Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(5), 14–24 https://doi.org/10.59139/ws.2024.05.2
Studia metodologiczne / Methodological studies
Zgłoszony/submitted: 7.12.2022, poprawiony/revised: 5.10.2023, zaakceptowany/accepted: 12.02.2024

Log-Quad divergence for Non-negative Matrix Factorization in multi-model prediction[1]

Ryszard Szupiluk,[a] Paweł Rubach[b]

Abstract. The aim of this paper is to present a new Non-negative Matrix Factorization (NMF) algorithm based on Log-Quad divergence, and to demonstrate its application to the separation of latent destructive components contained in prediction results in a multi-model approach. We provide an example of its application to a real economic problem, i.e. forecasting electricity consumption on the basis of information about hourly use of electricity in Poland in the period of 1988–1997. We evaluated and compared this method with other blind signal (source) separation techniques, such as Independent Component Analysis (ICA) and Algorithm for Multiple Unknown Signals Extraction (AMUSE). The results show that the NMF algorithm based on Log-Quad divergence has an interesting ability to improve predictions for small volumes of data.

Keywords: Non-negative Matrix Factorization, NMF, latent components identification, blind source separation, blind signal separation, prediction, ICA, AMUSE

JEL: C02, C50

Dywergencja Log-Quad dla nieujemnej faktoryzacji macierzy w predykcji za pomocą podejścia wielomodelowego

Streszczenie. Celem niniejszego artykułu jest przedstawienie nowego algorytmu nieujemnej faktoryzacji macierzy (Non-negative Matrix Factorization – NMF) opartego na dywergencji Log-Quad i możliwości zastosowania go do separacji ukrytych komponentów destrukcyjnych zawartych w predykcjach w ujęciu wielomodelowym. Jako przykład wykorzystania go w praktyce wybrano prognozowanie zużycia energii elektrycznej na podstawie danych dotyczących godzinowego zużycia energii elektrycznej w Polsce w latach 1988–1997. Proponowany algorytm oceniono i porównano z innymi technikami z obszaru ślepej separacji sygnałów (źródeł), takimi jak analiza składowych niezależnych (Independent Component Analysis – ICA) oraz algorytm

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na konferencji Multivariate Statistical Analysis MSA'2022, która odbyła się w dniach 7–9 listopada 2022 r. w Łodzi. / The article is based on a paper delivered at the Multivariate Statistical Analysis MSA'2022 Conference, held on 7–9 November 2022 in Łódź, Poland.

[a] Szkoła Główna Handlowa w Warszawie, Instytut Informatyki i Gospodarki Cyfrowej, Polska / Warsaw School of Economics, Institute of Information Systems and Digital Economy, Poland.
ORCID: https://orcid.org/0000-0002-4416-7131. E-mail: ryszard.szupiluk@sgh.waw.pl.

[b] Szkoła Główna Handlowa w Warszawie, Instytut Informatyki i Gospodarki Cyfrowej, Polska / Warsaw School of Economics, Institute of Information Systems and Digital Economy, Poland.
ORCID: https://orcid.org/0000-0001-5487-609X. Autor korespondencyjny / Corresponding author, e-mail: pawel.rubach@sgh.waw.pl.

dekorelacji wieloetapowej (Algorithm for Multiple Unknown Signals Extraction – AMUSE). Wyniki pokazują, że algorytm NMF oparty na dywergencji Log-Quad charakteryzuje się ciekawą właściwością poprawiania predykcji w przypadku niedużych wolumenów danych.

Słowa kluczowe: nieujemna faktoryzacja macierzy, NMF, identyfikacja ukrytych komponentów, ślepa separacja źródeł, ślepa separacja sygnałów, predykcja, ICA, AMUSE

1. Introduction

One of the common tasks in data analysis is to find a new interesting or useful representation of data. For this purpose, various transformations can be used. Over the last two decades, the area of multivariate methods developed particularly intensively, especially the methods related to the problem of blind signal separation. In such a problem, the goal is to separate (reproduce, estimate, reconstruct, identify) *a priori* unknown signals mixed in an unknown mixing system. Separation takes place using mixed data only. The application of a problem formulated in such a way may be found in many practical areas, such as: telecommunications, medicine, geology, acoustics, economics and business (Comon & Jutten, 2010). A classic example of blind source separation is the 'cocktail party' problem. In this kind of a party, many individuals are speaking simultaneously, thus generating noise, which we record using microphones. Based on such mixed signals, we want to isolate individual conversations. Classic signal filtering techniques based on the Fourier transform (FT) have limited application, because people speak in approximately the same frequency band (ca. 1 kHz), and there is no frequency diversity necessary to use the Fourier transform. In this type of problems, techniques currently referred to as blind signal separation, such as the Independent Component Analysis (ICA) and the Non-negative Matrix Factorization (NMF), have proven to be effective. They are based on other mathematical criteria.

As a result of intense research, in addition to the classic methods such as Principal Component Analysis (PCA) or factor analysis, a long list of new methods has emerged, such as the above-mentioned ICA and NMF, Smooth Principal Component Analysis (SmPCA), Sparse Component Analysis (SCA) and Independent Factor Analysis (IFA). The whole set of these methods is also referred to as Latent Component Analysis (LCA; Cichocki & Amari, 2002).

Although each of these methods is based on different mathematical criteria with accompanying assumptions specific to the given method, they have certain common features. First of all, they are most often considered in the context of a machine-learning approach, which entails the numerical optimization of the objective function that is the basis of a given method. Consequently, much of the literature on statistical methods, e.g. on the ICA or algebraic methods such as the NMF, concern the context of neural networks and/or machine learning.

Another similarity between the above-mentioned methods is their performance in the area of blind signal separation, namely the fact that they can be applied to the reconstruction of mixed-source signals. It is expected that all these methods, despite being mathematically distinct, lead to similar or even identical results. An example illustrating the whole issue can be the ‘cocktail party’ problem mentioned before, where the goal is to separate a single conversation from all the sounds produced by many people talking simultaneously. A success or a failure in solving such a problem is whether the statement is understandable or not.

At the same time, there are significant differences in the implementation of individual methods within the LCA when they are applied to blind separation. These differences exist both between the methods and within a given method between particular numerical algorithms. One of these differences is the amount of data necessary for a given method to operate effectively.

The aim of this paper is to present a new NMF algorithm based on the Log-Quad divergence (Cichocki et al., 2009), and to demonstrate its application to the separation of latent destructive components contained in prediction results in a multi-model approach. The test was performed on the basis of the forecast of energy consumption based on a dataset of hourly electricity consumption in Poland in the years 1988–1997. The goal was to compare the operation of the Log-Quad NMF-based system with the ICA-based one, as well as showing the differences in the operation of other NMF algorithms, i.e. the Image Space Reconstruction Algorithm (ISRA) developed by Lee and Seung (1999) and the version derived from it by Cichocki and Févotte (Cichocki et al., 2008; Févotte et al., 2009).

In machine learning systems and models, there are usually many arbitrarily selected parameters that affect the efficiency of algorithms. They include, for example, the selection of learning coefficients, nonlinearity forms, or the assumed number of iterations. Since the LCA methods belong to unsupervised learning, we generally cannot control the quality of operation of such algorithms except for easily-interpretable physical signals.

2. Non-negative Matrix Factorization

In its basic form, Non-negative Matrix Factorization can be defined as expressing a matrix **X**, with elements $x_{ik} \geq 0$, as a multiplication of two non-negative matrices **A**, **S** (Berry et al., 2007), i.e.

$$\mathbf{X} \simeq \mathbf{AS}, \tag{1}$$

where the elements of matrix **A** are non-negative: $a_{ij} \geq 0$, which is also denoted as: $\mathbf{A} \geq 0$. An analogous condition is applied to matrix **S**.

To assess the extent to which matrices $\mathbf{X}$, $\mathbf{A}$, $\mathbf{S}$ fulfill equation (1), the following objective function $J = L(\mathbf{X}, \mathbf{AS})$ is introduced, which at the same time is the objective function used to formulate the NMF algorithm. This function may take the classical form of a measure of the distance between $\mathbf{X}$ and $\mathbf{AS}$ expressed by formula $J = ||\mathbf{X} - \mathbf{AS}||_p$, where $||.||_p$ is the p-norm. For $p = 2$, we obtain the Euclidean distance, on the basis of which the ISRA algorithm, considered to be the first solution of the NMF in its current form, was derived by Lee and Seung (1999). This algorithm can be represented in a matrix form as:

$$\mathbf{A} \leftarrow \mathbf{A}.\times\mathbf{X}\mathbf{S}^{\mathrm{T}}./\mathbf{A}\mathbf{S}\mathbf{S}^{\mathrm{T}} \tag{2}$$

and

$$\mathbf{S} \leftarrow \mathbf{S}.\times\mathbf{A}^{\mathrm{T}}\mathbf{X}./\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S}, \tag{3}$$

where $.\times$ and $./$ mean element-wise multiplication (Hadamard product) and element-wise division, respectively.

The ISRA algorithm does not guarantee convergence to a local minimum, but rather to a certain stationary point (which does not have to be a minimum). For this reason, it is currently rarely used in real applications; more often it plays the role of a reference point, a solution against which new algorithms are compared. It also serves as an inspiration for the development of other approaches. The disadvantages of using this algorithm include its slow convergence (if it converges at all) and the fact that once it is set to 0, it stays that way. These limitations necessitated the search for alternative solutions. One of the possibilities is to use the divergence function $D(y||z)$ which became the basis for deriving NMF algorithms such that $J = D(\mathbf{X}||\mathbf{AS})$.

3. Log-Quad divergence

Divergence functions are among the most popular criteria for the derivation of NMF algorithms. Unlike ordinary distance measures, a divergence function does not have to satisfy the triangle inequality, and is usually asymmetric. Historically, the development of divergence measures involved assessing the similarity between distributions. Currently, they are used to assess the similarity (or lack thereof) between non-negative variables, vectors, matrices or functions. Divergences can be defined for both continuous and discrete quantities. Divergence can be accepted or interpreted as a specific measure of distance (quasi-distance; Amari, 1985;

Cichocki et al., 2009; Csiszár, 1978). One of the broader families of divergences are Bregman divergences defined as (Bregman, 1967):

$$D_{\varphi}(\mathbf{X}||\mathbf{AS}) = \sum_{i=1}^{m}\sum_{t=1}^{n}(\varphi(x_{it}) - \varphi([\mathbf{AS}]_{it}) - \varphi'(x_{it})(x_{it} - [\mathbf{AS}]_{it})), \tag{4}$$

where $[\mathbf{AS}]_{it}$ denotes the (i,t)-th element of the $m \times n$ matrix $\mathbf{AS}$, and $\varphi(u)$ is a strictly convex function that has a continuous derivative $\varphi'(u)$.

For non-linearity and $\varphi(u) = u^2 + u\ln(u),\ u > 0$, the Log-Quad algorithm obtains the following divergence (Cichocki et al., 2009):

$$D_{\mathrm{LQ}}(\mathbf{X}||\mathbf{AS}) = \sum_{i=1}^{m}\sum_{t=1}^{n}\left((x_{it} - [\mathbf{AS}]_{it})^2 + x_{it}\ln\frac{x_{it}}{[\mathbf{AS}]_{it}} - x_{it} + [\mathbf{AS}]_{it}\right). \tag{5}$$

Log-Quad divergence is one of the measures of similarity that are relatively rarely discussed or used. It is specific in that it is actually a combination of some popular similarity measures such as squared differences, Kuback-Leibler entropy and ordinary differences. While these are usually standalone measures, here they are combined in one formula. We define the NMF problem for divergence (5) as $\min\limits_{\mathbf{A},\mathbf{S}} D_{\mathrm{LQ}}(\mathbf{X}||\mathbf{AS})$, and use the formula of the alternating multiplicative NMF algorithm for Bregman divergence (Dhillon & Sra, 2005) in the form of:

$$\mathbf{A} \leftarrow \mathbf{A}.\frac{(\varphi''(\mathbf{AS}).\times\mathbf{X})\mathbf{S}^{\mathrm{T}}}{(\varphi''(\mathbf{AS}).\times(\mathbf{AS}))\mathbf{S}^{\mathrm{T}}}, \tag{6}$$

and for matrix $\mathbf{S}$:

$$\mathbf{S} \leftarrow \mathbf{S}.\frac{\mathbf{A}^{\mathrm{T}}(\varphi''(\mathbf{AS}).\times\mathbf{X})}{\mathbf{A}^{\mathrm{T}}(\varphi''(\mathbf{AS}).\times(\mathbf{AS}))}. \tag{7}$$

In the case of assuming the objective function in the form of Log-Quad divergence, the second derivative of the function $\varphi(u)$ takes the form of $\varphi''(u) = 2 + 1/u$, which leads to the following algorithm for estimating matrices $\mathbf{A}$ and $\mathbf{S}$:

$$\mathbf{A} \leftarrow \mathbf{A}.\frac{((2 + 1./(\mathbf{AS})).\times\mathbf{X})\mathbf{S}^{\mathrm{T}}}{((2 + 1./(\mathbf{AS})).\times(\mathbf{AS}))\mathbf{S}^{\mathrm{T}}}, \tag{8}$$

and

$$\mathbf{S} \leftarrow \mathbf{S}.\frac{\mathbf{A}^{\mathrm{T}}((2 + 1./(\mathbf{AS})).\times \mathbf{X})}{\mathbf{A}^{\mathrm{T}}((2 + 1./(\mathbf{AS})).\times (\mathbf{AS}))}. \tag{9}$$

The algorithm obtained in models (8)–(9) is one of many variants possible to obtain on the basis of the objective function in the form of Log-Quad divergence. These forms depend on the adopted method of minimizing the objective function.

4. Multi-model system for improving quality of prediction

The NMF model (1) can be interpreted in terms of a prediction-improvement system. Let us assume that the rows of matrix $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n]^{\mathrm{T}}$ contain the results of forecasts from different models. Another assumption is that they are a linear combination $\mathbf{X} = \mathbf{AS}$ of hidden components responsible for the correct results of the forecast as well as the components responsible for errors. So, we can write that matrix $\mathbf{S} = [\boldsymbol{s}_1, ..., \boldsymbol{s}_n]^{\mathrm{T}} = [\hat{\boldsymbol{s}}_1, ..., \hat{\boldsymbol{s}}_p, \tilde{\boldsymbol{s}}_{p+1}, ..., \tilde{\boldsymbol{s}}_n]^{\mathrm{T}}$. This means it consists of constructive components $\hat{\boldsymbol{s}}_i$ and destructive components $\tilde{\boldsymbol{s}}_j$. By identifying matrices $\mathbf{A}$ and $\mathbf{S}$ and then eliminating the destructive components (i.e. assuming $\tilde{\boldsymbol{s}}_j = \mathbf{0}$), we should obtain improvement in the prediction quality

$$\widehat{\mathbf{X}} = \mathbf{A}[\hat{\boldsymbol{s}}_1, ..., \hat{\boldsymbol{s}}_p, \mathbf{0}_{p+1}, ..., \mathbf{0}_n]^{\mathrm{T}}, \tag{10}$$

where $\widehat{\mathbf{X}}$ denotes a matrix whose rows contain corrected forecasts (Szupiluk et al., 2007).

5. Practical experiment

The verification and demonstration of the new NMF Log-Quad algorithm was performed using hourly energy consumption data for Poland in the period of 1988–1997. The dataset contains observations of hourly electricity consumption and forecasts generated every 24 hours on the basis of previous hourly consumption. There are six independent forecasts that were generated using Multilayer Perceptron neural networks, with 12, 18, 21, 24, 27 and 30 neurons in the hidden layer. Overall, the dataset contains 43,824 predictions and is divided into three arbitrarily divided subsets (Dataset A: 17,544 predictions, Dataset B: 17,544 predictions, and Dataset C: 8,736 predictions).

The goal of the experiment was to improve the forecasts of energy consumption by applying our multi-model prediction improvement schema to a given dataset. According to this schema, the initial predictions are decomposed using the NMF Log-Quad algorithm into six latent components. In the next step, a selected base latent component or a set of components is eliminated (replaced with zeros), and subsequently the latent components are recomposed into six forecasts. The Mean Squared Error (MSE) is computed for initial predictions as well as on the recomposed ones, and then compared.

In order to observe the performance of the algorithm in more detail, the whole dataset was divided into smaller frames with predicted consumption values of various sizes, namely 20, 40, 50, 75, 100, 500, 1,000, 2,000, 3,000, 4,000 and 5,000. Overall, there were 5,355 frames. For each frame, a full combination of different decompositions using different sets of base components were tested – starting from the elimination of a single base component up to the replacement of five out of six base components with zeros. There are 62 such combinations that were evaluated for each frame.

Table 1. Prediction improvements using the NMF Log-Quad algorithm

Frame size	Predictions range (dataset)	Identifier of base component removed	Initial MSE 10^{-3}	Final MSE 10^{-3}	Improvement in %
20	8560–8580 (C)	2, 6	24.32	19.50	19.8
40	12400–12440 (B)	4	14.06	11.33	10.6
50	5400–5450 (A)	5	3.79	3.36	11.2
75	8550–8625 (C)	6	9.51	6.67	29.9

Source: authors' computations based on an energy dataset, performed using Python and matlab implementations of the NMF Log-Quad algorithm.

Table 1 shows examples where improvement in forecasts was observed. The table features particular frames, ranges of predictions, the dataset identifier (A, B or C), the identifier of the base component that was removed, the values of initial and final MSE, and a percentage drop in the prediction error. Overall, 39 cases were found where a forecast improvement was observed. Interestingly, all of these results were achieved on relatively small frames (up to 75 values) and with just a single or at most two base components eliminated. These results prove that the NMF Log-Quad algorithm can be used in such a prediction-improvement system; however, as is shown in Table 2, it does not perform as effectively as other algorithms, especially those that use the ICA, e.g. the Joint Approximate Diagonalization of Eigenmatrices (JADE; Cardoso & Souloumiac, 1996; Rutledge & Jouan-Rimbaud Bouveresse, 2013) or Algorithm for Multiple Unknown Signals Extraction (AMUSE; Szupiluk & Rubach, 2020; Tong et al., 1990).

Table 2. Comparison of prediction improvement combinations for different algorithms

Frame size	NMF LQ		JADE (ICA)		AMUSE (SOS BSS)	
	number of cases	average improvement in %	number of cases	average improvement in %	number of cases	average improvement in %
20	27	28.2	1,330	25.3	975	6.8
40	9	21.7	702	21.9	993	6.9
50	1	11.2	408	20.4	841	6.6
75	2	31.2	320	16.0	1,016	5.9
100	.	.	302	13.1	1,215	6.0
500	.	.	50	7.7	1,410	4.8
1,000	.	.	18	6.0	957	4.5
2,000	.	.	18	3.1	508	4.3
3,000	.	.	5	2.2	332	4.6
4,000	.	.	10	1.8	295	4.0
5,000	.	.	4	2.1	208	4.4

Source: authors' computations based on an energy dataset, performed using standard Python, SciKit Learn and matlab implementations of the algorithm.

The same prediction improvement schema was tested with three other algorithms: the classical NMF implementation by Cichocki and Févotte (Cichocki et al., 2008; Févotte et al., 2009), available e.g. in the Python, SciKit Learn library, the ICA method using the JADE algorithm, and the Second Order Statistics (SOS) Blind Source Separation (BSS) AMUSE algorithm.

It is very important to underline that unlike the proposed NMF Log-Quad (NMF LQ), the classical NMF algorithm did not achieve a prediction improvement in any of the explored subframes or base component combinations. On the other hand, the other two algorithms achieved better results than the NMF LQ in the same setup. Table 2 shows the number of cases of frames and base component combinations that lead to lower values of the MSE. Not only were there more combinations that resulted in prediction improvement, but the JADE algorithm achieved a more significant MSE drop for all the frame sizes. Table 3, on the other hand, demonstrates that both JADE and AMUSE were often able to achieve prediction improvements in the cases where many base components were removed (sometimes four or even five components out of six). This proves their ability to separate the prediction from the noise that lowers its quality.

Table 3. Number of base components that were removed to obtain a prediction improvement – comparison of three algorithms: NMF Log-Quad, JADE and AMUSE

Number base components removed	NMF LQ (all cases: 39)	JADE (all cases: 3,148)	AMUSE (all cases: 9,025)
1	36	1,958	3,473
2	3	871	2,810
3	.	263	1,855
4	.	52	751
5	.	4	136

Source: authors' computations based on an energy dataset, performed using standard Python, SciKit Learn and matlab implementations of the algorithm.

6. Conclusions

This paper introduces a new Non-negative Matrix Factorization algorithm derived from Log-Quad divergence, and proposes its application to a prediction-improvement system. This system can be applied in various areas where different prediction techniques are possible, and might also be used as a model aggregation method. In particular, our approach can be used in specific cases where data contains physical type of noise (e.g. measurement errors). It is particularly adequate for models created in the data-mining methodology (using machine learning), where we do not assume that the model represents the true nature of the analysed phenomenon, but is only intended to fulfill an instrumental purpose, giving the best possible forecast.

We implemented the new NMF Log-Quad algorithm and the prediction improvement system to energy consumption forecasting, and its performance was compared with other methods used in blind signal separation. It should be noted that the NMF Log-Quad algorithm is characterized by a relatively high efficiency for signals with a small amount of data and, within our study, it showed an advantage over methods such as Independent Component Analysis (ICA) or Algorithm for Multiple Unknown Signals Extraction (AMUSE). However, it performed worse than the above-mentioned methods when processing larger amounts of data. These approaches can be considered complementary and used jointly to effectively improve prediction for both large and small data volumes. Compared to other NMF algorithms, such as the Image Space Reconstruction Algorithm (ISRA) and its derivatives, the NMF Log-Quad improves predictions to a greater degree. It is also a more stable and effective algorithm in terms of the Non-negative Matrix Factorization itself.

References

Amari, S. (1985). Differential Geometry of Statistical Models. In S. Amari (Ed.), *Differential-Geometrical Methods in Statistics* (pp. 11–65). Springer-Verlag. https://doi.org/10.1007/978-1-4612-5056-2_2.

Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., & Plemmons, R. J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, *52*(1), 155–173. https://doi.org/10.1016/j.csda.2006.11.006.

Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, *7*(3), 200–217. https://doi.org/10.1016/0041-5553(67)90040-7.

Cardoso, J.-F., & Souloumiac, A. (1996). Jacobi Angles for Simultaneous Diagonalization. *SIAM Journal on Matrix Analysis and Applications*, *17*(1), 161–164. https://doi.org/10.1137/S0895479893259546.

Cichocki, A., & Amari, S. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons. https://doi.org/10.1002/0470845899.

Cichocki, A., Lee, H., Kim, Y.-D., & Choi, S. (2008). Non-negative matrix factorization with α-divergence. *Pattern Recognition Letters*, *29*(9), 1433–1440. https://doi.org/10.1016/j.patrec.2008.02.016.

Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S. (2009). *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley & Sons. https://doi.org/10.1002/9780470747278.

Comon, P., & Jutten, C. (Ed.). (2010). *Handbook of Blind Source Separation: Independent Component analysis and Applications*. Academic Press. https://doi.org/10.1016/C2009-0-19334-0.

Csiszár, I. (1978). Information measures: A critical survey. In J. Kožešnik (Ed.), *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians held in Prague, 18 to 23 August 1974* (pp. 73–86). Springer. https://www.fuw.edu.pl/~kostecki/scans/csiszar1978.pdf.

Dhillon, I. S., & Sra, S. (2005). Generalized Nonnegative Matrix Approximations with Bregman Divergences. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems*, *18* (pp. 283–290). https://proceedings.neurips.cc/paper/2005/hash/d58e2f077670f4de9cd7963c857f2534-Abstract.html.

Févotte, C., Bertin, N., & Durrieu, J.-L. (2009). Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Computation*, *21*(3), 793–830. https://doi.org/10.1162/neco.2008.04-08-771.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565.

Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2013). Independent Components Analysis with the JADE algorithm. *TrAC Trends in Analytical Chemistry*, *50*, 22–32. https://doi.org/10.1016/j.trac.2013.03.013.

Szupiluk, R., & Rubach, P. (2020). *Identification of Delays in AMUSE Algorithm for Blind Signal Separation of Financial Data.* In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, & J. M. Zurada (Eds.), *Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science* (pp. 253–261). Springer. https://doi.org/10.1007/978-3-030-61534-5_23.

Szupiluk, R., Wojewnik, P., & Ząbkowski, T. (2007). Smooth Component Analysis as Ensemble Method for Prediction Improvement. In M. E. Davies, C. J. James, S. A. Abdallah, & M. D. Plumbley (Eds.), *Independent Component Analysis and Signal Separation. ICA 2007. Lecture Notes in Computer Science* (pp. 277–284). Springer. https://doi.org/10.1007/978-3-540-74494-8_35.

Tong, L., Soon, V. C., Huang, Y. F., & Liu, R. (1990). AMUSE: A new blind identification algorithm. In *IEEE International Symposium on Circuits and Systems* (pp. 1784–1787). https://doi.org/10.1109/ISCAS.1990.111981.