics between the pairs of vertices.

Note that in general, the sides of a triangle are not uniquely determined by the vertices.

Definition A.11. Let (X, d) be a geodesic metric space. A triangle is δ-*slim* if each side is contained in the closed δ-neighbourhood of the union of the two other sides, i.e., if every $w \in xy$ has distance $\leqslant \delta$ to $xz \cup yz$, and similarly for the two other sides.

Gromov [5] used a definition that is similar, but somewhat more technical. Note that in a triangle with vertices x, y, z, we have, by (6.3), $(y, z)_x + (x, z)_y = d_{x,y}$, and thus the side xy can be divided into two parts of lengths $(y, z)_x$ and $(x, z)_y$, containing x and y respectively; we call these, the x-*part* and y-*part* of xy. (In a tree, these are the parts of $[x, y]$ before and after $\gamma(x, y, z)$. The definition in [5] is stated in terms of isometric mappings of the three sides into a real tree with three endpoints having the same distances between each other as x, y, z.)

Definition A.12. Let (X, d) be a geodesic metric space. A triangle xyz is δ-*thin* if for each w in the x-part of the side xy, the point $v \in xz$ with $d(v, x) = d(w, x)$ satisfies $d(v, w) \leqslant \delta$, and similarly for the y-part and for the two other sides.

It is immediate that a δ-thin triangle is δ-slim.

Lemma A.13. *Let X be a geodesic metric space.*

(i) *If X is δ-hyperbolic, then every triangle in X is 2δ-thin.*
(ii) *If every triangle in X is δ-thin, then X is 2δ-hyperbolic.*

Proof. See [5, Proposition 6.3.C]. □

Lemma A.14. *Let X be a geodesic metric space.*

(i) *If X is δ-hyperbolic, then every triangle in X is 2δ-slim.*
(ii) *If every triangle in X is δ-slim, then X is 3δ-hyperbolic.*

Proof. (i) follows from Lemma A.13(i); see also [6, Theorems 2.35] (with 3δ).

(ii) is [6, Theorems 2.34 (with $h = 0$)]. □

Remark A.15. If every triangle in X is δ-slim (or δ-thin), then any two geodesics with the same endpoints are within distance 2δ of each other. (I.e., their Hausdorff distance is $\leqslant 2\delta$.) This follows by considering the triangle obtained by subdividing one of the geodesics at an arbitrary interior point. (Or, more bravely, by considering the geodesics as two sides of a degenerate triangle with two vertices coinciding.) □

Remark A.16. In particular, using Theorem A.2, a geodesic metric space is a real tree if and only if every triangle is 0-slim (or 0-thin).

In fact, if this holds, then by Remark A.15, geodesics are unique, and thus (T1) holds. Then, a triangle xyz is 0-slim if and only if (T2b) holds for all permutations of x, y, z. Conversely, each triangle in a real tree is 0-thin as a consequence of Lemmas 3.5 and 3.6. □

References

[1] M. Bonk & O. Schramm: Embeddings of Gromov hyperbolic spaces. *Geom. Funct. Anal.* **10** (2000), no. 2, 266–306.

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

[3] Andreas W. M. Dress: Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces. *Advances in Mathematics* **53**:3 (1984), 321–402.

[4] Andreas Dress, Vincent Moulton & Werner Terhalle: T-theory: an overview. *European J. Combin.* **17** (1996), no. 2-3, 161–175.

[5] M. Gromov. Hyperbolic groups. In *Essays in Group Theory*, 75–263, Math. Sci. Res. Inst. Publ., 8, Springer, New York, 1987.

[6] Jussi Väisälä: Gromov hyperbolic spaces. *Expo. Math.* **23** (2005), no. 3, 187–231.

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/`