



# Mulsetions and Intervalions: Multiset Generalizations of Functions

Luciano da Fontoura Costa

## ► To cite this version:

Luciano da Fontoura Costa. Mulsetions and Intervalions: Multiset Generalizations of Functions. 2023. ⟨hal-04201139v2⟩

**HAL Id: hal-04201139**

**<https://hal.science/hal-04201139v2>**

Preprint submitted on 11 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

# *Mulsetions and Interventions:* Multiset Generalizations of Functions

Luciano da F. Costa  
*luciano@ifsc.usp.br*

*São Carlos Institute of Physics – DFCM/USP*

1st Sep. 2023

## Abstract

Real-valued functions, arguably the main staple of mathematics, are characterized by mapping each domain scalar value into a single image scalar value. Sets represent another key concept in mathematics, corresponding to collections of elements appearing only once, in any order. Multisets are generalized sets allowing elements to appear more than once. The present work aims at generalizing functions as mappings from scalar domain values into single respective multisets, yielding mulsetions. The case in which a map is established from scalar values to intervals is shown to be a specific case of mulsetions, giving rise to respective intervalions. In addition to developing the basic theory of mulsetions and intervalions, the present work also addresses how these concepts underlie similarity indices, such as the Jaccard operator, generalized for the comparison between two nonzero multisets with elements possibly having negative multiplicities. This extension is here achieved by employing a special complex paired representation of multisets in which the positive and negative multiplicities are kept and handled independently one another. Some graphical examples of the presented concepts and methods are also provided.

## 1 Introduction

*Functions* and *sets* are arguably among the most ubiquitous central concepts in mathematics, reflecting their ability to model (e.g. [1]) a wide range of abstract and real-world structures and actions.

Introduced more recently, the concept of *multiset* (e.g. [2, 3, 4, 5, 6, 7]) provides an interesting generalization of traditional sets in the sense of allowing elements to appear several times, and not only once as required by sets.

Multisets are particularly interesting in the sense that they can be used to represent virtually any mathematical structure, including not only sets, but also vectors, functions, matrices, and graphs [8, 9, 10, 11]), so that these structures can be manipulated by analogues of set operations including union and intersection. As a consequence, multisets have been especially important in paving the way to extending *similarity indices* as means to compare, in terms of similarity, two generic mathematical structures.

The extension of several similarity indices, including the Jaccard operation, to non-zero vectors and functions containing non-negative multiplicities is straightforward, involving the use of maximum between multiplicities as a

means to implement multiset union and minimum in the case of intersection. However, the generalization of these operations and indices to generic-real valued multiplicities (possibly including negative multiplicities) requires additional manipulation of the multisets, as considered in [8, 9, 10, 11]. Related operations have also been described in the context of 1-norms [12, 13, 14].

The present work has two main objectives. First, it describes a further generalization of multisets in the sense that each of their elements can have one or two multiplicities, corresponding to positive and/or negative values, which are treated independently and kept separately. As it will be shown, this approach allows a simple definition of the union and intersection operations in the case of generic real-valued multisets, therefore paving the way to obtaining the Jaccard similarity index for comparison of two non-zero multisets with real-valued multiplicities. Second, we describe *mulsetions*, which are mappings taking real-valued scalars into multiset images. The particular case in which these multisets are intervals implies a particular case of mulsetions, which are here called *intervalions*. Being composed of multisets, mulsetions and intervalions can be readily combined in terms of respective union and intersection operations.

This work starts by reviewing the basic concepts of

sets, functions, and multisets, and then proceeds to defining a complex representation of multisets with possibly negative multiplicities, the concepts of mulsetions and intervals, and the generalization of the Jaccard index to multisets, mulsetions, and intervals, including some graphical case-examples.

## 2 Basic Concepts

A *set* of elements is a respective collection where each element can appear only once, in any order. Examples of *numeric* sets include:

$$A = \{-1, 0, 3, 7\} \quad (1)$$

$$B = [-1, 3] \quad (2)$$

$$C = A \cup B = \{-1, 0, 3, 7\} \cup [-1, 3] = \{7\} \cup [-1, 3] \quad (3)$$

Sets  $A$  and  $B$  are respectively said to be *discrete* and *continuous*, while set  $C$  can be said to be *hybrid*. Set  $B$  illustrates that intervals are a particular case of a sets (continuous).

A *real-valued function*  $f(x)$  maps from every value  $x$  in a given *domain*  $\mathcal{D} \in \mathbb{R}$  into a single respective scalar  $y$  in an *image* (or *range*) set  $I \in \mathcal{C} \subset \mathbb{R}$ :

$$f : x \in \mathcal{D} \subset \mathbb{R} \longrightarrow y = f(x) \in I \subset \mathcal{C} \subset \mathbb{R} \quad (4)$$

where the set  $\mathcal{C}$  is the *codomain* of  $f(x)$ .

A *multiset* (e.g. [2, 3, 4, 5, 6, 7]) is a set where any element can appear more than once. The number of times a given element  $i$  appears is called its respective multiplicity  $m_i$ .

It is possible to generalize multisets [8, 9, 10, 11] to any mathematical structure, including functions. In the latter case, given a function as in 25, we would have the following respective multiset representation:

$$f(x) = \{\llbracket x, f(x) \rrbracket\} \quad (5)$$

with  $x \in \mathcal{D} \subset \mathbb{R}$  and  $f(x) \in I \subset \mathcal{C} \subset \mathbb{R}$ .

Examples of multisets include:

$$A = \{-1, 0, 2, 2, 2\} = \{\llbracket 0, 1 \rrbracket; \llbracket 2, 3 \rrbracket\} \quad (6)$$

$$B = \{-1, 0, 1, 1, 1, 3\} = \{\llbracket 0, 1 \rrbracket; \llbracket 1, 3 \rrbracket; \llbracket 3, 1 \rrbracket\} \quad (7)$$

$$C = \{\llbracket x, x^2 \rrbracket\}, \quad x \in [-1, 1] \quad (8)$$

$$D = \{-1, 0, 1, 2, 3, 4\} \quad (9)$$

$$E = [-1, 3] = \{\llbracket x, 1 \rrbracket\}, \quad x \in [-1, 3] \quad (10)$$

Observe the representation of multisets adopted in 6, where  $\llbracket i, m_{i,A} \rrbracket$  is a tuple representing the element  $i$  and its respective multiplicity  $m_{i,A}$  in that same multiset. Example  $C$  can be understood as a piece of the function  $f(x) = x^2$  taken along the interval  $[-1, 1]$ . The examples

in 9 and 10 illustrate that traditional sets are particular cases of multisets.

The set  $S_A$  of the elements of a multiset  $A$  is henceforth said to be its *support*. Therefore, in the case of the above examples, we have:

$$S_A = \{-1, 0, 2\} \quad (11)$$

$$S_B = \{-1, 0, 1, 3\} \quad (12)$$

$$S_C = [-1, 1] \quad (13)$$

$$S_D = \{-1, 0, 1, 2, 3, 4\} \quad (14)$$

$$S_E = [-1, 3] \quad (15)$$

Given two multisets  $A$  and  $B$  with *non-negative entries*, their respective *union*  $A \cup B$  will have support  $S_{A \cup B}$  corresponding to the union of the respective original supports. The new multiset is then obtained by taking tuples respective to each element and the maximum between the respective original multiplicities. For instance, in the case of the above multisets  $A$  and  $B$ , we have:

$$\begin{aligned} A \cup B &= \{-1, 0, 1, 1, 1, 2, 2, 2, 3\} = \\ &= \{\llbracket -1, 1 \rrbracket; \llbracket 0, 1 \rrbracket; \llbracket 1, 3 \rrbracket; \llbracket 2, 3 \rrbracket; \llbracket 3, 1 \rrbracket\} \end{aligned}$$

with  $S_{A \cup B} = \{-1, 0, 1, 2, 3\}$ .

The *intersection* between two multisets  $A$  and  $B$  with *non-negative entries* has support  $S_{A \cap B}$  containing the elements resulting from the intersection of the respective supports. The resulting multiset contains the tuples of the elements in the support paired by the respective minimum between the original multiplicities. For instance, in the case of the multisets  $A$  and  $B$  above, we have:

$$A \cap B = \{-1, 0\} = \{\llbracket -1, 1 \rrbracket; \llbracket 0, 1 \rrbracket\}$$

The union and intersection of real-values multisets (i.e. allowing possible negative multiplicities) will be defined in terms of mulsetions in the next section.

## 3 Multisets with Negative Multiplicities

In case a multiset  $A$  contains negative multiplicities, it becomes interesting to represent positive multiplicities as real values, while negative multiplicities are represented in terms of imaginary numbers. For instance, we have the following multisets:

$$A = \{\llbracket -2, -3j \rrbracket; \llbracket -1, 1 \rrbracket; \llbracket 0, 1 \rrbracket; \llbracket 1, -2j \rrbracket; \llbracket 2, 1 \rrbracket\}$$

$$B = \{\llbracket -3, 1 \rrbracket; \llbracket -2, 3 \rrbracket; \llbracket -1, 1 \rrbracket; \llbracket 0, -1j \rrbracket; \llbracket 1, -1j \rrbracket\}$$

where  $j$  is the imaginary number. It should be kept in mind that the use of the above complex representation is henceforth adopted as a means of keeping and handling

the positive and negative multiplicities of each multiset element separate and independently. In this sense, the imaginary number can be thought only as a resource for storing these two values into the same scalar. In this work, no arithmetic operations are supposed to take place involving the imaginary number more directly.

Multisets which elements with all multiplicities being either non-negative or non-positive are henceforth called *single-valued*, being otherwise referred to as *double-valued*.

The introduction of the complex paired representation of the multiplicity of multisets is justified by the fact that, as we shall see, the union between two single-valued multisets can result in a double-valued multiplicity. Hence, the more general situation of double-valued multisets needs to be implemented from the beginning.

The positive multiplicity of an element  $i$  in a multiset  $\mathbf{A}$  will henceforth be expressed as  $m_{i,A}^P$ , while negative multiplicity of that same element will be represented as  $m_{i,A}^N$ . For instance, in the above examples we have:

$$\begin{aligned} m_{-2,A}^N &= -3; \\ m_{-1,A}^P &= 1; \\ m_{-2,B}^N &= 3; \\ m_{-1,B}^P &= 1. \end{aligned}$$

For generality's sake, it is henceforth allowed a same element  $i$  to have *both* positive and negative multiplicities. Therefore, we can write:

$$m_{i,A} = m_{i,A}^P + j m_{i,A}^N \quad (16)$$

Observe that the adoption of the complex paired representation implies the multiplicities effectively to become complex scalar values.

The cardinality of each individual multiplicity can then be expressed as:

$$\|m_{i,A}\| = m_{i,A}^P + |m_{i,A}^N| \quad (17)$$

where the symbols  $|\cdot|$  and  $\|\cdot\|$  are henceforth adopted for expressing absolute value and cardinality, respectively.

The cardinality of a real-valued multiset  $\mathbf{A}$  can now be defined as:

$$\|\mathbf{A}\| = \sum_i \|m_{i,A}\| = \sum_i \{m_{i,A}^P + |m_{i,A}^N|\}, \quad (18)$$

where  $i$  are the elements of the multiset.

The above introduced complex paired representation of multisets with possibly negative entries allows the respective operations of union and intersection to be defined in a simple manner. More specifically, if  $\mathbf{A}$  and  $\mathbf{B}$  are multisets with real entries, their respective *union* can be obtained from the respective complex paired representa-

tions as:

$$\begin{aligned} \mathbf{A} \cup \mathbf{B} &= \\ &= \bigcup_{i=1}^N \{[i, \max(m_{i,A}^P, m_{i,B}^P) + j \min(m_{i,A}^N, m_{i,B}^N)]\}, \end{aligned} \quad (19)$$

where  $i$  are the elements of the respective support.

In the case of the real-valued multisets  $\mathbf{A}$  and  $\mathbf{B}$  above, we have:

$$\begin{aligned} \mathbf{A} \cup \mathbf{B} &= \\ &= \{[-3, 1]; [-2, 2 - 3j]; [-1, 1]; [0, 1 - 1j]; [1, -2j]; [2, 1]\} \end{aligned}$$

The *cardinality* of the union of two multisets  $\mathbf{A}$  and  $\mathbf{B}$  can be expressed as:

$$\begin{aligned} \|\mathbf{A} \cup \mathbf{B}\| &= \sum_i \{\max(m_{i,A}^P, m_{i,B}^P) - \min(m_{i,A}^N, m_{i,B}^N)\} = \\ &= \sum_i \{\max(m_{i,A}^P, m_{i,B}^P) + \max(|m_{i,A}^N|, |m_{i,B}^N|)\} \end{aligned} \quad (20)$$

In case both multisets have all elements with only positive or negative multiplicities – therefore constituting single-valued multisets, the above expression simplifies to:

$$\|\mathbf{A} \cup \mathbf{B}\| = \sum_i \{\max(|m_{i,A}|, |m_{i,B}|)\} \quad (21)$$

where the terms  $|m_{i,A}|$  and  $|m_{i,B}|$  are calculated as expressed in Equation 18.

The intersection between two real-valued multisets  $\mathbf{A}$  and  $\mathbf{B}$  can be expressed as:

$$\begin{aligned} \mathbf{A} \cap \mathbf{B} &= \\ &= \bigcap_{i=1}^N \{[i, \min(m_{i,A}^P, m_{i,B}^P) - j \max(m_{i,A}^N, m_{i,B}^N)]\} \end{aligned} \quad (22)$$

where  $i$  are the elements of the respective support.

The cardinality of the above intersection can be written as:

$$\begin{aligned} \|\mathbf{A} \cap \mathbf{B}\| &= \sum_i \{\min(m_{i,A}^P, m_{i,B}^P) - \max(m_{i,A}^N, m_{i,B}^N)\} = \\ &= \sum_i \{\min(m_{i,A}^P, m_{i,B}^P) + \min(|m_{i,A}^N|, |m_{i,B}^N|)\} \end{aligned} \quad (23)$$

In case both multisets have all elements with only positive or negative multiplicities, the above expression simplifies to:

$$\|\mathbf{A} \cap \mathbf{B}\| = \sum_i \{\min(|m_{i,A}|, |m_{i,B}|)\} \quad (24)$$

## 4 Multisets and Intervalls

A *multisetion* (for *multiset function*)  $\mathbf{f}$  is henceforth understood to correspond to a mapping from a domain set

of scalar values into respective multisets contained in the respective range  $\mathcal{R}$ . More informally, we can write:

$$\mathbf{f} : x \in \mathcal{D} \subset \mathbb{R} \longrightarrow \mathbf{y} = \mathbf{f}(x) \in \mathcal{R} \subset \mathcal{M} \quad (25)$$

where  $\mathcal{M}$  is a specific multiset space.

As an example, let us consider  $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$ . A possible mulsetion is:

$$\mathbf{f} = \begin{cases} 0 \longrightarrow \{\llbracket -2, -3 \rrbracket; \llbracket -1, -3 \rrbracket; \llbracket 0, -3 \rrbracket\} \\ 1 \longrightarrow \{\llbracket -1, -2 \rrbracket; \llbracket 1, -2 \rrbracket\} \\ 2 \longrightarrow \{\llbracket 0, -1 \rrbracket\} \\ 3 \longrightarrow \{\llbracket 1, 1 \rrbracket; \llbracket 2, 1 \rrbracket; \llbracket 3, 1 \rrbracket\} \\ 4 \longrightarrow \{\llbracket 2, 2 \rrbracket; \llbracket 3, 2 \rrbracket; \llbracket 4, 2 \rrbracket\} \end{cases} \quad (26)$$

A particularly interesting specific case of mulsetions consists in *intervalions* (for *interval function*), which take scalars into respective *intervals*  $[g(t), f(t)]$  defined on the Cartesian plane by respective multisets of the type  $\llbracket t, f(t) + j g(t) \rrbracket$ , with  $g(t) \leq 0$  and  $f(t) \geq 0$ . For instance, assuming that  $t \in \mathcal{D} = [0, 4\pi]$ , we could have the following examples of intervalions:

$$\mathbf{f}(t) = \{t \longrightarrow [-|\cos(t)|, |\cos(t)|] \quad (27)$$

$$\mathbf{g}(t) = \{t \longrightarrow [-0.5|\cos(t)|, |\cos(t)|] \quad (28)$$

$$\mathbf{h}(t) = \{t \longrightarrow [0, |\cos(t)|] \quad (29)$$

Figure 1 illustrates the above intervalions (a-c), as well as an additional example (d).

It can be readily shown that two intervalions can be united and intersected in terms of the union and intersection between their respective multisets  $\llbracket t, f(t) + j g(t) \rrbracket$  and  $\llbracket t, h(t) + j k(t) \rrbracket$ , necessarily with  $g(t) \leq 0, f(t) \geq 0, k(t) \leq 0$  and  $h(t) \geq 0$ .

Observe that intervalions sampled along  $t$  can be understood as respective real-valued *multiset vectors*. In addition, more general mappings into intervals (without requiring  $g(t) \leq 0$  and  $f(t) \geq 0$ ) can be achieved by using mulsetions, in which case the union and intersection are applied over all elements of the two compared intervals.

## 5 Multiset Similarity Indices

Similarity indices (e.g. [15, 16, 17, 18, 19, 9]) have been used extensively in several theoretical and practical situations, typically involving the comparison between two given traditional sets. For instance, the Jaccard similarity index (e.g. [20, 8, 18, 21]) between two non-empty sets  $A$  and  $B$  can be defined as follows:

$$\mathcal{J}(A, B) = \frac{\|A \cap B\|}{\|A \cup B\|} \quad (30)$$

with  $\mathcal{J}(A, B) = \mathcal{J}(B, A)$  and  $0 \leq \mathcal{J}(A, B) \leq 1$ .

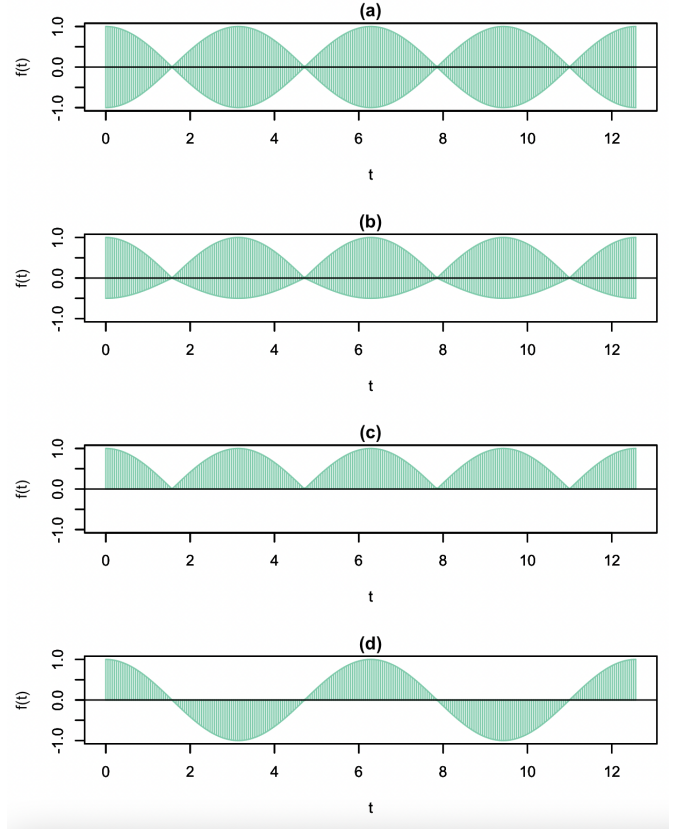


Figure 1: Visualization of the intervalion specified in Eqs. 27 (a), 28 (b), and 29 (c), as well as an additional example (d). Given that any domain point  $t$  is mapped into a whole respective interval, this structure is not a function, but a particular case of multiset. In addition, observe that every image is necessarily composed of a positive and a negative parts, respectively  $f^P(t) \geq 0$  and  $f^N(t) \leq 0$ . The dashed regions of the Cartesian plane indicate respective occupation by the intervalion (with multiplicity 1 along the  $z$  axis, not shown). Though single-valued intervals such as that shown in (d) are completely specified by respective real-valued functions, they correspond to whole respectively occupied regions of the Cartesian space, and not a single curve with infinitesimal width.

This expression can be readily rewritten for the comparison of two real-value multisets by employing Equations 20 and 23 as:

$$\begin{aligned} \mathcal{J}(\mathbf{A}, \mathbf{B}) &= \frac{\|\mathbf{A} \cap \mathbf{B}\|}{\|\mathbf{A} \cup \mathbf{B}\|} = \\ &= \frac{\sum_i \{\min(m_{i,A}^P, m_{i,B}^P) + \min(|m_{i,A}^N|, |m_{i,B}^N|)\}}{\sum_i \{\max(m_{i,A}^P, m_{i,B}^P) + \max(|m_{i,A}^N|, |m_{i,B}^N|)\}} \end{aligned} \quad (31)$$

with  $\mathcal{J}(\mathbf{A}, \mathbf{B}) = \mathcal{J}(\mathbf{B}, \mathbf{A})$  and  $0 \leq \mathcal{J}(\mathbf{A}, \mathbf{B}) \leq 1$ .

In the particular case in which all the elements in both compared multisets have only positive or negative multiplicities, the previous equation can be simplified as:

$$\mathcal{J}(\mathbf{A}, \mathbf{B}) = \frac{\sum_i \{\min(\|m_{i,A}\|, \|m_{i,B}\|)\}}{\sum_i \{\max(\|m_{i,A}\|, \|m_{i,B}\|)\}} \quad (32)$$

The multiset similarity index considered in [8, 9, 22], as well as an operation defined in analogy to 1-norms [13],

can be written as:

$$\mathcal{J}_{\otimes}(A, B) = \frac{\sum_i \{s_{i,AB} \min(|\tilde{m}_{i,A}|, |\tilde{m}_{i,B}|)\}}{\sum_i \{\max(|\tilde{m}_{i,A}|, |\tilde{m}_{i,B}|)\}} \quad (33)$$

where  $\tilde{m}$  indicates real-valued signed multiplicities, not using the complex paired representation, and  $s_{i,AB}$  is the product of the signs of  $\tilde{m}_{i,A}$  and  $\tilde{m}_{i,B}$ . This operator applies to multisets with elements having exclusively negative or positive multiplicities. We have that  $\mathcal{J}_{\otimes}(A, B) = \mathcal{J}_{\otimes}(B, A)$  and  $-1 \leq \mathcal{J}_{\otimes}(A, B) \leq 1$ .

The above operator corresponds to a kind of Jaccard-based cross-correlation, which is here referred to as *cross-Jaccard* similarity operator. Observe that this operator considers as positive the contribution of pairs multiplicities with the same sign, and as negative the contribution of pairs multiplicities with distinct signs. The nature of this operator allows it to be more directly compared to the Pearson correlation coefficient, as discussed in [8].

Observe that the Jaccard and cross-Jaccard similarity indices become identical in case both operands have only non-negative multiplicities. Also shown are the respective multiset union and intersection, as well as the resulting Jaccard and cross-Jaccard indices.

## 6 Some Case-Examples

Figure 2 presents two functions  $f(t)$  and  $g(t)$  with images taking exclusively positive, zero, and negative multiplicities.

As it can be readily inferred from this example, the union between the two original functions corresponds to their total area occupied in the respective Cartesian plane, being analogous to the union operation represented in Venn diagrams. Similarly, the intersection between the functions corresponds to their shared occupied areas. It is of particular interest to observe that the union of two functions can result in an interval, as is the case in the above example. Observe also that the cardinality of an interval can be understood as the area comprised between the curve and the horizontal axis.

Another example of comparing two real-valued functions by using intervals is presented in Figure 3. Smaller values of both Jaccard indices have been obtained as a consequence of these two real-valued functions being less similar than the situation shown in Figure 2.

Observe that the cardinality of an interval can be understood as the sum of the areas comprised between the horizontal axis and the positive and negative multiplicity curves.

A similar example, but now respective to two intervals, is depicted in Figure 4, also including the respectively obtained Jaccard similarity index.

Another example of multiset operations between intervals is presented in Figure 5.

## 7 Concluding Remarks

The present work described several contributions related to multisets and generalized maps. First, multisets having real-values multiplicities (possibly negative) have been represented in terms of a real and imaginary values corresponding to the respective positive and negative multiplicities of each element, understanding that these two types of multiplicities are to be kept and handled independently for each element in the multiset.

In addition to allowing multisets to have elements with both positive and negative multiplicities, the present work has also pursued the possibility to have mulsetions, namely mappings taking scalar values into multisets. Intervals are obtained in the particular case in which these multisets correspond to intervals.

Several important results have been reported and discussed. First, we have that real-valued multisets with possibly negative multiplicities can be more effectively expressed in terms of a complex form in which the positive and negative multiplicities of each of the elements are represented in terms of real and imaginary values, respectively. By keeping these two multiplicities independent one another, this representation allowed the union and intersection of multisets with real-valued multiplicities to be simply and effectively defined.

The complex paired representation of multiplicities also allowed mulsetions and intervals to be defined, corresponding to mappings from real-valued scalars into respective multisets with positive and/or negative multiplicities. As such, these structures can be combined by real-valued multiset union and intersection operations. Of particular importance is the fact that the union of two multisets having only positive or negative multiplicities can result in an interval, therefore justifying the adopted complex paired representation of multisets. The obtained results have also been applied to the derivation of similarity indices respective to multiset with real-values multiplicities.

Though the present work has focused on extending the Jaccard similarity index to real-valued multisets, the concepts and methods introduced here have much broader implications, with possible applications in several theoretical and applied scientific and technological areas, including signal processing, multivariate statistics, differential and integral calculus, to name but a few possibilities.

The reported concepts and results pave the way to a number of further studies. For instance, it would be interesting to obtain generalized multiset versions of other



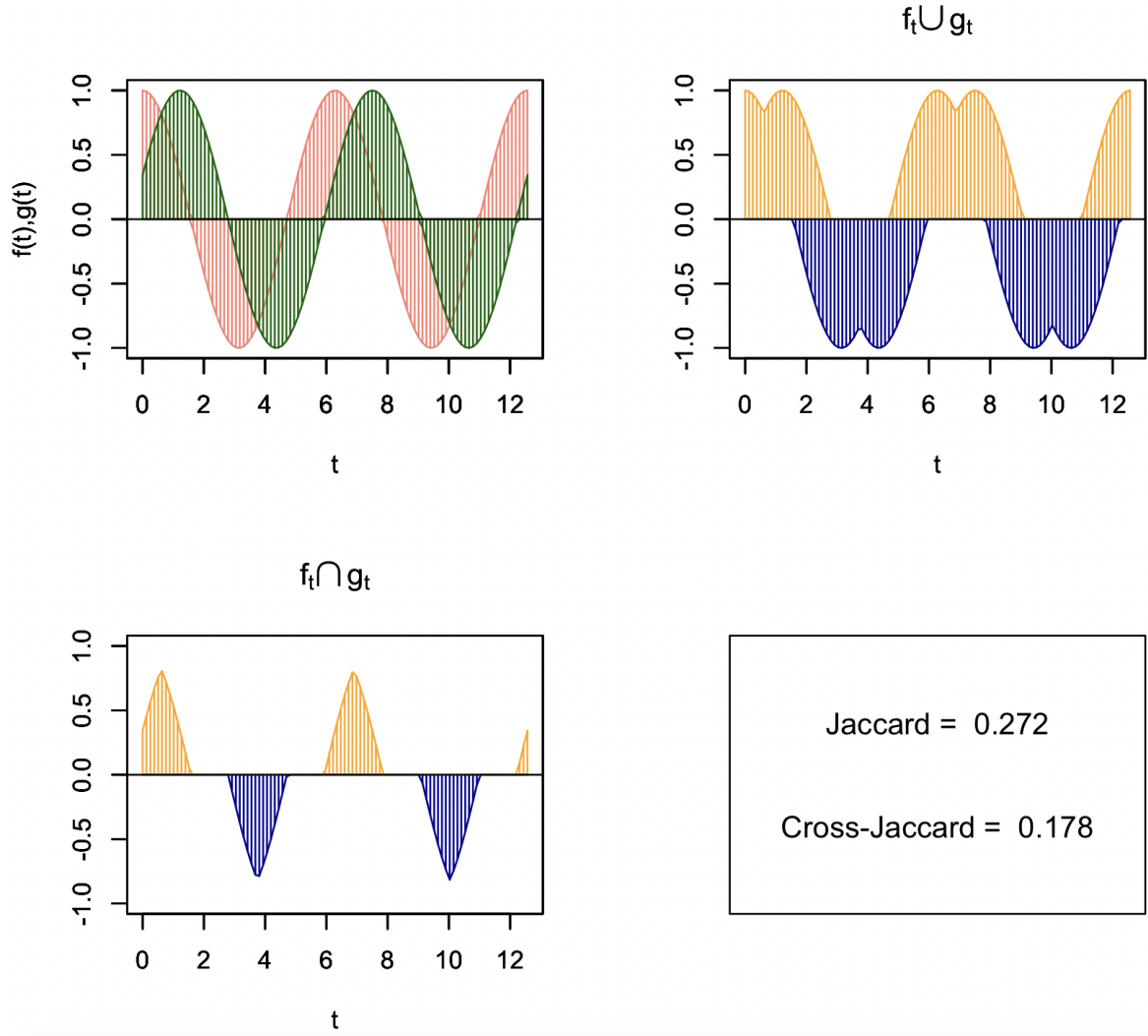


Figure 2: Example of two functions (particular case of real-valued multisets)  $f(t)$  and  $g(t)$ , as well as their union and intersection. The respectively obtained Jaccard and cross-Jaccard similarity indices are also indicated. The portions of the union and intersection results corresponding to the positive and negative multiplicities are shown in orange and blue, respectively. Observe that the union of these two functions is an interval, which justifies the complex paired approach adopted in the present work.

similarity indices. In addition, it would be interesting to compare in a more systematic manner how the Jaccard and cross-Jaccard similarity indices relate one another. Yet another particularly promising possibility is to extend the reported concepts to other multiset operations, including complementation, addition, and subtraction.

## Acknowledgments

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

## Observations

As all other preprints by the author, the present work contains preliminary work subject to further revision and validation. Respective modification, commercial use, or distribution of any of its parts are not possible, as this work has author copyright. Many of the preprints by the author are also available in Hal and/or arXiv. This work can also be cited by using the DOI number or article identification link. It is also interesting to check for new versions of this work, as they may become available. Thanks for reading.

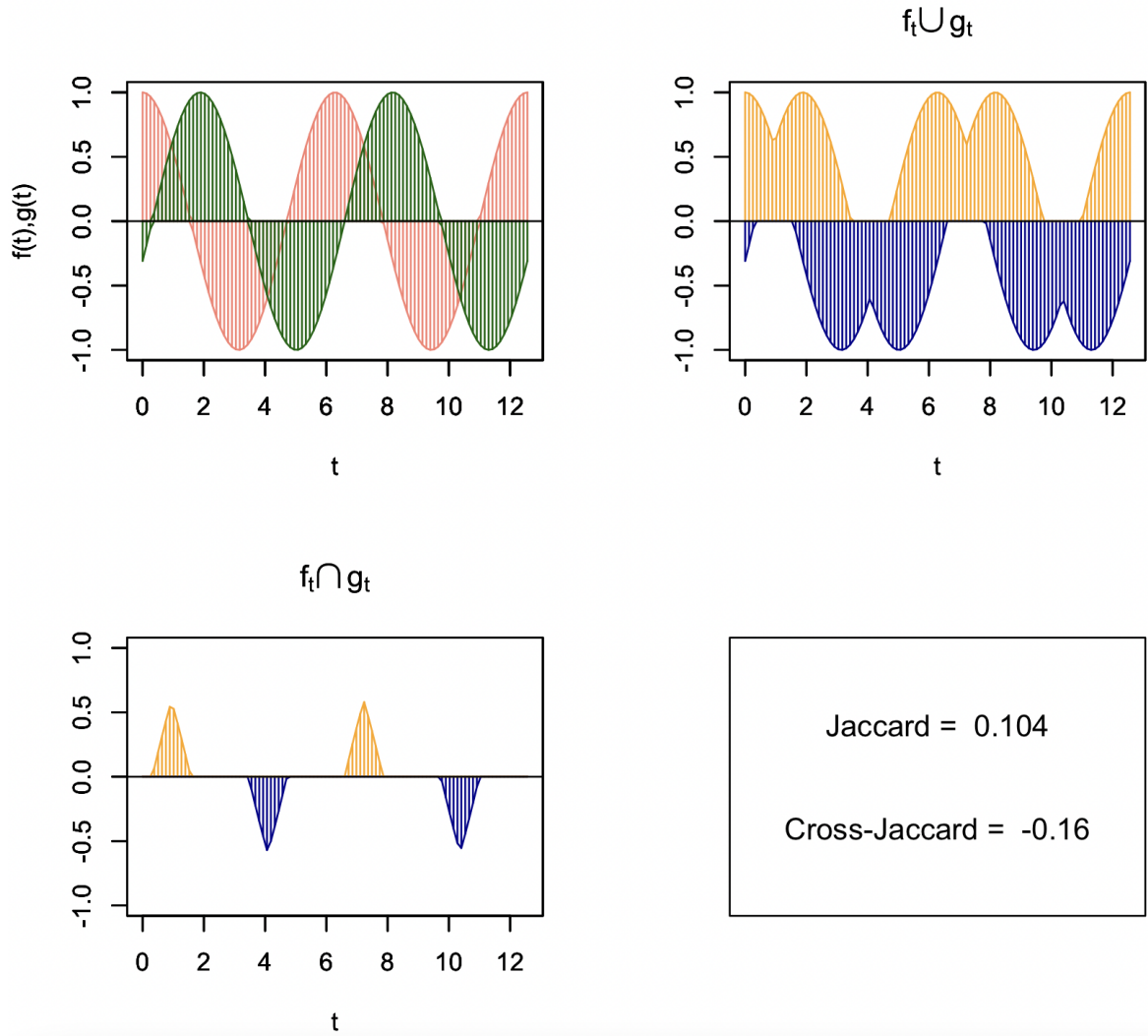


Figure 3: Another example of two functions (particular case of real-valued multisets)  $f(t)$  and  $g(t)$ , as well as their union and intersection. The respectively obtained Jaccard and cross-Jaccard similarity indices are also indicated. Observe that smaller value of the Jaccard indices have been obtained as a consequence of these two functions being less similar than in the previous example.

## References

- [1] L. da F. Costa. Modeling: The human approach to science. Researchgate, 2019. [https://www.researchgate.net/publication/333389500\\_Modeling\\_The\\_Human\\_Approach\\_to\\_Science\\_CDT-8](https://www.researchgate.net/publication/333389500_Modeling_The_Human_Approach_to_Science_CDT-8). [Online; accessed 1-Oct-2020].
- [2] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36–66, 1989.
- [3] W. D. Blizard. The development of multiset theory. *Modern Logic*, 4:319–352, 1991.
- [4] K. et al. Rana. Mathematics of multisets. *International Journal for Research in Applied Sciences and Biotechnology*, 9(3):102–106, 2022.
- [5] D. Singh, M. Ibrahim, T. Yohana, and J. N. Singh. Complementation in multiset theory. *International Mathematical Forum*, 38:1877–1884, 2011.
- [6] L. da F. Costa. Multisets. [https://www.researchgate.net/publication/355437006\\_Multisets](https://www.researchgate.net/publication/355437006_Multisets), 2021. [Online; accessed 21-Aug-2021].
- [7] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.
- [8] L. da F. Costa. Further generalizations of the Jaccard index. [https://www.researchgate.net/publication/355381945\\_Further\\_Generalizations\\_of\\_the\\_Jaccard\\_Index](https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index), 2021. [Online; accessed 21-Aug-2021].



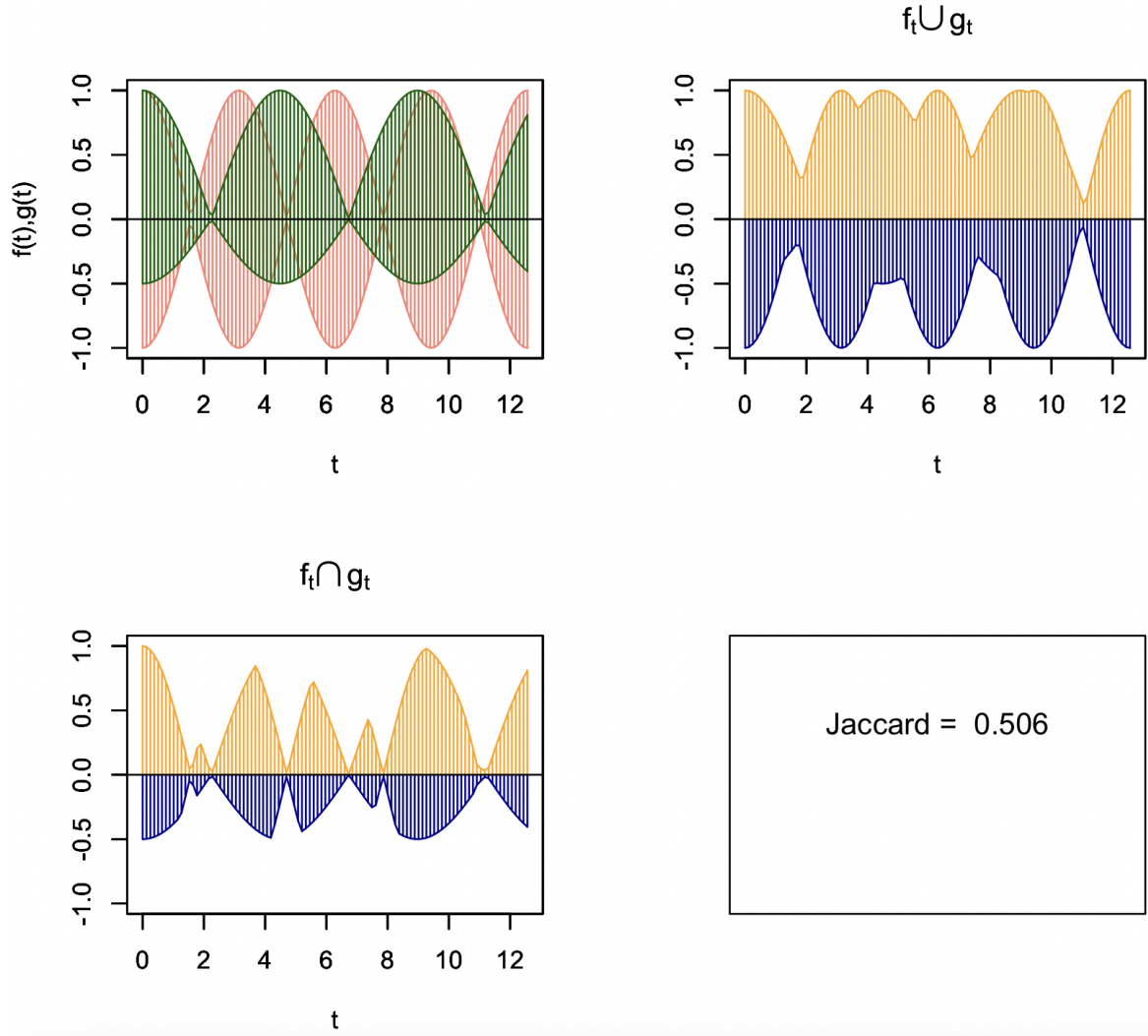


Figure 4: Example of two intervals  $f(t)$  and  $g(t)$ , their respective union and intersection, as well as the obtained Jaccard similarity index.

- [9] L. da F. Costa. On similarity. *Physica A*, 599:127456, 2022.
- [10] L. da F. Costa. Coincidence complex networks. <https://iopscience.iop.org/article/10.1088/2632-072X/ac54c3>, 2022. *J. Phys.: Complexity*, (3): 015012.
- [11] L. da F. Costa. Multiset neurons. *Physica A*, 609:128318, 2023.
- [12] B. Mirkin. *Mathematical Classification and Clustering*. Kluwer Academic Publisher, Dordrecht, 1996.
- [13] C. E. Akbas, A. Bozkurt, M. T. Arslan, H. Aslanoglu, and A. E. Cetin. L1 norm based multiplication-free cosine similarity measures for big data analysis. In *IEEE Computational Intelligence for Multimedia Understanding (IWCIM)*, France, Nov. 2014.
- [14] C. E. Akbas, A. Bozkurt, A. E. Cetin, R. Cetin-Atalay, and A. Uner. Multiplication-free neural networks. In *Signal Processing and Communications Applications Conference (SIU)*, Malatya, Turkey, May. 2015.
- [15] M. K. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications*, 3(1):19–28, 2016.
- [16] H. Wolda. Similarity indices, sample size and diversity. *Oecologia*, 50:296–302, 1981.
- [17] M. Brusco, J. D. Cradit, and D. Steinley. A comparison of 71 binary similarity coefficients: The effect of base rates. *PLOS One*, 16(4):e0247751, 2021.
- [18] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Kettters, H. Rousseau, and A. Vanhoutte. Similarity measures in scientometric research: The jaccard

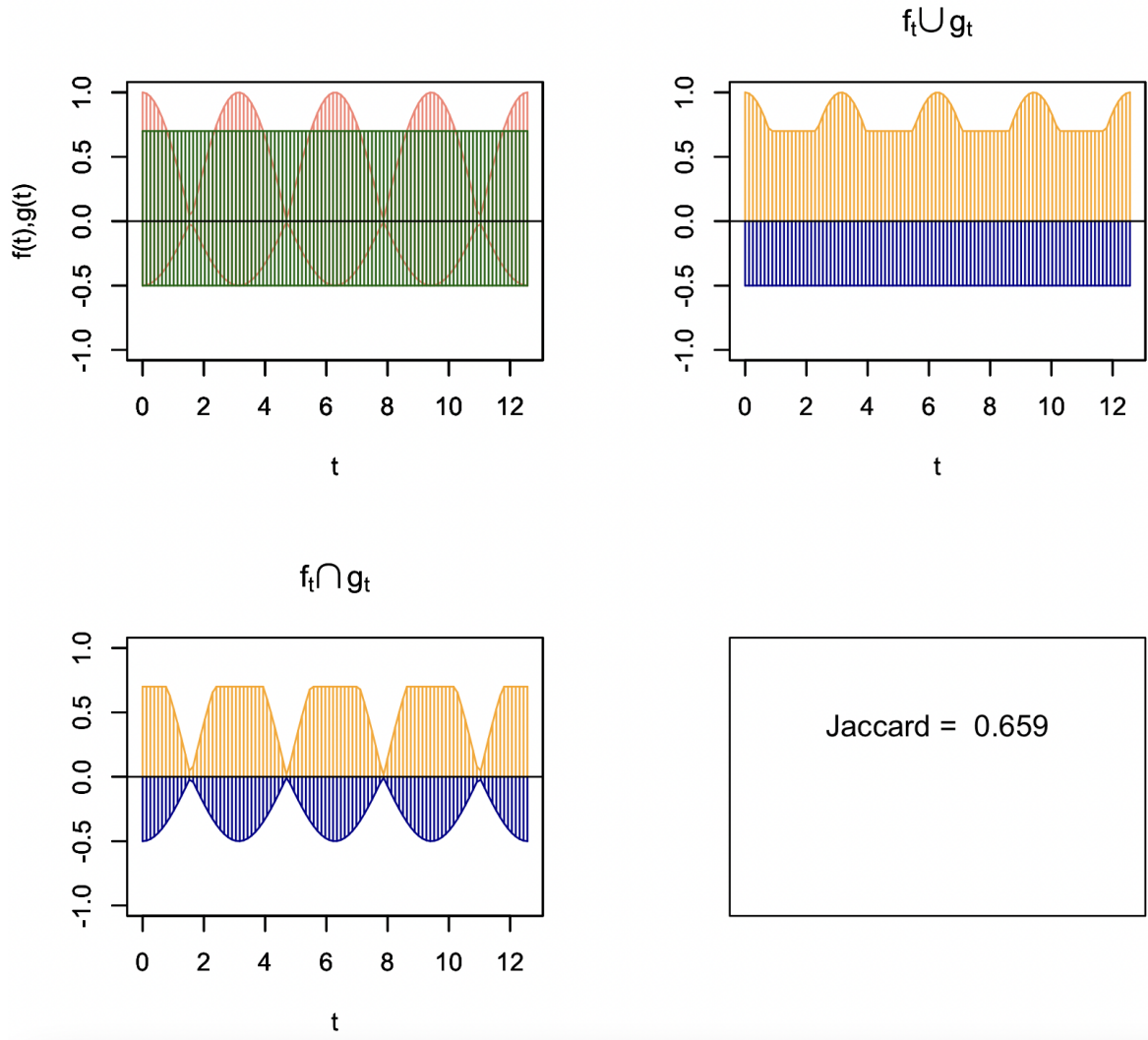


Figure 5: Another example of two interval functions  $f(t)$  and  $g(t)$ , their respective union and intersection, as well as the obtained Jaccard similarity index.

index versus salton's cosine formula. *Information Processing and Management*, 25(3):315–318, 1989.

- [19] S.-H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *Intl. J. Math. Models and Meths. in Appl. Sci.*, 1(4):300–307, 2007.
- [20] P. Jaccard. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*, 37:241–272, 1901.
- [21] Wikipedia. Jaccard index. [https://en.wikipedia.org/wiki/Jaccard\\_index](https://en.wikipedia.org/wiki/Jaccard_index). [Online; accessed 10-Oct-2021].
- [22] L. da F. Costa. Multiset neurons. [https://www.researchgate.net/publication/356042155\\_Multiset\\_Neurons](https://www.researchgate.net/publication/356042155_Multiset_Neurons), 2021.