

FUNKCJA LOG(COSH) I JEJ ROLA W ŚLEPEJ SEPARACJI SYGNAŁÓW

Ryszard Szupiluk  <https://orcid.org/0000-0002-4416-7131>

Kolegium Analiz Ekonomicznych
Szkola Główna Handlowa w Warszawie
e-mail: rszupi@sgh.waw.pl

Streszczenie: W niniejszym artykule przedstawimy szczególną właściwość algorytmów ślepej separacji jaką jest występowanie funkcji $\log(\cosh)$ w funkcjach celu, z których owe algorytmy są wyprowadzane. Wskażemy, że zakładane jako zasadniczo różne metody separacji, oparte na odmiennych kryteriach takich jak statystyczna niezależność, rzadkość lub gładkość, mogą lub są w praktyce sprowadzane do eksploracji w istocie tej samej matematycznej charakterystyki.

Słowa kluczowe: analiza składowych niezależnych, ślepa separacja, analiza składowych gładkich, rzadkie dane

JEL classification: C02, C50

WSTĘP

Ślepa separacja sygnałów, zwana także ślepą separacją źródeł (ang. Blind Singnal/Source Sepration–BSS), jest problemem intensywnie badanym w dwóch ostatnich dekadach [Comon, Jutten 2010; Piccolotto et al. 2022]. Zasadnicza idea BSS polega na odtworzeniu nieznanymi *a priori* sygnałów źródłowych, zmieszanych w także nieznanym systemie mieszającym. W zależności od kontekstu rozważań cały proces może być określany jako ślepa: separacja, identyfikacja lub estymacja [Cichocki et al. 2002]. Zagadnienie BSS może być rozważane w kategoriach praktycznych problemów lub czysto matematycznych przekształceń. W tym drugim przypadku BSS można przestawić jako zadanie poszukiwania interesujących transformacji i reprezentacji danych. W związku z tym zagadnieniem powstały, zostały rozwinięte lub zaadoptowane takie metody jak analiza składowych

<https://doi.org/10.22630/MIBE.2022.23.3.8>

niezależnych [Cardoso 1999], analiza składowych rzadkich [Zibulevsky, Pearlmutter 2001], analiza składowych gładkich [Szupiluk et al. 2006], nieujemna faktoryzacja macierzy [Berry et al. 2007], nieliniowa analiza składowych głównych [Oja 1995] i wiele innych. W większości przypadków ww. metody wymagają zastosowania metod numerycznych, co powoduje, że są, lub mogą być, klasyfikowane także jako metody uczenia maszynowego [Haykin 2009].

Cechą charakterystyczną dla problematyki BSS jest oczekiwanie, że niezależnie od wybranych metod separacji, otrzymane wyniki będą podobne lub identyczne. W przypadku praktycznych zastosowań przyjmuje się, że ze zmieszanych danych obserwowanych zostaną wyodrębnione, rzeczywiste, odrębne sygnały źródłowe. Przykładem może być tu odtworzenie pojedynczej wypowiedzi z gwaru rozmów (tzw. cocktail party problem) [Koutras, 2002]. Podobnie oczekiwania adresowane są w przypadku czysto analitycznych reprezentacji danych. Oznacza to, że zarówno w przypadkach praktycznych jak i symulacjach teoretycznych oczekuje się wyodrębnienia sygnałów (używane są także określenia: komponentów, składowych, zmiennych, szeregów czasowych) o niejako fizycznym charakterze. Osiągnięcie takich rezultatów, metodami o różnych właściwościach, bazującymi na różnych kryteriach i założeniach, zazwyczaj wyjaśnia się faktem istnienia różnych charakterystyk sygnałów, których eksploracja może doprowadzić do ich wyodrębnienia. Pojawia się jednak pytanie, w jakim stopniu poszczególne formalnie „różne” metody, są w istocie różne.

W niniejszym artykule wykażemy, że pozornie odmienne kryteria jak wzajemna niezależność sygnałów oraz indywidualne charakterystyki sygnałów jak rzadkość lub gładkość w kontekście ślepej separacji, na pewnym poziomie rozważań wykazują znaczące podobieństwa.

PROBLEM ŚLEPEJ SEPARACJI

Zadanie ślepej separacji można przedstawić następująco. Przyjmuje się, że pewne nieznanne a *priori* sygnały źródłowe zostały zmieszane w także nieznanym systemie mieszającym. Należy odtworzyć sygnały źródłowe dysponując jedynie obserwacjami zmieszanymi. Dla rozwiązania tego problemu przyjmowane są pewne robocze założenia dotyczące postaci systemu mieszającego oraz sygnałów źródłowych, co prowadzi do zdefiniowania systemu generującego. W zależności od przyjętego/założonego systemu mieszającego możemy mówić o BSS: statycznym, dynamicznym, nieliniowym statycznym lub nieliniowym dynamicznym [Cichocki, Amari 2002]. Najczęściej rozważanym jest przypadek liniowego statycznego systemu postaci

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \quad (1)$$

gdzie: $\mathbf{A} \in R^{N \times M}$, $\mathbf{x}(k) = [x_1(k), \dots, x_N(k)]^T$ jest wektorem sygnałów obserwowanych, zaś $\mathbf{s}(k) = [s_1(k), \dots, s_M(k)]^T$ źródłowych, k oznacza indeks czasu lub numer obserwacji. Z reguły przyjmuje się także, następujące dodatkowe założenia: $M \leq N$ - oznacza to, iż liczba sygnałów obserwowanych nie może być mniejsza od liczby sygnałów źródłowych, w naszych rozważaniach $N=M$, zaś kolumny macierzy \mathbf{A} są liniowo niezależne [Cardoso 1998].

W celu znalezienia \mathbf{x} , poszukuje się takiego systemu separującego, reprezentowanego przez macierz \mathbf{W} , że dla sygnałów separowanych $\mathbf{y}(k) = [y_1(k), \dots, y_M(k)]^T$ zachodzi

$$\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k) = \mathbf{W}\mathbf{A}\mathbf{s}(k) = \mathbf{P}\mathbf{D}\mathbf{s}(k), \quad (2)$$

gdzie: \mathbf{P} - macierz permutacji określająca kolejność estymowanych sygnałów, \mathbf{D} - diagonalna macierz skalująca. Równanie (3) oznacza, że estymowane sygnały źródłowe \mathbf{y} mogą być przeskalowaną oraz w innej kolejności uporządkowaną „wersją” sygnałów \mathbf{s} . Jest to nieunikniona konsekwencja faktu, iż przyjmując $N=M$, dla dowolnej odwracalnej macierzy $\mathbf{E} \in R^{N \times N}$ zachodzi $\mathbf{x} = \mathbf{A}\mathbf{E}\mathbf{E}^{-1}\mathbf{s}(k)$.

Rozwiązanie tak postawionego problemu wymaga z reguły jeszcze dodatkowych założeń związanych ze specyfiką danej metody separacji. W dalszych rozważaniach skupimy się na podejściach eksplorujących takie charakterystyki sygnałów jak statystyczna niezależność, gładkość o rzadkość danych.

ANALIZA SKŁADOWYCH NIEZALEŻNYCH

Analiza składowych niezależnych jest jedną z pierwszych metod BSS. Jest także metodą uważaną za podstawową w tym obszarze, a w początkowej fazie rozwoju problematyki ślepej separacji, metoda ICA oraz problem BSS były traktowane zgoła jako tożsame. Z formalnego punktu widzenia, analiza składowych niezależnych stawia sobie za cel wydzielenie z obserwacji komponentów (składników, sygnałów) statystycznie niezależnych. Oznacza to, że dla modelu (1)-(2) poszukujemy takiego \mathbf{W} , że transformacja $\mathbf{y}=\mathbf{W}\mathbf{x}$ prowadzi do spełnienia warunku statystycznej niezależności estymowanych danych [Comon 1994]

$$p_1(y_1)p_2(y_2)\dots p_n(y_n) = p_{1\dots n}(y_1, y_2, \dots, y_n), \quad (3)$$

gdzie $p_i(y_i)$ oznacza funkcję gęstości prawdopodobieństwa zmiennej y_i , zaś prawa strona równania (3) oznacza łączną funkcję gęstości prawdopodobieństwa sygnałów separowanych. Do oceny spełnienia równania (3) przyjmowane są różne praktyczne kryteria matematyczne. Jednym z nich może być wzajemna informacja [Amari et al. 1999]

$$I(\mathbf{y}) = \int_{-\infty}^{+\infty} p(y_1, \dots, y_N) \log \frac{p(y_1, \dots, y_N)}{\prod_i p(y_i)} dy. \quad (4)$$

Wyrażenie (4) może być przyjęte jako funkcja celu, co sprowadza zagadnienie znajdowania składników niezależnych do problemu optymalizacji [Cardoso 1997]

$$\mathbf{W}_{opt} = \min_{\mathbf{W}} I(\mathbf{y}) = \min_{\mathbf{W}} I(\mathbf{W}\mathbf{x}). \quad (5)$$

Minimalizacja (5) wiąże się zwykle z przekształceniem wyrażania (4) do postaci wygodniejszej dla wyprowadzenia algorytm uczenia. Jedną z możliwości jest wykorzystanie pojęcia negentropii [Girolami, Fyfe 1996]

$$J(\mathbf{y}) = H_G(\mathbf{y}) - H(\mathbf{y}), \quad (6)$$

gdzie $H(\mathbf{y}) = - \int_{-\infty}^{+\infty} p(\mathbf{y}) \log(p(\mathbf{y})) d\mathbf{y}$ jest entropią zmiennej \mathbf{y} zaś $H_G(\mathbf{y})$ jest entropią

zmiennej \mathbf{y} o rozkładzie Gaussowskim, z taką samą średnią oraz kowariancją jak $p(\mathbf{y})$. Wyrażenie (5) można traktować jako miarę niegaussowości zmiennej \mathbf{y} . Negentropia, w odróżnieniu od zwykłej entropii, jest niezmiennicza dla transformacji odwracalnych, czyli dla $\mathbf{y} = \mathbf{W}\mathbf{z}$ zachodzi $J(\mathbf{y}) = J(\mathbf{W}\mathbf{z}) = J(\mathbf{z})$. Można wykazać, że jeżeli dane \mathbf{x} poddamy procesowi dekorelacji $\mathbf{z} = \mathbf{Q}\mathbf{x}$, gdzie \mathbf{Q} jest macierzą dekorującą to otrzymamy następującą zależność [Lee 1998].

$$I(\mathbf{y}) = J(\mathbf{z}) - \sum_{i=1}^N J_i(y_i), \quad (7)$$

Przyjmując, że dla $\mathbf{y} = \mathbf{W}\mathbf{z} = \mathbf{W}\mathbf{Q}\mathbf{x}$, macierz \mathbf{W} nie wpływa na $J(\mathbf{z})$ problem ślepej separacji sprowadza się do następującego zadania optymalizacji

$$\min_{\mathbf{W}} I(\mathbf{y}) = \max_{\mathbf{W}} \sum_{i=1}^N J_i(y_i). \quad (8)$$

Dla rozwiązania problemu (8) można wykorzystać następującą aproksymację negentropii [Hyvarinen 1998]

$$J(y) \approx q_1(E\{G_1(y)\})^2 + q_2(E\{G_2(y)\} - E\{G_2(v_G)\})^2, \quad (9)$$

gdzie $E\{\cdot\}$ jest operatorem wartości oczekiwanej, $G_1(y)$, $G_2(y)$ są niekwadratowymi funkcjami nieliniowymi określającymi odpowiednio stopień niesymetryczności oraz stromości zmiennej y , v_G oznacza dowolną zmienną o

rozkładzie gaussowskim, q_1, q_2 są odpowiednio dobranymi stałymi. Wyrażenie (9) można uprościć dla rozkładów symetrycznych do

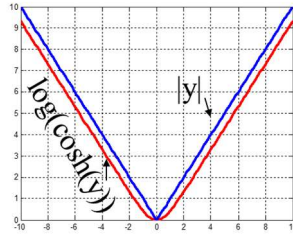
$$J(y) \approx [E\{G_2(y)\} - E\{G_2(v_G)\}]^2, \quad (10)$$

Dobór funkcji $G_2(y)$ może opierać się na różnych motywacjach, jedną z nich może być kontekst uczenia maszynowego w jakim postawiony może być problem optymalizacji funkcji (8). W takim kontekście funkcją, która spełnia warunki aproksymacji a jednocześnie ma sprawdzone i dobrze interpretowalne własności w kategoriach uczenia maszynowego jest funkcja [Oja 1995]

$$f(y) = \log \cosh(y). \quad (11)$$

Jest używana m.in. jako aproksymacja wartości bezwzględnej, co jest korzystne w przypadku gradientowych metod optymalizacji, rysunek 1.

Rysunek 1. Wykresy funkcji $\log(\cosh)$ oraz wartości bezwzględnej



Źródło: opracowanie własne

Wstawiając $G_2(y) = f(y)$ oraz uwzględniając człony zależne od w otrzymuje się następującą aproksymację negentropii

$$J(y) = [E\{\log(\cosh(y))\} - E\{\log(\cosh(v_G))\}]^2. \quad (12)$$

Prowadzi to ostatecznie do następującej funkcji celu

$$F_1 = \sum_{i=1}^N [E\{\log(\cosh(y_i))\} - E\{\log(\cosh(v_G))\}]^2, \quad (13)$$

stanowiącej podstawę algorytmów estymacji niezależnych komponentów a w kontekście ślepej separacji algorytmów odtwarzających zmieszane sygnały.

ANALIZA SKŁADOWYCH RZADKICH

Analiza składowych rzadkich (ang. Sparse Component Analysis–SCA) eksploruje charakterystykę danych określaną jako rzadkość [Zibulevsky, Pearlmutter 2001]. Za sygnały rzadkie uznaje się sygnały, w których dominuje jedna wartość np. zero, zaś

„rzadko” pojawia się wartość niezerowa. W ogólnym przypadku wartość dominująca może być dowolna. Rzadkość danych ma istotne znaczenie w zagadnieniach reprezentacji oraz kompresji danych. W przypadku metod analitycznych charakterystyka ta może pełnić pewną dodatkową rolę jako element kary, ograniczeń lub dodatkowych warunków. Przyjmując rzadkość jako główne kryterium wyodrębniania sygnałów można mówić o zdefiniowaniu problemu analizy składowych rzadkich. Może ona zostać sformowana następująco. Dla obserwacji $\mathbf{x}(k) = [x_1(k), \dots, x_N(k)]^T$ poszukujemy takiej kombinacji liniowej (2), że uzyskane zmienne y_i charakteryzuje się jak największą rzadkością (od największej do najmniejszej). Kluczowym elementem jest tu określenie kryterium rzadkości, które z reguły pełni jednocześnie rolę funkcji celu dla algorytmu separacji. Typowymi miarami rzadkości sygnałów są kurtoza, norma zero lub p -norma z $p=1$ postaci [Georgiev et al. 2005]

$$l_s = \sum_{k=1}^K |y(k)|, \quad (14)$$

Występująca w (14) wartość bezwzględna jest zwykle zastępowana wygodniejszą z punktu widzenia algorytmów uczenia aproksymacją za pomocą funkcji (11) co prowadzi do funkcji celu postaci

$$F_2 = E\{\log(\cosh(y(k)))\}, \quad (15)$$

Eksploracja rzadkiej charakterystyki danych, często wiązana jest i innymi kryteriami w efekcie mamy takie metody „mieszane” jak rzadka analiza składowych niezależnych [Khan, Kim 2009], rzadka analiza składowych głównych [Zou et al. 2006].

ANALIZA SKŁADOWYCH GŁADKICH

Analiza składowych gładkich (ang. Smooth Component Analysis – SmCA), eksploruje charakterystykę, którą możemy określić jako zmienność (lub zmiany) między kolejnymi wartościami sygnału, przy czym, jest to metoda adresowana do danych posiadających strukturę czasową tzn. w których kolejność obserwacji jest istotna (sygnały, szeregi czasowe) [Szupiluk et al. 2012]. W przypadku małych zmian, sygnał można potraktować jako gładki. Kryterium oceny gładkości stanowiące jednocześnie funkcję celu na bazie której wprowadzany jest algorytm SaCA można przyjąć postaci [Szupiluk et al. 2006]

$$P_1(y) = \frac{\frac{1}{N} \sum_{k=2}^N |y(k) - y(k-1)|}{\rho(\max(y) - \min(y))}, \quad (16)$$

gdzie

$$\rho(u) = \begin{cases} 1 & \text{dla } u = 0 \\ u & \text{dla } u \neq 0 \end{cases} \quad (17)$$

Miara (16) ma wartość maksymalną, kiedy zmiany dla kolejnych obserwacji, są równe zakresowi sygnału (maksymalne zmiany), zaś miara jest minimalna, kiedy sygnał jest stały. Możliwe wartości miary $P_1(y)$, zawierają się w przedziale od 1 do 0. Funkcja jedyńki $\rho(\cdot)$, w mianowniku, wprowadzona została aby uniknąć dzielenia przez zero. Miarę gładkości (16), można uogólnić, zastępując wartość średnią w liczniku wartością oczekiwaną.

Przyjmując, że mamy dane znormalizowane do przedziału $\langle -1, 1 \rangle$ oraz zastępując wartość bezwzględną funkcją $\log(\cosh)$, z powodów jak wyżej, otrzymujemy funkcję celu postaci $F_3 = E\{\log(\cosh(y(k)) - \log(\cosh(y(k-1))))\}$.

W przypadku sygnałów „wysoco” gładkich, gdzie $y(k) \approx y(k-1)$, możliwe jest wprowadzenie wag dla wartości w chwili k oraz $k-1$. Co prowadzi do

$$F_3 = (b_1 - b_2)E\{\log(\cosh(y(k)))\} \quad (18)$$

Dla $b_1 = 2, b_2 = 1$ funkcja (18) jest tożsama z (15).

ALGORYTMY SEPARACJI I ICH DZIAŁANIE

Mając zdefiniowaną funkcję celu, można na jej bazie wyprowadzić algorytm uczenia. Kolejne komponenty źródłowe, wyznaczone są, jako kolejne kombinacje liniowe sygnałów x_i spełniające kryteriom separacji. Dla modelu (2) poszukiwana jest więc taka macierz $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]$, że $y_1(k) = \mathbf{w}_1^T \mathbf{x}(k)$ maksymalizuje $F_1(y_1)$, czyli

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}\|=1} (F_1(\mathbf{w}^T \mathbf{x})) \quad (19)$$

W funkcji występuje warunek ograniczający $\|\mathbf{w}\|=1$, co zapobiega dążeniu wag do wartości nieskończonych.

Mając estymowane pierwszych $k-1$ komponentów źródłowych eliminujemy je z obserwacji poprzez ortogonalizację Gram-Schmidt'a i powtarzamy procedurę optymalizacji otrzymując kolejne komponenty

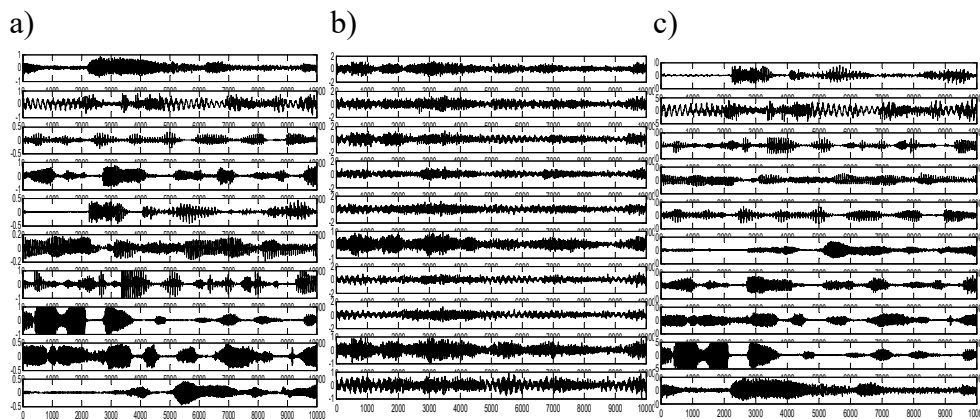
$$\mathbf{w}_k = \arg \max_{\|\mathbf{w}\|=1} (P(\mathbf{w}^T (\mathbf{x} - \sum_{i=1}^{k-1} \mathbf{y}_i \mathbf{y}_i^T \mathbf{x}))), \quad (20)$$

gdzie $\mathbf{y}_i = \mathbf{w}_i^T \mathbf{x}$, $i = 1 \dots k$.

Zwróćmy uwagę, że analogiczna procedura może być zastosowana do (15) oraz (18) z tym, że tym razem poszukiwane będzie minimum. Niemniej w każdym z tych przypadków, czyli: statystycznej niezależności sygnałów, gładkości sygnałów oraz rzadkości sygnałów podstawą algorytmu estymującego komponenty danej metody jest (lub może być) „podobna” funkcja celu.

Konkretne algorytmy separacji bazujące na (19)-(20) mogą mieć postać różnych procedur numerycznych, co zależy między innymi od przyjętej metody optymalizacji [Hyvariene et al., 2001, Cichocki i Amari 2001, Szupiluk et al. 2012] Niemniej uzyskane wyniki będą z reguły podobne. Przykładowe działanie algorytmu separacji prezentuje rysunek 2.

Rysunek 2. Sygnały a) źródłowe, b) zmieszane, c) separowane



Źródło: opracowanie własne

PODSUMOWANIE

W niniejszym artykule przedstawiliśmy szczególną rolę jaką odgrywa funkcja $\log(\cosh)$ w problemie ślepej separacji. Zasadniczo funkcja ta nie jest wiązana bezpośrednio z jakąś szczególną charakterystyką danych. A jednak, przy bliższym wglądzie w głąb algorytmów separacji pojawia się jako podstawa dla aproksymacji wydawałoby się odmiennych pojęć jak statystyczna niezależność, gładkość oraz rzadkość danych. Standardowa interpretacja osiągnięcia podobnych wyników separacji mówi o tym, że eksploracja różnych, odmiennych charakterystyk sygnałów powinna prowadzić do podobnych wyników. Jednak jak widzimy te odmienności na poziomie technicznych przekształceń prowadzących do końcowego algorytm znacząco się zacierają. Można powiedzieć, że w istocie eksplorowana jest właściwość $\log(\cosh(y))$, która nie ma bezpośredniej interpretacji fizycznej.

BIBLIOGRAFIA

- Amari S., Cichocki A., Yang H. (1999) Unsupervised Adaptive Filtering, chapter Blind Signal Separation and Extraction - Neural and Information Theoretic Approaches. John Wiley.
- Berry M., Browne M., Langville A., Pauca P., Plemmons R. (2007) Algorithms and Applications for Approximate Nonnegative Matrix Factorization. *Computational Statistics & Data Analysis*, 52(1), 155-173.
- Cardoso J. (1998) Blind Signal Separation: Statistical Principles. *Proceedings. of the IEEE*, 86(10), 2009-2025.
- Cardoso J. (1999) High-Order Contrasts for Independent Component Analysis. *Neural Computation*, 11(1), 157-192.
- Cardoso J. (1997) Infomax and Maximum Likelihood for Source Separation. *IEEE Letters on Signal Processing*, 4(4), 112-114.
- Cichocki A., Amari S. (2002) Adaptive Blind Signal and Image Processing. John Wiley, Chichester.
- Comon P., Jutten Ch. (2010) Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press.
- Georgiev P., Theis F., Cichocki A. (2005) Sparse Component Analysis and Blind Source Separation of Underdetermined Mixtures. *IEEE Transactions on Neural Networks*, 16(4), 992-996.
- Girolami M., Fyfe C. (1996) Negentropy and Kurtosis as Projection Pursuit Indices Provide Generalized ICA Algorithms. *Advances in Neural Information Processing Systems, NIPS'96 Workshop*, Snowmaas.
- Haykin S. (2009) Neural Networks and Learning Machines. Upper Saddle River: Pearson Education.
- Hyvarinen A. (1998) New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit. *Advances in Neural Information Processing Systems*, 10, 273-279, Cambridge, MA: MIT Press.
- Hyvarinen A., Karhunen J., Oja E. (2001) Independent Component Analysis. Wiley John, New York.
- Khan A., Kim I. (2009) Sparse Independent Component Analysis with Interpolation for Blind Source Separation. *2nd International Conference on Computer, Control and Communication*.
- Koutras A. (2002) Blind Speech Separation for Solving the Cocktail Party Problem in Real Room Environments. *Info-M Journal*, 1, 29-32.
- Lee T. (1998) Independent Component Analysis: Theory and Applications. Kluwer Academic Publishers.
- Oja E. (1995) The Nonlinear PCA Learning Rule and Signal Separation – Mathematical Analysis. Helsinki Univ. of Technology, Lab. of Computer and Information Science, Report A26.
- Piccolotto N., Bögl M., Muehlmann C., Nordhausen K., Filzmoser P., Miksch S. (2022) Visual Parameter Selection for Spatial Blind Source Separation. *Computer Graphics Forum*, 41,3, 157-168.

- Szupiluk R., Wojewnik P., Ząbkowski T. (2012) Smooth Component Analysis and MSE Decomposition for Ensemble Methods. *Lectures Notes in Artificial Intelligence*, 7327, 595-603.
- Szupiluk R., Wojewnik P., Ząbkowski T. (2006) Prediction Improvement via Smooth Component Analysis and Neural Network Mixing. *ICANN 2006, Lecture Notes in Computer Science*, 4132, 133-140, Springer, Berlin, Heidelberg.
- Szupiluk R., Wojewnik P., Ząbkowski T. (2007) Smooth Component Analysis as Ensemble Method for Prediction Improvement. *Lecture Notes in Computer Science*, 4666, 277-284. Springer, Berlin, Heidelberg.
- Szupiluk R. (2014) Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli Data Mining. Szkoła Główna Handlowa, Oficyna Wydawnicza.
- Zibulevsky M., Pearlmutter B. (2001) Blind Source Separation by Sparse Decomposition in a Signal Dictionary. *Neural Computation*, 13, 863-882.
- Zou H., Hastie T., Tibshirani R. (2006) Sparse Principal Component Analysis, *Journal of Computational and Graphical Statistics*, 15(2), 262-286.

THE LOG (COSH) FUNCTION AND ITS ROLE IN THE PROBLEM OF BLIND SEPARATION

Abstract: In this article, we will present a special feature of blind separation algorithms, which is the presence of $\log(\cosh)$ functions in the objective functions from which these algorithms are derived. We point out that, assumed as fundamentally different methods of separation, based on different criteria such as statistical independence, sparsity or smoothness, can or are in practice reduced to the exploration of essentially the same mathematical characteristic.

Keywords: independent component analysis, smooth component analysis, blind signals separation, sparse data

JEL classification: C02, C50