An Interleaving Distance for Ordered Merge Trees

Citation for published version (APA):

Document status and date:
Published: 01/03/2024

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
An Interleaving Distance for Ordered Merge Trees

Thijs Beurskens¹, Tim Ophelders¹,², Bettina Speckmann¹, and Kevin Verbeek¹

¹ Department of Mathematics and Computer Science, TU Eindhoven, The Netherlands
{t.p.j.beurskens|k.a.b.verbeek|b.speckmann}@tue.nl
² Department of Information and Computing Science, Utrecht University, The Netherlands
t.a.e.ophelders@uu.nl

Abstract

Merge trees are a common topological descriptor for data with a hierarchical component. The interleaving distance, in turn, is a common distance measure for comparing merge trees. In this abstract, we introduce a form of ordered merge trees and extend the interleaving distance to a measure that preserves orders. Exploiting the additional structure of ordered merge trees, we then describe an $O(n^2)$ time algorithm that computes a 2-approximation of this new distance with an additive term $G$ that captures the maximum height differences of leaves of the input merge trees.

1 Introduction

Merge trees are a common topological descriptor for data with a hierarchical component, such as terrains and scalar fields. However, standard merge trees focus solely on the hierarchy and do not represent other salient geometric features of the data. Specifically, our work is motivated by the study of braided rivers (see Figure 1). A braided river is a multi-channel river system, known to evolve rapidly [13, 17]. There exist methods to generate a river network from a snapshot of the terrain [7, 14, 16]. We model a river network as a hierarchy of braids. We use a merge tree to represent this hierarchy: each leaf represents a single channel in the network, and each internal vertex represents two braids merging (see Figure 1).

It is our goal to analyse the evolution of the channel network over time. The standard way to compare two merge trees is the interleaving distance [19]. However, the interleaving distance has two main drawbacks. Firstly, the standard interleaving distance is unable to capture any intrinsic order, e.g. from bank to bank in braided rivers, that might be present in

Figure 1 (Left) the Waimakariri River in New-Zealand is a braided river. Photo was taken by Greg O’Beirne [21]. (Right) representing a river network by a merge tree.
the data. Secondly, there is no known efficient algorithm to compute even an approximation of the interleaving distance. To tackle both issues, we introduce the monotone interleaving distance: an order-preserving distance measure on ordered merge trees.

Contributions. We show that the monotone interleaving distance can be defined in terms of a single map between the input merge trees, two maps, or a labelling. Moreover, we give an algorithm that computes an approximation of this distance in $O(n^2)$ time. Finding an efficient algorithm to compute the monotone interleaving distance exactly, or to prove NP-hardness of computing the distance, remains an open problem. We first review the relevant background in Section 2. In Section 3 we introduce a form of ordered merge trees and define monotone interleavings, monotone $\delta$-good maps and monotone labellings. We give constructions to prove that all of these lead to the same distance. Finally, in Section 4 we describe an efficient algorithm to approximate the monotone interleaving distance. All omitted proofs can be found in the full version on arXiv.

Related work. The interleaving distance was first introduced as a measure for persistence modules [8]. It has since been well-studied from a categorical point of view [3, 4, 5, 6, 9, 10, 11, 15, 22], and has been transferred to numerous topological descriptors [2, 18, 20]. Morozov et al. [19] defined the interleaving distance for merge trees. Agarwal et al. [1] established a relation between the interleaving distance and the Gromov-Hausdorff distance. The interleaving distance on merge trees was redefined by first Touli and Wang [23], and later Gasparovich et al. [12]. Touli and Wang also gave an FPT-algorithm to compute the interleaving distance. Recently, the result by Gasparovich et al. has been used to design algorithms for computing geometry aware labellings [24, 25].

2 Preliminaries

A merge tree is a pair $(T, f)$, where $T$ is a rooted tree and $f: T \to \mathbb{R} \cup \{\infty\}$ is a continuous height function that is increasing towards the root, with $f(v) = \infty$ if and only if $v$ is the root. Here, $f$ is defined not only on the vertices of $T$, but also on points of $T$ interior to the edges. Specifically, $f$ is linearly interpolated along the edges. For a point $x \in T$, we denote by $T_x$ the subtree of $T$ rooted at $x$. Furthermore, for a given value $\delta \geq 0$, we denote by $x^\delta$ the unique ancestor of $x$ with $f(x^\delta) = f(x) + \delta \in T$.

Now consider two merge trees $(T, f)$ and $(T', f')$ and fix a value $\delta \geq 0$. Intuitively, a $\delta$-interleaving describes a mapping $\alpha$ from $T$ to $T'$ that sends points exactly $\delta$ upwards, and a similar map $\beta$ from $T'$ to $T$, such that both compositions of $\alpha$ and $\beta$ send any point to its unique ancestor $2\delta$ higher. Figure 2 shows an example of a $\delta$-interleaving.

Definition 1 (Morozov et al. [19]). Given two merge trees $(T, f)$ and $(T', f')$, a pair of maps $\alpha: T \to T'$ and $\beta: T' \to T$ is called a $\delta$-interleaving if for all $x \in T$ and $y \in T'$:

\begin{align*}
(C1) \quad f'(\alpha(x)) &= f(x) + \delta, \quad & (C3) \quad f'(\beta(y)) &= f'(y) + \delta, \quad & (C4) \quad \alpha(\beta(y)) &= y^{2\delta}. \\
(C2) \quad \beta(\alpha(x)) &= x^{2\delta},
\end{align*}

The interleaving distance $d_I$ is defined as the smallest $\delta$ such that there exists a $\delta$-interleaving.

---

1 Agarwal et al. [1] actually prove that approximating the Gromov-Hausdorff distance with a factor better than 3 is NP-hard. As many have observed, this proof also applies to the interleaving distance.
The maps $\alpha$ and $\beta$ are both $\delta$-shift maps, i.e. continuous maps that send points exactly $\delta$ higher. Touli and Wang [23] give an alternative definition of the interleaving distance in terms of a single $\delta$-shift map with additional requirements. They call this map a $\delta$-good map. Gasparovich et al. [12] show that the interleaving distance can also be defined in terms of labelled merge trees: merge trees equipped with label-maps. Formally, let $n \geq 0$. We denote $[n] := \{1, \ldots, n\}$. A map $\pi: [n] \to T$ is called a label map if each leaf in $T$ is assigned at least one label. Note that $\pi$ is not restricted to vertices, and may map different labels to the same point. The induced matrix $M = M(T, f, \pi)$ of a labelled merge tree is defined by $M_{i,j} = f(lca(\pi(i), \pi(j)))$, where $lca(\cdot, \cdot)$ is the lowest common ancestor of two points. See Figure 3 for an example of a labelled merge tree and its induced matrix.

For a matrix $M$, the $\ell^\infty$-norm is defined as $\|M\|_\infty = \max_{i,j} |M_{i,j}|$. For two unlabelled merge trees $(T, f)$ and $(T', f')$, we refer to a pair of equally-sized label maps $(\pi, \pi')$ as a $\delta$-labelling if $\|M(T, f, \pi) - M(T', f', \pi')\|_\infty = \delta$. The $\delta$-good interleaving distance $d^G_\delta$ and the label interleaving distance $d^L_\delta$ are defined as the smallest $\delta$ such that there exists a $\delta$-good map or a $\delta$-labelling. It has been shown that $d_\delta = d^G_\delta = d^L_\delta$ [12, 23].

## 3 An Order-Preserving Interleaving Distance

We consider a new class of merge trees, which we call ordered merge trees. For a point $x \in T$ with $f(x) \leq h$, we denote by $x|h$ the unique ancestor of $x$ at height $h$. An ordered merge tree $(T, f, (\leq_h)_{h \geq 0})$ is a merge tree $(T, f)$ equipped with a set of total orders on the level sets of $T$, such that these orders are consistent (see Figure 4). Formally, for two heights $h_1 \leq h_2$
and two points $x_1, x_2$ in $f^{-1}(h_1)$, we require that $x_1 \leq_{h_1} x_2$ implies $x_1|h_2 \leq_{h_2} x_2|h_2$.

An ordered merge tree induces a binary relation $\sqsubseteq$ on the complete tree $T$: for $x_1, x_2 \in T$, we define $x_1 \sqsubseteq x_2$ if $x_1|h \leq_{h} x_2|h$, where $h = \max(f(x_1), f(x_2))$. This induced relation is not antisymmetric, and thus not a total order. If we restrict $\sqsubseteq$ to the set of leaves of $T$, denoted $\sqsubseteq_L$, we do obtain a total order. We refer to $\sqsubseteq_L$ as the induced leaf-order of $T$.

**Monotone interleaving distance.** We now define an order-preserving distance measure for ordered merge trees. Specifically, we say a $\delta$-shift map $\alpha: T \to T'$ is monotone if for all height values $h \geq 0$ and for any two points $x_1, x_2 \in f^{-1}(h)$ it holds that $x_1 \leq_{h} x_2$ implies $\alpha(x_1) \leq_{h+\delta} \alpha(x_2)$. A monotone $\delta$-interleaving is a $\delta$-interleaving $(\alpha, \beta)$ such that the maps $\alpha$ and $\beta$ are both monotone (see Figure 5). A monotone $\delta$-good map, in turn, is a $\delta$-good map $\alpha$ that is also monotone. Lastly, a $\delta$-labelling $(\pi, \pi')$ of size $n$ is monotone if for all $\ell_1, \ell_2 \in [n]$ it holds that $\pi(\ell_1) \sqsubseteq \pi(\ell_2)$ implies $\pi'(\ell_1) \sqsubseteq \pi'(\ell_2)$. The **monotone interleaving distance** $d_{MI}$ is defined as the smallest $\delta$ such that there exists a monotone $\delta$-interleaving.

Similarly, we can define the monotone $\delta$-good, and the monotone label interleaving distances, denoted $d_{GMI}$ and $d_{LMI}$. Our main result is the following.

**Theorem 2.** The distances $d_{MI}$, $d_{GMI}$ and $d_{LMI}$ are equal.

To prove Theorem 2, we describe constructions between monotone $\delta$-interleavings, monotone $\delta$-good maps, and monotone $\delta$-labellings. The first construction, from a monotone $\delta$-interleaving to a monotone $\delta$-good map, follows directly from the regular setting (by Touli and Wang [23]). The construction of a monotone $\delta$-labelling from a monotone $\delta$-good map follows from a refinement of the construction by Gasparovich et al. [12]. We use $y^\bullet$ to denote the lowest ancestor of $y \in T'$ in the image of $\alpha$. Moreover, for two distinct points $x_1, x_2 \in T$, we say $x_1$ is smaller than $x_2$ if $x_1 \sqsubseteq x_2$. The existing construction is as follows.

---

2 We remark that they use a slightly different (but equivalent) definition for a $\delta$-good map.
Figure 6: The refined step (S2). The grey parts of $T'$ do not lie in the image of $\alpha$. We add the pair $(\hat{x}, w_5)$ to $\Pi$. In this example, $w = w_5$, $S = \{1, 2, 3, 6\}$, $i = 5$, $S_i = \{1, 2, 3\}$ and $i = 3$.

(S1) For every leaf $u \in L(T)$, add $(u, \alpha(u))$ to an initially empty set $\Pi$.
(S2) For every leaf $w \in L(T')$, take an arbitrary point $x \in \alpha^{-1}(w)$. Add $(x, w)$ to $\Pi$.
(S3) Consider an arbitrary ordering $\Pi = \{(x, y) \mid \ell \in [n]\}$, and set $\pi(\ell) = x_\ell$, $\pi'(\ell) = y_\ell$.

We refine (S2), by choosing a specific $x \in \alpha^{-1}(w)$, intuitively, we first identify all points $\hat{x}$ that lead to a violation of the monotonicity property if we choose $x$ smaller than $\hat{x}$. Such a violation occurs if $\hat{x}$ is an ancestor of a labelled point whose corresponding labelled point in $T'$ is smaller than $w$. We then take $x$ to be the largest point among the points $\hat{x}$.

(S2) For every leaf $w \in L(T')$, sort the set of leaves in $T'_x$ by induced leaf-order, denoted $W = \{w_1, \ldots, w_m\}$. Define $S \subseteq [m]$ such that for $k \in S$, $w_k$ is a strict descendant of $w$. Fix $i \in [m]$ such that $w_i = w$, and define $S_i = \{k \in S \mid k < i\}$. Now consider the set $X = \alpha^{-1}(w)$.
- If $S_i$ is empty, take $x$ to be the smallest point in $X$ and add $(x, w)$ to $\Pi$.
- If $S_i$ is not empty, consider the largest index $i \in S_i$. Define $Y$ as the set of strict descendants of $w$ that were labelled in (S1). Consider the following height values:

$$h_1 := \max\{f'(w_k) \mid k \in S\}, \quad h_2 := \max\{f'(y) \mid y \in Y\}, \quad \hat{h} = \max(h_1, h_2) \quad (1)$$

Consider the unique ancestor $\hat{w}_i$ of $w_i$ at height $\hat{h}$. Note that $\hat{w}_i$ is a strict descendant of the point $w$ and that it lies in the image of $\alpha$. Let $X_1 \subset X$ be the set of ancestors of points in $\alpha^{-1}(\hat{w}_i)$. Take $\hat{x}$ to be the largest point in $X_1$ and add $(\hat{x}, w)$ to $\Pi$.

See Figure 6 for an illustration. We can show that the resulting $\delta$-labelling is monotone.

Lastly, we construct a monotone $\delta$-interleaving from a monotone $\delta$-labelling. To do so, we extend an existing construction of a $\delta$-good map $\alpha$ from a $\delta$-labelling by Gasparovich et al. [12]. Specifically, for all $x \in T$, they consider an arbitrary label $\ell$ from the subtree $T_x$ and set $y_\ell = \pi'(\ell)/f(x)+\delta$. Gasparovich et al. show that the point $y_\ell$ is well-defined, and argue that the resulting map $\alpha$ is a $\delta$-good map. We can construct a $\delta$-interleaving $(\alpha, \beta)$ by using this construction twice: first to build a map $\alpha: T \rightarrow T'$ and next to build a map $\beta: T' \rightarrow T$.

Monotone leaf-label interleaving distance. We now turn to a restriction of the (monotone) label interleaving distance. Specifically, if we restrict a label map to map only to the leaves of $T$, we obtain a leaf-label map. A $\delta$-leaf-labelling is a pair of leaf-label maps that is also a $\delta$-labelling. The leaf-label interleaving distance $d^l_{\alpha, \beta}$, in turn, is defined as the smallest $\delta$ for which there exists a $\delta$-leaf labelling. We can show that this distance is an approximation of the interleaving distance, in both the regular and monotone setting. We define the leaf-gap $G$ of two trees $T$ and $T'$ as the maximum height difference of any pair of leaves in $T$ and $T'$.
Theorem 3. The monotone leaf-label interleaving distance between \((T, f, (\leq_h))\) and \((T', f', (\leq'_h))\) is bounded by \(2\delta + G\), where \(\delta = d_{\text{MH}}(T, T')\) and \(G\) is the leaf-gap of \(T\) and \(T'\).

4 Approximating the Monotone Interleaving Distance

In this section we describe an algorithm to compute the monotone leaf-label interleaving distance between two ordered merge trees \((T, f, (\leq_h))\) and \((T', f', (\leq'_h))\). We denote the leaves of \(T\) and \(T'\), sorted by leaf-order, by \(L(T) = \{u_1, \ldots, u_n\}\) and \(L(T') = \{w_1, \ldots, w_m\}\), respectively. We can show that we can permute the labels of a monotone labelling \((\omega, \omega')\) on \(T\) and \(T'\), such that the resulting label maps respect the induced leaf-orders of their trees. That is, for any two labels \(i, j \in [n]\) with \(i < j\), we have \(\omega(i) \leq_L \omega(j)\) and \(\omega'(i) \leq'_L \omega'(j)\). Assume \((\omega, \omega')\) is such a monotone leaf-labelling. Let \(M = M(T, f, \omega)\) and \(M' = M(T', f', \omega')\) be the corresponding induced matrices and set \(M = M(T, T') := |M - M'|\). We refer to entries \(M_{i,i}\) as the diagonal of \(M\), and to entries \(M_{i,j}\) with \(j - i = 1\) as the upper-diagonal of \(M\).

Lemma 4. The maximum of \(M\) lies on the diagonal or upper diagonal of \(M\).

We now describe a dynamic program to compute the monotone leaf-label interleaving distance between \(T\) and \(T'\). We denote by \(T[i]\) the subtree of \(T\) consisting of only the first \(i\) leaves. For \(i \in [L(T)]\) and \(j \in [L(T')]\), we maintain a value \(\Delta[i, j]\) that stores the monotone leaf-label interleaving distance between \(T[i]\) and \(T'[j]\). Consider an optimal monotone leaf-labelling \((\omega, \omega')\) for \(T[i]\) and \(T'[j]\) with a minimum number of \(k\) labels, such that \(\omega\) and \(\omega'\) respect the leaf-orders of \(T\) and \(T'\), respectively. From Lemma 4 we know that to compute \(d_{\text{MH}}^k(T[i], T'[j])\), it suffices to compute the diagonal and upper-diagonal entries of \(M = M(T[i], T'[j])\). As \(\omega(k) = u_i\) and \(\omega'(k) = w_j\), we have \(M_{k,k} = |f(u_i) - f'(w_j)| = \varepsilon\). To compute the other relevant elements of \(M\), we consider the three options for label \(k - 1\):

\[
\begin{align*}
(1) & \quad \omega(k - 1) = u_i, & (2) & \quad \omega(k - 1) = u_{i-1}, & (3) & \quad \omega(k - 1) = u_{i-1}, \\
& \quad \omega'(k - 1) = w_{j-1}, & & \quad \omega'(k - 1) = w_j, & & \quad \omega'(k - 1) = w_{j-1}.
\end{align*}
\]

First assume case (1) applies. Then we know that \(M_{k-1,k} = |f(u_i) - f'(\text{lca}(w_{j-1}, w_j))|\). Furthermore, \(\Delta[i, j - 1]\) captures the remaining relevant entries of \(M\). We set \(\delta_1 = \max(\Delta[i, j - 1], M_{k-1,k})\). Similarly, we can set \(\delta_2\) and \(\delta_3\) for cases (2) and (3) respectively. Finally, at each iteration, we set \(\Delta[i, j] = \max(\varepsilon, \min(\delta_1, \delta_2, \delta_3))\). We can show that \(\Delta[i, j] = d_{\text{MH}}^k(T[i], T'[j])\). The algorithm returns \(\Delta[|L(T)|, |L(T')|]\).

We can use our algorithm to compute a monotone interleaving as follows. First, we compute the lowest common ancestors of all consecutive pairs of leaves in \(T\) or \(T'\). This allows us to construct \(\Delta\) in \(O(n^2)\) time, where \(n = |L(T)| + |L(T')|\). Recovering an optimal leaf-labelling from the dynamic program can be done in a standard way. Next, we can construct two partial maps \(\alpha_L: L(T) \rightarrow T'\) and \(\beta_L: L(T') \rightarrow T\) using the construction from Section 3 in \(O(n)\) time. Lastly, one can recover a complete interleaving \((\alpha, \beta)\) from \(\alpha_L\) and \(\beta_L\) using continuity and \(\delta\)-shift map properties. Our final result follows from Theorem 3:

Theorem 5. Given two ordered merge trees \(T\) and \(T'\), there exists an \(O(n^2)\) algorithm that computes a monotone \(\delta\)-interleaving between \(T\) and \(T'\), where \(\delta \leq 2d_{\text{MH}}(T, T') + G\).

5 Acknowledgments

Thijs Beurskens and Tim Ophelders are supported by the Dutch Research Council (NWO) under project numbers OCENW.M20.089 and VI.Veni.212.260, respectively.
References


2:8  An Interleaving Distance for Ordered Merge Trees

21  G. O’Beirne. Waimakariri River (New Zealand), 2017. Image is slightly cropped; accessed on 30/11/2023. Photo is licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/). URL: https://commons.wikimedia.org/w/index.php?curid=61302377.