ISSN Printed 1592-7415 ISSN Online 2282-8214

Number 30 - 2016

RATIO MATHEMATICA

Journal of Foundations and Applications of Mathematics

Publisher

A.P.A.V.

Accademia Piceno – Aprutina dei Velati in Teramo

Ratio Mathematica

Journal of Foundations and Applications of Mathematics

Aims and scope: The main topics of interest for Ratio Mathematica are:
Foundations of Mathematics, Applications of Mathematics, New theories and applications, New theories and practices for dissemination and communication of mathematics.
Papers submitted to Ratio Mathematica Journal must be original unpublished work and should not be in consideration for publication elsewhere.

Instructions to Authors: the language of the journal is English. Papers should be typed according to the style given in the journal homepage (http://eiris.it/ratio_mathematica_scopes.html)

Submission of papers: papers submitted for publication should be sent electronically in Tex/Word format to one of the following email addresses:

- fabmatu@gmail.com
- antomato75@gmail.com
- giuseppemanuppella@gmail.com
- sarka.mayerova@seznam.cz

Journal address: Accademia Piceno – Aprutina dei Velati in Teramo – Pescara (PE) - Italy, Via del Concilio n. 24
Web site: www.eiris.it – www.apav.it

Cover Making and Content Pagination: Fabio Manuppella
Editorial Manager and Webmaster: Giuseppe Manuppella
Legal Manager (Direttore Responsabile): Bruna Di Domenico

ISSN 1592-7415 [Printed version] | ISSN 2282-8214 [Online version]

RATIO MATHEMATICA
ISSUE N. 30 (2016) pp. 3-21

ISSN (print): 1592-7415
ISSN (online): 2282-8214

A geometric view on Pearson's correlation coefficient and a generalization of it to non-linear dependencies

Priyantha Wijayatunga
Department of Statistics, Umeå School of Business and Economics,
Umeå University, Umeå 901 87, Sweden
priyantha.wijayatunga@umu.se

Abstract

Measuring strength or degree of statistical dependence between two random variables is a common problem in many domains. Pearson's correlation coefficient ρ is an accurate measure of linear dependence. We show that ρ is a normalized, Euclidean type distance between joint probability distribution of the two random variables and that when their independence is assumed while keeping their marginal distributions. And the normalizing constant is the geometric mean of two maximal distances; each between the joint probability distribution when the full linear dependence is assumed while preserving respective marginal distribution and that when the independence is assumed. Usage of it is restricted to linear dependence because it is based on Euclidean type distances that are generally not metrics and considered full dependence is linear. Therefore, we argue that if a suitable distance metric is used while considering all possible maximal dependences then it can measure any non-linear dependence. But then, one must define all the full dependences. Hellinger distance that is a metric can be used as the distance measure between probability distributions and obtain a generalization of ρ for the discrete case.

Keywords: metric/distance; probability simplex; normalization.

2010 AMS subject classifications: 62H20

doi: 10.23755/rm.v30i1.5

1 Introduction

Measuring association between two random quantities is of interest in many types statistical analyses and applications in various disciplines. Pearson's product moment correlation coefficient is the standard in statistical textbooks and applications for measuring linear association. And Spearman's rank correlation coefficient is capable of measuring any monotonic dependence between two random variables. For two ordinal variables Cramér's V-statistic is widely used whereas Tchuprow's T-statistic is less-known and therefore less often used (see [14] and references therein). Furthermore, there are many other kinds of dependence measures used in statistical literature, especially in applied statistical analyses. In statistical genetics for evaluation of linkage disequilibrium between genetic markers, authors of [2] use volume tests that are discussed in [10] as a measures of dependence between ordinal variables with fixed margins. For massive datasets in [8] it is used mutual information dimension that is defined in terms of information dimension descried in [1].

In [9] it is said that *"although it is customary in bivariate data analysis to compute a correlation measure of some sort, one number (or index) alone can never fully reveal the nature of dependence; hence a variety of measures are needed"*. It is also stated therein that *"if (two quantities are) not totally dependent, then it may be helpful to find some quantities that can measure the strength or degree of dependence between them"*. In this article we try to develop a measure that can indicate 'the' degree or strength of association between two discrete variables. Our measure can be seen as a generalization of the Pearson's correlation coefficient ρ using a suitable distance metric between joint probability distributions, instead of simple Euclidean type distances that are used in ρ (see below). Given the joint probability distribution (jpd) of two discrete variables, say, X and Y, the degree of dependence (also called association) between them is expressed as the normalized distance between the jpd of them and that of when the independence of them is assumed. The associated normalizing constant is geometric mean of distances between the latter and all possible jpds where full dependence between X and Y is assumed while retaining each marginal distribution at a time. These latter distances are in fact the maximal distances since we obtain them by assuming full dependence. In the following we show that the Pearson's correlation coefficient is measure of this nature based on some Euclidean type distances. That is, it is the ratio of the distance between dependence and independence, and the geometric mean of the distances that are between full linear dependences and independence. Therefore, our measure can be regarded as a generalization of ρ using a suitable distance between probability distributions and considering non-linear dependencies. One thing that ρ shows us is that if we need to define a strength of a dependence then we must find or hypothesize the full dependence(s) corre-

sponding to the given dependence. This aspect can make numerical evaluation of the measure algorithmic or computational since sometimes it may not be possible to obtain the full dependences easily. However, here we do not deal with such computational issues but our consideration is on defining a measure following the structure of ρ. For a given dependence (in terms of a jpd) finding efficiently related jpds representing the full dependences that preserve either of the marginal is an open problem.

First we show that, in the simple case of binary X and Y, the ρ measures the degree of dependence with a certain type of Euclidean distance, but for multinary case (and also for continuous variables) a distance in terms another type of Euclidean area is used. But these Euclidean type distances are appropriate for measuring only linear dependences. Since we are interested in measuring any non-linear dependence we propose to use Hellinger distance between joint probability distributions, that is called as Matsusita distance in the discrete (see [6]). The Hellinger distance is a metric and it possesses the so-called linear invariance properties, so it is more suitable for measuring distances between the probability distributions. Therefore, it can be used to measure any type of dependence.

2 Pearson's correlation coefficient ρ

For random variables X and Y, the Pearson's correlation coefficient $\rho(X, Y)$ is such that $|\rho(X, Y)| \leq 1$. The equality holds if and only if X and Y are fully linearly dependent and $\rho(X, Y) = 0$ if they are linearly independent. And the converse of the latter is not always true unless X and Y are binary. Note that the full dependence is linear in the binary (also called 2×2) case where then the $\rho(X, Y)$ is often called ϕ-coefficient.

2.1 2×2 case: ϕ-coefficient

Let X and Y be two binary variables with a common state space $\{0, 1\}$ where their jpds and marginal probability distributions are written as $p_{xy} = p(X = x, Y = y)$, $p_x = p(X = x)$ and $q_y = p(Y = y)$ for $x, y = 0, 1$. Let $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ for short. As shown in [12], any such P can be represneted as a point in the probability simplex shown in the Figure 1. The jpd of X and Y under the assumption that they are independent while keeping the marginal distributions fixed is $P^I = \begin{pmatrix} p_0 q_0 & p_0 q_1 \\ p_1 q_0 & p_1 q_1 \end{pmatrix}$ and the set of such probability distributions for all P makes a surface (shown by lines) in the probability simplex. The ϕ-coefficient

of X and Y is defined by

$$\phi = \frac{p_{11} - p_1 q_1}{\sqrt{p_1(1-p_1)q_1(1-q_1)}},$$

which is a measure of degree of association between X and Y. Now let X and Y be positively correlated, then there are two jpds under the assumption that the two variables are fully dependent. They are $P^X = \begin{pmatrix} p_0 & 0 \\ 0 & p_1 \end{pmatrix}$ and $P^Y = \begin{pmatrix} q_0 & 0 \\ 0 & q_1 \end{pmatrix}$, where P^X is when the marginal distribution of X is preserved and P^Y is when the marginal distribution of Y is preserved. Note that each full dependence is obtained from P while preserving respective marginal distribution, then the marginal distribution of the other variable should be assumed by it. Therefore in these cases, the full dependence is essentially linear.

For a generalization of ρ to measure 'any' type of dependence we need to look at its structure and construction. First we consider the case of two binary variables by examining the ϕ-coefficient. Let $D_{P^I,P}$ be $p_{11} - p_1 q_1$ that is the $(2,2)^{th}$ component Euclidean distance between the two probability distributions P^I and P. It is a measure of how far the dependence (under P) from the independence (under P^I) when marginals of X and Y are fixed. Note that in the 2×2 case it is sufficient to consider a single component difference (between the two probability matrices) since all the components have same absolute difference. Similarly, we have $D_{P^I,P^X} = p_1(1-q_1)$ and $D_{P^I,P^Y} = q_1(1-p_1)$. Since P^X and P^Y are the two full dependences that we can obtain from P while preserving respective marginal in each case, we have that $D_{P^I,P} \leq D_{P^I,P^X}$ and $D_{P^I,P} \leq D_{P^I,P^Y}$. In fact $D_{P^I,P} = p_{11} - p_1 q_1 = p_1(p_{11}/p_1 - q_1) \leq p_1(1-q_1) = D_{P^I,P^X}$ since $p_1 \geq p_{11}$ and similarly the other inequality. It is easy to see that the denominator of the ϕ-coefficient is the geometric mean of D_{P^I,P^X} and D_{P^I,P^Y} (the two maximal distances) and the numerator is $D_{P^I,P}$. Therefore, the ϕ-coefficient can be thought of as the normalized distance between P and P^I where the normalizing constant is the geometric mean of the two maximal distances. Hence the ϕ-coefficient is 1 if and only if $P = P^X = P^Y$ (full dependence) and it is 0 if and only if $P^I = P$ (independence).

2.2 $n \times m$ case

Let X and Y be two multinary random variables where their state spaces are $\{0, 1, .., n-1\}$ and $\{0, 1, .., m-1\}$ respectively for $n, m > 2$. For any given jpd of X and Y, $P = (p_{00}, ..., p_{0(m-1)}; p_{10}, ..., p_{1(m-1)}; ...; p_{(n-1)1}, ..., p_{(n-1)(m-1)})$ where $p_{ij} = p(X = i, Y = j)$ for $i = 0, .., n-1$ and $j = 1, ...m-1$, we define the probability simplex, $\Delta = \{P = (p_{ij})_{n \times m} : \sum_{ij} p_{ij} = 1, p_{ij} \geq 0; i =$

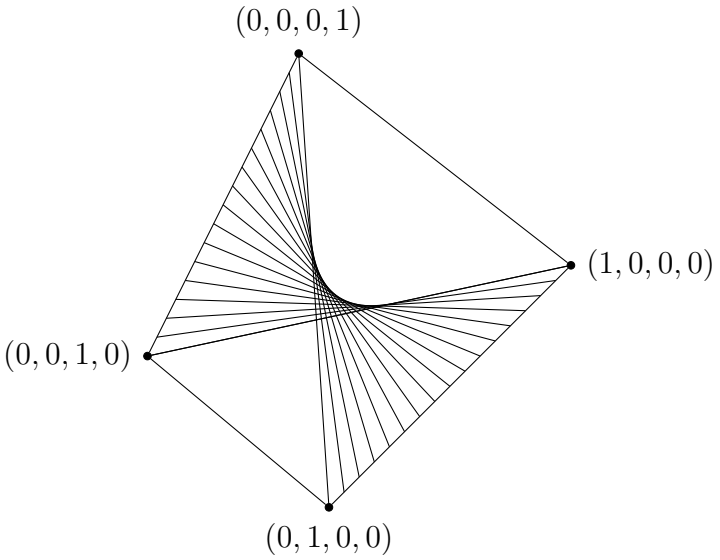

Figure 1: Probability simplex for binary X and Y where their jpd $P = (p_{00}, p_{10}, p_{01}, p_{11})$ is a point in it. Any jpd on surface shown by lines represents independence of X and Y.

$0, 1, .., n-1; j = 0, 1, ..., m-1\}$ similar to the case of two binary random variables. But here visualization of it is more difficult. Recall that $\rho(X,Y) = cov(X,Y)/\sqrt{var(X)Var(Y)}$, where

$$cov(X,Y) = \sum_{x,y} xyp(x,y) - \sum_{x} xp(x) \sum_{y} yp(y)$$

and

$$var(X) = \sum_{x} x^2 p(x) - \{\sum_{x} xp(x)\}^2.$$

In the following we try to visualize the ρ and its structure for understanding how it measures the dependence.

Let us take the case where $n = m$, thus allowing us to have perfect (one-to-one) dependence between X and Y, linear or non-linear. It can be seen that when X and Y are assigned to two perpendicular axes, $cov(X,Y)$ is area difference between two rectangular Euclidean areas, that is shown as the dark area in the Figure 2. The first area (i.e., $\sum_{x,y} xyp(x,y)$) is the weighted average area created by the values of X and Y, where, for each component area that is being weighted is with side lengths $X = x$ and $Y = y$ and its weight is the respective joint probability of $X = x$ and $Y = y$, i.e., $p(X = x, Y = y)$. This area represents the

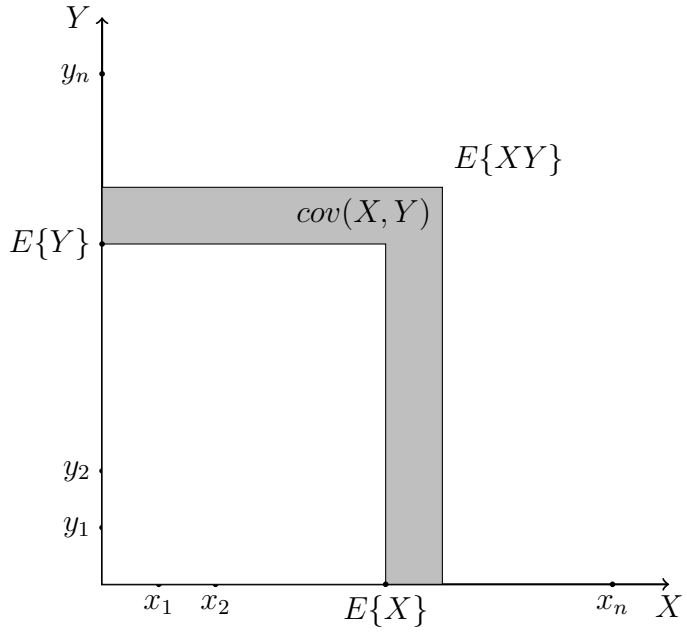

Figure 2: Covariance of X and Y is the weighted averaged Euclidean area difference.

dependence between X and Y. And the second area (i.e., $\sum_x xp(x) \times \sum_y yp(y)$) is the area created by the side lengths that are the weighted average of values of X (i.e., $E\{X\}$) and that of Y (i.e., $E\{Y\}$) where the weights are the respective marginal probabilities. Since the lengths or values $E\{X\}$ and $E\{Y\}$ are also on same axes as X and Y are, respectively, we can see the difference of the two areas. Note that it can be seen that the second area (i.e., $\sum_{x,y} xyp(x)p(y)$) is also calculated in the similar way as the first, but assuming the independence of X and Y, i.e., it is the weighted average area created by the values of X and Y, where for each component area that is being weighted is with side lengths $X = x$ and $Y = y$ and the weight associated with it is the respective joint probability of $X = x$ and $Y = y$ assuming independence $p(X = x, Y = y) = p(X = x)p(Y = y)$. So the second area represents the scenario of the independence of X and Y. Therefore one can view that the two areas refer to those when a dependence between X and Y is assumed and when their independence is assumed while keeping the marginal distributions fixed, therefore $cov(X, Y)$ is a 'distance' in terms of a Euclidean area difference between dependence and independence of the two variables.

Moreover $var(X)$ can be interpreted in the same way. Now X is assumed to be on both axes meaning that Y is replaced by X (taken as if Y were X). This is a context of assuming a full dependence of X and Y when the marginal of X is

preserved. Assuming one variable by the other is 'a way' to consider a case of full dependence between the two variables. Then we are assuming the marginal of Y by that of X. This assumption is easily seen when both variables have same sizes in their state spaces but it is hard to see when they are different. So the $E\{X^2\}$ is indicated by the weighted average area that we obtain when Y is X where weight for each component area x^2 is $p(x,y) = p(x)$, i.e., when the marginal of X is preserved. This is a sensible area under full dependence. And $E\{X\}^2$ is indicated by the area when the respective weight is $p(x)p(y) = p(x)^2$ where $x = y$. This is a hypothetical case where it is taken as if Y were X, yet their joint probability is taken as if they were independent. So, $var(X)$ is deviation of the full dependence from independence if Y were X. And the same interpretation applies for $var(Y)$.

Thus, $\rho(X,Y)$ is the normalized area difference referring to $cov(X,Y)$ with the normalizing constant being the geometric mean of the two maximal area differences referring to $cov(X,Y)$ where they are such that, one is when Y is assumed to be X (i.e., $var(X)$) and the other is when X is assumed to be Y (i.e., $var(Y)$). That is, the normalizing constant is obtained by assuming the full dependence between X and Y. However the full dependence quantified in this way is appropriate only for doing so for linear dependences. Since there are two such cases of full linear dependence the geometric mean of these two maximal area differences is taken. Note that the above interpretation is valid for the case of X and Y have continuous state spaces.

One thing that we need to show is that $cov(X,Y)$ is maximal (or minimal) when X and Y are strictly monotonically related, for example, linearly related positively (negatively), among all cases of full ono-to-one dependencies between X and Y for fixed maginals of X and Y. This indicates that ρ is not able to identify non-monotonic relations since their covariance values can not be ordered. To see that $cov(X,Y)$ is maximal when Y is strictly increasing with X, let $\mathcal{X} = \{a_1 < ... < a_n\}$ be the state space of X and $\mathcal{Y} = \{b_1 < ... < b_n\}$ be that of Y. Then considering inequalities $(a_i - a_j)(b_i - b_j) > 0$ for $i,j = 1,...,n$ (i.e., we have $a_ib_i + a_jb_j > a_ib_j + a_jb_i$) it can be shown that $\sum_i a_ib_i > \sum_{i,j:j=f(i)} a_ib_j$ where f is any one-to-one function from $\mathcal{X}$ to $\mathcal{Y}$ such that $f(i) \neq i$ for at least two distinct values of i (i.e., f is not a strictly increasing function of i). Now if the marginals of X and that of Y are $(p_1,...,p_n)$ and $(q_1,...,q_n)$, where $p_i = q_i$ for all $i = 1,...,n$ when Y is monotonically increasing with X and otherwise $p_i = q_j$ for some appropriate $i \neq j$ for $i,j = 1,....,n$, then $\sum_i a_ib_ip_i > \sum_{i,j:j=f(i)} a_ib_jp_i$ meaning that $E\{XY_M\} \geq E\{XY\}$ where Y_M is Y when it is strictly increasing with X. This implies that $cov(X,Y_M) \geq cov(X,Y)$ for fixed marginals of X and Y. Therefore, for discrete X and Y, $\rho(X,Y)$ is maximal when Y is strictly increasing in X, among all one-to-one relationships between them. So, if this is the case $\rho(X,Y) = 1$ (maximal) since $cov(X,Y) \leq var(X)$ and $cov(X,Y) \leq var(Y)$.

3 Some other popular measures of dependence

There are a few popular measures of dependence that have similar structure in their definition. We review them briefly by giving some interpretations that support our definition of dependence measure.

3.1 Spearman's rank correlation coefficient ρ^s

In many statistical analyses, especially for non-normal data a popular measure of dependence between two random variables, say, X and Y, is the Spearman's rank correlation coefficient.

$$\rho^s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where $d_i = x_{(i)} - y_{(i)}$ and $x_{(i)}$ is the i^{th} smallest value in the data sample of X and similarly for $y_{(i)}$. It is obvious that $\rho^s = 1$ if and only if two components of data pair (x_i, y_i) has the same ranking, for all data pairs since then $d_i = 0$ for all i. And one can see that for a perfect negative dependence $\sum_{i=1}^{n} d_i^2$ should be its maximal value that is $n(n^2-1)/3$ in order to get $\rho^s_{X,Y} = -1$. Therefore the normalizing constant is taken as $n(n^2-1)/6$ but due to the structure of the definition of the coefficient it is applied to the term $\sum_{i=1}^{n} d_i^2$. Therefore the ρ^s is an accurate measure any monotonic dependence between the two variables. However, when the two variables are not having a strictly monotonic relationship the measure can not give a correct picture of the dependence.

3.2 Information theoretic measures

Another popular measure of dependence, especially in machine learning literature and applied statistics is so-called mutual information (see, for example, [11]). For discrete random variables X and Y, it is defined as

$$I(X,Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}$$

and furthermore, conditional mutual information between X and Y given another variable Z is defined as

$$CI(X,Y,Z) = \sum_{x,y,z} p(x,y,z) log \frac{p(x,y|z)}{p(x|z)p(y|z)} \quad (1)$$

If X and Y are independent then the $I(X,Y) = 0$ and if X and Y are conditionally independent given Z then the $CI(X,Y,Z) = 0$. In fact, these dependence measures are also based on so-called Kullback-Leibler (KL) distance or

rather divergance, [13]. It is easy to see that $I(X,Y)$ is the KL divergence between the joint probability distribution of X and Y, and that when independence is assumed, therefore it measures the dependence in terms of 'departure' from independence. In fact, $I(X,Y)$ is the weighted average of Euclidean distance between logarithmic of the joint probability $p(x,y)$ and that when independence is assumed, where weights are the respective joint probabilities. That is, it is the expectation, under the joint probability, of the difference between the logarithmic of the joint probability $p(x,y)$ and that when independence is assumed. Note that though $0 \leq I(.,.) \leq 1$, there is no normalization (with respect to any maximal dependence) is involved.

Though these information measures are used to identify respective dependences they are not metrics since KL-divergance is not a true distance (metric), therefore they can not be used to measure the degree of dependence between variables. For example, as shown in [7] let $p(x,y)$ and $q(x,y)$ define two dependencies between X and Y where $p(x,y) = \begin{pmatrix} 3/8 & 1/8 \\ 1/8 & 3/8 \end{pmatrix}$ and $q(x,y) = \begin{pmatrix} 1/2 & 0 \\ 1/8 & 3/8 \end{pmatrix}$. Obviously probability distribution q shows a higher dependency than that of p but its mutual information is lower than that of p, $(MI_p(X,Y) > MI_q(X,Y))$. Note that q is obtained from p without preserving the marginal distributions of X and Y. Now let $r(u,v)$ and $s(u,v)$ define two dependencies between random variables U and V where $r(u,v) = \begin{pmatrix} 0 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 0 \end{pmatrix}$ and $s(u,v) = \begin{pmatrix} 0 & 0 & 2/7 \\ 1/7 & 2/7 & 0 \\ 1/7 & 1/7 & 0 \end{pmatrix}$. Then we have that $MI_r(U,V) < MI_s(U,V)$. Note that s shows a higher dependency than that of r and it is obtained from r by preserving the marginal distributions of U and V. Furthermore, all zeros in r are also in s. If this is the case then higher dependency implies higher mutual information. So mutual information is restricted measure of degree of dependence.

3.3 Chi squared test statistic χ^2

We can see that well-known Chi squared test statistic χ^2 that is used for testing independence of two discrete random variables uses a certain dependence measure in it for performing the test. Let X and Y take values $i = 1, ..., \alpha$ and $j = 1, ..., \beta$, respectively and let us write the joint probability of $X = i$ and $Y = j$ as p_{ij}, marginal probability of $X = i$ as $p_{i.}$ and that of $Y = i$ as $p_{.j}$. So, the conditional probability of $X = i$ given $Y = j$ is $p_{i|j} = p_{ij}/p_{.j}$ and similarly $p_{j|i}$ is defined.

Then,

$$\chi^2 = \sum_{i,j} n\frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}} = n\Big\{\sum_{i,j}\frac{p_{ij}^2}{p_{i.}p_{.j}} - 1\Big\} = n\Big\{\sum_{i,j} p_{ij}\frac{p_{ij} - p_{i.}p_{.j}}{p_{i.}p_{.j}}\Big\}$$
$$= n\Big\{\sum_{i,j} p_{ij}\frac{p_{i|j} - p_{i.}}{p_{i.}}\Big\} = n\Big\{\sum_{i,j} p_{ij}\frac{p_{j|i} - p_{.j}}{p_{.j}}\Big\} = nE\{A\}$$

where A is a random variable taking the value $\frac{p_{i|j}-p_{i.}}{p_{i.}} = \frac{p_{j|i}-p_{.j}}{p_{.j}}$ with probability p_{ij}, for $i = 1, ..., \alpha$ and $j = 1, ..., \beta$, and E denotes the expectation. That is, χ^2 is n-multiple of the expectation of a random variable whose $(i, j)^{th}$ value is a 'normalized' distance between the probability value $p_{i|j}$ and $p_{i.}$ where the normalizing constant is $p_{i.}$, for all i, j, and vice versa. Note that $\frac{p_{i|j}-p_{i.}}{p_{i.}}$ may be referred to as the 'degree' of dependence between the two events $X = i$ and $Y = j$. In fact, it is the certainty factor for the case $p_{i|j} < p_{i.}$, as described in [4] for measuring the dependency between the two events and it is a symmetric measure. However, here it is used without the condition. So, $E\{A\}$ is the expectation of a degree of dependence between the events $X = x$ and $Y = y$ for all x, y. Therefore, $E\{A\}$ can be thought of as measure of degree of dependence between X and Y. And the term n in χ^2 makes it a statistic. That is, a statistic for testing dependence between two variables can be seen as a product of two factors; one is a quantity related the degree of dependence between two variables and the other is that of total number of data cases that are used to estimate the probabilities related to them (i.e., sample information).

3.4 Test of two proportions

Sometimes one may be interested in testing equality of two proportions to see if given two variables are independent, for example, when the outcome (Y) of interest is binary, such as voting, denoted by $Y = 1$ (or not, denoted by $Y = 0$), for a political candidate in an election for two groups/populations (X) such as men, denoted by $X = 1$, and women, denoted by $X = 0$. Then one can test if two proportions are equal, i.e., $p(Y = 1|X = 1) = p(Y = 1|X = 0)$ (let us write it as $p_1 = q_1$) by the Z statistics

$$Z = \frac{1}{\sqrt{1/a + 1/b}}\frac{p_1 - q_1}{\sqrt{p(1-p)}}$$

where a and b are the sizes of the two samples of Y when $X = 1$ and $X = 0$, respectively, and $p = p(Y = 1)$. Now we can interpret that the factor $\frac{p_1-q_1}{\sqrt{p(1-p)}}$ as a measure of degree of dependence between the two variables due to the term

$(p_1 - q_1)$ in it, where the term $\sqrt{p(1-p)}$ should be taken as the normalizing constant. Note that the latter is constructed assuming full dependence between the two variables where, then their joint probability distribution is $P = \begin{pmatrix} 1-p & 0 \\ 0 & p \end{pmatrix}$ or similar. Instead of just using p which is the pooled proportion, the geometric mean of p and $(1-p)$ should be used as the normalizing constant. This is necessary to yield the same test statistic value for testing the same hypothesis with complementary probabilities i.e., $p(Y = 0|X = 1)$ and $p(Y = 0|X = 0)$. And the term $\frac{1}{\sqrt{1/a+1/b}}$ which is a function of sample sizes (sample information) makes Z a statistics. So, similar to χ^2 statistic, Z has a measure of degree of dependence between the two variables in it, in addition to information on the sample sizes.

4 Axioms of an ideal measure of dependence

Before we define our measure of strength/degree of dependence (or rather a generalization of ρ) it is appropriate to mention axioms that an ideal measure should possess as shown in [3]. However, it is hard to find dependence measures satisfying all these axioms. Our generalization of ρ seems to have a bigger potential in satisfying them, but we omit the discussion here. Following are the axioms;

1. It is well-defined for both continuous and discrete case
2. It is normalized such that its value 0 implies the independence and value 1 implies the full dependence (one variable is a deterministic function of the other), where all intermediate degrees of dependencies lie between 0 and 1
3. It is equal or has a simple relationship with the Pearson's correlation coefficient in the case of a bivariate normal distribution
4. It is a metric, i.e., it is a true measure of distance (between the independence and dependence of interest) not just a divergence
5. It is invariant under continuous and strictly increasing transformations.

These axioms are straightforward and require no further explanation.

In the following we define our measure following the structure and the construction of ρ but using a true distance metric. We propose to use so-called Hellinger distance but one may use another suitable distance metric. Since we are keeping the structure of the ρ the same but replacing its distance measure with a better one (a metric) when defining our dependence measure, we call it as a generalization of the ρ. This means that for any given dependence we should be

able to define the corresponding all possible full dependences, since the measure should be a ratio between a distance from independence to the given dependence and geometric average of distances from independence to the full dependences.

5 Defining a measure of degree of dependence

As we have seen earlier, in the two binary variables (2×2) case where only the linear dependence exists the dependence can be measured by using a single component Euclidean distance between joint probability distributions. However, in the case of two multinary variables ($n \times n$, where $n > 2$) we can have many types of dependences, and therefore distances among probability distributions can not be defined through only a single component or a weighted average area difference, that are Euclidean type distances and capable of measuring only linear dependences. Therefore we need to use some other suitable distance to measure any non-linear dependences. In the following we discuss a possible distance that is a true metric.

5.1 A metric distance between two probability distributions

We propose to use Hellinger distance between probability distributions (also called Matsushita distance for the discrete case) which is a metric in the probability simplex for our task of measuring dependence. Recall that our dependence measure should be the normalized distance between the given joint probability distribution of the two variables and that when their independence is assumed while preserving the marginals, where the normalizing constant is obtained by considering similar distances related to the all possible maximal dependences but preserving only one of the marginals at each time. Let Φ and Ψ be two discrete distribution functions (ϕ and ψ are probability distributions or mass functions) then the Hellinger distance between Φ and Ψ is defined as

$$M(\Phi, \Psi) = \left\{ \frac{1}{2} \sum_x \left\{ \sqrt{\phi(x)} - \sqrt{\psi(x)} \right\}^2 \right\}^{1/2}$$

In addition to satisfying properties of a metric $M(.,.)$ also satisfies the following properties: (1) $0 \leq M(\Phi, \Psi) \leq 1$, (2) $M(\Phi(T), \Psi(T)) = M(\Phi(T + a), \Psi(T + a))$ for any constant a, and (3) $M(\Phi(T), \Psi(T)) = M(\Phi(cT), \Psi(cT))$ for any constant $c \neq 0$ where the last two are called the *linear invariance* properties of the probability metric. Note that $\big(M(.,.)\big)^2$ is not a metric.

First we should have an idea about the furtherest jpd(s) for a given jpd that may represent independence. In fact we can see that the furtherest probability

distribution to a distribution that represent independence is not useful but those with fixed marginals, each at a time. For a given distribution function, say, Φ let us find the maximally Hellinger-distanced distribution function Ψ. The following proposition shows how to find it.

Proposition 5.1. *For positive probability distribution ϕ maximally Hellinger-distanced probability distribution ψ is given by*

$$\psi(t) = \begin{cases} 1, & \textit{if } t = \operatorname{argmin}_u \phi(u) \\ 0, & \textit{otherwise.} \end{cases}$$

and then, $M(\Phi, \Psi) = \left\{1 - \sqrt{min\{\phi(t) : t \in \mathcal{T}\}}\right\}^{1/2} < 1.$

Proof. Let $|\mathcal{T}| = n$, $\phi(t_i) = \phi_i$ and $\psi(t_i) = \psi_i$ for $i = 1, ..., n$. Let re-index all ϕ_i's such that $\phi_{(1)} \geq \phi_{(2)} \geq .... \geq \phi_{(n)}$ and possibly some of the ψ_i's can be zeros. $M(\Phi, \Psi)$ is maximal when $\sum_{t\in\mathcal{T}} \sqrt{\phi(t)\psi(t)}$ is minimal.

$$\begin{aligned} \sum_{i=1}^{n} \sqrt{\phi_i\psi_i} &= (\sqrt{\psi_1} + ... + \sqrt{\psi_n})\sqrt{\phi_{(n)}} \\ &\quad +(\sqrt{\psi_1} + ... + \sqrt{\psi_{n-1}})(\sqrt{\phi_{(n-1)}} - \sqrt{\phi_{(n)}}) \\ &\quad ... + \sqrt{\psi_1}(\sqrt{\phi_{(1)}} - \sqrt{\phi_{(2)}}) \geq \sqrt{\phi_{(n)}} \end{aligned}$$

That is, $\sum_{i=1}^{n} \sqrt{\phi_i\psi_i}$ is minimal when $\psi_1 = ... = \psi_{n-1} = 0$ and $\psi_n = 1$. So we obtain the maximally Hellinger-distanced distribution function Ψ and therefore $M(\Phi, \Psi)$.□

But then T is deterministic variable with respect to Ψ! This theorem says that for any given probability distribution, bivariate discrete in our case, the maximally Hellinger-distanced probability distribution is represented by a vertex of the probability simplex. All its component are zeros except for one place that has 1 that is corresponding to the smallest probability value of the reference probability distribution. This is a degenerate case as far as dependence of the two variables are concerned since it represents that both variables are deterministic and having full dependence. Therefore, such a full dependence can not be used for the normalization since it does not generally preserve the marginals.

For a given jpd P of X and Y, the dependence of them that it represents should be measured with a suitable normalized distance between P and P^I. It is clear from above that the normalizing constant should be the geometric mean of distances from independence to all possible full dependences where each such

full dependence should be preserving either of marginals. This rule is to follow the correlation coefficient definition. Therefore, an essential step is to find the two types of probability distributions P^X (jpd(s) representing full dependence when marginal of X is fixed) and P^Y (jpd(s) representing full dependence when marginal of Y is fixed) in order to find the normalizing constant. As you will see in some cases there may be multiple candidates for each of them. Therefore we have the following definition. Note that there are some instances such as in [3] and [5] where Hellinger distance between the jpd and that of when independence is assumed is used for measuring the dependence, but in such work no normalization is done. However, the above proposition implies that distance between any non-deterministic jpd representing independence and that representing a full dependence can be strictly less than 1 for two discrete random variables, therefore normalization is necessary if one wants to have a measure that shows strength of dependence.

Definition 5.1. *When M is a metric in the probability simplex of two discrete random variables X and Y, M-based measure of degree of dependence between X and Y represented by their joint distribution function P is defined as*

$$\rho^M(X,Y) = \frac{M(P^I,P)}{\prod_{P^X\in\mathcal{P}^X_{max}}\prod_{P^Y\in\mathcal{P}^Y_{max}}\left\{M(P^I,P^X)M(P^I,P^Y)\right\}^{1/|\mathcal{P}^Y_{max}+\mathcal{P}^X_{max}|}}$$

where P^I is the joint distribution function of X and Y when their independence is assumed, $\mathcal{P}^X_{max}$ denotes the set of all joint distribution functions, each representing a maximal dependence while preserving the marginal distribution of X and similarly for $\mathcal{P}^Y_{max}$, $|A|$ is the cardinality of the set A, and $M(P,Q)$ is the distance metric between two probability distributions P and Q.

Note that the denominator is the geometric mean of the maximal distances between full dependences and the independence. And we use Hellinger distance as the distance measure. Since ρ^M is defined following the structure of the Pearson's correlation coefficient it can be regarded as a generalization of it for the case of discrete variables.

For linear relationships measuring the dependence is relatively easy since both P^X and P^Y represent perfect linear dependence. This is when they have all their entries zero except for those, but may not be all, in each diagonal in respective case. For example, for a positive linear relation, P^X is obtained by assigning each main diagonal entry with the sum of all entries in the respective row. This assures that the marginal probability of X is preserved when obtaining full dependence, and similarly for P^Y. Note that positive linear relationship is selected if main diagonal entries are generally larger than the other entries in the joint probability value matrix P. But when we allow non-linear relationships between X and Y

there are no pre-specified P^X and P^Y, therefore multiple candidates may exist for each of them. We argue that they should be induced from the jpd in a similar way to the case of linear dependence. So we propose following simple rule for obtaining P^X and P^Y.

Definition 5.2. *For each x, when there exists a single value y' such that $y' = \mathrm{argmax}_y\, p(X = x, Y = y)$, then let $p^X(X = x, Y = y') = p(X = x)$ and $p^X(X = x, Y \neq y') = 0$ to obtain P^X. If there are multiple such y' values then obtain multiple P^X, each refering to one of those y' values, assuming that it is the only value where maxima exists. And similarly P^Y is defined.*

By this way, we get one or more jpds each representing a maximal dependence that preserves respective marginal.

6 Examples of $n \times n$ case where $n \geq 2$

Now we consider some different cases of P and demonstrate how we can calculate our measure and compare its value to those of some trational measures.

Case 1 Suppose a simple case of each row and column of P having a single maximal entry that is common to both its row and column. Then the other entries in the row are summed onto the maximal entry in the row for each row to yield P^X and similarly P^Y is obtained. Therefore, P^X and P^Y are on the boundary of Δ, so they are the furtherest probability distributions from P^I while preserving respective marginals. Then the degree of dependence between X and Y is defined as (since $|\mathcal{P}^X_{max}| = |\mathcal{P}^Y_{max}| = 1$)

$$\rho^M(X, Y) = \frac{M(P^I, P)}{\sqrt{M(P^I, P^X)M(P^I, P^Y)}}$$

Example 6.1. *For binary X and Y with $P = \begin{pmatrix} 0.3 & 0.2 \\ 0.1 & 0.4 \end{pmatrix}$, $\phi = 0.4082$ and $\rho^M = 0.2783$ (Cramer's V and Tschuprow's T are 0.4082). And interchanging off-main diagonal entries but keeping the main diagonal entries as they were, i.e., having $P = \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{pmatrix}$, gives the same results for all measures.*

Example 6.2. *Let state spaces of X and Y be $\{1, 2, 3\}$ and their joint probability $P = \begin{pmatrix} 0.05 & 0.03 & 0.20 \\ 0.30 & 0.07 & 0.05 \\ 0.04 & 0.20 & 0.06 \end{pmatrix}$ that is a non-linear dependence and then*

$P^I = \begin{pmatrix} 0.1092 & 0.084 & 0.0868 \\ 0.1638 & 0.126 & 0.1302 \\ 0.1170 & 0.090 & 0.0930 \end{pmatrix}$, $P^X = \begin{pmatrix} 0.00 & 0.00 & 0.28 \\ 0.42 & 0.00 & 0.00 \\ 0.00 & 0.30 & 0.00 \end{pmatrix}$ *and* $P^Y = \begin{pmatrix} 0.00 & 0.00 & 0.31 \\ 0.39 & 0.00 & 0.00 \\ 0.00 & 0.30 & 0.00 \end{pmatrix}$. *And then* $\rho = -0.2025$ *but* $\rho^M = 0.4113$ *(Cramer's* V *and Tschuprow's* T *are* 0.5472*). But had that* $P == \begin{pmatrix} 0.05 & 0.03 & 0.20 \\ 0.04 & 0.20 & 0.05 \\ 0.30 & 0.07 & 0.06 \end{pmatrix}$ *which is a linear dependence then* $\rho = -0.5474$ *and* $\rho^M = 0.4075$ *(Cramer's* V *and Tschuprow's* T *are* 0.5467*). Note the change in the degree of dependence is small since linear dependence is obtained from nonlinear case by just interchanging probability values in* P*.*

Case 2 When each row and column of P has a single maximal entry that may not be common to both its row and column we still can obtain a single P^X and a single P^Y. Therefore, we can apply the above definition.

Example 6.3. *When* $P = \begin{pmatrix} 0.30 & 0.03 & 0.20 \\ 0.05 & 0.07 & 0.05 \\ 0.04 & 0.20 & 0.06 \end{pmatrix}$ *we have* $\rho = 0.1383$ *and* $\rho^M = 0.450011$*. Note that here we have that Cramer's* V *and Tschuprow's* T *are* 0.4257843 *that are lesser than our measure.*

Case 3 When there are more than one maximal entry in a row or a column we have multiple P^X's and multiple P^Y's. Note that here we try to obtain a similar situation in the above two cases. That is, each row of P^X has only one non-zero element (it is obtained by summing up all entries in the corresponding row of P, thereby preserving the marginal probability distribution of X). Assume that we get a number of P^X's, say, $P^{X_1}, ..., P^{X_a}$ and b number of P^Y, say, $P^{Y_1}, ..., P^{Y_b}$. Let us consider the following example.

Example 6.4. *When* $P = \begin{pmatrix} 0.11 & 0.01 & 0.01 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.01 & 0.25 \\ 0.01 & 0.10 & 0.10 & 0.01 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.15 & 0.01 \\ 0.01 & 0.10 & 0.01 & 0.01 & 0.01 \end{pmatrix}$ *then we make two* P^X*'s;*

$$P^{X_1} = \begin{pmatrix} 0.15 & 0.000 & 0.000 & 0.00 & 0.00 \\ 0.00 & 0.000 & 0.000 & 0.00 & 0.29 \\ 0.00 & 0.230 & 0.000 & 0.00 & 0.00 \\ 0.00 & 0.000 & 0.000 & 0.19 & 0.00 \\ 0.00 & 0.140 & 0.000 & 0.00 & 0.00 \end{pmatrix} \textit{ and}$$

$$P^{X_2} = \begin{pmatrix} 0.15 & 0.000 & 0.000 & 0.00 & 0.00 \\ 0.00 & 0.000 & 0.000 & 0.00 & 0.29 \\ 0.00 & 0.000 & 0.230 & 0.00 & 0.00 \\ 0.00 & 0.000 & 0.000 & 0.19 & 0.00 \\ 0.00 & 0.140 & 0.000 & 0.00 & 0.00 \end{pmatrix}.$$

Therefore we have two maximal distances to these two full dependences. They are $M(P^I, P^{X_1})$ *and* $M(P^I, P^{X_2})$ *and similarly we obtain another two full dependences when marginal of* Y *is preserved. Therefore,*

$$\rho^M(X,Y) = \frac{M(P,P^I)}{\prod_{i=1}^{2}\prod_{j=1}^{2}\left\{M(P^{X_i},P^I)M(P^{Y_j},P^I)\right\}^{\frac{1}{4}}}$$

Then $\rho = -0.0491$ *and* $\rho^M = 0.5731$. *Note that here we have that Cramer's* V *and Tschuprow's* T *are* 0.6652.

7 Conclusion

We have looked at the structure and the construction of the Pearson's correlation coefficient ρ in order to have a generalization of it for measuring any non-linear dependence between two random variables. We have shown that it is simple do it geometrically for discrete variables. It can be shown that ρ is a normalized 'Euclidean' type distance between the joint probability distribution of the two random variables and that when their independence is assumed in the probability simplex of the two variables where normalizing constant is the geometric mean of two maximal such distances; each between full linear dependence of the two variables and their independence while preserving the marginal distribution of respective variable. So, we have shown that if we consider all possible full dependences and use an appropriate distance such as Hellinger then we can have a genaralization of ρ. But generally it is not easy to find all possible maximal distances, which is an open problem that may need algorithmic or computational solutions. However we have shown some examples after having defined a generalization.

Acknowledgments: Financial support for this research is from Swedish Research Council for Health, Working Life and Welfare (FORTE) and Swedish Initiative for Microdata Research in the Medical and Social Sciences (SIMSAM).

References

[1] A. Rényi, *Probability Theory* North-Holland Publishing Company and Akadémiai Kiadó, Publishing House of the Hungarian Academy of Sciences. Republished Dover USA, 2007.

[2] C. Sabatti, *Measuring dependency with volume tests*, The American Statistician 56 3 (2002), 191-195. **DOI**: 10.1198/000313002128.

[3] C. W. Granger, E. Maasoumi and J. Racine, *A Dependence Metric for Possibly Nonlinear Processes*, The Journal of Time Series Analysis 25 5 (2004), 649-669.

[4] F. Berzal, I. Blanco, D. Sanchez and M. -A. Vila, *Measuring the Accuracy and Interest of Association Rules: A New Framework*, Intelligent Data Analysis 6 3 (2002), 221-235.

[5] H. Skaug and D. Tjostheim, *Testing for serial independence using measures of distance between densities*, P. M. Robinson and M. Rosenblatt (Eds): Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis In Memory of E.J. Hannan, Springer Lecture Notes in Statistics 115 (1996), 363-377.

[6] K. Matsusita, *Decision rules, based on distance, for problems of fit, two samples, and estimation*, Annals of Mathematical Statistics 26 4 (1955), 631-640.

[7] M. Studeny and J. Vejnarova, *The Multiinformation Function as a Tool for Measuring Stochastic Dependence*, M. I. Jordan (Eds): Learning in Graphical Models, Kluwer Academic Publishers (1998), 261-297.

[8] M. Sugiyama and K. M. Borgwardt, *Measuring Statistical Dependence via the Mutual Information Dimension*, Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI'13) AAAI Press (2013), 1692-1698.

[9] N. Balakrishnan and C. -D. Lai, *Continuous Bivariate Distributions*, Springer, 2009.

[10] P. Diaconis and B. Efron, *Testing for independence in a two-way table: new interpretations of Chi-square statistics*, The Annals of Statistics 13 (1985), 845-874.

[11] P. Wijayatunga, S. Mase and M. Nakamura, *Appraisal of Companies with Bayesian Networks*, International Journal of Business Intelligence and Data Mining 1 3 (2006), 326-346.

[12] S. E. Fienberg and J. P. Gilbert, *The Geometry of a Two by Two Contingency Table*, Journal of the American Statistical Association 65 (1970), 694-701

[13] S. Kullback and R. A. Leibler, *On information and sufficiency*, The Annals of Mathematical Statistics 22 1 (1951), 79-86

[14] W. Bergsma, *A bias-correction for Cramér's V and Tschuprow's T*, Journal of the Korean Statistical Society 42 3 (2013), 323-328. http://dx.doi.org/10.1016/j.jkss.2012.10.002.

RATIO MATHEMATICA
ISSUE N. 30 (2016) 23-33

ISSN (print): 1592-7415
(online): 2282-8214

Verification of the mathematically computed impact of the relief gradient to vehicle speed

Martin Bureš, Filip Dohnal
Department of Military Geography and Meteorology
University of Defence in Brno, Czech Republic
martin.bures@unob.cz

Abstract

Terrain traficability is one of the key activities of military planning, firefighting and emergency interventions. Terrain traficability is affected by many factors and terrain slope is one of them. Deceleration ratio that represents the influence of slope inclination is dependent on a technical attributes of vehicle. The results of field terrain tests suggest that deceleration ratio established via calculation does not have to correspond with practical experience.

Keywords: cross-country movement, deceleration ratio, terrain slope

doi: 10.23755/rm.v30i1.6

1 Introduction

The basis for the planning of vehicle movement in terrain is the knowledge of natural conditions, which influence the movement itself. With respect to the driving characteristics, which are characterized by a whole range of technical parameters, there is a modelling process of the impact of natural conditions on the movement in the field [1], [2], [3]. Landscape represents very complicated system and therefore, during the modelling of the natural conditions impact on the movement, the landscape elements are evaluated separately. One of these elements is terrain relief, whose slope characteristics have a direct influence on the speed of a moving vehicle [4]. Compared to the other terrain characteristics,

the relief slope can be successfully analyzed with the GIS tools (considering the accuracy and quality of spatial data) [5].

2 Theory of Cross Country Movement

The vehicle mobility in the field is based on the mutual effect of the tree basic components, which influence: operation in terrain (maneuver), used technique and geographical conditions. The mutual influence of these components, related to the military operations, shows Fig. 1.

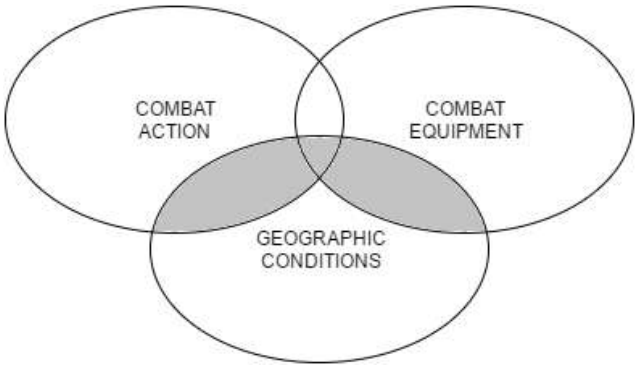

Fig. 1: The influence of geographic conditions to combat action and combat equipment [6].

Modern methods of conducting military operations are supported by a range of operational analysis. One of the very important area represent the terrain analysis, which are these days conducted especially with the usage of digital geographic data and GIS tools. In this area we classify also terrain trafficability analysis, which results may not be used only for military purposes, but also for the fulfilment of the tasks of IRS or emergency management authorities.

The terrain trafficability can be defined as a mobile ability of units, which is influenced especially by geographical factors of the territory and technical parameters of vehicles, or (according to [6]) as the level of technical competence of individual vehicles to move in terrain and overcame different geographic features and phenomena.

Evaluation of geographic factors which influence the terrain trafficability mainly concentrates on the impact of relief gradient, microrelief forms, soil condition, vegetation, waters, climate and weather condition, settlements and communications. These factors are later divided to other components [6]. The evaluation is also influenced by technical data of used vehicles and driver's capability. But it is very difficult to mathematically evaluate driver's influence.

All these factors are closely related and influence each other. Their combined influence on vehicle cause deceleration or even stopping. The real speed of the vehicle can be expressed by this formula [2]:

$$v_j = f(v_{max}, c_1, c_2, \dots c_n), \quad j = 1 \dots k \tag{1}$$

where v_j means vehicle speed at j-section of vehicle path [km·h^{-1}], v_{max} maximum vehicle speed at communications [km·h^{-1}], c_i i-coefficient of deceleration due to factor F_i computed for j-section with invariable values c_i, n number of geographic factors effecting at given section of terrain and k number of sections on vehicle path.

The terrain trafficability is very complex and it is not possible to identify effect of all the terrain factors, therefore it is necessary to proceed systematically. First comes identifying basic terrain factors influence, such as relief gradient, then comes their combined influence and last comes less important components.

The impact of relief gradient to cross-country movement

A relief gradient represents one of the most fundamental factor implicating cross-country movement. The calculation of total resulting coefficient of vehicle deceleration by relief and microrelief impact is given for determinate by relation as follows [7]:

$$c_1 = c_{11} c_{12}, \tag{2}$$

where c_{11} is deceleration coefficient by impact of gradient factor and c_{12} deceleration coefficient by impact of microrelief factor.

3 Calculation of coefficient of vehicle deceleration of impact of relief gradient (c_{11})

It is possible to express a relief gradient by various terrain models such as: raster model, TIN and others. The coefficient of deceleration of gradient factor c_{11} is determinable by three methods as follows [6]:

1) according to DMA method (Defence mapping Agency);
2) on the basic of tractive charts of particular vehicles;
3) by the terrain operation tests.

Determination of c11 according to DMA method:

According to the formula listed below, which contains values of relief gradient and parameters of vehicle, is possible to acquire deceleration ratio [8]:

$$c_{11} = \frac{GradT_{max} - SH}{GradK_{max}}, \tag{3}$$

where $GradT_{max}$ [%,°] is maximum climbing capability of a vehicle on terrain; $GradK_{max}$ [%,°] maximum climbing capability of a vehicle on road and SH [%,°] mean value of slope gradient obtained from the Table 1.

Category	Slope [%]	SH [%]	Slope [°]	SH [°]
1	< 0	0	< 0	0
2	0 – 3	1,5	0,00 – 1,35	0,68
3	3 – 10	6,5	1,35 – 4,50	2,93
4	10 – 20	15	4,50 – 9,00	6,75
5	20 – 30	25	9,00 – 13,50	11,25
6	30 – 45	37,5	13,50 – 20,25	16,88
7	> 45	slope [%]	> 20,25	slope [°]

Table 1: Determination of mean value of the slope gradient (SH) from the measured range of slopes [6].

Determination of c_{11} at the basis of tractive charts:

The deceleration ratio of impact of gradient factor can by also determinable on the basis of tractive charts of particular vehicles [6]. To calculate running characteristic on the route of vehicle it must be started from the presupposition that this route is described at particular section by longitudinal gradient (α), transversal inclination (β), coefficient of rolling resistance (f) and coefficient of static friction (φ). Particularly significant from the point of view of cross-country movements evaluation are also following data:

- attainable driving speed (eventually an acceleration);
- conditions whereat coming to a swerving either of longitudinal or transversal direction;
- conditions whereat coming to loss of maneuverability and longitudinal or transversal rollover.

A tractive chart is the formulation of tractive power dependence on vehicle driving speed. The driving speed is plotted on the horizontal axis on the chart and on the vertical axis are plotted tractive power and forces of resistance. The tractive power F_T depends on engine torque and total ratio, whereas both quantities are changeable in running. Providing that transmission efficiency is constant, the tractive power at particular speed gear is adequate to engine torque

at that moment. Considering total ratio changeability, we can say that each vehicle has as much tractive power curves as the number of vehicle speed gears.

The process of calculating the course of curves of tractive power has following parts [6]:

- number of points are selected on external torque characteristics of engine that are characterizing engine torque curve;
- from the characteristic we determine corresponding quantity of engine torque M_m and proper engine revolutions n_m;
- the coordinates (M_m, n_m) are read out of selected points on the engine torque characteristics;
- the coordinates (M_m, n_m) are then transformed to coordinates (F_T, V) by formula (4) and points with the coordinates (F_T, V) for particular speed gears create the curves of tractive power at the tractive chart.

$$F_T = \frac{M_m \eta_m i_{c(j)}}{r_d}; \quad V = 0{,}377 \frac{n_m r_d}{i_{c(j)}} \tag{4}$$

where F_T [N] is tractive power, M_m [Nm] engine torque, η_m [%] mechanical efficiency of transmissions, $i_{c(j)}$ total transmission ratio, r_d [m] wheel dynamic radius, V [km·h^{-1}] vehicle driving speed and n_m [min^{-1}] engine revolutions.

The curves of rolling resistance by even speed movements are marked by proper terrain gradient and tractive power curve is marked by relating speed gear. For the ideal course of tractive power each tractive power curve tangents a hyperbolic curve.

The contact points of both curves at every speed gear corresponds to engine revolutions at maximum power. The chart is completed under horizontal axis and scales of motor revolutions at given speed in particular speed gears. This diagram also presents a survey of driving characteristics of vehicle [6]:

- climb capability at particular speed gears (by the interpolation among curves of rolling resistance – slopes);
- what speed gear is to be used during uphill driving on particular slope;
- what speed is achievable on a particular slope;
- maximum speed on a plain field (V_{max}).

Determination of c_{11} at the basis of operational testing:

For the basic type of vehicles was relief gradient deceleration ratio determine based on terrain tests. For the particular vehicles was used following procedure [6]:

1. The tractive chart is calculated.

2. The readings of maximum available driving speeds and driving positions used were made from the tractive diagram for each partial parts of section given.
3. The passage time was calculated for each mentioned partial parts (at all 19 sections of terrain).
4. The results were compared with operational driving tests and on that basis; the resulting coefficient of deceleration was defined for each section (at all 19 sections of terrain).
5. There were calculated mean values of the multiple coefficients of deceleration for each vehicle for: terrain; cartways and forest ways; roads.

To calculate presupposed driving speed on communications, cartways and forest ways can be used following relation:

$$V_{EST} = V_{MAX} c_{11}\,, \tag{5}$$

where V_{EST} [km·h^{-1}] means estimated driving speed, V_{MAX} [km·h^{-1}] maximum driving speed indicated for a vehicle and c_{11} multiple coefficient of deceleration according to the Table 2.

Carriageway type	Passenger off-road vehicle	Medium off-road utility vehicles	Heavy off-road lorries	Infantry combat vehicles	Tanks
Terrain	0.22	0.31	0.28	0.42	0.41
Cartways and forest ways	0.43	0.53	0.52	0.58	0.53
Roads	0.72	0.86	0.84	0.72	0.72

Table 1: The mean multiple coefficients of deceleration of military vehicle movements on free terrain and on communications [6].

4 Field testing and data processing

To verify the theoretical values field tests were used. Tests were conducted in the military training area Libava in 2015, there were tested eight types of vehicles, including Tatra 810 6x6 (T810).

For analysis of the impact of the relief gradient there were selected rides on the training circuit, which contains a tank track. Unpaved surface of the tank track was not covered by vegetation and contained lots of micro-relief forms, especially the waves od soil that have approximately 20 m in length with an amplitude up to 1 m and ruts. The width of the tank tracks ranges from 10 to 30 m. The maximum slope of the test area reaches only to values of 16 °.

Verification of the mathematically computed impact of the relief gradient to vehicle speed

The vehicle routes were recorded by a GPS receiver Trimble Geoexplorer 3000 GeoXT equipped with an external antenna External Mini. The vehicle speed was calculated from locations and times of the records.

All records have been checked and the wrong or unnecessary ones have been removed (an error in position, parking, turning at the end of the route). Next step was to add the value of the terrain slope from the most precise digital elevation model of Czech Republic (DMR5G) [7] in the spot of each record with use of the ArcGIS 10.2.1 [8].

Correction of the estimated speed

Values of the speed calculated by formula (3) and derived from the traction diagram are acceptable only in case of ideal conditions, where the only factor influencing the drive is terrain slope. The analysed rides took place under invariant but still not ideal conditions, such as after rain with muddy and slippery surface. After the elimination of micro-relief affected records all other influences can be considered constant.

Maximum speed T810 vehicle in terrain mode (V_{MAX}) is 65 km·h^{-1}, the maximum reached speed in the given conditions was 36 km·h^{-1} ($V_{MAX'}$). Then the deceleration coefficient C_S (influence of surroundings) has the value 0.55. All the following values have been corrected by this coefficient (equation xxx).

$$V_{MAX'} = C_S \cdot V_{MAX} \tag{6}$$

Method of predicting the vehicle speed in general terrain, which neglects the influence of the slope, was not corrected, because all the factors have been already included.

Verification of theoretical values of speed

The results of all three methods of calculating the velocity field were compared with the measured data. Unfortunately, the measured data do not represent the whole range of terrain slope which T810 can pass through, for example up to 30 °. Frequency distribution of the slope gradient in the records is shown in Fig. 2. Small counts in the higher slopes reduce their credibility. However, at least in the lower slopes below 7 ° the data can be probably used to verify mathematical apparatus.

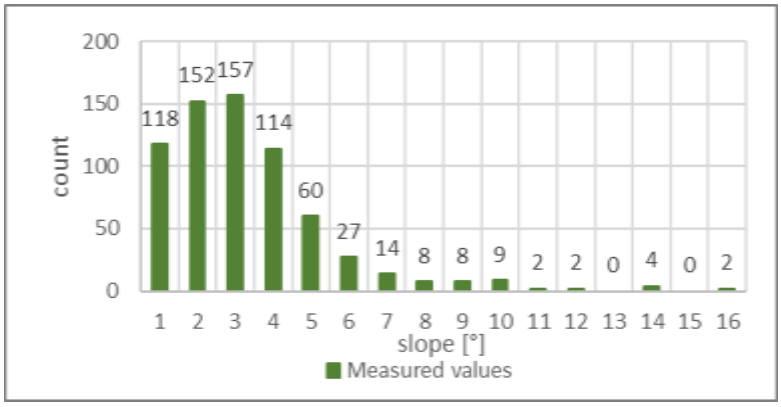

Fig. 2: Counts of measured values

The Table 3 compares the calculated values according to the methodology of DMA and speed read from T810 tractive chart with the measured speed. The same comparing is represented on the Fig. 3.

Category	Slope [°]	DMA [km·h⁻¹]	Tractive chart [km·h⁻¹]	Prediction at the basis of operational testing[km·h⁻¹]	Measured speed [km·h⁻¹]
1	< 0	-	-	-	-
2	0.00-1.35	35	35	-	21
3	1.35-4.50	32	33	-	24
4	4.50-9.00	28	24	-	24
5	9.00-13.50	23	18	-	23
6	13.50-20.25	16	12	-	21
7	> 20.25	-	-	-	-
Mean		29	27	28	23

Tab. 3. The comparison of the calculated and measured values.

The difference between both estimated value is not significant in small slopes, but with a growing slope the difference increase up to 5 km·h^{-1}. The measured speed is much lower in slopes 0 ° – 7 ° and the same situation applies to the predicted average speed, which is 28 km·h^{-1}, but the measured speed was 23 km·h^{-1}.

Verification of the mathematically computed impact of the relief gradient to vehicle speed

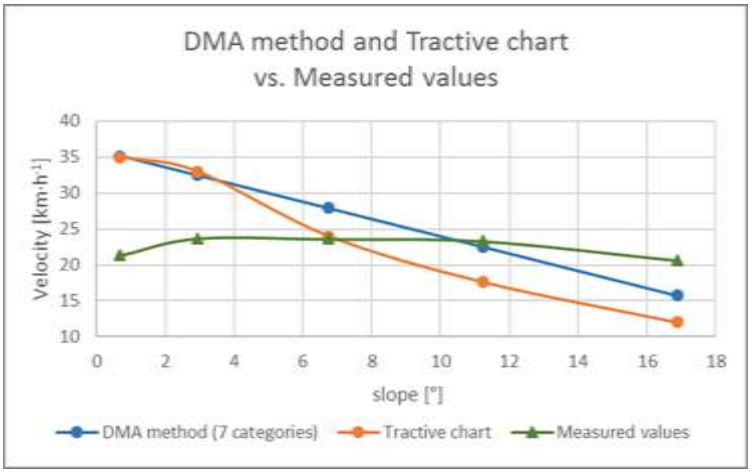

Fig. 3. The Illustration of the comparison of the calculated and measured values

The achieved results do not correspond with the expectations and it is probably not possible to use this data at this point of the research to verify impact of the slope gradient to the vehicle speed. The reason of very slow ride along the entire length of the route and a reason of unusable results seem to be less experienced driver, who drove a given car.

Significant distortion of measured data due to unexperienced driver was confirmed by comparison with the another lorry, Tatra 815 8x8. Its technical specifications are slightly different, but the maximum surmountable slope remains the same value. The Fig. 4 illustrates both measured rides – T810 and T815. Data measurement by other vehicles proves that the main impact on the T810 ride was the driver.

The relatively high speed at inclinations of 11 ° - 16 ° are caused by a too short climb to slowdown the vehicles marginally.

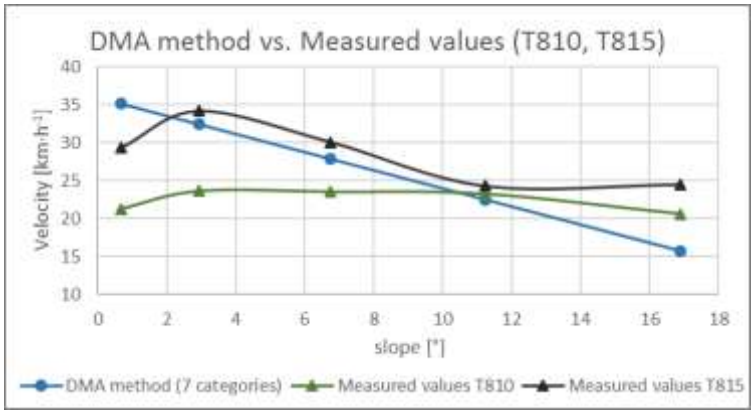

Fig. 4: The comparison of the calculated and measured values (T810, T815)

5 Conclusions

Unfortunately, the data from the T810 cannot be used to verify the mathematical apparatus used to calculate the impact of the slope gradient on vehicle speed. The first obstacle is the number of data from higher terrain slopes and the other is a distortion caused by inexperience of the driver. Even this result has a positive contribution in the form of experience needed for planning field tests and obtaining relevant data.
To determine the impact of the slope gradient on the speed of T810 is necessary to get more data from multiple passes through the high slopes near the limits of the vehicle. The next step to successful verification of mathematical calculations is testing several drivers. It will significantly reduce the influence of experience of the driver and also possibly his mental state.

6 Acknowledgement

The work presented in this paper was supported within the project for "Development of the methods of evaluation of environment in relation to defense and protection of the Czech Republic territory" (Project code NATURENVIR) by the Ministry of Defence the Czech Republic.

Bibliography

[1] Collins, J., M., (1998). Military geography for professionals and the public. 1st Brassey's ed. Washington, D. C.: Brassey's, xxiv, 437 p. ISBN 15-748-8180-9.

[2] Rybansky, M., Hofmann, A., Hubacek, M. et al., (2015). Modeling of Cross-Country Transport in Raster Format. Environ. Earth Sci., 74: 7049. doi:10.1007/s12665-015-4759-y.

[3] Hofmann, A., Hošková-Mayerová, Š., Talhofer, V. et al., (2015). Creation of Models for Calculation of Coefficients of Terrain Passability. Quality & Quantity, 49: 1679. doi:10.1007/s11135-014-0072-1.

[4] Vala, M., (1992). Evaluation of the Passability of the Military Vehicles (in Czech). /Habilitation Thesis/. VA Brno, 152 pp.

[5] Rybansky, M., Vala, M., (2010). Relief Impact to Cross-Country Movement. In: Proceedings of the Joint 9th Asia-Pacific ISTVS Conference, Sapporo, Japan, 16 pp.

[6] Rybansky, M., (2009). Cross-Country Movement: The Impact and Evaluation of Geographic Factors. 1st ed. Brno. Academical publishing CERM, 113 p. ISBN 978-80-7204-661-4.

[7] Talhofer, V., Hofmann, A., Kratochvil, V., Hubacek M., Zerzan, P., (2015). Verification of Digital Analytical Models - Case Study of the Cross-Country Movement. In: ICMT´15 – International Conference on Military Technologies 2015, Brno, (Czech Republic), 7 pp. ISBN 978-1-4799-7785-7.

[8] Rybansky, M., Hofmann, A., Hubacek, M. et al., (2015). Modelling of cross-country transport in raster format. Environ. Earth Sci., 74: 7049. doi:10.1007/s12665-015-4759-y.

[9] Hubacek, M., Kovarik, V., Kratochvil, V., (2016). Analysis of influence of terrain relief roughness on dem accuracy generated from lidar in the Czech Republic territory. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 41 25-30. doi:10.5194/isprsarchives-XLI-B4-25-2016.

[10] ESRI. (2013). User documentation. Copyright © 1995–2013 Esri.

RATIO MATHEMATICA
ISSUE N. 30 (2016) pp. 35-43

ISSN (print): 1592-7415
(online): 2282-8214

A Recursive Variant of Schwarz Type Domain Decomposition Methods

František Bubeník, Petr Mayer
Faculty of Civil Engineering, Czech Technical University in Prague, Czech Republic
Frantisek.Bubenik@cvut.cz, Petr.Mayer@cvut.cz

Abstract

In this paper a slightly different approach to the use of the domain decomposition method of the Schwarz type is proposed. Instead of the standard coarse space construction we propose to use a recursive solution on each domain. Thus we do not need to construct a coarse space but nevertheless we are still keeping $O(1)$ convergence speed. For local problems we use the standard iterative solvers for which the amount of the work for one step is $O(N)$, where N is the number of equations. Due to the fact that the overlapping is under our control we can keep total work in $O(N^{(1+\gamma)})$ operations with arbitrary positive γ.

Keywords: Domain Decomposition, Finite Element Method, Linear Systems.

2000 AMS subject classifications: 97U99

doi: 10.23755/rm.v30i1.4

1 Introduction

This paper deals with some aspects of the classic Schwarz alternating method. There are analyzed ways how to arrange with the deceleration of algorithms if the stepsize of a mesh for the finite element method is decreasing. The standard two-level method is described and some alternative approaches to solve a number of local problems by the same method are proposed.

2 The Schwartz alternating method

As a model problem we will solve a Poisson problem

$$-\triangle u(x) = f(x), \quad x \in \Omega \subseteq \mathbb{R}^d.$$

We use the finite element method to solve the problem. It leads to a linear system

$$Ax = b\,. \tag{1}$$

Consider functions $\varphi_1, \ldots, \varphi_n$ as a basis and denote by $V_n = span(\varphi_1, \ldots, \varphi_n)$ the linear hull of the functions, that is the set of all linear combinations of the functions. Let us remind that $A_{ij} = a(\varphi_i, \varphi_j)$ and

$$a(u, v) = \int_\Omega uv \ \mathrm{d}\Omega.$$

The matrix A for our problem is symmetric and positive definite. Moreover, for many interesting choices of the basis of V_n, the matrix A is sparse, but large.

One of the possible strategies to solve the system (1) is to use a sparse version of LU-decomposition. In most cases, fill-in which takes place along with Gaussian elimination makes such an approach unusable. A usual choice is then to use some iterative method. Since our matrix is a symmetric positive definite, it seems to be more advantageous to employ the conjugate gradient method. But this choice is still problematic because the amount of the work is rapidly increasing with the size of the problem. Therefore there is then more convenient to use a preconditioned conjugate gradient method with an appropriate preconditioner.

As a preconditioner we can choose a slightly modified the Schwarz alternatig method, see [2]. The original description is in [1]. We can see it as a kind of some block symmetrized Gauss-Seidel method.

2.1 Formulation of the algorithm

Let us denote by n_d the number of domains. For each $i \in I = \{1, 2, \ldots, n_d\}$ we define an index set $I_i = \left\{i_1^{(i)}, i_2^{(i)}, \ldots, i_{n_i}^{(i)}\right\}$. These index sets realize a covering of I, i. e. $I = \bigcup_{i=1}^{n_d} I_i$. This covering is not required to be disjoint. We define subspaces of V_n so that the subspace $V_n^{(i)}$ is the linear hull of a corresponding part of the basis of V_n, that is $V_n^{(i)} = \underset{j \in I_i}{span} \{\varphi_j\}$. Finally we

define subdomains

$$\Omega^{(i)} = \bigcup_{j \in I_i} supp\,(\varphi_j)\,, \tag{2}$$

where $supp\,(f)$ denotes the support of a function f. The sizes of individual domains are $n_1, \ldots, n_{n_d}$. We can write $A^{(i)} = A(I_i, I_i)$ in terms of Matlab-like notation. Matrix interpretation of local problems is $N^{(i)} = \{n^{(i)}_{k,l}\} \in R^{n \times n_i}$, where

$$n^{(i)}_{k,l} = \begin{cases} 1 & \text{if} \quad l = i^{(i)}_k, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$A^{(i)} = {N^{(i)}}^T A\, N^{(i)}. \tag{3}$$

The following algorithm describes the transition from $x^{(k)}$ to $x^{(k+1)}$.

Algorithm 2.1. *One step of the symmetrized Schwarz method*

$x^{(k+\frac{0}{2n_d})} := x^{(k)}$
for $i = 1, \ldots, n_d$
  $r := b - Ax^{(k+\frac{i-1}{2n_d})}$
  $\widetilde{r} := {N^{(i)}}^T r$
  $A^{(i)} := {N^{(i)}}^T A N^{(i)}$
(♣) *Solve* $A^{(i)} c = \widetilde{r}$, *i. e.* $c = {A^{(i)}}^{-1} \widetilde{r}$
  $x^{(k+\frac{i}{2n_d})} := x^{(k+\frac{i-1}{2n_d})} + N^{(i)} c$
end for (4)
for $i = 1, \ldots, n_d$ (5)
  $r := b - Ax^{(k+\frac{1}{2}+\frac{i-1}{2n_d})}$
  $\widetilde{r} := {N^{(n_d+1-i)}}^T r$
  $A^{(n_d+1-i)} := {N^{(n_d+1-i)}}^T A N^{(n_d+1-i)}$
(♣) *Solve* $A^{(n_d+1-i)} c = \widetilde{r}$
  $x^{(k+\frac{1}{2}+\frac{i}{2n_d})} := x^{(k+\frac{1}{2}+\frac{i-1}{2n_d})} + N^{(n_d+1-i)} c$
end for
end algorithm

The method used in the Algorithm 2.1 can also be viewed as a variant of the block Gauss-Seidel method, but with the fact that the individual blocks can overlap.
Put

$$P^{(i)} = A^{1/2} N^{(i)} ({N^{(i)}}^T A N^{(i)})^{-1} {N^{(i)}}^T A^{1/2}. \tag{6}$$

Then

$$\begin{aligned} {P^{(i)}}^2 &= A^{1/2}N^{(i)}({N^{(i)}}^T A N^{(i)})^{-1}{N^{(i)}}^T A^{1/2} A^{1/2}N^{(i)}({N^{(i)}}^T A N^{(i)})^{-1}{N^{(i)}}^T A^{1/2} \\ &= A^{1/2}N^{(i)}{A^{(i)}}^{-1}A^{(i)}{A^{(i)}}^{-1}{N^{(i)}}^T A^{1/2} \\ &= A^{1/2}N^{(i)}{A^{(i)}}^{-1}{N^{(i)}}^T A^{1/2} = P^{(i)}. \end{aligned}$$

It follows that $P^{(i)}$ is a projection, moreover A-orthogonal. Further $P^{(i)} = {P^{(i)}}^T$, then the projection is symmetric.

Let us denote

$$\varepsilon^{(k)} = x^{(k)} - x^*,$$

where x^* denotes the solution of the problem. Errors are analyzed in terms of the energy norm $||x||_A = \sqrt{(x,x)_A}$, where $(x,y)_A = x^T A\, y$.

Let us note that A is a symmetric positive definite matrix and $A^{(i)}$ is a principal minor of A. Then $A^{(i)}$ is also a symmetric positive definite matrix.

We have

$$||\varepsilon^{(k)}||_A^2 = {\varepsilon^{(k)}}^T A\, \varepsilon^{(k)} = {\varepsilon^{(k)}}^T A^{1/2}\, A^{1/2}\, \varepsilon^{(k)} = ||A^{1/2}\varepsilon^{(k)}||_A^2.$$

Thus

$$\begin{aligned} A^{1/2}\varepsilon^{(k-1)} &= (I - P^{(1)})\dots(I - P^{(n_d)})(I - P^{(n_d)})\dots(I - P^{(1)})A^{1/2}\varepsilon^{(k)} \\ &= M\, A^{1/2}\varepsilon^{(k)}. \end{aligned} \tag{7}$$

Since $I - P^{(i)}$ is a symmetric $A-$orthogonal projection then M is a symmetric matrix. And moreover, M is, according to the definition, a positive semi-definite matrix. It can be proved that M is even a positive definite matrix.

2.2 Dependence on the dimension of V_n

For the following considerations we suppose that piecewise linear finite elements are used. In that case $n = O(1/h^d)$, where h is the stepsize of a mesh. If we try to keep domains with the same geometry, the amount of elements will increase as $O((H/h)^d)$, where H is the typical size of a domain. Then, the amount of iterations is the same, but the amount of work for one step will increase.

On the other hand, when we keep equal the number of elements inside a domain, then the size of the domain will decrease and then the number of domains will increase. This leads to increasing amount of iterations and slightly increasing work for one full step.

Usual solution is to use a coarse space. It typically means to replace each domain by a base function. We create the coarse space and the solution of the

problem for the corrections on the coarse level is inserted between steps (4) and (5) of the Algorithm 2.1. A detailed analysis is introduced, for example, in [2]. When it is used in a right way, we get $O(1)$ convergence speed.

2.3 Basic convergence

Let us denote

$$E(x) = x^T Ax - 2x^T b. \tag{8}$$

It is known that $Ax = b$ if and only if $E(x)$ assumes its minimum at x. Each step in Algorithm 2.1 means the minimization of functional (8) on a corresponding subspace and the following inequality holds for the successive terms of the minimizing sequence

$$E(x^{(k+1)}) \leq E(x^{(k)}). \tag{9}$$

We prove the following equivalence:

Lemma 2.1. *The equality in* (9) *occurs* $\Longleftrightarrow$ $x^{(k)}$ *is the accurate solution of* $Ax^{(k)} = b$.

Proof. It is clear that an accurate solution is equivalent to $r^{(k)} = 0$, where $r^{(k)} = b - Ax^{(k)}$.

We prove one direction of the equivalence: Suppose that $x^{(k)}$ is an accurate solution and we prove the equality required. It is easy to see that if $x^{(k)}$ is an accurate solution then $r^{(k)} = 0$ and then the equality in (9) occurs.

Now we prove the opposite direction of the equivalence. We apply the proof by contradiction: suppose that $x^{(k)}$ is not an accurate solution and suppose that the equality in (9) holds. If $x^{(k)}$ is not an accurate solution then $r^{(k)} \neq 0$. Then there exists the least index i_m such that ${N^{(i_m)}}^T \; r^{(k)} \neq 0$. It causes a decrease at this step of Algorithm 2.1 and then the strict inequality $E(x^{(k+1)}) < E(x^{(k)})$. This contradicts to the assumption of equality in (9) and the proof of the equivalence in Lemma 2.1 is complete. □

3 Recursive approach

Another possibility comes from the idea that the local problems are conceptually identical as the original one. It opens a possibility to use the same Schwarz algorithm for solving them. It means to retain domains in the same geometry, then $\Omega^{(i)}$ in (2) remain unchanged. On the other hand it means that the number of degrees of freedom for a domain increases. In this case we recommend to use the same algorithm for each domain separately.

3.1 Two - level variant

We start from the Algorithm 2.1, and we replace both steps denoted by (♣) in the algorithm by an iterative solution for c. It means that we replace ${A^{(i)}}^{-1}\widetilde{r}$ by an approximate solution of the problem $A^{(i)}c = \widetilde{r}$. As the method we use again the algorithm 2.1 with ℓ steps.

Then the error operator has the form

$$\widetilde{M} = (I - \widetilde{P}^{(1)}) \dots (I - \widetilde{P}^{(n_d)})(I - \widetilde{P}^{(n_d)}) \dots (I - \widetilde{P}^{(1)}),$$

where

$$\widetilde{P}^{(i)} = A^{1/2} N^{(i)} \widetilde{Q}^{(i)} {N^{(i)}}^T A^{1/2}. \tag{10}$$

Expression $({N^{(i)}}^T A N^{(i)})^{-1}$ in (6) is for short denoted by $Q^{(i)}$ and it is replaced in (10) by

$$\widetilde{Q}^{(i)} = {A^{(i)}}^{-1/2} \left[I - \left(I - {A^{(i)}}^{1/2} M^{(i)} {A^{(i)}}^{1/2} \right)^{\ell} \right] {A^{(i)}}^{-1/2}, \tag{11}$$

where $A^{(i)}$ is from (3) and $M^{(i)}$ denotes the error operator to the algorithm 2.1 applied on the $i-$th domain. This replacement comes from the following:

Let us solve a problem $Ax = b$ and let us use the following iterative method

$$x^{(i+1)} = x^{(i)} + M(b - Ax^{(i)}),$$

where M is a symmetric positive definite matrix. The initial approximation is

$$x^{(0)} = 0. \tag{12}$$

Let $x^* = A^{-1}b$. Then

$$x^* - x^{(i+1)} = x^* - x^{(i)} - M(b - Ax^{(i)}) = (I - MA)(x^* - x^{(i)})$$

and then

$$A^{1/2}(x^* - x^{(i+1)}) = (I - A^{1/2} M A^{1/2}) A^{1/2}(x^* - x^{(i)}).$$

After $\ell-$iterations we get

$$A^{1/2}(x^* - x^{(\ell)}) = (I - A^{1/2} M A^{1/2})^{\ell} A^{1/2}(x^* - x^{(0)})$$

and thus

$$x^* - x^{(\ell)} = A^{-1/2}(I - A^{1/2} M A^{1/2})^{\ell} A^{1/2}(x^* - x^{(0)}).$$

Since $x^{(0)} = 0$ we get

$$x^* - x^{(\ell)} = A^{-1/2}(I - A^{1/2}MA^{1/2})^\ell A^{1/2}x^* = A^{-1/2}(I - A^{1/2}MA^{1/2})^\ell A^{-1/2}b.$$

Thus

$$\begin{aligned} x^{(\ell)} &= x^* - A^{-1/2}(I - A^{1/2}MA^{1/2})^\ell A^{-1/2}b \\ &= A^{-1/2}\left[I - \left(I - A^{1/2}MA^{1/2}\right)^\ell\right] A^{-1/2}b \end{aligned}$$

and that is why the form of $\widetilde{Q}^{(i)}$ in (11) and $\widetilde{P}^{(i)}$ in (10).

3.2 Recursive - multilevel method

When we use a two-level method we need to compute $\widetilde{P}^{(i)}$ in (10) and for it there is required to know $\widetilde{Q}^{(i)}$ from (11). For $\widetilde{Q}^{(i)}$ there is necessary to know $M^{(i)}$ and its realization comes from the solution of a local problem on subdomains of the $i-$th domain. In case when these local problems are still large then the process may be repeated again and a two-level method becomes a recursive-multilevel method.

As to convergence of a recursive variant of the method the same facts as in section 2.3 can be used. Then we can state that a recursive - multilevel method converges as well and we have proved the following theorem.

Theorem 3.1. *A recursive - multilevel method is convergent.*

4 Cost analysis

4.1 Two levels

The work required for solving a problem of size n is

$$W = K\, n^{(1+\beta)}$$

with K and β positive constants. Let α be a relative overlapping in one dimension. The number of domains is n_d. Then the size of a local problem is

$$n_{loc} = \frac{n}{n_d}(1+\alpha)^d$$

and the work needed for its solving

$$W = K\left(\frac{n}{n_d}(1+\alpha)^d\right)^{(1+\beta)}.$$

Thus the work for one iteration is

$$W = 2n_d\,K\,\left(\frac{n}{n_d}(1+\alpha)^d\right)^{(1+\beta)} = \frac{2(1+\alpha)^{d(1+\beta)}}{n_d^{\beta}}K\,n^{(1+\beta)}$$

and for ℓ iterations on this level

$$W = 2\ell\frac{(1+\alpha)^{d(1+\beta)}}{n_d^{\beta}}K\,n^{(1+\beta)}.$$

4.2 k levels

In the $k-$th level we repeat the previous considerations. We have n_d^k subdomains and the size of one subdomain is $n_{loc,k} = n\left(\frac{1+\alpha}{n_d}\right)^k$. The problem is solved on each subdomain $(2\ell)^k$ times. Total work is

$$W = (2\ell)^k n_d^k\,K\,\left(n\left(\frac{1+\alpha}{n_d}\right)^k\right)^{(1+\beta)} = K\left(2\ell\,n_d^{-\beta}(1+\alpha)^{(1+\beta)}\right)^k n^{(1+\beta)}.$$

4.3 Full recursion

We want to find such k that $n_{loc,k} = 1$. We take the greatest possible k. This is

$$k = \frac{\ln n}{\ln n_d - \ln(1+\alpha)}.$$

We obtain

$$W = K\left(2\ell\,n_d^{-\beta}(1+\alpha)^{(1+\beta)}\right)^{\frac{\ln n}{\ln n_d - \ln(1+\alpha)}} n^{(1+\beta)}$$

which is

$$W = K\exp^{(\ln 2+\ln\ell-\beta\ln n_d+(1+\beta)\ln(1+\alpha))\frac{\ln n}{\ln n_d-\ln(1+\alpha)}+(1+\alpha)\ln n}.$$

This expression can be improved and after some manipulations we get

$$W = K\,n^{(1+\gamma)},$$

with

$$\gamma = \frac{\ln(1+\alpha)+\ln 2+\ln\ell}{\ln n_d - \ln(1+\alpha)}. \tag{13}$$

We can see from (13) that it is possible to achieve γ arbitrary small by an appropriate choice of α, ℓ, n_d.

5 Conclusion

An alternative process to the classic two-level method with a coarse space is proposed in this paper. One of the significant advantages of the method presented here is the fact that we can extremely reduce the memory requirements if this is called for.

References

[1] H. A. Schwarz, *Gessamelte Mathematische Abhandlungen*, volume 2, pages 133 − 143, Springer, Berlin, (1890). First published in Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, *Über einen Grenzübergang durch alternierendes Verfahren*, volume 15, (1870), pp. 272 − 286.

[2] A. Toselli and O. Widlund, *Domain Decomposition Methods - Algorithms and Theory.* Springer-Verlag, Berlin Heidelberg, (2005).

[3] M. Brezina and P. Vaněk, *A black-box iterative solver based on a twolevel Schwarz method.* Computing 63(3), (1999), 233 − 263.

[4] R. Varga, *Matrix Iterative Analysis.* Prentice Hall, first edition, (1962).

RATIO MATHEMATICA
ISSUE N. 30 (2016) pp. 45-58

ISSN (print): 1592-7415
ISSN (online): 2282-8214

Dealing with randomness and vagueness in business and management sciences: the fuzzy-probabilistic approach as a tool for the study of statistical relationships between imprecise variables

Fabrizio Maturo
Department of Management and Business Administration
University G. d'Annunzio, Chieti - Pescara
f.maturo@unich.it

Abstract

In practical applications relating to business and management sciences, there are many variables that, for their own nature, are better described by a pair of ordered values (i.e. financial data). By summarizing this measurement with a single value, there is a loss of information; thus, in these situations, data are better described by interval values rather than by single values. Interval arithmetic studies and analyzes this type of imprecision; however, if the intervals has no sharp boundaries, fuzzy set theory is the most suitable instrument. Moreover, fuzzy regression models are able to overcome some typical limitation of classical regression because they do not need the same strong assumptions. In this paper, we present a review of the main methods introduced in the literature on this topic and introduce some recent developments regarding the concept of randomness in fuzzy regression.

Keywords: fuzzy data; fuzzy regression; fuzzy random variable; tools for business and management sciences

2010 AMS subject classifications: 62J05; 62J86; 03B52; 62A86; 97M10

doi: 10.23755/rm.v30i1.8

1 Introduction

Regression analysis offers a possible solution to study the dependence between two sets of variables. Standard classical statistical linear regressions take the form [27]:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + ... + b_j x_{ij} + ... + b_P x_{iP} + u_i \tag{1}$$

where:

- i=1,.....,N is the i-th observed unit;
- j=1,...,P is the j-th observed variable;
- y_i is the dependent variable, observed on N units;
- x_{ij} are the P independent variables observed on N units;
- b_0 is the crisp intercept and b_j are the P crisp coefficients of the P variables;
- u_i are the random error terms that indicate the deviation of Y from the model;
- y_i, x_{ij}, b_j, u_i are all crisp values.

In classical regression model it is assumed that:

- $E(u_i) = 0$
- $\sigma^2_{u_i} = \sigma^2$
- $\sigma_{u_i,u_j} = 0 \quad \forall \quad i,j \quad with \quad i \neq j$

In matrix form, the classical regression model is expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{2}$$

where $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_N})'$, $\mathbf{b} = (\mathbf{b_1}, \mathbf{b_2}, ..., \mathbf{b_P})'$, $\mathbf{u} = (\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_N})'$ are vectors and $\mathbf{X}$ is a matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & . & . & . & x_{1P} \\ 1 & x_{21} & . & & . & x_{2P} \\ 1 & . & . & . & . & . \\ 1 & . & . & . & . & . \\ 1 & x_{N1} & . & . & . & x_{NP} \end{pmatrix}$$

The aim of statistical regression is to find the set of unknown parameters so that the model gives is a good prediction of the dependent variable Y. The most widely used regression model is the Multiple Linear Regression Model (MLRM), as well as the Ordinary Least Squares (OLS) [12] is the most widespread estimation procedure. Under the OLS assumptions the estimates are BLUE (Best Linear Unbiased Estimator), as stated by the famous Gauss-Markov theorem.

OLS is based on the minimization of the sum of squared deviations:

$$\min \qquad (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} + \mathbf{Xb}) \tag{3}$$

The optimal solution of the minimization problem is the following vector:

$$\widehat{b} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{y} \tag{4}$$

The OLS model is comfortable but its assumptions are every restrictive. Several phenomena violate these assumptions causing biased and inefficient estimators [9]. In particular the assumptions $E(\mathbf{u}|\mathbf{X}) \approx \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I})$ is very strong and rarely it is respected in real phenomena. Moreover in case of "quasi" multicollinearity (many highly correlated explanatory variables), although this does not violate OLS assumption there is a bad impact on the variance of B. In these circumstance the OLS estimators are efficient and unbiased but have large variance, making estimation useless from a practical point of view.

The effects of the quasi multi-collinearity are more evident when the sample size is small [1]. The generally proposed solution consists in removing correlated exploratory variables. This solution is unsatisfying in many applications fields where the user would keep all variables in the model.

In general, we can observe that classical statistical regression has many useful applications but presents troubles in the following situations [26]:

- Number of observations is inadequate (small data set);
- Difficulties verifying distribution assumptions;
- Vagueness in the relationship between input and output variables;
- Ambiguity of events or degree to which they occur;
- Inaccuracy and distortion introduced by linearization;

Furthermore, there are many variables that, for their own nature, are better described by a pair of ordered values, like daily temperatures or financial data. By summarizing this measurement with a single value, there is a loss of information. In these situations data are better described by interval values rather than by single

values. Interval arithmetic studies and analyzes this type of imprecision; but if the intervals has no sharp boundaries, fuzzy set theory is the better tool. In particular fuzzy regression model are able to overcome some typical limitation of classical regression because they don't need the same strong assumptions. Furthermore, some nuanced concepts that exist in economic and social sciences, need to be necessarily treated with linguistic variables, which for their nature, are imprecise concepts.

2 Fuzzy Linear Regression Models (FLR)

There are two general ways, not mutually exclusive, to develop a fuzzy regression model:

- Models where the relationship of the variables is fuzzy;
- Models where the variables themselves are fuzzy;

Therefore fuzzy linear regression (FLR) can be classified in:

- Partially fuzzy linear regression (PFLR), that can be further divided into:
  - PFLR with fuzzy parameters and crisp data;
  - PFLR with fuzzy data and crisp parameters;
- Totally fuzzy linear regression (TFLR) where data and parameters are both fuzzy.

Fuzzy Least Squares Regression is more close to the traditional statistical approach. In fact, following the Least Squares line of thought [13], the aim is to minimize the distance between the observed and the estimated fuzzy data. This approach is referred as Fuzzy Least Squares Regression (FLSR).

In case of one independent variable, the model take the form:

$$\widetilde{y_i} = b_0 + b_1\widetilde{x_i} + \widetilde{u_i} \qquad \text{i=1,2,...,N} \tag{5}$$

where:

- i=1,.....,N is the i-th observed unit;
- y_i is the dependent fuzzy variable, observed on N units;
- x_i is the independent fuzzy variable, observed on N units;

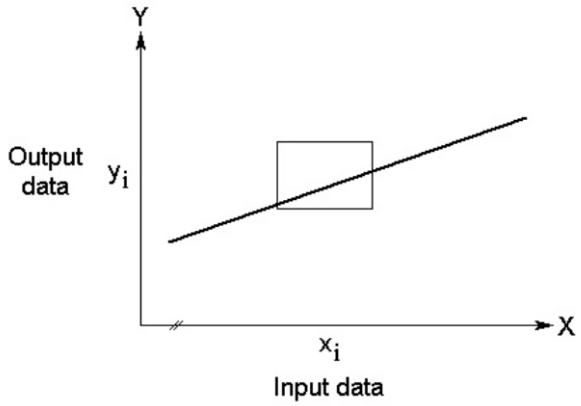

Figure 1: Relation between output and input variables

- b_0 and b_1 are the crisp intercept and the crisp regression coefficient;
- u_i are the fuzzy random error terms;

From a graphical point of view [26] the relation between output and input variables can be represented as shown in Fig.1

In case of several independent variables, the model take the form:

$$\widetilde{y_i} = b_0 + b_1\widetilde{x_{i1}} + b_2\widetilde{x_{i2}} + ... + b_j\widetilde{x_{ij}} + ... + b_P\widetilde{x_{iP}} + \widetilde{u_i} \tag{6}$$

where:

- i=1,.....,N is the i-th observed unit;
- j=1,...,P is the j-th observed variable;
- y_i is the dependent fuzzy variable, observed on N units;
- x_{ij} are the P independent fuzzy variables, observed on N units;
- b_0 is the crisp intercept and b_j are the P crisp regression coefficients measured for the P fuzzy variables;
- u_i are the fuzzy random error terms;

Limiting the reasoning to the first model, the error term can be expressed as follows:

$$\widetilde{u_i} = \widetilde{y_i} - b_0 - b_1\widetilde{x_i} \qquad \text{i=1,2,...,N} \tag{7}$$

Therefore, from a least square perspective, the problem becomes as follows:

$$\min \sum_{i=1}^{N} [\widetilde{y_i} - b_0 - b_1 \widetilde{x_i}]^2 \qquad \text{i=1,2,...,N} \tag{8}$$

Many criteria for measuring this distance have been proposed over the years; however, the most common are two methods:

- The Diamond's approach;
- The compatibility measures approach.

2.1 FLSR using distance measures

The Diamond's approach is also known as fuzzy least squares regression using distance measures. This is the most close approach to the traditional statistical one. Following the Least Squares line of thought, the aim is to minimize the distance between the observed and the estimated fuzzy data, by minimizing the output quadratic error of the model. Since the model contains fuzzy numbers the minimization problem considers distances between fuzzy numbers [5, 17, 20, 15, 19, 18].

Diamond defined an L^2-metric between two triangular fuzzy numbers; it measures the distance between two fuzzy numbers based on their modes, left spread and right spread as follows

$$\begin{gathered} d[(c_1, l_1, r_1), (c_2, l_2, r_2)]^2 = \\ = (c_1 - c_2)^2 + [(c_1 - l_1) - (c_2 - l_2)]^2 + [(c_1 + r_1) - (c_2 + r_2)]^2 \end{gathered} \tag{9}$$

The methods of Diamond are rigorously justified by a projection-type theorem for cones on a Banach space containing the cone of triangular fuzzy numbers, where a Banach space is a normed vector space that is complete as a metric space under the metric $d(x, y) = ||x - y||$ induced by the norm [25].

In the case of crisp coefficients and fuzzy variables, the problem is the following:

$$\min \quad \sum_{i=1}^{N} d[\widetilde{y_i}^* - \widetilde{y_i}]^2 \qquad \text{i=1,2,...,N} \tag{10}$$

where,

$$\widetilde{y_i}^* = b_0 + b_1 \widetilde{x_i} \tag{11}$$

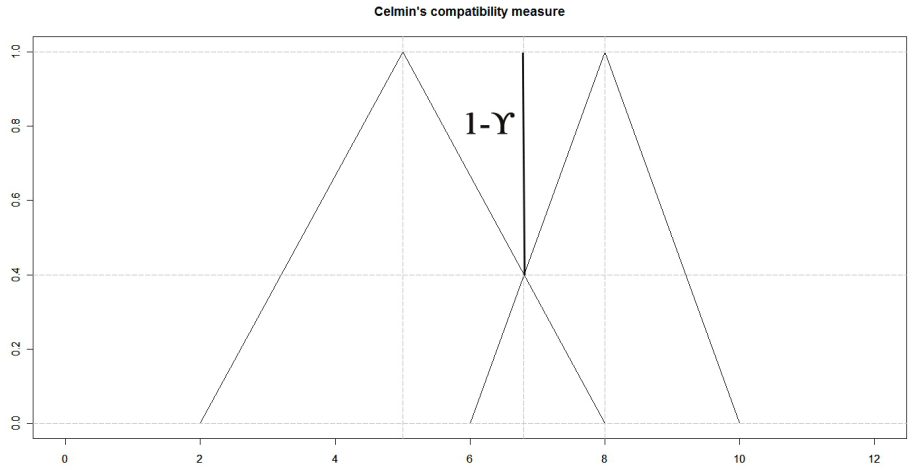

Figure 2: Compatibility measure

therefore the optimization problem can be written as follows:

$$\min \quad \sum_{i=1}^{N} d[b_0 + b_1\widetilde{x}_i - \widetilde{y}_i]^2 \qquad \text{i=1,2,...,N} \tag{12}$$

Using Diamond's difference in this minimization problem, we can obtain the parameters. If the solutions exist, it is necessary to solve a system of six equations in the same number of unknowns; of course, these equations arise from the derivatives being set equal to zero.

2.2 FLSR using compatibility measures

The second type of fuzzy least squares regression model is based on Celmins's compatibility measures [3]. A compatibility measure can defined by

$$\gamma(\widetilde{A}, \widetilde{B}) = maxmin(\mu_A(x), \mu_B(x)) \tag{13}$$

This index is included in the interval [0,1] as shown in Fig. 2. A value of "0" means that the membership functions of the fuzzy numbers A and B are mutually exclusive as shown in Fig. 3. A value of "1" means that the membership functions A coincides with that one of B as shown in Fig.4.

The basic idea is to maximize the overall compatibility between data and model. Thus, the objective may be reformulated in a minimization problem with the following objective function:

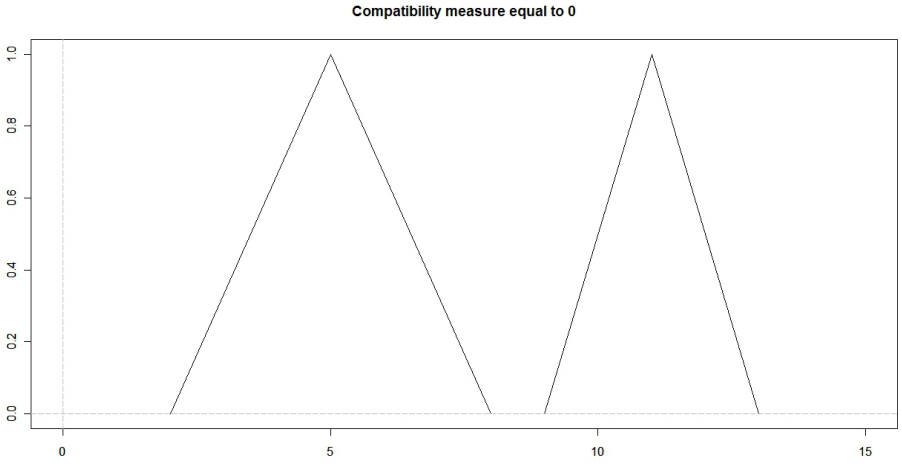

Figure 3: Zero compatibility

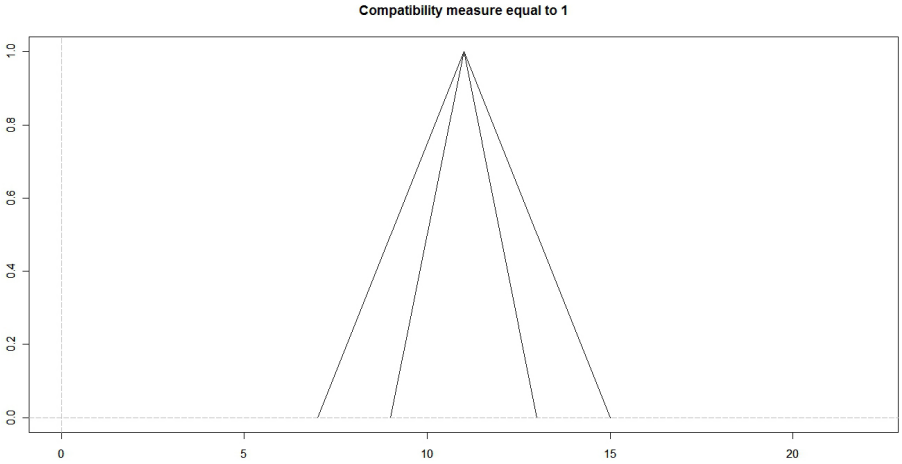

Figure 4: Max compatibility

$$\min \quad \sum_{i=1}^{N}[1-\gamma_i]^2 \qquad \text{i=1,2,...,N} \tag{14}$$

3 Fuzzy regression models with fuzzy random variables

Recent studies have reintroduced the concept of Fuzzy Random Variables (FRVs) [24] firstly introduced by Puri and Ralescu [23]. The need for FRVs arises when the data are not only affected by imprecision but also by randomness [11]. Several papers deal with this topic that it is called fuzzy-probabilistic approach. It consists in explicitly taking into account randomness for estimating the regression parameters and assessing their statistical properties [22, 7, 8].

The membership function of a fuzzy number can be expressed, in term of spreads as:

$$\mu_{\widetilde{A}}(x) = \begin{cases} L\frac{A_m - x}{A_l} & \text{for} \quad x \leq A_m, \quad A_l > 0 \\ 1 & \text{for} \quad x \leq A_m, \quad A_l = 0 \\ R\frac{x - A_m}{A_r} & \text{for} \quad x > A_m, \quad A_r > 0 \\ 0 & \text{for} \quad x > A_m, \quad A_r = 0 \end{cases} \tag{15}$$

where the functions L, R : $\Re - > [0,1]$ are convex upper semi-continuous functions so that L(0) = R(0) = 1 and L(z) = R(z) = 0, for all $z \in \Re/[0,1]$ [6] and A_m is the center, A_l and A_r are the left and the right spread. Of course these functions must be chosen by the researcher in advance and must be the same for all the data.

In particular, for a triangular fuzzy number we obtain:

$$\mu_{\widetilde{A}}(x) = \begin{cases} 0 & \text{for} \quad x \leq A_m - A_l \\ 1 - \frac{A_m - x}{A_l} & \text{for} \quad A_m - A_l \leq x \leq A_m \\ 1 - \frac{x - A_m}{A_r} & \text{for} \quad A_m \leq x \leq A_m + A_r \\ 0 & \text{for} \quad x \geq A_m + A_r \end{cases} \tag{16}$$

A distance for these functions [21] could be:

$$D^2(\widetilde{A}, \widetilde{B}) = (A_m - B_m)^2 + [(A_m - \lambda A_l) - (B_m - \lambda B_l)]^2 + [(A_m + \rho A_r) - (B_m + \rho B_r)]^2 \tag{17}$$

where,

$$\lambda = \int_0^1 L^{-1}(\alpha)d\alpha$$

$$\rho = \int_0^1 R^{-1}(\alpha)d\alpha$$

These functions consider the shape of the membership functions; for example, for triangular fuzzy numbers λ and ρ = 1/2.

To avoid the problem of the non-negativity of the spreads of $\widetilde{Y}$, it is possible to solve a non negative regression problem [14], or to transform the spreads of $\widetilde{Y}$ by means of the centers and the spreads of the P regressors X. In this context, we use the latter method introducing two invertible functions [7]:

$$g : (0, +\infty) \longrightarrow \Re$$

$$h : (0, +\infty) \longrightarrow \Re$$

Thus the linear regression model take the form

$$\begin{cases} Y_m = \mathbf{xb'_m} + \mathbf{a_m} + \mathbf{u_m} \\ g(Y_l) = \mathbf{xb'_l} + \mathbf{a_l} + \mathbf{u_l} \\ h(Y_r) = \mathbf{xb'_r} + \mathbf{a_r} + \mathbf{u_r} \end{cases} \tag{18}$$

where u_l,u_m,u_r are the real valued random variables with $E(u_l|(\mathbf{x})) = \mathbf{0}$, $E(u_m|(\mathbf{x})) = \mathbf{0}$, $E(u_r|(\mathbf{x})) = \mathbf{0}$.

The row vector of length 3p of all the components of the regressors is:

$$\mathbf{x} = (\mathbf{x_{m1}}, \mathbf{x_{l1}}, \mathbf{x_{r1}}, ..., \mathbf{x_{mP}}, \mathbf{x_{lP}}, \mathbf{x_{rP}})$$

The row vectors of length 3p of the parameters related to $\mathbf{x}$ are:

$$\mathbf{b_m} = (\mathbf{b_{mm1}}, \mathbf{b_{ml1}}, \mathbf{b_{mr1}}, ..., \mathbf{b_{mmP}}, \mathbf{b_{mlP}}, \mathbf{b_{mrP}})$$

$$\mathbf{b_l} = (\mathbf{b_{lm1}}, \mathbf{b_{ll1}}, \mathbf{b_{lr1}}, ..., \mathbf{b_{lmP}}), \mathbf{b_{llP}}, \mathbf{b_{lrP}})$$

$$\mathbf{b_r} = (\mathbf{b_{rm1}}, \mathbf{b_{rl1}}, \mathbf{b_{rr1}}, ..., \mathbf{b_{rmP}}, \mathbf{b_{rlP}}, \mathbf{b_{rrP}})$$

The generic element b_{ijt} is the regression coefficient between the component $i\epsilon[m, l, r]$ of $\widetilde{Y}$, where m,l,r refer to center and the transformed spread of $\widetilde{Y}$, and the component $j\epsilon[m, l, r]$ of the regressor $\widetilde{x}_t$ with t=1,....,P, where m,l,r refer to

the corresponding center, left spread and right spread. For example b_{mr2} is the relationship between the right spread of $\widetilde{x}_2$ and the center of Y. Of course, a_m, a_l, and a_r are the intercepts.

The covariance matrix of $\mathbf{x}$ is denoted by:

$$\Sigma_{(\mathbf{x})} = E[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{x} - \mathbf{E}(\mathbf{x}))] \tag{19}$$

The covariance matrix of u_m, u_l, u_r is indicated with Σ and contains the variances $\sigma^2_{u_m}$, $\sigma^2_{u_l}$ and $\sigma^2_{u_r}$.

The regression parameters can be expressed as:

$$\begin{aligned}
&\mathbf{b_m}' = [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{Y_m} - \mathbf{E}(\mathbf{Y_m})] \\
&\mathbf{b_l}' = [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{g(Y_l)} - \mathbf{E}(\mathbf{g(Y_l)})] \\
&\mathbf{b_r}' = [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{h(Y_r)} - \mathbf{E}(\mathbf{h(Y_r)})] \\
&a_m = E(Y_m|\mathbf{x}) - [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{Y_m} - \mathbf{E}(\mathbf{Y_m})] \\
&a_l = E(g(Y_l)|\mathbf{x}) - [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{g(Y_l)} - \mathbf{E}(\mathbf{g(Y_l)})] \\
&a_r = E(h(Y_r)|\mathbf{x}) - [\mathbf{\Sigma_{(x)}}]^{-1}\mathbf{E}[(\mathbf{x} - \mathbf{E}(\mathbf{x}))'(\mathbf{h(Y_r)} - \mathbf{E}(\mathbf{h(Y_r)})]
\end{aligned} \tag{20}$$

Since the total variation of the response can be written in terms of variances and covariances of real random variables, it can be decomposed in the variation not depending on the model and that explained by the model. Thus, we can obtain a determination coefficient for the fuzzy model based on the decomposition of the total variance given by:

$$\begin{aligned}
&E[D^2(Y_t, E(Y_t)] = E[D^2(Y_t, E(Y_t|\mathbf{x})]+ \\
&+ E[D^2(E(Y_t|\mathbf{x}, \mathbf{E}(\mathbf{Y_t}))]
\end{aligned} \tag{21}$$

Therefore, the linear determination coefficient R^2 can be defined as:

$$\begin{aligned}
R^2 &= \frac{E[D^2(E((Y_t|\mathbf{x}), \mathbf{E}(\mathbf{Y_t}))]}{E[D^2(Y_t, E(Y_t)]} = \\
&= 1 - \frac{E[D^2(Y_t, E(Y_t|\mathbf{x})]}{E[D^2(Y_t, E(Y_t)]}
\end{aligned} \tag{22}$$

The meaning of this index is the same of the classical regression model. The estimation problem of the regression parameters is faced by means of the LS criterion. As shown in [6], applying the appropriate substitutions and using the concept of distance between two fuzzy numbers, like in the Diamond's approach, it is possible to find the equation of the estimators of all parameters.

4 Conclusions

Fuzzy regression models are able to overcome some limitations of classical regression because they do not need the same strong assumptions. In this paper, we have presented a review of the main methods introduced in the literature on this topic and some recent developments regarding the concept of randomness in fuzzy regression. In practical applications relating to business and management sciences, fuzzy regression models with fuzzy random variables are more suitable for the characteristics of the data. However, some of the main issues of Zadeh's operations with these models are the following: the addition and the multiplication between fuzzy numbers lead to a considerable increase of the spreads; the multiplication of two symmetric fuzzy numbers does not provide a symmetric fuzzy number or at least a fuzzy number with equal spreads; spreads of Zadeh's product depend heavily on the modes of the numbers; some important algebraic properties, such as the distributive property, are valid only in particular circumstances; the product of two triangular fuzzy numbers does not provide a triangular fuzzy number. Therefore, alternative operations in order to overcome some problems connected to the addition and the product between fuzzy numbers in fuzzy linear regression models are strongly necessary. Moreover, our research prospects include considering finite geometric spaces [16, 2], multivalued functions [4] and algebraic hyperoperations [10] in fuzzy regression models.

References

[1] Achen, C., 1982. Interpreting and using regression. Sage Publications, California.

[2] Ameri, R., Nozari, T., 2012. Fuzzy hyperalgebras and direct product, Ratio Mathematica. 24.

[3] Celmins, A., 1987. Least squares model fitting to fuzzy vector data. Fuzzy Sets and Systems 22(3), 245–269.

[4] Corsini, P., Mahjoob, R., 2010. Multivalued functions, Fuzzy subsets and Join spaces. Ratio Mathematica 20, 1–41.

[5] Diamond, P., 1988. Fuzzy least squares. Information Sciences 46 (3), 141–157.

[6] Ferrraro, M., Colubi, A., Gonzales-Rodriguez, A., Coppi, R., 2010. A linear regression model for imprecise response. Int J Approx Reason 21, 759–770.

[7] Ferrraro, M., Colubi, A., Gonzales-Rodriguez, A., Coppi, R., 2011. A determination coefficient for a linear regression model with imprecise response. Environmetrics 22, 516–529.

[8] Gonzales-Rodriguez, A., Blanco, A., Colubi, A., Lubiano, M., 2009. Estimation of a simple linear regression model for fuzzy random variables. Fuzzy Sets Syst 160, 357–370.

[9] Gujarati, D., 2003. Basic Econometrics. McGraw-Hill, New York.

[10] Hošková-Mayerová, Š., Maturo, A., 2016. Fuzzy sets and algebraic hyperoperations to model interpersonal relations, in: Recent Trends in Social Systems: Quantitative Theories and Quantitative Models. Springer Nature, pp. 211–221. URL: http://dx.doi.org/10.1007/978-3-319-40585-8_19, doi:10.1007/978-3-319-40585-8_19.

[11] Klir, G., 2006. Uncertainty and information: foundations of generalized information theory. Wiley, New York.

[12] Kratschmer, V., 2006a. Strong consistency of least-squares estimation in linear regression models with vague concepts. J Multivar Anal 97, 633–654.

[13] Kratschmer, V., 2006b. Strong consistency of least-squares estimation in linear regression models with vague concepts. J Multivar Anal 97, 1044–1069.

[14] Lawson, C., R.J. Hanson, 1995. Solving least squares problems. Classics in applied mathematics 15.

[15] Maturo, A., Maturo, F., 2013. Research in social sciences: Fuzzy regression and causal complexity, in: Multicriteria and Multiagent Decision Making with Applications to Economics and Social Sciences. Springer, pp. 237–249. URL: http://dx.doi.org/10.1007/978-3-642-35635-3_18, doi:10.1007/978-3-642-35635-3_18.

[16] Maturo, A., Maturo, F., 2014. Finite geometric spaces, steiner systems and cooperative games. Analele Universitatii "Ovidius" Constanta - Seria Matematica 22. URL: http://dx.doi.org/10.2478/auom-2014-0015, doi:10.2478/auom-2014-0015.

[17] Maturo, A., Maturo, F., 2016. Fuzzy events, fuzzy probability and applications in economic and social sciences, in: Recent Trends in Social Systems: Quantitative Theories and Quantitative Models. Springer Nature, pp. 223–233. URL: http://dx.doi.org/10.1007/978-3-319-40585-8_20, doi:10.1007/978-3-319-40585-8_20.

[18] Maturo, F., 2016. FUZZINESS: TEORIE E APPLICAZIONI. Aracne Editrice, Roma, Italy. chapter LA REGRESSIONE FUZZY. pp. 99–110.

[19] Maturo, F., Fortuna, F., 2016. Bell-shaped fuzzy numbers associated with the normal curve, in: Topics on Methodological and Applied Statistical Inference. Springer. URL: http://dx.doi.org/10.1007/978-3-319-44093-4_13, doi:10.1007/978-3-319-44093-4_13.

[20] Maturo, F., Hošková-Mayerová, Š., 2016. Fuzzy regression models and alternative operations for economic and social sciences, in: Recent Trends in Social Systems: Quantitative Theories and Quantitative Models. Springer Nature, pp. 235–247. URL: http://dx.doi.org/10.1007/978-3-319-40585-8_21, doi:10.1007/978-3-319-40585-8_21.

[21] M.S. Yang, C.K., 1996. On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets and Syst 84, 49–60.

[22] Nather, W., 2006. Regression with fuzzy random data. Comput Stat Data Anal 51, 235–252.

[23] Puri, M., Ralescu, D., 1986. Fuzzy random variables. J Math Anal Appl 114, 409–422.

[24] Ramos-Guajardo, A., Colubi, A., Gonzales-Rodriguez, A., 2010. One-sample tests for a generalized frÃ¨chet variance of a fuzzy random variable. Metrika 71, 185–202.

[25] Shapiro, A., 2004. Fuzzy regression and the term structure of interest rates revisited. Proceedings of the 14th International AFIR Colloquium Vol.1, 29–45.

[26] Shapiro, A., 2005. Fuzzy regression models. ARC .

[27] Stock, J., Watson, M., 2009. Introduction to econometrics. Pearson Addison Wesley.

RATIO MATHEMATICA
ISSUE N. 30 (2016) pp. 59-66

ISSN (print): 1592-7415
ISSN (online): 2282-8214

The sum of the series of reciprocals of the quadratic polynomial with different negative integer roots

Radovan Potůček

Department of Mathematics and Physics, Faculty of Military Technology,
University of Defence, Brno, Czech Republic
Radovan.Potucek@unob.cz

Abstract

This contribution, which is a follow-up to author's paper [1] and [2] dealing with the sums of the series of reciprocals of some quadratic polynomials, deals with the series of reciprocals of the quadratic polynomials with different negative integer roots. We derive the formula for the sum of this series and verify it by some examples evaluated using the basic programming language of the CAS Maple 16.

Keywords: sequence of partial sums, telescoping series, harmonic number, computer algebra system Maple.

2010 AMS subject classifications: 40A05, 65B10

doi: 10.23755/rm.v30i1.12

1 Introduction and basic notions

Let us recall the basic terms, concepts and notions. For any sequence $\{a_k\}$ of numbers the associated *series* is defined as the sum

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \cdots .$$

The *sequence of partial sums* $\{s_n\}$ associated to a series $\sum_{k=1}^{\infty} a_k$ is defined for each n as the sum of the sequence $\{a_k\}$ from a_1 to a_n, i.e.

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \,.$$

The series $\sum_{k=1}^{\infty} a_k$ *converges* to a limit s if and only if the sequence of partial sums $\{s_n\}$ converges to s, i.e. $\lim_{n\to\infty} s_n = s$. We say that the series $\sum_{k=1}^{\infty} a_k$ has a *sum* s and write $\sum_{k=1}^{\infty} a_k = s$.

The *telescoping series* is any series where nearly every term cancels with a preceding or following term, so its partial sums eventually only have a fixed number of terms after cancellation. Telescoping series are not very common in mathematics but are interesting to study. The method of changing series whose terms are rational functions into telescoping series is that of transforming the rational functions by the method of partial fractions.

For example, the series $\sum_{k=1}^{\infty} \frac{1}{k^2+k}$ has the general nth term

$$a_n = \frac{1}{n(n+1)} = \frac{A}{n} + \frac{B}{n+1}.$$

After removing the fractions we get the equation $1 = A(n+1) + Bn$. Solving it for A and B we obtain $a_n = 1/n - 1/(n+1)$. After that we arrange the terms of the nth partial sum $s_n = a_1 + a_2 + \cdots + a_n$ in a form where can be seen what is cancelling. Then we find the limit of the sequence of the partial sums s_n in order to find the sum s of the infinite telescoping series as $s = \lim_{n\to\infty} s_n$. In our case we get

$$s_n = \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 1 - \frac{1}{n+1}\,.$$

So we have $s = \lim_{n\to\infty} \left(1 - \frac{1}{n+1}\right) = 1$.

The *nth harmonic number* is the sum of the reciprocals of the first n natural numbers:

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{k=1}^{n} \frac{1}{k}\,.$$

The values of the sequence $\{H_n - \ln n\}$ decrease monotonically towards the limit $\gamma \doteq 0.57721566$, which is so-called the *Euler-Mascheroni constant*. Basic information about harmonic numbers can be found e.g. in the web-sites [3] or [4], interesting information are included e.g. in the paper [5]. First ten values of the harmonic numbers are presented in this table:

n	1	2	3	4	5	6	7	8	9	10
H_n	1	$\frac{3}{2}$	$\frac{11}{6}$	$\frac{25}{12}$	$\frac{137}{60}$	$\frac{49}{20}$	$\frac{363}{140}$	$\frac{761}{280}$	$\frac{7129}{2520}$	$\frac{7381}{2520}$

2 The sum of the series of reciprocals of the quadratic polynomial with different negative integer roots

Now, we deal with the series formed by reciprocals of the normalized quadratic polynomial $(k-a)(k-b)$, where $a < b < 0$ are integers. Let us consider the series

$$\sum_{k=1}^{\infty} \frac{1}{(k-a)(k-b)}\,,$$

and determine its sum $s(a,b)$.

We express the nth term a_n of this series as the sum of two partial fractions

$$a_n = \frac{1}{(n-a)(n-b)} = \frac{A}{n-a} + \frac{B}{n-b}\,.$$

From the equality of two linear polynomials $1 = A(n-b) + B(n-a)$ for $n = a$ we get $A = 1/(a-b)$ and for $n = b$ we get $B = 1/(b-a) = -1/(a-b)$. So we have

$$a_n = \frac{1}{a-b}\left(\frac{1}{n-a} - \frac{1}{n-b}\right) = \frac{1}{b-a}\left(\frac{1}{n-b} - \frac{1}{n-a}\right). \tag{1}$$

For the nth partial sum of the given series so we get

$$s_n = \frac{1}{b-a}\left[\left(\frac{1}{1-b} - \frac{1}{1-a}\right) + \left(\frac{1}{2-b} - \frac{1}{2-a}\right) + \cdots \right.$$
$$\left. \cdots + \left(\frac{1}{n-1-b} - \frac{1}{n-1-a}\right) + \left(\frac{1}{n-b} - \frac{1}{n-a}\right)\right].$$

The first terms that cancel each other will be obviously the terms for which for the suitable index ℓ it holds $1/(1-a) = 1/(\ell-b)$. Therefore the last term from the beginning of the expression of the nth partial sum s_n, which will not cancel, will

be the term $1/(-a)$, so that the first terms from the beginning of the expression the sum s_n, which will not cancel, will be the terms generating the sum

$$\frac{1}{1-b}+\frac{1}{2-b}+\cdots+\frac{1}{-a}\,.$$

Analogously, the last terms that cancel each other will be the terms for which for the suitable index m it holds $1/(n-b) = 1/(m-a)$. Therefore the first term from the ending of the expression of the nth partial sum s_n, which will not cancel, will be the term $1/(n+1-b)$, so that the last terms from the ending in the expression of the sum s_n, which will not cancel, will be the terms generating the sum

$$-\frac{1}{n+1-b}-\frac{1}{n+2-b}-\cdots-\frac{1}{n-a}\,.$$

After cancelling all the inside terms with the opposite signs we get the nth partial sum

$$s_n=\frac{1}{b-a}\left(\frac{1}{1-b}+\frac{1}{2-b}+\cdots+\frac{1}{-a}-\frac{1}{n+1-b}-\frac{1}{n+2-b}-\cdots-\frac{1}{n-a}\right).$$

Because for integer c it holds $\lim\limits_{n\to\infty}\dfrac{1}{n+c}=0$, then the searched sum, where $a<b<0$, is

$$\begin{aligned}s(a,b)=\lim_{n\to\infty}s_n&=\frac{1}{b-a}\left(\frac{1}{1-b}+\frac{1}{2-b}+\cdots+\frac{1}{-a}\right)=\\&=\frac{1}{b-a}\left[\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{-a}-\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{-b}\right)\right],\end{aligned}$$

so we get

Theorem 2.1. The series $\displaystyle\sum_{k=1}^{\infty}\frac{1}{(k-a)(k-b)}$, where $a<b<0$ are integers, has the sum

$$s(a,b)=\frac{1}{b-a}\big(H_{-a}-H_{-b}\big),\tag{2}$$

where H_n is the nth harmonic number.

Corolary 2.1. For the sum $s(a,b)$ above it obviously hold:

1. $s(a,b)=s(b,a)$,
2. $s(a,a+1)=H_{-a}-H_{-a-1}=\dfrac{1}{-a}$,

3. $s(a,a+i)=\dfrac{1}{i}\big(H_{-a}-H_{-a-i}\big)=\dfrac{1}{i}\left(\dfrac{1}{-a-i+1}+\dfrac{1}{-a-i+2}+\cdots+\dfrac{1}{-a}\right)$, $i\in\mathbb{N}$.

Remark 2.1. *Let us note, that the formula* (2) *holds also in the case* $a<b=0$. *Because* H_0 *is defined as* 0, *it has the form*

$$s(a,0)=\frac{1}{0-a}\big(H_{-a}-H_0\big)=\frac{H_{-a}}{-a}\,. \tag{3}$$

Example 2.1. *The series*

$$\sum_{k=1}^{\infty}\frac{1}{\big(k-(-5)\big)\big(k-(-2)\big)}=\sum_{k=1}^{\infty}\frac{1}{(k+2)(k+5)}\,,$$

where $a=-5$, $b=-2$, *has the* n*th partial sum*

$$s_n=\frac{1}{3}\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{5}-\frac{1}{n+3}-\frac{1}{n+4}-\frac{1}{n+5}\right).$$

By the relation $s(-5,-2)=\lim\limits_{n\to\infty}s_n$, *since* $\lim\limits_{n\to\infty}\dfrac{1}{n+c}=0$, *or by Theorem* 2.1 *we get its sum*

$$s(-5,-2)=\frac{1}{3}\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right)=\frac{1}{3}\big(H_5-H_2\big)=\frac{1}{3}\left(\frac{137}{60}-\frac{3}{2}\right)=\frac{47}{180}=0.26\overline{1}.$$

Example 2.2. *The series*

$$\sum_{k=1}^{\infty}\frac{1}{\big(k-(-5)\big)k}=\sum_{k=1}^{\infty}\frac{1}{k(k+5)}\,,$$

where $a=-5$, $b=0$, *has the* n*th partial sum*

$$s_n=\frac{1}{5}\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}-\frac{1}{n+1}-\frac{1}{n+2}-\frac{1}{n+3}-\frac{1}{n+4}-\frac{1}{n+5}\right).$$

By the relation $s(-5,0)=\lim\limits_{n\to\infty}s_n$, *since* $\lim\limits_{n\to\infty}\dfrac{1}{n+c}=0$, *or by Theorem* 2.1, *or by the Remark* 2.1 *we get its sum*

$$s(-5,0)=\frac{1}{5}\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right)=\frac{H_5}{5}=\frac{137/60}{5}=\frac{137}{300}=0.45\overline{6}\,.$$

3 Numerical verification

We solve the problem to determine the values of the sum

$$s(a,b) = \sum_{k=1}^{\infty} \frac{1}{(k-a)(k-b)}$$

for $a = -1, -2, \ldots, -9$ and $b = a+1, a+2, \ldots, -8$. We use on the one hand an approximative direct evaluation of the sum

$$s(a,b,t) = \sum_{k=1}^{t} \frac{1}{(k-a)(k-b)},$$

where $t = 10^8$, using the basic programming language of the computer algebra system Maple 16, and on the other hand the formula (2) for evaluation the sum $s(a,b)$. We compare $45 = 9 + 8 + \cdots + 1$ pairs of these ways obtained sums $s(a,b,10^8)$ and $s(a,b)$ to verify the formula (2). We use following simple procedures `hnum`, `rp2abneg` and two **for** statements:

```
hnum:=proc(h)
    local i,s; s:=0;
    if h=0 then s:=0 else
    for i from 1 to h do
        s:=s+1/i;
    end do;
    end if;
end proc:
rp2abneg:=proc(a,b,n)
    local k,sab,sumab; sumab:=0;
    sab:=(hnum(-a)-hnum(-b))/(b-a);
    print("n=",n,"s(",a,b,")=",evalf[20](sab));
    for k from 1 to n do
        sumab:=sumab+1/((k-a)*(k-b));
    end do;
    print("sum(",a,b,")=",evalf[20](sumab),
        "diff=",evalf[20](abs(sumab-sab)));
end proc:
for i from -1 by -1 to -9 do
    for j from i+1 by -1 to -8 do
        rp2abneg(i,j,100000000);
    end do;
end do;
```

The approximative values of the sums $s(a,b)$ rounded to 3 decimals obtained by these procedures are written into the following table:

$s(a,b)$	$a=-1$	$a=-2$	$a=-3$	$a=-4$	$a=-5$	$a=-6$	$a=-7$	$a=-8$	$a=-9$
$b=0$	1.000	0.750	0.611	0.521	0.457	0.408	0.370	0.340	0.314
$b=-1$	$\times$	0.500	0.417	0.361	0.321	0.290	0.266	0.245	0.229
$b=-2$	$\times$	$\times$	0.333	0.292	0.261	0.238	0.219	0.203	0.190
$b=-3$	$\times$	$\times$	$\times$	0.250	0.225	0.206	0.190	0.177	0.166
$b=-4$	$\times$	$\times$	$\times$	$\times$	0.2000	0.183	0.170	0.159	0.149
$b=-5$	$\times$	$\times$	$\times$	$\times$	$\times$	0.167	0.155	0.145	0.136
$b=-6$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0.143	0.134	0.126
$b=-7$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0.125	0.118
$b=-8$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0.111

Computation of 45 couples of the sums $s(a,b,10^8)$ and $s(a,b)$ took over 18 minutes. The absolute errors, i.e. the differences $\left|s(a,b)-s(a,b,10^8)\right|$, have here place value about 10^{-8}.

4 Conclusion

We dealt with the sum of the series of reciprocals of the quadratic polynomials with different negative integer roots a and b, i.e. with the series

$$\sum_{k=1}^{\infty} \frac{1}{(k-a)(k-b)}\,,$$

where $a < b < 0$ are integers. We derived that the sum $s(a,b)$ of this series is given by the formula

$$s(a,b) = \frac{1}{b-a}\big(H_{-a} - H_{-b}\big),$$

where H_n is the nth harmonic number. We verified this result by computing 45 various sums by using the CAS Maple 16.

We also stated that this formula holds also for $a < b = 0$, when it has the form

$$s(a,0) = \frac{1}{0-a}\big(H_{-a} - H_0\big) = \frac{H_{-a}}{-a}\,.$$

The series of reciprocals of the quadratic polynomials with different negative integer roots so belong to special types of infinite series, such as geometric and telescoping series, which sums are given analytically by means of a simple formula.

References

[1] R. Potůček, *The sums of the series of reciprocals of some quadratic polynomials*. In: Proceedings of AFASES 2010, 12th International Conference "Scientific Research and Education in the Air Force" (CD-ROM). Brasov, Romania, 2010, p. 1206-1209. ISBN 978-973-8415-76-8.

[2] R. Potůček, *The sum of the series of reciprocals of the quadratic polynomials with double non-positive integer root*. In: Proceedings of the 15th Conference on Applied Mathematics APLIMAT 2016. Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, 2016, p. 919-925. ISBN 978-80-227-4531-4.

[3] Wikipedia contributors: *Harmonic number*. Wikipedia, The Free Encyclopedia, [online], [cit. 2016-09-01]. Available from: https://en.wikipedia.org/wiki/Harmonic_number.

[4] E. W. Weisstein, *Harmonic Number*. From MathWorld – A Wolfram Web Resource, [online], [cit. 2016-09-01]. Available from: http://mathworld.wolfram.com/HarmonicNumber.html

[5] A. T. Benjamin, G. O. Preston, and J. J. Quinn, *A Stirling Encounter with Harmonic Numbers*. Mathematics Magazine 75 (2), 2002, p. 95 –103, [online], [cit. 2016-09-01]. Available from: https://www.math.hmc.edu/∼benjamin/papers/harmonic.pdf

RATIO MATHEMATICA
ISSUE N. 30 (2016) pp. 67-81

ISSN (print): 1592-7415
ISSN (online): 2282-8214

On Hyper Hoop-algebras

Rajabali Borzooei[(a)], **Hamidreza Varasteh**[(b)], **Keivan Borna**[(c)]

[(a)] *Department of Mathematics, Shahid Beheshti University, Tehran, Iran*
borzooei@sbu.ac.ir

[(b)] *Department of research and technology, Kharazmi University, Tehran, Iran*
varastehhamid@gmail.com

[(c)] *Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran*
borna@khu.ac.ir

Abstract

In this paper, we apply the hyper structure theory to hoop-algebras and introduce the notion of (quasi) hyper hoop-algebra which is a generalization of hoop-algebra and investigate some related properties. We also introduce the notion of (weak)filters on hyper hoop-algebras, and give several properties of them. Finally, we characterize the (weak) filter generated by a non-empty subset of a hyper hoop-algebra.

Keywords: Hoop-algebra, (quasi) hyper hoop-algebra, (weak) filter.

2010 AMS subject classifications: 20N20, 14L17, 97H50, 03G25

doi: 10.23755/rm.v30i1.14

1 Introduction

Hoop-algebras or Hoops are naturally ordered commutative residuated integral monoids were originally introduced by Bosbach in [7] under the name of complementary semigroups. It was proved that a hoop is a meet-semilattice. Hoop-algbras then investigated by Büchi and Owens in an unpublished manuscript [8] of 1975, and they have been studied by Blok and Ferreirim [2],[3], and Aglianò et.al. [1], among others. The study of hoops is motivated by their occurrence both in universal algebra and algebraic logic. Typical examples of hoops include both Brouwerian semilattices and the positive cones of lattice ordered abelian groups,

while hoops structurally enriched with normal multiplicative operators naturally generalize the normal Boolean algebras with operators. In recent years, hoop theory was enriched with deep structure theorems. Many of these results have a strong impact with fuzzy logic. Particularly, from the structure theorem of finite basic hoops one obtains an elegant short proof of the completeness theorem for propositional basic logic introduced by Hájek in [12]. The algebraic structures corresponding to Hájek's propositional (fuzzy) basic logic, BL-algebras, are particular cases of hoops and MV-algebras, product algebras and Gödel algebras are the most known classes of BL-algebras. Hypersructure theory was introduced in 1934[13], when Marty at the 8th congress of scandinavian mathematicians, gave the definition of hypergroup and illustrated some applications and showed its utility in the study of groups, algebraic functions, and rational fraction. Till now, the hyperstructures have been studied from the theoretical point of view for their applications to many subject of pure and applied mathematics. Some fields of applications of the mentioned structures are lattices, graphs, coding, ordered sets, median algebra, automata, and cryptography[9]. Many researchers have worked on this area. R.A.Borzooei et al. introduced and studied hyper residuated lattices and hyper K-algebras in [4],[6] and S.Ghorbani et al.[11], applied the hyper structures to MV-algebras and introduced the concept of hyper MV-algebra, which is a generalization of MV-algebra.
In this paper we construct and introduce the notion of (quasi) hyper hoop-algebra which is a generalization of hoop-algebra. Then we study some properties of this structure. We also introduce the notion of (weak)filters on hyper hoop-algebras, and give several properties of them. Finally, we characterize the (weak) filter generated by a non-empty subset of a hyper hoop-algebra.

2 Preliminaries

In this section, we recall some definitions and theorems in hoop algebras which will be needed in this paper.

Definition 2.1. [1] A *hoop-algebra* or a *hoop* is an algebra $(A, *, \to, 1)$ of the type (2, 2, 0) such that, for all $x, y, z \in A$:
(H1) $(A, *, 1)$ is a commutative monoid,
(H2) $x \to x = 1$,
(H3) $(x \to y) * x = (y \to x) * y$,
(H4) $x \to (y \to z) = (x * y) \to z$.

On the hoop A, if we define $x \leq y$ iff $x \to y = 1$, for any $x, y \in A$, it is proved that $\leq$ is a partial order on A. A hoop A is bounded if there is an element

$0 \in A$ such that $0 \leq x$ for all $x \in A$.

Proposition 2.2. [1] Let A be a hoop-algebra. Then for every $a, b, c \in A$ the following hold:
(i) $(A, \leq)$ is a $\wedge$-semilattice and $a \wedge b = a * (a \to b)$,
(ii) $a \leq b \to c$ iff $a * b \leq c$,
(iii) $1 \to a = a$,
(iv) $a \to 1 = 1$, i.e. $a \leq 1$,
(v) $a \to b \leq (c \to a) \to (c \to b)$,
(vi) $a \leq b \to a$,
(vii) $a \leq (a \to b) \to b$,
(viii) $a \to (b \to c) = b \to (a \to c)$,
(ix) $a \to b \leq (b \to c) \to (a \to c)$,
(x) $a \leq b$ implies $b \to c \leq a \to c$ and $c \to a \leq c \to b$.

Now, we recall some basic notions of the hypergroup theory from [9]:
Let H be a non-empty set. A hypergroupoid is a pair $(H, \odot)$, where $\odot : H \times H \longrightarrow P(H) \setminus \emptyset$ is a binary hyperoperation on H. If $a \odot (b \odot c) = (a \odot b) \odot c$ holds, for all $a, b, c \in H$ then $(H, \odot)$ is called a semihypergroup, and it is said to be commutative if $\odot$ is commutative. An element $1 \in H$ is called a unit, if $a \in 1 \odot a \cap a \odot 1$, for all $a \in H$ and is called a scaler unit, if $\{a\} = 1 \odot a = a \odot 1$, for all $a \in A$. If the reproduction axiom $a \odot H = H = H \odot a$, for any element $a \in H$ is satisfied, then the pair $(H, \odot)$ is called a hypergroup. Note that if $A, B \subseteq H$, then $A \odot B = \bigcup_{a \in A, b \in B} (a \odot b)$.

3 Hyper hoop-algebras

Definition 3.1. A*quasi hyper hoop-algebra* or briefly, a *quasi hyper hoop* is a non-empty set A endowed with two binary hyperoperations $\odot, \to$ and a constant 1 such that, for all $x, y, z \in A$ satisfying the following conditions:
(HHA1) $(A, \odot, 1)$ is a commutative semihypergroup with 1 as the unit,
(HHA2) $1 \in x \to x$,
(HHA3) $(x \to y) \odot x = (y \to x) \odot y$,
(HHA4) $x \to (y \to z) = (x \odot y) \to z$,
A quasi hyper hoop $(A, \odot, \to, 1)$ is called a *hyper hoop* if the following hold;
(HHA5) $1 \in x \to 1$,
(HHA6) if $1 \in x \to y$ and $1 \in y \to x$ then $x = y$,
(HHA7) if $1 \in x \to y$ and $1 \in y \to z$ then $1 \in x \to z$.

In the sequel we will refer to the (quasi) hyper hoop $(A, \odot, \to, 1)$ by its universe A. On (quasi) hyper hoop A, for any $x, y \in A$, we define $x \leq y$ if and

only if $1 \in x \to y$. If A is a hyper hoop, it is easy to see that $\leq$ is a partial order relation on A. Moreover, for all $B, C \subseteq A$ we define $B \ll C$ iff there exist $b \in B$ and $c \in C$ such that $b \leq c$ and define $B \leq C$ iff for any $b \in B$ there exists $c \in C$ such that $b \leq c$. A (quasi) hyper hoop A is bounded if there is an element $0 \in A$ such that $0 \leq x$, for all $x \in A$.

In the following examples, we will show that the conditions (HHA5), (HHA6), and (HHA7) are independent from the other conditions.

Example 3.2. (i) Let $A = \{1, a, b\}$. Define the hyperoperations $\odot$, and $\to$ on A as follows:

$\odot$	1	a	b
1	$\{1\}$	$\{a\}$	$\{a, b\}$
a	$\{a\}$	$\{a\}$	$\{a, b\}$
b	$\{a, b\}$	$\{a, b\}$	$\{b\}$

$\to$	1	a	b
1	$\{1\}$	$\{a, b\}$	$\{b\}$
a	$\{b\}$	$\{1, a, b\}$	$\{b\}$
b	$\{1, a, b\}$	$\{1, a, b\}$	$\{1, a, b\}$

Then $(A, \odot, \to, 1)$ is a quasi hyper hoop, but doesn't satisfy the condition (HHA5). Since $1 \notin a \to 1$.

(ii) Let $A = \{1, a, b\}$. Define the hyperoperations $\odot$ and $\to$ on A as follows:

$\odot$	1	a	b
1	$\{1\}$	$\{a\}$	$\{b\}$
a	$\{a\}$	$\{a\}$	$\{a\}$
b	$\{b\}$	$\{a\}$	$\{1\}$

$\to$	1	a	b
1	$\{1, b\}$	$\{a\}$	$\{1, b\}$
a	$\{1, b\}$	$\{1, b\}$	$\{1, b\}$
b	$\{1, b\}$	$\{a\}$	$\{1, b\}$

Then $(A, \odot, \to, 1)$ is a quasi hyper hoop, but doesn't satisfy the condition (HHA6). Since $1 \in b \to 1$ and $1 \in 1 \to b$, but $1 \neq b$.

(iii) Let $A = \{1, a, b, c\}$. Define hyperoperations $\odot$ and $\to$ on A as follows:

$\odot$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{a\}$	$\{a\}$	$\{a, b\}$	$\{a, b\}$
b	$\{b\}$	$\{a, b\}$	$\{b\}$	$\{b\}$
c	$\{c\}$	$\{a, b\}$	$\{b\}$	$\{c\}$

$\to$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{1\}$	$\{a, 1\}$	$\{1, b, c\}$	$\{c\}$
b	$\{1\}$	$\{a\}$	$\{1, b, c\}$	$\{1, b, c\}$
c	$\{1\}$	$\{a\}$	$\{b\}$	$\{1, b, c\}$

Then $(A, \odot, \to, 1)$ is a quasi hyper hoop, but doesn't satisfy the condition (HHA7). Because $1 \in a \to b$ and $1 \in b \to c$ but $1 \notin a \to c$.

In the following, we give some examples of (quasi) hyper hoop algebras.

Example 3.3. (i) In any (quasi) hyper hoop $(A, \odot, \to, 1)$, if $x \odot y$ and $x \to y$ are singletons, for any $x, y \in A$, then $(A, \odot, \to, 1)$ is a hoop. Then (quasi) hyper hoops are generalizations of hoops.

(ii) Let $A = \{1\}$. If we consider $1 \to 1 = \{1\}$, $1 \odot 1 = \{1\}$, then it is clear that $A = (A, \odot, \to, 1)$ is a (quasi) hyper hoop.

(iii) Let $A = \{1, a\}$. Define the hyperoperations $\odot$ and $\to$ on A as follows:

$\odot$	1	a
1	$\{1\}$	$\{1, a\}$
a	$\{1, a\}$	{a}

$\to$	1	a
1	$\{1, a\}$	$\{a\}$
a	$\{1\}$	$\{1, a\}$

Then $(A, \odot, \to, 1)$ is a bounded (quasi) hyper hoop.

(iv) Let $A = \{1, a, b\}$. Define the hyperoperations $\odot$ and $\to$ on A as follows,

$\odot$	1	a	b
1	$\{1\}$	$\{a\}$	$\{b\}$
a	$\{a\}$	$\{a, b\}$	$\{a, b\}$
b	$\{b\}$	$\{a, b\}$	$\{b\}$

$\to$	1	a	b
1	$\{1\}$	$\{a\}$	$\{b\}$
a	$\{1\}$	$\{1, a, b\}$	$\{1, b\}$
b	$\{1\}$	$\{a\}$	$\{1, b\}$

Then $(A, \odot, \to, 1)$ is a bounded (quasi) hyper hoop.

(v) Let $A = \{1, a, b, c\}$. Define the hyperoperations $\odot$ and $\to$ on A as follows:

$\odot$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{a\}$	$\{a\}$	$\{a, b, c\}$	$\{a, c\}$
b	$\{b\}$	$\{a, b, c\}$	$\{b, c\}$	$\{b, c\}$
c	$\{c\}$	$\{a, c\}$	$\{b, c\}$	$\{c\}$

$\to$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{1\}$	$\{1, a\}$	$\{1, b, c\}$	$\{1, c\}$
b	$\{1\}$	$\{a\}$	$\{1, b, c\}$	$\{b, c\}$
c	$\{1\}$	$\{a\}$	$\{b\}$	$\{1, b, c\}$

Then $(A, \odot, \rightarrow, 1)$ is a bounded (quasi) hyper hoop.

(vi) Let $A = \{1, a, b, c\}$. Define the hyperoperations $\odot$ and $\rightarrow$ on A as follows:

$\odot$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{a\}$	$\{a\}$	$\{a, b, c\}$	$\{a, c\}$
b	$\{b\}$	$\{a, b, c\}$	$\{b, c\}$	$\{b, c\}$
c	$\{c\}$	$\{a, c\}$	$\{b, c\}$	$\{c\}$

$\rightarrow$	1	a	b	c
1	$\{1\}$	$\{a\}$	$\{b\}$	$\{c\}$
a	$\{1\}$	$\{1, a\}$	$\{b\}$	$\{1, c\}$
b	$\{1\}$	$\{a\}$	$\{1, b, c\}$	$\{b, c\}$
c	$\{1\}$	$\{a\}$	$\{b\}$	$\{1, b, c\}$

Then $(A, \odot, \rightarrow, 1)$ is an unbounded (quasi) hyper hoop. Hence, (quasi) hyper hoops may not be bounded, in general.

(vii) Let $A = [0, 1]$. Define the hyperoperations $\odot$ and $\rightarrow$ on A as follows:

$$x \odot y = \{1, x, y\} \qquad x \rightarrow y = \begin{cases} \{1, y\} & \text{, if } x \leq y, \\ \{y\} & \text{, otherwise.} \end{cases}$$

Then $(A, \odot, \rightarrow, 1)$ is an infinite (quasi) hyper hoop.

Proposition 3.4. Let A be a quasi hyper hoop. Then the following hold, for all $x, y, z \in A$ and $B, C, D \subseteq A$:
(HHA8) $B \ll C \Leftrightarrow 1 \in B \rightarrow C$,
(HHA9) $(B \odot C) \rightarrow D = B \rightarrow (C \rightarrow D)$,
(HHA10) $x \odot y \ll \{z\} \Leftrightarrow \{x\} \leq y \rightarrow z$,
(HHA11) $B \odot C \ll D \Leftrightarrow B \ll C \rightarrow D$,
(HHA12) $x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z)$,
(HHA13) $\{x\} \leq y \rightarrow z \Leftrightarrow \{y\} \leq x \rightarrow z$,
(HHA14) $\{x\} \leq (x \rightarrow y) \rightarrow y$,
(HHA15) $x \odot (x \rightarrow y) \ll \{y\}$.

Proof. Let $x, y, z \in A$ and $B, C, D \subseteq A$. Then,
(HHA8): $B \ll C \Leftrightarrow$ there exist $b \in B$ and $c \in C$ such that $b \leq c$ i.e. $1 \in b \rightarrow c \Leftrightarrow 1 \in B \rightarrow C$.

(HHA9): By (HHA4), the proof is clear.
(HHA10): $x \odot y \ll \{z\} \Leftrightarrow$ by (HHA8), $1 \in (x \odot y) \to z \Leftrightarrow$ by (HHA4), $1 \in x \to (y \to z) \Leftrightarrow$ by (HHA8), $\{x\} \leq y \to z$.
(HHA11): The proof is similar to the proof of (HHA10).
(HHA12): By (HHA4) and (HHA1),

$$x \to (y \to z) = (x \odot y) \to z = (y \odot x) \to z = y \to (x \to z).$$

(HHA13): $\{x\} \leq y \to z \Leftrightarrow$ by (HHA10), $x \odot y \ll \{z\} \Leftrightarrow$ by (HHA1), $y \odot x \ll \{z\} \Leftrightarrow$ by (HHA10), $\{y\} \leq x \to z$.
(HHA14): Since $x \to y \ll x \to y$, by (HHA1)and (HHA11), $x \odot (x \to y) \ll \{y\}$ and so by (HHA11), $\{x\} \leq (x \to y) \to y$.
(HHA15): By (HHA10) and (HHA14), the proof is clear. □

Proposition 3.5. Let A be a hyper hoop. Then the following hold, for all $x, y, z, t \in A$ and $B, C, D \subseteq A$,
(HHA16) $x \odot y \ll \{x\}, \{y\}$,
(HHA17) $\{y\} \leq x \to y$,
(HHA18) if $1 \in 1 \to x$, then $x = 1$,
(HHA19) $x \in 1 \to x$, and x is the maximum element of $1 \to x$,
(HHA20) $1 \odot 1 = \{1\}$,
(HHA21) if A is bounded, then $0 \in x \odot 0$,
(HHA22) if $B \ll C \leq D$, then $B \ll D$, and $\{x\} \leq B \leq \{y\}$ implies $x \leq y$,
(HHA23) if $B \leq C \leq D$, then $B \leq D$, and $\{x\} \leq \{y\} \leq B$ implies $\{x\} \leq B$,
(HHA24) if $B \ll \{x\} \ll C$, then $B \ll C$, and $B \ll \{x\} \leq C$ implies $B \ll C$,
(HHA25) if $x \leq y$, then $z \to x \leq z \to y$,
(HHA26) if $x \leq y$, then $y \to z \leq x \to z$,
(HHA27) $z \to y \leq (y \to x) \to (z \to x)$,
(HHA28) $z \to y \ll (x \to z) \to (x \to y)$,
(HHA29) if $x \leq y$, then $x \odot z \ll y \odot z$,
(HHA30) if $x \leq y$ and $z \leq t$, then $x \odot z \ll y \odot t$,
(HHA31) $(x \to y) \odot z \ll x \to (y \odot z)$.

Proof. (HHA16): By (HHA2)and(HHA5), $\{y\} \leq x \to x$ and so by (HHA10), $x \odot y \ll \{x\}$. Moreover by (HHA5), $\{x\} \leq y \to y$ and so by (HHA10), $x \odot y \ll \{y\}$.
(HHA17): By (HHA16) and (HHA10), the proof is clear.
(HHA18): Let $1 \in 1 \to x$. Since by (HHA5), $1 \in x \to 1$, by (HHA6), $1 = x$.
(HHA19): For all $u \in 1 \to x$ by (HHA2), $1 \in u \to (1 \to x)$. Then by (HHA12), $1 \in 1 \to (u \to x)$ and so there exists $v \in u \to x$ such that $1 \in 1 \to v$. Then by (HHA18), $v = 1$. Hence $1 \in u \to x$ and so $u \leq x$. On the other hand, by (HHA17), $\{x\} \ll 1 \to x$. Then there exists a $t \in 1 \to x$ such that $x \leq t$. Since for all $u \in 1 \to x$ we have $u \leq x$, by considering $u = t$, we have $t \leq x \leq t$ and

so by (HHA6), $x = t$. Hence $x \in 1 \to x$ and so x is the maximum element of $1 \to x$.
(HHA20): By (HHA1), 1 is the unit and so $1 \in 1 \odot 1$. Let $1 \neq a \in 1 \odot 1$. Then $1 \odot 1 \ll a$ and so by (HHA10), $1 \leq 1 \to a$. Hence $1 \in 1 \to a$ and by (HHA18), $a = 1$. Then $1 \odot 1 = \{1\}$.
(HHA21): Let A be bounded. Since by (HHA2), $1 \in 0 \to 0$, we get $\{x\} \leq 0 \to 0$, for all $x \in A$. Then by (HHA10), $x \odot 0 \ll \{0\}$. Hence since A is bounded, we get $0 \in x \odot 0$.
(HHA22): Straightforward, by (HHA7).
(HHA23): Straightforward, by (HHA7).
(HHA24): Straightforward, by (HHA7).
(HHA25): Let $x \leq y$. For all $u \in z \to x$ we have $\{u\} \leq (z \to x)$ and so by (HHA10), $u \odot z \ll \{x\}$. Since $x \leq y$, by (HHA24), $u \odot z \ll \{y\}$ and so by (HHA10), $\{u\} \leq z \to y$. Hence $z \to x \leq z \to y$.
(HHA26): Let $x \leq y$. For all $u \in y \to z$ we have $\{u\} \ll (y \to z)$ and so by (HHA13), $\{y\} \ll u \to z$. Since $x \leq y$, by (HHA23), $\{x\} \ll (u \to z)$. Hence by (HHA13), $\{u\} \ll (x \to z)$ and so $y \to z \leq x \to z$.
(HHA27): For all $u \in z \to y$ we have $\{u\} \ll z \to y$ and so by (HHA10) and (HHA14), $u \odot z \ll \{y\} \ll (y \to x) \to x$. Hence by (HHA24) and (HHA10), $\{u\} \ll z \to ((y \to x) \to x)$ and so by (HHA12), $\{u\} \ll (y \to x) \to (z \to x)$. Therefore, $z \to y \leq (y \to x) \to (z \to x)$.
(HHA28): By (HHA27), $(x \to z) \ll (z \to y) \to (x \to y)$. Hence by (HHA13), $(z \to y) \ll (x \to z) \to (x \to y)$.
(HHA29): Let $x \leq y$. Since $y \odot z \ll y \odot z$, by (HHA10), $\{y\} \leq z \to (y \odot z)$. Hence by (HHA23), $\{x\} \ll z \to (y \odot z)$ and so by (HHA10), $(x \odot z) \ll (y \odot z)$.
(HHA30): Let $x \leq y$ and $z \leq t$. Since $z \leq t$, by (HHA29), $y \odot z \ll y \odot t$. Then by (HHA10), $\{y\} \leq z \to (y \odot t)$. Hence by (HHA23), $\{x\} \leq z \to (y \odot t)$ and so by (HHA10), $x \odot z \ll y \odot t$.
(HHA31): Since $x \to y \ll x \to y$, by (HHA10), $(x \to y) \odot x \ll \{y\}$. Hence by (HHA29), $(x \to y) \odot x \odot z \ll y \odot z$. Therefore, by (HHA10), $(x \to y) \odot z \ll x \to (y \odot z)$. □

Notation: Let A be a bounded (quasi) hyper hoop. Then for any $x \in A$, we consider $x' = x \to 0$.

Proposition 3.6. Let A be a bounded quasi hyper hoop. Then $1 \in 0'$ and for any $x \in A, \{x\} \leq x''$.

Proof. By (HHA2), $1 \in 0 \to 0$. Then $1 \in 0'$. Since by (HHA12),

$$(x \to 0) \to (x \to 0) = x \to ((x \to 0) \to 0) = x \to x''$$

and by (HHA2), $1 \in (x \to 0) \to (x \to 0)$. Then $1 \in x \to x''$ and so, $\{x\} \leq x''$. □

Proposition 3.7. Let A be a bounded hyper hoop. Then the following hold, for any $x, y \in A$,
(i) $x \leq y$, implies that $y' \leq x'$,
(ii) $x' \leq x \to y$,
(iii) $x \to y \leq y' \to x'$.

Proof. (i) If $x \leq y$, then by (HHA26), $y \to 0 \leq x \to 0$. Hence $y' \leq x'$.
(ii) Since $0 \leq y$, by (HHA25), $x \to 0 \leq x \to y$. Hence $x' \leq x \to y$.
(iii) By Proposition 3.6 , $y \leq y''$. Then by (HHA25) and (HHA12),

$$x \to y \leq x \to y'' = x \to ((y \to 0) \to 0) = (y \to 0) \to (x \to 0) = y' \to x'.$$

□

Theorem 3.8. *Any (quasi) hyper hoop of order n, can be extend to a (quasi) hyper hoop of order $n+1$, for any $n \in \mathbb{N}$.*

Proof. Let A be a (quasi) hyper hoop of order $n \in \mathbb{N}$, e be an element such that $e \notin A$ and $A_1 = A \cup \{e\}$. Then we define two hyperoperations $\odot'$ and $\to'$ on A_1 by:

$$a \odot' b = \begin{cases} a \odot b & \text{if } a, b \in A, \\ \{a\} & \text{if } a \in A, b = e, \\ \{b\} & \text{if } b \in A, a = e \end{cases} \qquad a \to' b = \begin{cases} a \to b \cup \{e\} & \text{if } a, b \in A, 1 \in a \to b, \\ a \to b & \text{if } a, b \in A, 1 \notin a \to b, \\ \{e\} & \text{if } b = e, \\ \{b\} & \text{if } a = e \end{cases}$$

By some modification we can prove that $(A_1, \odot', e)$ is a commutative semihypergroup with e as the unit and satisfies the conditions (HHA2), (HHA3), (HHA4), (HHA5), (HHA6), and (HHA7). Therefore, $(A_1, \odot', \to', e)$ is a (quasi) hyper hoop and e is the unit element of it.

□

Corollary 3.9. There exist at least one (quasi) hyper hoop of order n, for any $n \in \mathbb{N}$

Proof. By Theorem 3.8 and Example 3.3 (ii), the proof is clear.

□

Note: From now on, we let A be a hyper hoop, unless otherwise is stated.

4 Some filters on hyper hoop-algebras

In this section we define the concepts of some filters on hyper hoops and we get some properties.

Definition 4.1. Let F be a non-empty subset of A. Then F is called an upset of A, if $x \in F$ and $x \leq y$ imply $y \in F$, for all $x, y \in A$,

Definition 4.2. Let F be a non-empty subset of A. Then:
(i) F is called a weak filter of A, if F is an upset and for all $x, y \in F$, $x \odot y \cap F \neq \emptyset$.
(ii) F is called a filter of A, if F is an upset and for all $x, y \in F$, $x \odot y \subseteq F$.

Note: Let F be a (weak) filter of A and $x \in F$. Since F is an upset and $x \leq 1$, we get $1 \in F$.

Example 4.3. (i) In Example 3.3(iv), $F = \{b, 1\}$ is a filter.
(ii) In Example 3.3(v), $F = \{b, 1\}$ is a weak filter.

Example 4.4. It is clear that A is a (weak) filter of A. By (HHA20), $\{1\} = 1 \odot 1$ and so $1 \odot 1 \subseteq \{1\}$. Then $\{1\}$ is a (weak)filter of A.

Proposition 4.5. Any filter of A is a weak filter.

Proof. Let F be a filter of A. Then F is an upset and $x \odot y \subseteq F$, for all $x, y \in F$. Hence $(x \odot y) \cap F \neq \emptyset$, for all $x, y \in F$. Then F is a weak filter. □

Note: Any weak filter is not a filter, in general. It can be verified by the following Example.

Example 4.6. In Example 3.3(vi), $F = \{b, 1\}$ is a weak filter, but it is not a filter.

Theorem 4.7. *Let F be a non-empty subset of A. Then F is a weak filter of A if and only if F is an upset and $F \ll x \odot y$, for all $x, y \in F$.*

Proof. ($\Rightarrow$) Straightforward.
($\Leftarrow$) Let F be an upset and $F \ll x \odot y$, for all $x, y \in F$. Hence there exist $u \in F$ and $v \in x \odot y$ such that $u \leq v$. Since F is an upset and $u \in F$, then $v \in F$ and so $x \odot y \cap F \neq \emptyset$. Hence F is a weak filter of A. □

Theorem 4.8. *Let F be a filter of A. Then for all $x, y, z \in A$,*
(i) if $x \to y \subseteq F$ and $x \in F$, then $y \in F$,
(ii) If $x \to y \subseteq F$ and $x \odot z \subseteq F$, then $y \odot z \subseteq F$,
(iii) If $x, y \in F$ and $x \ll y \to z$, then $z \in F$.

Proof. (i) Let $x \in F$ and $x \to y \subseteq F$, for $x, y \in A$. Then $x \odot (x \to y) = \bigcup_{u \in x \to y} x \odot u \subseteq F$. On the other hand, since $x \to y \ll x \to y$, by (HHA11), $(x \to y) \odot x \ll y$. Therefore, there is $v \in (x \to y) \odot x$ such that $v \leq y$. Since $v \in F$, we get $y \in F$.

(ii) By (HHA16), $x \odot z \ll x, z$. Then there exists $u \in x \odot z \subseteq F$ such that $u \leq x, z$. Since $u \in F$ and F is a filter, we get $x, z \in F$. Now, since $x \in F$ and $x \to y \subseteq F$, by (i) $y \in F$. Finally, since $y, z \in F$ and F is a filter, $y \odot z \subseteq F$.

(iii) Let $x, y \in F$. Since F is a filter, $x \odot y \subseteq F$ and since $x \ll y \to z$, by (HHA10), $x \odot y \ll z$. Then there exists $u \in x \odot y \subseteq F$ such that $u \leq z$. Since F is a filter and $u \in F$, we get $z \in F$. □

Theorem 4.9. *Let F be a non-empty subset of A. Then F is a filter of A if and only if $1 \in F$ and $F \ll x \to y$ and $x \in F$ implies $y \in F$, for any $x, y \in A$.*

Proof. ($\Rightarrow$) Let F be a filter, $F \ll x \to y$ and $x \in F$, for $x, y \in A$. Hence there exist $u \in F$ and $v \in x \to y$ such that $u \leq v$. Since $u \in F$ and F is an upset, we get $v \in F$ and since F is a filter, we get $x \odot v \subseteq F$. By $v \in x \to y$ we have $\{v\} \leq x \to y$. Then by (HHA10), $v \odot x \ll y$ and so there exists $t \in v \odot x \subseteq F$ such that $t \leq y$. Since F is an upset, we get $y \in F$.
($\Leftarrow$) Let $x \leq y$ and $x \in F$, for $x, y \in A$. Then $1 \in x \to y$ and since $1 \in F$, we get $F \ll x \to y$. Then, by hypothesis $y \in F$ and so F is an upset. Now, let $x, y \in F$ and $u \in x \odot y$. Then $x \odot y \ll u$ and so by (HHA10), $\{y\} \leq x \to u$. Since $y \in F$, we get $F \ll x \to u$ and so by hypothesis, $u \in F$. Hence $x \odot y \subseteq F$ and so F is a filter of A.

□

Definition 4.10. Let S be a non-empty subset of A. If S is a hyper hoop with respect to the hyperoperations $\odot$ and $\to$ on A, we say that S is a hyper hoop-subalgebra of A.

Theorem 4.11. *Let S be a non-empty subset of A. Then S is a hyper hoop-subalgebra of A iff $x \odot y \subseteq S$ and $x \to y \subseteq S$, for all $x, y \in S$.*

Proof. ($\Rightarrow$) The proof is clear.

($\Leftarrow$) Let $x \in S$. By (HHA2), $1 \in x \to x$ and by assumption, $x \to x \subseteq S$. Hence $1 \in S$. It is easy to show that $(S, \odot, \to, 1)$ is a hyper hoop. Then S is a hyper hoop-subalgebra of A.

□

Example 4.12. (i) In Example 3.3(iv), $F = \{b, 1\}$ is a hyper hoop-subalgebra.
(ii) In Example 3.3(iii), $F = \{1\}$ is a (weak)filter, but it is not a hyper hoop-subalgebra.
(iii) In Example 3.3(vi), $F = \{a, 1\}$ is a hyper hoop-subalgebra, but it is not a

(weak)filter. Since $a \leq c$ and $a \in F$, but $c \notin F$ and so F is not an upset.

Theorem 4.13. *If $\{F_i\}$ is a finite family of filters of A, then $\cap\{F_i\}$ is a filter of A.*

Proof. The proof is easy. □

Definition 4.14. Let D be a subset of A. The intersection of all (weak) filters of A containing D is called the (weak) filter generated by D. The filter generated by D denoted by $[D)$ and the weak filter generated by D denoted by $[D)_w$. It is trivial to verify that $[D)$ is the least filter containing D and $[D)_w$ is the least weak filter containing D.

Theorem 4.15. *If $\emptyset \neq D \subseteq A$, then*

$$[D)_w \subseteq \{x \in A | \exists\, a_1, ..., a_n \in D, \textit{ s.t. } a_1 \odot .... \odot a_n \ll \{x\}\}$$

Proof. Let

$$F = \{x \in A | \exists\, a_1, ..., a_n \in D, \text{ s.t. } a_1 \odot a_2 \odot ..... \odot a_n \ll \{x\}\}$$

It is sufficient to show that F is a weak filter containing D. Let $x \leq y$ and $x \in F$, for $x, y \in A$. Then there exist $a_1, ..., a_n \in D$, such that, $a_1 \odot ..... \odot a_n \ll \{x\}$. Since $x \leq y$, by (HHA23), $a_1 \odot ..... \odot a_n \ll \{y\}$ and so $y \in F$. Hence F is an upset. Now, let $x, y \in F$. Then there exist $a_1, ..., a_n, b_1, ..., b_m \in D$, such that, $a_1 \odot .....\odot a_n \ll \{x\}$ and $b_1 \odot .....\odot b_m \ll \{y\}$. Hence there exist $u \in a_1 \odot .....\odot a_n$ and $v \in b_1 \odot ..... \odot b_m$, such that $u \leq x$ and $v \leq y$. By (HHA30) $u \odot v \ll x \odot y$. Then $a_1 \odot ..... \odot a_n \odot b_1 \odot ..... \odot b_m \ll x \odot y$. Hence there exists $s \in x \odot y$ such that $a_1 \odot ..... \odot a_n \odot b_1 \odot ..... \odot b_m \ll \{s\}$ and so $x \odot y \cap F \neq \emptyset$. Thus F is a weak filter of A. For all $d \in D$ we have $\{d\} \ll \{d\}$, and so $d \in F$. Therefore F is a weak filter of A containing D.

□

Note: In the following Example we will show that the equation, $[D)_w = F$ is not true, in general, where

$$F = \{x \in A | \exists\, a_1, ..., a_n \in D, \text{ s.t. } a_1 \odot .... \odot a_n \ll \{x\}\}$$

Example 4.16. In Example 3.3(v), if we take $D = \{b\}$ then it follows that $F = \{1, b, c\}$, that is a weak filter containing D, but $[D)_w = \{1, b\}$. Hence in this Example $[D)_w \neq F$.

Theorem 4.17. *If $\emptyset \neq D \subseteq A$, then*

$$[D) = \{x \in A | \exists\, a_1, ..., a_n \in D, \textit{ s.t. } a_1 \odot .... \odot a_n \ll \{x\}\}$$

Proof. Let

$$F = \{x \in A | \exists\ a_1, ..., a_n \in D, \text{ s.t. } a_1 \odot a_2 \odot ..... \odot a_n \ll \{x\}\}$$

Let $x \leq y$ and $x \in F$, for $x, y \in A$. Then there exist $a_1, ..., a_n \in D$, such that,

$$a_1 \odot ..... \odot a_n \ll \{x\}$$

Since $x \leq y$, by (HHA24), $a_1 \odot ..... \odot a_n \ll \{y\}$ and so $y \in F$. Hence F is an upset. Now, let $x, y \in F$. Then there exist $a_1, ..., a_n, b_1, ..., b_m \in D$, such that, $a_1 \odot ..... \odot a_n \ll x$ and $b_1 \odot ..... \odot b_m \ll \{y\}$. For all $u \in x \odot y$, $x \odot y \ll \{u\}$. Then by(HHA10), $\{x\} \leq y \to u$. Since $a_1 \odot ..... \odot a_n \ll \{x\}$ and $\{x\} \leq y \to u$ by (HHA24), $a_1 \odot ..... \odot a_n \ll y \to u$. Since $b_1 \odot ..... \odot b_m \ll y$ by (HHA26), $y \to u \leq (b_1 \odot ..... \odot b_m) \to u$. Hence

$$a_1 \odot ..... \odot a_n \ll y \to u \leq (b_1 \odot ..... \odot b_m) \to u$$

and so by (HHA22), $a_1 \odot ..... \odot a_n \ll (b_1 \odot ..... \odot b_m) \to u$. Then by (HHA11), $(a_1 \odot ..... \odot a_n) \odot (b_1 \odot ..... \odot b_m) \ll \{u\}$ and so $u \in F$. Therefore $x \odot y \subseteq F$ and so F is a filter. Since $d \ll d$, for all $d \in D$, we have $d \in F$ and so F is a filter of A containing D. Let $D \subseteq C$ and C be a filter of A. For all $x \in F$, there exist $a_1, ..., a_n \in D$, such that

$$a_1 \odot .... \odot a_n \ll \{x\}$$

Then there exists $v \in a_1 \odot .... \odot a_n$, such that $v \leq x$. By $a_1, ..., a_n \in D \subseteq C$ and C is a filter, it follows that $a_1 \odot ..... \odot a_n \subseteq C$ and so $v \in C$. Since C is an upset we have $x \in C$ and so $F \subseteq C$. Therefore $[D) = F$.

□

Definition 4.18. Let A be bounded. Then $D \subseteq A$ is said to have *the finite intersection property* if $a_1 \odot a_2 ..... \odot a_n \cap \{0\} = \emptyset$, for all $a_1, ..., a_n \in D$.

Theorem 4.19. *Let A be bounded and $D \subseteq A$. Then $[D)$ is a proper filter of A if and only if D has the finite intersection property.*

Proof. Let $[D)$ be a proper filter of A and D has not the finite intersection property, by the contrary. Then there exist $a_1, ..., a_n \in D$ such that $0 \in a_1 \odot a_2 ..... \odot a_n$. Hence $a_1 \odot a_2 ..... \odot a_n \ll \{0\}$ and so by Theorem 4.17, $0 \in [D)$. Since $0 \leq x$, for all $x \in A$ and $[D)$ is a filter, we have $x \in [D)$ and so $[D) = A$, which is a contradiction. Hence D has the finite intersection property.
Conversely, let D has the finite intersection property and $[D)$ is not a proper filter, by the contrary. Then $[D) = A$ and so $0 \in [D)$. Then by Theorem 4.17, there exist $a_1, ..., a_n \in D$ such that $a_1 \odot a_2 ..... \odot a_n \ll \{0\}$ and so $0 \in a_1 \odot a_2 ..... \odot a_n$. Then D has not the finite intersection property, which is a contradiction. Hence $[D)$ is a proper filter.

□

Theorem 4.20. *If F is a filter of A and $a \in A$, then*

$$[F \cup \{a\}) = \{x | x \in A, \exists n \in \mathbb{N}, s.t., a^n \to x \cap F \neq \emptyset\}$$

Proof. Suppose that $x \in [F \cup \{a\})$. By Theorem 4.17, there exist $b_1, ..., b_m \in F$ and $n \in \mathbb{N}$ such that

$$b_1 \odot ..... \odot b_m \odot a^n \ll \{x\}$$

By (HHA11), we have $b_1 \odot ..... \odot b_m \ll a^n \to x$. Then there exists $u \in b_1 \odot ..... \odot b_m$ and $v \in a^n \to x$ such that $u \leq v$. Since F is a filter and $b_1, ..., b_m \in F$, we get $b_1 \odot ..... \odot b_m \subseteq F$ and so $u \in F$. Now, since F is a filter, we get $v \in F$. Hence $a^n \to x \cap F \neq \emptyset$.
Conversely, let there exists $n \in \mathbb{N}$ such that $a^n \to x \cap F \neq \emptyset$. If $s \in a^n \to x \cap F$, then $1 \in s \to (a^n \to x)$. Hence by (HHA4), $1 \in (s \odot a^n) \to x$. Therefore, $s \odot a^n \ll \{x\}$ and so by Theorem 4.17, $x \in [F \cup \{a\})$.

□

5 Conclusion

In this paper, we applied the hyper structure theory to the hoop algebras and introduced the notion of (quasi) hyper hoop algebra which is a generalization of hoop-algebra. Then we studied some properties and filter theory of this structure. Topological and categorical properties, quotient structures and relation with the other hyperstructures can be studied for the future researches.

References

[1] P. Aglianò, I. M. A. Ferreirim, F. Montagna, *Basic hoops: an algebraic study of continuous t-norms*, Studia Logica, 87(1),2007, 73 - 98.

[2] W. J. Blok, I. M. A. Ferreirim, *Hoops and their Implicational Reducts (Abstract)*, Algebraic Methods in Logic and Computer Science, Banach Center Publications, 28, 1993, 219 - 230.

[3] W.J. Blok, I.M.A. Ferreirim, *On the structure of hoops*, Algebra Universalis, 43, 2000, 233 - 257.

[4] R. A. Borzooei, M. Bakhshi, O. Zahiri, *Filter Theory On Hyper Residuated Lattices*, Quasigroups and Related Systems, 22 (2014),33-50.

[5] R. A. Borzooei, W. A. Dudek, O. Zahiri, A. Radfar, *Some Remarks on Hyper MV-algebras*, Journal of Intelligent and Fuzzy Systems, to appear DOI:10.3233/IFS-141258.

[6] R. A. Borzooei, A. Hasankhani, M. M. Zahedi,Y.B. Jun, *On hyper K-algebras*, Mathematica Japonica, 52 (2000), 113-121.

[7] B. Bosbach, *Komplementare halbgruppen. Axiomatik und arithmetik*, Fundamenta Mathematicae, Vol. 64 (1969), 257-287.

[8] J.R. Büchi, T.M. Owens, *Complemented monoids and hoops*, Unpublished manuscript.

[9] P.Corsini, V.Leoreanu, *Applications of hyperstructure theory*, Advances in Mathematics, Kluwer Academic Publishers, Dordrecht, 2003.

[10] G. Georgescu, L. Leustean, V. Preoteasa, *Pseudo-hoops*, Journal of Multiple-Valued Logic and Soft Computing, 11 (2005), 153-184.

[11] Sh. Ghorbani, A. Hasankhani, E. Eslami, *Hyper MV-algebras*, Set-Valued Math, Appl, 1, 2008, 205-222.

[12] P. Hájek, *Metamathematics of fuzzy logic*, Trends in Logic-Studia Logica Library, Dordrecht/Boston/London,(1998).

[13] F. Marty, *Sur une generalization de la notion de groupe*, 8^{th} Congress Mathematiciens Scan- dinaves, Stockholm, (1934), 45-49.

Publisher:
Accademia Piceno – Aprutina dei Velati in Teramo (A.P.A.V.)

Periodicity:
every six months

Printed in 2016 in Pescara (Italy)

Autorizzation n. 9/90 of 10/07/1990 released by Tribunale di Pescara
ISSN: 1592-7415 (printed version)

Autorizzation n. 16 of 17/12/2013 released by Tribunale di Pescara
ISSN: 2282-8214 (online version)

Accademia Piceno - Aprutina dei Velati in Teramo
ACCADEMIA DI SCIENZE, LETTERE, ARTI E TECNOLOGIA

www.eiris.it – www.apav.it

Ratio Mathematica, 30, 2016

Contents

Published by Accademia Piceno - Aprutina dei Velati in Teramo
(A.P.A.V.)