A Quadratic Programming Approach in Estimating Similarity Relations

Evangelos Triantaphyllou

Web: http://www.imse.lsu.edu/vangelis

E-mail: trianta@lsu.edu

Abstract—**This paper examines the problem of estimating how similar N objects are when they are compared with each other. The proposed approach uses as data comparative judgments of all possible pairs of the N objects. Pairwise comparisons have long been used with success in determining the relative importance of individual members in a group of objects. In the proposed approach the pairwise comparisons focus on the similarity relations instead of the relative importance of each object. A quadratic programming model is also proposed. This model processes the similarity-based pairwise comparisons and determines the similarity relations among the N objects. The proposed quadratic programming model has linear constraints; therefore it can be solved easily by transferring it into a system of linear equations.**

I. Introduction

Consider a nonempty set of possible worlds U which is introduced to represent different states of the system being modeled by a set of sentences. Then a similarity function maps pairs of possible worlds into a number in the interval [0, 1]. In other words, a similarity function

$$s : U \times U \rightarrow [0, 1]$$

assigns to each pair of entities (W, W') a unique degree of similarity between 0 and 1 [1, p. 51]. In this paper we assume that the value 1 corresponds to the **maximum similarity** between two entities, and the value 0 to the **maximum dissimilarity**. It should be stated here that this assignment of the values 1 and 0 to the maximum similarity and dissimilarity, respectively, is quite arbitrary. However, here we use 1 for the maximum similarity in order to capture the intuitive feeling that the degree of similarity between any world and itself should be as high as possible.

In general, a function $s : U \times U \rightarrow R^+$ is a **similarity** measure if it has the following properties

i) $s(W_i, W_j) = s(W_j, W_i)$ for all $W_i, W_j \epsilon U$
ii) $s(W_i, W_i) \geq s(W_j, W_i)$ for all $W_i, W_j \epsilon U$.

On the other hand, a **dissimilarity** relation $d : U \times U \rightarrow R^+$ satisfies the properties

i) $d(W_i, W_j) = d(W_j, W_i)$ for all $W_i, W_j \epsilon U$
ii) $d(W_i, W_j) \geq 0$ for all $W_i, W_j \epsilon U$
iii) $d(W_i, W_i) = 0$ for all $W_i \epsilon U$.

Examples of similarity and dissimilarity measures can be found in any book on data analysis (e.g. Anderberg [2] and Sneath and Sokal [3]). A dissimilarity function which has also the properties

iv) if $d(W_i, W_j) = 0$, then $W_i = W_j$ (definiteness)
v) $d(W_i, W_j) \leq d(W_i, W_k) + d(W_k, W_j)$ for all $W_i, W_j, W_k \epsilon U$ (triangular inequality).

is called a distance. This is the case presented in this paper. Throughout the paper the similarities among the entities are measured by a dissimilarity (not similarity) relation, particularly by the distance referred to in the literature as the city-block metric [4].

From the previous expressions it follows that the function d is a **symmetric** one. Furthermore, in this paper it is assumed that entities can be compared because their similarities can be measured by means of difference of degrees of a **common characteristic** or **feature**. We denote by A_W and $A_{W'}$ the degree that this common feature is present in the two entities W and W', respectively. The values A_W and $A_{W'}$ can also be viewed as membership values of the members of a fuzzy set defined in terms of a common feature. More on fuzzy sets can be found, for instance, in the works by Dubois and Prade [5] and Kaufmann [6]. Usually, these membership values take values in the interval [0, 1]. A value of 1 indicates that the feature is fully present, while a value of 0 indicates that the feature is completely absent.

Using the definition of the two values A_W and $A_{W'}$, it follows that the closer the two values A_W and $A_{W'}$ are to each other, the more similar the two entities W and W' should be. Therefore, a dissimilarity function can also be defined in terms of the degree to which two entities share a common feature as follows:

$$d(W, W') = |A_W - A_{W'}| \quad \text{for any entities } W, W' \in U.$$

The main problem that this paper examines is how to estimate the similarity relation among any pair of N entities or objects. It is assumed that these N entities share a common feature and that these similarity relations are viewed in terms of that common feature.

As input data for determining the previous similarity relations we use a set of pairwise comparisons. Since there are $N(N-1)/2$ possible pairs of entities, the number of pairwise comparisons is $N(N-1)/2$ as well.

For each comparison the decision maker is asked to do his best in estimating the similarity **only between two objects W and W' at a time**. The answer of the decision maker is a phrase from a finite set of linguistic phrases. Each such

Manuscript received May 29, 1992; revised August 14, 1992.
old! The author is with the Department of Industrial Engineering, Kansas State University, Manhattan, KS 66506.
IEEE Log Number 9205331.

now: Louisiana State University

linguistic phrase is preassigned to a numerical value which attempts to capture the numerical value of the difference $|A_W - A_{W'}|$ (where A_W and $A_{W'}$ are as defined above). This is done by using a **scale**. Pairwise comparisons and scales have been proposed for use in solving multicriteria decision making problems.

In the past (e.g., [7, p. 165], [8]–[10], [11, p. 201]) pairwise comparisons have been used to estimate the **relative importance** of N objects when they are examined in terms of a common feature. In that context pairwise comparisons express **ratios** of relative importance. Usually, the values of these comparisons are **not** in the interval [0, 1]. In the present treatment, however, pairwise comparisons express the **relative similarity** of a pair of objects. Furthermore, they express **differences** instead of ratios and take values in the interval [0, 1]. The issue which is raised in the light of these pairwise comparisons of relative similarity is how to combine them all and estimate the actual similarity relations among any pair of the N objects.

This paper is organized as follows. Section II presents briefly how pairwise comparisons of relative importance have been used in decision making problems. Section III introduces the concept of pairwise comparisons of relative similarity. This section also illustrates how a scale of discrete choices can be used to quantify these types of comparisons. Section IV describes the main contribution of this paper. It presents the methodology for processing pairwise comparisons of relative similarity. The comparisons form the data of a quadratic programming problem which minimizes the errors associated with these comparisons. In this way, the actual similarity relations among a number of objects can be estimated efficiently. Finally, the last section presents a summary of the contributions and some possible extensions for more research in this area.

II. Decision Making with Pairwise Comparisons

Probably the most critical step in any decision making problem is how to estimate the pertinent data. Very often these data cannot be known in terms of absolute values. For instance, what is the worth of the Kth alternatives in terms of a political impact criterion? Although information about questions such as the previous one is vital in deriving the correct decision, it is very difficult, if not impossible, to quantify it correctly. For this reason, many decision making methods attempt to determine the **relative importance**, or weight, of the alternatives in terms of each criterion involved in a given decision making problem. An approach based on pairwise comparisons proposed by Saaty [12], [7] has long attracted the interest of many researchers, both because of its easy applicability and its interesting mathematical properties.

In decision problems of this type, pairwise comparisons are used to determine the relative importance of each alternative in terms of each decision criterion. In this approach the decision maker has to express his opinion about the value of one pairwise comparison at a time. Usually, the decision maker has to choose his answer among 10–17 discrete choices. Each choice is a linguistic phrase. Some examples of such linguistic phrases are "A is more important than B," "A is of the same importance as B," and "A is a little more important than B." There are two critical problems raised when dealing with pairwise comparisons. The first is how to quantify the linguistic choices selected by the decision maker during the evaluation of the pairwise comparisons. All the methods that use pairwise comparisons eventually express the qualitative answers of a decision maker in numerical values. The second problem is how to process the numerical values of the pairwise comparisons and thus estimate the actual relative importance of a number of objects regarding a single criterion.

A. Quantifying Pairwise Comparisons of Relative Importance

Pairwise comparisons are usually quantified by using a **scale**. Such a scale is nothing but a one-to-one mapping between the set of discrete linguistic choices available to the decision maker and a discrete set of numbers that represent the importance, or weight, of these linguistic choices. There are two major approaches in developing such scales. The first is based on the linear scale proposed by Saaty as part of the analytic hierarchy process (AHP) [7, p. 54]. The second approach is due to Lootsma [13], who attempts to determine **exponential scales**. Both approaches are based on psychological theories. These two types of scales are presented briefly in the following paragraphs.

In 1846 Weber stated his law regarding a stimulus of measurable magnitude. According to this law, a change in sensation is noticed if the stimulus is increased by a constant percentage of the stimulus itself [7, p. 54]. That is, **people are unable to make choices from an infinite set**. For example, people cannot distinguish between two very close values of importance, say 3.00 and 3.04. Psychological experiments by Miller [14] have also shown that individuals **cannot simultaneously compare more than seven objects (plus or minus two)**.

Saaty uses 9 as the upper limit of his scale, 1 as the lower limit of his scale and a unit difference between successive scale values. According to the Saaty scale the values of the pairwise comparisons are determined according to the instructions depicted in Table I [7]. Therefore, based on this scale the possible values for the pairwise comparisons of relative importance should be members of the set $\{9, 8, 7, 6, 5, 4, 3, 2, 1, 1/2, 1/3, 1/4, 1/5, 1/6, 1/7, 1/8, 1/9\}$.

On the other hand, a class of exponential scales is described in [15] and [13]. The development of these scales is based on an observation in psychology made by Roberts [16] about stimulus perception (denoted as e_i). According to that observation the difference $e_{n+1} - e_n$ must be greater than or equal to the smallest perceptible difference, which is proportional to e_n. The permissible choices by the decision maker are summarized in Table II. As a result of Robert's observation the numerical equivalents of these linguistic choices need to satisfy the following relations:

$$e_{n+1} - e_n = \epsilon e_n \text{ or:}$$
$$e_{n+1} = (1+\epsilon)e_n = (1+\epsilon)^2 e_{n-1} = \cdots = (1+\epsilon)^{n+1} e_0 \text{ or:}$$
$$e_n = \exp(\gamma * n) e_0 \qquad \text{for } n = 0, 1, 2, 3, \cdots.$$

TABLE I
SCALE OF RELATIVE IMPORTANCES (ACCORDING TO SAATY [12])

Intensity of Importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Demonstrated importance	An activity is strongly favored and its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals of above nonzero	If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*.	

TABLE II
EXPONENTIAL SCALES OF RELATIVE IMPORTANCES (ACCORDING TO LOOTSMA *et al.* [6])

Intensity of Importance	Definition
e_0	Indifference between A_i and A_j
e_1	Indifference threshold toward A_i
e_2	Weak preference for A_i
e_3	Commitment threshold toward A_i
e_4	Strong preference for A_i
e_5	Dominance threshold toward A_i
e_6	Very strong preference for A_i
Reciprocals of above nonzero	If member *i* has one of the above nonzero numbers assigned to it when compared with member *j*, then *j* has the reciprocal value when compared with *i*.

The previous expressions constitute a sequence with geometric progression. The initial step is e_0, while $(1+\epsilon)$ is the progression factor. The parameter γ is **unknown**. Using this approach, pairwise comparisons of relative importance are quantified using Table II. Table III presents two exponential scales when $\gamma = 1.00$ and $\gamma = 0.50$. Apparently, different exponential scales can be generated by assigning different values to the γ parameter.

Lootsma has also observed that in many occasions humans categorize certain intervals of interest by using an approximated geometric progression. Examples include the classification of written history of Europe into historic periods and the categorization of nations on the basis of the size of their population, as well as the classification of different sound and light intensities.

TABLE III
TWO EXPONENTIAL SCALES

Normal ($\gamma = 1/2$)	Stretched ($\gamma = 1$)	Definition
$\exp(\gamma^*0) = 1.00$	$= 1.00$	e_0
$\exp(\gamma^*1) = 1.65$	$= 2.72$	e_1
$\exp(\gamma^*2) = 2.72$	$= 7.39$	e_2
$\exp(\gamma^*3) = 4.48$	$= 20.09$	e_3
$\exp(\gamma^*4) = 7.39$	$= 54.60$	e_4
$\exp(\gamma^*5) = 12.09$	$= 148.41$	e_5
$\exp(\gamma^*6) = 20.09$	$= 403.43$	e_6

In [17] a total of 78 scales are examined in terms of three evaluative criteria. These scales were derived from the original scale proposed by Saaty and the exponential scales proposed by Lootsma. The findings of that work reveal that there is no single scale that can outperform all other scales. Furthermore, the same findings strongly indicate that a few scales are very efficient under certain conditions. Therefore, for a successful application of pairwise comparisons of relative importance the appropriate scale needs to be selected and applied.

B. Processing Pairwise Comparisons of Relative Importance

Let $A_1, A_2, \cdots, A_n$ be N entities (for instance, alternatives in a decision problem). The problem is how to evaluate their relative importance in terms of a decision criterion. Saaty in [7] proposes the use of a matrix A of rational numbers taken from the finite set $\{1/9, 1/8, \cdots, 1, 2, \cdots, 8, 9\}$. It should be stated here that the problem of how to process the pairwise comparisons is **independent** of the scale used in quantifying these pairwise comparisons. Each entry of the above matrix A represents a pairwise comparison. Specifically, the entry a_{ij} denotes the number which estimates the relative importance of entity A_i when it is compared with entity A_j. Obviously, $a_{ij} = 1/a_{ji}$ and $a_{ii} = 1$. That is, matrix A is a **reciprocal** one.

Researchers have proposed a number of methods for processing these reciprocal matrices. In [7, p. 49], an eigenvalue-based approach is used. For an evaluation of this method see [18]. However, other researchers have proposed least-squares models (e.g., [19], and [11]). In certain other contributions, regression models (e.g., [20] and [10]) have been proposed as well.

III. PAIRWISE COMPARISONS OF RELATIVE SIMILARITY

In the case of using pairwise comparisons of relative importance, the comparative judgments express **ratios** of relative importance among pairs of objects. For this reason the matrices which contain these pairwise comparisons are **reciprocal**. That is, the relations: $a_{ji} = 1/a_{ij}$ and $a_{ii} = 1.00 (N \geq i, j \geq 1)$ hold. However, in the case of pairwise comparisons of relative similarity, the comparative judgments express the **difference** of the degree to which a certain feature is present in pairs of objects. Therefore, the matrices which contain these types of pairwise comparisons are **symmetric**. That is, the following expression is true:

$$a_{ij} = a_{ji} \qquad (N \geq i, j \geq 1).$$

Every pairwise comparison of relative similarity, a_{ij} ($N \geq i, j \geq 1$), represents the assessment by the decision maker of the absolute difference $|A_i - A_j|$ of the degree to which a certain feature is present in the ith and the jth object, respectively. For these types of comparisons the decision maker focuses directly on the similarity relations among pairs of objects. The interest here is not how to estimate the values A_i but rather how to estimate the previous differences $|A_i - A_j|$.

Since the decision maker is restricted to using a similarity scale with discrete choices, two problems arise. The first is how to quantify the similarity comparisons. The second problem is how to combine all $N(N-1)/2$ possible pairwise comparisons and **estimate the actual** similarity relations among the N objects. It should be stated here that regardless of the numerical values associated with the choices given by a discrete similarity scale, there is always a need to combine the similarity pairwise comparisons and estimate the actual similarity relations. These two problems are similar to the two problems described in the previous section.

Quantifying Pairwise Comparisons of Relative Similarity

As was seen in the case with pairwise comparisons of relative importance, a scale is needed to quantify pairwise comparisons of relative similarity. A decision maker cannot directly assign numerical values to his judgments. Instead, he can use linguistic phrases to assess his comparative judgments efficiently and effectively. The observations made by Weber in 1846 and Miller, which were used in developing scales for quantifying pairwise comparisons of relative importance, are also applicable when dealing with similarity-based comparisons.

Although the main goal of this paper is not the development of a similarity-based scale, such a scale is presented in Table IV. This scale uses as linguistic choices an extension of the symbolic structures highlighted by Ruspini in [1]. A close examination of the scales depicted in Tables I, II, and III with this scale reveals that the proposed scale focuses explicitly on similarity relations as opposed to the relative importance of objects.

TABLE IV
PROPOSED SIMILARITY SCALE

Intensity of Similarity	Definition
0	The two objects are identical
0.10	Almost identical
0.20	Very similar
0.30	Almost very similar
0.40	Almost similar
0.50	Similar
0.60	Almost dissimilar
0.70	Almost very dissimilar
0.80	Very dissimilar
0.90	Almost completely dissimilar
1.00	Completely dissimilar

IV. PROCESSING PAIRWISE COMPARISONS OF RELATIVE SIMILARITY

Suppose that the **real** (and hence **unknown** to the decision maker) value of the $i-j$ pairwise comparison is equal to α_{ij} (where $\alpha_{ij} \geq 0$). This value α_{ij} is equal to the absolute value of the difference $(A_i - A_j)$, where A_i and A_j are the degree to which a certain feature is present in the ith and the jth object, respectively. That is, the following is true:

$$\alpha_{ij} = \alpha_{ji} = |A_i - A_j|. \tag{1}$$

Since the decision maker (in his assessment of the value of the $i-j$ pairwise comparison) has to use a similarity scale with discrete numerical values, most likely he will select a linguistic choice (such as "very similar" or "almost similar") which is associated with a numerical value (denoted as a_{ij}), which, hopefully, will be very close to the actual value, α_{ij}. Therefore, there is an error factor, X_{ij}, introduced with each comparison. Clearly, the following relation is true:

$$X_{ij}a_{ij} = X_{ji}a_{ji} = \alpha_{ij} = \alpha_{ji} = |A_i - A_j|. \tag{2}$$

From the previous relation it follows that the error factor, X_{ij}, is equal to 1.00 if and only if the value a_{ij} (given by the decision maker) and the actual value α_{ij} are identical and $a_{ij} > 0$. Otherwise, the farther X_{ij} is from 1.00, the more different the two values a_{ij} and α_{ij} will be.

At this point, without loss of generality, suppose that the following ordering exists among the values $A_1, A_2, A_3, \cdots, A_N$:

$$A_1 \geq A_2 \geq A_3 \geq \cdots \geq A_N. \tag{3}$$

This ordering is always possible because if the $A_i (N \geq i \geq 1)$ values of the N objects are not as in (3), above, then a rearranging of their indices can achieve the ordering expressed in (3).

Consider all possible pairwise comparisons among any three objects O_i, O_j, and O_k (where $N \geq i > j > k \geq 1$). Then, by combining the previous expressions (2) and (3), the following expressions are derived:

$$X_{ik}a_{ik} = |A_i - A_k| = A_i - A_k$$
$$X_{kj}a_{kj} = |A_k - A_j| = -(A_k - A_j) = A_j - A_k$$

and

$$X_{ji}a_{ji} = |A_j - A_i| = -(A_j - A_i) = A_i - A_j.$$

By adding up the previous three expressions, the following expression is derived:

$$X_{ik}a_{ik} + X_{kj}a_{kj} + X_{ji}a_{ji} = 2(A_i - A_k) \quad \text{or}$$
$$X_{ik}a_{ik} + X_{kj}a_{kj} + X_{ji}a_{ji} = 2X_{ik}a_{ik} \quad \text{or}$$
$$X_{kj}a_{kj} + X_{ji}a_{ji} = X_{ik}a_{ik} \quad \text{for any} N \geq i > j > k \geq 1. \tag{4}$$

Given N objects there are $C_3^N = N(N-1)(N-2)/6$ possible expressions like the previous one. These expressions involve $N(N-1)/2$ variables (note that $X_{ij} = X_{ji}$ and $a_{ij} = a_{ji}$, for any $N \geq i,\ j \geq 1$). An obvious solution to system (4) is $X_{ij} = 0$, for any $N \geq i,\ j \geq 1$. However, it makes sense here to seek to determine the X_{ij} values which are **as close to 1.00 as possible**. That is, to find those X_{ij} values which minimize the following sum of squares:

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (1 - X_{ij})^2 \quad \text{or, equivalently:}$$

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X_{ij}^2 - 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X_{ij} \tag{5}$$

subject to constraints (4).

The concept of minimizing a sum of squared errors is very common in many error estimation problems in science and engineering. Expression (5) reaches an optimal value of 0 if and only if all the X_{ij} variables are equal to 1. From the previous discussion it follows that the $N(N-1)(N-2)/6$ expressions of type (4), above, consist of the body of constraints, while expression (5) is the objective function **of a quadratic problem with linear constraints**. This quadratic programming problem can be easily transformed into an equivalent **system of linear equations**. This system of linear equations has a special structure which can be used to solve it very efficiently.The previous considerations are explained further via the next example.

An Example of Processing Pairwise Comparisons of Relative Similarity

Suppose that a decision maker has to estimate the similarity relations among the four (i.e., $N = 4$) objects: O_1, O_2, O_3, and O_4. Furthermore, assume that the **real** (and hence **unknown** to the decision maker) values A_1, A_2, A_3, and A_4 are equal to 0.92, 0.74, 0.53, and 0.28, respectively. In other words, the real (and hence unknown) similarity pairwise comparisons are as follows:

$$A = \begin{bmatrix} 0 & 0.18 & 0.39 & 0.64 \\ 0.18 & 0 & 0.21 & 0.46 \\ 0.39 & 0.21 & 0 & 0.25 \\ 0.64 & 0.46 & 0.25 & 0 \end{bmatrix}$$

In this matrix the entry (1,2) is equal to 0.18 because 0.92 - 0.74 = 0.18. A similar explanation holds for the remaining entries in matrix A.

The decision maker cannot determine the exact values of the previous comparisons. However, he can use the scale depicted in Table IV to quantify his judgments. If we **assume** that the decision maker is always able to make that selection from the scale which has a numerical value **closest** to the corresponding actual value in matrix A, then the following matrix, B, presents the pairwise comparisons which we assume that the decision maker derives for this example:

$$B = \begin{bmatrix} 0 & 0.20 & 0.40 & 0.60 \\ 0.20 & 0 & 0.20 & 0.50 \\ 0.40 & 0.20 & 0 & 0.30 \\ 0.60 & 0.50 & 0.30 & 0 \end{bmatrix}$$

In this matrix the entry (1,2) is equal to 0.20 because this value is the closest value to 0.18 when the scale in Table IV is used. A similar explanation holds for the remaining entries in matrix B.

Matrix A corresponds to the notion of the **real continuous pairwise** (or RCP) matrix, while matrix B corresponds to the notion of the **closest discrete pairwise** (or CDP) matrix, described in more detail by Triantaphyllou *et al.* [21]. These two classes of matrices were originally introduced in order to study certain phenomena in decision making problems which use pairwise comparisons as the input data.

Furthermore, suppose that the decision maker has determined that the ranking of the values A_i (the values A_i are **unknown** to the decision maker) is as follows:

$$A_1 \geq A_2 \geq A_3 \geq A_4.$$

The decision maker can reach the above conclusion by asking first which of the N objects has the highest degree of the similarity feature, then which object has the second highest, etc. The decision maker **does not have to** estimate the values of $A_1, A_2, A_3, \cdots, A_N$. He simply needs to determine **only** their **ranking**.

The corresponding quadratic problem with the $4(4-1)(4-2)/6 = 4$ linear constraints has the following general structure:

$$\begin{aligned} \text{MINIMIZE} \quad & f(X_{12}, X_{13}, X_{14}, X_{23}, X_{24}, X_{34}) \\ & = X_{12}^2 + X_{13}^2 + X_{14}^2 + X_{23}^2 + X_{24}^2 + X_{34}^2 \\ & - 2X_{12} - 2X_{13} - 2X_{14} - 2X_{23} - 2X_{24} - 2X_{34} \end{aligned}$$

subject to

$$\begin{aligned} & g_1(X_{12}, X_{13}, X_{14}, \ldots, X_{34}) \\ & \qquad = X_{32}a_{32} + X_{21}a_{21} - X_{13}a_{13} = 0 \\ & g_2(X_{12}, X_{13}, X_{14}, \ldots, X_{34}) \\ & \qquad = X_{42}a_{42} + X_{21}a_{21} - X_{14}a_{14} = 0 \\ & g_3(X_{12}, X_{13}, X_{14}, \ldots, X_{34}) \\ & \qquad = X_{43}a_{43} + X_{31}a_{31} - X_{14}a_{14} = 0 \\ & g_4(X_{12}, X_{13}, X_{14}, \ldots, X_{34}) \\ & \qquad = X_{43}a_{43} + X_{32}a_{32} - X_{24}a_{24} = 0 \end{aligned}$$

and all X_{ij}'s are real positive numbers.

In general, the quadratic programming problem takes the following form:

$$\text{MINIMIZE} \quad f = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X_{ij}^2 - 2\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} X_{ij}$$

subject to:

$$X_{kj}a_{kj} + X_{ji}a_{ji} - X_{ik}a_{ik} = 0$$
$$(\text{for any} N \geq i > j > k \geq 1)$$
$$\text{and all } X_{ij}\text{'s} \geq 0. \tag{I}$$

As stated in the previous subsection, this quadratic programming problem has $C_3^N = N(N-1)(N-2)/6$ linear constraints. Furthermore, its objective function is always convex. To find the optimal solution of this problem we first need to associate a Lagrangian multiplier, λ_i, with the ith constraint and form the Lagrangian (see, for instance, [22]).

For the current example the Lagrangian is

$$\begin{aligned} L(X_{12}, X_{13}, X_{14}, \cdots, X_{34}, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \\ = f(X_{12}, X_{13}, X_{14}, \ldots, X_{34}) \\ - \sum_{i=1}^{4} \lambda_i (g_i(X_{12}, X_{13}, X_{14}, \cdots, X_{34})). \end{aligned}$$

Then, any point $(\overline{X}_{12}, \overline{X}_{13}, \overline{X}_{14}, \cdots, \overline{X}_{34}, \overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3, \overline{\lambda}_4)$ which satisfies the following relations:

$$\frac{\theta L}{\theta X_{12}} = \cdots = \frac{\theta L}{\theta X_{34}} = \frac{\theta L}{\theta \lambda_1} = \cdots = \frac{\theta L}{\theta \lambda_4} = 0 \quad (6)$$

is also an optimal solution to the previous quadratic programming problem. For the present example, expressions (6) indicate that the quadratic programming system (I) is equivalent to the system of linear equations given in matrix form as (I.a) at the bottom of the page.

In general, this system of linear equations has the following structure:

$$\begin{bmatrix} I_m, & -A^T \\ A, & O_n \end{bmatrix} \begin{bmatrix} X \\ \lambda \end{bmatrix} = \begin{bmatrix} \tilde{1} \\ \tilde{0} \end{bmatrix}, \quad \text{(II)}$$

where I_m is the identity matrix of order m, A is an $n \times m$ matrix with the coefficients of expressions (4), A^T is the transpose of A, O_n is a square matrix of order n and all entries equal to 0, X is a vector of size m with the X_{ij} variables $(N \geq j > i >\geq 1)$, λ is a vector of size n with the Lagrangian coefficients $\lambda_i (n \geq i \geq 1)$, $\tilde{1}$ is a vector of size m with all entries equal to 1, and $\tilde{0}$ is a vector of size n with all entries equal to 0 (where $n = N(N-1)(N-2)/6$ and $m = N(N-1)/2$).

System (II) leads to the following derivations:

$$\left.\begin{aligned} I_m X - A^T \lambda = \tilde{1} \\ AX + O_n \lambda = \tilde{0} \end{aligned}\right\} \quad \text{or:}$$

$$\left.\begin{aligned} X - A^T \lambda = \tilde{1} \\ AX = \tilde{0} \end{aligned}\right\} \quad \text{or:}$$

$$\left.\begin{aligned} X = \tilde{1} + A^T \lambda \\ -AA^T \lambda = A\tilde{1} \end{aligned}\right\}. \quad \text{(III)}$$

The matrix $-AA^T$ is always symmetric of order n. For the current example this matrix takes the form (III.a) shown at the bottom of the page. This matrix has rank $n - 1$. This is true because any column (or row) is linearly dependent on the remaining columns (or rows). Therefore, any of the λ_i variables can be set to an **arbitrary** value and then solve for the remaining $n - 1$ variables. Then the X_{ij} variables can be determined from the first relation in (III).

For instance, suppose that in the current example we set $\lambda_4 = 0$. When the numerical data (that is, matrix B) of this example are used, the linear system defined by the second relation in (III) yields the solution $\overline{\lambda}_1 = -0.087015$, $\overline{\lambda}_2 = -0.066934$, $\overline{\lambda}_3 = -0.147256$ (and $\overline{\lambda}_4 = 0$).

From the previous $\overline{\lambda}_i$ values and the first relation in (III), the following optimal solution of the original quadratic programming problem is derived:

$$\begin{bmatrix} \overline{x}_{12} \\ \overline{x}_{13} \\ \overline{x}_{14} \\ \overline{x}_{23} \\ \overline{x}_{24} \\ \overline{x}_{34} \end{bmatrix} = \begin{bmatrix} 0.96921 \\ 0.97590 \\ 1.12851 \\ 0.98260 \\ 0.96653 \\ 0.95582 \end{bmatrix}$$

It should be emphasized here that this optimal solution is **independent** of the $\overline{\lambda}_i$ values. To see this, suppose that λ' and λ'' (where $\lambda' \neq \lambda''$) are the solutions to equations denoted

$$\begin{bmatrix}
1.00 & 0 & 0 & 0 & 0 & 0 & -a_{12} & -a_{12} & 0 & 0 \\
0 & 1.00 & 0 & 0 & 0 & 0 & a_{13} & 0 & -a_{13} & 0 \\
0 & 0 & 1.00 & 0 & 0 & 0 & 0 & a_{14} & a_{14} & 0 \\
0 & 0 & 0 & 1.00 & 0 & 0 & -a_{23} & 0 & 0 & -a_{23} \\
0 & 0 & 0 & 0 & 1.00 & 0 & 0 & -a_{24} & 0 & a_{24} \\
0 & 0 & 0 & 0 & 0 & 1.00 & 0 & 0 & -a_{34} & -a_{34} \\
a_{12} & -a_{13} & 0 & a_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\
a_{12} & 0 & -a_{14} & 0 & a_{24} & 0 & 0 & 0 & 0 & 0 \\
0 & a_{13} & -a_{14} & 0 & 0 & a_{34} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_{23} & -a_{24} & a_{34} & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} X_{12} \\ X_{13} \\ X_{14} \\ X_{23} \\ X_{24} \\ X_{34} \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}
=
\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{(I.a)}$$

$$\begin{bmatrix}
-(a_{12}^2 + a_{13}^2 + a_{23}^2) & -a_{12}^2 & +a_{13}^2 & -a_{23}^2 \\
-a_{12}^2 & -(a_{12}^2 + a_{14}^2 + a_{24}^2) & -a_{14}^2 & +a_{24}^2 \\
+a_{13}^2 & -a_{14}^2 & -(a_{13}^2 + a_{14}^2 + a_{34}^2) & -a_{34}^2 \\
-a_{23}^2 & +a_{24} & -a_{34}^2 & -(a_{23}^2 + a_{24}^2 + a_{34}^2)
\end{bmatrix} \quad \text{(III.a)}$$

by the second relation in (III). Then the following derivations are true:

$$AA^T\lambda' = AA^T\lambda'' = -A.\tilde{1}, \text{ or: } AA^T(\lambda' - \lambda'') = 0. \quad (7)$$

From the structure of the AA^T matrix (as it was depicted earlier) it follows that expression (7) is true if and only if the difference $(\lambda' - \lambda'')$ is equal to the following vector:

$$(\lambda' - \lambda'') = \begin{bmatrix} +1.00 \\ -1.00 \\ +1.00 \\ -1.00 \end{bmatrix} t,$$

where t is any real number.

Given the previous observation on the difference $(\lambda' - \lambda'')$, the structure of the matrix A^T, and relations (III), it follows that the following relations are also true:

$$X' - X'' = A^T(\lambda' - \lambda'') = 0, \text{ or: } X' = X'',$$

where $X' = \tilde{1} + A^T\lambda'$, and $X'' = \tilde{1} + A^T\lambda''$. In other words, although system (III) may have infinitely many $\overline{\lambda}_i$ solutions, the optimal solution, $\overline{x}_{ij}$, is unique.

As stated earlier, this is also the **optimal solution** of the original quadratic programming problem. In general, if there are N objects, then the resulting system of linear equations has $C_3^N - 1 = N(N-1)(N-2)/6 - 1$ variables and same number of equations.

By using expression (2) and the previous optimal values $\overline{x}_{ij}$, the decision maker can determine $\hat{a}_{ij}$, the **estimated** similarity relations among the N objects, as follows:

$$\begin{bmatrix} \hat{a}_{12} \\ \hat{a}_{13} \\ \hat{a}_{14} \\ \hat{a}_{23} \\ \hat{a}_{24} \\ \hat{a}_{34} \end{bmatrix} = \begin{bmatrix} X_{12}a_{12} \\ X_{13}a_{13} \\ X_{14}a_{14} \\ X_{23}a_{23} \\ X_{24}a_{24} \\ X_{34}a_{34} \end{bmatrix} = \begin{bmatrix} 0.193842 \\ 0.390360 \\ 0.677106 \\ 0.196520 \\ 0.483265 \\ 0.286746 \end{bmatrix}$$

It is interesting to observe here that these estimates are closer to the actual values in matrix A than the original input data (i.e., the pairwise comparisons of relative similarity) presented in matrix B.

A critical issue which arises here is whether the proposed pairwise comparison approach always works. There is only one situation in which the pairwise comparisons may yield the wrong results. This is the case when the **triangular property** $a_{ij} \leq a_{ik} + a_{kj}$ does not hold for all possible combinations of the i, k, and j indices. From the definition of the pairwise comparisons of relative similarity, however, it was assumed that the triangular inequality should always be satisfied. This is introduced in order to capture the intuitive feeling by many decision makers that there is a close relation between the notion of similarity and that of distance (see, for example, [1, p. 54]). Therefore, if the decision maker reaches a situation in which the triangular property does not hold for all possible combinations, then some or all of his comparative judgments need to be revised until the triangular property holds true.

Another interesting issue here is to observe that the proposed approach **always reaches a feasible solution**. This is the case because from the transformation of problem (I) into its equivalent form (III) it follows that the variables $\lambda_i (n \geq i \geq 1)$ can always be calculated. Furthermore, the solution vector $\overline{x}$ **cannot** have a negative element (and hence be infeasible).

To see this consider the two relations in (III). If some element in the solution vector $\overline{x}$ is negative, then (III), together with the fact that matrix A is formed from the coefficients of constraints (4), implies the expression (that is, an element of the vector $A\tilde{1}$)

$$a_{kj} + a_{ij} - a_{ik} \text{ is negative for some } N \geq i > j > k \geq 1.$$

However, the above situation never occurs because from the triangular property $a_{ik} \leq a_{ij} + a_{jk}$ (for any $N \geq i, j, k \geq 1$) and the fact that $a_{ij} = a_{ji}$ (for any $N \geq i, j \geq 1$) it follows that the expression $a_{kj} + a_{ij} - a_{ik}$ can never be negative. Therefore, the proposed approach always reaches a feasible solution.

V. Summary and Discussion

The main objective of this paper is the development of an approach for estimating similarity relations among N objects. Pairwise comparisons have been used intensively as the means for extracting the pertinent data for many decision making problems (see, for example, [7], [23], and [24]). In this way, imprecise judgments of an expert can be processed and accurate estimates of the unknown parameters of a problem can be derived.

In the past, pairwise comparisons have been used in estimating the relative importance among the members of a set. In that context pairwise comparisons estimate the ratio of the relative importance of two objects when they are considered in terms of a property which is present in both objects.

On the other hand, the pairwise comparisons used in this paper refer to the relative similarity among pairs of objects. At each comparison between two objects the decision maker is asked to estimate the difference of the degrees that a given feature is present in these two objects. In this way, the resulting similarity relations exhibit the triangular inequality. Furthermore, these types of pairwise comparisons focus directly on the similarity relations among the objects.

In order for the proposed type of pairwise comparisons to be quantified, a discrete scale is defined in this paper. Finally, a quadratic programming formulation is proposed as the means for estimating the desired similarity relations among a number of objects. This formulation minimizes a sum of squared errors. The proposed method is very efficient because the quadratic programming problem can be reduced in the problem of solving a system of linear equations.

An interesting question at this point is why one needs all the $N(N-1)/2$ pairwise comparisons. It is easy to verify that only $N-1$ independent comparisons are enough to determine the rest of the $N(N-1)/2$ comparisons. The reason for seeking to evaluate all possible comparisons is that in this way the proposed approach can use information from many more sources (e.g. comparative judgments) in order to effectively estimate the similarity relations among N objects. $N-1$ comparisons would be enough if the decision maker were **perfectly accurate** in all his judgments. However, if the decision maker is heavily inaccurate in a few comparisons, then the negative impact is diminished if all the independent $N(N-1)/2$ comparisons are incorporated into the estimation process. This justification is very similar to the case of using pairwise comparisons in solving decision making problems.

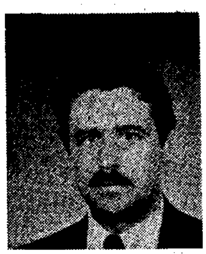
A number of extensions are possible at this point. An interesting issue is to devise an approach for estimating similarity relations among objects even if some of the pairwise comparisons are missing. Problems with incomplete data are sometimes common in many real life applications. A related issue is to examine how the order of deriving pairwise comparisons affects the final results. In other words, if the decision maker can only estimate, say, 80% of the total number of comparisons, then which comparisons should these be? The present paper provides the necessary foundations for investigating the previous two problems. The general problem of estimating similarity relations among a number of objects is an important one and more research in this area is needed.

ACKNOWLEDGMENT

The author would like to thank the referees for their thoughtful comments, which significantly improved the quality of this paper.

REFERENCES

[1] E. H. Ruspini, "On the semantics of fuzzy logic," *Int. J. Approximate Reasoning*, vol. 5, no. 1, pp. 45–88, 1991.
[2] M. R. Anderberg, "Cluster analysis for applications," *Probability and Mathematical Statistics*. New York: Academic Press, 1973, p. 19.
[3] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*. San Francisco: W. H. Freeman, 1973.
[4] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic Press, 1979.
[5] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic Press, 1980.
[6] A. Kaufman, *Introduction to the Theory of Fuzzy Sets*. New York: Academic Press, 1975.
[7] T. L. Saaty, *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.
[8] J. I. Khurgin, and V. V. Polyakov, "Fuzzy analysis of the group concordance of expert preferences, defined by saaty matrices," in *Fuzzy Sets Applications, Methodological Approaches and Results*. Berlin: Akademie-Verlag, 1986, pp. 111–115.
[9] J. I. Khurgin and V. V. Polyakov, "Fuzzy approach to the analysis of expert data," in *Fuzzy Sets Applications, Methodological Approaches and Results*. Berlin: Akademie-Verlag, 1986, pp. 116–124.
[10] F. A. Lootsma, T. C. A. Mensch, and F. A. Vos, "Multi-criteria analysis and budget reallocation in long-term research planning," *European J. Operational Research*, vol. 47, pp. 293–305, 1990.
[11] E. Triantaphyllou, P. M. Pardalos, and S. H. Mann, "A minimization approach to membership evaluation in fuzzy sets and error analysis," *Optimization Theory and Applications*, vol. 66, no. 2, pp. 275–287, 1990.
[12] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *J. Math. Psychol.*, vol. 15, no. 3, pp. 57–68, 1977.
[13] F. A. Lootsma, "Scale sensitivity and rank preservation in a multiplicative variant of the analytic hierarchy process," Rep. 91-20, Faculty of Technical Mathematics and Informatics, University of Delft, The Netherlands, 1991.
[14] G. A. Miller, "The magical number seven plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81–91, Mar. 1956.
[15] F. A. Lootsma, "Numerical scaling of human judgment in pairwise-comparison methods for fuzzy multi-criteria decision analysis," in *Mathematical Models for Decision Support* (NATO ASI Series F, Computer and System Sciences). Berlin: Springer, vol. 48, 1988, pp. 57–88.
[16] F. S. Roberts, *Measurement Theory*. Reading, MA: Addison-Wesley, 1979.
[17] E. Triantaphyllou, F. A. Lootsma, P. M. Pardalos, and S. H. Mann, "On the evaluation and application of different scales for quantifying pairwise comparisons in fuzzy sets," Multi-Criteria Decision Analysis, under review, 1993.
[18] E. Triantaphyllou, and S. H. Mann, "An evaluation of the eigenvalue approach for determining the membership values in fuzzy sets," *Fuzzy Sets and Systems*, vol. 35, no. 3, pp. 295–301, 1990.
[19] A. T. W. Chu, R. E. Kalaba, and K. Spingarn, "A comparison of two methods for determining the weights of belonging to fuzzy sets," *J. Optimization Theory and Applications*, vol. 27, pp. 531–538, 1979.
[20] V. V. Federov, V. B. Kuz'min, and A. I. Vereskov, "Membership degrees determination from Saaty matrix totalities," Institute for System Studies, Moscow, USSR. Paper appeared in *Approximate Reasoning in Decision Analysis*, M. M. Gupta, and E. Sanchez, Eds. New York: North-Holland, 1982, pp. 23–30.
[21] E. Triantaphyllou, P. M. Pardalos, and S. H. Mann, "The problem of determining membership values in fuzzy sets in real world situations," in *Operations Research and Artificial Intelligence: The Integration of Problem Solving Strategies*, D. E. Brown and C. C. White III, Eds. Boston: Kluwer Academic Publishers, 1990, pp. 197–214.
[22] W.L. Winston, *Operations Research, Applications and Algorithms*, 2nd ed. PWS-Kent Publishing, 1991.
[23] G. Shrestha, and S. Rahman, "A statistical representation of imprecision in expert judgments," *Int. J. Approximate Reasoning*, vol. 5, no. 1, pp. 1–25, 1991.
[24] E. Triantaphyllou, and S. H. Mann, "An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox," *Int. J. Decision Support Systems*, vol. 5, pp. 303–312, 1989.



Evangelos Triantaphyllou received a dual Ph.D. in industrial engineering and operations research from the Pennsylvania State University in 1990. He also has a M.S. degree in computer science from the same university.

Currently he is an Assistant Professor at Kansas State University. His research interests include interfaces of AI and OR, machine learning, decision making, and fuzzy systems. He is the author of more than 13 papers in the above areas.