Fast Approximations and Coresets for (k, ℓ)-Median under Dynamic Time Warping

Jacobus Conradi[1], Benedikt Kolbe[2], Ioannis Psarros[3], and Dennis Rohde[1]

1 **Department of Computer Science, University of Bonn, Germany**
2 **Hausdorff Center for Mathematics, University of Bonn, Germany**
3 **Archimedes, Athena Research Center, Greece**

Abstract

We present algorithms for the computation of ε-coresets for k-median clustering of point sequences in $\mathbb{R}^d$ under the p-dynamic time warping (DTW) distance. Coresets under DTW have not been investigated before, and the analysis is not directly compatible with existing methods as DTW is not a metric. We achieve our results by investigating approximations of DTW that provide a trade-off between the provided accuracy and amenability to known techniques. In particular, we observe that given n curves under DTW, one can directly construct a metric that approximates DTW on this set, permitting the use of the wealth of results on metric spaces for clustering purposes. The resulting approximations are the first with polynomial running time and achieve a very similar approximation factor compared to state-of-the-art techniques.

1 Introduction

One of the core challenges of contemporary data analysis is the handling of massive data sets. A powerful approach to clustering problems involving such sets is data reduction, and ε-coresets offer a popular approach that has received substantial attention [4, 5, 14]. An ε-coreset is a problem-specific condensate of the given input set of reduced size which captures its core properties towards the problem at hand and can be used as a proxy to run an algorithm on, producing a solution with a relative error of $(1 \pm \varepsilon)$.

Clustering and especially k-median represent fundamental tasks in classification problems, where they have been extensively studied for various spaces. With the growing availability of e.g. geospatial tracking data, clustering problems for time series or curves have received growing attention both from a theoretical and applied perspective. In practice, time series classification largely relies on the dynamic time warping (DTW) distance and is widely used in the area of data mining. Simple nearest neighbor classifiers under DTW are considered hard to beat [17, 24] and much effort has been put into making classification using DTW computationally efficient [16, 19, 20, 21].

For time series and curves, k-median takes the shape of the (k, ℓ)-median problem, where the sought-for center curves are restricted to have a complexity (number of vertices) of at most ℓ, with a two-fold motivation. First, the otherwise NP-hard problem becomes tractable, and second, it suppresses overfitting.

The construction of ε-coresets for the (k, ℓ)-median problem for DTW is precisely what this paper will address. To this end, we adapt the framework of *sensitivity sampling* by Feldman and Landberg [13] to our setting. We rely on approximations of nearly all objects involved in our inquiry, thereby improving the bounds we obtain for the VC dimension of the range spaces in question and broadening the scope of our approach.

All presently known approaches to the approximation of the (k, ℓ)-median problem are based on an approximation scheme [6, 10, 12, 1, 7, 18]. For DTW, the best algorithm [8]

40th European Workshop on Computational Geometry, Ioannina, Greece, March 13–15, 2024.
This is an extended abstract of a presentation given at EuroCG'24. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear eventually in more final form at a conference with formal proceedings and/or in a journal.

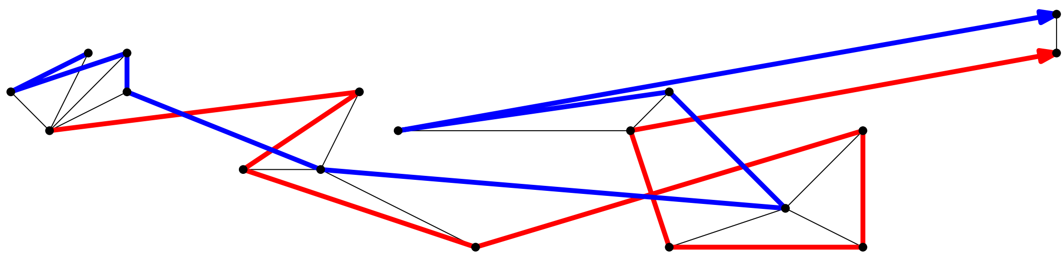

Figure 1 Example of a traversal between the red and blue curve realizing the dynamic time warping distance. The sum of the black distances is minimized.

has running time exponential in k, roughly with a dependency of $\widetilde{O}((32k^2\varepsilon^{-1})^{k+2}n)$.

Our results and methods We derive a bound on the VC dimension of a range space that approximates that of closed balls under DTW, obtained from a distance function approximating DTW. We modify and apply the sensitivity sampling framework by Feldman and Langberg [13], which relies on the bounds of the VC dimension and requires a first rough approximation of the (k, ℓ)-median problem under DTW, to construct coresets for (k, ℓ)-median under DTW. To adapt the sensitivity sampling framework to our (non-metric) setting, we investigate weaker versions of the triangle inequality for DTW and find a generalized iterated triangle inequality (Lemma 4.1). This novel inequality allows the approximation of DTW with a metric and thus the application of metric clustering algorithms.

2 Preliminaries

We think of a sequence $(p_1, \dots, p_m) \in \left(\mathbb{R}^d\right)^m$ of points in $\mathbb{R}^d$ as a (polygonal) *curve*, with complexity m. We denote by $\mathbb{X}^d_{=m}$ the space of curves in $\mathbb{R}^d$ with complexity exactly m and by $\mathbb{X}^d_m$ the space of curves with complexity at most m.

▶ **Definition 2.1** (p-Dynamic Time Warping)**.** For given $m, \ell > 0$ we define the space $\mathcal{T}_{m,\ell}$ of (m, ℓ)-*traversals* as the set of sequences $((a_1, b_1), (a_2, b_2), \dots, (a_l, b_l))$, such that

- $a_1 = 1$ and $b_1 = 1$; and $a_l = m$ and $b_l = \ell$,
- for all $i \in [l-1] := \{1, \dots, l-1\}$ it holds that $(a_{i+1}, b_{i+1}) - (a_i, b_i) \in \{(1,0), (0,1), (1,1)\}$.

For $p \in [1, \infty)$ and two curves $\sigma = (\sigma_1, \dots, \sigma_m) \in \mathbb{X}^d_{=m}, \tau = (\tau_1, \dots, \tau_\ell) \in \mathbb{X}^d_{=\ell}$ the (p-)Dynamic Time Warping distance (p-DTW) is defined as

$$\mathrm{dtw}_p(\sigma, \tau) = \min_{T \in \mathcal{T}_{m,\ell}} \left(\sum_{(i,j) \in T} \|\sigma_i - \tau_j\|_2^p \right)^{1/p},$$

where $\|\cdot\|_2^p$ is the Euclidean norm raised to the p-th power.

The central focus of the paper is the following clustering problem.

▶ **Definition 2.2** (Problem definition)**.** The (k, ℓ)-median problem for $\mathbb{X}^d_m$ and $k \in \mathbb{N}$ is the following: Given a set of $n \in \mathbb{N}$ input curves $T = \{\tau_1, \dots, \tau_n\} \subset \mathbb{X}^d_m$, identify k center curves $C = \{c_1, \dots, c_k\} \subset \mathbb{X}^d_\ell$ that minimize $\mathrm{cost}(T, C) = \sum_{\tau \in T} \min_{c \in C} \mathrm{dtw}(\tau, c)$.

An influential approach to solving k-median problems is to construct a point set that acts as proxy on which to run computationally more expensive algorithms that yield solutions with approximation guarantees. The condensed input set is known as a coreset.

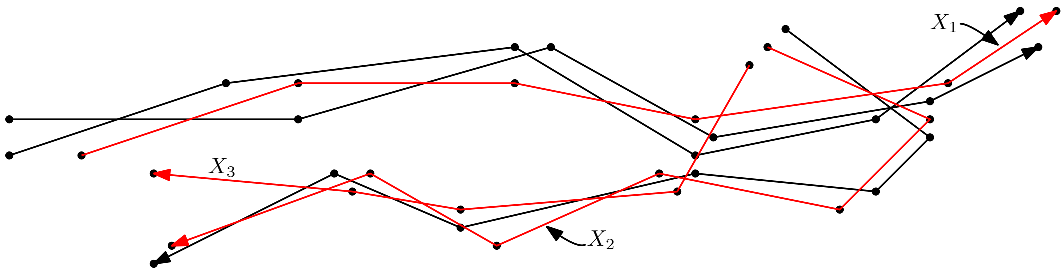

Figure 2 Illustration of a coreset (red), i.e. a weighted sparse representation of the original set of curves (in red and black). The weights in this case are $w(X_1) = 3$, $w(X_2) = 2$ and $w(X_3) = 1$.

▶ **Definition 2.3** (ε-coreset)**.** Let $T \subset \mathbb{X}_m^d$ be a finite set and $\varepsilon \in (0, 1)$. Then a weighted multiset $S \subset \mathbb{X}_m^d$ with weight function $w : S \to \mathbb{R}_{>0}$ is a weighted ε-coreset for (k, ℓ)-median clustering of T under dtw_p if for all $C \subset \mathbb{X}_\ell^d$ with $|C| = k$

$$(1-\varepsilon)\operatorname{cost}(T, C) \leq \sum_{s\in S} w(s) \min_{c\in C} \mathrm{dtw}_p(s, c) \leq (1+\varepsilon)\operatorname{cost}(T, C).$$

▶ **Definition 2.4** ((α, β)-approximation)**.** Let a set of $n \in \mathbb{N}$ input curves $T = \{\tau_1, \ldots, \tau_n\} \subset \mathbb{X}_m^d$ be given. A set $\hat{C} \subset \mathbb{X}_\ell^d$ is called an (α, β)-approximation of (k, ℓ)-median, if $|\hat{C}| \leq \beta k$ and $\sum_{\tau\in T} \min_{c\in\hat{C}} \mathrm{dtw}_p(\tau, c) \leq \alpha \sum_{\tau\in T} \min_{c\in C} \mathrm{dtw}_p(\tau, c)$ for any $C \subset \mathbb{X}_\ell^d$ of size k.

Focusing on approximations allows us to pass through simplifications of the input curves.

▶ **Definition 2.5** ($(1 + \varepsilon)$-approximate ℓ-simplifications)**.** Let $\sigma \in \mathbb{X}_m^d$, $\ell \in \mathbb{N}$ and $\varepsilon > 0$. We call $\sigma^* \in \mathbb{X}_\ell^d$ an $(1 + \varepsilon)$-*approximate* ℓ-*simplification* if

$$\inf_{\sigma_\ell\in\mathbb{X}_\ell^d} \mathrm{dtw}_p(\sigma_\ell, \sigma) \leq \mathrm{dtw}_p(\sigma^*, \sigma) \leq (1+\varepsilon) \inf_{\sigma_\ell\in\mathbb{X}_\ell^d} \mathrm{dtw}_p(\sigma_\ell, \sigma).$$

A *range space* is defined as a pair of sets $(X, \mathcal{R})$, where X is the *ground set* and $\mathcal{R}$ is the *range set* which is a subset of the power set $\mathcal{P}(X) = \{X' | X' \subset X\}$. Let $(X, \mathcal{R})$ be a range space. For $Y \subseteq X$, we denote: $\mathcal{R}_{|Y} = \{R \cap Y \mid R \in \mathcal{R}\}$. If $\mathcal{R}_{|Y} = \mathcal{P}(Y)$, then Y is *shattered* by $\mathcal{R}$. A key property of range spaces is the so called Vapnik-Chernovenkis dimension [22, 23, 25] (VC dimension) which for a range space $(X, \mathcal{R})$ is the maximum cardinality of a shattered subset of X.

3 VC Dimension bounds and coresets for DTW

We now derive bounds on the VC dimension of a range space that approximates the range space induced by all closed balls in $\mathbb{X}_m^d$ centered at curves in $\mathbb{X}_\ell^d$ under p-DTW. The following lemma shows that one can determine (approximately) the p-DTW between two sequences, based solely on the signs of certain polynomials, that are designed to provide an approximation of all point-wise distances and forms the basis for the results in this section. Missing proofs and statements can be found in the full version [11].

▶ **Lemma 3.1.** *Let $\tau \in \mathbb{X}_{=\ell}^d$, $\sigma \in \mathbb{X}_{=m}^d$, $r > 0$ and $\varepsilon \in (0, 1]$. For each $i \in [\ell]$, $j \in [m]$ and $z \in [\lfloor \varepsilon^{-1} + 1 \rfloor]$ define $f_{i,j,z}(\tau, r, \sigma) = \|\tau_i - \sigma_j\|^2 - (z \cdot \varepsilon r)^2$. There is an algorithm that, given as input the values of $\operatorname{sign}(f_{i,j,z}(\tau, r, \sigma))$, for all $i \in [\ell], j \in [m]$ and $z \in [\lfloor \varepsilon^{-1} + 1 \rfloor]$, outputs a value in $\{0, 1\}$ such that if $\mathrm{dtw}_p(\tau, \sigma) \leq r$ then it outputs 1 and if $\mathrm{dtw}_p(\tau, \sigma) > (1 + (m + \ell)^{1/p}\varepsilon)r$ then it outputs 0.*

The algorithm of Lemma 3.1 essentially implements approximate p-DTW balls membership and satisfies the requirements set by previous results that upper bound the VC dimension by decomposing the underlying predicate to sign evaluations of polynomials (Theorem 8.3 [2]). However, it is only defined on curves in $\mathbb{X}^d_{=\ell}$ and $\mathbb{X}^d_{=m}$. We extend the approach to all curves in $\mathbb{X}^d_m$, which provides the basis for a distance function $\widetilde{\mathrm{dtw}}_p$ between elements of $\mathbb{X}^d_m$ and $\mathbb{X}^d_\ell$ that approximates dtw_p with a relative error of $(1+\varepsilon)$. The properties of $\widetilde{\mathrm{dtw}}_p$ culminate in the following theorem giving a bound on the VC dimension on the approximate range space of p-DTW induced by $\widetilde{\mathrm{dtw}}_p$.

▶ **Theorem 3.2.** *Let $\varepsilon \in (0,1]$ and $\widetilde{\mathcal{R}}^p_{m,\ell} = \{\{x \in \mathbb{X}^d_m \mid \widetilde{\mathrm{dtw}}_p(x,\tau) \leq r\} \subset \mathbb{X}^d_m \mid \tau \in \mathbb{X}^d_\ell, r > 0\}$ be the range set consisting of all balls centered at elements of $\mathbb{X}^d_\ell$ under $\widetilde{\mathrm{dtw}}_p$ in $\mathbb{X}^d_m$. The VC dimension of $(\mathbb{X}^d_m, \widetilde{\mathcal{R}}^p_{m,\ell})$ is in $O(d\ell\log(\ell m\varepsilon^{-1}))$.*

Sensitivity bounds and coresets for DTW To make use of the sensitivity sampling framework for coresets by Feldman and Langberg [13], we recast the input set $T \subset \mathbb{X}^d_m$ as a set of functions. Consider for any $y \in \mathbb{X}^d_m$ and $\varepsilon > 0$ the real-valued function $\widetilde{f}_y$ defined on (finite) subsets of $\mathbb{X}^d_\ell$ by $\widetilde{f}_y : \mathcal{P}(\mathbb{X}^d_\ell) \setminus \{\emptyset\} \to \mathbb{R}$ with $\widetilde{f}_y(C) = \min_{c\in C} \widetilde{\mathrm{dtw}}_p(y,c)$, transforming T into $\widetilde{F}_T = \{\widetilde{f}_\tau \mid \tau \in T\}$. To construct a coreset, one draws elements from T according to a fixed probability distribution over T, and reweighs each drawn element. Both the weight and sampling probability are expressed in terms of the *sensitivity* of the drawn element t, which describes the maximum possible relative contribution of t to the cost of any query evaluation. We bound the sensitivity of each $\widetilde{f}_\tau \in \widetilde{F}_T$ by a function $\gamma(\widetilde{f}_\tau)$, which solely depends on a (α,β)-approximation, m, ℓ and p. The sensitivity sampling framework and Theorem 3.2 then yield Theorem 3.3.

▶ **Theorem 3.3.** *For $\widetilde{f} \in \widetilde{F}$, let $\lambda(\widetilde{f}) = 2^{\lceil \log_2(\gamma(\widetilde{f}))\rceil}$, with $\gamma(\widetilde{f})$ the aforementioned sensitivity bound, associated to an (α,β)-approximation consisting of $\hat{k} \leq \beta k$ curves, for (k,ℓ)-median for curves in $\mathbb{X}^d_m$ under dtw_p, $\Lambda = \sum_{\widetilde{f}\in\widetilde{F}} \lambda(\widetilde{f})$, $\psi(\widetilde{f}) = \frac{\lambda(\widetilde{f})}{\Lambda}$ and $\delta, \varepsilon \in (0,1)$. A sample S of*

$$\Theta\Big(\varepsilon^{-2}\alpha\hat{k}(m\ell)^{1/p}\left((d\ell\log(\ell m\varepsilon^{-1}))k\log(k)\log(\alpha n)\log(\alpha\hat{k}(m\ell)^{1/p}) + \log(1/\delta)\right)\Big)$$

elements $\tau_i \in T$, drawn independently with replacement with probability $\psi(\widetilde{f}_i)$ and weighted by $w(\widetilde{f}_i) = \frac{\Lambda}{|S|\lambda(\widetilde{f}_i)}$ is a weighted ε-coreset for (k,ℓ)-median clustering of T under dtw_p with probability at least $1-\delta$.

4 Linear time approximation algorithm for (k,ℓ)-median

As Theorem 3.3 requires an initial (α,β)-approximate solution of the (k,ℓ)-median problem to compute the bounds $\gamma(\cdot)$ of the sensitivities, we turn to developing approximation algorithms for (k,ℓ)-median for a set $T \subset \mathbb{X}^d_m$ of n curves. For this, we approximate dtw_p on T by a metric using a new inequality for dtw_p (Lemma 4.1). This allows the use of any approximation algorithm for metric k-median, leading to an initial approximation algorithm of the original problem. Combined with a k-median algorithm in metric spaces [15], we obtain a linear time $(O((m\ell)^{1/p}), 1)$-approximation algorithm, which in turn allows us to compute a coreset in linear (in n) time.

Metrification of p-DTW We begin with the following more general triangle inequality for dtw_p, which motivates analysing the metric closure of the input set. While dtw_p does not

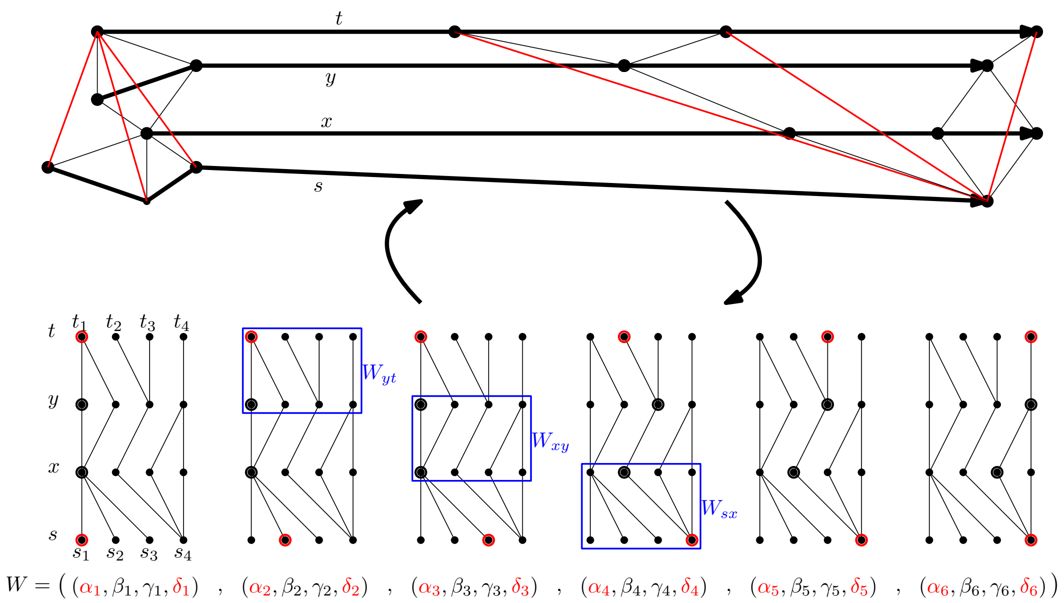

Figure 3 Illustration of how the optimal traversals W_{sx}, W_{xy} and W_{yt} of visited curves can be 'composed' to yield a set W that induces a traversal $\widetilde{W}$ (in red) of s and t. Any single matched pair of vertices in W_{sx}, W_{xy} or W_{yt} is at most $|W| \leq \ell + \ell'$ times a part of W.

satisfy the triangle inequality, the inequality shows it is never 'too far off'. Remarkably, the inequality does not depend on the complexity of the curves 'visited'. Missing proofs of this section can be found in the full version [11]. Figure 3 illustrates Lemma 4.1.

▶ **Lemma 4.1** (Iterated triangle inequality). *Let $s \in \mathbb{X}^d_\ell$, $t \in \mathbb{X}^d_{\ell'}$ and $X = (x_1, \ldots, x_r)$ be an arbitrary ordered set of curves in $\mathbb{X}^d_m$. Then*

$$\mathrm{dtw}_p(s,t) \leq (\ell + \ell')^{1/p} \left(\mathrm{dtw}_p(s, x_1) + \sum_{i<r} \mathrm{dtw}_p(x_i, x_{i+1}) + \mathrm{dtw}_p(x_r, t) \right).$$

▶ **Definition 4.2** (metric closure). Let (X, ϕ) be a finite set endowed with a distance function $\phi : X \times X \to \mathbb{R}$. The metric closure $\overline{\phi}$ of ϕ is the function

$$\overline{\phi} : X \times X \to \mathbb{R}, (s,t) \mapsto \min_{\substack{r \geq 2, \{\tau_1, \ldots, \tau_r\} \subset X \\ s = \tau_1, t = \tau_r}} \sum_{i<r} \phi(\tau_i, \tau_{i+1}).$$

▶ **Lemma 4.3.** *For any set of curves X and two curves $\sigma, \tau \in X$ of complexity at most m it holds that the metric closure $\overline{\mathrm{dtw}_p|_X}$ of the restriction of dtw_p onto the set X is bounded by $\mathrm{dtw}_p(\sigma,\tau) \leq (2m)^{1/p}\overline{\mathrm{dtw}_p|_X}(\sigma,\tau) \leq (2m)^{1/p}\,\mathrm{dtw}_p(\sigma,\tau)$.*

Linear time algorithm Naïvely, we would like to apply linear time algorithms [9] to the metric closure of dtw_p. However, constructing the metric closure usually takes cubic time resulting in cubic time algorithms. We circumvent this by applying Indyk's sampling technique for bicriteria k-median approximation [15], which reduces a k-median instance with n points to two k-median instances with $O(\sqrt{n})$ points, simply by sampling. We apply this technique twice, so that we only compute the metric closure on four sampled subsets of size $O(n^{1/4})$, resulting in the following theorem.

▶ **Theorem 4.4.** *For any $\varepsilon > 0$ there is an algorithm which computes a $(O(1+\varepsilon)(m\ell^3)^{(1/p)}, 4)$-approximation for (k,ℓ)-median for an input set X of n curves of complexity m under dtw_p in time*

$$O(nm^3d + nk\log(k)\ell^2 d + nk^2\log^2(k)\varepsilon^{-4}\log^2(\varepsilon^{-1}) + k^7\varepsilon^{-5}\log^5(n)).$$

5 Coreset Application

The theoretical derivations of the previous sections culminate in an approximation algorithm (Theorem 5.1) to (k,ℓ)-median that is particularly useful in the big data setting, where $n \gg m$. Our strategy is to first compute an efficient but not very accurate approximation(Theorem 4.4) of (k,ℓ)-median, which we use to construct a coreset. By virtue of the size reduction we greatly reduce the running time of slower more accurate algorithms, yielding a better approximation. Missing proofs can be found in the full version [11].

Algorithm 1 $((32+\varepsilon)(4m\ell)^{1/p})$-approximate (k,ℓ)-median

```
procedure (k,ℓ)-MEDIAN(X ⊂ 𝕏^d_m, p, ε)
    ε′ ← ε/46
    Compute (O((16mℓ³)^{1/p}), 4)-approximation C′ (Theorem 4.4)
    Compute bound γ(f̃_x) of sensitivity for each curve x ∈ X from C′
    Compute sample size s ← O(ε^{-2} dℓk² (m²ℓ⁴)^{1/p} log³(mℓ) log²(k) log(ε^{-1}) log(n))
    Sample and weigh ε′-coreset S of X of size s (Theorem 3.3)
    Compute a 2-simplification for every curve in S resulting in the set S* of curves
    Compute metric closure φ̄(x, y) = \overline{dtw_p |_{S*}}(x, y) for every x, y ∈ S*
    Return (5 + ε′, 1)-approximation of weighted metric k-median in (S*, φ̄) (c.f. [3, 9])
end procedure
```

▶ **Theorem 5.1.** *Let $0 < \varepsilon \leq 1$. There is an $((32+\varepsilon)(4m\ell)^{1/p}, 1)$-approximate algorithm with constant success probability ((k,ℓ)-MEDIAN in Algorithm 1) for (k,ℓ)-median on curves under dtw_p with a running time of $\widetilde{O}\left(n(m^3d + k^2 + k\ell^2 d) + \varepsilon^{-6}d^3\ell^3k^7\sqrt[p]{m^6\ell^{12}}\right)$, where $\widetilde{O}$ hides polylogarithmic factors in n, m, ℓ, k and ε^{-1}.*

▶ **Corollary 5.2.** *There is an algorithm that computes an ε-coreset for (k,ℓ)-median in time $\widetilde{O}\left(n(m^3d + k^2 + k\ell^2 d) + \varepsilon^{-6}d^3\ell^3k^7\sqrt[p]{m^6\ell^{12}}\right)$ with constant success probability of size*

$$O(\varepsilon^{-2}d\ell k^2(m^2\ell^2)^{1/p}\log^3(m\ell)\log^2(k)\log(\varepsilon^{-1})\log(n)).$$

Acknowledgments. Jacobus Conradi was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 313421352 (FOR 2535 Anticipating Human Behavior) and the iBehave Network: Sponsored by the Ministry of Culture and Science of the State of North Rhine-Westphalia. Benedikt Kolbe was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 459420781. Ioannis Psarros has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program. Dennis Rohde was partially supported by the Hausdorff Center for Mathematics and the iBehave Network: Sponsored by the Ministry of Culture and Science of the State of North Rhine-Westphalia.

References

1 Marcel R. Ackermann, Johannes Blömer, and Christian Sohler. Clustering for metric and nonmetric distance measures. *ACM Transactions on Algorithms*, 6(4):59:1–59:26, 2010.

2 Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999. `doi:10.1017/CBO9780511624216`.

3 Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local Search Heuristics for k-Median and Facility Location Problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.

4 Olivier Bachem, Mario Lucic, and Andreas Krause. Scalable k-Means Clustering via Lightweight Coresets. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, (KDD)*, pages 1119–1127, 2018.

5 Vladimir Braverman, Vincent Cohen-Addad, Shaofeng H.-C. Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Toftrup, and Xuan Wu. The power of uniform sampling for coresets. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 462–473. IEEE, 2022. `doi:10.1109/FOCS54457.2022.00051`.

6 Maike Buchin, Anne Driemel, and Dennis Rohde. Approximating (k, ℓ)-Median Clustering for Polygonal Curves. In Dániel Marx, editor, *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 2697–2717, Virtual Conference, January 2021. SIAM.

7 Maike Buchin, Anne Driemel, and Dennis Rohde. Approximating (k, ℓ)-median clustering for polygonal curves. *ACM Trans. Algorithms*, 19(1):4:1–4:32, 2023.

8 Maike Buchin, Anne Driemel, Koen van Greevenbroek, Ioannis Psarros, and Dennis Rohde. Approximating length-restricted means under dynamic time warping. In Parinya Chalermsook and Bundit Laekhanukit, editors, *Approximation and Online Algorithms - 20th International Workshop, WAOA 2022, Potsdam, Germany, September 8-9, 2022, Proceedings*, volume 13538 of *Lecture Notes in Computer Science*, pages 225–253. Springer, 2022.

9 Ke Chen. On Coresets for k-Median and k-Means Clustering in Metric and Euclidean Spaces and Their Applications. *SIAM Journal on Computing*, 39(3):923–947, 2009.

10 Siu-Wing Cheng and Haoqiang Huang. Curve simplification and clustering under fréchet distance. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1414–1432. SIAM, 2023.

11 Jacobus Conradi, Benedikt Kolbe, Ioannis Psarros, and Dennis Rohde. Fast approximations and coresets for (k, l)-median under dynamic time warping. *CoRR*, abs/2312.09838, 2023. URL: `https://doi.org/10.48550/arXiv.2312.09838`, `arXiv:2312.09838`, `doi:10.48550/ARXIV.2312.09838`.

12 Anne Driemel, Amer Krivosija, and Christian Sohler. Clustering time series under the Fréchet distance. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 766–785, 2016.

13 Dan Feldman and Michael Langberg. A unified framework for approximating and clustering data. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing*, pages 569–578. ACM, 2011.

14 Lingxiao Huang, Shaofeng H.-C. Jiang, Jian Li, and Xuan Wu. Epsilon-coresets for clustering (with outliers) in doubling metrics. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 814–825, 2018.

15 Piotr Indyk. Sublinear time algorithms for metric space problems. In Jeffrey Scott Vitter, Lawrence L. Larmore, and Frank Thomson Leighton, editors, *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing, May 1-4, 1999, Atlanta, Georgia, USA*, pages 428–434. ACM, 1999. `doi:10.1145/301250.301366`.

16 Youngseon Jeong, Myong Kee Jeong, and Olufemi A. Omitaomu. Weighted dynamic time warping for time series classification. *Pattern Recognit.*, 44(9):2231–2240, 2011.

17 Rohit J. Kate. Using dynamic time warping distances as features for improved time series classification. *Data Min. Knowl. Discov.*, 30(2):283–312, 2016.

18 Amit Kumar, Yogish Sabharwal, and Sandeep Sen. A Simple Linear Time $(1+\varepsilon)$-Approximation Algorithm for k-Means Clustering in Any Dimensions. In *45th Symposium on Foundations of Computer Science (FOCS), 17-19 October, Rome, Italy, Proceedings*, pages 454–462. IEEE Computer Society, 2004.

19 François Petitjean, Germain Forestier, Geoffrey I. Webb, Ann E. Nicholson, Yanping Chen, and Eamonn J. Keogh. Dynamic time warping averaging of time series allows faster and more accurate classification. In Ravi Kumar, Hannu Toivonen, Jian Pei, Joshua Zhexue Huang, and Xindong Wu, editors, *2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, December 14-17, 2014*, pages 470–479. IEEE Computer Society, 2014.

20 François Petitjean, Germain Forestier, Geoffrey I. Webb, Ann E. Nicholson, Yanping Chen, and Eamonn J. Keogh. Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm. *Knowl. Inf. Syst.*, 47(1):1–26, 2016.

21 Thanawin Rakthanmanon, Bilson J. L. Campana, Abdullah Mueen, Gustavo E. A. P. A. Batista, M. Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn J. Keogh. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Trans. Knowl. Discov. Data*, 7(3):10:1–10:31, 2013.

22 Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory Series A*, 13:145–147, 1972.

23 Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1), 1972.

24 Tuan Minh Tran, Xuan-May Thi Le, Hien T. Nguyen, and Van-Nam Huynh. A novel non-parametric method for time series classification based on k-nearest neighbors and dynamic time warping barycenter averaging. *Eng. Appl. Artif. Intell.*, 78:173–185, 2019.

25 Vladimir Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.