DOI: https://doi.org/10.15276/hait.08.2025.17
UDC 519.6:004.93

Improving the performance of clustering with wavelet function by using inequality constraints

Galina Y. Shcherbakova[1)]
ORCID: https://orcid.org/0000-0003-0475-3854; galina.sherbakova@op.edu.ua. Scopus Author ID: 27868185600
Daria V. Koshutina[1)]
ORCID: https://orcid.org/0009-0004-1326-8775; d.v.koshutina@op.edu.ua. Scopus Author ID: 58289385400
[1)] Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

Clustering methods based on gradient estimation are common in automated control and diagnostic systems, where reliable data processing is needed under noise and multimodality. Their use, however, is constrained by low robustness and high computational costs. Wavelet-based approaches are relevant because they enhance noise immunity and improve efficiency. The purpose of this work is to develop and study a clustering method that employs wavelet functions to introduce constraints and ensure stable performance under noisy conditions. The research included analysis of existing approaches, development of a wavelet-based method with inequality-type second-order constraints, creation of an algorithm for its implementation, and experimental evaluation. The proposed method relies on wavelet transforms with hyperbolic functions, which reduce the number of oracle calls, decrease computational stages, and accelerate convergence in classification and clustering problems. Experiments show that the method shortens the search time for the optimum by about one and a half to seven times at different signal-to-noise ratios, with a moderate increase in error of roughly five to fifteen percent for the De Jong test function. On synthetic datasets, the gain in computation time exceeded one point one compared with the baseline method. In a practical case of reliability assessment for resistors used in critical equipment, efficiency improved by nearly eight percent. Finally, the novelty lies in the clustering method with constraints-inequalities defined by wavelet processing. This can increase the computational speed in conditions of high noise levels, asymmetric objective functions, and small data samples.

Keywords: Clustering; wavelet transform; second-order constraints; optimization; oracle calls; noise immunity.

***For citation*: Shcherbakova G. Y., Koshutina D.V. "Improving the performance of clustering with wavelet function by using inequality constraints". *Herald of Advanced Information Technolog*. 2025; Vol.8 No.3: 277–287. DOI: https://doi.org/10.15276/hait.08.2025.17**

INTRODUCTION

The tasks of intelligent data analysis are addressed in a variety of application areas, including biomedical research [1], equipment condition assessment [2], [3], systems with visual information processing [4], smart region support systems [5], [6], and technical diagnostic systems [7]. These diverse applications share a common requirement: the need to extract meaningful patterns and insights from large and often noisy datasets. In many cases, pattern recognition methods serve as the primary tool for understanding data structures, and classification techniques with self-training mechanisms are employed. Such techniques typically involve two complementary procedures: clustering and classification. In clustering, the membership of a pattern, represented as a point in the feature space, is initially unknown. Therefore, it becomes necessary to determine the boundaries and relationships initially unknown. Therefore, it becomes necessary to determine the boundaries and relationships between clusters based on intrinsic data characteristics.

Within the field of intelligent data analysis, clustering algorithms generally group data into clusters based on measures of compactness and similarity. A considerable subset of these algorithms is designed around the optimization of a specific objective function, which quantifies the quality or validity of the clustering solution. The effectiveness of clustering critically depends on the choice of optimization method, which in turn is influenced by the mathematical properties of the objective function. In practical scenarios, this this function is often not explicitly defined; it may present a multimodal surface, exhibit irregularities, and contain significant noise, particularly in cases where clusters are poorly separable.

Additionally, real-world datasets frequently display large imbalances in cluster sizes, which furthe complicates the accurate delineation of cluster boundaries.

To overcome these challenges, methods based on wavelet functions have been proposed. The use of wavelet transforms allows for the introduction of constraints that enhance noise immunity, ensuring

© Shcherbakova G., Koshutina D., 2025

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/deed.uk)

that the clustering procedure remains robust even under adverse conditions [8]. Moreover, these constraints contribute to improved computational efficiency by reducing the number of necessary evaluations of the objective function. Building on these advantages, the present study proposes a clustering method that leverages wavelet-based optimization with inequality constraints. This approach provides both sufficient noise immunity for applied tasks and enhanced computational speed, making it suitable for practical applications where both accuracy and efficiency are critical.

PROBLEM STATEMENT

The purpose of this work is to develop and conduct a comprehensive study of a clustering method based on wavelet functions. The key idea is to enable the introduction of constraints through wavelet processing, which in turn ensures improved performance and enhanced robustness to noise in various applied tasks. The proposed method aims to address common limitations of conventional clustering algorithms, particularly their sensitivity to noisy data and computational inefficiency when dealing with complex, multimodal objective functions.

To achieve this goal, the following tasks were systematically addressed:

– Analysis of modern clustering methods: A detailed review of existing clustering approaches was carried out to identify their strengths and limitations, particularly with respect to noise resilience, computational requirements, and adaptability to poorly separable or imbalanced data clusters. This analysis provides the theoretical foundation for the development of an improved clustering approach.

– Development and justification of a clustering method incorporating constraints derived from wavelet function processing. The core idea of the proposed method lies in using wavelet-based transformations to introduce inequality-type constraints. These constraints guide the clustering process, enhancing both accuracy and stability. A careful justification of the approach was provided, highlighting why wavelet functions are well-suited for improving noise immunity and convergence speed.

– Development of a methodology and algorithm for the practical implementation of the proposed clustering method: A systematic methodology was formulated, detailing the algorithmic steps necessary to apply the proposed method in real-world scenarios. This includes the procedures for wavelet transformation, constraint application, and iterative optimization to achieve efficient clustering outcomes.

– Experimental evaluation of the proposed method to assess its computation speed and noise immunity: To validate the computation speed of the proposed approach, a series of experiments was conducted.

ANALYSIS OF MODERN CLUSTERING METHODS

The initial data for clustering consists of a set of data, each represented by a vector of characteristics in the feature space. That is, the realizations of images in the feature space correspond to geometrically close points, forming so-called “compact” clots [9], [10], [11].

The compactness hypothesis operates using absolute distances between vectors in the feature space. In cases where data must be divided into clusters of complex shapes, the λ-compactness hypothesis is often applied, taking into account normalized distances between images. However, when the number of points in the feature space is large (on the order of thousands), computational costs increase significantly. Additional complications arise when it is necessary to obtain the number of clusters $k >> 2$.

In such situations, dividing the data into clusters requires breaking $k-1$ edges. Enumerating all possible variants, the number of which equals the number of combinations from $(m-1)$ to $(k-1)$, also increases computational costs and reduces performance.

Considering that, when dividing into clusters with non-intersecting convex hulls, the application of compactness or λ-compactness hypotheses produces the same results [12], this work adopts the hypothesis of compactness of the initial data in the feature space.

The clustering problem is formulated as partitioning a set of data points in the feature space into clusters based on their inherent similarity. In a metric space, similarity is usually defined by distance. Distance can be calculated both between data points and between points and the cluster center. Typically, the coordinates of cluster centers are not known in advance and are determined simultaneously with the division of the data into clusters.

To solve the clustering problem, it is necessary to determine an optimal vector $c = c_{opt}$ that, while satisfying the constraints, provides an extremum of the functional $J(\boldsymbol{c}) = M_x\{Q(\boldsymbol{x},\ \boldsymbol{c})\}$ [10], where $J(\boldsymbol{c})$ is the quality criterion, and $Q(\boldsymbol{x},\ \boldsymbol{c})$ is the functional

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

of the variable vector $\boldsymbol{c} = (c_1, \dots, c_N)$, depending on the vector of random variables $\boldsymbol{x} = (x_1, \dots, x_M)$.

Often, the sum of squared errors criterion is used for clustering [13]:

$$J = \sum_{k=1}^{M} \sum_{x \in X_k} \|\boldsymbol{x} - \boldsymbol{c}_k\|^2,$$

where n_k is the number of elements in cluster k, and $\boldsymbol{c}_k = \frac{1}{n_k} \sum_{x \in X_k} \boldsymbol{x}$ is the mean of the cluster.

Thus, J measures the total squared error resulting from representing the data by k clusters with centers $\boldsymbol{c}_k$. Clustering of this type is known as minimum variance clustering. This criterion is suitable when the data form compact clouds in the feature space, well separated from each other, and when the number of clusters is small (typically two or three). However, for elongated, noisy data with widely separated subgroups of points, this criterion may not yield satisfactory results. When there is a large difference in the number of elements among separated clusters, the larger cluster may be split [13].

In such cases, related minimum variance criteria are recommended, particularly a criterion based on $\bar{s}_k$ – the mean squared distance between points in the *k*-th cluster [13]:

$$J = \frac{1}{2} \sum_{k=1}^{M} n_k \bar{s}_k,$$

where $\bar{s}_k = \frac{1}{n_k{}^2} \sum_{\boldsymbol{x} \in X_k} \sum_{\boldsymbol{x}' \in X_k} \|\boldsymbol{x} - \boldsymbol{x}'\|^2$,

Alternatively, $\bar{s}_k$ may be replaced by the mean, median, or maximum distance between points in the cluster [13]. For further studies, and in the absence of prior information about cluster shapes, a criterion from the minimum variance–related group is adopted [10].

The choice of a clustering method is often problem-oriented. Methods also differ in whether the number of clusters is predetermined. If the number of clusters is not known in advance, it can be determined during algorithm execution based on the distribution of the initial data, the compactness and separability of individual clusters, or by starting with a sufficiently large number of clusters and sequentially combining them according to selected similarity criteria.

At present, clustering methods are generally divided into two groups. The first group includes hierarchical methods of successive partitions, which are based on data from the proximity matrix. Two main strategies exist for initializing the initial partition.

The first strategy is agglomerative: clustering begins by assuming that each point in the feature space forms a separate cluster. In this case, when sequentially dividing n points in the feature space into *k* clusters, the first partition produces n clusters (each containing only one point). The next partition produces $n - 1$ clusters, then $n - 1$, and so on, until a single cluster containing all points is obtained. That is, the first merging level corresponds to n clusters, while the *k*-th level corresponds to a single cluster. At any level, any two points in the feature space will be grouped into the same cluster. If the sequence has the property that, when two points are combined into a cluster at level *k*, they remain together at higher levels, such a sequence is called hierarchical clustering.

The second strategy is divisive: initially, all points in the feature space belong to a single cluster. The partitioning of points is stopped once the desired number of clusters is reached.

In addition to hierarchical clustering strategies, there are other popular clustering methods.

The first of these popular methods can be considered the nearest neighbor method. The minimum (nearest neighbor) algorithm is one of the classical approaches to clustering. Its basic principle lies in successively merging points that have the minimum distance between them, thus gradually forming larger and larger groups.

If the algorithm terminates when the distance between the nearest clusters exceeds a predefined threshold, the procedure is known as the single-linkage algorithm. This variation is widely used because of its simplicity and ability to handle clusters of arbitrary shapes. However, it may also be sensitive to noise and outliers, since even a single anomalous point can connect otherwise distinct clusters.

To improve the likelihood of obtaining a correct partitioning, a modification is introduced in which not just the single closest point is considered, but distances to several of the nearest neighbors. This method is commonly referred to as the k-nearest neighbors algorithm. By analyzing distances to *k* points instead of just one, the algorithm reduces the effect of accidental proximities and increases the stability of the clustering process. Such an approach makes it possible to obtain more reliable clusters, especially in datasets where points are unevenly distributed or where noise is present.

Thus, the minimum (nearest neighbor) algorithm and its modifications form the foundation

for many practical clustering methods, combining conceptual simplicity with flexibility of application.

The next known method can be considered the maximum (farthest neighbor) algorithm. That method is based on the principle that points with the maximum distance between them are assigned to different clusters. The main advantage of this method is its ability to form compact and spherical clusters, ensuring that all points within a cluster remain relatively close to each other. Consequently, it is effective for datasets where groups of points are clearly separated and do not exhibit elongated or irregular shapes.

This approach has limitations. In particular, it may produce unsatisfactory results when dealing with elongated or chain-like clusters, as the method tends to over-segment such structures into smaller groups. Moreover, similar to the nearest-neighbor method, that algorithm is highly sensitive to outliers and noise, which can distort the cluster boundaries and reduce the overall accuracy of partitioning [13].

To address these drawbacks, especially the sensitivity to deviations and noisy data, other algorithms such as the average-linkage method and Ward's method are commonly employed. These approaches balance the influence of individual distances, reducing the impact of extreme values and improving the stability of clustering results.

When implementing the method, calculate

$$d_{avg}(X_i, X_j) = \frac{1}{n_i n_j} \sum_{x \in X_i} \sum_{x' \in X_j} \|x - x'\|,$$

where: x are points from cluster X_i; x' are points from cluster X_j and n_i and n_j are the numbers of points in clusters X_i and X_j, respectively. Advantage: it can be used when distances are defined based on similarity measures such as the angle between two vectors [13].

There is also an algorithm that is based on the calculation

$$d_{mean}(X_i, X_j) = \|m_i - m_j\|,$$

where: m_i and m_j are the means of clusters X_i and X_j, respectively.

Advantage: the simplest measure. Disadvantage: for certain similarity measures (e.g., angle between two vectors), the distance may be difficult or impossible to determine [13]. Other general drawbacks of these methods include computational complexity for large datasets and sensitivity to noise depending on the distance measure [13].

In iterative algorithms, data is divided into several clusters, and then elements are moved between clusters to minimize a objective function. Main drawbacks: sensitivity to initial conditions, sensitivity to noise, and convergence to a local rather than global extremum. To assess sensitivity to the initial point, clustering is repeated with different starting points, or the initial point is selected using the result of hierarchical clustering [13].

Considering the above, both groups of methods are sensitive to noise; therefore, it is advisable to develop techniques that reduce noise sensitivity during clustering.

Minimization of a quality objective function represents an optimization problem. Iterative regular and subgradient optimization methods have been developed for such problems. However, both approaches have certain trade-offs. Regular search methods have high accuracy but low noise immunity and high sensitivity to local extrema and the initial search point. Subgradient methods are more noise-resistant but have higher errors [7].

To reduce the influence of noise in such problems, a wavelet function (WF)–based approach is applied [1], [4], [7]. The objective function for optimization is typically spatially nonuniform, with localized global and local extrema. Wavelet transforms provide an adequate tool for analyzing such functions [1], [4], [7], [14].

However, for instance, the method proposed in [14] based on hyperbolic wavelet functions (HWF) may exhibit insufficient noise resistance for practical clustering tasks. Therefore, in [1], [4], [7], the direction of movement toward the target function extremum was estimated by sequential application of Haar WF and HWF, which may reduce the speed of clustering procedures.

To improve performance by reducing the number of calls to the target function, [8] proposed searching for the extremum using constraints in the form of inequalities [10].

Since the search for methods that reduce the impact of the above problems remains relevant, this work proposes to develop and study a clustering method using wavelet functions based on the above approach [8].

DEVELOPMENT AND JUSTIFICATION OF THE CLUSTERING METHOD

Initial stages of the constrained optimization method

This optimization method is defined by an iterative scheme [2]:

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

$$c[n] = c[n-1] - \\ -\gamma[n]\sum_{m=1}^{s_\alpha}\alpha_m[n]\tilde{\nabla}_{c^\pm}Q(x[n],c[n-1],a[n-m]), \quad (1)$$

where $\sum_{m=1}^{s_\alpha}\alpha_m[n]\tilde{\nabla}_{c^\pm}Q(x[n],c[n--1],a[\text{n-m}])$ is the wavelet transform of the implementation $Q(x,c)$ of $c_i,\ i=1,\dots,\ N$; $Q(x,c)$ is the functional of the vector $c=(c_1,\dots,c_N)$, depending on the vector of random sequences or processes $x=(x_1,\dots,x_M)$; $c[n-1]$ is the coordinate of the minimum; $\alpha_m[n], m=1,\dots,s_a$ are the components of the vector $\alpha[n]$, obtained as a result of the discretization of the wavelet function.

To refine the coordinate of the extremum, it is advisable to use a wavelet transform with spatial-frequency localization [1], [4], [7]. Real wavelets in the form of odd symmetric functions with compact or effective support possess this property. Examples include Haar, Gaussian, split Gaussian basis functions (GBF), bounded sine wavelet, hyperbolic basis functions (HWT), and others. These functions satisfy the necessary conditions (localization, admissibility, oscillation, and boundedness) [17]. However, they differ in frequency-selective properties; for instance, the Gaussian wavelet "blurs" details compared to Haar and hyperbolic wavelets [18]. The extremum can be treated as a highly localized phenomenon in space.

The use of Haar wavelet allows saving multiplication operations. However, due to asymmetry in the objective function, the extremum coordinate may be shifted. For more accurate determination, the hyperbolic wavelet function is recommended [4].

After initializing the method parameters in scheme (1), convolution with Haar's wavelet function is used in a neighborhood determined by the length of its carrier to estimate the search direction. The minimum obtained at this stage is adjusted for asymmetry in the functional. Based on the minimum coordinate obtained after convolution of the functional under study with the Haar's wavelet function, according to the technique described in [4], the new goal (objective) function $Q_1(\cdot)$ is formed that allows this coordinate to be taken into account as a constraint,

$$Q_1\big(x[n],c[n],g(x[n],c_H^*[n-1])\big) = \\ = -ln(\,Q(x[n],c[n]) - Q(\boldsymbol{x}[n],\boldsymbol{c}_H^*[n-1])) - \\ - ln(\,g(x[n],c_H^*[n-1])),$$

where $Q(x[n],c[n])$ is the initial goal function at the search step n; $Q(x[n],c_H^*[n-1])$ is the value of the goal function at the minimum point obtained after convolution with the Haar's wavelet function; $g(x[n],c_H^*[n-1]) = c_H^*[n-1] - x[n]$.

As defined later in this work, the relative error of determining the extremum of the asymmetric function during weighted summation with the Haar wavelet is directly proportional to the coefficient of asymmetry of the objective function in the search area.

The iterative scheme (1), using the gradient method [16] as the basic approach to optimization. At this stage, the weighted sum of values $Q_1(\cdot)$ with the hyperbolic wavelet function $\Psi(j)=1/\alpha x$ is weighted, unlike previous works, not regularized using the lifting scheme [5], but, to increase the speed, is calculated only with the initial scale $\alpha=1$.

The optimum is then searched according to scheme (1), using the gradient method [16] as the basic optimization approach. At this stage, the weighted sum of $Q_1\big(x[n],c[n],g(x[n],c_H^*[n-1])\big)$values with the hyperbolic wavelet function $\Psi(j)=\frac{1}{\alpha x}$, regularized via the lifting scheme [5], is calculated at the initial scale $\alpha=1$. The starting point is the previous stage's minimum obtained with Haar's wavelet.

Improving performance for optimization at the stage with HWT

In the basic WT optimization method, if the minimum coordinate differs from the previous stage by no more than δ, the search ends; otherwise, the scale for the hyperbolic wavelet function is increased by 1 (up to scale = 5). This transition moves from noise-immune Haar optimization to high-accuracy hyperbolic differentiation (if $\alpha\to\infty$, then $\frac{1}{\alpha x}$ tends to the differentiator).

Since multi-stage evaluation of the search direction with HWT is used (five stages in [5], α increasing from 1 to 5), the number of oracle calls to obtain functional values remains significant and depends on the wavelet carrier length N. During Haar processing, the carrier length ensures noise immunity. In the hyperbolic stage, extremum refinement allows reducing oracle calls and adjusting computation time based on the noise level.

The aim of this study is to investigate the optimization method with wavelet functions under second-kind constraints to improve search speed and reduce oracle invocations.

Main stages of the clustering method

In clustering [10], based on the feature vectors $x\in X$, the centers of the sets X_k and their boundaries $J(\boldsymbol{c})$ are determined so that.

$$J(c) = E\{Q(x,\ c_{1,\ \dots,}\ c_M)\}, \quad (2)$$

where $Q(\boldsymbol{x},\ c_1, \dots,\ c_M) = \sum_{k=1}^{M} \varepsilon_k(\boldsymbol{x},\ c_1, \dots,\ c_M)\, F_k(\boldsymbol{x},\ c_1, \dots,\ c_M)$ is the implementation of the quality functional; $F_k(\mathrm{x},\ c_1, \dots,\ c_M)$ is the distance function of elements x of the set X from the "centers" c_k of the subsets X_k (clusters); $\varepsilon_k(\cdot)$ are characteristic functions [19].

$$\varepsilon_k(\mathrm{x},\ c_1, \dots,\ c_M) = \begin{cases} 1, & \text{if } \mathrm{x} \in X_k, \\ 0, & \text{if } \mathrm{x} \notin X_k. \end{cases} \quad (3)$$

Since the optimality criterion is given in an implicit form, only implementations of the quality functional *Q(x, c)* are known. Therefore, $\nabla_c J(c)$, the gradient of the functional, is unknown and can only be estimated using the gradient of the implementation $\nabla_c Q(x,c) = \left(\frac{\partial Q(x,c)}{\partial c_1}, \cdots, \frac{\partial Q(x,c)}{\partial c_M}\right)$. Because in noisy conditions or in the case of discontinuities it is often impossible to calculate, search methods for clustering are used.

For example, for the number of clusters *M=2*, the search clustering algorithm for determining the values of cluster centers c_1^*, and c_2^* is:

$$\left.\begin{aligned} c_1[n] &= c_1[n-1] - \gamma_1[n]\tilde{\nabla}_{c1+}Q(\boldsymbol{x}[n],\ c_1[n-1],\ c_2[n-1]) \\ c_2[n] &= c_2[n-1] - \gamma_2[n]\tilde{\nabla}_{c1+}Q(\boldsymbol{x}[n],\ c_1[n-1],\ c_2[n-1]) \end{aligned}\right\}, \quad (4)$$

where $\gamma_k[n]$ (for k = 1,2) is the step size; n is the iteration number; $\tilde{\nabla}_{c1+}Q(x[n],\ c_1[n-1],\ c_2[n-1])$ and $\tilde{\nabla}_{c2+}Q(x[n],\ c_1[n-1],\ c_2[n-1])$ are estimates of the implementation gradient [10].

Regular iterative search methods [20], [21] are based on using gradient estimation. An approximate gradient estimate in search methods is obtained by the difference method [10, 15]:

$$\nabla_{c\pm}J(c,a) = \frac{J_+(c,a) - J_-(c,a)}{2a}, \quad (5)$$

where $J_+(c,a) = (J(c + ae_1), \dots, J(c + ae_N))$, $J_-(c,a) = (J(c - ae_1), \dots, J(c - ae_N))$ are the values of the functional at modified vectors *c*; *a* is scalar; e_v are basis vectors $e_1 = (1,0, \dots,0)$; $e_2 = (0,1, \dots,0)$; ...; $e_N = (0,0, \dots,1)$.

The main disadvantages of regular iterative search methods are sensitivity to local extrema and the initial search point, low convergence rate, and low noise immunity due to the low robustness of gradient estimation by the difference method [10], [15].

As the basic algorithm for estimating the extremum coordinates, the method [16], [24] was used. The initial data for the algorithm are: the initial minimum value of the function, initial step size γ = *1*, the coefficient β = *0.5* defining step size change near the minimum, the accuracy of gradient estimation ε, and the number of iterations *i*.

The procedure for computing the minimum during clustering includes: selection of the initial coordinate of the optimum, calculation of the gradient estimate using wavelet functions;

- if the gradient estimate is less than ε – stop;
- calculation of the step size: the initial step γ = *1*; compute the auxiliary function increment Δ [16]; if Δ<*0* – $\gamma[n] = \gamma$ and proceed to the next stage, otherwise $\gamma[n] = \gamma$ and return to the previous stage;
- calculation of the optimum coordinate at iteration *i*;
- *i = i + 1* and return to the initial stage.

According to the iterative scheme (1), at the first stage, to determine the gradient estimate, a weighted summation of the values of the function to be minimized using Haar wavelets is applied [25]. This allows the search to move to the extremum region with an error determined by the asymmetry of the objective function in this region. Studies of a synthesized test function showed that the relative error of determining the extremum of an asymmetric function with weighted summation using [22], [23] Haar wavelets is directly proportional to the asymmetry coefficient of the objective function in the search area.

Due to this property, at the second stage of clustering, to find the coordinate of the cluster center, the procedure of weighted summation of the function to be minimized with a hyperbolic function $\Psi(i) = \frac{1}{\alpha x}$ at scale α = *1* is used.

The search includes: weighting the function to be minimized $J[x]$ with the function *Ψ(i)*:

$$HWT(x) = J[x] * \Psi(i), \quad (6)$$

where * is the weighted summation operation; determination of the cluster center coordinate using the hyperbolic wavelet function according to the scheme:

$$x[k+1] = x[k] + \gamma[k]HWT(x[k]), \quad (7)$$

where *HWT(x[k])* is the value of the weighted sum with the wavelet function at *x[k], γ[k]* is the step.

If the cluster center coordinate found at this stage differs from the coordinate found at the previous stage by no more than δ, the search process ends. Here δ is the specified search accuracy of the cluster center coordinate. After performing the conditions for determining constraint (2), the search for the extremum coordinate using the hyperbolic wavelet function at scale α = *1* is performed. If the stopping condition for the cluster center coordinate

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

at $\alpha = 1$ is not reached, the estimate is performed according to scheme (4), after which the search ends.

During such a search, a sequential transition from the search for the cluster center coordinate using Haar wavelets, which provide high noise immunity [26], to a search using a differentiator, which provides maximum accuracy, is carried out.

The clustering method using wavelet functions with parameters: initial values of cluster centers $c_1[1]$, $c_2[1]$; scale of the hyperbolic wavelet function $\alpha = 1$; optimum search accuracy; length of the wavelet function support N; maximum number of iterations; is as follows:

- initialize method parameters;
- for each of i elements of the weighted sum with the wavelet function, determine the values of characteristic functions $\varepsilon_l(x, c_1, c_2)$, $l = 1,2$ (3). For this, according to [10], the pairs of values $c_1[n-1]$, $c_2[n-1]$; $c_1[n-1] \pm ie_1 a[n]$, $c_2[n-1]$, $c_1[n-1]$, $c_2[n-1] \pm ie_2 a[n]$ $i = \overline{1,N}$ for a given $x[n]$ are substituted into $f(x, c_1, c_2) = \|x[n]\text{-}c_1\|^2 - \|x[n]\text{-}c_2\|^2$ 2. The function $f(x, c_1, c_2)$ equals zero at the boundary and has different signs in different regions. Therefore, if $f(x, c_1, c_2) < 0$, $\varepsilon_1 = 1$, $\varepsilon_2 = 0$; if positive, $\varepsilon_1 = 0$, $\varepsilon_2 = 1$ [10];
- calculate the value of the Haar function and, if necessary, the hyperbolic function $\Psi(i)$ at $\alpha = 1$, and $i = \overline{1,N}$ determine the weighted sum of the function to be minimized (objective function) with this function $\Psi(i)$;
- determine the approximation of the cluster center value according to scheme (4);
- check the above-mentioned condition of cluster center accuracy (if not reached – transition from weighted summation with Haar wavelet to weighted summation with hyperbolic function, and further according to scheme (5) by discrete differentiation; if reached – stop).

To study the increase in performance of the developed method during clustering, a set of unnamed data was formed, consisting of two groups of 15 points in a two-dimensional feature space (Fig. 1).

As seen from Fig. 1, the clusters are linearly separable in the feature space.

When data clustering using the developed method with second-kind constraints, compared to the basic clustering method, the gain in computation time was more than 1.14 times. During the search for the optimum, the discretization step $dx = 6$ and the length of the Haar wavelet function support when searching for the optimum with the hyperbolic wavelet function $dx = 1$ were adopted.

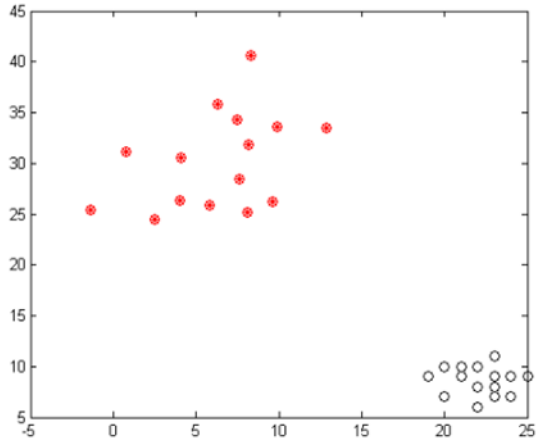


Fig. 1. **Data in clustering**

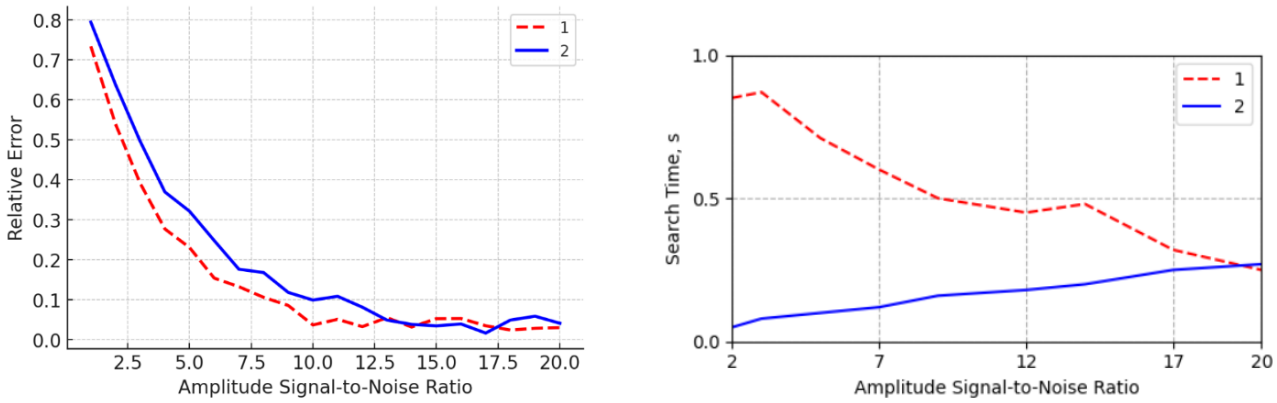
***Source*: compiled by the authors**

Furthermore, the dependence of the relative error of the minimum search on amplitude was investigated for both optimization approaches using wavelet functions (the basic and the constrained versions). The results of this analysis are presented in Fig. 2a. A performance evaluation in terms of computational time (in seconds) for both approaches is given in Fig. 2b, where curves 1 and 2 correspond to the basic and constrained methods, respectively.

The analysis revealed that for signal-to-noise ratios ranging from 2 to 14, the time required to determine the optimum decreased significantly – by a factor of 7 to 1.5 compared with the basic optimization method. At the same time, this improvement in speed was accompanied by a moderate increase in the relative error of minimum determination for the De Jong test function, which rose from 5 % to 15 %. Such a trade-off between computational efficiency and accuracy is common in optimization problems; however, the results suggest that the constrained method provides a reasonable balance, making it suitable for practical applications where reduced processing time is of high priority.

CONCLUSIONS

A clustering method based on wavelet functions has been successfully developed and thoroughly investigated. During the course of this work, a complete methodology and algorithm for implementing the proposed method were designed, ensuring a systematic approach for practical applications. In addition, extensive experimental studies were carried out to assess the method's noise immunity, accuracy, and computational efficiency under various conditions.



a b

Fig. 2. **Results of evaluating the properties of the optimization method with wavelet functions using iterative constraint estimation and the basic method:**
a – error of minimum search depending on the signal-to-noise ratio by amplitude;
b – time of minimum search in seconds depending on the signal-to-noise ratio by amplitude.
***Source*: compiled by the authors**

It has been established that, for a signal-to-noise ratio (SNR) in amplitude ranging from 2 to 14, the optimization search time is significantly reduced, ranging from a factor of 7 at low SNR values to 1.5 at higher SNR values, compared to the basic optimization method. At the same time, the relative error of minimum estimation for the De Jong test function increases moderately from 5% to 15%, demonstrating that the proposed method provides a favorable trade-off between speed and accuracy, particularly in noisy environments. This finding confirms the method's suitability for tasks where rapid clustering is critical, even when data contain substantial noise.

For synthesized datasets, clustering using the developed method with the consideration of second-type constraints showed a notable gain in calculation time, exceeding 1.14 compared to the basic clustering approach. This improvement indicates that the introduction of constraints via wavelet processing effectively reduces redundant computational steps and accelerates convergence, making the method more practical for real-time or large-scale applications.

The method was also validated in a practical clustering problem. The proposed method was applied to a batch of resistors intended for mission-critical equipment [27]. Predictive parameters included the noise level and the expected variation in resistance within groups. Using the first control data collected after 24 hours of operation under load, the proposed method successfully clustered the resistors into two distinct groups: the first cluster included groups 1–8, and the second cluster included group 9, based on the measured noise level. Calculated failure rates were determined for both clusters and the overall batch, highlighting the practical applicability of the method. Compared to the basic clustering method using wavelet functions, the procedure's speed was increased by 8%, confirming the advantage of incorporating constraints for computational speed improvement.

These results collectively demonstrate that the proposed clustering method is effective and reliable, providing both computational speed and robust performance under challenging conditions. Consequently, the developed method can be confidently recommended for a wide range of practically significant classification and clustering tasks, particularly those involving high noise levels, asymmetric objective functions, and small sample sizes, where conventional methods may fail or become computationally prohibitive.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

REFERENCES

1. Hemanth, D. J., Gupta, D. & Balas, V. E. “Intelligent data analysis for biomedical applications”. *Academic Press*. 2019, https://www.sciencedirect.com/book/9780128155530/intelligent-data-analysis-for-biomedical-applications. DOI: https://doi.org/10.1016/C2017-0-03676-5.

2. Kharchenko, V., Fesenko, H. & Illiashenko, O. “Quality models for artificial intelligence systems: Characteristic-based approach, development and application”. *Sensors*. 2022; 22: 4865. DOI: https://doi.org/10.3390/s22134865.

3. Illiashenko, O., Kharchenko, V., Babeshko, I., Fesenko, H. & Di Giandomenico, F. “Security-informed safety analysis of autonomous transport systems considering AI-powered cyberattacks and protection”. *Entropy*. 2023; 25: 1123. DOI: https://doi.org/10.3390/e25081123.

4. Huan, W., Shcherbakova, G., Sachenko, A., Yan, L., Volkova, N., Rusyn, B. & Molga, A. “Haar wavelet-based classification method for visual information processing systems”. *Applied Sciences*. 2023; 13 (9): 5515. DOI: https://doi.org/10.3390/app13095515.

5. Zheng, Y., Shcherbakova, G., Rusyn, B., Sachenko, A., Volkova, N., Kliushnikov, I. & Antoshchuk, S. “Wavelet transform cluster analysis of UAV images for sustainable development of smart regions due to inspecting transport infrastructure.” *Sustainability*. 2025; 17: 927. DOI: https://doi.org/10.3390/su17030927.

6. Fesenko, H., Kharchenko, V., Sachenko, A., Hiromoto, R. & Kochan, V. “Dependable an internet of drone-based multi-version post-severe accident monitoring system: structures and reliability.” *IoT for Human and Industry: Modeling, Architecting, Implementation.* 2018. p. 197–218. DOI: https://doi.org/10.1201/9781003337843.

7. Shcherbakova, G., Krylov, V., Qinqi, W., Rusyn, B. & Sachenko, A. “Optimization methods on the wavelet transformation base for technical diagnostic information systems.” *11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Cracow, Poland. 2021; 2: 767–773. DOI: https://doi.org/10.1109/IDAACS53288.2021.9660927.

8. Shcherbakova, G., Antoshchuk, S., Koshutina, D., Sakhno K. & Kondratiev, S. “Adaptive clustering for distribution parameter estimation in technical diagnostics”. In: *Dovgyi, S., Siemens, E., Globa, L., Kopiika, O., Stryzhak, O. (eds) Applied Innovations in Information and Communication Technology. ICAIIT. Lecture Notes in Networks and Systems. Springer, Cham*. 2024; 1338. DOI: https://doi.org/10.1007/978-3-031-89296-7_19.

9. Jianyun, L. & Shiguo, L. “Wavelet clustering based on cluster signal”. *19th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. Chengdu, China. 2022. p. 1–4. DOI: https://doi.org/10.1109/ICCWAMTIP56608.2022.10016578.

10. Tsypkin, Y. Z. “Adaptation and learning in the automatic systems”. *New York and London*: *Academic Press, Inc.*

11. Radovanovic, A., Li, J., Milanovic, J. V., Milosavljevic, N. & Storchi, R. “Application of agglomerative hierarchical clustering for clustering of time series data”. *IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. 2020. p. 640–644. DOI: https://doi.org/10.1109/ISGT-Europe47291.2020.9248759.

12. Zagoruiko, N. G. “Problems in constructing an empirical theory of data mining.” In: *Wolff, K. E., Palchunov, D. E., Zagoruiko, N. G. & Andelfinger, U. (eds) Knowledge Processing and Data Analysis.* Lecture Notes in Computer Science, Springer, Berlin, Heidelberg. 2011; 6581. DOI: https://doi.org/10.1007/978-3-642-22140-8_16.

13. Bhuyan, M. K. “Computer vision and image processing: Fundamentals and applications”. *Boca Raton: CRC Press/Taylor & Francis Group*. 2019. DOI: https://doi.org/10.1201/9781351248396.

14. Krylov, V. N. & Volkova, N. P. “Vector-difference texture segmentation method in technical and medical express diagnostic systems”. *Herald of Advanced Information Technology*. 2020; 3 (4): 226–239. DOI: https://doi.org/ 10.15276/hait.04.2020.2.

15. Yuan, Q., Yang, Z., & Xiao, Y. “Solving the newsvendor problem using stochastic approximation: A Kiefer-Wolfowitz Algorithm Approach”. *American Journal of Applied Mathematics and Statistics*. 2024; 12 (2): 24–27. DOI: https://doi.org/10.12691/ajams-12-2-1.

16. Polak, E. “Computational Methods in Optimization: A Unified Approach”. *New York, NY: Academic Press*. 1971.

17. Martinez-Ríos, E. A., Bustamante-Bello, R., Navarro-Tuch S. & Perez-Meana, H. “Applications of the Generalized Morse Wavelets: A Review”. *IEEE Access*. 2023; 11: 667–688. DOI: https://doi.org/10.1109/ACCESS.2022.3232729.

18. Prashar, N., Sood, M. & Jain, S. “Design and implementation of a robust noise removal system in ECG signals using dual-tree complex wavelet transform.” *Biomedical Signal Processing and Control*. 2021; 63. DOI: https://doi.org/10.1016/j.bspc.2020.102212.

19. Singh, D., Singh, A., Singh, A. P., Garg, R., Prashar, D. , Khan, A. A. & Yimam, F. A. “A smooth penalty function algorithm for computing nonlinear inequality constraint optimization problem”. *Mathematical Problems in Engineering*. 2024. DOI: https://doi.org/10.1155/2024/1224408

20. Emiola, I. & Adem, R. “Comparison of minimization methods for rosenbrock functions.” *arXiv*. 2021. DOI: https://arxiv.org/abs/2101.10546.

21. Rao, S. S. “Engineering Optimization Theory and Practice”. *John Wiley & Sons, Inc.* 2020. DOI: https://doi.org/10.1002/9781119454816.

22. Stoer, J. “On the relation between quadratic termination and convergence properties of minimization algorithms”. *Numer. Math*. 1977; 28: 343–366. DOI: https://doi.org/10.1007/BF01389973.

23. Berahas, A. S., Cao, L., Choromanski, K. & Scheinberg, K. “A theoretical and empirical comparison of gradient approximations in derivative-free optimization.” *ISE Technical Report 19T-004, Lehigh University*. 2019.

24. Xi, M., Sun, W., & Chen, J., “Survey of derivative-free optimization”. 2020; 10 (4): 537–555. DOI: https://doi.org/10.3934/naco.2020050.

25. Liu, W. Li, S. Chen, M. Fang, Y. Cha L. & Wang, Z. “Fault diagnosis for attitude sensors based on analytical redundancy and wavelet transform”. *Chinese Automation Congress (CAC)*. Shanghai, China. 2020. p. 6471–6476. DOI: https://doi.org/ 10.1109/CAC51589.2020.9327370.

26. Kumar, M., Pachori, R. B. & Acharya, U. R. “Automated diagnosis of myocardial infarction ECG signals using sample entropy in flexible analytic wavelet transform framework.” *Entropy*. 2017; 19 (9): 488. DOI: https://doi.org/10.3390/e19090488.

27. Shcherbakova, G., Antoshchuk, S., Koshutina, D. & Sakhno, K. “Adaptive clustering for distribution parameter estimation in technical diagnostics”. *Proceedings of International Conference on Applied Innovation in IT*.. Koethen, Germany. 2024; 12 (1): 123–127. DOI: https://doi.org/10.25673/115650.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article
Author Galina Y. Shcherbakova is a member of the Editorial Board of this journal. This role had no influence on the peer review process or editorial decision regarding this manuscript

Received 22.07.2025
Received after revision 04.09.2025
Accepted 18.09.2025

DOI: https://doi.org/10.15276/hait.08.2025.17
УДК 519.6:004.93

Підвищення ефективності кластеризації у вейвлет-домені із застосуванням нерівнісних обмежень

Щербакова Галина Юріївна[1)]
Д-р техн. наук, професор, кафедри Інформаційних систем
ORCID: https://orcid.org/0000-0003-0475-3854; galina.sherbakova@op.edu.ua. Scopus Author ID: 27868185600
Кошутіна Дар’я Валеріївна[1)]
PhD студент, кафедри Інформаційних систем
ORCID: https://orcid.org/0009-0004-1326-8775; d.v.koshutina@op.edu.ua. Scopus Author ID: 58289385400
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, Україна, 65044

АНОТАЦІЯ

Методи кластеризації, засновані на оцінюванні градієнта, є поширеними в автоматизованих системах керування та діагностики, де потрібна надійна обробка даних за наявності шуму та мультимодальності. Проте їх використання обмежується низькою стійкістю та високими обчислювальними витратами. Хвильлетні підходи є актуальними, оскільки вони підвищують завадостійкість та покращують ефективність. Метою цієї роботи є розробка та дослідження методу кластеризації, що використовує хвильлетні функції для введення обмежень і забезпечує стабільну роботу в умовах шуму. Дослідження включало аналіз існуючих підходів, розробку хвильлетного методу з обмеженнями другого порядку нерівнянного типу, створення алгоритму його реалізації та експериментальну оцінку. Запропонований метод ґрунтується на хвильлетних перетвореннях із гіперболічними функціями, що дозволяють зменшити кількість викликів оракула для оцінки цільової функції, скоротити обчислювальні етапи та прискорити збіжність у задачах класифікації та кластеризації. Експерименти показали, що метод скорочує час пошуку оптимуму приблизно від півтора до семи разів за різних відношень сигнал/шум при помірному зростанні похибки на п’ять–п’ятнадцять відсотків для тестової функції Де Йонга. На синтетичних наборах даних виграш у часі перевищив одну цілу одну десяту порівняно з базовим методом. У практичному випадку оцінювання надійності резисторів для критичного обладнання ефективність підвищилася майже на вісім відсотків. У підсумку, новизна полягає в методі кластеризації з обмеженнями-нерівностями, що визначаються вейвлет-обробкою. Це може збільшити обчислювальну швидкість в умовах високого рівня шуму, асиметричних цільових функцій та малих вибірок даних.

Ключові слова: кластеризація; вейвлет-перетворення; обмеження другого порядку; оптимізація; виклики оракула; шумостійкість; обчислювальна ефективність

ABOUT THE AUTHORS



Galina Y. Shcherbakova - Doctor of Technical Sciences, Professor, Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0003-0475-3854; galina.sherbakova@op.edu.ua. Scopus Author ID: 27868185600
***Research field*:** data science; machine learning; computer vision; IoT; neural networks; intellectual data analysis

Щербакова Галина Юріївна - доктор технічних наук, професор кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна



Daria V. Koshutina - PhD Student, Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine.
ORCID: https://orcid.org/0009-0004-1326-8775; d.v.koshutina@op.edu.ua; Scopus Author ID: 58289385400
***Research field*:** Data science; machine learning; computer vision; IoT; neural networks

Кошутіна Дар’я Валеріївна - PhD студент, кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, Україна, 65044