Strong Separation of Two Convex polytopes in Machine Learning*

Irina Sokolinskaya, Galina Yadryshnikova
School of Electrical Engineering and Computer Science
South Ural State University (national research university)
Chelyabinsk, Russian Federation
Irina.Sokolinskaya@susu.ru, Galina.Yadryshnikova@inbox.ru

***Abstract*—The problem of strong separation of two convex disjoint polytopes is considered. This problem is important for many industrial applications using machine learning, image analysis and pattern recognition. An iterative algorithm based on the Fejerian mappings is proposed. This algorithm is based on the Eremin's method. An important property of this method is the robustness of the computational process in presence dynamic data changes. The two types of Fejerian mappings are investigated: single-valued Fejerian mapping and multi-valued Fejerian mapping. To study the behavior of the proposed algorithm, the computational experiments with scalable *Model-n* problems and *Random* problems were performed. The conducted experiments confirmed the effectiveness of the proposed approach.**

Keywords— problem of strong separating, Fejerian mappings, iterative process, pattern recognition

Industry 4.0 is the current trend of automation and data exchange in manufacturing technologies. It includes cyber-physical systems, the Internet of things, cloud computing and cognitive computing. Cognitive computing describes technology platforms that are based on the scientific disciplines of artificial intelligence and signal processing. These platforms encompass machine learning, reasoning, natural language processing, speech recognition and vision (object recognition), human–computer interaction. The crucial areas of the Industry 4.0 are artificial intelligence, machine learning [1] and the large-scale linear programming [2] problems, such as industrial optimization [3]. One of the most important problems in machine learning is the classification. Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. One of the promising classification methods is the strong separation method. Another important application area of the strong separation method is the pattern recognition including the following problems: detecting oil/gas reservoirs in sand/shale sediments [4], signal processing [5], 3D face recognition [6], classifying visual motion patterns [7], handwritten Japanese character recognition [8], recognition of 1-D barcodes [9] and others.

The problem of strong separation can be solved by an iterative process using the projecting operation. However, in practice, the application of this method is limited to the fact that it is not always possible to construct the correct equation for calculating the projection of a point onto a convex set [10]. Therefore, it is expedient to replace the projecting operation by a sequence of Fejerian mappings [11]. This method was proposed by Eremin in [12].

Fejer's methods are a class of iterative projection-type methods used to solve systems of linear inequalities and linear programming problems. These methods allow efficient parallelization and therefore can be used for solving large-scale systems of linear inequalities on multiprocessor systems with distributed memory [13]. The construction of the corresponding Fejer operators is based on a superposition of elementary projections, namely, projections onto half-spaces. The projection onto a half-space given by a linear inequality is the basis of the procedure for generating a sequence $\{x_k\}$ that solves the problem in the limit: whether it is simply a system of linear inequalities or a linear programming problem. This class of methods is interesting from different points of view, in particular, from the point of view of their use for non-stationary (evolutionary) modeling [14]. Examples of such non-stationary problems are, for example, the problem of a securities portfolio [15] and modeling for asset-liability management [16], the problem of spam filtering [17], [18] and the problem of classification in meteorology [19]. To overcome the problem of non-stationarity of input data, the Pursuit algorithm for solving non-stationary linear programming problems on cluster computing systems was proposed in [20]. The Pursuit algorithm uses Fejer maps to build a pseudo-projection onto the convex bounded set. The pseudo-projection operator is like the projection, but, in contrast to the last, the pseudo-projection is stable for dynamic change of input data. In the paper [21], the author investigated the efficiency of using multi-core processors Intel Xeon Phi to calculate the pseudo-projections.

Fejer introduced the notion of convergence of the sequence $x_k, k = 0,1,2,\ldots$ with respect to the set $M \subset \mathrm{R}^n$: $\|x_{k+1} - y\| < \|x_k - y\|$ where k and $y \in M$.

---

* The reported study was partially supported by RFBR according to the research project № 17-07-00352 a, the Government of the Russian Federation according to Act 211 (contract № 02.A03.21.0011) and the Ministry of Science and Higher Education of the Russian Federation (government order 2.7905.2017/8.9).

In this paper, an iterative algorithm $\mathfrak{F}$ is constructed and investigated to solve the problem of strong separation. This algorithm is based on the Eremin's method. The paper describes the theoretical and practical aspects of the algorithm $\mathfrak{F}$. The rest of the paper is organized as follows. Section I gives a formal statement of a problem of strong separation and presents the definitions of Fejer process and the pseudo-projection onto a polyhedron. Section II describes a method for solving the problem of strong separation using an iterative algorithm based on Fejer mappings. Section III is dedicated to the tasks used to test the effectiveness of the proposed approaches and the results of computational experiments are discussed. In conclusion, the obtained results are summarized and directions for further research are outlined.

I. Problem Statement

Let us formulate the problem of strong separation. Let us consider two convex disjoint polytopes $M \subset \mathrm{R}^n$ and $N \subset \mathrm{R}^n$ given by systems of linear inequalities:

$$M = \{x \mid Ax \le b\} \ne \emptyset,$$
$$N = \{x \mid Bx \le d\} \ne \emptyset, \tag{1}$$
$$M \cap N = \emptyset.$$

The problem of strong separation is the problem of finding a layer of the greatest thickness separating M and N. Strong separation is the problem of finding the distance between M and N, which is equivalent to

$$\rho(M,N) = \min\{\|x-y\| \mid x \in M, y \in N\}. \tag{2}$$

where $\|x-y\|$ is the Euclidean distance on $\mathbb{R}^n$.

Let us draw tangent hyperplanes in terms of $x \in M$ and $y \in N$ for which condition (2) is satisfied, so that these hyperplanes are parallel. Then a perpendicular dropped from one hyperplane to another will determine the layer of greatest thickness that separates the polytopes *M* and *N*.

The problem of strong separation can be solved using an iterative algorithm based on Fejerian mappings. Let us give a definition of the Fejerian mapping.

Let we be given $\phi \in \{\mathrm{R}^n \to \mathrm{R}^n\}$. The map ϕ is said to be *M-Fejerian* if $\phi(y) = y$, $\forall y \in M$; $\|\phi(x) - y\| < \|x - y\|$, $\forall y \in M$, $\forall x \notin M$.

A sequence $\{x_k\} \subset \mathrm{R}^n$ is said to be *M-Fejerian* if $\|x_{k+1} - y\| < \|x_k - y\|$, $\forall k$, $\forall y \in M$.

A multi-valued (point-set) mapping $\phi \in \{\mathrm{R}^n \to 2^{\mathrm{R}^n}\}$ is said to be *M-Fejerian* if $\phi(y) = y$, $\forall y \in M$; $\|z - y\| < \|x - y\|$, $\forall y \in M$, $\forall x \notin M$, $\forall z \in \phi(x)$.

By definition, put

$$\varphi^s(x) = \underbrace{\varphi \ldots \varphi(x)}_{s}.$$

Fejerian process generated by the map φ for an arbitrary initial approximation $x_0 \in \mathbb{R}^n$ is the sequence $\{\varphi^s(x_0)\}_{s=0}^{+\infty}$.

Let us consider two types of Fejerian mappings given in [10]: single-valued and multi-valued Fejerian mappings.

Type 1 (single-valued Fejerian mapping). Let we be given a finite system of linear inequality in the vector space $\mathbb{R}^n$:

$$Ax \le b:\ l_j(x) = (a_j, x) - b_j \le 0,\ j = 1, \ldots, m, \tag{3}$$

where $a_j \ne 0$ for every j, (a_j, x) is the Euclidean dot product of a_j and x in $\mathbb{R}^n$, $b_j \in \mathbb{R}^n$. Let us define $l_j^+(x)$ as follows:

$$l_j^+(x) = \max\{l_j(x), 0\},\ j = 1, \ldots, m. \tag{4}$$

Then a *Fejerian mapping of the first type* can be defined as the follows:

$$\phi(x) = x - \sum_{j=1}^{m} \alpha_j \lambda_j \frac{l_j^+(x)}{\|a_j\|^2} a_j \tag{5}$$

for any system of positive coefficients $\{\alpha_j > 0\}$, $j = 1, \ldots, m$ such that $\sum_{j=1}^{m} \alpha_j = 1$, and relaxation factor $0 < \lambda_j < 2$. Here $\|\cdot\|$ is the Euclidean norm.

Type 2 (multi-valued Fejerian mapping). Let us construct the Fejer mapping of the second type as follows:

$$\psi(x) = x - \lambda \frac{\max\limits_{(j)} l_j^+(x)}{\|a_{j_x}\|^2} a_{j_x}, \tag{6}$$

here j_x is any of the indices where $\max\limits_{(j)} l_j^+(x)$ is reached, $0 < \lambda < 2$ is the relaxation factor.

Let *M* be a convex closed set given by a system of linear inequalities. Let ϕ be a continuous M-Fejerian mapping. Let us denote $\{x_i\}$ - the sequence generated by the mapping ϕ. Let $x_0 \in \mathrm{R}^n$ be an arbitrary initial approximation. The following fundamental theorem of convergence holds [10].

Theorem 1. The process $\{x_{i+1} \in \phi(x_i)\}$, $\forall i$ generated by any Fejerian mapping ϕ for any initial x_0, converges to a solution x' of the system of convex inequalities: $\forall i\ \{x_{i+1} \in \phi(x_i)\} \to x' \in M$.

Let us consider how Fejerian mappings can be used to solve the problem of strong separation of two convex disjoint polytopes.

II. *The Method of Solving the Problem of Strong Separation*

Let us define the algorithm $\mathfrak{F}$ for the separation of convex polytopes by using Fejerian mappings.

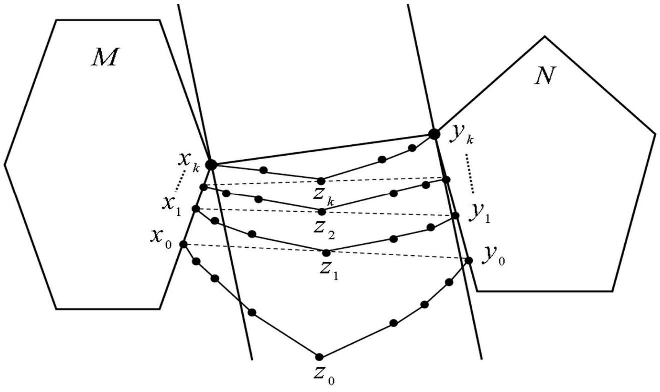

Fig. 1. The iterations of the algorithm $\mathfrak{F}$.

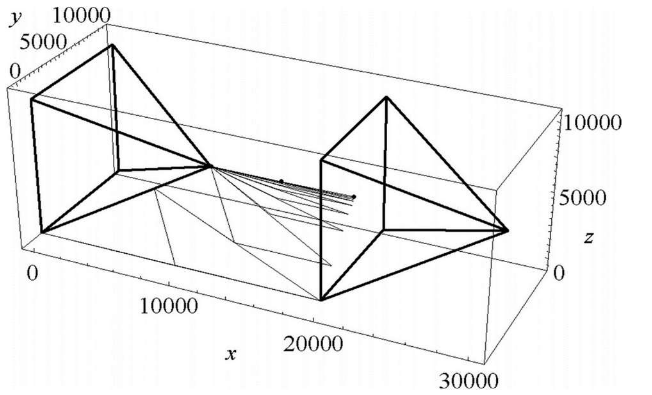

Fig. 2. Iterations of the algorithm $\mathfrak{F}$ for the problem *Model*-3.

Algorithm $\mathfrak{F}$. Let we be given two convex disjoint polytopes $M \subset \mathrm{R}^n$ and $N \subset \mathrm{R}^n$ given by systems of linear inequalities (1). Let ϕ_M and ϕ_N be continuous M and N-Fejerian mappings. Method starts with an arbitrary point z_0 in $\mathbb{R}^n$ as an initial approximation. The algorithm consists of the following steps:

Step 0. $k := 0$.

Step 1. Find the point x_k as the result of the repeated successive application of the mapping ϕ_M to the point z_k:

$$x_k := \lim_{u\to\infty} \phi_M^u\left(z_k\right).$$

Step 2. Find the point y_k as the result of the repeated successive application of the mapping ϕ_N to the point z_k:

$$y_k := \lim_{u\to\infty} \phi_N^u\left(z_k\right).$$

Step 3. Designate:

$$z_{k+1} := \frac{x_k + y_k}{2}.$$

Step 4. $k := k+1$.

Step 5. Go to Step 1.

The iterations of algorithm $\mathfrak{F}$ are shown schematically in Fig. 1.

The use of the algorithm $\mathfrak{F}$ to find the layer of the greatest thickness that separates two convex polytopes *M* and *N* consists of the sequential computation of the points $x_k \in M$ and $y_k \in N$. This iterative process ends when:

$$\max\left\{\left\|x_k - x_{k-1}\right\|, \left\|y_k - y_{k-1}\right\|\right\} < \varepsilon,$$

where $\varepsilon > 0$ is a positive real number being a parameter of the algorithm.

It is known that if the points x_k and y_k are obtained as projections of the points x_{k-1} and y_{k-1} onto the polytopes M and N respectively then for the case of convex polytopes the process converges to the stability loop. In this case, in the limit, we obtain a distance defining the layer of greatest thickness. Eremin suggested that the algorithm F using Fejér's approximations would also converge to the stability loop.

In order to verify the analytical results, we implemented the algorithm $\mathfrak{F}$ for the separation of convex polytopes by using Fejerian mappings in C++ language. Using this program, an experimental study of the behavior of the algorithm $\mathfrak{F}$ for various convex polytopes was made. The convergence of the process to the stability loop was observed in all cases. As a result, we obtained a distance defining the layer of the greatest thickness with a specified accuracy. The results of computational experiments are described in detail in the next section.

III. Numerical Experiments

During the computational experiments the behavior of the algorithm $\mathfrak{F}$ for Fejerian mappings of the first and second types was investigated. Two classes of tasks were used. The first class is the model scalable problem *Model-n*. For all such problems, it is easy to calculate the exact value of the thickness of the maximal separating layer analytically. Therefore, they are suitable for checking the correctness of the algorithm and investigating its scalability.

The second class is the *Random* problems which are randomly generated by a special algorithm. This class of problems allows us to confirm the applicability of the algorithm $\mathfrak{F}$ for arbitrary problems of strong separation.

Let us consider an experiment with a model problem. The *Model-n* problem has the following form (*n* is the dimension of the problem):

Polytope *M*	Polytope *N*
$\begin{cases} x_1 - 2x_2 \le 0 \\ x_1 - 2x_3 \le 0 \\ \dots \\ x_1 - 2x_n \le 0 \\ x_1 + 2x_2 \le 20000 \\ x_1 + 2x_3 \le 20000 \\ \dots \\ x_1 + 2x_n \le 20000 \\ -x_1 \le 0 \\ -x_2 \le 0 \\ \dots \\ -x_n \le 0 \end{cases}$	$\begin{cases} x_1 - 2x_2 \le 20000 \\ x_1 - 2x_3 \le 20000 \\ \dots \\ x_1 - 2x_n \le 20000 \\ x_1 + 2x_2 \le 40000 \\ x_1 + 2x_3 \le 40000 \\ \dots \\ x_1 + 2x_n \le 40000 \\ -x_1 \le -20000 \\ -x_2 \le 0 \\ \dots \\ -x_n \le 0 \end{cases}$

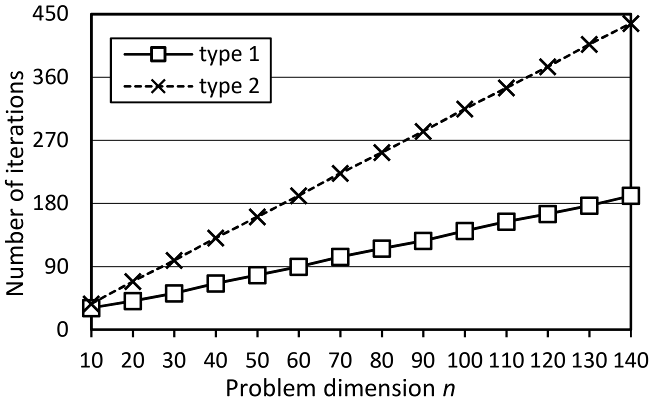

Fig.3. Dependence of the number of iterations on the dimension n for the *Model-n* problem.

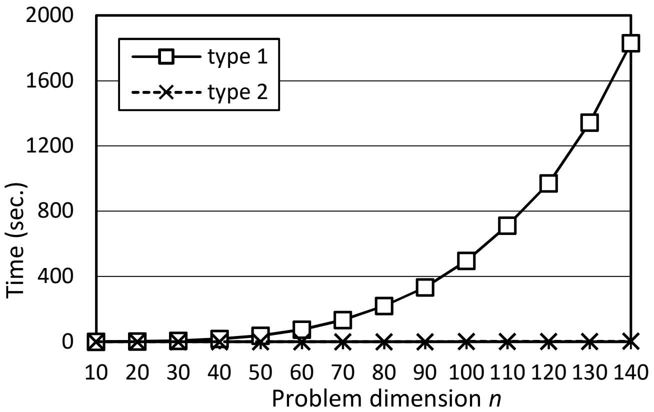

Fig.4. Dependence of the time for solving the *Model-n* problem on the dimension n.

The polytopes M and N given by the *Model-3* ($n = 3$) problem and the iterations of the algorithm $\mathfrak{F}$ leading to the stability loop are shown in Fig. 2. For all *Model-n* problems the thickness of the maximum separating layer is 10000.

A series of computational experiments with *Model-n* problems was performed. The dependence of the number of iterations and the time of solving the problem on the dimension n was investigated. The calculations were performed for the dimensions from 10 to 140.

Fejerian mappings of the 1st and 2nd types were used. The results of the experiments show (see Fig. 3 and Fig. 4) that the number of iterations for both types of Fejerian mappings rises with increasing dimension. In this case, the algorithm using a Fejerian mapping of the 1st type finds the stability loop for a smaller number of iterations. However, an algorithm using the 2nd type wins significantly by the time of solving the problem for all dimensions.

In the second part of the computational experiment, studies were carried out on random problems. To generate problems of the *Random* class, a special program was used. This program randomly constructs two systems of linear inequalities that define two convex disjoint polytopes of any dimension.

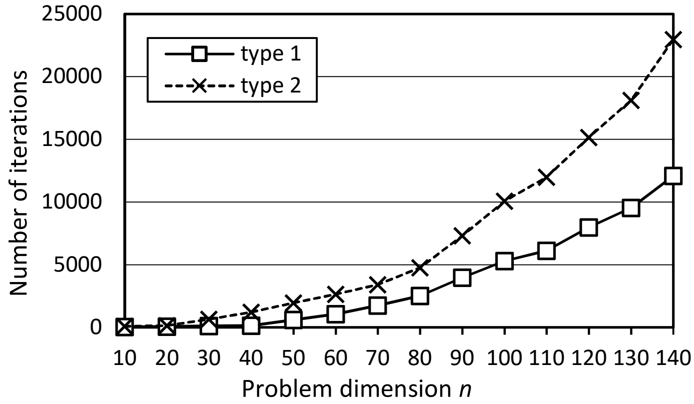

Fig.5. Dependence of the number of iterations on the dimension n for the *Random* problem.

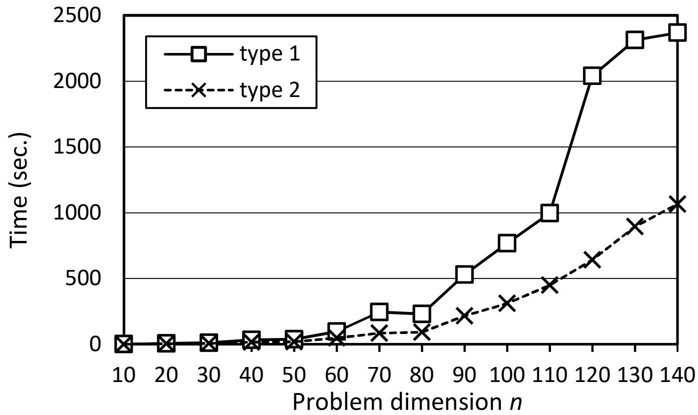

Fig.6. Dependence of the time for solving the *Random* problem on the dimension n

In a series of computational experiments with *Random* problems the dependence of the number of iterations and the time of solving the problem on the dimension n was investigated. The dimension of the problem was ranged from 10 to 140. Fejerian mappings of the 1st and 2nd types were used. During the computational experiment a series of 10 random problems was generated for each dimension. As a result, we used the averaged values of the number of iterations and the time of solving the problem in each series.

The results of the experiments are shown in Fig. 5 and Fig. 6. They show that the behavior of the algorithm $\mathfrak{F}$ on random problems is similar to the behavior on model problems. However, the difference in the rate of convergence of two types of Fejerian mappings for random problems is no longer so significant.

IV. Conclusion

In this paper, the iterative algorithm $\mathfrak{F}$ for solving the separation problem of two convex disjoint polytopes by a layer of greatest thickness was considered. This algorithm is based on the use of Fejerian mappings as an analog of the design operation. To study the behavior of the algorithm $\mathfrak{F}$ the computational experiments with scalable *Model-n* problems and *Random* problems were conducted. The conducted experiments confirmed the effectiveness of the proposed approach. Our future goal is a parallel implementation of the algorithm $\mathfrak{F}$ in C++ language using MPI library, and computational experiments on a cluster computing system using synthetic and real problems.

References

[1] K. Schwab, The fourth industrial revolution. Geneva: World Economic Forum, 2016.

[2] W. Chung, "Applying large-scale linear programming in business analytics," in 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2015, pp. 1860–1864.

[3] H. Tipi, "Solving Super-Size Problems with Optimization," 2010. [Online]. Available: http://nymetro.chapter.informs.org/prac_cor_pubs/06-10 Horia Tipi SolvingLargeScaleXpress.pdf. [Accessed: 21-Apr-2017].

[4] Zheng-He Yao and Li-De Wu, "A pattern recognition approach to detect oil/gas reservoirs in sand/shale sediments," Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems, The Hague, Netherlands, 1992, pp. 462-465. doi: 10.1109/ICPR.1992.201818

[5] V. Neagoe, "Seeking pattern recognition principles for intelligent detection of FSK signals," Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems, The Hague, Netherlands, 1992, pp. 721-724. doi: 10.1109/ICPR.1992.201878

[6] Cheng Zhong, Zhenan Sun and Tieniu Tan, "Learning efficient codes for 3D face recognition," 2008 15th IEEE International Conference on Image Processing, San Diego, CA, 2008, pp. 1928-1931. doi: 10.1109/ICIP.2008.4712158

[7] Ming-Hsuan Yang and N. Ahuja, "Extraction and classification of visual motion patterns for hand gesture recognition," Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231), Santa Barbara, CA, 1998, pp. 892-897. doi: 10.1109/CVPR.1998.698710

[8] B. Chen, B. Zhu and M. Nakagawa, "Effects of Generating a Large Amount of Artificial Patterns for On-line Handwritten Japanese Character Recognition," 2011 International Conference on Document Analysis and Recognition, Beijing, 2011, pp. 663-667. doi: 10.1109/ICDAR.2011.139

[9] S. Wachenfeld, S. Terlunen and X. Jiang, "Robust recognition of 1-D barcodes using camera phones," 2008 19th International Conference on Pattern Recognition, Tampa, FL, 2008, pp. 1-4. doi: 10.1109/ICPR.2008.4761085

[10] I.I. Eremin, The theory of linear optimization. Ekaterinburg: Publishing House of the "Yekaterinburg", 1999. (in Russian)

[11] I.I. Eremin, Fejer's Methods for Problems of Convex and Linear Optimization. Chelyabinsk, Publishing of the South Ural State University, 2009. (in Russian)

[12] I.I. Eremin, Theory of Linear Optimization / Inverse and Ill-Posed Problems. VSP: Utrecht, Boston, Koln, Tokyo, 2002.

[13] I. Sokolinskaya and L. B. Sokolinsky, "On the Solution of Linear Programming Problems in the Age of Big Data," Parallel Computational Technologies. PCT 2017. Communications in Computer and Information Science, vol. 753, pp. 86–100, 2017.

[14] I. Sokolinskaya, "Parallel Method of Pseudoprojection for Linear Inequalities," in Parallel Computational Technologies. PCT 2018. Communications in Computer and Information Science, vol. 910, Cham: Springer, 2018, pp. 216–231.

[15] H. Markowitz, "Portfolio selection," Journal of Finance, vol.7, no. 1, pp. 77-91, 1952

[16] M. S. Sodhi, "LP modeling for asset-liability management: A survey of choices and simplifications," Operations Research, vol. 53, no. 2, pp. 181–196, 2005.

[17] T.S. Guzella, W.M. Caminhas, "A review of machine learning approaches to Spam filtering," Expert Systems with Applications, vol. 36, no. 7, pp. 10206-10222, 2009.

[18] L. Zhang, J. Zhu, T. Yao, "An Evaluation of Statistical Spam Filtering Techniques," Transactions on Asian Language Information Processing, vol. 3, no. 4, pp. 243-269, 2004.

[19] S.Q. Kidder, T.H. Vonder Haar. Satellite meteorology. Academic Press: London, 1995.

[20] I. Sokolinskaya, L. Sokolinsky, "Solving unstable linear programming problems of high dimension on cluster computing systems," 1st Russian Conference on Supercomputing Days 2015. CEUR Workshop Proceedings, vol. 1482, pp. 420-427, 2015.

[21] I. Sokolinskaya and L. Sokolinsky, "Revised Pursuit Algorithm for Solving Non-stationary Linear Programming Problems on Modern Computing Clusters with Manycore Accelerators," Supercomputing. RuSCDays 2016. Communications in Computer and Information Science, vol. 687, pp. 212–223, 2016.