Alpha-cut based compositional representation of fuzzy sets and exploration of associated fuzzy set regression

Direnc Pekaslan, *Member, IEEE,* and Christian Wagner, *Member, IEEE*

Abstract—The compositional representation of data and associated statistical approaches is a powerful framework for modelling and reasoning about quantities which reflect proportions of a whole. Recently, an increasing body of work has started exploring the adoption of a compositional representation for modelling interval-valued data reflecting uncertainty or vagueness, for example interval-valued questionnaire responses. Results have flagged the intriguing potential of this approach, such as the elegant handling of traditional inference challenges, including implicitly ensuring coherence in linear regression for interval data, i.e. ensuring the estimated left bound of intervals is smaller than the right one. Building on these insights, extending the compositional representation via alpha-cut decomposition to fuzzy sets is an intuitive next step. In this paper, we discuss this compositional representation of fuzzy sets, building on prior interval work. We proceed to explore the adoption of compositional regression approaches to conduct linear regression on fuzzy set valued data sets. We demonstrate the approach, discuss results and in particular flag shortcomings and the challenges for next steps.

I. INTRODUCTION

Compositional data convey the structural information reflecting the quantitative parts of a whole, such that when features of a dataset inherently influence each other, they are expressed in proportions where all components are non-negative and are summed up to a constant, e.g., 1, 24 hours or 100% etc. [1], [2] For instance, data on time spent in daily (24 hours) physical activity. This inherent dependency of such compositional data is usually not taken into consideration in *standard* multivariate statistical approaches [1], [2], [3], [4], [5]. Compositional statistical approaches–such as compositional linear regression– are powerful tools for modelling and reasoning about data which captures proportions of a whole [6], [7], [8], [9], [10], [11], [12].

Recently, an increasing body of work has started exploring the adoption of a compositional representation for modelling interval-valued data reflecting uncertainty or vagueness–for example, interval-valued questionnaire responses [13], [14], [14]. Due to the nature of the mutual dependency of interval endpoint parameters, the compositional transformation is proposed in [15]. The authors articulate how, why and when a compositional representation of interval-valued data may be appropriate, and further demonstrate compositional linear regression applied to interval-valued data [15].

Direnc Pekaslan and Christian Wagner are with Intelligent Modelling and Analysis (IMA) Group and Lab for Uncertainty in Data and Decision Making (LUCID), School of Computer Science, University of Nottingham, Nottingham, United Kingdom. (e-mail:direnc.pekaslan1@nottingham.ac.uk; christian.wagner@nottingham.ac.uk)

The literature on interval-valued regression studies has so far afforded improved model accuracy [16], [17], [18] and increased resilience to parameter flipping or 'loss of mathematical coherence' [19], [20], [21], [22], [23]. The latter is a major challenge for interval-regression as models are required to maintain the mathematical structure of intervals with a left and a right endpoint, or a centre and (positive) range.

Compositional representation of interval-valued data and compositional interval regression results have flagged intriguing potential, such as the elegant handling of such traditional inference challenges, including implicitly ensuring coherence [15]. Building on these insights, extending the compositional representation to fuzzy sets (FSs) is an intuitive next step.

In this paper, we discuss an α-cut based compositional representation of FSs, building on prior interval work [13], [15]. We proceed to explore the adoption of compositional regression approaches to conduct linear regression on the compositional representations of FS valued data. We demonstrate the approach, discuss results and in particular flag shortcomings and the challenges for next steps.

The structure of this paper is as follows. Section II provides background information on the α-cut interval-valued representation of FSs, compositional data and compositional linear regression. Section III presents the α-cut based compositional representation of FSs and the overall methodology for exploring the associated FS regression. Section IV demonstrates the initial experiments of the compositional linear model on synthetically generated FSs interval data and the results illustrations are given. Section V provides conclusions and a reflection on future work.

II. BACKGROUND

A. FS and α-cut decomposition

A FS A is defined on a universe of discourse X and characterised by MF $\mu_A(x)$ that takes values in the interval $[0, 1]$. A FS A in X can be represented as a set of ordered pairs of a generic elements x, and their grade of MF is shown as follows:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{1}$$

If membership grades are constrained to be either 0 or 1, then a crisp set is obtained. Otherwise, membership grades $\mu_A(x)$ takes the value in the interval of $[0, 1]$ for each element $x \in X$.

The principal role of α-cuts and strong α-cuts in FS theory is their capability to represent FSs via α-cut decomposition [24]. The general idea of alpha cut (α)-cut decomposition is

to decompose FSs into a collection of crisp sets (intervals) related together via the α levels [25], [24]. For all membership degrees α level is defined in $[0,1]$ and given a FS A, an (α) cut is defined as follows:

$$^{\alpha}[\overline{A}] = \{\forall x \in X| \ \mu(x) \geq \alpha\}, \quad (2)$$

and the strong α-cuts $(^{\alpha+}[\overline{A}])$ are defined as:

$$^{\alpha+}[\overline{A}] = \{\forall x \in X| \ \mu(x) > \alpha\}, \quad (3)$$

To illustrate the application of α-cut, the following FS set A is given with the $\alpha = 0.5$ in Fig. 1.
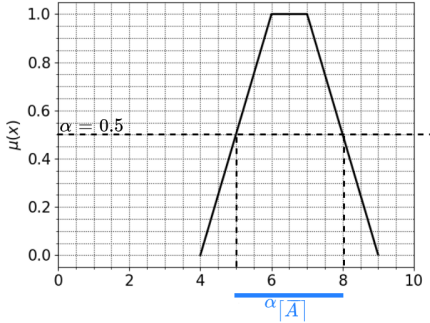

Fig. 1: $\alpha = 0.5$ decomposition which leads to interval $[5, 8]$.

The given 0.5-cut level $(^{0.5}[\overline{A}])$ on the FS A provides the closed interval $[5, 8]$. The collection of all (α)-cuts provides the (α)-cut representation of that fuzzy set. Note that for normal, convex FSs, for each $\alpha \in (0, 1]$, the α-cut is a closed interval.

B. Interval-Valued Data

A closed interval $\overline{a}$ is formed by two endpoints as lower a^- and upper a^+, with the condition of $a^- \leq a^+$.

In this paper, the intervals are formed from the defined α-cuts on FSs $(^{\alpha}[\overline{A}])$, and vice-versa, on a finite domain $X = [X^-, X^+]$, i.e. $^{\alpha}[\overline{A}] \subseteq X$. The lower endpoint if each such interval is denoted as $^{\alpha}[\overline{A}]^-$, while the upper endpoint is denoted as $^{\alpha}[\overline{A}]^+$.

At times, we refer to a set of n intervals over an α-cut $^{\alpha}[\mathbf{A}] = \left(^{\alpha}[\overline{A_1}], ^{\alpha}[\overline{A_2}], ..., ^{\alpha}[\overline{A_n}]\right)$. The cardinality or size each interval $|^{\alpha}[\overline{A}]|$ of an interval $^{\alpha}[\overline{A}]$ is given by $|^{\alpha}[\overline{A}]| = ^{\alpha}[\overline{A}]^+ - ^{\alpha}[\overline{A}]^-$.

C. Compositional Data

A data set is called compositional if the quantities reflect proportions of a whole which is a fixed total sum κ, e.g., percentages of workers in different sectors, portions of the chemical elements in a mineral, concentrations of nutrients in a beverage, portions of working time spent on different tasks etc. [5]. More formally, a (row) vector, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_D)$ is a D-part composition where all the components are strictly positive real numbers and reflect relative information. A sample space of this compositional structure is called a simplex S^D, which is expressed as follows:

$$S^D = \{\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_D)|\eta_i > 0, i = 1, 2, ..., D; \sum_{i=1}^{D} \eta_i = \kappa]\}. \quad (4)$$

As a common practice, the three-part ($D = 3$) compositional data can be illustrated on a ternary diagram to show the components' compositional structure and inherent dependency. In a ternary diagram, triangle vertices represent the three elements of the composition. While high proportions are close to a vertex, low proportions are further, and equal proportions lie on the triangle's centre. For instance, the composition $\boldsymbol{\eta} = (0.5, 0.3, 0.2)$ is shown on the illustrative ternary diagram in Fig. 2.
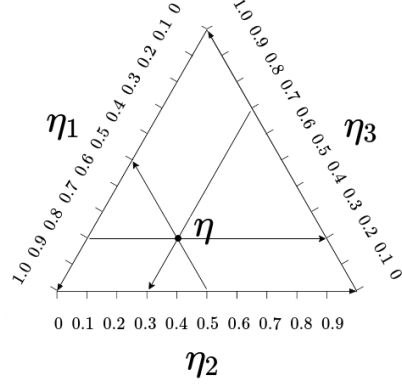

Fig. 2: Ternary diagram of the composition $(0.5, 0.3, 0.2)$.

D. Log-ratio Transformation

To enable standard unconstrained multivariate statistic operations, compositional data is commonly mapped from the simplex (S^D) to the real space (R^D). For this purpose, in the 1980s, the log-ratio transformation was proposed [1], [2] which provides a one-to-one mapping on to real values (between $-\infty$ and $+\infty$) where any compositions can be reformulated in terms of log-ratios, and vice versa.

In the literature, several algorithms for log-transformation of compositional data have been proposed (e.g., the centered log-ratio, the additive log-ratio or the isometric log-ratio [2], [26]). As one of the most common transformations, the centred log-ratio (clr) is used in this paper, in order to transform compositions (e.g., $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_D) \in S^D$) into coordinates as $S^D \rightarrow R^D$. The clr transformation is employed as follows:

$$clr(\boldsymbol{\eta}) = \boldsymbol{\eta}^{clr} = \left(\log(\frac{\eta_1}{g(\boldsymbol{\eta})}), \ldots, \log(\frac{\eta_D}{g(\boldsymbol{\eta})}) \right), \quad (5)$$

where $g(\boldsymbol{\eta})$ is the geometric mean of the composition vector $\boldsymbol{\eta}$:

$$g(\boldsymbol{\eta}) = \left(\prod_{i=1}^{D} \eta_i \right)^{1/D} \quad (6)$$

The inverse operations clr^{-1} is done as follows:

$$clr^{-1}(\boldsymbol{\eta}^{clr}) = \boldsymbol{\eta} = C\left(exp(\eta_1^{clr}), exp(\eta_2^{clr}), ..., exp(\eta_D^{clr}) \right) \quad (7)$$

As explained in the next subsection, these transformed log-ratio data can be used in traditional regression models to investigate relationships between variables.

E. Compositional Linear Regression Models

In exploring relationships, compositional data can play the role of both independent and dependent variables through constructing an appropriate regression model on the log-ratio scale.

Most methods for classical linear regression have a close analogue in the form of compositional linear regression models [5]. In this paper, compositions (both dependent and independent variables [26]) are used as an alternative to interval regression models where both sets of variables are interval-valued [17], [20], [21], [22], [23].

In general, in univariate (non-compositional) linear regression models, the dependence of one variable Y on another variable X is modelled as follows:

$$Y = a + bX + \epsilon, \tag{8}$$

where Y and X are dependent and independent variables, respectively. The a is the intercept, b is the regression coefficient and ϵ an error term with -generally- 0 mean and variance σ^2.

The aforementioned regression model is built as follows. First, the dependent and independent variables are expressed as compositions. The data samples contain n observations of a D part compositions, resulting in a $n \times D$ matrix. In this paper, the dependent variables are denoted as $\boldsymbol{\Upsilon}$, and independent variables are denoted as $\boldsymbol{X}$. Both dependent and independent variables are transformed into coordinates $\boldsymbol{X}^{clr}, \boldsymbol{\Upsilon}^{clr} \in R^D$ by using the chosen log-ratio transforms (clr in our case) and the following statistical model is constructed:

$$\boldsymbol{\Upsilon}^{clr} = \mathbf{a} + \boldsymbol{X}^{clr}\mathbf{B} + \boldsymbol{\epsilon} \tag{9}$$

The regression parameters can be estimated in the standard way by the least squares method [27].

More details on the log-ratio transformations and compositional linear regression operations can be found in [2], [5], [15], [26].

After having a brief overview of α-cut representation of FSs, compositional data and operations, we now proceed to the motivation and methodology of this paper.

III. MOTIVATION AND METHODOLOGY

In recent years, regression models for data sets where both dependent and independent variables are interval-valued have attracted increasing interest, with a view to improving regression model accuracy and mathematical in-coherency, i.e. ensuring the estimated left endpoint of intervals is smaller than the right endpoints. While the latter poses one of the key challenges for interval regression models, recent studies have started exploring the adoption of a compositional representation to address it and unlock further potential advantages such as ensuring estimations remain within a given variable's domain [13], [15].

In this paper, building on these insights, we extend the compositional representation to FSs by using alpha-cut decomposition. Further, we explore the use of the obtained compositional representations of the FSs to facilitate linear regression of FS-valued data.The steps of this approach are as follows.

A. FSs to Compositional Data

In [15], closed intervals are transformed into their 3-part compositional representations adopting cardinality of each individual intervals on a fixed domain. The authors [15], further applied the (interval) compositional representations on linear regression, articulating how, why and when the approach may be appropriate in terms of maintaining mathematical coherency and avoiding estimations outside of the fixed domain.

Adopting the approach in [15], in this paper, we extend the transformation of compositional data via α-cut decomposition in FSs. As mentioned in Section II-A, the α-cut decomposition theorem is used to obtain closed intervals ($^\alpha[\overline{A}]$) from FSs where $\alpha \in [0,1]$. Thus, first, intervals are obtained via the α-cut decomposition theorem and these obtained intervals are transformed in compositional representation (e.g., $^\alpha\boldsymbol{\eta}$) following the approach in [15], where the corresponding α level is denoted as the left superscript.
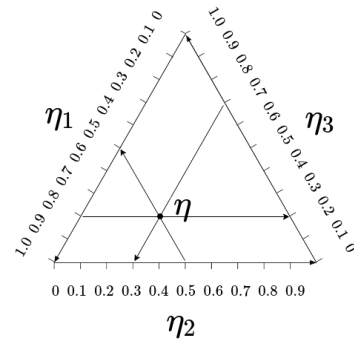
As an illustrative example, in Fig. 3a, on a finite domain X,
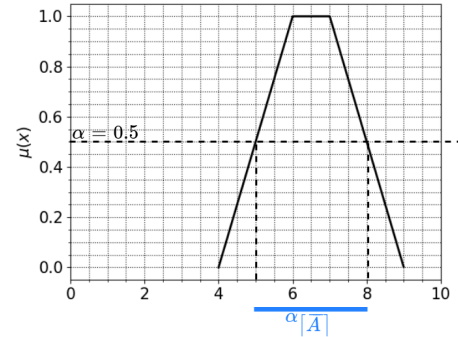
(a) α-cut decomposition of the FS A leads to the interval between 5 and 8
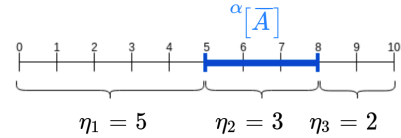
(b) Transformation from the interval-valued data, on a fixed scale, to a compositional representation.

(c) Ternary diagram of the composition $(0.5, 0.3, 0.2)$ here, resulting from the transformation of the interval $5, 8$ to a compositional representation.

Fig. 3: Compositional representation of FSs.

the FS A is decomposed by the $\alpha = 0.5$ to obtain the interval $^{0.5}[\overline{A}] = [5, 8]$. Later, this obtained interval (Fig. 3b on the domain X) is represented as a 3-part composition which is denoted by $^{0.5}\boldsymbol{\eta}$ on a 3-dimensional simplex (Fig. 3c), i.e. $^{0.5}\boldsymbol{\eta} \in S^3$ as follows:

$$^{0.5}\boldsymbol{\eta} = (^{0.5}\eta_1, {}^{0.5}\eta_2, {}^{0.5}\eta_3)|^{0.5}\eta_i > 0, i = 1, 2, 3; \sum_{i=1}^{3} {}^{0.5}\eta_i = \kappa,$$
(10)

where $\kappa = X^+$, i.e. the parts sum to the maximum of the domain and the parts themselves are defined as follows:

$$
\begin{aligned}
^{0.5}\eta_1 &= |\quad X^-, {}^{0.5}[\overline{A}]^- \;| = |[0, 5]| = (5.0) \\
^{0.5}\eta_2 &= |\;{}^{0.5}[\overline{A}]^-, {}^{0.5}[\overline{A}]^+| = |[5, 8]| = (3.0) \\
^{0.5}\eta_3 &= |\;{}^{0.5}[\overline{A}]^+, \quad X^+ \;| = |[8, 10]| = (2.0).
\end{aligned}
$$
(11)

where $\alpha = 0.5$ that decomposes $^{0.5}[\overline{A}]$ from the FS A. The left and right endpoint of this interval are denoted as $^{0.5}[\overline{A}]^-$ and $^{0.5}[\overline{A}]^+$, respectively. Note the cardinality of each component can not be equal or less than 0. This restricts this transformation to intervals $^{0.5}[\overline{A}]$, where $^{0.5}[\overline{A}]^- > X^-$ and $^{0.5}[\overline{A}]^+ < X^+$.

The procedures in [15] is extended to α-cut decomposition, providing intervals of FSs which in turn transform into compositional data as exemplified above. As the next step of this paper methodology, the transformed compositional representations are processed through to the compositional linear.

B. Compositional Linear (Interval) Regression

As detailed in Section II-E, the transformed compositional data are used to generate coefficients of compositional linear regression models.

Consider two sets of FSs (dependent and independent) are transformed to their compositional representation as outlined in Section III-A. The thus obtained compositional data are transformed by using the log-ratio transforms. As mentioned in section II-D, various log-ratio transform techniques exist in the literature and in this paper, the *clr* transformation is performed, as it focuses on geometric mean proportion of three compositions (5).

The approach can be divided into seven main steps and is illustrated in the flow chart in Fig. 4. Let two FSs A and B, instances of independent and dependent variables respectively, for the compositional linear regression. First, α-cut decomposition intervals ($^{\alpha}[\overline{A}]$ and $^{\alpha}[\overline{B}]$ where $\alpha \in [0, 1]$) are obtained (Step 1 in Fig. 4). These intervals are transformed to their compositional representation, denoted by $^{\boldsymbol{\alpha}}\boldsymbol{H}$ and $^{\boldsymbol{\alpha}}\boldsymbol{Z} \in S^D$, following the procedures in Section III-A (Step 2 in Fig. 4).

Next, the *clr* log-ratio transforms are applied on the obtained compositions representations ($^{\boldsymbol{\alpha}}\boldsymbol{H}^{clr}$ and $^{\boldsymbol{\alpha}}\boldsymbol{Z}^{clr} \in R$), as Step 3 in Fig. 4. These log-ratio values are processed in the regression model to calculate coefficients (Step 4 in 4), as detailed in Section II-E. Lastly, the estimations of the regression models are inverted back to compositional data (Step 5 in Fig. 4) which in turn is transformed back to intervals (Step 6 in Fig. 4) and 're-assembled' in conjunction with their α-level to FSs (Step 7 in Fig. 4). Note that in this initial exploration of the compositional representation of FSs for linear regression, we adopt a naive approach, where α-cuts are obtained and processed independently for each α-cut level.

IV. EXPERIMENTS AND ILLUSTRATIONS

In Section III, we discussed the compositional representation of FSs and adopting the compositional regression on those calculated representations. In order to explore the behaviour of the model, two experiments are carried out. In the first experiment, the coefficients are calculated and the model is tested by using the same FSs which are used in the training phase. In the second experiment, the obtained same coefficients are tested by a different FS which is not involved in the coefficient generation phase.

A. Experiment 1

To illustrate the approach, we generate two sets of five synthetic FSs as independent and dependent variables denoted
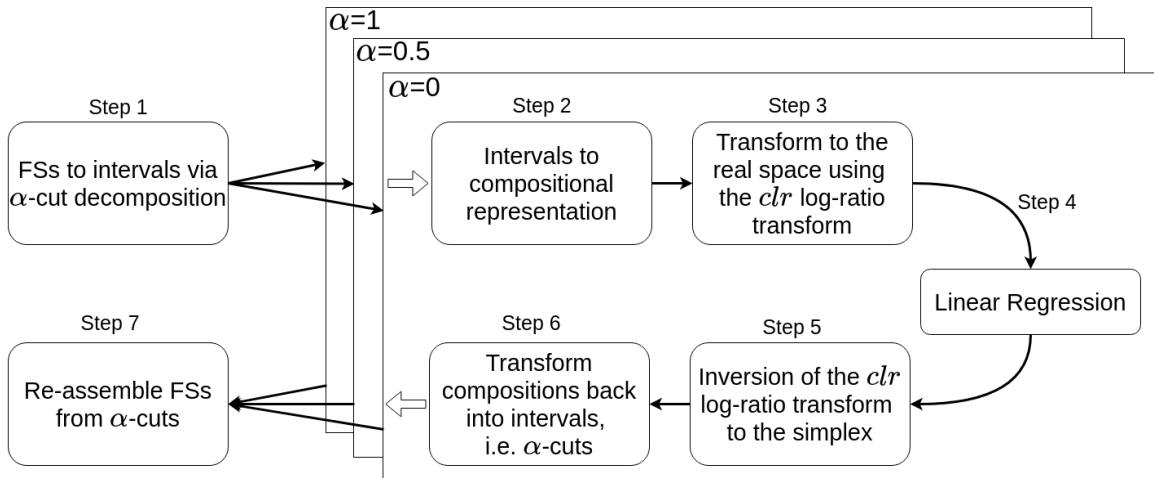

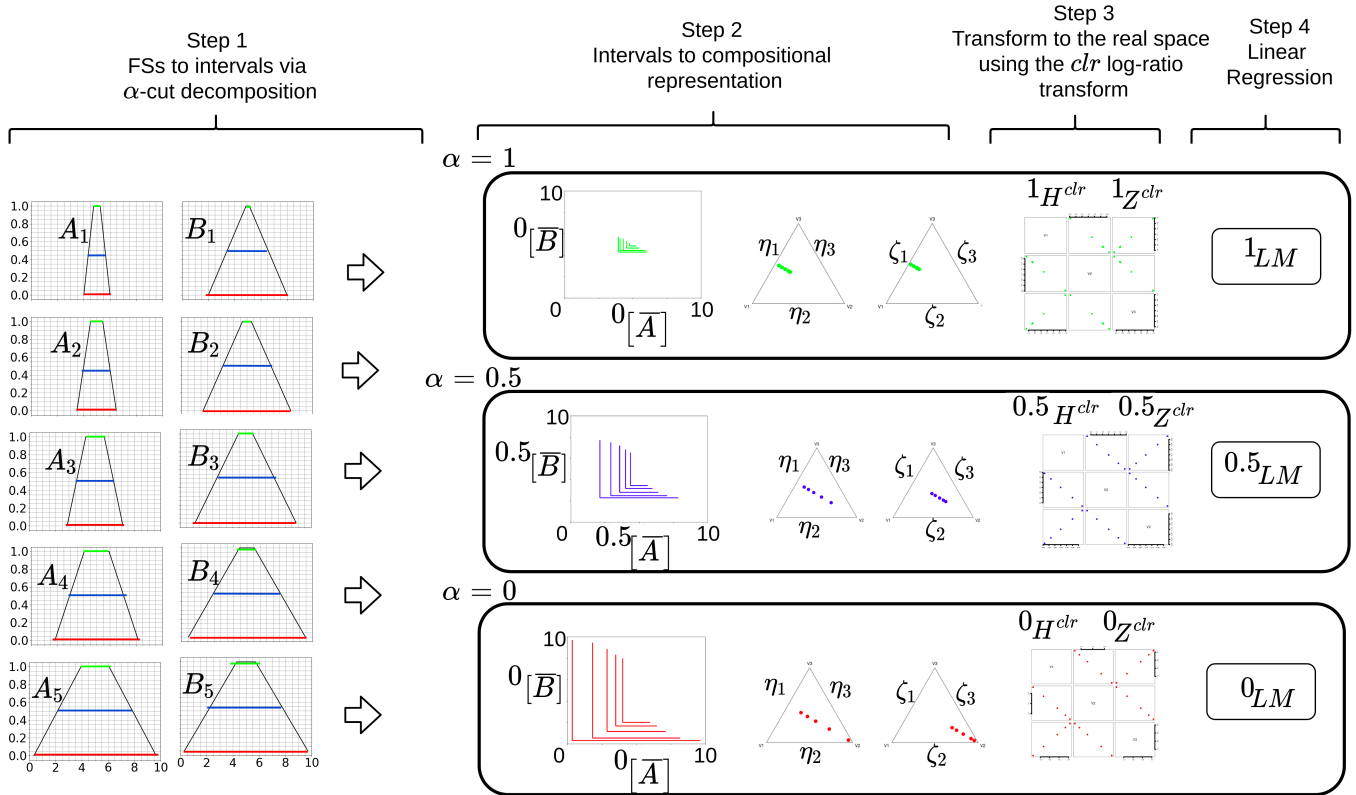Fig. 4: The demonstration of the proposed approach.

Fig. 5: The first four experimental steps of the proposed approach.

A and **B**, shown in Fig. 5. We explore adopting linear regression on the compositional representation of the α-cut decompositions of the FSs to conduct regression effectively, mitigating some of the challenges of maintaining domain and parameter coherence.

As can be seen in Fig. 5, each dependent and independent variable FSs' support is gradually increased to explore and to clearly communicate the behaviour of the regression models in respect to different properties of the FSs. In other words, beyond considering model performance, we are conducting an analysis on the resulting regression models for systematic variations of the FSs.

The specific experimental steps are as follows, and each is visually illustrated in Fig. 5:

Step 1: To enable visualisation within the paper, we conduct α-cut decomposition, using only three levels (at $\alpha = 0+, \alpha = 0.5$ and $\alpha = 1$), thus obtaining three set of intervals for each instance of the independent and dependent variables ($^{\alpha}[\overline{A}], ^{\alpha}[\overline{B}]$).

Step 2: The intervals from Step 1 are transformed to the compositional representation (denoted by $^{\alpha}H$ and $^{\alpha}Z \in S^3$). Compositional linear regression is performed by using the 'Compositional' library in R language [28].

Step 3: The clr log-ratio transform is applied on the obtained compositional representations and the $^{\alpha}H^{clr}$ and $^{\alpha}Z^{clr} \in R$ are calculated.

Step 4: Following the procedure in Section II-E, the co-

efficients are obtained using the linear regression approach applied to the transformed valued generated in Step 3. Note that for each α-cut level, a different linear regression model (denoted by $^{\alpha}LM$) is optimised and, thus, a different set of coefficients is calculated. As an example, the $^{0.5}LM$ is given in (12).

$$
\begin{aligned}
^{0.5}\widehat{\zeta_1^{clr}} &= -0.1641 + 0.4420 * {}^{0.5}\eta_1^{clr} \\
^{0.5}\widehat{\zeta_2^{clr}} &= 0.3282 + {-0.8840} * {}^{0.5}\eta_1^{clr} \\
^{0.5}\widehat{\zeta_3^{clr}} &= -0.1641 + 0.4420 * {}^{0.5}\eta_1^{clr}
\end{aligned} \quad (12)
$$

Note that η_2^{clr} and η_3^{clr} are excluded/dropped to avoid having mutually dependent components in the regression stage in (12). η_2^{clr} and η_3^{clr} could be excluded and later recreated at the estimated unconstrained real space stage–as for clr, all three components add up to zero. Excluding η_2^{clr} and η_3^{clr} would be computationally more efficient as it avoids the estimation of its model parameters, but it makes the pipeline more cumbersome as the recreation requires adjustment depending on the transform used.

With the compositional regression model in place, we can now estimate an output for a given FSs **A**. In order to so, first, we follow the same four steps in Fig. 5 to obtain the $^{\alpha}H^{clr}$.

Step 5: We use the models ($^{\alpha}LM$) in conjunction with the $^{\alpha}H^{clr}$ to estimate the $^{\alpha}\widehat{Z^{clr}}$ for each α-cut level decomposition. Later, the inverse operation (7) is applied on the estima-
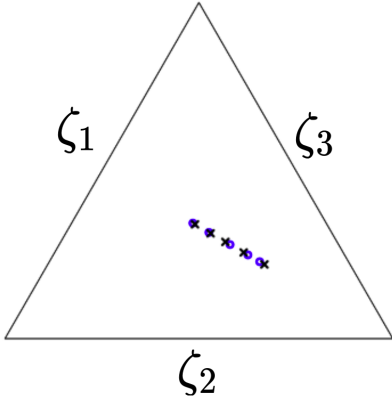
Fig. 6: The compositional linear regression $^{0.5}LM$ estimation results (black crosses $\widehat{^{0.5}Z}$) on ground truth (blue circle $^{0.5}Z$), correspond to Step 5 in Fig. 4

Fig. 7: Back transformation of $^{0.5}LM$ estimation results into interval-valued data (black dashed lines $\widehat{^{0.5}[B]}$) over the ground truth intervals (blue lines $^{0.5}[B]$), correspond to Step 6 in Fig. 4.

tions $clr^{-1}(\widehat{\alpha Z^{clr}})$ and the compositional representation $\widehat{\alpha Z}$ is obtained.

To provide visual insight of the quality of the regression model, the estimated compositions ($\widehat{^{0.5}Z}$) from the $^{0.5}LM$ are illustrated as black crosses in the ternary diagram in Fig.6, with the blue circles representing the ground truth compositions $^{0.5}Z$.

Step 6: Each estimated composition is transformed into intervals or more specifically α-cut $\widehat{^{\alpha}[B]}$, where $\alpha = (0+, 0.5, 1)$. The estimated $\widehat{^{0.5}[B]}$ intervals are illustrated as black dashed-lines in Fig. 7 where the blue lines represent the ground truth $\widehat{^{0.5}[B]}$.

Step 7: As the final step, all the α-cut levels ($\alpha = (0+, 0.5, 1)$) estimations are combined, resulting in the FSs $\widehat{\mathbf{B}}$, as shown in Fig. 6.

In the testing phase of the generated coefficients, the same FSs ($\mathbf{A}$) are used as input to the built linear regression models and the estimations results (black-dashed intervals) are visualised on the given input FSs in Fig. 8. As the initial results indicate that all the estimations are in-line with the fixed domain and follow the mathematical coherency (the left endpoints are smaller than the right endpoints) where the models meet intuitive expectations.

B. Experiment 2

In experiment 1, the generated FS pairs ($\mathbf{A} - \mathbf{B}$) are used to optimise coefficients and the same FSs $\mathbf{A}$ are tested on the generated coefficients to explore the estimation behaviour of the model.

In experiment 2, we conduct the analysis with a different FS to further explore the behaviour of the models (e.g., $^{0.5}LM$ is given in (12)). First, we generate the Gaussian shape FS I (shown on the left-hand side of the Fig. 9) which is not involved in the coefficient generation process. Following the same 7 steps in Fig. 4, the I transform into the clr log-ratio and the same linear regression model coefficients are used to generate the estimation. Later, the back transformation is carried out, and estimations are re-assemble to visualise the
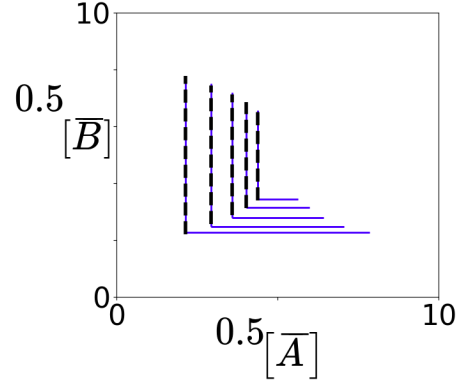
result ($\hat{I}$), which is shown as black dashed lines on the right-hand side of Fig. 9.

Overall, based on the experiment, it can be observed that the estimations go beyond the given FS (I). We highlight the expected potential of the approach: having the estimations in-line with the fixed domain $[0, 10]$ and avoiding the risk of parameter flipping where both of the key issues are addressed in the given results.

V. CONCLUSIONS

Recent studies have started exploring adopting a compositional representation for data reflecting uncertainty or vagueness, for example, interval-valued questionnaire responses. The further effort has been made in articulating how regression for these compositional representations can be conducted and how this can provide advantages in mitigating some of the challenges faced by traditional approaches. This paper proposes the α-cut based compositional representation of FSs and conducts an initial empirical exploration of linear regression using the resulting FS representations.

In the experiments, initial explorations carry out to examine the behaviour of the proposed approach which shows promising results in principle. However, we highlight that substantial questions and challenges remain. For example, further experiments and a formal exploration of whether the compositional representation and associated regression addresses key concerns such as mathematical coherence are needed. Moreover, the simple approach of generating independent regression models for each α-cut, as outlined here, is expected to risk challenges, such as resulting in α-cuts at a higher α level not being a subset of lower α-cuts.

Thus, intuitively, in future work, we will further explore the expected potential of this approach in addressing two key challenges of the risk of parameter flipping and the unexpected generation of estimations outside the dependent variable's domain. Furthermore, we will examine *dependency* across multiple linear regressions associated with different α levels.
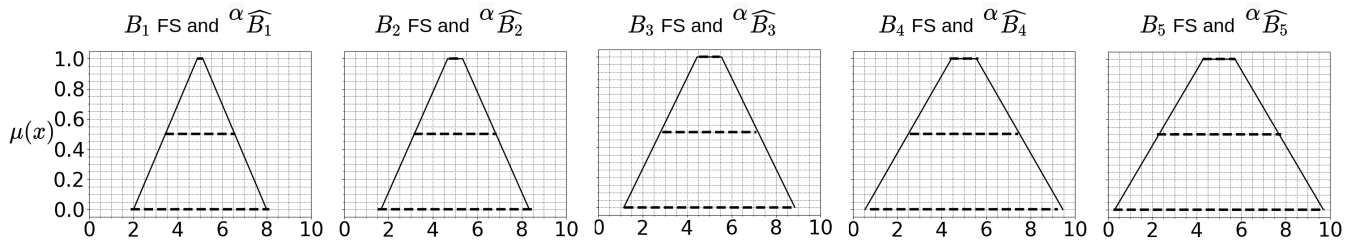
Fig. 8: Re-assembling FSs from α-cuts from $^{\alpha}LM$ estimations into interval-valued data (black dashed lines), correspond to Step 7 in Fig. 4.
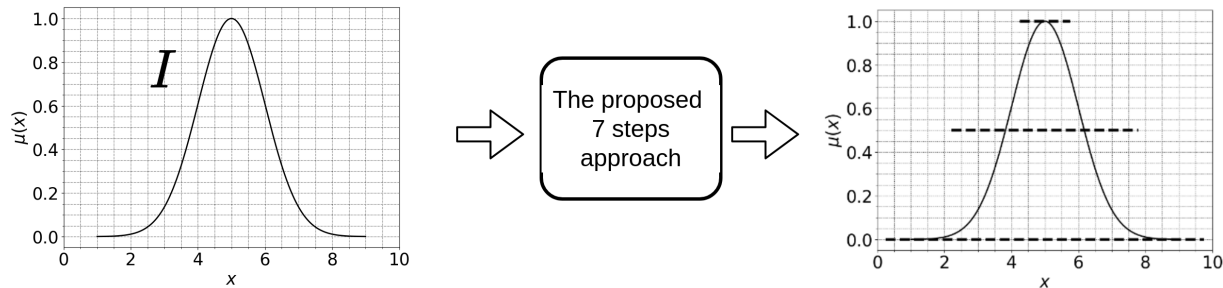

Fig. 9: Experiment 2 - Re-assembling estimation results from α-cuts in $^{\alpha}LM$ to interval-valued data (black dashed lines), correspond to Step 7 in Fig. 4.

REFERENCES

[1] P. Lewi, "Spectral mapping, a technique for classifying biological activity profiles of chemical compounds," *Arzneimittel-Forschung*, vol. 26, no. 7, p. 1295—1300, 1976. [Online]. Available: http://europepmc.org/abstract/MED/1036914

[2] J. Aitchison, "The statistical analysis of compositional data," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 44, no. 2, pp. 139–160, 1982.

[3] Ž. Pedišić, "Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research—the focus should shift to the balance between sleep, sedentary behaviour, standing and activity," *Kinesiology*, vol. 46, no. 1., pp. 135–146, 2014.

[4] A. Van Wamel, H. Kroon, and S. Van Rooijen, "A concise guide to compositional data analysis," *Mental Health and Substance Use: Dual Diagnosis*, vol. 2, no. 2, p. 101–110, 2009.

[5] K. G. Van den Boogaart and R. Tolosana-Delgado, *Analyzing compositional data with R*.   Springer, 2013, vol. 122.

[6] E. J. M. Carranza, "Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values," *Journal of Geochemical Exploration*, vol. 110, no. 2, pp. 167–185, 2011.

[7] G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, and J. J. Egozcue, "Microbiome datasets are compositional: and this is not optional," *Frontiers in microbiology*, vol. 8, p. 2224, 2017.

[8] R. Knight, A. Vrbanac, B. C. Taylor, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L.-I. McCall, D. McDonald *et al.*, "Best practices for analysing microbiomes," *Nature Reviews Microbiology*, vol. 16, no. 7, pp. 410–422, 2018.

[9] P. Vermeesch, "Multi-sample comparison of detrital age distributions," *Chemical Geology*, vol. 341, pp. 140–146, 2013.

[10] G. J. Weltje and R. Tjallingii, "Calibration of xrf core scanners for quantitative geochemical logging of sediment cores: Theory and application," *Earth and Planetary Science Letters*, vol. 274, pp. 423–438, 2008.

[11] S. F. Chastin, J. Palarea-Albaladejo, M. L. Dontje, and D. A. Skelton, "Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach," *PloS one*, vol. 10, no. 10, p. e0139984, 2015.

[12] V. Carson, M. S. Tremblay, J.-P. Chaput, and S. F. Chastin, "Associations between sleep duration, sedentary time, physical activity, and health indicators among canadian children and youth using compositional analyses," *Applied Physiology, Nutrition, and Metabolism*, vol. 41, no. 6, pp. S294–S302, 2016.

[13] M. Smithson and S. Broomell, "Compositional data analysis tutorial," *submitted to Psychological Methods*, 2021.

[14] Z. Ellerby, C. Wagner, and S. B. Broomell, "Capturing richer information: On establishing the validity of an interval-valued survey response mode," *Behavior Research Methods*, pp. 1–23, 2021.

[15] D. Pekaslan and C. Wagner, "Compositional linear regression on interval-valued data," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*.   IEEE, 2021, pp. 01–08.

[16] L. Billard and E. Diday, "Regression analysis for interval-valued data," in *Data Analysis, Classification, and Related Methods*.   Springer, 2000, pp. 369–374.

[17] ——, "Symbolic regression analysis," in *Classification, Clustering, and Data Analysis*.   Springer, 2002, pp. 281–288.

[18] E. d. A. L. Neto and F. d. A. de Carvalho, "Centre and range method for fitting a linear regression model to symbolic interval data," *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1500–1515, 2008.

[19] S. Kabir and C. Wagner, "Interval-valued regression-sensitivity to data set features," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.   IEEE, 2021, pp. 1–7.

[20] S. Kabir, C. Wagner, and Z. Ellerby, "Towards handling uncertainty-at-source in ai – a review and next steps for interval regression," 2021.

[21] E. d. A. L. Neto and F. d. A. de Carvalho, "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics & Data Analysis*, vol. 54, no. 2, pp. 333–347, 2010.

[22] Y. Sun and D. Ralescu, "A linear model for interval-valued data," 2015.

[23] L. C. Souza, R. M. Souza, G. J. Amaral, and T. M. Silva Filho, "A parametrized approach for linear regression of interval data," *Knowledge-Based Systems*, vol. 131, pp. 149–159, 2017.

[24] G. J. Klir and B. Yuan, "Fuzzy sets and fuzzy logic: theory and applications," *Upper Saddle River*, p. 563, 1995.

[25] H. Hamrawi, S. Coupland, and R. John, "Type-2 fuzzy alpha-cuts," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 3, pp. 682–692, June 2017.

[26] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelo-Vidal, "Isometric logratio transformations for compositional data analysis," *Mathematical geology*, vol. 35, no. 3, pp. 279–300, 2003.

[27] N. R. Draper and H. Smith, *Applied regression analysis*.   John Wiley & Sons, 1998, vol. 326.

[28] M. Tsagris, G. Athineou, A. Alenazi, C. Adam, M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *Compositional: Compositional Data Analysis*, 2022.