Which Interval Is the Closest to a Given Set?

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

Abstract

In some practical situations, we know a set of possible values of a physical quantity – a set which is not an interval. Since computing with sets is often complicated, it is desirable to approximate this set by an easier-to-process set: namely, with an interval. In this paper, we describe intervals which are the closest approximations to a given set.

1 Formulation of the Problem

Why do we need 1-D sets beyond intervals. For each physical quantity, we would like to know which real numbers are possible values of this quantity.

For some quantities, the possible values come from fundamental physics. For example, the possible values of velocity form an interval $[0, c]$, where c is the speed of light. For other quantities – e.g., for size of some insect species – we have to rely on experts.

An expert usually provides us with a possible interval range. Different experts can provide different range. Some of these ranges may be non-intersecting: for example, one expert provides a range of adult cockroach sizes in St. Petersburg, Russia, where they are reasonably small, while another expert provides a range of adult cockroach sizes in El Paso, Texas, where the cockroaches are much larger.

If we trust all the experts, then we consider all the values supplied by all the experts as possible values of the corresponding quantity. Thus, after asking all the experts, we end up with the union of their answer sets as the desired set of possible values of the quantity – and this union of intervals is not necessarily an interval itself:

- In some cases, if two values $a < b$ are possible, then all intermediate values from the interval $[a, b]$ are also possible.
- However, this is not always the case: when a species consists of two populations of different sizes, the set of possible sizes is the union of two intervals; see, e.g., [5] and references therein.

It is often desirable to approximate a set by an interval. Processing sets is sometimes computationally complicated; see, e.g., [2].

It is therefore desirable to approximate the given set by an easier-to-process set, i.e., by an interval.

From the interval computation viewpoint, a reasonable approximation is the interval hull. In the spirit of interval computation (see, e.g., [1, 4]), a reasonable idea is to produce an interval enclosure for the desired set S. The narrowest such enclosure is the interval hull $[\inf S, \sup S]$.

What if we simply want to approximate? In some applications, it is not necessary to have an enclosure, we simply want an approximation – but, of course, we want the closest possible approximation. In this paper, we describe a solution to this approximation problem.

2 Definitions and the Main Result

How to gauge distance between sets. As distance between two sets A and B, we will take the *Hausdorff distance*

$$d_H(A, B) = \max\left(\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right),$$

where the distance $d(a, B)$ between an element a and a set B is defined in the usual way, as the smallest of the distances $d(a, b)$ between a and different elements $b \in B$:

$$d(a, B) = \inf_{b\in B} d(a, b).$$

Hausdorff distance can be alternatively defined as the infimum of all the values $\varepsilon > 0$ for which the following two properties are satisfied:

- for every element $a \in A$, there exists an ε-close element $b \in B$, i.e., an element b for which $d(a, b) < \varepsilon$, and
- for every element $b \in B$, there exists an ε-close element $a \in A$.

Intuitive meaning of Hausdorff distance in this case. Let us assume that the actual range of values of the physical quantity is an interval I. The experts do not provide us with exactly this interval, since their estimates are approximate, with some accuracy $\varepsilon > 0$. Because of this approximate character of expert estimates, they provide us with a set S which is somewhat different from the interval I.

We would like to make sure, however:

- that the experts' answers are *correct* – i.e., that each value $s \in S$ provided by an expert is ε-close to some possible value $x \in I$, and
- the experts' answers are *complete* – i.e., that each actually possible value $x \in I$ is ε-close to one of the values $s \in S$ provided by the experts.

As we can see, this means exactly that $d_H(S, I) \le \varepsilon$.

We want the closest possible interval. We do not know the experts' accuracy $\varepsilon > 0$. In such situations, it is reasonable to use the Maximum Compatibility Method [3, 6, 7], and look for the smallest possible value ε, i.e., to look for the interval I for which the distance $d_H(S, I)$ is the smallest possible:

$$d_H(S, I) = \min_J d_H(S, J),$$

where the minimum is taken over all intervals J.

Main result. Now, we are ready to formulate our main result.

Definition.

- *Let S be a bounded set, and let $\inf S \le a < b \le \sup S$. We say that the open interval (a, b) is a* gap *in the set S if $(a, b) \cap S = \emptyset$.*
- *For each set S, by $\delta(S)$, we mean the supremum of the half-widths $\dfrac{b-a}{2}$ of all the gaps of the set S.*

Proposition. *Let S be a bounded set. Then, an interval $I = [\underline{x}, \overline{x}]$ is the closest to the set S if and only if $|\inf S - \underline{x}| \le \delta(S)$ and $|\sup S - \overline{x}| \le \delta(S)$. For each of these intervals, we have $d_H(S, I) = \delta(S)$.*

Comment. The proof of this result is given in the following section.

Discussion. The above conditions are equivalent to the enclosures

$$[\inf S + \delta(S), \sup S - \delta(S)] \subseteq [\underline{x}, \overline{x}] \subseteq [\inf S - \delta(S), \sup S + \delta(S)].$$

Out of all the intervals that satisfy this condition, the narrowest is the interval

$$[\inf S + \delta(S), \sup S - \delta(S)]$$

and the widest is the interval

$$[\inf S - \delta(S), \sup S + \delta(S)].$$

If we take an arithmetic average of these two intervals, we get exactly the interval hull $[\inf S, \sup S]$ of the set S.

Alternatively, we can describe this result by saying that we have a *twin interval* (an "interval" whose endpoints are intervals), in which:

- the lower endpoint is the interval $[\inf S - \delta(S), \inf S + \delta(S)]$, and
- the upper endpoint is the interval $[\sup S - \delta(S), \sup S + \delta(S)]$.

2-D case is different. In the 1-D case, the interval hull $[\inf S, \sup S]$ of the set S is (one of the) closest intervals to this set.

In the 2-D case, we can formulate a similar problem by considering multi-D intervals (boxes) $B = [\underline{x}, \overline{x}] \times [\underline{y}, \overline{y}]$ approximating a given set S. Interestingly, in this case, the interval hull is not always the closest box to the original set S.

As a simple example, we can take the set S consisting of three line segments:

- a horizontal segment: $-1 \le x \le 1$ and $y = 0$,
- the first vertical segment: $x = -1$ and $-1 \le y \le 1$, and
- the second vertical segment: $x = 1$ and $-1 \le y \le 1$.

For this set, the interval hull is the box $B = [-1,1] \times [-1,1]$, for which $d_H(S,B) = 1$.

Indeed, $S \subseteq B$ and every point $(x,y) \in B$ is ≤ 1-close to the point $(x,0) \in S$, so $d_H(S,B) \le 1$. In particular, for the point $(0,1) \in B$, the closest points in S are at distance 1, so $d_H(S,B) = 1$.

On the other hand, one can easily check that for the box

$$B' = [-1,1] \times [-0.5, 0.5],$$

we have $d_H(S,B') = 0.5 < 1 = d_H(S,B)$. Thus, here, the multi-interval enclosure B is indeed *not* the closest multi-interval.

3 Proof of the Proposition

1°. Let us first prove, by contradiction, that for an interval I, we cannot have $d_H(S,I) < \delta(S)$.

Indeed, suppose that for some interval $I = [\underline{x}, \overline{x}]$, we have $d_H(S,I) < \delta(S)$.

1.1°. Let us first prove that in this case, we have $\underline{x} \le \inf S + \delta(S)$.

By definition of the infimum $\inf S$, for every n, there exists a point $s_n \in S$ for which $\inf S \le s_n \le \inf S + 2^{-n}$. By the definition of the Hausdorff distance, we have $d(s_n, I) \le \delta(S)$, so $|s_n - x| \le \delta(S)$ for some $x \in I$. For this x, we thus have $x \le s_n + \delta(S)$. Since $x \in I = [\underline{x}, \overline{x}]$, we have $\underline{x} \le x \le s_n + \delta(S)$. From $s_n \le \inf S + 2^{-n}$, we now conclude that $\underline{S} \le \inf S + 2^{-n} + \delta(S)$. In the limit $n \to \infty$, we get $\underline{x} \le \inf S + \delta(S)$.

1.2°. Similarly, by considering the points $s_n \in S$ which are close to $\sup S$, we conclude that $\overline{x} \ge \sup S - \delta(S)$.

1.3°. Let us now get the desired contradiction.

By definition, $\delta(S)$ is the supremum of half-widths of all the gaps of the set S. By definition of the supremum, this means that for every $\varepsilon > 0$, there exists a gap (a,b) for which $\delta \stackrel{\text{def}}{=} \dfrac{b-a}{2} \ge \delta(S) - \varepsilon$.

In particular, this is true for any positive value $\varepsilon < \dfrac{\delta(S) - d_H(S,I)}{2}$, for which $d_H(S,I) < \delta(S) - 2\varepsilon$.

Let us consider the midpoint $m = \dfrac{a+b}{2}$ of the corresponding gap. Then, we have three options:

- $m < \underline{x}$,

- $m > \overline{x}$, and
- $m \in I = [\underline{x}, \overline{x}]$.

Let us consider these three options one by one.

1.3.1°. Let us first consider the case when $m < \underline{x}$.

Here, $m = a + \delta$, where $m \geq \inf S$, thus, $m \geq \inf S + \delta \geq \inf S + \delta(S) - \varepsilon$. On the other hand, from Part 1.1 of this proof, we conclude that $\underline{x} \leq \inf S + \delta(S)$. Thus, we have

$$\inf S + \delta(S) - \varepsilon \leq m < \underline{x} \leq \inf S + \delta(S).$$

Both values m and $\underline{x}$ belong to the same interval $[\inf S + \delta(S) - \varepsilon, \inf S + \delta(S)]$ of width ε, hence $|m - \underline{x}| \leq \varepsilon$.

Since the point m is in the middle of the gap of radius δ, the closest point from the set S is at least at the distance δ: $d(m, S) \geq \delta$. Thus, by the triangle inequality, we have $d(\underline{x}, S) \geq d(m, S) - d(m, \underline{x}) \geq \delta - \varepsilon$. Since $\delta \geq \delta(S) - \varepsilon$, this implies that $d(\underline{x}, S) \geq \delta(S) - 2\varepsilon$.

We have selected ε for which $\delta(S) - 2\varepsilon > d_H(S, I)$. Thus, we have found an element $\underline{x} \in I = [\underline{x}, \overline{x}]$ for which $d(m, S) > d_H(S, I)$. This inequality contradicts to the definition of $d_H(S, I)$, according to which the Hausdorff distance is the largest of all the values $d(x, S)$ for $x \in I$ and values $d(s, I)$ for $s \in S$.

1.3.2°. A similar contradiction can be obtained when $m > \overline{x}$.

1.3.3°. To complete the proof of Part 1.3, it is now sufficient to consider the case when $m \in I = [\underline{x}, \overline{x}]$.

We have already mentioned that $d(m, S) \geq \delta > \delta(S) - \varepsilon$. By our choice of ε, we have $\delta(S) - \varepsilon > d_H(S, I)$, so we get $d(m, S) > d_H(S, I)$ for some point $m \in I$. Similarly to Part 1.3.1, this inequality leads to a contradiction.

1.3.4°. In all three possible cases, the assumption that $d_H(S, I) < \delta(S)$ for some interval I leads to a contradiction. Thus, for every interval I, we have $d_H(S, I) \geq \delta(S)$.

2°. Let us now prove that if $|\inf S - \underline{x}| \leq \delta(S)$ and $|\sup S - \overline{x}| \leq \delta(S)$, then for the interval $I = [\underline{x}, \overline{x}]$, we have $d_H(S, I) = \delta(S)$. To prove this, we need to prove that:

- for every $x \in I$, we have $d(x, S) \leq \delta(S)$, and
- for every $s \in S$, we have $d(s, I) \leq \delta(S)$.

Let us prove these two statements one by one.

2.1°. Let us first prove that if $x \in I$, then $d_H(x, S) \leq \delta(S)$.

To prove this, we will consider three possible cases:

- $x < \inf S$,
- $x > \sup S$, and

- $\inf S \le x \le \sup S$.

In the third case, we will consider two subcases:

- when x belongs to the closure $\overline{S}$ of the set S and
- when $x \notin \overline{S}$.

So, we need to consider four cases. Let us consider these four cases one by one.

2.1.1°. Let us first consider the case when $x < \inf S$.

Since $x \in I$, we have $\underline{x} \le x < \inf S$. By our assumption, $\inf S - \delta(S) \le \underline{x}$, hence $\inf S - \delta(S) \le x < \inf S$. Thus, $|x - \inf S| \le \delta(S)$, hence $d(x, S) \le d(x, \inf S) \le \delta(S)$.

2.1.2°. Similarly, we can prove that if $x > \sup S$, then $d(x, S) \le \delta(S)$.

2.1.3°. If $\inf S \le x \le \sup S$ and $x \in \overline{S}$, then clearly $d(x, S) = 0$ and thus, $d(x, S) \le \delta(S)$.

2.1.4°. Finally, let us consider the case when $\inf S \le x \le \sup S$ and $x \notin \overline{S}$.

In this case, x belongs to the complement to the closure of S. This complement is an open set, so from the fact that x belongs to this complement, we can conclude that x belongs to some gap of the set S.

By definition of $\delta(S)$, the width of this gap is $\le 2\delta(S)$. Thus, either to the left of x, or to the right of x, there must be a point $s \in \overline{S}$ for which $d(x, s) \le \delta(S)$ – otherwise, if both distances are larger, we would have a gap of width $> 2\delta(S)$. From the fact that $d(x, s) \le \delta(S)$ for some $s \in \overline{S}$, it follows that $d(x, S) \le \delta(S)$.

2.1.5°. In all four cases, we concluded that if $x \in I$, then $d(x, S) \le \delta(S)$.

2.2°. Let us now prove that if $s \in S$, then $d(s, I) \le \delta(S)$.

In this case, we can also consider three possible cases:

- case when $s < \underline{x}$,
- case when $s > \overline{x}$, and
- case when $\underline{x} \le s \le \overline{x}$.

Let us consider these three cases one by one.

2.2.1°. Let us first consider the case when $s < \underline{x}$.

From $s \in S$, we conclude that $\inf S \le s$. By our assumption, we have $\underline{x} \le \inf S + \delta(S)$. Thus, $\inf S \le s < \underline{x} \le \inf S + \delta(S)$.

So, both values s and $\underline{x}$ belong to the same interval $[\inf S, \inf S + \delta(S)]$, thus the difference between these two values cannot exceed the width $\delta(S)$ of this interval: $d(s, \underline{x}) \le \delta(S)$. Since $\underline{x} \in I$, we therefore have $d(s, I) \le d(s, \underline{x}) \le \delta(S)$ and so, indeed, $d(s, I) \le \delta(S)$.

2.2.2°. Similarly, we can prove that if $s > \overline{x}$, then $d(s, I) \le \delta(S)$.

2.2.3°. Finally, if $\underline{x} \le s \le \overline{x}$, this means that $s \in I = [\underline{x}, \overline{x}]$ thus $d(s, I) = 0$ and therefore, $d(s, I) \le \delta(S)$.

3°. To complete the proof, we need to prove that if for some interval $I = [\underline{x}, \overline{x}]$, we have $d_H(S, I) = \delta(S)$, then $|\inf S - \underline{x}| \le \delta(S)$ and $|\sup S - \overline{x}| \le \delta(S)$.

Without losing generality, let us prove that $|\inf S - \underline{x}| \le \delta(S)$. (The inequality $|\sup S - \overline{x}| \le \delta(S)$ is proved similarly.) To prove this double inequality, we must prove:

- that $\underline{x} \ge \inf S - \delta(S)$ and
- that $\inf S \ge \underline{x} - \delta(S)$.

Let us prove these two statements one by one.

3.1°. Let us first prove that $\underline{x} \ge \inf S - \delta(S)$.

Indeed, if $d_H(S, I) \le \delta(S)$, this means, in particular, that $d(\underline{x}, S) \le \delta(S)$. By the definition of the distance between an element and the set, this means that for every n, there exists a value $s_n \in S$ for which $|\underline{x} - s_n| \le \delta(S) + 2^{-n}$ and therefore, for which $\underline{x} \ge s_n - \delta(S) - 2^{-n}$. Since $s_n \ge \inf S$, this implies that $\underline{x} \ge \inf S - \delta(S) - 2^{-n}$. In the limit $n \to \infty$, we get $\underline{x} \ge \inf S$. The inequality is proven.

3.2°. Let us now prove that $\inf S \ge \underline{x} - \delta(S)$.

By definition of the infimum, for every n, there exists a value $s_n \in S$ for which $\inf S \le s_n \le \inf S + 2^{-n}$. Since $d_H(S, I) \le \delta(S)$, this implies that $d(s_n, I) \le \delta(S)$. So, $d(s_n, x) \le \delta(S)$ for some $x \in I = [\underline{x}, \overline{x}]$. Thus, for this x, we have $s_n \ge x - \delta(S)$. Since $x \ge \underline{x}$ this implies that $s_n \ge \underline{x} - \delta(S)$. Here, $\inf S \ge s_n - 2^{-n}$, therefore $\inf S \ge \underline{S} - \delta(S) - 2^{-n}$. In the limit $n \to \infty$, we get the desired inequality $\inf S \ge \underline{x} - \delta(S)$.

The proposition is proven.

Acknowledgments

This work was supported in part by the National Science Foundation grants 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.

The authors are thankful to all the participants of the 2016 IEEE World Congress on Computational Intelligence (Vancouver, Canada, July 24–29, 2016) for valuable suggestions.

References

[1] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*, Springer, London, 2001.

[2] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.

[3] V. Kreinovich and S. Shary, "Interval methods for data fitting under uncertainty: a probabilistic treatment", *Reliable Computing*, 2016, Vol. 23, pp. 105–141.

[4] R. E. Moore, R.B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.

[5] G. Ruiz, H. Pomares, I. Rojas, and H. Hagras, "Towards general forms of interval type-2 fuzzy logic systems", *Proceedings of the IEEE International Conference on Fuzzy Systems FUZZ-IEEE'2016*, Vancouver, Canada, July 24–29, 2016.

[6] S. P. Shary, "Solvability of interval linear equations and data analysis under uncertainty", *Automation and Remote Control*, 2012, Vol. 73, No. 2, pp. 310–322.

[7] S. P. Shary, "Maximum consistency method for data fitting under interval uncertainty", *Journal of Global Optimization*, published online July 18, 2015, pp. 1–16.