Binary Fuzzy Measures and Choquet Integration for Multi-Source Fusion

Derek T. Anderson*, Muhammad Aminul Islam*, Roger King†, Nicolas H. Younan*, Joshua R. Fairley‡, Stacy Howington‡, Frederick Petry§, Paul Elmore§ and Alina Zare¶

*Electrical and Computer Engineering, Mississippi State University, USA,
†Center for Advanced Vehicular Systems, Mississippi State University, USA,
‡U.S. Army Engineer Research and Development Center, Geotechnical and Structures Laboratory, Vicksburg, USA,
§Geospatial Sciences and Technology Branch, Naval Research Laboratory, MS, USA,
¶Electrical and Computer Engineering, University of Florida, FL, USA

***Abstract*— Countless challenges in engineering require the intelligent combining (aka fusion) of data or information from multiple sources. The *Choquet integral* (ChI), a parametric aggregation function, is a well-known tool for multi-source fusion, where source refers to sensors, humans and/or algorithms. In particular, a selling point of the ChI is its ability to model and subsequently exploit rich interactions between inputs. For a task with N inputs, the ChI has 2^N interaction variables. Therefore, the ChI becomes intractable quickly in terms of storage and its data-driven learning. Herein, we study the properties of an efficient to store, compute, and ultimately optimize version of the ChI based on a *binary fuzzy measure* (BFM). The BFM is further motivated by empirical observations in the areas of multi-sensor fusion and hyperspectral image processing. Herein, we provide a deeper understanding of the inner workings, capabilities and underlying philosophy of a *BFM ChI* (BChI). We also prove that two fuzzy integrals, the ChI and the Sugeno integral, are equivalent for a BFM. Furthermore, only a small subset of BFM variables need be stored, which reduces the BChI to a relatively simple look up operation.**

***Keywords*— Choquet integral, fuzzy measure, multi-sensor fusion, binary fuzzy measure, binary Choquet integral**

I. INTRODUCTION

Engineering applications like Big Data, remote sensing, unmanned vehicles, robotics, signal and image processing, and machine learning, to name a few, all have something in common. They all typically require the intelligent combining (aka fusion) of multiple *sources*. Here, the term source is used to refer to the "generator" of data or information, e.g., sensors, humans and/or algorithms. In general, the idea of fusion is to obtain a "better" result than if we only used the individual inputs. However, better is not a well defined concept. In some applications, better might mean taking a set of inputs and reducing them into a single result that can be more efficiently or effectively used for visualization. Better could also refer to obtaining more desirable properties such as higher information content or lower conflict. In areas like machine learning and pattern recognition, theories that power the majority of engineering applications mentioned above, better often refers to some desirable property like more robust and generalizable solutions (e.g., classifiers). Regardless of the task at hand or the particular application, fusion is a core tool at the heart of numerous modern scientific thrusts.

Whereas the implications of this article are far reaching, we narrow its vision and scope, without loss of generality, to the particular problem of multi-sensor fusion. It is practically impossible nowadays to provide an unbiased and comprehensive review of multi-sensor fusion as there have been a massive number of publications on this topic ranging from correspondence to aggregation mathematics, various types of uncertainty, missing data, tracking, and numerous other factors. A recent review is [1]. Multi-sensor fusion is also challenging to summarize because when multiple sensors are operating in different portions of the electromagnetic spectrum important domain information is relevant related to the physics of the sensors, phenomena (e.g., objects), environments and environmental conditions that they are operating in. Herein, we focus on the underlying set of mathematics that enable fusion across this broad spectrum of challenges. In general, an aggregation operator f is a mapping that can take different *forms* (branches of mathematics). Usually, its a mapping of data (e.g., sensor measurements) from N inputs, $X = \{x_1, ..., x_N\}$, to a single result, i.e., $f(h(x_1), ..., h(x_i), ..., h(x_N)) \in \Re$, where $h(x_i)$ is the data provided by input x_i. An introductory aggregation function text and guide for practitioners is [2].

Herein, we focus on the *Choquet integral* (ChI) [3]–[6], a well-known and demonstrated parametric function for data and information fusion. The ChI is a generator function that is parametrized by the so-called *fuzzy measure* (FM), a monotone and normal capacity. Once the FM has been determined, either by an expert or learned from data, the ChI turns into a specific aggregation operator, e.g., a *linear combination of order statistics* (LCOS) [7]. The ChI has been used for numerous applications, to name a few; humanitarian demining [8], computer vision [7], pattern recognition [9]–[13], multi-criteria decision making [14], [15], control theory [16], and multiple kernel learning [8], [17]–[20]. Different approaches exist to learn the FM from data, e.g., [17], [21]–[26].

In this article, we have two primary motives. The first is an application driven one. In [27], we studied the task of learning a ChI fusion for binary decision making problems, e.g., target versus non-target classification, relative to uncertainty in the labeling of training data. Our approach was to extend the ChI via *multiple instance learning* (MIL). Results were demonstrated for the fusion of multiple classifiers from one or more sensors relative to *explosive hazard detection* (EHD) for humanitarian demining and the fusion of RX detectors for hyperspectral image processing. In particular, we observed that the data learned ChI solutions almost always preferred answers that had FM variables in (or approximately) $\{0,1\}$ versus a more arbitrary expected number on the standard $[0,1]$ FM interval. An advantage is that learning the 2^N FM variables on $\{0,1\}$ is a drastically simpler problem than learning on $[0,1]$. This leads us to our second motivation. It is not only difficult to learn on $[0,1]^{2^N}$ versus $\{0,1\}^{2^N}$, but it is also a naturally intractable problem. Meaning, even for relatively small N it can be a game stopper. Therefore, we are interested in multi-source problems that utilize a $\{0,1\}$ versus $[0,1]$ variable range since it is efficient to store only the one-valued terms and the ChI can easily be computed via a look up table on the fly.

Herein, we put forth the following contributions. First, we investigate a *binary FM* (BFM). Second, we define and study a *BFM ChI* (BChI). Third, we prove that for a BFM, two well-known *fuzzy integrals* (FIs), the *Sugeno integral* (SI) and ChI, are equivalent and therefore selection of integral is irrelevant. Fourth, we explain the underlying aggregation philosophy for a BChI, a procedure we call the "best pessimistic agreement". Last, we show that only a small subset of one-valued variables need be stored, which leads to an efficient to store, learn and simple to compute via a look up table version of the ChI.

The remainder of this article is organized as such. In Section II we summarize the FM and ChI, Section III is an alternative way to think about the ChI, Section IV describes a BFM and Section V outlines the BChI. Next, a semantic description of the BChI is provided (Section VI), followed by BChI memory savings (Section VII) and a fast way to calculate the BChI (Section VIII).

II. Choquet Integral (ChI)

The reader can refer to [6] for a recent survey and description of theory, applications and important extensions of the *fuzzy integral* (FI) [3]. Let $X = \{x_1, x_2, ..., x_N\}$ be a set of N inputs, e.g., two radar systems and an infrared sensor. A FM is a monotonic function defined on the power set of X, 2^X, as $\mu : 2^X \to \Re^+$ that satisfies the following two properties: (i) boundary condition, $\mu(\emptyset) = 0$ and $\mu(X) > 0$ and (ii) monotonicity property, if $A, B \subseteq X$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$. Often an additional constraint is imposed on the FM to limit the upper bound to 1, i.e., $\mu(X) = 1$. Figure (1) is an illustration of the FM. FM variables of different cardinality are displayed with different colors.

Let $h(x_i)$ be the data/information, e.g., sensor measurement, from input i. The discrete ChI (finite X) is

$$\int_C h \circ \mu = C_\mu(h) = \sum_{i=1}^{N} h(x_{\pi(i)}) \left[\mu(A_i) - \mu(A_{i-1})\right], \tag{1}$$

where π is a permutation of X, such that $h(x_{\pi(1)}) \geq h(x_{\pi(2)}) \geq \ldots \geq h(x_{\pi(N)})$, $A_i = \{x_{\pi(1)}, \ldots, x_{\pi(i)}\}$, and $\mu(A_0) = 0$. Based on selection of μ, the ChI turns into a specific aggregation operator. For example, when $\mu(A) = 0, \forall A \in 2^X \setminus X$, the ChI is equivalent to the minimum operator. When $\mu(A) = 1, \forall A \in 2^X \setminus \emptyset$, we obtain the maximum operator. When $\mu(A) = \frac{|A|}{N}, \forall A$ we obtain the arithmetic mean. More generally, when $\mu(A) = \mu(B)$ $\forall A, B \in 2^X$ such that $|A| = |B|$, we recover the familiar class of LCOS, e.g., min, max, soft min and max, mean, median, trimmed statistics, etc.

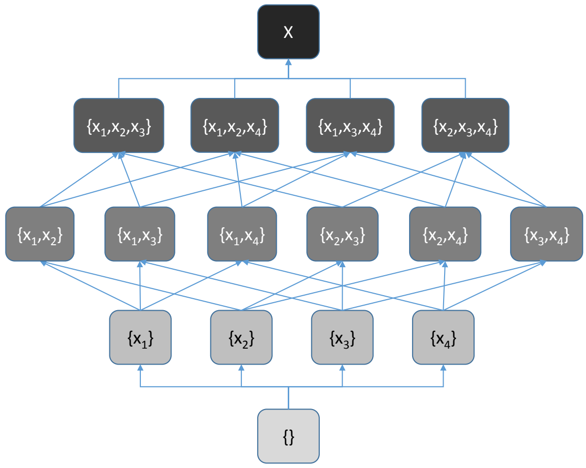

Fig. 1. Illustration of variables and monotonicity constraints in the FM. Higher cardinality sets are shaded darker and arrows denote subsethood.

III. Lattice Representation

The first step in the discrete (finite X) ChI is to sort the inputs (h values). In total, there are $N!$ possible sorts. An alternative way to think about the ChI is in terms of its underlying lattice, shown in Figure (1). The lattice is a way to relate the FM variables with respect to monotonicity, of which there are $N(2^{N-1}-1)$ monotonicity constraints. An important concept is a "walk" through the lattice. Figure (2) illustrates two different walks. Per a given input sorting, the ChI (Equation (1)) induces a walk, meaning it visits or uses $N+1$ variables, i.e., $\mu(A_i) - \mu(A_{i-1})$. Formally, this is exactly one variable from each "level" in the FM, where level refers to all variables that have the same cardinality. This provides another convenient way to think about the ChI. The ChI is actually $N!$ different aggregation operators wrapped into one function—one for each walk through the lattice. In the case of operators like the minimum, maximum, average, and other LCOSs, each layer in the FM has constant measure value and therefore the operator is fixed and independent of which walk is observed. However, there are countless custom FMs (and therefore ChIs) that do not have this behavior, e.g., most

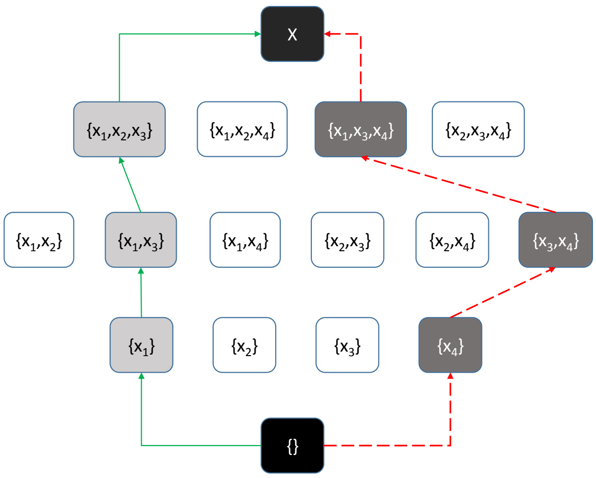

Fig. 2. Example of two "walks" through the FM lattice with respect to the ChI. Variables visited in the first walk, i.e., for an input sorting of $h(x_1) \geq h(x_3) \geq h(x_2) \geq h(x_4)$, are shown in light gray and their order visited is indicated by green arrows. Variables visited in the second walk, i.e., $h(x_4) \geq h(x_3) \geq h(x_1) \geq h(x_2)$, are shown in dark gray and their order visited is indicated by dashed red lines. Note, the empty set and X are both shaded in black and visited by both walks. Each walk is the set of weights used in the respective ChI calculation.

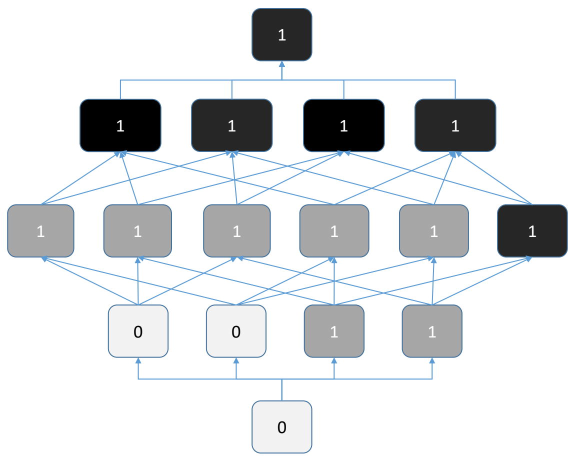

Fig. 3. Example BFM—white cells have value zero, gray cells have value one (but one or more subsets, if we remove a single element, are zero) and black cells have value one (all direct subsets have value one).

cases of a BFM which we investigate later, and walks result in different aggregations per input value sort.

IV. Binary Fuzzy Measure (BFM)

A BFM is trivial, it is a restriction of $\mu \in \{0,1\}$ instead of $\mu \in [0,1]$. While simple, a BFM has big implications. As already discussed, it simplifies the space for data-driven optimization, $\{0,1\}^{2^N}$ versus $[0,1]^{2^N}$, and it often leads to more compact solutions. When the FM is allowed to take any value in $[0,1]$, we must generally store all variables. However, if we work with a BFM then naturally a number of these values, the exact number of depends on the target FM, are zero. Instead of storing all $2^N - 2$ variables (since $\mu(\emptyset) = 0$ and $\mu(X) = 1$), we can instead record just one-valued variables (which leads to savings in terms of storage). Figure (3) shows an example BFM.

V. BFM Choquet Integral (BChI)

Mathematically speaking, a *binary ChI* (BChI) is Equation (1). There is nothing new. However, what is interesting is what happens as a result of using a BFM. We can expand Equation (1) and observe the following (due to the monotonicity property of the FM).

$$\int_C h \circ \mu = \sum_{i=1}^{N} h(x_{\pi(i)}) \left[\mu(A_i) - \mu(A_{i-1})\right],$$
$$= \left(\sum_{i=1}^{n_1} h(x_{\pi(i)}) \left[0-0\right]\right) + h(x_{\pi(n_1+1)}) \left[1-0\right]$$
$$+ \left(\sum_{i=n_1+2}^{N} h(x_{\pi(i)}) \left[1-1\right]\right),$$

where n_1 is the number of inputs whose corresponding $\mu(A_i)$ terms are zero, $h(x_{\pi(n_1+1)})$ is the first term with $\mu(A_{n_1+1}) = 1$ and the remaining set are all one-valued as well. Note, without loss of generality, we partitioned the sum up as such, however we are more than aware that the first term may have an empty number of terms if $\mu(A_1) = 1$. What is important to note here is that due to monotonicity, a BFM will result in zero to $N-1$ $(0-0)$ terms, a single $(1-0)$ weight and anywhere from zero to $N-1$ $(1-1)$ weights. Thus, only one weight is one and all other weights are zero, resulting in the selection of a single input value, $h(x_{\pi(n_1+1)})$.

This leads us to Proposition 1. However, we first review the definition of the *Sugeno integral* (SI) [3],

$$\int_S h \circ \mu = S_\mu(h) = \bigvee_{i=1}^{N} \left(h(x_{\pi(i)}) \wedge \mu(A_i)\right), \quad (2)$$

where $\vee$ is the maximum operator and $\wedge$ is the minimum operator. However, these two operators, a t-conorm and t-norm can, and have, been extended beyond the scope of minimum and maximum. The SI and ChI are both FIs. We will not go into full depth here about similarities and differences between the two integrals. The reader can see [5] for more details. In general, the ChI is desirable because of factors such as; it does not use the minimum and maximum operators and is differentiable (a property that has been exploited before in data-driven learning), for an additive (probability) measure one recovers the classical Lebesgue integral, etc.

Proposition 1: The ChI and SI are equal for a BFM.

Proof: This proof is trivial. As already noted, the BChI has a single non-zero weight and we get $h(x_{\pi(n_1+1)})$. When it comes to the SI, we take the minimum of each $h(x_{\pi(i)})$ with their $\mu(A_i)$. The SI can be expanded as such,

$$\int_S h \circ \mu = \bigvee_{i=1}^{N} \left(h(x_{\pi(i)}) \wedge \mu(A_i)\right),$$
$$= \left((h(x_{\pi(1)}) \wedge 0) \vee (h(x_{\pi(2)}) \wedge 0) \vee ...\right)$$
$$\vee \left(h(x_{\pi(n_1+1)}) \wedge 1\right)$$
$$\vee \left((h(x_{\pi(n_1+2)}) \wedge 1) \vee ... \vee (h(x_{\pi(N)}) \wedge 1)\right),$$

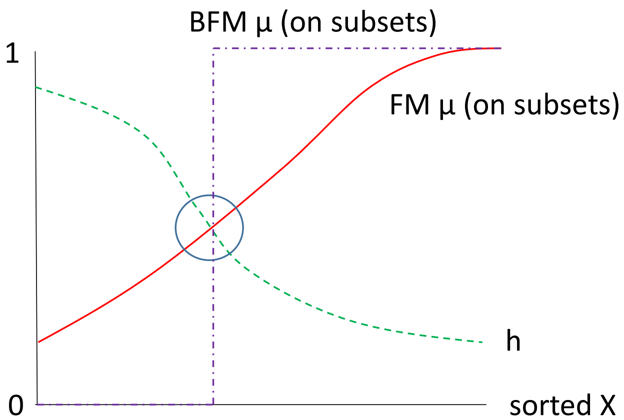

Fig. 4. "Best pessimistic agreement" behavior of the BChI. Green dashed curve is the sorted h values, red is the respective FM μ relative to sorted sets of increasing size and purple is a BFM μ relative to sorted sets of increasing size. The blue circle is the output of the BChI.

where once again, without loss of generality, we denote a partitioning of the integral into parts, specifically two, where the first have zero-valued FM values and the second has one-valued FM values. However, if the first A_1 has value one then there are of course no zero terms. Note, $h(x_{\pi(n_1+1)})$ is the largest of the set $h(x_{\pi(n_1+1)})$ to $h(x_{\pi(N)})$. Therefore, it is selected as the result of the SI. This means that both the ChI and the SI both select the same input with respect to a BFM and are therefore equal, completing the proof. ■

What is interesting about Proposition 1 is it informs us that whereas there are numerous mathematical differences in general between the SI and ChI, many of which often impact a domain and dictate which integral is selected, they do not matter for the case of a BFM. The SI and the ChI are completely equal under this restriction.

VI. What is the BChI Really Doing?

In this section, we explore what the BChI is doing in terms of underlying data/information aggregation philosophy. The BChI breaks down into the following steps.

1) Sort the input values from largest to smallest, i.e., $h(x_{\pi(1)}) \geq ... \geq h(x_{\pi(N)})$.
2) Find the smallest i such that $\mu(A_i) = 1$.
3) Return $h(x_{\pi(i)})$.

Figure (4) is an illustration of the BChI. Note, the h values are sorted in decreasing order and g is monotonic. In addition, we can describe what the BChI is doing in a few (be it equivalent) ways.

Interpretation 1: The BChI is looking for the *first* A_i (i.e., smallest i), relative to the h sort, that has non-zero "worth" (FM value). The result is the "best pessimistic agreement" with respect to h and μ.

Interpretation 2: Another way to think about the BChI is the following. Each BFM variable that is equal to zero indicates a more-or-less *do not care* condition. Meaning, the input by itself should not be considered, more input (evidence) is needed before we are willing to make a decision. However, BFM variables equal to one indicate a group of inputs is worth considering. Specifically, they all agree at least as much as their minimum value.

Interpretation 3: We can also talk about the BChI in terms of a lattice walk. The $N+1$ variables in a walk (the constants, i.e., non-h terms, in the BChI) are a monotonic sequence of zeros followed by ones. Once we reach the first one value we take the minimum input for that set. Conversely, it is the maximum of all the inputs with corresponding BFM value one.

Example 1: As already stated, the focus of this article is the underlying mathematics of aggregation that are used by applications like multi-sensor fusion. For conceptual sake, consider a problem that requires interrogation in three different parts of the *electromagnetic* (EM) spectrum. For example, assume we have a task where x_1 is Radar, x_2 is *infrared* (IR) and x_3 is *EM induction* (EMI) for the handheld detection of buried explosive hazards in humanitarian demining [28], [29]. However, the following applies to any N different spectral bands in a hyperspectral sensor or combination of single spectral band sensors in different parts of EM. Next, let there be one, for simplicity sake, respective algorithm on each of the above three sensors (GPR, IR and EMI) and let each sensor produce an output indicating target (value one) or non-target (value zero). The algorithms can be a binary decision, $\{0,1\}$, or a $[0,1]$ value generated by a probabilistic support vector machine, Bayes decision theoretic classifier, or any other number of countless pattern recognition or machine learning algorithms. Assume we already have our BFM—i.e., it was specified by an expert or learned from data (using any of the methods referenced in Section I). Let the BFM be $\mu(x_1)$=$\mu(x_2)$=$\mu(x_3)$=0 (no source is fully trustworthy by itself) and $\mu(\{x_2,x_3\})$=0. Let all other variables be value one. Now, consider a particular set of inputs be $h(x_1) = 0.8$, $h(x_2) = 0.5$ and $h(x_3) = 0.01$. The BChI assigns a value of 0.5 because when Radar is the highest value we require input and confirmation from IR (and disregard EMI). However, for an input of $h(x_3) = 1$, $h(x_2) = 0.9$ and $h(x_1) = 0.01$, the BChI gives us a value of 0.01 because our BFM (relative to the ChI) informs us that when EMI has the strongest detection and IR is next, we need confirmation from all three sensors. The point is, the BFM, relative to the BChI, encodes a wealth of aggregation information across different walks about a particular task at hand.

VII. Compression of BFM for BChI

In Section V, we discussed that for a BFM only one-valued terms need be stored. This can lead to big memory savings as there are otherwise $2^N - 2$ non-constant variables. For a relatively small problem with 10 inputs we already have $1,022$ variables. For 20 inputs, we have $1,048,574$ variables. However, if a problem has a good number of one values in the lower layers of the lattice then savings diminish. This cannot be predicted, the exact amount of savings, in general. Each problem likely requires a different aggregation operator and therefore measure. Thus, the amount of savings depends on the problem at hand.

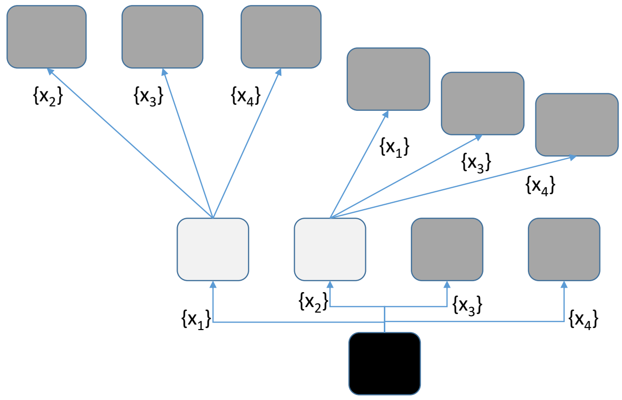

Fig. 5. Example BChI link list data structure corresponding to Figure (3). The first step is to sort the inputs. Next, we start at the black node and move the direction of the largest input variable. If we ever encounter a "gray cell" (terminal node) then we take the minimum of the set of numbers up to that point. Otherwise, we keep following edges with respect to our sorting order until we hit a terminal node.

However, Figure (3) reveals a further savings. Variables shown in white (call it set $W \in 2^X$) are all zero-valued BFM terms. Values in gray (call it set $G \in 2^X$) are one-valued BFM terms that have at least input (all subsets if we remove one element) with value zero. Values in black (set $B \in 2^X$) are one-valued BFM terms whose entire input set are one-valued. It is trivial to show that $2^X = W \cup G \cup B$ and $W \cap G = W \cap B = G \cap B = \emptyset$. Furthermore, G creates a partitioning between otherwise uninformative zero and one valued variables. What this means is, we have redundancy and given G we can perfectly reconstruct the entire BFM. Furthermore, G will often in practice be $|G| \ll |B \cup G \cup W|$. The take away is, one can identify and just store G and perform the BChI.

VIII. BChI Data Structure

The ChI is typically calculated with respect to Equation (1), which has N subtractions and N multiplications and $N-1$ additions. Alternatively, we can pre-compute all differences in μ variables (if storage is not an issue), which leads to only N multiplications and $N-1$ additions.

For the BChI, one option is to store the reduced set G and on the fly, post sorting of our inputs, identify the corresponding term in G and take the minimum of that set. There are a number of algorithms that can be employed. A computationally efficient, but possibly not the most memory efficient, scheme is to use a link list data structure. At most, we would take N "steps" (traversals) in such a data structure (but likely far less on average). Figure (5) is an example for the BFM in Figure (3).

IX. Conclusion and Future Work

In summary, the focus of this article is the underlying mathematics of data and information fusion. The tool selected is the *Choquet integral* (ChI), a flexible parametric aggregation function that morphs into a class of aggregation operators based on specification of the *fuzzy measure* (FM), a normal and monotone capacity. In practice, the FM is specified by an expert or learned by data. However, tractability (both in learning but also memory storage) of the FM, and therefore the ChI, is of concern. In addition, applications like our discussed multiple instance learning ChI for binary target detection in multi-sensor systems (e.g., humanitarian demining and hyperspectral image processing) often learns or naturally prefers what we call a *binary FM* (BFM). Herein, we studied a *BFM ChI* (BChI). We discovered that a BFM renders the *Sugeno integral* (SI) and ChI equal, which is not often the case, and the underlying aggregation philosophy is the "best pessimistic agreement". We also observed that drastically fewer BFM variables need be stored and the BChI is nothing more than a simple look up strategy via some scheme like a link list data structure. Overall, we encounter both memory and computational savings. Last, learning a BFM is a much simpler task as the space is $\{0,1\}^{2^N}$ versus $[0,1]^{2^N}$.

Whereas the ChI is already a well-demonstrated theory for fusion, in future work we will explore the empirical benefits of using the proposed BChI for data-driven learning, ChI compression and efficient calculation in the context of specific applications for multi-sensor fusion.

Acknowledgment

This material is based upon work supported by the U.S. Army TACOM Life Cycle Command under Contract No. W56HZV-08-C-0236. In addition, we would like to thank the Naval Research Laboratory's Base Program, Program Element No. 0602435N for sponsoring this research.

References

[1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[2] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*, 1st ed. Springer Publishing Company, Incorporated, 2008.

[3] M. Sugeno, "Theory of fuzzy integrals and its applications," *Ph.D. thesis, Tokyo Institute of Technology*, 1974.

[4] G. Choquet, "Theory of capacities," in *Annales de l'institut Fourier*, vol. 5. Institut Fourier, 1954, pp. 131–295.

[5] M. Grabisch, H. T. Nguyen, and E. A. Walker, *Fundamentals of uncertainty calculi with applications to fuzzy inference*. Springer Science & Business Media, 2013, vol. 30.

[6] D. T. Anderson, T. C. Havens, C. Wagner, J. M. Keller, M. F. Anderson, and D. J. Wescott, "Extension of the fuzzy integral for general fuzzy set-valued information," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1625–1639, 2014.

[7] H. Tahani and J. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Transactions System Man Cybernetics*, vol. 20, pp. 733–741, 1990.

[8] S. R. Price, B. Murray, L. Hu, D. T. Anderson, T. C. Havens, R. H. Luke, and J. M. Keller, "Multiple kernel based feature and decision level fusion of ieco individuals for explosive hazard detection in flir imagery," in *SPIE*, vol. 9823, 2016, pp. 98 231G–98 231G–11.

[9] M. Grabisch and J. Nicolas, "Classification by fuzzy integral: performance and tests," *Fuzzy sets and systems*, vol. 65, no. 2-3, pp. 255–271, 1994.

[10] M. Grabisch and M. Sugeno, "Multi-attribute classification using fuzzy integral," in *IEEE International Conference on Fuzzy Systems*. IEEE, 1992, pp. 47–54.

[11] A. Mendez-Vazquez, P. Gader, J. M. Keller, and K. Chamberlin, "Minimum classification error training for choquet integrals with applications to landmine detection," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 1, pp. 225–238, 2008.

[12] J. M. Keller, P. Gader, H. Tahani, J. Chiang, and M. Mohamed, "Advances in fuzzy integration for pattern recognition," *Fuzzy sets and systems*, vol. 65, no. 2-3, pp. 273–283, 1994.

[13] P. D. Gader, J. M. Keller, and B. N. Nelson, "Recognition technology for the detection of buried land mines," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 1, pp. 31–43, 2001.

[14] M. Grabisch, "The application of fuzzy integrals in multicriteria decision making," *European journal of operational research*, vol. 89, no. 3, pp. 445–456, 1996.

[15] C. Labreuche, "Construction of a choquet integral and the value functions without any commensurateness assumption in multi-criteria decision making," in *EUSFLAT Conf.*, 2011, pp. 90–97.

[16] L. Tomlin, D. T. Anderson, C. Wagner, T. C. Havens, and J. M. Keller, *Fuzzy Integral for Rule Aggregation in Fuzzy Inference Systems*. Springer International Publishing, 2016, pp. 78–90.

[17] A. J. Pinar, J. Rice, L. Hu, D. T. Anderson, and T. C. Havens, "Efficient multiple kernel classification using feature and decision level fusion," *IEEE Trans. Fuzzy Syst.*, vol. PP, no. 99, pp. 1–1, 2016.

[18] A. Pinar, T. C. Havens, D. T. Anderson, and L. Hu, "Feature and decision level fusion using multiple kernel learning and fuzzy integrals," in *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Aug 2015, pp. 1–7.

[19] L. Hu, D. T. Anderson, T. C. Havens, and J. M. Keller, *Efficient and Scalable Nonlinear Multiple Kernel Aggregation Using the Choquet Integral*. Springer International Publishing, 2014, pp. 206–215.

[20] L. Hu, D. T. Anderson, and T. C. Havens, "Multiple kernel aggregation using fuzzy integrals," in *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2013, pp. 1–7.

[21] D. T. Anderson, S. R. Price, and T. C. Havens, "Regularization-based learning of the choquet integral," in *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2014, pp. 2519–2526.

[22] J. M. Keller and J. Osborn, "Training the fuzzy integral," *International Journal of Approximate Reasoning*, vol. 15, no. 1, pp. 1–24, 1996.

[23] ——, "A reward/punishment scheme to learn fuzzy densities for the fuzzy integral," *Proc. Int. Fuzzy Sys. Assoc. World Cong*, pp. 97–100, 1995.

[24] A. Mendez-Vazquez and P. Gader, "Sparsity promotion models for the choquet integral," in *Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium on*. IEEE, 2007, pp. 454–459.

[25] D. T. Anderson, J. M. Keller, and T. C. Havens, "Learning fuzzy-valued fuzzy measures for the fuzzy-valued sugeno fuzzy integral," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2010, pp. 502–511.

[26] G. Beliakov, "Construction of aggregation functions from data using linear programming," *Fuzzy Sets and Systems*, vol. 160, no. 1, pp. 65–75, 2009.

[27] X. Du, A. Zare, J. M. Keller, and D. T. Anderson, "Multiple instance choquet integral for classifier fusion," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, July 2016, pp. 1054–1061.

[28] M. Lee, D. Anderson, J. E. Ball, and J. White, *Background adaptive division filtering for hand-held ground penetrating radar*. SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI, 2016.

[29] J. White, D. T. Anderson, J. E. Ball, and B. Parker, *Curvelet filter based prescreener for explosive hazard detection in hand-held ground penetrating radar*. SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI, 2016.