

Distances Parameterized by Size: Models and Adaptation Techniques

Archil Maysuradze¹[0000-0002-1757-556X] and Daria
Petrenko¹[0000-0003-4306-9035]

Lomonosov Moscow State University, Moscow, Russia
maysuradze@cs.msu.ru
daria.petrenko@bk.ru

Abstract. In many domains, the concept of distance is used for initial formulation and subsequent formalization of problems and solution methods. However, for an adequate representation of complex situations, the traditional concept of distance is insufficient, and more expressive families of models are required. In this paper, we propose and investigate theoretically and empirically one of the families — distances parameterized by size. We also introduce the generalized metric axioms as a set of natural requirements in many domains. As examples of applied domains, we can consider transport systems, in which the transportation time depends on the mass of the cargo, or message passing networks, in which the transfer delay depends on the length of the message. The number of combinations of pairs of object and sizes is huge, so the complete description of all the situations is data intensive. The problem of modelling and approximating the collected dissimilarity tensor is posed and solved in various ways. Several models of distances parameterized by size are proposed in the work. For each of the models, sufficient conditions are found on the parameters (theorems on sufficient conditions) that ensure the fulfillment of all the generalized metric axioms. To adapt each of the models, we propose a specific method of conditional optimization. The idea of methods is in iterative conditional minimization of the variational upper bound for the stress function. All the proposed models and methods were implemented and tested on real data on message passing delays between processes in the Lomonosov supercomputer system. Experiments have shown a good quality of approximation for models with a small number of parameters (that is, a high degree of data compression), as well as comparability of losses with unconditional problem statements in which the generalized metric axioms are ignored.

Keywords: Data models · Distance modeling · Data compression · Metric Extraction · MPI time delays

1 Introduction

Nowadays there is a large number of subject areas in which a part of a modeled system by its properties resembles a transport system, and the researcher is

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

interested in some conditional laboriousness of transition from object to object. For example, in case of tasks related to trucking, objects correspond to cities, and labour intensity to time or cost of cargo transportation between them. Another example is a computing cluster in which objects correspond to processes, and laboriousness to message passing time delays.

One can see that for the provided examples the laboriousness of transition between objects is actually determined not only by the objects themselves. In the first example it can also depend on the mass of the transported cargo, in the second one — on the message length. Therefore, in order to formally describe such a laboriousness, the traditional definition of distance as a function of two objects is not enough. We suggest to use a function of three arguments, two of which take objects, and the third one — some conditional “cargo size”. Such a function we propose to call a distance parameterized by size.

Note that the majority of normally functioning transport systems meet a natural set of requirements connecting various objects, sizes and distances. For example, speaking about the task of trucking, in the case of one particular cargo transportation, metric axioms become a mathematical formalization of such requirements. The transportation time of any cargo is non-negative, and if the cities of departure and arrival coincide, then the transfer time equals to zero regardless of the cargo size. For any pair of cities the time of cargo transportation between them does not depend on which of the cities was a starting point and which was a final one. While trucking it is disadvantageous to leave the shortest route to visit an additional object that doesn't require to be visited. The requirements get more complicated when the goods of different sizes are transported. A heavy cargo takes more time to transfer, and transportation of any cargo as a whole takes no more time than sequentially in parts.

Non-compliance of a transport system with any of the described requirements may indicate the presence of malfunctions or potential opportunities of routing improvement. Therefore, it is important to be able to verify the fulfillment of the specified requirements and to estimate how much the system differs from the “correct” one. Accordingly, the formalization of concept of distance parameterized by size itself gives us a tool for quantitative measurement of system quality.

While collecting initial information on transport systems almost always a large amount of data appears. In its original form, it is hard to store and analyze. Therefore, in this paper we propose and investigate approaches to modeling such systems, in which the initial information is significantly “compressed” by approximation by interpreted models of distances parameterized by size.

The problems of modeling distances by various formal systems in the literature are sometimes called distance realization problems [2]. In particular, the problems of approximating data on differences in continuous spaces belong to a class of multidimensional scaling problems [1, 3, 4]. The works on multidimensional scaling also describe situations when for the same pair of objects there exist several distance values (e.g. individual differences models from [3], three-way MDS models from [1]). An example is obtaining information on objects

similarity from several independent experts. However, the existing works do not assume the presence of any relations or operations on a set of quantities describing various distances for a fixed pair of objects. This paper substantially uses the fact that the set of sizes has an order relation and an addition operation. Therefore, the proposed approach to distance modeling is fundamentally new.

The rest of this paper is structured as follows. In section 2 we provide a formal definition of distance parameterized by size and some of its expansions and contractions. In sections 3 and 4 we describe specific models of distances parameterized by size and methods of their adaptation. In section 4 we describe our experimental setup and provide empirical results.

2 Basic Definitions and Notations

Let us introduce a formal definition of distance parameterized by size. Let X be an arbitrary set, S — a partially ordered set with an addition operation.

Definition 1. *Distance parameterized by size is a function $\rho(x_1, x_2, s) : X \times X \times S \rightarrow \mathbb{R}$ that satisfies the following system of axioms:*

- *distance axioms* $\forall x_1, x_2 \in X, \forall s \in S$
- D1. $\rho(x_1, x_1, s) = 0$ (*reflexivity*)
- D2. $\rho(x_1, x_2, s) = \rho(x_2, x_1, s)$ (*symmetry*)
- D3. $\rho(x_1, x_2, s) \geq 0$ (*non-negativity*)
- *size axioms* $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S$
- S1. $s_1 \leq s_2 \implies \rho(x_1, x_2, s_1) \leq \rho(x_1, x_2, s_2)$ (*monotonicity*)
- S2. $\rho(x_1, x_2, s_1 + s_2) \leq \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2)$ (*indivisibility*)

Similar to normal distances, the above definition of distances parameterized by size can be expanded by introduction of additional axioms.

Definition 2. *A function $\rho : X \times X \times S \rightarrow \mathbb{R}$ is called a pseudometric parameterized by size if it satisfies the definition 1 and $\forall x_1, x_2, x_3 \in X, \forall s \in S$*

- D4. $\rho(x_1, x_2, s) \leq \rho(x_1, x_3, s) + \rho(x_2, x_3, s)$ (*triangle inequality*)

Definition 3. *A function $\rho : X \times X \times S \rightarrow \mathbb{R}$ is called a metric parameterized by size if it satisfies the definition 2 and $\forall x_1, x_2 \in X, \forall s \in S$*

- D5. $\rho(x_1, x_2, s) = 0 \iff x_1 = x_2$ (*identity of indistinguishable*)

Due to measurement errors or peculiarities of systems functioning, the real collected data often do not satisfy some of the conditions described. The functions with a relaxed set of requirements in the case of normal distances are often referred to as dissimilarities. Therefore, in some cases in this paper we will talk about dissimilarities parameterized by size.

Definition 4. *A function $\rho : X \times X \times S \rightarrow \mathbb{R}$ is called a dissimilarity parameterized by size if it satisfies the axioms of reflexivity and non-negativity.*

Let us call a function $f : S \rightarrow \mathbb{R}$ subadditive if it satisfies $\forall s_1, s_2 \in S$

$$f(s_1 + s_2) \leq f(s_1) + f(s_2)$$

The set of natural numbers $1, \dots, N$ is denoted by $[1, N]$. For the euclidean metric we will use the notation $\text{eucl}(x, y) = \sqrt{\sum_{k=1}^K (x_k - y_k)^2}$.

Let a finite set of objects of cardinality N and a finite set of sizes of cardinality K be given. Further, for simplicity, we will assume that objects are numbered from 1 to N and sizes are proportional to their indices. Let $\Delta \in \mathbb{R}^{N \times N \times K}$ be a tensor such that δ_{ijk} contains a measured initial difference between the i -th and the j -th objects parameterized by size with index k .

3 Proportional Distances Model

Let us define a proportional distances model and prove that it satisfies all the axioms of distance parameterized by size. Then a method of its adaptation is described.

3.1 Model Definition

Let S be a partially ordered set of sizes with an addition operation. Let X be a set with the defined distance dist . Let r be a function from sizes to real numbers. For $x_1, x_2 \in X$, $s \in S$ consider the function

$$\rho(x_1, x_2, s) = r(s) \text{dist}(x_1, x_2) \quad (1)$$

Theorem 1. *If dist is a distance, and function $r(s)$ is non-negative, monotonous and subadditive, i.e the following is satisfied*

$$\forall s \in S \quad r(s) \geq 0, \quad (2)$$

$$\forall s_1, s_2 \in S \quad s_1 \leq s_2 \Rightarrow r(s_1) \leq r(s_2), \quad (3)$$

$$\forall s_1, s_2 \in S \quad r(s_1 + s_2) \leq r(s_1) + r(s_2), \quad (4)$$

then the function (1) satisfies all the axioms of distance parameterized by size. If dist is a pseudometric then ρ is a pseudometric parameterized by size. If dist is a metric and additionally r is positive then ρ is a metric parameterized by size.

Proof. In [7] it is shown that multiplication of a distance (pseudometric) by a non-negative value guarantees the preservation of distance (pseudometric) axioms. Multiplication of a metric by a positive value guarantees the preservation of metric axioms.

Let us show the fulfillment of the monotonicity axiom (S1). Due to (3) and non-negativity of dist , for any pair of objects $x_1, x_2 \in X$ provided $s_1, s_2 \in S$, $s_1 \leq s_2$ the following is fulfilled

$$\begin{aligned} \rho(x_1, x_2, s_2) - \rho(x_1, x_2, s_1) &= r(s_2) \text{dist}(x_1, x_2) - r(s_1) \text{dist}(x_1, x_2) = \\ &= (r(s_2) - r(s_1)) \text{dist}(x_1, x_2) \geq 0 \end{aligned}$$

Let us show the fulfillment of the indivisibility axiom (S2). Due to (4) and non-negativity of $\text{dist} \forall x_1, x_2 \in X, \forall s_1, s_2 \in S$

$$\begin{aligned} \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2) &= r(s_1) \text{dist}(x_1, x_2) + r(s_2) \text{dist}(x_1, x_2) = \\ &= (r(s_1) + r(s_2)) \text{dist}(x_1, x_2) \geq r(s_1 + s_2) \text{dist}(x_1, x_2) = \rho(x, y, s_1 + s_2) \end{aligned}$$

3.2 Adaptation Method

The proportional distances model with $\text{dist} = \text{eucl}$ is considered. Let us look for the solution in the following form: let us build a uniform “group” configuration $X = [x_1, \dots, x_N]^T$ in a space of a given dimension L and a weight vector $r \in \mathbb{R}^K$. In order to ensure the fulfillment of the axioms of distance parameterized by size, the vector r can be required to satisfy the sufficient conditions of the theorem 1. That is, for the case of sizes being proportional to indices of the input tensor, it is suffice to require

$$r_k \geq 0 \quad \forall k \in [1, K] \quad (5)$$

$$r_{k_2} - r_{k_1} \geq 0 \quad \forall k_1 \leq k_2 \in [1, K] \quad (6)$$

$$r_{k_1} + r_{k_2} - r_{k_1+k_2} \geq 0 \quad \forall k_1, k_2 \in [1, K], k_1 + k_2 \leq K \quad (7)$$

Let W be the predefined objects weights. Let the optimal solution be the one that minimizes the stress function

$$S(r, X) = \sum_{k=1}^s \sum_{i=1}^n \sum_{j=1}^n w_{ijk} (r_k \text{eucl}(x_i, x_j) - \delta_{ijk})^2 \quad (8)$$

and satisfies the conditions (5) — (7).

We will optimize using a coordinate descent method with respect to X and r , minimizing with respect to one group of variables while fixing the values of another one. Let $r^{[t]}, X^{[t]}$ be the values of parameters at the t -th step of iterative process.

Let us fix the value of X . Taking into account the requirements (5) — (7), we get a problem of minimization of a quadratic function with respect to r with linear constraints. The optimal values of parameters can be found using the appropriate methods of constrained optimization. E.g. the active set method can be used [8]. It should be noted that with the change in X only the target function changes in the problem with respect to r , and the constraints remain unchanged. Therefore, the previous approximation $r^{[t]}$ remains in the feasible region after the iteration over X and can be used as a new initial approximation during the next step.

For a fixed r we will optimize with respect to X using the SMACOF algorithm [3]. The stress function can be expressed as

$$S(r, X) = \text{const}(X, r) + \text{tr}(X^T V X) - 2 \text{tr}(X^T B(X) X) \quad (9)$$

where $V = \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^s r_k^2 w_{ijk} \right) A_{ij}$ assuming $A_{ij} = (e_i - e_j)(e_i - e_j)^T$ (10)

$$B(X) = \sum_{i=1}^N \sum_{j=1}^N s_{ij} A_{ij}, \quad s_{ij}(X) = \begin{cases} \frac{\sum_{k=1}^s r_k w_{ijk} \delta_{ijk}}{\text{eucl}(x_i, x_j)}, & \text{eucl}(x_i, x_j) \neq 0 \\ 0, & \text{eucl}(x_i, x_j) = 0 \end{cases} \quad (11)$$

The variational upper bound for it is the function

$$T(r, X, Y) = \text{const}(X, r) + \text{tr}(X^T V X) - 2 \text{tr}(X^T B(Y) Y), \quad (12)$$

The variational upper bound can be minimized as follows:

step 1: $Y^{[t+1]} : T(r, Y^{[t+1]}, X^{[t]}) = S(r, X^{[t]})$, that is, $Y^{[t+1]} = X^{[t]}$

step 2: $X^{[t+1]} = \arg \min_X T(r, X, Y^{[t+1]})$

$T(r, X, Y)$ is quadratic with respect to X , therefore, it reaches a minimum with respect to X at a single point where the derivative equals to 0.

$$\frac{\partial T(r, X, Y)}{\partial X} = 2VX - 2B(Y)Y = 0$$

The matrix V is degenerate as the sum of its elements in each row and in each column equals to 0. We will use the Moore-Penrose inversion. Let V^+ be a pseudoinverse of V . Then

$$\hat{X} = V^+ B(Y) Y \quad (13)$$

Combining the two steps, the next approximation of X can be obtained by the equation:

$$X^{[t+1]} = V^+ B(X^{[t]}) X^{[t]} \quad (14)$$

4 Individual Differences Model

Let us, similarly, give a definition of individual differences model and prove that it satisfies the axioms of distance parameterized by size. We also provide a method of its adaptation.

4.1 Model Definition

Now let X be a set with defined distance dist such that its elements are vectors of length L . Let S be a partially ordered set of sizes with an addition operation. Let $r : S \rightarrow \mathbb{R}^L$ be a function from sizes to vectors of real numbers of length L , and for all elements of X an operation of multiplication by real matrix $\in \mathbb{R}^{L \times L}$ is defined. Then $\forall x_1, x_2 \in X, \forall s \in S$ consider a function

$$\rho(x_1, x_2, s) = \text{dist} \left(x_1 \times \text{diag}(r(s)), x_2 \times \text{diag}(r(s)) \right) \quad (15)$$

One may state that if dist is a distance (a pseudometric), then for all r function ρ satisfies the axioms of distance (pseudometric). Multiplication of a vector by a diagonal matrix is equivalent to multiplication of its i -th component by the i -th diagonal element $\forall i \in [1, L]$. Such a transformation converts different elements of X to equal (with zero components corresponding to $r_i(s) = 0$) or different elements.

If dist is a metric, then if

$$r_i(s) \neq 0 \quad \forall s \in S, \forall i \in [1, L],$$

then ρ satisfies the metric axioms. Multiplication of each component of a vector by a non-zero scalar converts different vectors to different ones.

The constraints that should be imposed on a function r to ensure the fulfillment of size axioms depend on a function dist . Let us consider a special case.

Theorem 2. *If in individual differences model $\text{dist} = \text{eucl}$, and r is non-negative, monotonous and subadditive, i.e it satisfies $\forall s_1, s_2 \in S$*

$$r(s_1) \geq 0 \quad (16)$$

$$s_1 \leq s_2 \Rightarrow r(s_1) \leq r(s_2) \quad (17)$$

$$r(s_1 + s_2) \leq r(s_1) + r(s_2) \quad (18)$$

then ρ satisfies the size axioms (S1) – (S2).

Proof. Let us show the fulfillment of the monotonicity axiom (S1). Let $x_1, x_2 \in X, s_1, s_2 \in S$ be arbitrary elements. It is true that

$$s_1 \leq s_2 \Rightarrow \{(17)\} \Rightarrow r(s_1) \leq r(s_2) \Rightarrow \{(16)\} \Rightarrow r^2(s_1) \leq r^2(s_2) \quad (19)$$

Since the function eucl is non-negative, it is also true that

$$\begin{aligned} \rho(x_1, x_2, s_1) \leq \rho(x_1, x_2, s_2) &\Leftrightarrow \text{eucl} \left(x_1 \times r(s_1), x_2 \times r(s_1) \right) \leq \\ &\leq \text{eucl} \left(x_1 \times r(s_2), x_2 \times r(s_2) \right) \Leftrightarrow \rho^2(x_1, x_2, s_1) \leq \rho^2(x_1, x_2, s_2) \end{aligned} \quad (20)$$

Then $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S, s_1 \leq s_2$

$$\begin{aligned}\rho^2(x_1, x_2, s_1) - \rho^2(x_1, x_2, s_2) &= \sum_{l=1}^L r_l^2(s_1)(x_{1l} - x_{2l})^2 - \sum_{l=1}^L r_l^2(s_2)(x_{1l} - x_{2l})^2 = \\ &= \sum_{l=1}^L (r_l^2(s_1) - r_l^2(s_2))(x_{1l} - x_{2l})^2 \leq \{(19)\} \leq 0\end{aligned}$$

And according to (20) it is true that

$$\rho^2(x_1, x_2, s_1) - \rho^2(x_1, x_2, s_2) \leq 0 \Leftrightarrow \rho(x_1, x_2, s_1) - \rho(x_1, x_2, s_2) \leq 0$$

Now let us show the fulfillment of the axiom (S2). $\forall x_1, x_2 \in X, \forall s_1, s_2 \in S$

$$\begin{aligned}\rho^2(x_1, x_2, s_1 + s_2) &= \sum_{l=1}^L r_l^2(s_1 + s_2)(x_{1l} - x_{2l})^2 \leq \{(16), (18)\} \leq \\ &\leq \sum_{l=1}^L (r_l(s_1) + r_l(s_2))^2 (x_{1l} - x_{2l})^2 = \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2))(x_{1l} - x_{2l})^2 + \\ &+ 2 \sum_{l=1}^L r_l(s_1)r_l(s_2)(x_{1l} - x_{2l})^2 \leq \{(16)\} \leq \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2))(x_{1l} - x_{2l})^2\end{aligned}$$

Since both expressions are non-negative, it is true that

$$\begin{aligned}\rho^2(x_1, x_2, s_1 + s_2) &\leq \sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2))(x_{1l} - x_{2l})^2 \Leftrightarrow \\ &\Leftrightarrow \rho(x_1, x_2, s_1 + s_2) \leq \sqrt{\sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2))(x_{1l} - x_{2l})^2}\end{aligned}$$

Since the square root of a sum of non-negative elements does not exceed the sum of roots of these elements, then

$$\begin{aligned}\rho(x_1, x_2, s_1 + s_2) &\leq \sqrt{\sum_{l=1}^L (r_l^2(s_1) + r_l^2(s_2))(x_{1l} - x_{2l})^2} \leq \\ &\leq \sqrt{\sum_{l=1}^L r_l^2(s_1)(x_{1l} - x_{2l})^2} + \sqrt{\sum_{l=1}^L r_l^2(s_2)(x_{1l} - x_{2l})^2} = \\ &= \rho(x_1, x_2, s_1) + \rho(x_1, x_2, s_2)\end{aligned}$$

4.2 Adaptation Method

The individual differences model with $\text{dist} = \text{eucl}$ is considered. Let us look for the solution in the following form: let us build a uniform “group” configuration

$X = [x_1, \dots, x_N]^T$ in a space of a given dimension L , and a set of weight vectors $r = [r_1, \dots, r_K]^T \in \mathbb{R}^{K \times L}$. In order to ensure the fulfillment of the axioms of distance parameterized by size, the vectors r can be required to satisfy the sufficient conditions of the theorem 2. That is, for the case of sizes being proportional to indices of the input tensor, it is suffice to require

$$r_k \geq 0 \quad \forall k \in [1, K] \quad (21)$$

$$r_{k_2} - r_{k_1} \geq 0 \quad \forall k_1 \leq k_2 \in [1, K] \quad (22)$$

$$r_{k_1} + r_{k_2} - r_{k_1+k_2} \geq 0 \quad \forall k_1, k_2 \in [1, K], k_1 + k_2 \leq K \quad (23)$$

Arithmetic operations and comparisons are element-wise.

We will consider the optimal solution the one that minimizes the stress function

$$S(X, r) = \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N w_{ijk} \left(\text{eucl}(x_i \times \text{diag}(r_k), x_j \times \text{diag}(r_k)) - \delta_{ijk} \right)^2 \quad (24)$$

and satisfies the conditions (21) — (23).

We will optimize using a modified generalization of SMACOF method for weighted euclidean model [3].

Let us denote $X_k = X \times \text{diag}(r_k)$. The stress function can be rewritten as follows:

$$S(X, r) = \sum_{k=1}^K \left(\text{const}(X, r) + \text{tr}(X_k^T V_k X_k) - 2 \text{tr}(X_k^T B(X_k) X_k) \right)$$

Let $Y = \{Y_1, \dots, Y_K\}$. A variational upper bound for the stress function is

$$T(X, Y, r) = \sum_{k=1}^K \left(\text{const}(X, r) + \text{tr}(X_k^T V_k X_k) - 2 \text{tr}(X_k^T B(Y_k) Y_k) \right) \quad (25)$$

Let $\bar{X}_k = V_k^+ B(Y_k) Y_k$. Then

$$T(X, Y, r) = \sum_{k=1}^K \left(\text{const}(X, r) + \text{tr}((X_k - \bar{X}_k)^T V_k (X_k - \bar{X}_k)) - \text{tr}(\bar{X}_k^T V_k \bar{X}_k) \right)$$

We will minimize the variational lower bound, similarly to the method of proportional distances model adaptation, in two steps. Since

$$T(X, \{X \times \text{diag}(r_1), \dots, X \times \text{diag}(r_K)\}, r) = S(X, r),$$

then $Y_k^{[t+1]} = X_k^{[t]}$.

For any fixed Y only the second term depends on X, r , and the equation is quadratic with respect to these variables. The function can be minimized with respect to one group of parameters for fixed values of others. Minimization with

respect to X can be performed by the least squares method, and with respect to r by one of the methods of optimization of quadratic equation with linear constraints (e.g. by the active set method). It should be noted that the optimization with respect to r can be performed independently for each dimension.

Combining the two steps, we get

$$\bar{X}_k = V_k^+ B(X_k^{[t]}) X_k^{[t]} \quad (26)$$

$$X^{[t+1]}, r^{[t+1]} = \arg \min_{X, r} \sum_{k=1}^K \text{tr}((X \times \text{diag}(r_k) - \bar{X}_k)^T V_k (X \times \text{diag}(r_k) - \bar{X}_k)) \quad (27)$$

5 Experiments

The experiments were carried out on data on message passing delays in the Lomonosov supercomputer system. Note that the methods of collecting this information themselves require significant scientific and technological efforts. One can read about the methods of collecting and primary processing of the mentioned data in the papers [5, 6, 9].

In parallel programming using the MPI standard, the program is divided into processes that can interact with each other by exchanging messages. The information about message passing delays is useful to collect and model, as it can help to improve the efficiency of the computing system, in particular, to solve the problems of dynamic scheduling of programs execution, as well as to diagnose the communication environment.

In our experiment the data was collected for 78 processes and message lengths in range from 0 to 10000 in increments of 100 bytes. For every ordered pair of processes and every message length several measurements of delays were taken, after which the median of the obtained empirical distribution was calculated. As a result, a tensor $\Delta \in \mathbb{R}^{78 \times 78 \times 100}$ of dissimilarities parameterized by size was obtained.

To assess how real data correspond to the proposed axiomatic model, for each axiom, all pairs (triples) of values were selected from the input tensor, for which it is correct to check its fulfillment. Then the fraction of pairs (triplets), for which this axiom is incorrect, was calculated. The obtained values are shown in table 1. One can see that the axioms are violated only for an insignificant part of the data, which indicates that the proposed concept quite well describes real systems.

Table 1. Fraction of data that does not satisfy the axioms

axiom	D1	D2	D3	S1	S2
fraction	0	0.181	0	0.008	0.0005

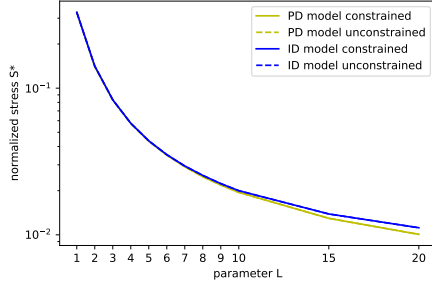


Fig. 1. Normalized stress S^* value depending on parameter L

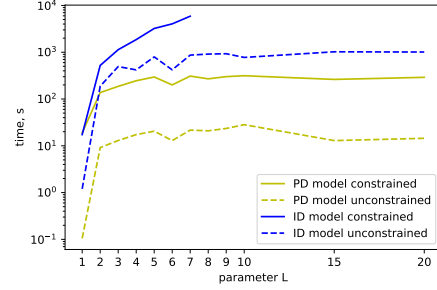


Fig. 2. Models tuning time depending on parameter L value

During further experiments, the proportional distances (PD) model and individual differences (ID) model were adapted to data for different values of L — the solution space dimension. For comparison, these models were also tuned without regard to parameters restrictions. Under this condition they equal to the identity model and the weighted euclidean model from [1], and their adaptation methods coincide with SMACOF generalizations for these models.

Figure 1 shows the final values of normalized stress function

$$S^*(X, r) = \frac{S(X, r)}{\sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \delta_{ijk}^2} \quad (28)$$

for PD and ID models. The stress values for the conditional and unconditional optimization of the models are indistinguishable, thus they look like 2 lines. Figure 2 contains the plots of the dependence of models tuning time on the values of L .

It can be noticed that for the adapted models the normalized stress function takes small (about 10^{-2}) values, which means that the real system is close to the “correct” one. This is also evidenced by the similarity of stress values for conditional and unconditional models optimization: the requirements for the parameters do not impose significant restrictions on the quality of model adaptation. Small values of stress are achieved for rather low values of L , that is, the input tensor allows efficient compression by the proposed models.

The values of stress function for two models substantially differ only for large values of L , and the use of individual differences model can be redundant for our data. While the number of parameters and tuning time for this model is significantly larger.

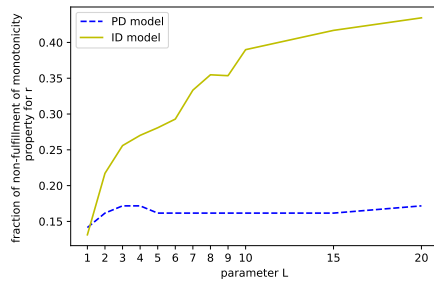


Fig. 3. Fraction of non-fulfillment of monotonicity property for parameter r depending on parameter L value for unconstrained models

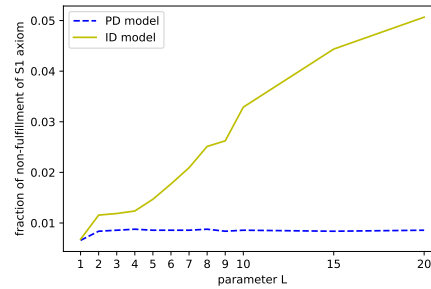


Fig. 4. Fraction on non-fulfillment of S1 axiom depending on the value of parameter L value for unconstrained models

The results of constrained models adaptation satisfy all the axioms of distance parameterized by size. For the results of unconstrained adaptation, the fulfillment of the distance axioms is guaranteed for any parameter values. For the axioms of size, we measured the fraction of the resulting tensor for which these axioms are not fulfilled. The fraction of the vector (matrix) of parameters r that does not satisfy the conditions of non-negativity, monotonicity and subadditivity was also measured. Non-negative values were obtained only for monotonicity constraint and S1 axiom, the plots of dependence on L are shown in figures 3 and 4.

6 Conclusion

We have formally introduced the concept of 'distances parameterized by size' and proposed several specific models of such distances. For each model we have found the sufficient conditions on the parameters that guarantee the fulfillment of all the axioms from the definition, and have provided a specific method of model adaptation. We have demonstrated empirically that the proposed models allow to approximate real data with a high degree of compression.

Acknowledgements. The study was partially supported by RFBR (project no. 20-01-00664-a) and state-financed research work no. 5.1.21 of Lomonosov Moscow State University.

References

1. Borg, I., Groenen, P.J.F.: Modern multidimensional scaling: Theory and applications. Springer Science & Business Media (2005)
2. Chung, F., Garrett, M., Graham, R., Shallcross, D.: Distance realization problems with applications to internet tomography. *Journal of Computer and System Sciences* **63**(3), 432–448 (2001)

3. Cox, T.F., Cox, M.A.: *Multidimensional Scaling*. New York: Chapman and Hall/CRC (2000)
4. Davison, M.L.: *Multidimensional scaling: methods of data visualization*. (in Russian). Finance and Statistics (1988)
5. Kozlov, V., Maysuradze, A.: Separation of a mixture of three-parameter lognormal distributions in the analysis of communication environments. *Computational Mathematics and Modeling* **30**(3), 311–319 (2019)
6. Maysuradze, A., Kozlov, V.: Modeling message passing delays in a computer cluster to monitor its network. In: *CEUR Workshop Proceedings*. pp. 93–99 (2017)
7. Maysuradze, A.I.: Homogeneous and rank bases in spaces of metric configurations (in russian). *Journal of Computational Mathematics and Mathematical Physics* **46**(2), 344–360 (2006)
8. Nocedal, J., Wright, S.: *Numerical optimization*. Springer Science & Business Media (2006)
9. Salnikov, A., Begaev, A., Maysuradze, A.: Similarity mining of message passing delays in supercomputer networks based on peak and step detection. In: *Russian Supercomputing Days*. pp. 486–499. Springer (2020)