sciendo
Control and Cybernetics
vol. **52** (2023) No. 3
pages: 291-334
DOI: 10.2478/candc-2023-0040

Metric correctness of pairwise comparisons in intelligent data analysis*

by

Sergey D. Dvoenko

Tula State University, Tula, Russia

Abstract: In modern data analysis and machine learning, data are often represented in the form of pairwise comparisons of the elements of the data set. The pairwise comparisons immediately correspond to the similarity or dissimilarity of objects under investigation, and such a situation regularly arises in the domains of image and signal analysis, bioinformatics, expert evaluation, etc. The practical pairwise comparison functions may be incorrect in terms of potentially using them as scalar products or distances. In contrast to other approaches, we develop in this paper a technique based on the so-called metric approach, which proposes to modify the values of empirical functions so as to get scalar products or distances. The methods for obtaining the correct matrices of pairwise comparisons and for improving their conditionality are developed here.

Keywords: similarity, dissimilarity, distance, conditionality

1. Introduction. Data representation by pairwise comparisons

Traditionally, data analysis is based on a general model in the form of a multidimensional feature space as in the case of the three-dimensional space of our real world, which corresponds to Euclidean geometry. The feature space is formed by the characteristics of objects measured in one way or another, or additionally calculated based on measurements in some natural way.

This model is really convenient because it allows us to stay within the framework of the natural ideas about the behavior of research objects by analogy with the behavior of objects in the real world. In this case, the elements of the set are represented in the multidimensional coordinate space of features (attributes) by their vectors, between which mutual distances and scalar products can be

*Submitted: October 2023; Accepted: April 2024.

calculated based, for instance, on the well-known law of cosines. These values are easily calculated based on the measurement results. From a mathematical point of view, such experimental data are considered to be immersed in the coordinate (Euclidean) space by the measurement process.

It should be noted that the modern understanding of the object of research has significantly expanded. It is believed that in the general case, this object is some kind of structure. The types of structures and their representation are really different. Unlike the traditional situation, when the object under investigation is represented by a set of characteristics, it is often inconvenient and even impossible to determine what should be measured as characteristics of such complex objects.

Therefore, in modern intelligent data analysis, machine learning, etc., experimental results are often immediately presented in the form of pairwise comparisons, representing the similarities or dissimilarities (differences) of the objects under study. In practice, it is more convenient to compare such objects directly with each other in one way or another in order to evaluate their similarity or dissimilarity with respect to each other.

From the mathematical point of view, the pairwise comparisons should be immersed in some metric space. Obviously, the configuration of a set of objects is determined by their mutual locations in a metric space.

Here, it should be said foremost that based on the idea of paired comparisons, such an approach requires, at least, solutions in three interrelated areas: first, developing the algorithms for data analysis and machine learning as analogues of the algorithms for data matrices; second, developing the methods for immersing paired comparisons in metric space (the problem of correcting the violations of metricity); and third, developing the methods for solving the specific practical problems.

The first problem was considered in some previous publications of the present author (Dvoenko, 2009; 2022). It is assumed that this problem is solved for data immersed in a metric space without violations and involves the use of special techniques for pairwise representation of data (Dvoenko, 2022). The use of respective techniques makes it possible to determine the specific conditions and conclude that algorithms for pairwise comparisons are equivalent in terms of the results obtained to the same algorithms for the vector form (based on the examples of the well-known algorithms of cluster analysis). The third kind of problems is expected to be considered later in subsequent future publications here.

In this article, the second problem is considered. The metric immersion problem is formulated here as follows. The pairwise comparison functions commonly used in practice can often be developed “ad hoc”. Here, it is suitable

to refer to them as "empirical". In modern conditions, such pairwise comparisons can even require developing the special algorithms to obtain them (for example, evaluating the similarity of signals or symbolic chains based on their alignment, evaluating the similarity of graph structures in an image and text analysis). So, sometimes it may happen that they are not the formally correct functions of similarity or distance. The use of such comparison functions can lead to violations of the configuration of the set elements in a metric space, due to inconsistencies in their relative positions. What we mean is that, in this case, the elements of the set cannot be immersed in the metric space. As a result, formally, we cannot correctly solve the first task mentioned above.

In order to immerse the elements of a set into metric space, it is necessary to eliminate violations of the relations, associated with their relative positions. So, there is a need to develop the methodology of eliminating metricity violations.

The metric immersion problem was, definitely, known and approached before, see, for instance, Bishop and Crittenden (1964). In modern conditions, the problem of metric immersion of pairwise comparisons has become practically important in data analysis, machine learning, signal and image processing (Pekalska and Duin, 2005), etc.

It is known that the condition for immersion of a set in a metric (Euclidean) space consists of the non-negative definiteness of the scalar product matrix of its elements (Young and Householder, 1938). It is known that the configuration violations occur when the triangle inequality is violated on some triplets of elements. A stronger violation consists of a violation of the law of cosines on triplets of elements. In all such cases, the matrix of scalar products of the corresponding triple of elements is non-positive definite.

In our approach, we would like to change the results of pairwise comparisons possibly minimally to recover the metric configuration. If there is a pairwise similarity matrix (for any similarity function), then the correction problem does not arise if such a matrix is non-negatively definite. As we exclude the coinciding elements of the set, the respective matrix is positive definite. This means that all its eigenvalues are positive. Then we consider that the set is immersed in some metric space or an equivalent space formed by the corresponding eigenvectors. It is then convenient to consider such a similarity function as a scalar product. It has positive values in this case, since the elements of the set are located in the same quadrant of the metric space. Note that on another limited set, the similarity matrix may turn out to be non-positive definite.

To obtain a positive definite matrix, we can proceed in any way that leads to this goal. In the present article, we demonstrate the methods to do this. They are based on the law of cosines, relating Euclidean distances and scalar products, and are called here the metric approach. If the similarity matrix is

correct in the above sense, then any transformation to dissimilarities gives a correct matrix of the corresponding empirical distances.

If there is only a matrix of pairwise dissimilarities (also any function), then any suitable transformation to the similarity function that gives a positive definite similarity matrix allows us to conclude that such a distance matrix is correct. If the transformation to the similarity function gives a non-positive definite matrix, then it is necessary to correct it and then restore the correct distances. Note that different similarity transformations can give different results in the form of positive (or non-positive) definiteness of the similarity matrix.

Hence, the problem of correction does not arise if it is possible to propose the appropriate empirical (in general) similarity function. In a general case, we can consider various similarity and distance functions, their properties, etc. But this is not the topic of this article.

Note that the idea of correcting the initial data represented by pairwise comparisons is practically the same as the well-known problem of the expert's ranking of alternatives. In such a task, it is sometimes necessary to present the evaluation results in the form of pairwise comparisons of the preference of one alternative over another. Usually, this is necessary for the consideration of multiple alternatives. In this case, it is more suitable for an expert to compare alternatives by pairs, without paying attention to the other ones. It is less suitable to immediately arrange them, taking into account the position relative to all other alternatives. At the same time, experts are not required to have their opinions transitive. As a result, the ranking method needs to be "responsible" for the correction of violations.

Ranking means the result of measurements on some ordinal scale. The limited nature of permitted operations on ordinal scales allows us to apply simple conditions for the transitivity of preferences (Luce's axiom) and obtain the so-called supertransitive matrices of pairwise comparisons of preference for alternatives (Luce, 1959; Mirkin, 1974).

When the elements of a set are immersed in a metric space, the quantitative measurements are supposed to be produced on measurement scales that are more powerful than ordinal or interval ones. This allows us to perform the usual transformations in data analysis on measurement results.

It is necessary to note, in addition, that in the well-known multidimensional scaling analysis, the interpretable space of the so-called "stimuli" is reconstructed with the additional minimization of its dimension (Torgerson, 1958; Cox and Cox, 2001). In the metric scaling problem, the elements of a set are immersed in Euclidean space. An appropriate dimension is determined, for example, based on the discrete Karhunen-Loeve expansion, preserving not less than 80% of the variance of the original data (Tou and Gonzalez, 1977). In

the non-metric scaling problem, it is required only to maintain the triangle inequality (Cox and Cox, 2001).

In our metric approach to immersing a set in a coordinate space, we do not need to restore the Euclidean coordinate space of "stimuli" explicitly for the subsequent development of data processing algorithms (Dvoenko, 2009; 2022). On the other hand, unlike in non-metric scaling, we need to comply with the law of cosines (Dvoenko and Pshenichny, 2018).

It needs to be noted that we have already encountered the situations when the feature space appears to be unnecessary for further processing. In data analysis, the SVM problem (Vapnik and Chervonenkis, 1974; Vapnik, 1998) of searching a separating hyperplane when learning to recognize two classes is considered. This is a quadratic programming task. In the dual formulation, its solution is based on a matrix of scalar products of the vectors of elements of the training set in Euclidean space. Therefore, the feature space itself is no longer needed in the dual quadratic programming problem.

The so-called "kernels" are usually used in the SVM method. This is a class of the potential functions. It can be shown that a potential function in a finite-dimensional space corresponds to an inner product in a countable-dimensional Hilbert space (Aizerman, et al., 1970) in general.

A similar situation, when the feature space itself is no longer required, arises in a number of learning algorithms, clustering, and factor analysis: the algorithm for determining the nearest points of convex hulls (Kozinets, 1973), algorithms of the Forel family (Zagoruiko, 1999), the k-means algorithm for similarity and the algorithm for the centroid main factor using the k-means algorithm (Dvoenko, 2009).

Finally, when correcting violations of the relative arrangement of the elements of a set in a metric space, we try, if possible, to minimize differences between the original and the modified values of pairwise comparisons. Elimination of violations leads to the elimination of negative eigenvalues of the corresponding scalar product matrix. But positive eigenvalues close to zero inevitably lead to ill-conditioned matrices of pairwise scalar products.

The thus appearing conditionality problem is well-known (Boyd, Ghaoui, Feron, and Balakrishnan, 1994). When working with such matrices, for example, to obtain the inverse matrix, the very advanced computational methods are used. Here, it is proposed additionally to regulate the correction degree of the original pairwise comparisons to achieve some acceptable level of conditionality.

When considering the applied problems, we pay attention both to the quality of their solution and to the volume and speed of calculations. This is because modern data analysis requires the processing of truly large-scale data sets. It should be noted that in this article the actual applied problems (cluster analysis,

machine learning, etc.) are not considered. Also, this article does not address the problem of processing large-scale data. Here, we discuss a mathematical method, illustrated with small examples.

It should be noted that the methods developed in the article are based on the calculation of the determinants of square symmetric matrices, their eigenvalues, and eigenvectors. These computational methods are well-known, and their applicability to large-scale data (advantages and disadvantages) is also known. To process large volumes of data, as is known, it is necessary to decompose the calculation process, determine parallel intervals, etc. So, it is necessary to apply some technology for organizing large-scale computing. Methods for organizing large-scale calculations are known. For example, we can refer to Kalyaev et al. (2012) for a technology for automating massive calculations based on the creation of supercomputers on FPGA elements (field-programmable gate array) to solve problems of different classes that arise in modern machine learning.

Let us consider a set $\Omega = \{\omega_1, ..., \omega_m\}$ of elements immersed into a multidimensional metric space by the process of measurements $\mathbf{x}_i = \mathbf{x}(\omega_i)$, which gives a traditional matrix of experimental data, $X(m, n)$. Here m is the number of acts of measurements or experiments, considered as objects. Then, n is the number of characteristics of the phenomenon under study, considered to be object's features. On the one hand, these data are being treated as row vectors $\mathbf{x}_i = (x_{i1}, ..., x_{in})$, $i = 1, ..., m$. Based on measurements, Euclidean distances, forming the matrix $D(m, m)$, are usually calculated between objects, where $d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$, or perhaps less traditionally, scalar products, contained in matrix $S(m, m)$, where $s_{ij} = \mathbf{x}_i \mathbf{x}_j^T$.

On the other hand, the same data can be considered as column vectors $X_j = (x_{1j}, ..., x_{mj})^T$, $j = 1, ..., n$. Based on measurements, scalar products $S(n, n)$ are traditionally calculated between features, where $s_{ij} = X_i^T X_j$. In the probabilistic approach, normalized scalar products of features are also considered as correlations $R(n, n) = (1/m)X^T X$.

Note that the elements from Ω can be both perceived as the objects and the features, pairwise comparisons of which can be performed, for example, as shown above, if their characteristics are available for measurement.

A numerical example of how to immerse measurements and pairwise comparisons in Euclidean space is discussed in Appendix 1.

Now, in the next, second section, the metric violations in pairwise comparisons are investigated. It is shown that they arise not only on triples of elements but also on subsets with more than three elements.

The third section discusses the methods for eliminating violations of pairwise similarity. The corrections of normalized scalar products and similarities are considered. A numerical example of the optimal correction is discussed in

Appendix 2. The fourth section discusses a method for eliminating violations of pairwise differences. The fifth section considers the conditionality optimization problem of the corrected pairwise comparison matrix.

2. Metric violations in pairwise comparisons

2.1. Violations on triples of elements

Let us consider the set $\Omega = \{\omega_1, ..., \omega_m\}$ of elements, represented only by a symmetric matrix of pairwise comparisons, with values $S(m, m)$. The pairwise comparisons, with values $-1 < s_{ij} < 1$, are understood as the normalized scalar products of elements from Ω, and the pairwise comparisons, having values $0 \le s_{ij} < 1$, are also understood as non-negative similarities, i.e. the normalized scalar products for elements located in the same quadrant of a metric space.

If the pairwise comparison matrix $S(m, m)$ is positive definite, then it represents the normalized scalar products of elements from $\Omega = \{\omega_1, ..., \omega_m\}$ in some metric space of dimensionality m (Young and Householder, 1938). Then, the values s_{ij} are the cosines of the angles between elements ω_i, ω_j, represented by vectors in a metric space. So, in a triplet of elements ω_i, ω_j and ω_k we get $s_{ii} = 1$ (as the similarity of the element ω_i with itself), $s_{ij} = \cos\alpha$ (as the similarity ω_i with ω_j) and $s_{ik} = \cos\beta$ (as the similarity ω_i with ω_k). All values $s_{ij},\ j = 1, ..., m$ define the positions of the elements ω_j relative to ω_i as a hypercone based on a hypercircle centered on the axis of the vector that represents the element ω_i in a metric space. The values $s_{ik},\ k = 1, ..., m$ are represented in the same way.

If an element ω_i is given, then relative to it all the relative positions of pairs of elements ω_j, ω_k are given by the corresponding hypercones. Then, the values s_{jk} determine the cosines of the angles between ω_j and ω_k. These elements are the closest to each other when they are located in the plane with ω_i on the one side of it, where $s_{jk} = \cos(\alpha - \beta) = c_1^{ijk}$. These elements are most distant with respect to each other when they are located in the plane with ω_i on the opposite sides of it, where $s_{jk} = \cos(\alpha + \beta) = c_2^{ijk}$. If similarity is considered as a non-negative function, then $\alpha + \beta \le \pi/2$.

Based on the formulas for transforming the cosine of the difference and the sum of the arguments, we obtain the boundaries of the interval

$$c_{1,2}^{ijk} = s_{ij}s_{ik} \pm \sqrt{(1 - s_{ij}^2)(1 - s_{ik}^2)}.$$

The metric violation means the violation of the interval $c_2^{ijk} \le s_{jk} \le c_1^{ijk}$, meaning that the values of s_{jk} are outside of it. This means that the matrix $S(m, m)$ becomes non-positive definite.

2.2. Violations on the sets of elements

However, it may turn out that all of the submatrices of $S(m,m)$, consisting of triplets of elements from the set Ω, are positive definite, but the matrix itself is still non-positive definite. This seems a little surprising, but helps us to understand the problem better.

Metric violations, on the one hand, can occur on triplets of elements, both when the law of cosines is violated, and when the triangle inequality, as a weak version of the law, is violated. On the other hand, metric violations can occur on configurations containing more than three elements, in general.

Theorem 1 *If a metric violation occurs, then the corresponding set of elements cannot define the set of hyperspheres in a special coordinate space.*

Proof Let the set $\Omega = \{\omega_1, ..., \omega_m\}$ be considered in some order, for example, according to a given arbitrary numbering. Let some of its elements be already considered. At the step $k = 1, ..., m$, the set of already scanned elements is represented by a principal minor $S(k,k)$, where $S(1,1) = s_{11} = 1$ is the first (top left) principal minor, $S(2,2) = \begin{pmatrix} 1 & s_{12} \\ s_{21} & 1 \end{pmatrix}$ is the second principal minor, etc., of the matrix $S(m,m)$. Let us denote the values of the principal minors $S(k,k)$ by $S_k = \det S(k,k)$, where $S_1 = 1$, $S_2 = 1 - s_{12}^2$, etc.

The purpose of the proof is to define a set of hyperspheres in a special coordinate space. It is originally proven in Dvoenko and Pshenichny (2018) and revised here. Let the ends of all vectors in the m-dimensional space be located on the hypersphere of the unit radius. We sequentially add each next element of the set in the form of a vector to this space.

Let the first vector be directed along the first axis and represented by the coordinates $(u_1, ..., u_m) = (1, 0, ..., 0)$. According to the law of cosines, the position of the second vector relative to the first one is determined by the similarity s_{12} and the distance $d_{12} = \sqrt{2 - 2s_{12}}$ between them. All possible positions of the end of the second vector relative to the end of the first vector define the m-dimensional hypersphere of the radius d_{12}, subject to conditions

$$\begin{cases} u_1^2 + ... + u_m^2 = 1 \\ (u_1 - 1)^2 + ... + u_m^2 = 2 - 2s_{12} \end{cases}.$$

By subtracting the second equation from the first, we obtain $u_1 = s_{12}$ and the first hypersphere in the form of equation

$$u_2^2 + ... + u_m^2 = 1 - s_{12}^2 = S_2/S_1.$$

Consequently, all possible positions of the end of the second vector are determined by the $(m-1)-$dimensional hypersphere with the center at the origin

and the radius $\sqrt{S_2/S_1}$, where $u_1 = s_{12}$. Thus, the second vector is represented by the coordinates $(s_{12}, \sqrt{S_2/S_1}, 0, ..., 0)$.

Let us consider the similarity of the second and third vectors, s_{23}, and the distance $d_{23} = \sqrt{2 - 2s_{23}}$ between them. The correct positions of the third vector relative to the first and second vectors are determined by the conditions

$$\begin{cases} u_1 = s_{13} \\ u_2^2 + ... + u_m^2 = 1 - s_{13}^2 \\ (u_1 - s_{12})^2 + \left(u_2 - \sqrt{1 - s_{12}^2}\right)^2 + u_3^2 + ... + u_m^2 = 2 - 2s_{23} \end{cases}.$$

By subtracting the third equation from the second and taking into account the first, we get

$$u_2 = (s_{23} - s_{12}s_{13})/\sqrt{1 - s_{12}^2} = (S_3)_3^2/\sqrt{S_1 S_2},$$

where the notation $(S_k)_j^i = \det(S(k,k))_j^i$ represents the value of the additional minor $(S(k,k))_j^i$, which is obtained from the principal minor $S(k,k)$ by crossing out the i-th row and the j-th column.

After substituting the found component into the second equation, we obtain the second hypersphere in the form of the following equation

$$u_3^2 + ... + u_m^2 = \frac{1 + 2s_{12}s_{13}s_{23} - s_{12}^2 - s_{13}^2 - s_{23}^2}{1 - s_{12}^2} = S_3/S_2.$$

Consequently, all possible positions of the end of the third vector are defined by the $(m-2)-$dimensional hypersphere with the center at the origin and the radius $\sqrt{S_3/S_2}$. Thus, the third vector is represented by the coordinates

$$(s_{13},\ \ (S_3)_3^2/\sqrt{S_1 S_2},\ \ \sqrt{S_3/S_2},\ \ 0, ..., 0).$$

Continuing this reasoning for the fourth and subsequent vectors, we define the subsequent components $u_{k-1} = (S_k)_k^{k-1}/\sqrt{S_{k-2}S_{k-1}}$ and $(m-k+1)-$dimensional hyperspheres of radii $\sqrt{S_k/S_{k-1}}$ in the form of equations $u_k^2 + ... + u_m^2 = S_k/S_{k-1}$. As a result, the coordinates of the k-th vector in such a special m-dimensional space are determined by the expressions

$$\begin{cases} u_1 = s_{1k} \\ u_t = (S_k)_{t+1}^t/\sqrt{S_{t-1}S_t}, \ \ t = 2, ..., k-1 \\ u_k = \sqrt{S_k/S_{k-1}} \\ u_{k+1} = ... = u_m = 0 \end{cases},$$

where the coordinates u_i with indices outside of the range $1 \le i \le m$ do not exist, and $S_0 = 1$.

Obviously, for a normalized matrix $S(m,m)$, the sequence of its principal minors determines the sequence of decreasing values of their determinants $S_1 = 1 > S_2 = \sqrt{1-s_{12}^2} > ... > S_m = \det S(m,m)$.

A violation occurs when the current principal minor appears to be negative. In this case, the radius squared of the corresponding hypersphere in a special coordinate space turns out to be negative, and the radius itself turns out to be a complex value. In this case, the end of the added vector is not situated on the corresponding hypersphere. Hence, the proof is complete. ■

Corollary 1 *Suppose that on the set $\Omega = \{\omega_1, ..., \omega_m\}$, represented by the matrix $S(m,m)$, the law of cosines is not violated on all triples of its elements. If there is an element represented by a vector, whose end in the special coordinate space is not located on the corresponding hypersphere with a radius defined relative to all previous vectors, then this matrix has at least one negative eigenvalue.*

Corollary 2 *The values of the principal minors of a positive definite normalized matrix of scalar products $S(m,m)$ decrease, starting from the unit value, while remaining positive. If there are metric violations, then the values of the principal minors decrease in absolute value, alternating their signs. The number of sign changes is determined by the number of negative eigenvalues, according to Sylvester's law of inertia (Horn and Johnson, 1990). Let a metric violation occur in the sequence of principal minors after a new element is added. Then the current principal minor reverses its sign relative to the previous principal minor.*

Corollary 3 *If the current principal minor becomes negative, then a metric violation occurs. To eliminate this, we need to restore the positivity of this principal minor by adjusting the pairwise comparisons of the current element of the set, which are represented by the last row and column of this minor.*

3. Correction of violations in pairwise scalar products and similarities

3.1. Individual and vector correction

A simple method of correction immediately follows from the idea of metric violations on triples of elements of the set $\Omega = \{\omega_1, ..., \omega_m\}$, which is represented by the matrix $S(m,m)$. It is easy to see that in this matrix each of its rows, $s_{ij},\ j = 1, ..., m$ determines, as shown above, the acceptable intervals $c_2^{ijk} \le s_{jk} \le c_1^{ijk}$ for all other elements $s_{jk},\ j = 1, ..., m,\ k = 1, ..., m$.

Let us define m symmetric normalized similarity matrices $S^i(m,m),\ i = 1, ..., m$ with elements s_{jk}^i, where in each of them the row $s_{ik}^i,\ k = 1, ..., m$

coincides with the same row $s_{ik},\ k = 1, ..., m$ in the original matrix $S(m, m)$. As a result, for each similarity s_{jk}, the minimal acceptable interval $\max_i c_2^{ijk} \leq s_{jk} \leq \min_i c_1^{ijk}$ is determined. If any of these intervals is violated, a new matrix, $\tilde{S}(m, m)$, is defined, where for each violated interval a new value $\tilde{s}_{jk}$ is selected within it.

The problem consists of finding an appropriate value inside the acceptable interval. It is quite obvious that the corrected values should differ from the original ones as little as possible. Hence, the heuristic rule is as follows: since the original value is outside the acceptable interval to the left or right of it, the corrected value should be as close as possible to the corresponding interval boundary.

Let us denote the values of the principal minors $S(k, k),\ k = 1, ..., m$ of the normalized matrix $S(m, m)$ as $S_k = \det S(k, k)$, where $S_1 = 1,\ S_2 = 1 - s_{12}^2$, etc. According to the corollaries of Theorem 1, the sequence of principal minors defines the sequence of decreasing values of their determinants $S_1 = 1 > S_2 = \sqrt{1 - s_{12}^2} > ... > S_m = \det S(m, m)$. Therefore, the slower the values of the determinants decrease in this sequence, the fewer opportunities there are for the next element of the set to make the determinant zero or negative, i.e. provoke a metric violation.

Then, in the case of violation, it is necessary to change the negative value of the determinant of the current principal minor to a positive value as close as possible to the value of the determinant of the previous principal minor. Let us consider the individual correction of separate elements of the current principal minor subject to these conditions. Let us perform the decomposition of the minor $S(k, k)$ over the elements of the k-th row

$$S_k = \sum_{p=1}^{k} (-1)^{k+p} s_{kp} (S_k)_p^k = \sum_{p=1}^{k-1} (-1)^{k+p} s_{kp} (S_k)_p^k + S_{k-1},$$

where $(S_k)_p^k$ is the value of the additional minor $(S(k, k))_p^k$ for the minor $S(k, k)$ after excluding the k-th row and the p-th column,

$$S_{k-1} = (-1)^{k+k} s_{kk} (S_k)_k^k = (S_k)_k^k.$$

Next, we perform the decomposition of the minors $(S(k, k))_p^k$ by the elements of the k-th column, while we preserve the indexing relative to the minor $S(k, k)$:

$$S_k = S_{k-1} + \sum_{p=1}^{k-1} (-1)^{k+p} s_{kp} \left(\sum_{q=1}^{k-1} (-1)^{(q+k)-1} s_{qk} ((S_k)_p^k)_k^q \right) =$$

$$= S_{k-1} + \sum_{p=1}^{k-1} \sum_{q=1}^{k-1} (-1)^{(2k-1)+p+q} s_{kp} s_{qk} (S_{k-1})_p^q =$$

$$S_{k-1} - \sum_{p=1}^{k-1}\sum_{q=1}^{k-1}(-1)^{p+q}s_{kp}s_{qk}(S_{k-1})^q_p.$$

Since $s_{kl} = s_{lk}$ for a symmetric similarity matrix (of scalar products), we represent the values S_k as positive functions $S_k(s_{kl}) > 0,\ \ l = 1, ..., k-1$, satisfying the following conditions for each element of the last row:

$$\begin{cases} S_k(s_{kl}) = A_l s_{kl}^2 + B_l s_{kl} + C_l \\ A_l = -(S_{k-1})^l_l \\ B_l = -2\sum\limits_{q=1, q\neq l}^{k-1}(-1)^{q+l}s_{qk}(S_{k-1})^q_l \\ C_l = S_{k-1} - \sum\limits_{p=1,\, p\neq l}^{k-1}\ \sum\limits_{q=1,\, q\neq l}^{k-1}(-1)^{p+q}s_{kp}s_{qk}(S_{k-1})^q_p \end{cases}.$$

It is easy to see that $(S_{k-1})^l_l > 0$ for all $l = 1, ..., k-1$, because $S_{k-2} > 0$. Therefore, $A_l < 0$ and symmetrical elements can be corrected in the interval $c_2^l \le s_{kl} \le c_1^l$, if $B_l^2 - 4A_lC_l > 0$, where $c^l_{1,2} = \frac{1}{2A_l}\left(-B_l \pm \sqrt{B_l^2 - 4A_lC_l}\right)$.

It is easy to verify that on its boundaries $S_k(s_{kl}) = 0$, as $S_k(c_1^l) = 0$ and $S_k(c_2^l) = 0$, and $S_k(s_{kl}) > 0$ inside the interval. Indeed, let us consider the middle of this interval $s_{kl} = -B_l/(2A_l)$. Then, inside the interval $c_2^l \le s_{kl} \le c_1^l$, we get

$$S_k(s_{kl}) = S_k\left(\frac{-B_l}{2A_l}\right) = \frac{A_l(-B_l)^2}{4A_l^2} + \frac{-B_l^2}{2A_l} + C_l = \frac{B_l^2 - 4A_lC_l}{-4A_l} > 0,$$

since the numerator and denominator of this expression are positive.

Let us define the corrected value as $s_{kl} = -B_l/(2A_l)$, since in this case the value S_k is the closest to S_{k-1} under the condition $S_k < S_{k-1}$. In general, more than one element from $s_{kl},\ l = 1, ..., k-1$ can be modified. Then, among them, it must be specified which one results in the biggest value $S_k(s_{kl})$ during correction, since a positive minor $S(k,k)$ with the biggest value of its determinant $S_k = \det S(k,k)$ will be obtained. Such a minor secures a minimal decrease in the determinant $S_{k-1} = \det S(k-1, k-1)$ of the previous positive minor $S(k-1, k-1)$.

In a situation, when there is $B_l^2 - 4A_lC_l \le 0$ for some s_{kl} or for all of them in a line $l = 1, ..., k-1$, the individual correction is not possible. In such a case, all these comparisons (composing a vector) have to be modifed. Note that in the general case, the problem of simultaneous correction of several pairwise comparisons arises: of pairs, triplets, etc. We consider this problem below.

Here it is quite obvious that individual modifications in separate pairwise comparisons are always stronger than modifications in several pairwise comparisons of the element that caused the metric violation. Therefore, the vector

correction (the vector of pairwise comparisons) generally makes less changes to the values of individual pairwise comparisons.

Let the elements of the set $\Omega = \{\omega_1, ..., \omega_m\}$ be scanned in a certain order and define the set of the already visited elements, represented at the k-th step by the current minor $S(k,k),\ k = 1, ..., m$. If its determinant is negative, $S_k < 0$, then we consider that the current element $\omega_k \in \Omega$ introduces the metric violation. Pairwise comparisons of a given element define the k-th row and the k-th column of the minor.

If we replace the vector of comparisons of this element with previous elements with the orthogonal vector of unit length, then the determinant of such a minor will coincide with the value of the determinant of the previous minor $S_k^{ort} = S_{k-1} > 0$. Indeed, in a minor $S^{ort}(k,k)$, the last row and column are zero with the unit on the main diagonal of the minor. Then, calculation of its determinant by expanding over the elements of the last row only leads to the S_{k-1}.

However, the orthogonal vector of pairwise comparisons is too far from the vector of pairwise comparisons of the element $\omega_k \in \Omega$ that caused the violation. Therefore, the orthogonal vector must be rotated in the direction of the original comparison vector until a positive value S_k is obtained.

The procedure is as follows. For a given threshold $\varepsilon > 0$ of deviation from zero, the new value of the determinant S'_k is defined as $S'_k = (P_1 + P_2)/2$, where at first $P_1 = S_k < 0$ and $P_2 = S_k^{ort} = S_{k-1} > 0$. At each step, the following is checked: if $S'_k \le 0$, then $P_1 = S'_k$, if $S'_k > 0$, then $P_2 = S'_k$. If $0 \le S'_k \le \varepsilon$, then stop.

3.2. Optimal correction

Let us consider the inverse matrix $R(k-1,k-1) = S(k-1,k-1)^{-1}$, where its elements are calculated as $r_{pq} = (-1)^{p+q}(S_{k-1})_p^q / S_{k-1}$. It is easy to see that the decomposition of the minor $S(k,k)$ over the elements of the k-th row and the k-th column, discussed above for individual correction, now takes the form

$$S_k = S_{k-1} - \sum_{p=1}^{k-1}\sum_{q=1}^{k-1} s_{kp} s_{qk} r_{pq} S_{k-1} = S_{k-1}\left(1 - \sum_{p=1}^{k-1}\sum_{q=1}^{k-1} s_{kp} s_{qk} r_{pq}\right).$$

If $S_k < 0$, then we find its new value $0 \le c \le S_{k-1}$. Let the variables $x_p = s_{kp} = s_{pk},\ p = 1, ..., k-1$ determine the new values of the elements of the corrected minor $S(k,k)$, which satisfies the condition

$$c = S_{k-1}\left(1 - \sum_{p=1}^{k-1}\sum_{q=1}^{k-1} s_{kp} s_{qk} r_{pq}\right) = S_{k-1}(1 - C),$$

where $C = 1 - c/S_{k-1}$ for a predefined value $0 \le c \le S_{k-1}$ of the determinant S_k. Since $S_k \le S_{k-1}$, it is convenient to define $c = \tau S_{k-1}$ as a fraction of the value S_{k-1}, where $0 \le \tau \le 1$.

Let us formulate the problem of optimal correction with the ability to correct a subset of pairwise comparisons of the element $\omega_k \in \Omega$ that caused the metric violation. Let us denote by $P = \{1, ... k-1\}$ the set of indices of all elements. Let us denote by $I \subseteq P$ a subset of indices of the modified elements from P. Then, the indices of unchanged elements define a subset $P \backslash I$. Let us consider the constrained optimization problem

$$\sum_{p \in I} (s_{pk} - x_p)^2 \to \min, \text{ s.t. } \sum_{p \in P} \sum_{q \in P} s_{kp} s_{qk} r_{pq} = C.$$

The solution by the Lagrange multiplier method gives a system of equations, where the number of equations is determined by the indices $p \in I$ of the modified elements

$$\begin{cases} \lambda \sum_{i \in I} x_i r_{ip} + \sum_{i \in P \setminus I} s_{ki} r_{ip} = s_{kp} - x_p, \quad p \in I \\ \sum\limits_{i \in I} \sum\limits_{j \in I} x_i x_j r_{ij} + \sum\limits_{i \in I} \sum\limits_{j \in P \setminus I} x_i s_{jk} r_{ij} + \sum\limits_{i \in P \setminus I} \sum\limits_{j \in I} s_{ki} x_j r_{ij} + \sum\limits_{i \in P \setminus I} \sum\limits_{j \in P \setminus I} s_{ki} s_{jk} r_{ij} = C. \end{cases}$$

Upon solving of this system of nonlinear equations by a suitable numerical method, we obtain the optimally adjusted row and column of the minor $S(k,k)$. Appendix 2 presents a numerical example of that optimal correction.

3.3. Locating the negative eigenvalues

Let there be metric violations in the configuration of the elements of the set $\Omega = \{\omega_1, ..., \omega_m\}$. It is easy to imagine that the elements of this set can be viewed in a different order.

It may turn out that a different sequence of principal minors requires a correction of the negative values of the determinants of other principal minors and leads to a smaller magnitude of total changes in the values of the elements of the original pairwise comparison matrix.

It is easy to conceive that among all sequences of principal minors, there can be one that provides the minimum of such total changes, both due to a smaller number of negative minors and due to smaller corrections in each case. The search for such sequences in the general case is a combinatorial problem.

On the other hand, the order, in which the elements of a set Ω are scanned does not affect the properties of the matrix of their pairwise comparisons, $S(m,m)$. In particular, it is known that simultaneous rearrangement of the rows and columns of a matrix $S(m,m)$ does not change its eigenvalues.

According to Sylvester's law of inertia (Horn and Johnson, 1990), the number of sign changes when considering the determinants of principal minors coincides with the number of negative eigenvalues of the matrix $S(m,m)$.

Consequently, the number of changes in the signs of the determinants when considering the principal minors defines the number of elements of the set that cause metric violations.

However, in practice, the number of metric violations usually appears to be bigger and even significantly bigger. In this case, violations are usually grouped so that some of them cause a trail of subsequent violations after they are corrected.

Note that Sylvester's law of inertia establishes the number of sign changes of determinants in the sequence of principal minors, but does not determine the location of such changes.

Let us find the corresponding ordering and renumber the elements of the set $\Omega = \{\omega_1, ..., \omega_m\}$ so that changes in the signs of the determinants S_k, $k = 1, ..., m$ occur generally at the end in the sequence of principal minors. Let the matrix $S(m,m)$ have v negative eigenvalues. In an ideal case, the corresponding ordering gives a sequence of minors, where for the first time the value $S_{m-v+1} < 0$ is negative, and the signs of the remaining $v-1$ determinants start to alternate. Therefore, no more than v elements of the set violate the metricity. We introduced the term "the localization of negative eigenvalues in a non-positive definite matrix" (Dvoenko and Pshenichny, 2018).

The ordering is performed by the following locally optimal procedure. Let the matrix $S(m,m)$ determine the corresponding sequence of principal minors. The determinant S_m of this matrix $S(m,m)$ is equal to the product of its eigenvalues, where for an odd number of negative eigenvalues there is $S_m < 0$, and for an even number there is $S_m > 0$.

Let us look at the determinants of the principal minors in the opposite direction, S_k, $k = m, ..., 1$. At the current step k, we calculate the determinants $(S_k)^q_q$, $q = 1, ..., k$ of all additional minors $S(k,k)^q_q$ of the current principal minor $S(k,k)$. Let us find among them a minor $S(k,k)^{q_k}_{q_k}$, whose determinant $(S_k)^{q_k}_{q_k}$ changes its sign relative to S_k and appears to be the biggest in absolute value. If the sign does not change, then we find only the determinant $(S_k)^{q_k}_{q_k}$ with the largest absolute value. Let us rearrange the row and column with the index q_k to the last place k in the minor $S(k,k)$.

The resulting permutation of the rows and columns of the matrix $S(m,m)$ determines the optimal sequence of its principal minors, where the values of their determinants decrease in absolute value most slowly, and the alternation of their signs is concentrated at the end of the sequence.

However, it usually appears that the alternation of signs at the end of such

a sequence occurs over an interval greater than v, because sometimes it is not possible to obtain a change in the sign of the determinant at every step of the procedure considered.

Let u be the additional number of determinants without changing signs for all v negative eigenvalues. Then, in the procedure discussed above, no more than $v+u$ last minors at the end of the optimal sequence are adjusted.

As a result, elements of the set that violate the metricity are concentrated at the end of the optimal sequence, reducing, often significantly, the total number of necessary corrections to the original similarity matrix.

In general, a technique for correcting the results of pairwise measurements is presented here, where metric violations are associated with specific elements of the set. This is a new understanding of the role of negative eigenvalues, in contrast to, for example, the discrete Karhunen-Loeve expansion (Tou and Gonzalez, 1977). In order to reduce the number of corrections, it is first necessary to determine the optimal sequence of the elements of the set, and then produce the correction using one of the methods described above. The measurement results are understood as scalar products. The case of correction of pairwise differences is discussed below.

3.4. Direct change of eigenvalues

Recall that the well-known Karhunen-Loeve decomposition is an example of a spectral decomposition of a square matrix in a system of eigenvectors.

Let the set $\Omega = \{\omega_1, ..., \omega_m\}$ be represented by a normalized matrix $S(m,m)$ of pairwise comparisons with elements $s_{ii} = 1$ on the main diagonal and the values of other elements $0 \le s_{ij} < 1$ or $-1 < s_{ij} < 1$.

The spectral decomposition of a non-singular matrix $S(m,m)$ has the form $S = ALA^T$, where $A(m,m) = (\mathbf{a}_1, ..., \mathbf{a}_m)$ is the orthogonal matrix $A^TA = AA^T = E$ of unit-length column eigenvectors $\mathbf{a}_i = (a_{1i}, ..., a_{mi})^T$, $|\mathbf{a}_i| = 1$, $E(m,m) = diag(1, ..., 1)$, $L(m,m) = diag(\lambda_1, ..., \lambda_m)$, where $\lambda_1 \ge ... \ge \lambda_m$.

As before, the original data matrix $X(m,n)$, representing the set $\Omega = \{\omega_1, ..., \omega_m\}$, in the corresponding feature space of dimension n, is not available. Otherwise, there is no problem with calculating the scalar products $S(m,m) = (1/n)XX^T$.

It is easy to see that as a result of the decomposition of the non-singular normalized matrix S, we obtain $L = A^TSA$, where $trL = tr\, S = m$.

Note that traditionally in data analysis, the decomposition in a system of orthogonal vectors is applied to the correlation matrix of features $R(n,n) = (1/m)X^TX$ with eigenvalues $\mu_1 \ge ... \ge \mu_n$ targeted at reducing the dimensionality of the space of eigenvectors of the matrix $R(n,n)$ to a value $n' < n$. In

particular, an incorrect matrix $R(n,n)$ has negative eigenvalues. Therefore, a projection of the objects of the original data matrix $X(m,n)$ into a new orthogonal subspace of dimension $n' < n$ is applied. The new dimension is determined only by the positive eigenvalues $\mu_1 > ... > \mu_{n'} > 0$ of the original orthogonal decomposition.

This is the well-known discrete Karhunen-Loeve expansion (Tou and Gonzalez, 1977), which is widely used in data analysis. The new correlation matrix is diagonal $R'(n',n') = diag(\mu_1, ..., \mu_{n'})$, where $tr\ R' = \sum_{i=1}^{n'} \mu_i = n' < n$. Note that to project objects into a new subspace, the matrix $X(m,n)$ is needed to obtain a new data matrix $X'(m,n')$. In this case, the total variance of the normalized data is reduced to $n' < n$.

Unlike the traditional approach, the spectral decomposition is used here for a set, whose elements are represented only by pairwise comparisons (i.e., scalar products) as $S(m,m)$. It does not matter here whether the elements of the set are objects or features (attributes).

In the case of metric violations in the configuration of elements, the spectral decomposition $L = A^T S A$ of the matrix $S(m,m)$ has negative eigenvalues. To eliminate such violations, it is proposed here not to reduce the original pairwise comparisons, but to directly replace the negative eigenvalues with suitable positive values, resulting in a new matrix $\tilde{L}(m,m)$ of the same dimension. After that, the matrix $\tilde{S}(m,m)$ of pairwise comparisons is restored by the transformation $\tilde{S} = A\tilde{L}A^T$.

Note that the new matrix, $\tilde{S}(m,m)$, appears to be non-normalized with diagonal elements bigger than the unit value. That is why $tr\tilde{L} = tr\ \tilde{S} > m$. Let us normalize its values $\hat{s}_{ij} = \tilde{s}_{ij}/\sqrt{\tilde{s}_{ii}\tilde{s}_{jj}}$ and obtain the final decomposition of the matrix $\hat{S}(m,m)$, where $\hat{S} = \hat{A}\hat{L}\hat{A}^T$, $tr\hat{L} = tr\ \hat{S} = m$.

It is quite obvious that based on this approach we can modify any eigenvalues, not just the negative ones. The question is what do we want to get as a result? Obviously, to avoid uncontrolled modifications, suitable restrictions must be formulated. For example, it is necessary to provide an appropriate level of conditionality for the modified pairwise comparison matrix (this problem is discussed below).

Note that the matrix of pairwise similarity may have other types of violations, for example, when non-diagonal elements exceed the diagonal values in absolute value, another case is the asymmetry. Direct correction of the eigenvalues may eliminate such violations, too.

4. Correction of violations in arbitrary pairwise comparisons

4.1. Offset of the origin

Unfortunately, for arbitrary matrices of pairwise comparisons (similarity or dissimilarity), we cannot form a special coordinate space according to the theorem on metric violations, demonstrated above.

Let the set $\Omega = \{\omega_1, ..., \omega_m\}$ be represented by a non-normalized matrix $S'(m,m)$ of pairwise similarities. Based on the law of cosines, we consider similarities as scalar products $s'_{ij} = (d^2_{0i} + d^2_{0j} - d^2_{ij})/2$ concerning the origin, which is represented by the element ω_0, where $d_{0i} = d(\omega_0, \omega_i)$, and the mutual distances $d_{ij} = d(\omega_i, \omega_j)$ between the elements ω_i and ω_j.

Note that such a matrix $S'(m,m)$ represents the configuration of a set in a metric space by its distances to the origin, since $s'_{ii} = d^2_{0i}$. Before correction, the matrix $S'(m,m)$ is normalized by the transformation $s_{ij} = s'_{ij}/\sqrt{s'_{ii}s'_{jj}}$. The normalized matrix $S(m,m)$ now represents only the normalized configuration, where the elements of the set are located on the unit hypersphere.

Assuming that the distances to the origin remain unchanged after the correction, the original matrix $S'(m,m)$ is restored by an inverse transformation.

Let the set $\Omega = \{\omega_1, ..., \omega_m\}$ be represented by a matrix $D(m,m)$ of pairwise differences, which are conveniently considered as distances. To understand whether there are metric violations, it is necessary to convert distances into scalar products. To do this, we must specify the origin of the coordinates, which can be located anywhere in the metric space.

Let us define it by the Torgerson's method of principal projections (Torgerson, 1958), where the origin of coordinates as the element ω_0 is represented by its distances to other elements of the set $\Omega = \{\omega_1, ..., \omega_m\}$ and is located at its center

$$d^2_{0i} = \frac{1}{m}\sum_{p=1}^{m} d^2_{ip} - \frac{1}{2m^2}\sum_{p=1}^{m}\sum_{q=1}^{m} d^2_{pq}.$$

In the case of metric violations, such a representation may be inappropriate, since some distances from the origin, located in the center, to other elements of the set may turn out to be complex. This happens if the second term exceeds the first. To prevent this from happening, we need to move the origin of coordinates, for example, to move it beyond the convex hull of the set. This position of the origin determines its correct distances to the elements of the set and the correct normalized scalar products

$$s_{ij} = \frac{1}{2d_{0i}d_{0j}} \left(d_{0i}^2 + d_{0j}^2 - d_{ij}^2\right).$$

Let us look at the first offset method. Namely, let us move the origin of coordinates beyond the convex hull of the set. The second term in Torgerson's formula above determines the dispersion σ^2 of the set relative to the origin ω_0, since

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} d_{0i}^2 = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{m}\sum_{p=1}^{m} d_{ip}^2 - \frac{1}{2m^2}\sum_{p=1}^{m}\sum_{q=1}^{m} d_{pq}^2\right) = \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{p=1}^{m} d_{ip}^2.$$

Let us consider the distances $d_{ip},\ p = 1, \dots m$ from the object ω_i to other elements of the set as components of a vector in m-dimensional space. Then, the first distance in Torgerson's formula represents the squared norm of the mean vector in such a space. We assume that the object $\omega_{0'}$ is represented by the distances $d_{0'i}^2,\ i = 1, ..., m$ to the other elements of the set. Consequently, the origin of coordinates, as the Torgerson's center, is represented as the element ω_0 by distances $d_{0i}^2 = d_{0'i}^2 - \Delta$ to other elements, where $\Delta = \sigma^2$. The value of Δ should be determined in order to move the Torgerson's origin beyond the convex hull of the set. Then the correct values determine the interval $0 \le \Delta \le \sigma^2$.

It is easy to see that for $\Delta = 0$ the origin of the coordinates is placed outside of the convex hull of the set. This corresponds to the idea that the dispersion of the set is small compared to the distances to the origin. In this case, all distances are positive and real. This property becomes stronger if the origin of the coordinates is moved even further for $\Delta < 0$. On the contrary, for $\Delta > \sigma^2$, a non-metric configuration obligatorily arises since some distances squared $d_{0i}^2 = d_{0'i}^2 - \Delta$ must become negative and distances themselves become complex.

As a result, to correct the difference matrix, it is necessary to obtain a normalized matrix of scalar products $S(m, m)$ relative to the corresponding origin, modify scalar products, and restore the correct difference matrix $D(m, m)$ based on the inverse transformation $d_{ij}^2 = s_{ii} + s_{jj} - 2s_{ij} = d_{0i}^2 + d_{0j}^2 - 2s_{ij}d_{0i}d_{0j}$.

In practice, other irregularities often arise that also require correction, even if metric correction is not required. Usually, we need to correct asymmetry or incorrect values of the non-diagonal elements of the similarity matrix that exceed the values of some of the diagonal elements.

In the first case, for symmetrization, the transformation $\tilde{d}_{ij} = (d_{ij} + d_{ji})/2$ for differences and the transformation $\tilde{s}_{ij} = (s_{ij} + s_{ji})/2$ for similarities are usually performed. The asymmetry usually results in complex eigenvalues, and its correction can result in negative eigenvalues.

If some of the non-diagonal elements exceed in module the values of the corresponding diagonal elements in the similarity matrix, then during normalization they exceed the unit in a module. It should be noted that this can happen if there was an incorrect difference matrix before. Such a matrix may appear, for example, due to the peculiarities of measuring (calculating) differences in pairs of objects or due to errors. If the similarity matrix is immediately obtained as a result of measurements (calculations), then such rough errors are usually corrected immediately. A simple correcting method is as follows.

Let us consider the matrix of differences $D(m,m)$, which we consider to be a matrix of distances. Let us transform it to a normalized matrix of scalar products $S(m,m)$ relative to the origin of coordinates constituted by an object ω_0:

$$d_{0i}^2 = \frac{1}{m}\sum_{p=1}^{m} d_{ip}^2 - \sigma^2,\ i = 1,...,m;$$

$$s_{ij} = \frac{1}{2d_{0i}d_{0j}}\left(d_{0i}^2 + d_{0j}^2 - d_{ij}^2\right);\ i,j = 1,...,m.$$

As noted, the second term in the expression for distances represents the dispersion of the set relative to the origin ω_0. In the case of violation, both terms may turn out to be inconsistent, where, for example, the second will exceed the first. Let us represent the distances in the form $d_{0i}^2 = \frac{1}{m}\sum\limits_{p=1}^{m} d_{ip}^2 - \Delta,\ i = 1,...,m.$ When $\Delta = \sigma^2$, we obtain Torgerson's formula.

When we assume $\Delta = 0$, the origin of coordinates is far outside of the set, and the dispersion of its elements can be ignored. In this case, the scalar products s_{ij} are already corrected. In general, not all of them become positive if it appears that $d_{0i}^2 + d_{0i}^2 < d_{ij}^2$. This means that in a given direction to the origin, the "width" of the set appears greater than the distance to the origin. For the entire set of elements to be in the same quadrant of the metric space, we need to move the origin of coordinates even further beyond the convex hull of the set, that is, to take $\Delta < 0$. Next, the corrected distances are restored, as before, in the form

$$d_{ij}^2 = s_{ii} + s_{jj} - 2s_{ij} = d_{0i}^2 + d_{0j}^2 - 2s_{ij}d_{0i}d_{0j}.$$

The second offset method is as follows. Here the origin of coordinates ω_0 is not moved out of the convex hull of the set, but remains inside it, to be slightly moved due to a slight change in dispersion

$$d_{0i}^2 = \frac{1}{m}\sum_{p=1}^{m} d_{ip}^2 - \alpha\sigma^2,\ 0 \le \alpha \le 1,\ i = 1,...,m.$$

It is easy to see that for $\alpha = 0$ we obtain the origin of coordinates located beyond the convex hull of the set. It is necessary to specify such α that the non-diagonal scalar products are corrected and do not exceed by their module the unit. It should be noted that such corrections do not eliminate the negative eigenvalues if there are metricity violations.

4.2. Experiment on correction of pairwise comparisons

Experimental data are presented in the form of the list of 14 investment projects (Gazprom, 2000), compiled based on materials from the open press for the year 2000. All projects are ordered according to eight criteria of positive and negative effects.

To use these data for the experiment in this work, for each criterion, the most preferred project is given the rank of 1, the next one is given the rank of 2, etc. Some projects have the same rank, so in Table 1 below the so-called standard ranks are used.

Table 1. Project ranks

	Projects\Criteria	1	2	3	4	5	6	7	8
1	South Pars	5	4	14	8	11	12	4.5	5.5
2	Blue Stream	1	1	3.5	4	12.5	14	9.5	10
3	Yamal-Europe	3	3	8.5	3	14	3	11	7
4	Pskov Power Plant	9	7	10	8	9	4.5	4.5	2
5	Kuzbass Methane	8	9.5	5.5	8	2.5	7.5	4.5	11
6	Prirazlom Field	6	6	8.5	8	9	9	12.5	13
7	Transbalkan Pipeline	10	5	12.5	13.5	6	4.5	4	4
8	Kondopoga Pipeline	11	12.5	11	1	2.5	2	4.5	8.5
9	Ecology	13	12.5	5.5	12	2.5	2	4.5	3
10	Energy degasification	12	12.5	1	8	12.5	6	14	1
11	Shtokman Field	2	2	2	2	6	7.5	12.5	14
12	Vehicle Gasification	4	8	12.5	8	9	10.5	4.5	5.5
13	Space Communications	7	9.5	7	8	6	10.5	4.5	12
14	Corporate ACS	14	12.5	3.5	13.5	2.5	2	9.5	8.5

Each ranking is represented by a relationship matrix $B_i,\ i = 1, ..., m,\ m = 8$, with the size $N \times N,\ N = 14$ of the elements

$$b_{kl}^i = \begin{cases} 1,\ a_k \succ a_l \\ 0,\ a_k \sim a_l \\ -1,\ a_k \prec a_l \end{cases},$$

where a_k is a project with the number k, the symbol "$\succ$" means "better", the symbol "$\prec$" means "worse", and the symbol "$\sim$" means "the same". Based on the relationship matrices, the distances between rankings are calculated as

$$d_{ij} = \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}|b_{kl}^{i} - b_{kl}^{j}|\,;\, i,j = 1,...,m,$$

and are represented by the matrix:

$$\begin{pmatrix} 0 & 27 & 98 & 50 & 131 & 142 & 102 & 124 \\ 27 & 0 & 105 & 65 & 128 & 123 & 101 & 113 \\ 98 & 105 & 0 & 88 & 83 & 82 & 128 & 106 \\ 50 & 65 & 88 & 0 & 109 & 102 & 96 & 112 \\ 131 & 128 & 83 & 109 & 0 & 55 & 61 & 101 \\ 142 & 123 & 82 & 102 & 55 & 0 & 86 & 76 \\ 102 & 101 & 128 & 96 & 61 & 86 & 0 & 74 \\ 124 & 113 & 106 & 112 & 101 & 76 & 74 & 0 \end{pmatrix}.$$

It is important to note the following. Rankings, as discrete structures immersed in Euclidean space, are represented by a set of isolated objects in it, since there are no other objects between them. The distances between rankings are also discrete. Therefore, it is necessary to check whether the configuration of objects given by distances $D(m,m)$ can be correctly immersed in a continuous metric space. Correct immersion allows for generation of new objects (Dvoenko and Pshenichny, 2021) that are not presented before (averages, closest, or other objects that have some extreme properties).

As a result, the correct immersion of discrete objects in a metric space makes it possible to use the well-known or new cluster analysis and machine learning algorithms for data processing (Dvoenko, 2009), and not be obliged to apply exhaustive discrete optimization methods. In general, this problem is known as the metrization of binary relations (Kemeny and Snell, 1963; Mirkin, 1974; Litvak, 1982). This, however, is a topic for a separate discussion.

Let us place the origin of coordinates at the center of the set of rankings, as shown above, and obtain a normalized, but incorrect matrix of scalar products, since it contains elements that exceed by their module the values of the diagonal elements:

$$\begin{pmatrix} 1 & 0.927 & 0.002 & 0.722 & -0.795 & \mathbf{-1.128} & -0.190 & -0.524 \\ 0.927 & 1 & -0.234 & 0.449 & -0.842 & -0.717 & -0.263 & -0.357 \\ 0.002 & -0.234 & 1 & -0.041 & 0.216 & 0.228 & \mathbf{-1.059} & -0.208 \\ 0.722 & 0.449 & -0.041 & 1 & -0.616 & -0.428 & -0.399 & -0.602 \\ -0.795 & -0.842 & 0.216 & -0.616 & 1 & 0.651 & 0.534 & -0.102 \\ \mathbf{-1.128} & -0.717 & 0.228 & -0.428 & 0.651 & 1 & 0.059 & 0.371 \\ -0.190 & -0.263 & \mathbf{-1.059} & -0.399 & 0.534 & 0.059 & 1 & 0.356 \\ -0.524 & -0.357 & -0.208 & -0.602 & -0.102 & 0.371 & 0.356 & 1 \end{pmatrix}.$$

The first correction method, based on moving the origin of coordinates out of the convex hull of the set gives the correct matrix, yields:

$$\begin{pmatrix} 1 & 0.961 & 0.473 & 0.858 & 0.055 & \mathbf{-0.116} & 0.402 & 0.176 \\ 0.961 & 1 & 0.370 & 0.741 & 0.062 & 0.130 & 0.388 & 0.290 \\ 0.473 & 0.370 & 1 & 0.518 & 0.603 & 0.611 & \mathbf{0.010} & 0.371 \\ 0.858 & 0.741 & 0.518 & 1 & 0.257 & 0.346 & 0.392 & 0.241 \\ 0.055 & 0.062 & 0.603 & 0.257 & 1 & 0.824 & 0.776 & 0.428 \\ \mathbf{-0.116} & 0.130 & 0.611 & 0.346 & 0.824 & 1 & 0.550 & 0.675 \\ 0.402 & 0.388 & \mathbf{0.010} & 0.392 & 0.776 & 0.550 & 1 & 0.680 \\ 0.176 & 0.290 & 0.371 & 0.241 & 0.428 & 0.675 & 0.680 & 1 \end{pmatrix}.$$

Here, some of the scalar products appeared to be negative, but we do not move the origin of coordinates even further, because it does not matter now.

It turned out that in this configuration there are still metric violations, since the corrected matrix has some negative eigenvalues: 4.026, 2.167, 1.058, 0.579, 0.299, 0.167, –0.097, –0.198.

Table 2 shows the original and the optimal sequences of principal minors according to the optimal permutation. As we can see, there are two alternations of the signs of the determinants according to the number of negative eigenvalues. In the case of the optimal permutation, two corrections have been produced for the original 7^{th} rank (now also before the last) and for the original 1^{st} rank (now occupying the last position).

Table 2. Determinants of principal minors

Original sequence	Determinants	Optimal sequence	Determinants
1	1	2	1
2	0.0768	5	0.9962
3	0.0526	8	0.7443
4	0.0090	3	0.3901
5	0.0023	4	0.1451
6	-0.0143	6	0.0236
7	0.0045	7	-0.0149
8	0.0051	1	0.0051

The modification eliminates negative eigenvalues, so that we obtain: 3.946, 2.050, 0.959, 0.551, 0.270, 0.149, 0.047, 0.029. After the inverse transformation, we obtain the correct distance matrix:

$$\begin{pmatrix} 0 & 48.35 & 105.54 & 61.46 & 130.36 & 133.63 & 108.52 & 124.25 \\ 48.35 & 0 & 105 & 65 & 128 & 123 & 101.01 & 113 \\ 105.54 & 105 & 0 & 88 & 83 & 82 & 121.06 & 106 \\ 61.46 & 65 & 88 & 0 & 109 & 102 & 98.75 & 112 \\ 130.36 & 128 & 83 & 109 & 0 & 55 & 77.26 & 101 \\ 133.63 & 123 & 82 & 102 & 55 & 0 & 83.05 & 76 \\ 108.52 & 104.01 & 121.06 & 98.75 & 77.26 & 83.05 & 0 & 81.24 \\ 124.25 & 113 & 106 & 112 & 101 & 76 & 81.24 & 0 \end{pmatrix}.$$

The second correcting method, based on a small offset of the origin for $\alpha = 0.8$, also yields the correct matrix:

$$\begin{pmatrix} 1 & 0.938 & 0.153 & 0.767 & -0.521 & \mathbf{-0.801} & 0.008 & -0.302 \\ 0.938 & 1 & -0.035 & 0.551 & -0.545 & -0.437 & -0.041 & -0.148 \\ 0.153 & -0.035 & 1 & 0.156 & 0.344 & 0.355 & \mathbf{-0.693} & -0.020 \\ 0.767 & 0.551 & 0.156 & 1 & -0.307 & -0.153 & -0.110 & -0.308 \\ -0.521 & -0.545 & 0.344 & -0.307 & 1 & 0.709 & 0.617 & 0.070 \\ \mathbf{-0.801} & -0.437 & 0.355 & -0.153 & 0.709 & 1 & 0.228 & 0.470 \\ 0.008 & -0.041 & \mathbf{-0.693} & -0.110 & 0.617 & 0.228 & 1 & 0.465 \\ -0.302 & -0.148 & -0.020 & -0.308 & 0.070 & 0.470 & 0.465 & 1 \end{pmatrix}.$$

Since there are metricity violations in the configuration again, this corrected matrix also has negative eigenvalues: 3.585, 1.795, 1.372, 0.940, 0.520, 0.278, -0.161, and -0.330. The correction eliminates negative eigenvalues, so that we get: 3.423, 1.650, 1.245, 0.910, 0.459, 0.239, 0.048, 0.027. After the inverse transformation, we again obtain the correct distance matrix:

$$\begin{pmatrix} 0 & 45.93 & 105.42 & 59.44 & 130.21 & 134.19 & 108.06 & 124.22 \\ 45.93 & 0 & 105 & 65 & 128 & 123 & 103.62 & 113 \\ 105.42 & 105 & 0 & 88 & 83 & 82 & 121.57 & 106 \\ 59.44 & 65 & 88 & 0 & 109 & 102 & 98.25 & 112 \\ 130.21 & 128 & 83 & 109 & 0 & 55 & 76.10 & 101 \\ 134.19 & 123 & 82 & 102 & 55 & 0 & 82.98 & 76 \\ 108.06 & 103.62 & 121.57 & 98.25 & 76.10 & 82.98 & 0 & 80.79 \\ 124.22 & 113 & 106 & 112 & 101 & 76 & 80.79 & 0 \end{pmatrix}.$$

As a result, two objects introduce metric violations, these are the first and the seventh ranking. The corrected distance matrices are almost identical. Note that the optimal correction method allows for immediate correction of the incorrect scalar product matrix.

5. Optimization of the conditionality of pairwise comparisons

5.1. The problem of conditionality

The problem of matrix conditionality is a well-known problem. Its importance is determined by the fact that the traditional sources of coefficient matrices are usually systems of equations and inequalities, which arise when solving applied problems.

The condition number $cond(A)$ of a certain matrix A shows the degree of its degeneracy. If, for example, in a system of linear equations $A\mathbf{x} = \mathbf{b}$ the matrix A is ill-conditioned, then small changes in A or $\mathbf{b}$ cause large, incommensurate changes in the solution $\mathbf{x}$. If A is well conditioned, then small changes in A or $\mathbf{b}$ produce only correspondingly small changes in the solution $\mathbf{x}$.

A well-conditioned matrix is characterized by a small condition number. For example, the identity matrix $E(m,m) = diag(1,\ldots,1)$ has the best conditionality, namely $cond(E) = 1$, where m is the dimension of the matrix.

Consider a positive definite square matrix $S(m, m)$ of pairwise comparisons. It is known that its condition number can be defined as the product of norms of this matrix and of its inverse, $cond(S) = ||S|| \cdot ||S^{-1}||$. The norm of a matrix can be defined in different ways, for example, as the maximal module of the eigenvalue, i.e. $||S|| = \max|\lambda|$. Then the norm of the inverse matrix is $||S^{-1}|| = 1/\min|\lambda|$, because the eigenvalues of the inverse matrix are the inverses of the eigenvalues of the original matrix. Consider for a positive definite matrix S its condition number as $cond(S) = \lambda_1/\lambda_m$, where $\lambda_1 = \lambda_{\max}$, $\lambda_m = \lambda_{\min}$. The appearance of a negative eigenvalue does not suit us, because in this case, we consider the conditionality to be negative.

The problem of minimizing the matrix conditionality is inevitably associated with changing the matrix $S(m, m)$ in one way or another. This imposes restrictions on the degree of modifications of the matrix $S(m, m)$ elements. It may become that the required level of conditionality while minimizing it, is in conflict with the magnitude of change in the elements of the matrix S, which is unacceptable in practice. For example, this happens in the case of strong distortions in experimental data, with which the researcher cannot agree.

On the other hand, the solution to the problem

$$cond(S) = \lambda_1/\lambda_m \to \min$$

is not directly related to the elements of the matrix $S(m, m)$. It is only known that

$\det S = \prod_{i=1}^{m} \lambda_i$ and $\mathrm{tr}S = \sum_{i=1}^{m} \lambda_i$, where $\mathrm{tr}S = m$ for the normalized matrix S.

We can only state that modification of the elements s_{ij} while tr $S = m$ remains unchanged should lead to such a redistribution of the eigenvalues that $\lambda_{\max} = \lambda_1 > 0$ decreases and $\lambda_{\min} = \lambda_m > 0$ increases, decreasing $cond(S)$. Additionally, the degree of redistribution of eigenvalues should not lead to unacceptably large changes to the original pairwise comparison matrix.

5.2. Optimization of conditionality

Methods for minimizing the condition number of matrices are known, see, for instance, Boyd, Ghaoui, Feron and Balakrishnan (1994). For example, a solution method is to represent the original matrix (generally rectangular) as a part of a decomposition that includes left and right non-singular diagonal matrices. As a result, the problem reduces to the generalized eigenvalue problem (GEVP).

Let us consider a simple heuristic approach to minimizing the condition number of a positive definite normalized scalar product matrix $S(m,m)$.

Instead of looking for a representation of the original matrix based on the optimal decomposition and, ultimately, optimizing its conditionality by minimizing the maximal eigenvalue, here we propose, on the contrary, to maximize the minimal eigenvalue. If the eigenvalues are ordered in descending order, then it is natural that the minimal eigenvalue cannot be greater than the previous one. Therefore, we propose the direct changing of the smaller eigenvalues, ordered in descending order.

It is known that the matrix $S(m,m)$ of scalar products is stratified into contributions from eigenvectors proportional to the corresponding eigenvalues. Usually, the distribution of eigenvalues in descending order is not uniform, namely – a small part of large values holds the bulk of the variance of the normalized data, which is equal to the dimensionality m of the metric space. The contributions from the small eigenvalues are very limited under these conditions. Therefore, their modification does not lead to a significant change in the original matrix of pairwise comparisons.

Here, for the matrix $S(m,m)$ of scalar products, it is proposed to increase the small eigenvalues, increasing the contributions of their corresponding eigenvectors to the variance of the normalized data. Since the sum of the eigenvalues is constant and equal to m, then the contribution of large eigenvalues decreases. Consequently, the value of λ_1 also decreases, which, in total, leads to a decrease in the conditionality $cond(S) = \lambda_1/\lambda_m$. Let us note at this point that it does not matter to us whether the elements of the set considered are objects or attributes.

Let us consider the following procedure. Define for the original matrix $S(m,m)$ an ordered sequence of its eigenvalues $\lambda_1 > ... > \lambda_m > 0$ and its conditionality $cond(S) = \lambda_1/\lambda_m$.

Let us look at the eigenvalues one by one from the end, $\lambda_i,\ i = m-1, ..., 1$. Let us define new values, $\lambda_j = \lambda_i,\ j = i+1, ..., m;\ i = m-1, ..., 1$ for the eigenvalues $\lambda_j,\ j = i+1, ..., m$.

As before, $L(m,m) = diag(\lambda_1, ..., \lambda_m)$ is the matrix of eigenvalues, while $A(m,m) = (\mathbf{a}_1, ..., \mathbf{a}_m)$ is the matrix of normalized eigenvectors $\mathbf{a}_i = (a_{i1}, ..., a_{im})^T$, $\mathbf{a}_i^T\mathbf{a}_i = 1$, $\mathbf{a}_i^T\mathbf{a}_j = 0,\ i \neq i$.

At the current step i, after determining new corresponding eigenvalues, we obtain the matrix $\tilde{L}(m,m) = diag(\tilde{\lambda}_1, ..., \tilde{\lambda}_m)$ and restore the matrix $\tilde{S} = A\tilde{L}A^T$. After such a restoration tr $\tilde{S}$ = tr $\tilde{L} > m$, where the diagonal elements are $\tilde{s}_{ii} > 1$. After the transformation $\hat{s}_{ij} = \tilde{s}_{ij}/\sqrt{\tilde{s}_{ii}\tilde{s}_{jj}}$ we obtain a normalized matrix $\hat{S}(m,m)$, where tr $\hat{S} = m$.

Let us perform the decomposition again and obtain $\hat{L} = \hat{A}^T\hat{S}\hat{A}$, where tr $\hat{S}$ = tr $\hat{L} = m$. The method of searching for such a decomposition gives again an ordered sequence of different eigenvalues $\hat{\lambda}_1 > ... > \hat{\lambda}_m > 0$, where $\hat{L}(m,m) = diag(\hat{\lambda}_1, ..., \hat{\lambda}_m)$. Let us define a new conditionality $cond(\hat{S}) = \hat{\lambda}_1/\hat{\lambda}_m < cond(S)$. Let us repeat the process at the next step i again for the original matrix $S(m,m)$.

It is evident that at the last step, $i = 1$, all new eigenvalues appear to be the same $\tilde{\lambda}_1 = \tilde{\lambda}_2 = ... = \tilde{\lambda}_m$. After restoration $\tilde{S}(m,m)$ and normalization $\hat{S}(m,m)$, we obtain the identity matrix $\hat{S}(m,m) = E(m,m)$, where $cond(\hat{S}) = 1$. In this case, the matrix $\hat{S}(m,m)$ has a minimal conditionality, but it is, definitely, radically different from the original matrix $S(m,m)$.

Therefore, it is necessary to decide at what step in determining new eigenvalues we need to stop. It is proposed to control the degree of change in smaller eigenvalues using a heuristic procedure based on checking the statistical significance of the values of the elements of the normalized matrix. To do this, we consider its elements as correlation coefficients when testing the corresponding statistical hypotheses about their significance.

5.3. Stop condition of the optimization procedure

Let us consider a positive definite normalized similarity matrix $S(m,m)$. We consider the mutual normalized scalar products of the elements of the set as being the correlation coefficients. Let us consider the well-known Student's t-test for checking the statistical significance of a correlation. In the original matrix S, some relations s_{ij} appear to be significant. As it is shown above, the procedure for minimizing the condition number of the matrix S leads, ultimately, to the identity matrix. But in this case, all scalar products between the elements of the set appear not to be significant.

Thus, the condition number optimization procedure reduces the number of

significant relations. Under these conditions, the stopping criterion is a situation, where the number of significant relations very clearly and perhaps even sharply decreases.

According to the Student's criterion, it is necessary to check the significance of $m(m\text{-}1)/2$ elements s_{ij}, $i < j$, in the matrix $S(m,m)$. Let $x_q = s_{ij}$, $i = 1,\ldots,m\text{-}1$, $j = i+1,\ldots, m$, that is, $q = 1,\ldots, m(m-1)/2$. If the value of the Student's t-test is significant, then

$$|T| \;=\; |x_q\sqrt{r-2}/\sqrt{1-x_q^2}| > t_{\alpha,r-2},$$

where r is the sample size, $r-2$ is the number of degrees of freedom, and $\alpha = 0.01$ is the level of significance of the one-sided critical point.

Let

$$T^2 = x_q^2(r-2)/(1-x_q^2).$$

It is true that $x_q = x_{crit}$ at the critical point. Then we get

$$t_{crit}^2 = x_{crit}^2(r-2)/(1-x_{crit}^2).$$

Let us apply the following transformation:

$$x_{crit}^2(r-2) = t_{crit}^2(1-x_{crit}^2) = t_{crit}^2 - t_{crit}^2 x_{crit}^2.$$

Next, we get

$x_{crit}^2(r-2+t_{crit}^2) = t_{crit}^2$, and $x_{crit} = \sqrt{t_{crit}^2/(t_{crit}^2+r-2)}$.

If $|x_q| > x_{crit}$, then the relationship x_q is significant.

However, the sample size r, which is used in the Student's t-test (it can be interpreted as the number of trials in which the pairs of corresponding observed values are compared), is unknown to us in this case. So, we do not know how many times the pairs of certain quantities are compared for purposes of calculating the relationship value s_{ij}.

When analyzing the critical values of the Student's criterion, we assume that the unknown sample volume (unknown number of tests) can be defined in the following manner: the standard table of critical points of the Student's t-distribution usually contains tabulated values for up to 120 degrees of freedom. Therefore, we believe that a sample size of $r \geq 122$ could be statistically sufficient. On the other hand, it is believed that for a set of at least 30 independent random variables the properties of a normal probability distribution already appear. Since the Student's t-distribution is based on the normal distribution, we believe that the sample size $r \geq 32$ can be considered statistically small. Then, the sample size $r \geq 62$ can be considered as average. For a one-sided critical

Table 3. Critical values

Sample (number of trials)	Size $r \geq$	t_{crit}	x_{crit}
Sufficient	122	2.33	0.2080
Average	62	2.39	0.2948
Small	32	2.46	0.4097

region at the significance level $\alpha = 0.01$, the critical values for different sample sizes are shown in Table 3.

Then, for each value of r, we establish the diagrams of changes in the number of significant relations at all steps $i = m$-1,. . . ,1 of the conditionality optimization procedure and determine the threshold, after which a sharp drop in the number of significant relations occurs.

6. Experiments on conditionality optimization

In the first example, we consider the data from the psychologist V. D. Nebylitsyn (Nebylitsyn, 1990) on the experiments, carried out in the 1960s to study the influence of rhythmic light flashes on human brain. His experimental data are represented by the correlation matrix $S(11,11)$ of statistical relationships between the energy of vibrations at 11 different frequencies, representing, respectively, theta-rhythms (1 – 3), low-frequency (4 – 5), and high-frequency alpha-rhythms (6 – 7), as well as beta-rhythms (8 – 11) of a brain on electroencephalograms:

$$\begin{pmatrix}
1 & 0.562 & 0.568 & 0.152 & 0.347 & 0.250 & 0.264 & -0.020 & -0.212 & -0.086 & -0.076 \\
0.562 & 1 & 0.784 & 0.057 & 0.196 & 0.218 & 0.009 & -0.017 & -0.002 & 0.163 & 0.284 \\
0.568 & 0.784 & 1 & 0.288 & 0.475 & 0.264 & 0.066 & 0.144 & 0.114 & 0.228 & 0.151 \\
0.152 & 0.057 & 0.228 & 1 & 0.686 & 0.293 & 0.034 & 0.048 & -0.069 & -0.064 & 0.175 \\
0.347 & 0.196 & 0.475 & 0.686 & 1 & 0.429 & 0.070 & 0.152 & 0.036 & 0.028 & 0.216 \\
0.250 & 0.218 & 0.264 & 0.293 & 0.429 & 1 & 0.788 & 0.197 & 0.154 & 0.109 & 0.035 \\
0.264 & 0.009 & 0.066 & 0.034 & 0.070 & 0.788 & 1 & 0.109 & 0.054 & -0.002 & -0.018 \\
-0.020 & -0.017 & 0.144 & 0.048 & 0.152 & 0.197 & 0.109 & 1 & 0.807 & 0.830 & 0.699 \\
-0.212 & -0.002 & 0.114 & -0.069 & 0.036 & 0.154 & 0.054 & 0.807 & 1 & 0.904 & 0.728 \\
-0.086 & 0.163 & 0.228 & -0.064 & 0.028 & 0.109 & -0.002 & 0.830 & 0.904 & 1 & 0.768 \\
-0.076 & 0.284 & 0.151 & 0.175 & 0.216 & 0.035 & -0.018 & 0.699 & 0.728 & 0.768 & 1
\end{pmatrix}.$$

This matrix has ten positive eigenvalues, namely 3.636340, 2.827085, 1.611613, 1.358204, 0.515165, 0.412792, 0.278171, 0.164165, 0.151054, 0.069977, and the last is –0.024566.

Replacing a negative eigenvalue with a positive value very close to zero gives, after normalization, a positive definite matrix with eigenvalues equal 3.629217, 2.821157, 1.605834, 1.355937, 0.514139, 0.411444, 0.277702, 0.163727, 0.150881, 0.069932, 0.000030, where the condition number is 120973.9. It is easy to see that this matrix differs slightly from the previous one:

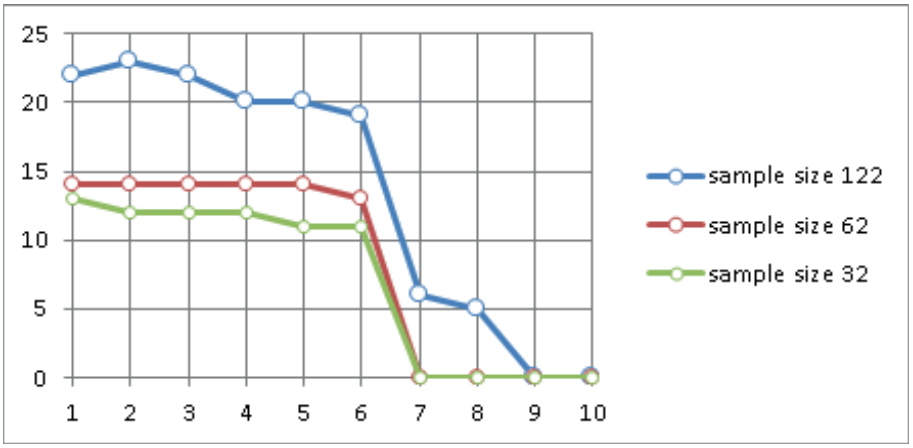

Figure 1. Numbers of statistically significant relations for the "Rhythms" data set

$$\begin{pmatrix}
1 & 0.558 & 0.568 & 0.152 & 0.346 & 0.251 & 0.263 & -0.021 & -0.212 & -0.086 & -0.074 \\
0.558 & 1 & 0.776 & 0.058 & 0.198 & 0.212 & 0.013 & -0.015 & -0.0001 & 0.164 & 0.277 \\
0.568 & 0.776 & 1 & 0.286 & 0.472 & 0.267 & 0.063 & 0.142 & 0.112 & 0.227 & 0.154 \\
0.152 & 0.058 & 0.286 & 1 & 0.686 & 0.291 & 0.035 & 0.048 & -0.069 & -0.064 & 0.173 \\
0.346 & 0.198 & 0.472 & 0.686 & 1 & 0.425 & 0.072 & 0.153 & 0.037 & 0.029 & 0.212 \\
0.251 & 0.212 & 0.267 & 0.291 & 0.425 & 1 & 0.782 & 0.195 & 0.152 & 0.108 & 0.039 \\
0.263 & 0.013 & 0.063 & 0.035 & 0.072 & 0.782 & 1 & 0.110 & 0.055 & -0.001 & -0.021 \\
-0.021 & -0.015 & 0.142 & 0.048 & 0.153 & 0.195 & 0.110 & 1 & 0.807 & 0.830 & 0.696 \\
-0.212 & -0.0001 & 0.112 & -0.069 & 0.037 & 0.152 & 0.055 & 0.807 & 1 & 0.904 & 0.724 \\
-0.086 & 0.164 & 0.227 & -0.064 & 0.029 & 0.108 & -0.001 & 0.830 & 0.904 & 1 & 0.765 \\
-0.074 & 0.277 & 0.154 & 0.173 & 0.212 & 0.039 & -0.021 & 0.696 & 0.724 & 0.765 & 1
\end{pmatrix}.$$

In this experiment, the sample size is unknown to us. The steps of the condition number optimization procedure are shown in Table 4 (values are rounded). At step 0, the eigenvalues are not changed; the condition number of a positive definite matrix $S(11,11)$ is given.

At each step of the procedure, starting from the first, the corresponding eigenvalues (*Eval nums* in the table) are changed. Next, the maximal (*Max eval*) and minimal (*Min eval*) eigenvalues, the condition number (*Cond*), the gain in conditionality (*Ratio* is the ratio of the initial condition number at step 0 to the condition number), the number of significant relations for three sample sizes (*Size 122, Size 62, Size 32*) are determined. Changes in the number of significant relations are also shown in Fig. 1.

A sharp drop in the number of significant relations for samples of all sizes occurs after step 6. As a result, it is quite acceptable to reduce the condition number to 7.503, i.e. the conditionality can be improved by approximately 16 000 times relative to the original value of 120973.9.

Table 4. Condition numbers ("Rhythms")

Step	Eval nums	Max eval	Min eval	Cond	Ratio	Size 122	Size 62	Size 32
0	-	3.629217	0.00003	120973.9	1	24	14	13
1	11	3.609	0.069	52.304	2313	22	14	13
2	10-11	3.536	0.145	24.386	4961	23	14	12
3	9-11	3.519	0.156	22.558	5363	22	14	12
4	8-11	3.348	0.247	13.555	8925	20	14	12
5	7-11	3.151	0.338	9.322	12977	20	14	11
6	6-11	2.986	0.398	7.503	16123	19	13	11
7	5-11	2.031	0.714	2.845	42522	6	0	0
8	4-11	1.840	0.771	2.387	50680	5	0	0
9	3-11	1.236	0.948	1.304	92771	0	0	0
10	2-11	1	1	1	120973.9	0	0	0

The pairwise comparison matrix for the condition number 52.304 is as follows:

$$\begin{pmatrix}
1 & 0.548 & 0.569 & 0.150 & 0.342 & 0.253 & 0.258 & -0.022 & -0.214 & -0.087 & -0.069 \\
0.548 & 1 & 0.754 & 0.062 & 0.205 & 0.195 & 0.024 & -0.010 & 0.005 & 0.166 & 0.258 \\
0.569 & 0.754 & 1 & 0.282 & 0.462 & 0.274 & 0.055 & 0.138 & 0.108 & 0.223 & 0.164 \\
0.150 & 0.062 & 0.282 & 1 & 0.686 & 0.286 & 0.037 & 0.050 & -0.067 & -0.063 & 0.169 \\
0.342 & 0.205 & 0.462 & 0.686 & 1 & 0.414 & 0.077 & 0.155 & 0.040 & 0.030 & 0.203 \\
0.253 & 0.195 & 0.274 & 0.286 & 0.414 & 1 & 0.765 & 0.189 & 0.146 & 0.104 & 0.051 \\
0.258 & 0.024 & 0.055 & 0.037 & 0.077 & 0.765 & 1 & 0.113 & 0.059 & 0.001 & -0.031 \\
-0.022 & -0.010 & 0.138 & 0.050 & 0.155 & 0.189 & 0.113 & 1 & 0.807 & 0.830 & 0.686 \\
-0.214 & 0.005 & 0.108 & -0.067 & 0.040 & 0.146 & 0.059 & 0.807 & 1 & 0.904 & 0.714 \\
-0.087 & 0.166 & 0.223 & -0.063 & 0.030 & 0.104 & 0.001 & 0.830 & 0.904 & 1 & 0.756 \\
-0.069 & 0.258 & 0.164 & 0.169 & 0.203 & 0.051 & -0.031 & 0.686 & 0.714 & 0.756 & 1
\end{pmatrix}.$$

Then, the pairwise comparison matrix for the ultimate condition number 7.503 is:

$$\begin{pmatrix}
1 & 0.478 & 0.482 & 0.161 & 0.275 & 0.272 & 0.191 & -0.083 & -0.138 & -0.067 & -0.039 \\
0.478 & 1 & 0.538 & 0.079 & 0.215 & 0.134 & 0.037 & 0.060 & 0.029 & 0.123 & 0.140 \\
0.482 & 0.538 & 1 & 0.232 & 0.359 & 0.219 & 0.072 & 0.115 & 0.063 & 0.147 & 0.195 \\
0.161 & 0.079 & 0.232 & 1 & 0.524 & 0.232 & 0.044 & 0.048 & -0.028 & -0.031 & 0.099 \\
0.275 & 0.215 & 0.359 & 0.524 & 1 & 0.309 & 0.110 & 0.112 & 0.030 & 0.047 & 0.163 \\
0.272 & 0.134 & 0.219 & 0.232 & 0.309 & 1 & 0.519 & 0.162 & 0.097 & 0.079 & 0.074 \\
0.191 & 0.037 & 0.072 & 0.044 & 0.110 & 0.519 & 1 & 0.088 & 0.047 & 0.016 & -0.038 \\
-0.083 & 0.060 & 0.115 & 0.048 & 0.112 & 0.162 & 0.088 & 1 & 0.581 & 0.576 & 0.546 \\
-0.138 & 0.029 & 0.063 & -0.028 & 0.030 & 0.097 & 0.047 & 0.581 & 1 & 0.592 & 0.547 \\
-0.067 & 0.123 & 0.147 & -0.031 & 0.047 & 0.079 & 0.016 & 0.576 & 0.592 & 1 & 0.564 \\
-0.039 & 0.140 & 0.195 & 0.099 & 0.163 & 0.074 & -0.038 & 0.546 & 0.547 & 0.564 & 1
\end{pmatrix}.$$

For a normalized matrix of scalar products $S(m,m)$, it is known that

$$\text{tr } S = \text{tr L} = m,$$

where its eigenvalues show the distribution of data dispersion $\sigma^2 = m$ between eigen directions.

Compared to the original matrix of paired comparisons with conditionality 120973.9, the sums of the squared deviations of the values of the elements of the two other matrices from their elements are: $D = 0.006298$ for conditionality 52.304, and $D = 1.194265$ for conditionality 7.503. Thus, relative to the variance of the data $\sigma^2 = 11$, the weighted total deviation ($100\% \cdot D/\sigma^2$) is approximately 0.06%, for conditionality 52.304, and almost 11%, for conditionality 7.503, of the data variance.

Thus, distortion of the original data by no more than 0.1% of the data variance makes it possible to improve the conditionality by the factor exceeding 2 000 (for the conditionality of 52.304). In the extreme case, the original data are distorted by almost 11%, allowing the conditionality to be improved at least by the factor of 16 000 (for the conditionality of 7.503).

In the second example, we consider K. Holzinger's data on 145 Chicago schoolchildren, who were subject in 1935 to psychological tests in order to evaluate their intellectual development. Experimental data are represented by a correlation matrix $S(24, 24)$ of test results and are given in Harman (1976).

All eigenvalues of this matrix are positive: 8.129618, 2.075669, 1.679600, 1.507071, 1.027226, 0.941721, 0.891253, 0.811778, 0.796299, 0.707289, 0.645087, 0.546671, 0.524084, 0.513335, 0.469656, 0.401345, 0.375913, 0.367564, 0.354657, 0.312689, 0.292714, 0.281258, 0.220879, and 0.126624. The condition number is 64.203 (rounded). Although it does not matter here, the data are considered complex, since researchers before were unable to satisfactorily solve the so-called problem of bi-factor analysis to identify groups of tests evaluating certain aspects of intelligence development (Table 5 below). For more details about the experiment a Reader is referred to Harman (1976).

In this experiment, the sample size is 145 trials. According to Table 3 of critical values, this is a sample of sufficient statistical size. The steps of the optimization procedure are shown in Table 6 (values are rounded). At step 0, the eigenvalues are not changed, and the condition number of the original matrix $S(24, 24)$ is given.

Table 5. Psychological tests

Group number	Group	Test numbers
1	Spatial relations	1 – 4
2	Verbal	5 – 9
3	Perceptual speed	10 – 13
4	Memory	14 – 19
5	Deduction	20 – 24

As before, at each step of the procedure, starting from the first step, the values of the corresponding eigenvalues (*Eval nums*) were changed, the maximal (*Max eval*) and minimal (*Min eval*) eigenvalues, the condition number (*Cond*), and the gain in conditionality (*Ratio* is the ratio of the initial condition number at the step 0 to the condition number), the number of significant relations for three possible sample sizes (*Size 122, Size 62, Size 32*) are determined. The change in the number of significant relations is also shown in Fig. 2.

Table 6. Condition numbers (Chicago schoolchildren)

Step	Eval nums	Max eval	Min eval	Cond	Ratio	Size 122	Size 62	Size 32
0	-	8.129618	0.126624	64.203	1	213	137	47
1	24	8.089	0.217	37.276	1.72	213	137	47
2	23-24	8.040	0.272	29.559	2.17	212	134	46
3	22-24	8.027	0.282	28.465	2.26	212	133	46
4	21-24	7.998	0.299	26.749	2.40	212	131	45
5	20-24	7.922	0.332	23.861	2.69	211	131	40
6	19-24	7.895	0.342	23.085	2.78	210	128	40
7	18-24	7.875	0.349	22.564	2.85	210	125	39
8	17-24	7.807	0.367	21.272	3.02	209	123	35
9	16-24	7.609	0.413	18.424	3.48	206	113	30
10	15-24	7.475	0.440	16.989	3.78	205	108	24
11	14-24	7.439	0.446	16.679	3.85	202	106	24
12	13-24	7.359	0.459	16.033	4.00	201	103	21
13	12-24	7.003	0.510	13.731	4.68	194	87	12
14	11-24	6.782	0.539	12.583	5.10	189	79	8
15	10-24	6.468	0.577	11.210	5.73	175	58	5
16	9-24	6.414	0.583	11.002	5.84	172	56	1
17	8-24	6.132	0.612	10.020	6.41	149	39	0
18	7-24	5.958	0.629	9.472	6.78	135	24	0
19	6-24	5.670	0.656	8.643	7.43	111	11	0
20	5-24	4.422	0.773	5.721	11.22	14	0	0
21	4-24	4.086	0.803	5.088	12.62	0	0	0
22	3-24	3.457	0.847	4.081	15.73	0	0	0
23	2-24	1	1	1	64.203	0	0	0

In this experiment, the decrease in the number of significant relations occurs more smoothly. If we focus on all sample sizes, then empirically we can consider the threshold step to be step 12, where the number of significant rela-

tions clearly drops. If we consider only a sample of sufficient size, then a more conservative estimation would point to step 4, when the number of significant relations decreased by only 1.

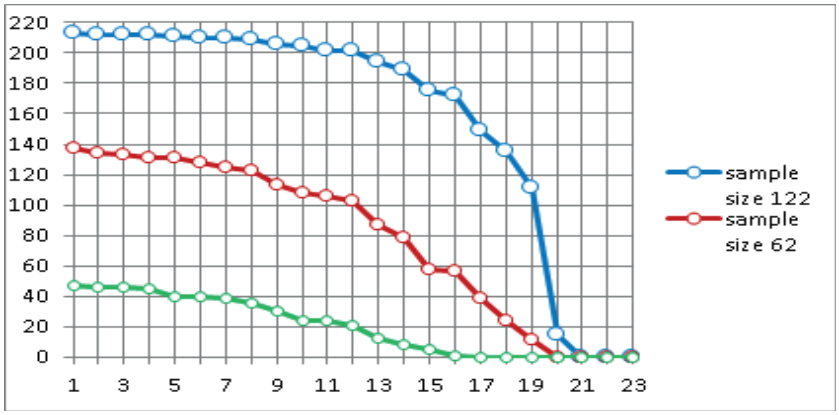

Figure 2. Numbers of statistically significant relations for "Chicago schoolchildren" data

As a result, it is quite acceptable to reduce the condition number to 16.033, i.e. conditionality can be improved 4 times over compared to the value of 64.203. A more careful evaluation shows that it is quite acceptable to reduce the condition number to 26.749, i.e. conditionality can be improved by the factor of 2.4.

Compared to the original matrix of paired comparisons with conditionality 64.203, the sum of squared deviations of the values of the elements of the other two matrices from its elements is $D = 0.075812$ for conditionality 26.749, and $D = 1.173715$ for conditionality 16.033. Therefore, relative to the variance $\sigma^2 = 24$, the weighted total deviation $(100\% \cdot D/\sigma^2)$ is approximately 0.3% for condition number 26.749 and almost 5% for condition number 16.033.

Thus, the distortion of the initial data by no more than 0.3% of the variance improves the conditionality 2.4 times (for conditionality 26.749). In the extreme case, the original data is distorted by close to 5%, allowing the conditionality to be improved 4 times (for conditionality 16.033).

In the third example, we consider data on the correlations of eight physical parameters measured for 305 girls in Chicago in 1935 (Harman, 1976). The first four parameters characterize "lankiness" (height, arm span, forearm length, lower leg length), and the next four parameters characterize "stockiness" (weight, bi-trochanteric diameter, chest girth, chest width). Correlations of these physical parameters are represented by the matrix $S(8, 8)$:

$$\begin{pmatrix} 1 & 0.846 & 0.805 & 0.859 & 0.473 & 0.398 & 0.301 & 0.382 \\ 0.846 & 1 & 0.881 & 0.826 & 0.376 & 0.326 & 0.277 & 0.415 \\ 0.805 & 0.881 & 1 & 0.801 & 0.380 & 0.319 & 0.237 & 0.345 \\ 0.859 & 0.826 & 0.801 & 1 & 0.436 & 0.329 & 0.327 & 0.365 \\ 0.473 & 0.376 & 0.380 & 0.436 & 1 & 0.762 & 0.730 & 0.629 \\ 0.398 & 0.326 & 0.319 & 0.329 & 0.762 & 1 & 0.583 & 0.577 \\ 0.301 & 0.277 & 0.237 & 0.327 & 0.730 & 0.583 & 1 & 0.539 \\ 0.382 & 0.415 & 0.345 & 0.365 & 0.629 & 0.577 & 0.539 & 1 \end{pmatrix}.$$

All eigenvalues of this matrix are positive: 4.672880, 1.770983, 0.481035, 0.421441, 0.233221, 0.186674, 0.137304, and 0.096463. The condition number is 48.442.

In this experiment, the sample size is 305. According to Table 3 of critical values, this is a sample of sufficient statistical size. The steps of the optimization procedure are shown in Table 7 (values are rounded). As before, at step 0 the eigenvalues are not changed, the condition number of the original matrix is given. The change in the number of significant relations is also shown in the diagram of Fig. 3.

Table 7. Condition numbers (Physical variables)

Step	Eval nums	Max eval	Min eval	Cond	Ratio	Size 122	Size 62	Size 32
0	-	4.67288	0.096463	48.4422	1	28	26	15
1	8	4.645	0.135	34.407	1.41	28	26	14
2	7-8	4.5799	0.1795	25.515	1.90	28	26	14
3	6-8	4.496	0.218	20.624	2.35	28	26	14
4	5-8	4.105	0.349	11.762	4.12	28	23	12
5	4-8	3.975	0.384	10.352	4.68	28	22	12
6	3-8	2.176	0.796	2.734	17.72	0	0	0
7	2-8	1	1	1	48.4422	0	0	0

A sharp drop in the number of significant relations for samples of all sizes occurs after step 5. As a result, it is acceptable to reduce the condition number to 10.352, so that conditionality can be improved approximately 4.5 times relative to the original value of 48.44. The pairwise comparison matrix for the condition number 34.407 is shown below:

$$\begin{pmatrix} 1 & 0.823 & 0.806 & 0.855 & 0.467 & 0.397 & 0.302 & 0.383 \\ 0.823 & 1 & 0.854 & 0.820 & 0.379 & 0.322 & 0.270 & 0.407 \\ 0.806 & 0.854 & 1 & 0.795 & 0.373 & 0.318 & 0.238 & 0.346 \\ 0.855 & 0.820 & 0.795 & 1 & 0.437 & 0.329 & 0.326 & 0.364 \\ 0.467 & 0.379 & 0.373 & 0.437 & 1 & 0.761 & 0.728 & 0.627 \\ 0.397 & 0.322 & 0.318 & 0.329 & 0.761 & 1 & 0.583 & 0.577 \\ 0.302 & 0.270 & 0.238 & 0.326 & 0.728 & 0.583 & 1 & 0.539 \\ 0.383 & 0.407 & 0.346 & 0.364 & 0.627 & 0.577 & 0.539 & 1 \end{pmatrix}.$$

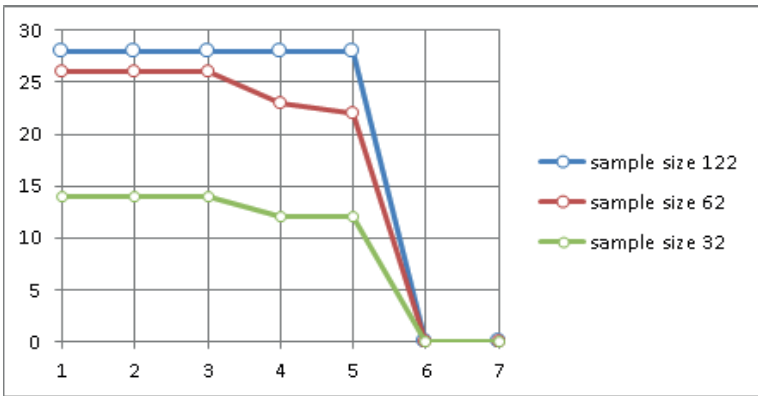

Figure 3. Numbers of statistically significant relations for the "Physical variables" data

The pairwise comparison matrix for the limit condition number 10.352 is as follows:

$$\begin{pmatrix} 1 & 0.614 & 0.608 & 0.611 & 0.362 & 0.321 & 0.276 & 0.360 \\ 0.614 & 1 & 0.613 & 0.612 & 0.329 & 0.288 & 0.241 & 0.330 \\ 0.608 & 0.613 & 1 & 0.607 & 0.312 & 0.271 & 0.224 & 0.314 \\ 0.611 & 0.612 & 0.607 & 1 & 0.346 & 0.305 & 0.260 & 0.345 \\ 0.362 & 0.329 & 0.312 & 0.346 & 1 & 0.578 & 0.567 & 0.560 \\ 0.321 & 0.288 & 0.271 & 0.305 & 0.578 & 1 & 0.553 & 0.539 \\ 0.276 & 0.241 & 0.224 & 0.260 & 0.567 & 0.553 & 1 & 0.527 \\ 0.360 & 0.330 & 0.314 & 0.345 & 0.560 & 0.539 & 0.527 & 1 \end{pmatrix}.$$

Compared to the original matrix of pairwise comparisons with conditionality of 48.4422, the sum of the squared deviations of the values of the elements of the two other matrices from its elements is $D = 0.003145$ for the conditionality of 34.407 and $D = 0.861147$ for the conditionality of 10.352. Thus, relative to the

variance $\sigma^2 = 8$, the weighted total deviation ($100\% \cdot D/\sigma^2$) is approximately 0.04% for the conditionality 34.407 and about 11% for the conditionality 10.352.

A well-known peculiarity of these data is that the physical parameters of the body naturally fall into two groups of parameters that are strongly correlated within a group and weakly correlated between groups. It is easy to see that for all correlation matrices with conditionality from 48.4422 to 10.352, the partition into two groups of strongly correlated features is preserved.

It should be noted that all correlations between physical parameters are positive. Therefore, these data clearly show that improving the conditionality of the matrix of scalar products reduces the modules of its elements. In data mining, the mutual relationships between the elements of the set (similarity) are ultimately analyzed. This follows from the law of cosines in Euclidean space when the elements of a set are immersed in it without metric violations.

Therefore, from the point of view of practical requirements in data mining, the problem of improving the conditionality of the pairwise comparison matrix should be solved taking into account the acceptable threshold for changes in the values of scalar products. In this case, their statistical interpretation as variations allows us to obtain such an evaluation.

7. Conclusion

The processing of pairwise comparisons in modern conditions is often necessary because it is convenient to present experimental data in this form. Under these conditions, it is necessary to ensure that the processing results coincide with the results as if the data were represented in classical vector form by measurements.

From the mathematical point of view, the elements of a set represented by pairwise comparisons must be immersed in a metric space. Here, we discuss the properties of experimental data represented only by pairwise comparisons. In particular, the problems of eliminating metric violations and improving the conditionality of pairwise comparison matrices are considered.

References

Aizerman, M. A., Braverman, E. M. and Rozonoer, L. I. (1970) *The Method of Potential Functions in Machine Learning Theory* [in Russian]. Nauka, Moscow.

Bishop, R. L. and Crittenden, R. J. (1964) *Geometry of Manifolds.* Academic Press, NY.

Boyd, S., Ghaoui, L., Feron, E. and Balakrishnan, V. (1994) *Linear Matrix Inequalities in System and Control Theory.* SIAM, Philadelphia.

Cox, T. F. and Cox, M. A. A. (2001) *Multidimensional Scaling.* Chapman & Hall/CRC.

Dvoenko, S. D. (2009) Clustering and separating of a set of members in terms of mutual distances and similarities. *Trans. on MLDM.* IBaI Publishing, **2**(2), 80–99.

Dvoenko, S. D. and Pshenichny, D. O. (2018) On metric correction and conditionality of raw featureless data in machine learning. *Pattern Recognit. Image Anal.* **28**, 595–604. https://doi.org/10.1134/S1054661818040089

Dvoenko, S. D. and Pshenichny, D. O. (2021) Rank aggregation based on new types of the Kemeny's Median. *Pattern Recognit. Image Anal.* **31**, 185–196. https://doi.org/10.1134/S1054661821020061

Dvoenko, S. (2022) Clustering of data represented by pairwise comparisons. *Control and Cybernetics.* **51**(3) 343–387. https://doi.org/10.2478/candc-2022-0021

Duda R. O. and Hart P. E. (1973) *Pattern Classification and Scene Analysis.* Wiley, NY.

Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics.* **7**, 179–188.

Gazprom (2000) http://www.bastion.ru/research/am/economy/

Harman, H. H. (1976) *Modern Factor Analysis.* Univ. of Chicago Press, Chicago.

Horn, R. and Johnson, C. (1990) *Matrix Analysis.* Cambridge University Press.

Kalyaev, I. A., Levin, I. I., Semernikov, E. A. and Shmoilov, V. I. (2012) *Reconfigurable Multipipeline Computing Structures.* Nova Science Publishers, Inc. USA.

Kemeny, J. and Snell, J. (1963) *Mathematical Models in the Social Sciences.* Blaisdell, New York.

Kozinets, B. N. (1973) A recurrent algorithm for separating convex hulls of two sets. *Pattern recognition learning algorithms* [in Russian], V. N. Vapnik, editor. Soviet Radio, Moscow, 43–50.

Luce, R. D. (1959) *Individual Choice Behavior: A Theoretical Analysis.* Wiley, NY.

Litvak, B. G. (1982) *Expert Information: Methods of Acquisition and Analysis* [in Russian]. Radio i Svyaz, Moscow.

Mirkin, B. G. (1974) *A Group Choice Problem* [in Russian]. Nauka, Moscow.

Nebylitsyn, V. D. (1990) *Selected Psychological Proceedings* [in Russian]. Pedagogika, Moscow.

Pekalska, E. and Duin, R. P. W. (2005) *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications.* World Scientific, Singapore.

Torgerson, W. S. (1958) *Theory & Methods of Scaling*. Wiley, NY.
Tou, J. T. and Gonzalez, R. C. (1977) *Pattern Recognition Principles*. Addison-Wesley.
Vapnik, V. N. and Chervonenkis, A. Y. (1974) *Pattern Recognition Theory* [in Russian]. Nauka, Moscow.
Vapnik, V. N. (1998) *Statistical Learning Theory. Adaptive and Learning Systems*. Wiley, NY.
Young, G. and Householder, A. S. (1938) Discussion of a set of points in terms of their mutual distances. *Psychometrika*. **3**(1). 19–22. https://doi.org/10.1007/BF02287916
Zagoruiko, N. G. (1999) A*pplied Methods of Data and Knowledge Analysis* [in Russian]. IM SBRAS, Novosibirsk.

Appendices

Appendix 1. Numerical example of immersion of measurements and pairwise comparisons in Euclidean space

If a set of elements $\Omega = \{\omega_1, ..., \omega_m\}$ is represented by the measurements of n features, then it is usually considered to be immersed in Euclidean feature space of dimension n. For this situation, there are no problems in calculating, for example, scalar products and developing algorithms that analyze similarity (as positive scalar products of the vectors located in the same quadrant of a coordinate space).

The situation when $m > n$ appears to be preferable since in this case, we do not face the so called "curse of dimensionality". This is a well-known problem, since in data mining and machine learning problems the lack of measurements (objects) in a multidimensional space, when there is $m \leq n$, leads, for example, to excess power of the linear decision rule. Let the objects be concentrated into two subsets. An insufficient number of observations (objects) usually leads to the possibility in a given space of dividing them into various pairs of subsets, where the desired partition (concentrations to be identified) turns out to be only one of them. Therefore, as follows from the well-known idea of VC-dimension, it is necessary to limit the separating ability of the decision rule. It is known as the fact that real data should not be separable in all possible ways (Vapnik and Chervonenkis, 1974; Vapnik, 1998). Then the partitions can more reasonably characterize in a coordinate space the configurations of the mutual arrangement of the elements of a set.

It is known (Young and Householder, 1938) that pairwise comparisons of a set of m elements can be immersed in a metric space of dimension no bigger than the set cardinality m.

If there are only pairwise comparisons represented in a matrix form, it is not known from which coordinate space the data are extracted. In this case, we fall into a "bad" situation, because the dimension m of a metric space always coincides with the cardinality of the set immersed in it, represented only by pairwise comparisons of its elements.

But, of course, we can suppose the expected dimension of the space inaccessible to us by the eigenvalues of the similarity matrix $S(m,m)$. Then, if there really is $m > n$, the corresponding matrix of scalar products $S(m,m)$ appears to be positive semidefinite since the set is located in a coordinate space of smaller dimension than m. The rank of a matrix is less than m if some eigenvalues are zero or "sufficiently" small, which allows them to be considered zero. What is considered as "sufficiently" small is usually decided in an appropriate optimization problem, such as a multidimensional scaling problem (Cox and Cox, 2001), etc.

However, unlike the problem of multidimensional scaling, we consider a situation where we do not need to explicitly restore the so-called stimuli space. We believe that pairwise comparisons are sufficient to analyze the relative positions of the set elements in metrically correct configurations. But in this case, the following contradiction arises.

On the one hand, the condition $rank\ S \leq m$ allows us to assume that there are enough dimensions in the reconstructed space of stimuli. The goal of multidimensional scaling is to reduce the dimensionality of data as much as possible.

On the other hand, it is necessary to immerse the set represented by pairwise comparisons $S(m,m)$ into metric space without violations. In this case, the matrix $S(m,m)$ is considered incorrect due to zero eigenvalues and has infinite conditionality. Since the task to restore the coordinate space of stimuli is not defined here, the matrix $S(m,m)$ should be corrected to provide the full rank. Thus, it is necessary here to solve the opposite problem and maximize the data dimensionality. It is obvious that in this case, formally, we do not have enough measurements. It is known that in such a situation it is necessary to implement the "regularization of a decision rule". Then, during the correction, the additional requirement to get the optimal conditionality should be considered as some sort of specific regularization at the data level.

Let us consider the well-known Iris data (Fisher, 1936). They represent measurements of four characteristics (sepal length and width, petal length and width) from 50 plant specimens of each of three plant species (Iris Setosa, Iris Versicolor, Iris Virginica), 150 specimens in total. It is known that the first class (Iris Setosa) is well separated from the other two (Iris Versicolor, Iris Virginica), which partially overlap. In the published original, the data on instances 102 and 143 are the same. Both of them belong to the species Iris Virginica. Therefore,

in order not to lose the object, here we simply slightly changed the feature values of instance 143. It is easy to see that the feature correlation matrices

$$\begin{pmatrix} 1 & -0.1176 & 0.8718 & 0.8179 \\ -0.1176 & 1 & -0.4284 & -0.3661 \\ 0.8718 & -0.4284 & 1 & 0.9629 \\ 0.8179 & -0.3661 & 0.9629 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & -0.1178 & 0.8715 & 0.8186 \\ -0.1178 & 1 & -0.4278 & -0.3649 \\ 0.8715 & -0.4278 & 1 & 0.9629 \\ 0.8186 & -0.3649 & 0.9629 & 1 \end{pmatrix}$$

and their eigenvalues before (2.9185, 0.9140, 0.1468, 0.0207) and after (2.9182, 0.9141, 0.1467, 0.0209) such a change in the data are practically the same. Note that the variance of the normalized data is 4, with the first three eigenvalues explaining 99.48% of the data variance. That is, the measurement data are almost three-dimensional.

The matrix of scalar products of objects of size 150 x 150 consists of their positive values, which can be conveniently considered as similarities (due to its size, it is not shown here). Theoretically, it has four positive eigenvalues, the sum of which is 150, and all the remaining 146 eigenvalues are zero. The variance of the data is 150.

When calculating eigenvalues, different computational methods give different results for small eigenvalues. As a result, this similarity matrix has three large eigenvalues (rounded): 128.51048, 21.12805, and 0.36148, and for all other eigenvalues, there is no more than 10^{-5} to 10^{-4} of data variance, where the fourth eigenvalue does not exceed 10^{-6}. Among the eigenvalues from the 4^{th} to the 150^{th} there are positive and negative ones with a modulus of no more than 10^{-6}, i.e. this is a computational garbage, the errors of computational methods.

But for the correction methods discussed in this paper, this similarity matrix needs to be corrected to fit pairwise comparisons into 150-dimensional metric space. Methods of individual optimal correction are not entirely suitable here, because it is necessary to correct violations for 146 objects, i.e. for almost every one of them. It is better to apply the correction to all pairwise comparisons at once.

Therefore, we apply a correction method based on direct changes in eigenvalues. It is easy to see that due to the large difference between the third and fourth eigenvalues by almost 5 orders of magnitude, it is necessary to immediately provide a significantly smaller difference while doing the replacement. In fact, it is possible to get rid of very small positive and negative eigenvalues only by starting with the correction of the eigenvalues from the 5^{th} to the 150^{th}

by the value $6.8 \cdot 10^{-7}$ of the fourth eigenvalue, where the condition number 21493798795.15 appears to be unacceptably large.

When correcting the eigenvalues from the 4^{th} to the 150^{th} by a value of no more than 10^{-3}, the conditionality turns out to be five orders of magnitude smaller and already appears to be equal 128599.2, where the first four eigenvalues have the values (rounded): 128.3846, 21.1074, 0.3611 and $\approx 10^{-3}$.

Compared to the original matrix of pairwise comparisons with theoretically infinite conditionality, the sum of squared deviations of the values of the elements of the modified matrix from the original values is $D = 0.016418$. Thus, relative to the $\sigma^2 = 150$, the weighted total deviation $(100\% \cdot D/\sigma^2)$ is approximately 0.01%. Hence, these pairwise comparisons become immersed in 150-dimensional space practically unchanged.

It should be also noted that the results of cluster analysis in our experiments, based on the specially developed k-means algorithm for pairwise comparisons (Dvoenko, 2022), in both cases (incorrect Iris similarity matrix and corrected matrix) appear to be the same. In both cases, the first class (Iris Setosa) is completely separated (objects 1 through 50). Because the second (objects 51-100) and third (obcects 101-150) classes partially overlap, they are separated with 6 errors, where three objects from the second class (69, 73, 84) are assigned to the third class, and three objects from the third class (111, 139, 142) are assigned to the second class.

We showed that modifying pairwise comparisons in order to immerse them in a multidimensional metric space allowed us to obtain the same processing results in this case. Note that if there are only pairwise comparisons, immersing in a metric space may require changing some values. If we imagine that they were previously deliberately distorted, then their restoration does not at all mean the restoration of unknown "true" values. Only those values are modified that make it possible to eliminate metric violations.

It should be noted that this clustering result in a 150-dimensional metric space is somewhat different from the clustering result in a 4-dimensional space of normalized features. Here, the first class is completely separated again (objects 1-50). Next, the second and third classes are separated with 25 errors, where 11 objects from the second class (51, 52, 53, 57, 66, 71, 76, 77, 78, 86, 81) are assigned to the third class, and 14 objects from the third class (102, 107, 114, 115, 120, 122, 124, 127, 134, 135, 139, 143, 147, 150) are assigned to the second class.

Note that the set of objects, represented by the corrected matrix of their scalar products of the size 150 x 150, turns out to be, in the end, correctly immersed in the 150-dimensional metric space. The "curse of dimensionality" can be overcome by now projecting these normalized pairwise comparisons into

the space of the first few principal components, which explain, for example, 80% of the variance of pairwise comparisons. In other words, we talk here about an analogue of the well-known discrete Karhunen-Loeve expansion, which is usually applied to a traditional data matrix.

It should be noted that this approach and further comparison of clustering results require a separate discussion. We only note the following. On the one hand, it is known that changing the results of measurements to obtain normalized features usually affects the result of partitioning in complex cases (Duda and Hart, 1973). On the other hand, in the case of a lack of measurements, it is necessary to use various methods to regularize the processing results. For example, improving the conditionality of the pairwise comparison matrix may make sense as one of them.

Appendix 2. Numerical example of optimal corrections

Let us consider the normalized matrix of scalar products $S = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & -0.9 \\ 0.5 & -0.9 & 1 \end{pmatrix}$ and calculate its principal minors: $S_1 = 1$, $S_2 = \det \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} = 0.75$, $S_3 = -0.76$.

Let us correct the last row and the last column entirely, so that the last minor has the value $c = 0.1$. Let us calculate the inverse matrix

$$R = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1} = \frac{1}{0.75} \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix} = \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix}$$

and calculate $C = 1 - 0.1/0.75 = 13/15$. We obtain the equations:

$$\begin{cases} \lambda (x_1 r_{11} + x_2 r_{21}) = s_{31} - x_1 \\ \lambda (x_1 r_{12} + x_2 r_{22}) = s_{32} - x_2 \\ x_1^2 r_{11} + x_1 x_2 r_{12} + x_2 x_1 r_{21} + x_2^2 r_{22} = C \end{cases} = \begin{cases} \lambda \left(\frac{4}{3} x_1 - \frac{2}{3} x_2\right) = 0.5 - x_1 \\ \lambda \left(-\frac{2}{3} x_1 + \frac{4}{3} x_2\right) = -0.9 - x_2 \\ \frac{4}{3} x_1^2 - \frac{4}{3} x_1 x_2 + \frac{4}{3} x_2^2 = \frac{13}{15} \end{cases}$$

with the result $x_1 = 0.285487$, $x_2 = -0.624637$, $\lambda = 0.269126$. Substituting this result into the matrix S allows us to verify that its determinant is equal to 0.1.

Let us now correct only the first element of the last row and the last column so that again the last minor has the value of 0.1. We get the system of equations:

$$\begin{cases} \lambda x_1 r_{11} + s_{32} r_{21} = s_{31} - x_1 \\ x_1^2 r_{11} + x_1 s_{23} r_{12} + s_{32} x_1 r_{21} + s_{32}^2 r_{22} = C \end{cases}$$

$$= \begin{cases} \frac{4}{3} \lambda x_1 - \frac{2}{3}(-0.9) = 0.5 - x_1 \\ \frac{4}{3} x_1^2 - \frac{4}{3} x_1 (-0.9) + \frac{4}{3} (-0.9)^2 = \frac{13}{15} \end{cases}$$

with the result $x_1 = -0.243845, \quad \lambda = -0.442427$. Substituting this result into the matrix S also allows us to verify that its determinant is equal to 0.1.

Finally, let us correct only the second element of the last row and last column so that the last minor has the value of 0.1. We get the system of equations:

$$\left\{ \begin{array}{c} \lambda x_2 r_{22} + s_{31} r_{12} = s_{32} - x_2 \\ x_2^2 r_{22} + x_2 s_{13} r_{21} + s_{31} x_2 r_{12} + s_{31}^2 r_{11} = C \end{array} \right. = \left\{ \begin{array}{l} \frac{4}{3}\lambda x_2 - \frac{2}{3}0.5 = -0.9 - x_2 \\ \frac{4}{3}x_2^2 - \frac{4}{3}0.5 x_2 + \frac{4}{3}0.5^2 = \frac{13}{15} \end{array} \right.$$

with the result $x_2 = -0.430074, \quad \lambda = 0.238203$. Substituting this result into the matrix S again allows us to verify that its determinant is equal to 0.1.

This example clearly shows that adjusting all pairwise comparisons (the vector correction) produces corrected values that are more similar to the original values than it would be otherwise. Corrections for individual pairwise comparisons are always stronger. And, as shown in this example, they may appear to be completely different from the original values.