Operations with Fuzzy Numbers Explain Heuristic Methods in Image Processing

O. Kosheleva
Department of Electrical
and Computer Engineering
University of Texas
El Paso, TX 79968, USA
olga@ece.utep.edu

V. Kreinovich
Dept. of Computer Science
University of Texas
El Paso, TX 79968, USA
vladik@cs.utep.edu

B. Bouchon-Meunier
LIP6, UPMC, Case 169
4 place Jussieu
75252 Paris Cédex 05
France
bouchon@poleia.lip6.fr

Abstract

Maximum entropy method and its heuristic generalizations are very useful in image processing. In this paper, we show that the use of fuzzy numbers enables us to naturally explain these heuristic methods.

1 Introduction to the problem

1.1 The main objective of image processing

In many problems of image processing, e.g., in optical and radio astronomy, we want to reconstruct the *image*. In precise terms, we want to know the function $I(\vec{\sigma})$ that describes how brightness depends on the coordinates $\vec{\sigma} = (x, y)$.

Most of the time, we are only interested in the values of brightness over a sufficiently dense rectangular *grid* that is formed by the points $\vec{\sigma}_{ij} = (x_i, y_j)$ with coordinates $x_i = x_0 + i \cdot h_x$ and $y_j = y_0 + j \cdot h_y$. The corresponding values $I(\vec{\sigma}_{ij})$ of the brightness function form a *matrix*; the components of this matrix are, for simplicity, usually denoted as I_{ij}.

1.2 The main problem of image processing: general formulation

Real-life measurements are usually inaccurate and incomplete. Due to these inaccuracy and incompleteness, there are usually *many* different images which are consistent with the same set of observations. One of the main problems of image processing is, thus, to select the *best* (most appropriate) image among all images which are consistent with the known measurement results.

1.3 Examples

The following two simple examples illustrate the above two points:

- that image reconstruction is not unique, and
- that, often, some images, which are consistent with the measurement results, are reasonable for this particular imaging problem, while other images, which are also consistent with the same measurements, are not reasonable.

The first example is from *astronomy*; it deals with the simplest possible source, that consist of only one bright point. An arbitrary not-100%-accurate observation of this single point source is consistent not only with the corresponding one-point image, but also with an image, in which, in addition to this point source, there is a very weak second source elsewhere, so weak that, due to measurement inaccuracy, its brightness is indistinguishable from 0. Since there is no physical or observational reason for this second point, we would definitely prefer the one-point image.

The second example is from *mammography*. The main objective of mammography is to detect possible nonhomogeneities that can be indications of a growing tumor. These nonhomogeneities appear as spots on the image. The idea is that if such a spot appears, then more accurate (more costly, and often painful) testing is done to check whether it actually is a malignant tumor. If the actual tissue does not contain any nonhomogeneities, then, due to the inevitable measurement inaccuracy, the measurement results are still consistent with the assumption that we have a very small tumor. However, if we always return an image with a spot, then we would have to use the more costly and painful procedure always, and this defeats the main purpose of mammography as a "gateway" painless and cheap test. It is therefore desirable to return an image with a spot only if there are indeed reasonable grounds for suspecting that there may be an nonhomogeneity in the actual tissue.

1.4 Best image: in what sense?

We want to select, from all the images that are consistent with the given observations, the most appropriate (best) image. How can we do that?

In every application area, there are experts who know which images are more appropriate. So, in principle, if we only had a few possible images to compare, we could simply ask an expert to compare them. However, for each set of measurement results, there are very many possible images that are consistent with theses results, and therefore, we cannot ask an expert to compare all of them, we must do this selection *automatically*. Thus, we need to formulate the expert's preferences in precise terms that a computer will be able to understand.

Situations in which expert themselves are able to formulate their preferences in precise terms are extremely rare, and definitely in both of our examples (astronomical and medical images) experts have not been able to provide us with such a formalization. Therefore, we must either somehow extract this preference relation from the experts, or use heuristic methods and hope that they will be in good accordance with the expert preferences.

1.5 Maximum entropy and generalized maximum entropy as useful heuristic methods

Among successful heuristic methods that are used to reconstruct the image are:

- the *maximum entropy* method, according to which we select, from all the images that are consistent with the observations, the image $I(\vec{\sigma})$ for which the entropy if the largest possible:

$$-\int I(\vec{\sigma})\cdot\log(I(\vec{\sigma}))\,d\vec{\sigma}\to\max; \qquad (1)$$

  and

- *generalized maximum entropy methods*, in which we choose the image $I(\vec{\sigma})$ for which

$$\int F(I(\vec{\sigma}))\,d\vec{\sigma}\to\max \qquad (2)$$

  for some function $F(x)$ (most frequently, $F(x)=\log_2(x)$ and $F(x)=|x|^p$ for some real number p).

In [1], [8], and [2], we have successfully used these methods for *radar imaging* (including *planetary radar imaging*).

1.6 Why maximum entropy?

These methods often work well, so the natural question is: why? This question is caused not only by natural scientific curiosity, but we also want to know:

- whether this success is caused by the specifics of the problems to which these methods have so far been applied, and these methods may fail for other problems,
- or these methods are universally valid, and therefore, we can fearlessly use them in future applications as well.

There exist interesting and convincing probabilistic justifications of the maximum entropy method based on reasonable prior distributions (see, e.g., [13]), but this justification is not that naturally applicable to *generalized* entropy methods.

It is therefore desirable to either find a new justification of these methods, or, if these methods cannot be naturally justified, provide justifiable working methods.

1.7 What we are planning to do

In this paper, we show that, if we use fuzzy numbers to describe the informal ideas behind image reconstruction, then we get exactly maximum entropy and generalized maximum entropy techniques.

Comments.

- Our result complements a similar conclusion about *data* processing obtained in [9].
- An alternative justification of generalized entropy methods is described in [5], [6], [4].

2 Justification of generalized maximum entropy

2.1 Image reconstruction: main idea

The main reason why we need to use computers to reconstruct images is the presence of noise. Even in the absence of any actual image, when $I(\vec{\sigma})=0$, due to the noise, sensors do sense some signal. Therefore:

- If the observations are consistent with the absence of image, then it makes sense to simply return 0.
- Accordingly, it makes sense to only return non-0 brightness at a certain point $\vec{\sigma}$ only if the observations are inconsistent with the brightness being 0, i.e., only if the observations *imply* that the brightness is non-zero.
- If the observations imply that the brightness is non-zero, then we should return only the brightness that is definitely implied by the observations.

Let us clarify these conclusions. If the measured brightness value is $\tilde{I}$, and the accuracy of brightness measurement is $\Delta>0$, this means that the actual brightness I can take an arbitrary value from the interval $[\tilde{I}-\Delta,\tilde{I}+\Delta]$.

- If this interval contains 0, then we should return 0 as a reconstructed brightness value.
- If this interval $[\tilde{I}-\Delta,\tilde{I}+\Delta]$ does not contain 0, i.e., if $\tilde{I}-\Delta>0$, then we should return, as a reconstructed brightness value, the smallest possible value of brightness from this interval, i.e., $\tilde{I}-\Delta$.

2.2 Image reconstruction: main problem

If the image consisted of only one pixel, i.e., if the brightness distribution consisted of only one number I_{ij}, then the above idea would lead to the desired solution: we pick the smallest possible value of I_{ij} that is consistent with the observations.

In reality, there are many pixels, and so, we would like all of them to be the smallest.

At first, it may seem natural to formalize this requirement literally: that we must choose the reconstructed image for which *all* the brightnesses take the smallest possible value. Unfortunately, this approach does not work: e.g., we can find the image for which I_{11} takes the smallest value, but for that image, in general, e.g., I_{12} will not take the smallest possible value.

So, instead of requiring that all the brightnesses take the *smallest* possible value, we can only require that these brightnesses are all *small*, and that this condition (that they are all small) should be satisfied to the largest possible degree.

2.3 Image reconstruction: informal solution

In other words, we require that *all* the brightnesses are small, i.e., that I_{11} is small, and I_{12} is small, and ...

How can we formalize this requirement?

2.4 Fuzzy numbers and fuzzy logic: a natural formalization of the above idea

A natural way to formalize this requirement is to use *fuzzy logic*.

Let $\mu(x)$ be a membership function that describes the natural-language term "small". (Intuitively, $\mu(x)$ should be a fuzzy number.) Then,

- our degree of belief that I_{11} is small is equal to $\mu(I_{11})$;
- our degree of belief that I_{12} is small is equal to $\mu(I_{12})$,
- etc.

To get the degree of belief d that all conditions are satisfied, we must use a t-norm (a fuzzy analogue of "and"), i.e., use a formula $d = \mu(I_{11})\&\mu(I_{12})\&\ldots$, where & is this t-norm.

2.5 This formalization leads to generalized maximum entropy methods: discrete case

In [10], we have shown that within an arbitrary accuracy, an arbitrary t-norm can be approximated by a strictly Archimedean t-norm[1]. Therefore, for all practical purposes, we can assume that the t-norm that describes the experts' reasoning, is strictly Archimedean and therefore, has the form $a\&b = \varphi^{-1}(\varphi(a)+\varphi(b))$ for some strictly decreasing function φ [3], [11]. Thus, $d(I) = \varphi^{-1}(\varphi(\mu(I_{11})) + \varphi(\mu(I_{12})) + \ldots)$.

We want to find the image I for which our degree of belief $d(I)$ (that the image is good) is the largest possible. Since the function φ is strictly decreasing, $d(I)$ attains its maximum if and only if the auxiliary characteristic $D(I) = \varphi(d(I))$ attains its minimum. From the formula that describe $d(I)$, we can conclude that $D(I) = \varphi(\mu(I_{11}))+\varphi(\mu(I_{12}))+\ldots$ Thus, the condition $D(I) \to \min$ takes the form $F(I_{11}) + F(I_{12}) + \ldots \to \min$, with $F(x) = \varphi(\mu(x))$.

2.6 Generalized entropy methods: continuous case

So far, we were analyzing the problem of how to compare different pixel-by-pixel images. In real-life, the object whose image we want to describe is continuous, pixels are simply a useful approximation. It is, therefore, desirable to reconstruct not just the values on a grid, but also the entire brightness distribution, i.e., the values of $I(\vec{\sigma})$ for every point $\vec{\sigma}$. To achieve this goal, we must be able to compare the quality of different possible reconstructed images, i.e., of different functions $I(\vec{\sigma})$.

The denser the pixels (i.e., the smaller the distances h_x and h_y between the neighboring pixels), the closer the pixel-by-pixel image to the continuous one. Therefore, as a characteristic $D(I)$ of a function $I(\vec{\sigma})$, we can take the *limit* of the utilities of its pixel-by-pixel representation as $h_x \to 0$ and $h_y \to 0$.

This limit is easy to describe because the sum $D(I)$ is, in effect, an integral sum, and therefore, as the pixels get denser, this sum tends to the integral $D(I) = \int F(I(\vec{\sigma}))\,d\vec{\sigma}$.

3 Conclusions

3.1 Main conclusion

So, we have indeed justified the generalized entropy method.

3.2 An important auxiliary conclusion

This justification enables us not only to explain why generalized entropy methods are useful and successful, but it also enables us to answer an important auxiliary question: which function $F(x)$ should we choose.

The answer is: We should base this choice on the opinion of the experts. Namely, from these experts, we extract:

[1] This result may be of independent interest, and therefore, we present its exact formulation and the idea of the proof in the special Appendix.

- the membership function $\mu(x)$ that corresponds to "small", and
- the function $\varphi(x)$ that best describes the experts' "and",

and then choose $F(x) = \varphi(\mu(x))$.

Acknowledgments.

This work was supported in part by NASA under cooperative agreement NCCW-0089, by NSF grants No. DUE-9750858, EEC-9322370, and CDA-9522207, and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

References

[1] B. C. Flores, A. Ugarte, and V. Kreinovich, Choice of an entropy-like function for range-Doppler processing. In: *Proceedings of the SPIE/International Society for Optical Engineering*, 1993, Vol. 1960, *Automatic Object Recognition III*, pp. 47–56.

[2] B. C. Flores, V. Kreinovich, and R. Vasquez, Signal design for radar imaging and radar astronomy: genetic optimization, In: D. Dasgupta and Z. Michalewicz, eds., *Evolutionary Algorithms in Engineering Applications*, Springer-Verlag, Berlin-Heidelberg, 1997, pp. 406–423.

[3] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, NJ, 1995.

[4] O. Kosheleva, Symmetry-group justification of maximum entropy method and generalized maximum entropy methods in image processing, In: G. Erickson (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1997 (to appear).

[5] O. M. Kosheleva and V. Kreinovich, A letter on maximum entropy method, *Nature*, 1979, Vol. 281, No. 5733 (Oct. 25), pp. 708–709

[6] O. M. Kosheleva and V. Kreinovich, *Utility functions that describe invariant preferences*, Technical Report, Center for New Information Technology "Informatika", Leningrad, 1989 (in Russian).

[7] C. H. Ling, Representation of associative functions, *Publ. Math. Debrecen,* 1965, Vol. 12, pp. 189–212.

[8] J. L. Mora, B. C. Flores, and V. Kreinovich, Suboptimum binary phase code search using a genetic algorithm. In: S. D. Udpa and H. C. Han, eds., *Advanced Microwave and Millimeter-Wave Detectors.* Proceedings of the SPIE/International Society for Optical Engineering, Vol. 2275, San Diego, CA, 1994, pp. 168–176.

[9] H. T. Nguyen, V. Kreinovich, and B. Bouchon-Meunier, Soft Computing Explains Heuristic Numerical Methods in Data Processing and in Logic Programming, *Working Notes of the AAAI Symposium on Frontiers in Soft Computing and Decision Systems*, Boston, MA, November 8–10, 1997, pp. 40–45; detailed version to appear in *Heuristics: The Journal of Intelligent Technology.*

[10] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, Strict Archimedean t-Norms and t-Conorms as Universal Approximators, *International Journal of Approximate Reasoning*, 1997 (to appear; a preliminary version of this paper can be downloaded via a link in the second author's homepage http://cs.utep.edu/csdept/faulty/kreinovich.html).

[11] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic,* CRC Press, Boca Raton, Florida, 1997.

[12] B. Schweizer and A. Sklar, Associative functions and abstracts semigroups, *Publ. Math. Debrecen,* 1963, Vol. 10, pp. 69–81.

[13] J. Skilling, In: G. Erickson (ed.), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1997 (to appear).

4 Appendix

4.1 Definitions

Definition 1. (see, e.g., [3, 11]) *A function $f_\&: [0,1]\times[0,1]\to[0,1]$ is called a t-norm if it satisfies the following four conditions:*

- *$f_\&(1,a) = a$ for all a;*
- *$f_\&(a,b) = f_\&(b,a)$ for all a and b;*
- *$f_\&(a, f_\&(b,c)) = f_\&(f_\&(a,b),c)$ for all a, b, and c;*
- *if $a \le a'$ and $b \le b'$, then $f_\&(a,b) \le f_\&(a',b')$.*

Definition 2. (see, e.g., [3, 11]) *A function $f_\vee : [0,1]\times[0,1]\to[0,1]$ is called a t-conorm if it satisfies the following four conditions:*

- *$f_\vee(1,a) = a$ for all a;*
- *$f_\vee(a,b) = f_\vee(b,a)$ for all a and b;*
- *$f_\vee(a, f_\vee(b,c)) = f_\vee(f_\vee(a,b),c)$ for all a, b, and c;*
- *if $a \le a'$ and $b \le b'$, then $f_\vee(a,b) \le f_\vee(a',b')$.*

It is also usually required that a t-norm and a t-conorm are *continuous* functions.

Of all possible continuous t-norms and t-conorms, the most widely used are the *idempotent* operations

$f_{\&}(a,b) = \min(a,b)$ and $f_{\vee}(a,b) = \max(a,b)$ and *Archimedean* t-norms and t-conorms that are defined as follows:

Definition 3. [3, 11]

- *A t-norm $f_{\&}(a,b)$ is called Archimedean if it is continuous and $f_{\&}(a,a) < a$ for all $a \in (0,1)$.*
- *An Archimedean t-norm is called strictly Archimedean if it is strictly increasing in each variable for $a, b \in (0,1)$.*

Definition 4. [3, 11]

- *A t-conorm $f_{\vee}(a,b)$ is called Archimedean if it is continuous and $f_{\vee}(a,a) > a$ for all $a \in (0,1)$.*
- *An Archimedean t-norm is called strictly Archimedean if it is strictly increasing in each variable for $a, b \in (0,1)$.*

Strictly Archimedean t-norms and t-conorms are easy to represent:

Proposition 1. [12, 7, 3, 11]

- *For every continuous strictly increasing function $\psi : [0,1] \to [0,1]$, the function $f_{\&}(a,b) = \psi^{-1}(\psi(a) \cdot \psi(b))$ is a strictly Archimedean t-norm.*
- *If $f_{\&}(a,b)$ is a strictly Archimedean t-norm, then there exists a continuous strictly increasing function $\psi : [0,1] \to [0,1]$ for which $f_{\&}(a,b) = \psi^{-1}(\psi(a) \cdot \psi(b))$.*

A similar representation exists for strictly Archimedean t-conorms.

4.2 Main results

Definition 5. *We say that two functions $f(a,b)$ and $f'(a,b)$ are $\varepsilon-$close if for every a and b, we have $|f(a,b) - f(a,b)| \le \varepsilon$.*

Theorem 1. *For every continuous t-norm $f_{\&}$, and for every $\varepsilon > 0$, there exists a strictly Archimedean t-norm $f'_{\&}$ that is $\varepsilon-$close to $f_{\&}$.*

Theorem 2. *For every continuous t-conorm $f_{\vee}$, and for every $\varepsilon > 0$, there exists a strictly Archimedean t-norm $f'_{\vee}$ that is $\varepsilon-$close to $f_{\vee}$.*

Comment. Since the real data always come with some accuracy, these results mean that whatever empirical data we have about the actual expert's use of "and" and "or", and however accurate these data are, these data can always be explained within an assumption that both the "and"-operation (t-norm) and the "or"-operation (t-conorm) are strictly Archimedean.

4.3 General idea of the proof

The proof of Theorems 1 and 2 is based on the *classification theorem* for t-norms and t-conorms that was first proven in [7]. According to this theorem, for every t-norm $f_{\&}(a,b)$, on the interval $[0,1]$, there exists finitely or countably many (possibly none) non-intersecting intervals I_α such that:

- on each of these intervals I_α, $f_{\&}(a,b)$ is:
  - either isomorphic to $a \cdot b$, i.e., has the form $\psi^{-1}(\psi(a) \cdot \psi(b))$ for some strictly increasing function ψ),
  - or isomorphic to $\max(a+b-1, 0)$, i.e., has the form
  $$\psi^{-1}(\max(\psi(a) + \psi(b) - 1, 0))$$
  for some strictly increasing function ψ;
- if a and b do not belong to the same interval I_α, or if one of the values a, b does not belong to any of the intervals I_α at all, then $f_{\&}(a,b) = \min(a,b)$.

Comment. In particular, if we have no intervals at all, we get a t-norm $f_{\&}(a,b) = \min(a,b)$; to get a t-norm $f_{\&}(a,b) = a \cdot b$, we must take the entire interval $[0,1]$ as the only interval I_α.

A similar classification theorem for t-conorms can be easily deduced from the fact that:

- for every t-norm $f_{\&}(a,b)$, its *dual* $f_{\vee}(a,b) = 1 - f_{\&}(1-a, 1-b)$ is a t-conorm; and
- vice versa, for every t-conorm $f_{\vee}(a,b)$, its *dual*
  $$f_{\&}(a,b) = 1 - f_{\vee}(1-a, 1-b)$$
  is a t-norm.

The desired approximation result says that an arbitrary (and arbitrarily complicated) t-norm can be approximated, with an arbitrary accuracy, by a strictly Archimedean t-norm. We will prove this result step-by-step:

- First, we will show that an arbitrary t-norm can be approximated, with an arbitrary accuracy, by a t-norm that has only finitely many intervals.
- Then, we will show that an arbitrary t-norm with finitely many intervals can be approximated, with an arbitrary accuracy, by a t-norm in which these intervals constitute the entire interval $[0,1]$, and in which on each interval, the t-norm is isomorphic to $a \cdot b$.
- Finally, we will show that a t-norm with $k > 1$ intervals on each of which this t-norm is isomorphic to $a \cdot b$, can be approximated, with an arbitrary accuracy, by a t-norm with the same property, but with only $k-1$ intervals. By repeating the last reduction finitely many times, we will finally get an approximating t-norm that has only one interval: $[0,1]$, and that is isomorphic to $a \cdot b$, i.e., that is strictly Archimedean.

If, on each of these three mega-steps, we choose an approximation with an accuracy $\delta = \varepsilon/3$, then after these three steps, we get a t-norm that approximates the original one with the desired accuracy ε.

Similarly, to achieve the accuracy $\varepsilon/3$ on the their megastep, we must, on each substep of this mega-step, take an approximation with an accuracy $\varepsilon/(3N)$, where N is the number of intervals at the beginning of this mega-step.

Comment. It is sufficient to be able to approximate t-norms. Indeed, if we can approximate an arbitrary t-norm $f_\&$ by an ε-close strictly Archimedean t-norm $f'_\&$, then, given an arbitrary t-conorm $f_\vee$, we will be able to approximate its dual $f_\&(a,b) = 1 - f_\vee(1-a, 1-b)$ by an ε-close strictly Archimedean t-norm $f'_\&(a,b)$. One can then easily show that the dual $f'_\vee$ to $f'_\&$ is a strictly Archimedean t-conorm that is ε-close to the original t-conorm $f_\vee(a,b)$ (because two t-conorms are ε-close iff their duals are ε-close, and vice versa).

4.4 Step 1: Reduction to finitely many intervals

Let us show how to approximate an arbitrary t-norm $f_\&$ with an arbitrary accuracy $\delta > 0$, by a t-norm whose classification requires only finitely many intervals.

Indeed, since the intervals I_α that characterize the original t-norm are all located within the interval $[0,1]$, and these intervals do not intersect with each other, the total number of intervals I_α whose length is $\ge \delta$ is finite ($\le 1/\delta$).

We can thus define a new t-norm $f'_\&(a,b)$ as follows:

- if in the characterization of $f_\&$, the numbers a and b belong to the same interval I_α of length $\ge \delta$, then $f'_\&(a,b) = f_\&(a,b)$;
- for all other pairs (a,b), $f'_\&(a,b) = \min(a,b)$.

It is clear that the new t-norm $f'_\&$ can be characterized in the same manner as the original t-norm $f_\&(a,b)$, but with only finitely many intervals I'_α. So, to prove that this first step does do the desired approximation, it is sufficient to show that the new t-norm $f'_\&(a,b)$ is δ-close to the original one, i.e., that $|f'_\&(a,b) - f_\&(a,b)| \le \delta$ for all a and b.

Indeed, the only case when the difference $f'_\&(a,b) - f_\&(a,b)$ is different from 0 (i.e., for which $f'_\&(a,b) \ne f_\&(a,b)$) is when both a and b belong to one of the original intervals $[a^-, a^+]$ of width $a^+ - a^- < \delta$. In this case, $a^- \le f_\&(a,b) \le a^+$. Similarly, $f'_\&(a,b) = \min(a,b)$ also belongs to the interval $[a^-, a^+]$. So, $f_\&(a,b)$ and $f'_\&(a,b)$ are two numbers on the same interval $[a^-, a^+]$ of width $< \delta$. Thus, the difference between these two numbers cannot exceed the width of this interval, and is, therefore $< \delta$.

So, $f_\&$ and $f'_\&$ are, indeed, δ-close. The first part is proven.

4.5 Step 2: Reduction to t-norms that are strictly Archimedean on each interval

Let us start with a t-norm $f_\&$ that has finitely many intervals I_α. Since there are finitely many intervals, the space between and outside these intervals I_α (if there is any space left) is also a union of finitely many intervals, on each of which $f_\&(a,b) = \min(a,b)$. Let us add these new intervals to the intervals I_α that characterize the t-norm $f_\&(a,b)$. When combined, the intervals from this enlarged set $\{J_\alpha\}$ cover the entire interval $[0,1]$.

We will now show that it is possible to approximate the t-norm $f_\&$ by a new t-norm $f'_\&$, with the same (extended) set of intervals $\{J_\alpha\}$, but for which on each of these intervals, the t-norm is isomorphic to $a \cdot b$.

We will approximate the original t-norm interval-by-interval. (This is OK, since the values of the two t-norms that are characterized by the same intervals are only different when both a and b belong to the same interval; otherwise, we have $f_\&(a,b) = f'_\&(a,b) = \min(a,b)$.) These intervals $[a^-, a^+]$ are of two types:

- intervals on which $f_\&(a,b) = \min(a,b)$;
- intervals on which $f_\&(a,b)$ is isomorphic to $\max(a+b-1, 0)$.

Let us show how we can approximate intervals of both types.

First, we reduce a t-norm defined on each interval to a t-norm defined on the interval $[0,1]$. Indeed, there exists an easily computable linear transformation $L(x) = (x - a^-)/(a^+ - a^-)$ that maps the interval $[a^-, a^+]$ onto $[0,1]$:

- if $a \in [a^-, a^+]$, then
$$L(a) = \frac{a - a^-}{a^+ - a^-} \in [0,1];$$
and, vice versa,
- if $A \in [0,1]$, then
$$L^{-1}(A) = a^- + A \cdot (a^+ - a^-) \in [a^-, a^+].$$

Thus:

- if $f_\&(a,b)$ is a t-norm on the interval $[a^-, a^+]$ (i.e., a function $[a^-, a^+] \times [a^-, a^+] \to [a^-, a^+]$), then the operation
$$F_\&(A,B) = L(f_\&(L^{-1}(A), L^{-1}(B))$$
is a t-norm on the interval $[0,1]$; and, vice versa,
- if $F_\&(A,B)$ is a t-norm on the interval $[0,1]$, then the operation $f_\&(a,b) = L^{-1}(F_\&(L(a), L(b))$ is a t-norm on $[a^-, a^+]$.

Hence, if we will be able to approximate the t-norm $F_\&(A,B)$ on the interval $[0,1]$ by a close strictly Archimedean t-norm $F'_{\&}(A,B)$, then the corresponding operation $f'_{\&}(a,b) = L^{-1}(F'_{\&}(L(a),L(b))$ on $[a^-,a^+]$ will be close to the original t-norm.

So, it is sufficient to approximate the t-norm $F_{\&}(A,B)$ defined on the interval $[0,1]$. Depending on whether $f_{\&}$ (and, hence, $F_{\&}$) is isomorphic to min or to

$$\max(A+B-1,0),$$

we get two different approximations:

- The function $F_{\&}(A,B) = \min(A,B)$ can be represented as

  $$\exp(-\max(|\ln(A)|,|\ln(B)|).$$

  Since $\max(x,y) = \lim_{p\to\infty}(x^p+y^p)^{1/p}$, we can, with an arbitrary accuracy, approximate $\min(A,B)$ by

  $$F'_{\&}(A,B) = \exp(-(|\ln(A)|^p + |\ln(B)|^p)^{1/p}).$$

  (this approximation was proposed by B. Schweizer and A. Sklar in [12]). This new function is isomorphic to $A\cdot B$, with the isomorphism given by a function $\psi(A) = \exp(-|\ln(A)|^p)$. The larger p, the better the approximation. So, for sufficiently large p, we can get an arbitrarily close approximation.
- For operations that are isomorphic to

  $$\max(A+B-1,0),$$

  it is somewhat easier to describe an approximating t-norm by describing a dual approximation: to the dual t-conorm that is isomorphic to $N(A,B) = \min(A+B,1)$.
  Isomorphic means that we have a function

  $$\psi:[0,1]\to[0,1]$$

  that implements the desired isomorphism, i.e., for which,

  $$F_{\&}(A,B) = \psi^{-1}(N(\psi(A),\psi(B))) =$$
  $$\psi^{-1}(\min(\psi(A)+\psi(B),1)).$$

  It is easy to see that if we find a sequence $N_n(A,B)$ of strictly Archimedean t-norms that tend to $N(A,B)$ (in the uniform metric), then the corresponding isomorphic operations $\psi^{-1}(N_n(\psi(A),\psi(B)))$ will tend to $\psi^{-1}(N(\psi(A),\psi(B))) = F_{\&}(A,B)$. Thus, to be able to approximate an arbitrary t-norm that is isomorphic to N, it is sufficient to be able to approximate $N(A,B)$ itself.
  This can be done as follows: we choose $\alpha\to 0$, and approximate $N(a,b)$ by a strict Archimedean operation $G^{-1}(G(A)+G(B))$, where

  $$G(A) = \frac{A}{1-\alpha}$$

  for $A\le 1-\alpha$ and

  $$G(A) = 1-\alpha+\frac{\alpha}{1-A}$$

  for $A\ge 1-\alpha$. This operation coincides with $\min(A+B,1)$ when $A+B\le 1-\alpha$, and leads to the results between $1-\alpha$ and 1 when $A+B\ge 1-\alpha$. Thus, when $\alpha\to 0$, this operation tends to $N(A,B)$.
  From this approximation of a dual operation, we can easily obtain the approximation of the original t-norm.

Step 2 is proven.

4.6 Step 3: Reduction to a t-norm with one fewer interval

We want to get a reduction from a t-norm that has k intervals to a t-norm that has $k-1$ intervals. To achieve this goal, it is sufficient to show that a t-norm that has two intervals can be approximated by a t-norm that has only one interval. By using this construction, we will be able to "merge" the two neighboring intervals and thus, reduce the number of intervals by one.

Let us consider the case when on two neighboring intervals, we have strictly Archimedean operations. Similarly to Step 2, we can prove that it is sufficient to consider the case when these two intervals form the interval $[0,1]$, i.e., when the first interval is $[0,p]$ and the second interval is $[p,1]$ for some boundary point $p\in(0,1)$.

Since the operation $f_{\&}$ is strictly Archimedean on both subintervals, it is isomorphic to $a\cdot b$ on both of them. In other words, there exist functions $\psi_1:[0,p]\to[0,1]$ and $\psi_2:[p,1]\to[0,1]$ such that for a,b from the first interval $[0,p]$, we have $f_{\&}(a,b) = \psi_1^{-1}(\psi_1(a)\cdot\psi_1(b))$, while for a and b from the second interval $[p,1]$, we have $f_{\&}(a,b) = \psi_2^{-1}(\psi_2(a)\cdot\psi_2(b))$.

We want to "merge" these two representations into a single formula that is close to the original two-part operation. For that merger, we will take into consideration the fact that a function ψ_i is not uniquely determined by the t-norm $f_{\&}$: the same t-norm can be obtained if we use a function $\psi'_i(x) = (\psi_i(x))^{r_i}$ for any positive real number r_i.

When $r_i\to\infty$, we have $(\psi_i(x))^{r_i}\to 0$; when $r_i\to 0$, we have $(\psi_i(x))^{r_i}\to 1$. Thus, to achieve a merger, we choose r_1 large enough so that $(\psi_1(x))^{r_1}<\delta$ for all $x\in[0,p-\delta]$ for some small δ, and we choose r_2 small enough so that $(\psi_2(x))^{r_2}>1-\delta$ for all $x\in[p+\delta,1]$.

Then, we take a monotonic function $\psi(x)$ that is:

- equal to $(\psi_1(x))^{r_1}$ for $x\in[0,p-\delta]$,
- equal to $(\psi_2(x))^{r_2}$ for $x\in[p+\delta,1]$, and
- linear on the remaining small interval $[p-\delta,p+\delta]$,

and define the new operation

$$f'_{\&}(a,b) = \psi^{-1}(\psi(a)\cdot\psi(b)).$$

From the construction of $f'_{\&}$, one can see that:

- If a and b belong to the same half-interval, then the new t-norm coincides with the old one everywhere except for the narrow interval $[p-\delta, p+\delta]$, in which case it is δ-close to the old one.
- If a belong to the first interval, and b belongs to the second one, then we have $f'_\&(a, p-\delta) = f_\&(a, p-\delta) \leq f'_\&(a, b) \leq a$. When $\delta \rightarrow 0$, we have $f_\&(a, p-\delta) \rightarrow f_\&(a, p) = a$, hence, $f'_\&(a, b) \rightarrow a = \min(a, b) = f_\&(a, b)$. This convergence is uniform in a and b, so, for sufficiently small $\delta > 0$, the resulting t-norm $f'_\&(a, b)$ is indeed close to the original t-norm $f_\&(a, b)$.

Step 3 is proven, and so is the theorem.