Chain development of metric compacts

Yu.V. Malykhin†, E.V. Schepin‡

April 5, 2016

Notions and basic facts. Let $(X, d)$ be a metric space. We call a sequence of points $x = x_0, x_1, x_2, \ldots, x_n = y$ an $\varepsilon$-chain if $d(x_i, x_{i+1}) \leq \varepsilon$ for all $i$. Define chain distance $c(x, y)$ as the infimum of $\varepsilon$ such that there exists an $\varepsilon$-chain from $x$ to $y$.

Chain distance satisfies strong triangle inequality: $c(x, z) \leq \max(c(x, y), c(y, z))$; hence it is ultrametric if it does not degenerate. Obviously, $c = d$ if $d$ is already ultrametric.

Definition. A function $f : X \to \mathbb{R}$ is called chain development if $f$ preserves chain distance:

$$c(x, y) = \tilde{c}(f(x), f(y)) \quad \text{for } x, y \in X,$$

where $c$ is the chain distance on $(X, d)$ and $\tilde{c}$ is the chain distance on the set $f(X)$ with usual distance $\tilde{d}(s, t) = |s - t|$.

Chain development was firstly introduced by E.V. Schepin for finite sets as a tool for fast hierarchical cluster analysis. Note that chain development always exists for finite spaces and can be effectively constructed using minimum weight spanning tree of the corresponding graph; see [1] and [2] Section 4 for more details. An equivalent construction appeared in the paper [3] by A.F. Timan and I.A. Vestfold: they proved that points of any finite ultrametric space can be enumerated in a sequence $x_1, \ldots, x_n$ such that $c(x_i, x_j) = \max(c(x_i, x_j), c(x_j, x_j))$ for $i < j < k$.

*This research was supported by Russian Science Foundation (project 14-50-00005)
†Steklov Mathematical Institute
‡Steklov Mathematical Institute
The goal of this paper is to discuss some properties of chain development for infinite spaces. So, there are compacts with no chain developments, e.g. the square $C 	imes C$ of a Cantor set. Necessary and sufficient condition of existence of chain developments is given below in Theorem 2.

By diameter of a chain development $f : X \to \mathbb{R}$ we mean $\text{diam } f(X) = \sup f(X) - \inf f(X)$. It is proven in [1] that for finite spaces $X$ the diameter of chain developments is determined uniquely. It turns out that this is not true in general case.

**Theorem 1.** Let $(X, d)$ be a compact metric space. Then the diameter of chain developments (if there are any) is determined uniquely if and only if $X$ is countable.

Throughout this paper by $(Z, d)$ we denote a zero-dimensional compact metric space. We focus on such spaces because study of chain developments for arbitrary compacts essentially reduces to the zero-dimensional case.\footnote{One can identify points of $(X, c)$ with $c(x, y) = 0$ to obtain zero-dimensional ultra-metric compact $(Z_X, c)$; a chain development of $(X, d)$ exists if and only if there is a chain development of $(Z_X, c)$.} We have the following property:

(i) $(Z, c)$ is an ultrametric space, i.e. chain distance does not degenerate. Indeed, take $x, y \in Z$. The set $\{x\}$ is a connected component, hence $x \in U \not\ni y$ for some closed open set $U$, so

$$c(x, y) \geq \min_{u \in U} d(u, v) > 0.$$ 

The transition from metric $d$ to ultrametric $c$ (which can be seen as a functor) preserves topology:

(ii) The identity map $\text{id}: Z \to Z$ is a homeomorphism between $(Z, d)$ and $(Z, c)$.

Indeed, $\text{id}$ is 1-Lipshitz ($c(x, y) \leq d(x, y)$), hence it is a continuous bijection from compact to Hausdorff space, hence a homeomorphism.

(iii) Any chain development $f : Z \to \mathbb{R}$ is continuous (with usual topology on $\mathbb{R}$). Hence, $f(Z)$ is compact and $f$ is a homeomorphism between $Z$ and $f(Z)$. 

Theorem 1. Let $(X, d)$ be a compact metric space. Then the diameter of chain developments (if there are any) is determined uniquely if and only if $X$ is countable.
Let $x_n \to x^*$ in $Z$; prove that $t_n := f(x_n) \to t^* =: f(x^*)$. Suppose that $t_n > t^* + \varepsilon$ for some $\varepsilon > 0$. If there are no points of $f(Z)$ in $(t^*, t^* + \varepsilon)$, then $\tilde{c}(t_n, t^*) \geq \varepsilon$ (where $\tilde{c}$ is the chain distance on $f(Z)$). And if there is some $t = f(x) \in (t^*, t^* + \varepsilon)$, then $\tilde{c}(t_n, t^*) = \tilde{c}(t, t^*) = c(x, x^*) > 0$. In both cases $\tilde{c}(t_n, t^*) \not\to 0$, which contradicts that $\tilde{c}(t_n, t^*) = c(x_n, x^*) \leq d(x_n, x^*) \to 0$. So, $f$ is continuous.

The chain distance on a compact $K \subset \mathbb{R}$ is determined by the lengths of the intervals of the open set $U_K := [\min K, \max K] \setminus K$.

(iv) Chain distance between points $s, t$ of $K$ is equal to the maximal length of the intervals of $U_K$, lying between $s$ and $t$.

**Existence of chain development.** There is a well-known correspondence between ultrametric spaces and labeled trees; here we describe it for our purposes. Let $(X, d)$ be a compact metric space; we will construct a labeled tree $T(X, d)$ with a vertex set $V$ and a labeling function $r : V \to \mathbb{R}$. We take an arbitrary point $v_0$ as a root of our tree and assign to it the $c$-diameter of $X$, i.e. $r(v_0) = \max_{x, y \in X} c(x, y)$. The relation $c(x, y) < r(v_0)$ is an equivalence relation; hence, $X$ breaks into finite number of “clusters” $Q_1, \ldots, Q_n$ of points with pairwise chain distance less than $r(v_0)$. Next, we connect the root with $n$ children, say $v_1, \ldots, v_n$, with $v_j$ corresponding to $Q_j$. We repeat the construction for each of $Q_j$: we assign $r(v_j) = \max_{x, y \in Q_j} c(x, y)$, and connect $v_j$ with children corresponding to the clusters $Q_{j,k} \subset Q_j$ with $c(x, y) < r(v_j)$, $x, y \in Q_{j,k}$. And so on. The process stops if $c$-diameter of a cluster becomes zero.

So, with each vertex $v$ of $T(X, d)$ we associate:

- $n(v)$ — the number of children of $v$;
- $C(v)$ — the set of children of $v$;
- $Q(v)$ — the cluster of points, corresponding to $v$; e.g. $Q(v_0) = X$;
- $r(v)$ — the $c$-diameter of $Q(v)$.

**Definition.** The width of the space $(X, d)$ is defined as

$$w(X, d) := \sum_v r(v)(n(v) - 1),$$

where the sum is over all vertices of the tree $T(X, d)$. 3
Theorem 2. Let \((X, d)\) be a compact metric space. Then there exists a chain development \(f : X \to \mathbb{R}\) if and only if \(w(X, d) < \infty\). Moreover, \(w(X, d)\) is the minimal possible diameter of a chain development of \(X\).

The construction of the tree uses only the chain distance, so \(T(Z, d) = T(Z, c)\) and \(w(Z, d) = w(Z, c)\). On the other hand, the ultrametric structure is fully captured by the tree \(T(Z, d)\). Each point \(x \in Z\) lies in some sequence of clusters; hence, it corresponds to a path in the tree.

Lemma 1. Let \(x, y \in Z\). If \(x \neq y\), then they lie in different path of the tree, and \(c(x, y)\) is equal to \(r(v)\), where \(v\) is the lowest common ancestor of \(x, y\), i.e. the farthest from root vertex lying on both paths.

Proof. Assume \(x, y\) lie in the same path \(\{v_0, v_1, \ldots\}\) of the tree. The compactness of \(Z\) implies that diameters of the clusters \(Q(v_j)\) tend to zero. Then \(c(x, y)\) is less than any diameter of the corresponding clusters, hence, \(c(x, y) = 0\), and \(x = y\).

Let \(v\) be the lowest common ancestor of \(x\) and \(y\). Then \(c(x, y) \leq r(v)\) by the definition of \(r(v)\) and \(c(x, y) = r(v)\) because \(x, y\) lie in different sub-clusters of \(Q(v)\).

Let us prove Theorem 2.

Proof. Consider the case of zero-dimensional ultrametric compact space \((Z, c)\). The construction of the set \(f(Z)\) is equivalent to the construction of the tree \(T(Z, c)\). Pick an interval \([a, b]\) of length \(w(Z, c)\); we know that

\[ w(Z, c) = \sum_{v \in C(v_0)} w(Q(v), c) + (n(v_0) - 1)r(v_0). \]

One can remove \(n(v_0) - 1\) disjoint open intervals of length \(r(v_0)\) from \([a, b]\) so that the remaining \(n(v_0)\) closed intervals will have lengths \(\{w(Q(v), c)\}_{v \in C(v_0)}\). Those closed intervals correspond to each of \(Q(v)\) and we proceed with them as with \([a, b]\).

After removal all of the open intervals we arrive at some closed set \(K \subset [a, b]\). Every point \(x \in Z\) corresponds to a path in \(T(Z, c)\) and to a nested sequence of closed intervals with non-empty intersection \(t \in K\); we put \(f(x) = t\) (intersection is always a point because \(\mu(K) = 0\)). The proof that \(f\) is chain development is straight-forward using Lemma 1 and property (iv). Note that \(\text{diam } f(Z) = w(Z, c)\).
Now, let \( f : Z \to \mathbb{R} \) be a chain development. Define \( U_{f(Z)} := [\min f(Z), \max f(Z)] \setminus f(Z) \). We prove that

\[
w(Z, c) = \mu(U_{f(Z)}) = \text{diam} f(Z) - \mu(f(Z)).
\] (1)

Remind that \( r(v_0) \) is the \( c \)-diameter of \( Z \) and the \( \tilde{c} \)-diameter of \( f(Z) \). It is obvious from (iv) that there are exactly \( n(v_0) - 1 \) intervals of \( U \) of length \( r(v_0) \). Repeating this argument with sets \( f(Q(v)), v \in C(v_0) \), we will count all of the intervals of \( U \) and found that each vertex \( v \) corresponds to \( n(v) - 1 \) intervals of \( U \) of length \( r(v) \). That implies (1). Hence, \( w(Z, c) < \infty \) and \( \text{diam} f(Z) \geq w(Z, c) \).

The general case follows easily. \( \square \)

Me will make use of the following standard construction.

**Lemma 2.** Let \( K \) be an uncountable compact in \([a, b]\). Then for any \( c > 0 \) there is a continuous increasing function \( \theta : [a, b] \to \mathbb{R} \) such that \( \mu(\theta(K)) = \mu(K) + c \) and \( \mu(\theta(I)) = \mu(I) \) for any interval \( I \subset [a, b] \setminus K \).

**Proof.** Write \( K \) as \( N \cup P \), where \( N \) is countable and \( P \) is perfect. Let \( \kappa : [a, b] \to [0, 1] \) be an analog of the Cantor's ladder for the set \( P \); we need that \( \kappa \) is continuous and non-decreasing, \( \kappa([a, b]) = [0, 1] \) and \( \kappa|_I \equiv \text{const} \) for any interval \( I \subset [a, b] \setminus P \). It remains to take \( \theta(t) = t + c\kappa(t) \). \( \square \)

Now we are ready to prove Theorem 1.

**Proof.** We consider only the zero-dimensional case. If \( Z \) is countable, then \( \mu(f(Z)) = 0 \) and from (1) we get \( \text{diam} f(Z) = w(Z, c) \). Suppose \( Z \) is uncountable. Take any chain development \( f : Z \to \mathbb{R} \) and apply Lemma 2 to \( K = f(Z) \) with some \( c > 0 \). Then \( \theta \circ f \) gives us a chain development with another diameter. \( \square \)

It appears that the diameter of a chain development of an uncountable compact may be any number greater or equal than \( w(X, d) \).

**Example.** Consider the set \( C \times C \), where \( C \subset [0, 1] \) is the usual Cantor set. Let \( d((x_1, y_1), (x_2, y_2)) = \max(|x_1 - x_2|, |y_1 - y_2|) \) for \((x_i, y_i) \in C \times C \). Then there is no chain development for the space \((C \times C, d)\).
Proof. Let us compute $w(C \times C, d)$. In the tree $T(C \times C, d)$ each node has four children; for example, the children of the root correspond to the clusters

$$(C \cap \left[\frac{2i}{3}, \frac{2i+1}{3}\right]) \times (C \cap \left[\frac{2j}{3}, \frac{2j+1}{3}\right]), \quad i, j = 0, 1. \quad (2)$$

We have $r(v_0) = 1/3$ for the root $v_0$ and $r(u) = \frac{1}{3}r(v)$ for each children $u$ of $v$, by self-similarity of $C$. Hence, $w(C \times C, d) = \sum_{k=0}^{\infty} 4^k 3^{-k} = \infty$ and the claim follows from Theorem 2.

\[\square\]

**Measure of disconnectivity.**

**Definition.** Let $(X, d)$ be a metric space. Define measure of disconnectivity of $(X, d)$ as

$$\text{dis}(X, d) = \inf_{x_i \sim y_i} \sum_i d(x_i, y_i),$$

where the infimum is taken over sequences (finite or infinite) or pairs $(x_i, y_i) \in X \times X$, such that the space $(X, d)$ with identified points $x_i \sim y_i$ is a connected topological space.

This notion is closely related to the minimum spanning trees of graphs. Indeed, if $X$ is finite, then $\text{dis}(X, d)$ is equal to the weight of a minimum spanning tree for $X$ (we regard points of $X$ as vertices and take weights of edges equal to the corresponding distances).

**Theorem 3.** Let $(X, d)$ be a compact metric space. Then $\text{dis}(X, d) = w(X, d)$.

We need one more notation for vertices of a tree $T(X, d)$: by level($v$) we denote the length of the path from the root to $v$.

**Proof.** Note that for finite sets $X$ the theorem follows from [1]. We prove there that $w(X, d)$ is the diameter of any chain development of $X$, and it is clear from the proof that it is equal to the weight of a minimum spanning tree of $X$.

Let us prove that $\text{dis}(X, d) \geq w(X, d)$. Pick some $N \in \mathbb{N}$ and consider all clusters $Q(v)$ with either level($v$) = $N$ or level($v$) < $N$ and $r(v) = 0$. We denote by $(X_N, c_N)$ the ultrametric space, which comes from $(X, c)$ when we identify points in each cluster. To make $X$ connected, we should connect all of the mentioned clusters, so $\text{dis}(X, d) \geq \text{dis}(X_N, c_N)$. For finite sets, $\text{dis} = w$, so $\text{dis}(X_N, c_N) = w(X_N, c_N)$. Obviously, $T(X_N, c_N)$ is obtained.
from $T(X, d)$ by deleting vertices of level $> N$, and assigning $r(v) = 0$ for the new leaves. So

$$w(X_N, c) = \sum_{\text{level}(v)<N} r(v)(n(v) - 1) \to w(X, c) \text{ as } N \to \infty,$$

hence $\text{dis}(X, d) \geq w(X, d)$.

Let us prove that $\text{dis}(X, d) \leq w(X, d)$. For each vertex $v$ we connect the clusters $\{Q(u)\}_{u \in C(v)}$ to each other by picking appropriate pairs $(x_i, y_i) \in C(u') \times C(u'')$. It is easy to show that one can make the set of that clusters connected using pairs with $\sum d(x_i, y_i) = r(v)(n(v) - 1)$. In total, the sum is $w(X, d)$. Let us prove that the image $\tilde{X}$ of $X$ after projection $\pi: X \to \tilde{X}$ is connected. If $\tilde{U} \subset \tilde{X}$ is non-empty, open and closed, then $U = \pi^{-1}\tilde{U} \subset X$ is also non-empty, open and closed; besides that, if $x_i \sim y_i$ and $x_i \in U$, then $y_i \in \tilde{U}$. It remains to prove that $U = X$.

If $x \in U$, then $x \in Q(v) \subset U$ for some $v$. Indeed, $\delta := \min_{u \in U, v \in X \setminus U} d(u, v) > 0$, so if we take $Q(v) \ni x$ with sufficiently small diameter, $r(v) < \delta$, then $Q(v) \subset U$. So, $U$ is a union of clusters; since $U$ is compact, it is a finite union. Now one can prove via induction on $N$ that for all $v$ of level $\geq N$ either $Q(v) \subset U$ or $Q(v) \cap U = \emptyset$. Indeed, $U$ is a union of finite number of clusters, so this is true for large $N$. Let us make an induction step from $N$ to $N - 1$. Suppose there is $Q(v)$, $\text{level}(v) = N - 1$, with $Q(v) \cap U \neq \emptyset$. We have $Q(v) = \bigsqcup_{u \in C(v)} Q(u)$ so $Q(u') \cap U \neq \emptyset$ for some $u' \in C(v)$. As $\text{level}(u') = N$, $Q(u') \subset U$. There is some $u'' \in C(v)$ and a pair $x_i \sim y_i$, $(x_i, y_i) \in Q(u') \times Q(u'')$. As $x_i \in U$, we have $y_i \in \tilde{U}$ and $Q(u'') \subset \tilde{U}$. As all the clusters $\{Q(u)\}_{u \in C(v)}$ are connected, we will prove that $Q(u) \subset U$ for all $u \in C(v)$, i.e. $Q(v) \subset U$. The claim follows.

Finally, $Q(v_0) \subset U$ so $U = X$ and $\tilde{X}$ is connected.

**Corollary.** For any metric compact $(X, d)$ three quantities are equal:

- the minimal diameter of a chain development of $X$;
- the width $w(X, d)$;
- the measure of disconnectivity $\text{dis}(X, d)$.

Note that first two quantities definitely have ultrametric nature, but this is not obvious for the third quantity.
References

[1] Yu.V. Malykhin, E.V. Shchepin, “Chain Development”, Proceedings of the Steklov Institute of Mathematics, 290:1 (2015), 300–305.

[2] V.A. Lemin, “Finite ultrametric spaces and computer science”, Categorical Perspectives. – Birkhäuser Boston, 2001. p.219–241.

[3] A.F. Timan, I.A. Vestfrid, “Any separable ultrametric space can be isometrically imbedded in $\ell_2$”, Functional Analysis and Its Applications, 17:1 (1983), 70–71.