

Why Hausdorff Distance Is Natural in Interval Computations

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
'500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

Several different metrics have been proposed to describe distance between intervals and, more generally, between compact sets. In this paper, we show that from the viewpoint of interval computations, the most adequate distance is the Hausdorff distance $d_H(A, A')$ – the smallest value $\varepsilon > 0$ for which every element $a \in A$ is ε -close to some element $a' \in A'$, and every element $a' \in A'$ is ε -close to some element $a \in A$.

1 Formulation of the Problem

Uncertainty-motivates sets as extensions of points. One of the main objectives of interval computations is to deal with uncertainty. Because of uncertainty, instead of the exact value a of a physical quantity, we only know a set A (usually, an interval) of possible values. In the cases of several variables, instead of a tuple $a = (a_1, \dots, a_n)$ consisting of their exact values, we only know a set A of possible tuples. In general, instead of the exact state a of the corresponding system, we only know a set A of possible states.

The case of complete knowledge can be viewed as a “degenerate” case, when the corresponding set A consists of a single element a : $A = \{a\}$.

Need to define distance between sets. In many practical situations, on the set X of all possible states, we have a physically meaningful distance $d(a, b)$. For example, on the set of real numbers, a natural distance is usually $d(a, b) = |a - b|$.

It is desirable to extend this distance from elements (i.e., degenerate sets) to a more general case of distance between sets.

There are many ways to define the distance between sets. In general, there are many ways to extend a function to its original domain to a larger domain. In particular, there are many ways to extend distance from a given set X (i.e., from the class of all 1-element subsets of the set X) to a larger class of subsets of X .

One way is to define *Hausdorff distance* $d_H(A, A')$ (see, e.g., [1]), the infimum of all the values $\varepsilon > 0$ for which:

- the set A is contained in the ε -neighborhood $\{x : d(x, A') \leq \varepsilon\}$ of the set A' , and
- the set A' is contained in the ε -neighborhood $\{x : d(x, A) \leq \varepsilon\}$ of the set A ,

where the distance $d(x, S)$ between an element x and the set S is defined as

$$d(x, S) \stackrel{\text{def}}{=} \inf_{s \in S} d(x, s).$$

The Hausdorff distance is usually defined for *compact* sets A and A' (see, e.g., [1]). For such sets,

- for every $a \in A$, there is a point $a' \in A'$ for which $d(a, a') \leq d_H(A, A')$, and
- for every $a' \in A'$, there exists a point $a \in A$ for which $d(a, a') \leq d_H(A, A')$.

However, other definitions are also possible. For example, since an interval $[\underline{x}, \bar{x}]$ can be naturally represented by a point (\underline{x}, \bar{x}) on a plane, many paper papers define the distance between intervals as the Euclidean distance between the corresponding 2-D points (see, e.g., [2]), i.e., as

$$d([\underline{a}, \bar{a}], [\underline{a'}, \bar{a'}]) = \sqrt{(\underline{a} - \underline{a'})^2 + (\bar{a} - \bar{a'})^2}.$$

Which definition is most adequate from the viewpoint of interval computation? Which definition is more adequate depends on the problems that we intend to solve. In this paper, we analyze which definitions are most appropriate from the viewpoint of interval computations.

One of the main objective of interval computations is:

- given a function $f : X \rightarrow \mathbb{R}$ and a set $A \subseteq X$,
- to compute the range $f(A) \stackrel{\text{def}}{=} \{f(x) : x \in A\}$ of the function $f(x)$ on the set A .

The functions $f(x)$ are usually continuous, and the set A is usually compact; a typical example is a box (multi-D interval) in an n -dimensional space. For a continuous function $f(x)$, its range on a compact set is also a compact set. It is also known that every real-valued continuous function on a compact set is *uniformly continuous*, i.e., for every $\varepsilon > 0$ there exists a $\delta > 0$ for which $d(a, a') \leq \delta$ implies that $|f(a) - f(a')| \leq \varepsilon$.

The corresponding sets A are also usually connected. It is known that the range of a continuous function on a connected set is always connected and

is, thus, an interval. In such situations, computing the range simply means computing the two endpoints of this interval:

$$\underline{f}(A) \stackrel{\text{def}}{=} \inf f(A) \text{ and } \overline{f}(A) \stackrel{\text{def}}{=} \sup f(A).$$

Towards a precise formalization of this idea. Let us summarize the above situation. For an exact point a , we compute a single value $f(a)$. We know that to compute this value $f(a)$ with accuracy $\varepsilon > 0$, we need to determine a with accuracy $\delta > 0$.

For a compact set A , we need to compute *two* values $\underline{f}(A)$ and $\overline{f}(A)$. It is therefore reasonable to define the distance $d(A, A')$ between the sets in such a way that to determine each of these two values $\underline{f}(A)$ and $\overline{f}(A)$ with accuracy ε , we need to define the set A with accuracy δ – in the sense of this yet-to-be-defined distance.

Let us describe this idea in precise terms.

2 Definition and the Main Result

Definition. Let $\varepsilon > 0$ and $\delta > 0$ be real numbers. We say that a continuous function $f : X \rightarrow \mathbb{R}$ from a metric space X to real numbers is (ε, δ) -continuous if for every two points x and x' for which $d(x, x') \leq \delta$, we have $|f(x) - f(x')| \leq \varepsilon$.

The following proposition shows that the Hausdorff metric $d_H(A, B)$ is – in the sense of the above idea – the most natural one for interval computations.

Proposition. Let $\varepsilon > 0$ and $\delta > 0$ be real numbers. For every two compact subsets A and A' of a metric space X , the following two conditions are equivalent to each other:

- for every (ε, δ) -continuous function $f : X \rightarrow \mathbb{R}$, we have

$$|\underline{f}(A) - \underline{f}(A')| \leq \varepsilon \text{ and } |\overline{f}(A) - \overline{f}(A')| \leq \varepsilon;$$

- $d_H(A, A') \leq \delta$.

Comment. So, for the Hausdorff distance:

- if $d(a, a') \leq \delta$ implies that the values $f(a)$ and $f(a')$ are ε -close,
- then whenever $d_H(A, A') \leq \delta$, both endpoints of the ranges $f(A)$ and $f(A')$ are also ε -close.

The Proposition also says that the Hausdorff metric is uniquely determined by this property.

Proof.

1°. Let us first prove that if $d_H(A, A') \leq \delta$ and $f(x)$ is an (ε, δ) -continuous function, then indeed $|\underline{f}(A) - \underline{f}(A')| \leq \varepsilon$ and $|\overline{f}(A) - \overline{f}(A')| \leq \varepsilon$.

1.1°. Let us first prove that the lower endpoints $\underline{f}(A)$ and $\underline{f}(A')$ are ε -close.

Indeed, since $f(x)$ is a continuous function on a compact set A , its infimum is attained, i.e., there exists a point $a_m \in A$ for which

$$f(a_m) = \underline{f}(A).$$

Since $d_H(A, A') \leq \delta$, there exists a point $a' \in A'$ for which $d(a_m, a') \leq \delta$. Since the function $f(x)$ is (ε, δ) -continuous, this implies that $|f(a_m) - f(a')| \leq \varepsilon$. Thus,

$$f(a') \leq f(a_m) + \varepsilon = \underline{f}(A) + \varepsilon.$$

By definition,

$$\underline{f}(A') = \inf\{f(a') : a' \in A'\}.$$

By definition of the infimum, this implies that

$$\underline{f}(A') \leq f(a').$$

Since we have proven that $f(a') \leq \underline{f}(A) + \varepsilon$, this implies that

$$\underline{f}(A') \leq \underline{f}(A) + \varepsilon.$$

We can repeat the same argument starting with the fact that $f(x)$ is a continuous function on a compact set A' , this would imply that

$$\underline{f}(A) \leq \underline{f}(A') + \varepsilon.$$

These two inequalities indeed imply that $|\underline{f}(A) - \underline{f}(A')| \leq \varepsilon$.

1.2°. Similarly, we can prove that the upper endpoints $\overline{f}(A)$ and $\overline{f}(A')$ are also ε -close.

Indeed, since $f(x)$ is a continuous function on a compact set A , its supremum is attained, i.e., there exists a point $a_M \in A$ for which $f(a_M) = \overline{f}(A)$. Since $d_H(A, A') \leq \delta$, there exists a point $a' \in A'$ for which $d(a_M, a') \leq \delta$. Since the function $f(x)$ is (ε, δ) -continuous, this implies that $|f(a_M) - f(a')| \leq \varepsilon$. Thus,

$$f(a') \geq f(a_M) - \varepsilon = \overline{f}(A) - \varepsilon.$$

By definition,

$$\overline{f}(A') = \sup\{f(a') : a' \in A'\}.$$

By definition of the supremum, this implies that

$$\overline{f}(A') \geq f(a').$$

Since we have proven that $f(a') \geq \bar{f}(A) - \varepsilon$, this implies that

$$\bar{f}(A') \geq \bar{f}(A) - \varepsilon.$$

We can repeat the same argument starting with the fact that $f(x)$ is a continuous function on a compact set A' , this would imply that

$$\bar{f}(A) \geq \bar{f}(A') - \varepsilon.$$

These two inequalities indeed imply that $|\bar{f}(A) - \bar{f}(A')| \leq \varepsilon$.

2°. Let us now assume that we have two sets A and A' for which, for every (ε, δ) -continuous function $f : X \rightarrow \mathbb{R}$, we have $|\underline{f}(A) - \underline{f}(A')| \leq \varepsilon$ and $|\bar{f}(A) - \bar{f}(A')| \leq \varepsilon$. Let us then prove that $d_H(A, A') \leq \delta$, i.e., that:

- for every point $a \in A$, we have $d(a, A') \leq \delta$, and
- for every point $a' \in A'$, we have $d(a', A) \leq \delta$.

Without losing generality, it is sufficient to prove the first statement. Let a be any point from the set A . Let us then take

$$f(x) \stackrel{\text{def}}{=} \frac{\varepsilon}{\delta} \cdot d(a, x).$$

Due to the triangle inequality, for every two points $x, x' \in X$, we have

$$|d(a, x) - d(a, x')| \leq d(x, x').$$

Multiplying both sides of this inequality by the ratio $\frac{\varepsilon}{\delta}$, we conclude that

$$|f(x) - f(x')| \leq \frac{\varepsilon}{\delta} \cdot d(x, x').$$

This implies that the function $f(x)$ is continuous.

When $d(x, x') \leq \delta$, the above inequality implies that $|f(x) - f(x')| \leq \varepsilon$, so the function $f(x)$ is indeed (ε, δ) -continuous. Thus, our assumption implies that the values $\underline{f}(A)$ and $\underline{f}(A')$ are ε -close.

The above function $f(x) = \frac{\varepsilon}{\delta} \cdot d(a, x)$ is always non-negative, and it is equal to 0 when $x = a$. Since $a \in A$, we thus conclude that $\underline{f}(A) = \inf\{f(x) : x \in A\} = 0$. From the fact that the values $\underline{f}(A)$ and $\underline{f}(A')$ are ε -close, we can therefore conclude that

$$\underline{f}(A') = \inf\{f(x) : x \in A'\} \leq \varepsilon.$$

By definition of $f(x)$, we have

$$\inf_{x' \in A'} f(x) = \frac{\varepsilon}{\delta} \cdot \inf_{x' \in A'} d(a, x'),$$

i.e., by definition of the distance between an element a and a set A' ,

$$\inf_{x' \in A'} f(x') = \frac{\varepsilon}{\delta} \cdot d(a, A').$$

Thus, the condition that

$$\inf_{x \in A'} f(x) = \frac{\varepsilon}{\delta} \cdot d(a, A') \leq \varepsilon$$

implies that $d(a, A') \leq \delta$.

The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants CAREER 0953339, HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award “UTEP and Prudential Actuarial Science Academy and Pipeline Initiative” from Prudential Foundation.

References

- [1] J. L. Kelley, *General Topology*, Ishi Press, New York, 2008.
- [2] *Proceedings of the 2016 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2016*, Vancouver, Canada, July 24–29, 2016.