Log-Quad Divergence For Nonnegative Matrix Factorization Used In Multi-Model Prediction

Ryszard Szupiluk[a], Paweł Rubach[b]

Abstract. The aim of this paper is to present a new Nonnegative Matrix Factorization (NMF) algorithm based on Log-Quad divergence and demonstrate its application to separate latent destructive components contained in prediction results in a multi-model approach. In this article we present mainly a methodological view which may have applications in different areas, however, we also provide a selected practical example of a real economic problem – the prediction of electricity consumption using a dataset of hourly usage in Poland in years 1988–1997. We evaluate and compare this method with other blind signal/source separation techniques such as the ICA and AMUSE algorithms. The results show that the new algorithm has an interesting ability to improve predictions for small volumes of data.
Keywords: nonnegative 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 stosowana w predykcji za pomocą podejścia wielomodelowego

Streszczenie. Celem niniejszego artykułu jest przedstawienie nowego algorytmu nieujemnej faktoryzacji macierzy opartego na dywergencji Log-Quad i zaprezentowanie jego zastosowania w celu separacji ukrytych komponentów destrukcyjnych zawartych w predykcjach w ujęciu wielomodelowym. W artykule prezentujemy głównie koncepcję metodologiczną, która może mieć zastosowania w różnych obszarach, ale także przedstawiamy wybrany praktyczny przykład realnego problemu ekonomicznego – predykcji zużycia energii elektrycznej na podstawie danych z godzinowego zużycia w Polsce w latach 1988-1997. Nową metodę zbadano i porównano z innymi algorytmami z obszaru ślepej separacji takimi jak ICA oraz algorytm AMUSE. Wyniki pokazują, że nowy algorytm charakteryzuje się ciekawą właściwością poprawiania predykcji w przypadku zastosowania do niedużych wolumenów danych.
Słowa kluczowe: nieujemna faktoryzacja macierzy, NMF, identyfikacja ukrytych komponentów, ślepa separacja źródeł, ślepa separacja sygnałów, predykcja, ICA, AMUSE

[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.

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, a particularly intensive development took place in the area of multivariate methods, especially related to the problem of blind signal / source separation. In this problem, the goal is to separate (reproduce, estimate, reconstruct, identify) *a priori* unknown signals mixed in an unknown mixing system. Separation takes place using only mixed data. The application of the problem formulated this 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 case we have a party where many individuals are simultaneously speaking and 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 for using the Fourier transform. In this type of problems, techniques currently referred to as blind signal/source separation, such as Independent Component Analysis (ICA) and Nonnegative 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 Independent Component Analysis (ICA), Smooth Principal Component Analysis (SmPCA), Sparse Component Analysis (SCA), Independent Factor Analysis (IFA) and Nonnegative Matrix Factorization (NMF). The whole set of these methods is also referred to as Latent Component Analysis (LCA) (Cichocki & Amari, 2002).

Although particular methods are based on different mathematical criteria with accompanying assumptions specific to a given method, they have certain common features. First of all, they are most often considered in the context of a machine learning approach, what is associated with numerical optimization of the objective function which is the basis of a given method. Consequently, much of the literature for a statistical method which by definition is i.e. ICA or an algebraic method such as NMF takes place in the context of neural networks and / or machine learning.

Another common element for the above-mentioned methods is the effect of their operation in the context of blind signal separation: they may be applied to reconstruct mixed source signals. It is expected that the methods used here, despite being mathematically distinct, lead to similar or even identical results. An example illustrating the whole issue can be the so-called cocktail party problem in which the goal is to separate a single conversation from all sounds produced by multiple people talking simultaneously. The final assessment of the success of solving such a problem is a purely human perception of whether the statement is understandable. One can say that the signals received here are, in a way, physically distinct.

At the same time, the individual methods of LCA used in the context of blind separation are characterized by significant implementation differences. These differences exist both between specific methods and within a given method between particular numerical algorithms. One of the significant differences is the amount of data at which the effective operation of a given method is achieved.

The aim of this study is to derive a new nonnegative matrix factorization algorithm based on the Log-Quad divergence (Cichocki et al., 2009), and then to investigate its effectiveness and performance properties in the context of the proposed forecast improvement system based on a multi-model approach. The test will be performed on the basis of the energy consumption prediction based on a dataset of hourly electricity usage in Poland in years 1988-1997. The goal is to compare the performance of the Log-Quad NMF-based system to the ICA-based system, as well as to show the differences to other NMF algorithms, the Image Space Reconstruction Algorithm (ISRA) developed by Lee Seung (Lee & 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, among others, selection of learning coefficients, nonlinearity forms, assumed number of iterations etc. Since LCA methods belong to unsupervised learning, except in the case of easily interpreted physical signals, we generally cannot control the quality of operation of such algorithms.

2. Nonnegative Matrix Factorization

In its basic form, nonnegative 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 elements of matrix **A** are nonnegative: $a_{ij} \geq 0$, what is also denoted as: $\mathbf{A} \geq 0$. An analogous condition is applied to matrix **S**. To assess the extent to which matrices **X**, **A**, **S** fulfill the 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 **X** and **AS** expressed by the formula J=$||\boldsymbol{X} - \boldsymbol{AS}||_p$ where $||.||_p$ is the p-norm. For $p = 2$ we obtain the Euclidean distance, on the basis of which the ISRA algorithm was derived by Lee and Seung (Lee & Seung, 1999), considered to be the first solution of NMF in its current form. This algorithm can be represented in matrix form as:

$$\mathbf{A} \leftarrow \mathbf{A}. \times \mathbf{XS}^{\mathrm{T}}./\mathbf{ASS}^{\mathrm{T}}, \tag{2}$$

and

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

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 only to a certain stationary point (which does not have to be a minimum). For this reason, it is currently rarely used in real applications, it rather plays the role of a reference point, a solution against which new algorithms are compared. It is also a kind of an inspiration for the development of other approaches. The disadvantage of this algorithm is not only its slow convergence (if it converges at all), but also the fact that once it is set to zero, it stays that way. These limitations resulted in 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 one of the most popular criteria based on which NMF algorithms are derived. Unlike ordinary distance measures, the divergence function does not have to satisfy the triangle inequality and is usually asymmetric. Historically, the development of divergence measures has 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):

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

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

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

Log-Quad divergence is one of the relatively rarely discussed and used measures of similarity. 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. These measures are usually used standalone while here they are combined in one formula. We define the NMF problem for divergence (5) as $\min_{\mathbf{A},\mathbf{S}} D_{\mathrm{LQ}}(\mathbf{X}||\mathbf{AS})$ and use the formula of the alternating multiplicative NMF algorithm for Bregman divergence (Dhillon & Sra, 2005) of the form:

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

and for matrix **S**

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

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 $\varphi''(u) = 2 + 1/u$, which leads to the following algorithm for estimating the matrices **A** and **S**:

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

and

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

The obtained algorithm (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 The Quality Of Prediction

The non-negative matrix factorization model (1) can be interpreted in terms of a prediction improvement system. Let us assume that the rows of the matrix $\mathbf{X} = [x_1, x_2,..., x_n]^{\mathrm{T}}$ contain the results of forecasts from different models. Another assumption is that they are a linear combination $\mathbf{X} = \mathbf{AS}$ of the hidden components responsible for the correct results of the forecast as well as the components responsible for errors. So, we can write the matrix $\mathbf{S} = [s_1,..., s_n]^{\mathrm{T}} = [\hat{s}_1,..., \hat{s}_p, \tilde{s}_{p+1}...., \tilde{s}_n]^{\mathrm{T}}$. That is, it consists of constructive components $\hat{s}_i$ and destructive components $\tilde{s}_j$. By identifying the matrices **A** and **S** and then eliminating the destructive components (i.e. assuming $\tilde{s}_j = 0$) we should receive an improvement in the prediction quality

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

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

5. Practical Experiment

The verification and demonstration of the new algorithm NMF Log Quad has been performed using hourly energy consumption data of the overall usage in Poland in years 1988-1997. The dataset used contains hourly consumption observations and predictions generated every 24h based on past hourly consumption. There are 6 independent predictions that were generated using Multilayer Perceptron neural networks with 12, 18, 21, 24, 27, 30 neurons in the hidden layer, respectively. Overall, the dataset contains 43824 predictions and is divided into 3 arbitrarily divided subsets (Dataset A: 17544 predictions, Dataset B: 17544 and Dataset C: 8736).

The goal of the experiment was to improve the predictions of energy consumption by applying our multi-model prediction improvement schema on a given dataset. According to this schema the initial predictions are decomposed using the proposed NMF Log Quad algorithm into 6 latent components. In the next step a chosen base latent component or a set of those components is eliminated (replaced with zeros) and then the latent components are recomposed into 6 predictions. The Mean Squared Error (MSE) is computed on initial predictions as well as on the recomposed ones and then compared.

In order to better observe the performance of the algorithm the whole dataset was divided into smaller frames with predicted consumption values of various sizes: 20, 40, 50, 75, 100, 500, 1000, 2000, 3000, 4000, 5000. Overall there were 5355 frames. For each frame the 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 with zeros of 5 of the 6 base components. 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 comp. removed	Initial MSE 10^{-3}	Final MSE 10^{-3}	Improvement
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: own computations based on the energy dataset realized using Python and matlab implementations of the NMF Log Quad algorithm.

Table 1 shows examples where a prediction improvement has been observed. In detail, one can see the particular frame, the range of predictions the dataset identifier (A, B or C), the identifier of the base component(s) that was removed and the values of initial and final MSE as well as a percentage drop of the prediction error. Overall, 39 cases were found where a prediction 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 2 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 so well as other algorithms, in particular the ones that use the Independent Component Analysis (ICA) such as Joint Approximate Diagonalization of Eigenmatrices (JADE) (Cardoso & Souloumiac, 1996; Rutledge & Jouan-Rimbaud Bouveresse, 2013) or AMUSE (Szupiluk & Rubach, 2020; Tong et al., 1990).

Table 2. Comparison of prediction improvement combinations for different algorithms. (Source: Own elaboration)

Frame size	NMF LQ		JADE (ICA)		AMUSE (SOS BSS)	
	# of cases	Average improvement	# of cases	Average improvement	# of cases	Average improvement
20	27	28.2%	1330	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%	1016	5.9%
100	.	.	302	13.1%	1215	6.0%
500	.	.	50	7.7%	1410	4.8%
1000	.	.	18	6.0%	957	4.5%
2000	.	.	18	3.1%	508	4.3%
3000	.	.	5	2.2%	332	4.6%
4000	.	.	10	1.8%	295	4.0%
5000	.	.	4	2.1%	208	4.4%

Source: own computations based on the energy dataset realized using standard Python, SciKit Learn and matlab implementations of the algorithm.

The same prediction improvement schema has been tested with 3 other algorithms the classical NMF implementation by Cichocki and Fevotte (Cichocki et al., 2008; Févotte et al., 2009) which is available i.a. 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 and base component combinations. On the other hand, the other 2 algorithms achieved, in fact, better results in the same setup as the NMF LQ. Table 2 shows the number of cases of frames and base component combinations that lead to lower MSE values. It is worth noting, that not only were there more combinations that gave prediction improvement but the JADE algorithm has achieved higher MSE drop for all frame sizes. Another observation that is visible in Table 3 was that both JADE as well as AMUSE, in particular, have achieved prediction improvements often in cases where many base components were removed (sometimes 4 or even 5 out of the 6). This demonstrates 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 3 algorithms: NMF Log Quad, JADE and AMUSE.

# base components removed	NMF LQ (all cases: 39)	JADE (all cases: 3148)	AMUSE (all cases: 9025)
1	36	1958	3473
2	3	871	2810
3	.	263	1855
4	.	52	751
5	.	4	136

Source: own computations based on the energy dataset realized using standard Python, SciKit Learn and matlab implementations of the algorithm.

6. Conclusions

This paper introduced a new Nonnegative Matrix Factorization algorithm derived from Log-Quad divergence and proposed its application in a prediction improvement system. This system can be applied in various areas where different prediction techniques are accepted and may also be interpreted as a model aggregation method. In particular, this approach can be used in specific cases where the data may contain physical type of noise (e.g. measurement errors). This system 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 analyzed phenomenon, but is only intended to fulfill an instrumental purpose, giving the best possible forecast.

The new NMF Log-Quad algorithm and the prediction improvement system has been implemented for energy consumption forecasting and its performance was compared to other methods used in blind signal separation. It should be noted that the Log-Quad NMF algorithm is characterized by a relatively high efficiency for signals with a small amount of data and, in the scope of the experiment, it has an advantage over methods such as Independent Component Analysis (ICA) or AMUSE (Blind Source Separation based on Second Order Statistics). However, it performs inferior when processing larger amounts of data. These approaches can be considered complementary and used in a joint way 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 Log Quad NMF achieves a higher degree of prediction improvement. It is also a more stable and effective algorithm in terms of the nonnegative 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. 19th International Conference, ICAISC 2020, Zakopane,* *Poland, October 12-14, 2020, Proceedings, Part II* (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. 7th International Conference, ICA 2007, London, UK, September 9-12, 2007, Proceedings* (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*. IEEE International Symposium on Circuits and Systems, New Orleans. https://doi.org/10.1109/ISCAS.1990.111981.