

Poznań, 26 marca 2019 r.

Wniosek o nadanie stopnia naukowego doktora habilitowanego

Tytuł:

Opracowanie algorytmu wyznaczania wymiaru korelacyjnego
dla chaotycznych sygnałów wysokowymiarowych

Załącznik: Autoreferat

Krzysztof Piotr Michalak

dr n. med. inż. informatyk, lek. neurolog
Uniwersytet Adama Mickiewicza w Poznaniu
Wydział Fizyki, Pracownia Fizyki Widzenia i Optometrii
ul. Umultowska 85, 61-614 Poznań

kmichalak@amu.edu.pl

ORCID:	0000-0002-0267-5605
Scopus Author ID:	55980875700
ResearcherID:	O-7865-2018
PBN ID:	1761776

SPIS TREŚCI

I.	Wykształcenie, tytuły, zatrudnienie	4
1.	Posiadane dyplomy i stopnie naukowe	4
2.	Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych	4
II.	Osiągnięcia naukowe wynikające z art.16 ust.2 ustawy z dnia 14 III 2003	5
III.	Lista publikacji podlegająca ocenie:.....	6
IV.	Szczegółowe omówienie wyników przedstawionych w ww. publikacjach wraz ze streszczeniem osiągniętego celu	7
1.	Wstęp.....	7
	Algorytm Grassbergera-Proccacia i Takensa-Ellnera.....	9
2.	Predictability of posturographic signals: Influence of window width and embedding dimension,.....	10
3.	Distinguishing Separate Components in High-Dimensional Signals by Using the Modified Embedding Method and Forecasting,	11
4.	Modifications of the Takens-Ellner algorithm for medium- and high-dimensional signals.....	13
	Problem odstępu B.....	13
	Problem optymalnej szerokości okna W (W_{opt}) dla wyznaczenia złożoności sygnału.....	14
	Rozwinięcie problemu interpolacji sygnału	14
	Potwierdzenie efektu saturacji d_w dla $m > m^*$ przy użyciu stałej szerokości okna W	15
	Problem optymalnego P_{max}	15
	Problem P_{min}	16
	Wpływ zaszumienia sygnału na poprawność oszacowania d	16
5.	How to estimate the correlation dimension of high-dimensional signals	17
	Porównanie algorytmu Grassbergera-Proccacia i Takensa-Ellnera.....	17
	Wyznaczenie wzoru na dokładność oszacowania wartości d_w	18
	Wyznaczenie parametrów P_{max} i δ_{max} określających koniec obszaru skalowania liniowego całki korelacyjnej.	19
	Zwiększanie błędu odchylenia e dla rosnących wartości szerokości okna W	22
	Oszacowanie maksymalnej liczby powtórzeń procedury łączenia punktów w pary (<i>Random Joining Procedure, RJP</i>).....	24
	Oszacowanie maksymalnej złożoności wymiarowej d_{max} , którą można wyznaczyć zmodyfikowanym algorytmem.....	25
6.	Estimating correlation dimension of high-dimensional signals – quick algorithm	26
	Dodanie możliwości dopasowania do relacji $d_w = fn(W)$ innych funkcji niż wielomian 4-tego stopnia.....	26
	Wprowadzenie wielostopniowego odrzucania zbyt dużych odległości między punktami, przekraczających wartość δ_{max}	26
	Sygnały o rozkładzie mocno nie-gaussowskim, wstępna normalizacja sygnału	27
	Poprawki do podstawowego wzoru Takensa-Ellnera	28

Sygnał o dwóch obszarach plateau na wykresie $d_w = \text{fn}(W)$	29
7. Podsumowanie powyższych publikacji	30
8. The influence of the vision on the complexity of the body balance regulation	30
Przykładowe zastosowanie algorytmu do sygnałów posturograficznych	30
Wyniki	32
Podsumowanie.....	33
9. Podsumowanie prezentowanych publikacji	33
10. Literatura.....	34
11. Funkcje wchodzące w skład pakietu do wyznaczania wymiaru korelacyjnego: HDS Toolkit 2:	35
Funkcje podstawowe	35
Funkcje wyznaczające histogramy odległości między punktami H, H_c, Hsl oraz kolejne ich sploty oraz funkcje operujące na nich tych histogramach i splotach.	36
Wyznaczanie odległości między punktami w przestrzeni m -wymiarowej (wektor: $deltastar / \delta$).....	36
Wyznaczanie wartości wymiaru korelacyjnego d dla określonych kombinacji $W, Pmin, Pmax$ na podstawie wyznaczonego wektora δ	37
Dopasowanie wielomianów i innych funkcji do relacji $d=\text{fn}(Pmax,W)$. Wyznaczanie $dopt, Wopt$	37
Funkcje rysowania wykresów pojawiających się w trakcie obliczeń.....	38
Pomocnicze funkcje matematyczne.....	38
Funkcje dokonujące monotonicznych nieliniowych przekształceń sygnału.....	39
V. Omówienie pozostałych osiągnięć naukowo-badawczych	40
1. Patobiochemia a odżywianie	41
2. Regulacja produkcji energii w mitochondriach ze szczególnym uwzględnieniem komórek nowotworowych.....	43
Regulacja produkcji energii a toksyczność fluorochinolonów.....	44
3. Biologia kwantowa, czyli fizyka kwantowa zastosowana do układów biologicznych.....	45
Podsumowanie.....	46
4. Pozostały dorobek naukowy i dydaktyczny w obszarze informatyki	47
Pozostałe publikacje naukowe i wystąpienia konferencyjne	47
Oprogramowanie do ćwiczeń laboratoryjnych i eksperymentów badawczych w środowiskach Matlab i Presentation.....	48
VI. Podsumowanie dorobku naukowego:.....	49

I. Wykształcenie, tytuły, zatrudnienie

1. Posiadane dyplomy i stopnie naukowe

WYKSZTAŁCENIE:

- | | |
|---|------------|
| 1. Uniwersytet Medyczny w Poznaniu – Wydział Lekarski | 1986 -1992 |
| 2. Politechnika Poznańska – kier. Informatyka, 5-letnie studia magisterskie | 1992 -1997 |

SPECJALIZACJA LEKARSKA:

- | | |
|------------------------|------|
| Neurologia - I stopień | 2000 |
|------------------------|------|

STOPNIE NAUKOWE I TYTUŁY ZAWODOWE:

- | | |
|---------------------------------|------|
| 1. lekarz medycyny | 1992 |
| 2. magister inżynier informatyk | 1997 |

Praca magisterska: Algorytmy dopasowania sekwencji białkowych

Miejsce: Politechnika Poznańska, Wydział Elektryczny, kierunek: Informatyka

- | | |
|---------------------------|------|
| 3. doktor nauk medycznych | 2005 |
|---------------------------|------|

Tytuł pracy doktorskiej: Wykorzystanie teorii dynamiki nieliniowej do badania właściwości sygnału posturo graficznego

Miejsce: Uniwersytet Medyczny im. K. Marcinkowskiego w Poznaniu

2. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- | | |
|---|------------------------------------|
| Szpital im. F. Raszei w Poznaniu – staż podyplomowy | XII 1992 – XII 1993 |
| Klinika Endokrynologii w Poznaniu, Pracownia Med. Nuklearnej | IV 1994 – VI 1995 |
| Politechnika Poznańska, Instytut Informatyki | IX 1996 – VII 1997 |
| Klinika Neurologii, Akademii Medycznej w Poznaniu (staż specjalizacyjny) | X 1997 – X 2000 |
| Katedra i Zakład Biofizyki Uniw. Medycznego w Poznaniu - <i>adiunkt</i> | VII 1997 – IX 2013 |
| Pracownia Fizyki Widzenia i Optometrii, Wydział Fizyki,
Uniwersytet Adama Mickiewicza w Poznaniu | - <i>adiunkt</i> X 2013 – do teraz |

II. Osiągnięcia naukowe wynikające z art.16 ust.2 ustawy z dnia 14 III 2003

Teoria chaosu deterministycznego zwana też teorią dynamiki nieliniowej to dziedzina wiedzy stosunkowo nowa i mało wykorzystywana w naukach biologicznych. Badania wskazują jednak, że regulacja układów biologicznych ma bardzo często charakter nieliniowy o wysokim stopniu złożoności, czyli podlega pod wiedzę z zakresu teorii dynamiki nieliniowej. Bardzo wiele sygnałów biologicznych cechuje się wysoką złożonością (np. sygnały posturograficzne badające zmysł równowagi, zmienność ciśnienia krwi, zmienność częstości akcji serca, sygnały EEG, magnetoencefalograficzne), przez co trudno poddają się badaniom z użyciem konwencjonalnych metod analizy statystycznej. Bardzo często, trudno też w takim przypadku o określenie parametrów diagnostycznych dla określonych sygnałów, które są z reguły celem ostatecznym prowadzenia badań nad sygnałami biologicznymi.

Moje badania w zakresie wykorzystania metod dynamiki nieliniowej do analizy sygnałów biologicznych prowadzone były w dwóch kierunkach: (a) badanie właściwości sygnałów posturograficznych oraz (b) rozwój metodyki analizy nieliniowej w celu lepszego dopasowania tych metod do potrzeb badania wysokowymiarowych sygnałów biologicznych.

Spośród różnych kierunków badań, jako swoje główne osiągnięcie w postępowaniu habilitacyjnym chciałbym przedstawić publikacje, w których rozwinąłem i zoptymalizowałem algorytm wyznaczania wymiaru korelacyjnego będącego jednym z parametrów określających złożoność wysokowymiarowych sygnałów chaotycznych. Algorytm ma charakter uniwersalny, czyli może być zastosowany do analizy wszelkich sygnałów ciągłych reprezentujących zachowanie złożonych układów samoregulacyjnych.

III. Lista publikacji podlegająca ocenie:

1. Predictability of posturographic signals: Influence of window width and embedding dimension, Michalak KP, Jaśkowski P, Biocybernetics and Biomedical Engineering, 2007, *Vol.27, Iss.4, pp.71-83, MNiSW: 6*

Udział własny **90%**.

Mój udział polegał na przygotowaniu procedur obliczeniowych w środowisku Matlab, przeprowadzeniu obliczeń, przygotowaniu rycin, napisaniu manuskryptu, przygotowaniu i sformatowaniu literatury.

Udział współautora polegał na merytorycznym nadzorze nad całością prac i korekcie manuskryptu.

Współautor zmarł.

2. Distinguishing Separate Components in High-dimensional Signals by Using the Modified Embedding Method and Forecasting, Michalak KP, Annals of Biomedical Engineering, 2010, *Vol. 38, Iss. 1, pp. 200-207, IF: 2,376, MNiSW: 30*

Udział własny : **100%**.

3. Modifications of the Takens-Ellner algorithm for medium- and high-dimensional signals, Michalak KP, Physical Review E, 2011, *Vol. 83, Iss. 2, pp. 026206, IF: 2,255, MNiSW: 35*

Udział własny : **100%**.

4. How to estimate the correlation dimension of High-dimensional signals?, Michalak KP, Chaos **24**, 033118 (2014); <https://doi.org/10.1063/1.4891185>, *IF: 1.954, MNiSW: 40*

Udział własny : **100%**.

5. Estimating correlation dimension of high-dimensional signals – quick algorithm, Michalak KP, AIP Advances **8**, 105201 (2018), <https://doi.org/10.1063/1.5013255>, *IF 1,653 MNiSW 25*

Udział własny : **100%**.

6. The influence of the vision on the complexity of the body balance regulation, Michalak KP, Przekoracka-Krawczyk A., Optyka **2019**, 1(56), pp. 38-42, *MNiSZW 2*

Udział własny : **90%**.

Mój udział polegał na przygotowaniu procedur obliczeniowych w środowisku Matlab, przeprowadzeniu obliczeń, przygotowaniu rycin, napisaniu manuskryptu, przygotowaniu i sformatowaniu literatury.

Udział współautora polegał na przeprowadzeniu analizy statystycznej i korekcie manuskryptu.

IV. Szczegółowe omówienie wyników przedstawionych w ww. publikacjach wraz ze streszczeniem osiągniętego celu

1. Wstęp

Sygnały chaotyczne to sygnały opisane przez deterministyczne układy równań różniczkowych, których rozwiązaniem jest nieregularny przebieg zmiennych w czasie.

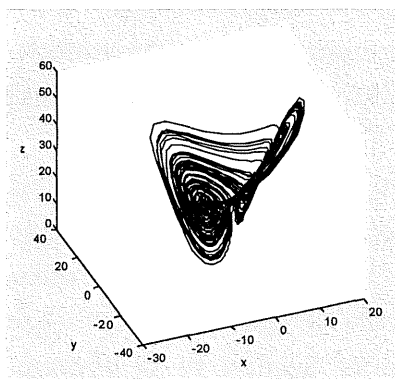
Przykładem jest układ równań Lorenza:

$$dx / dt = 10z(y - x)$$

$$dy / dt = -xz + 28x - y$$

$$dz / dt = xy - 8/3z$$

R. 1



Rycina 1. Przebieg zmiennych x, y, z z układu równań Lorenza będący rozwiązaniem tego układu.

Jak widać na rycinie 1, rozwiązaniem tego deterministycznego układu równań jest przebieg 3 zmiennych w czasie, który przebiega podobnie w kolejnych quasi-cyklach atraktora, jednak nigdy nie jest to przebieg identyczny z takim, który pojawił się w przeszłości.

Przebieg każdej ze zmiennych reprezentuje w pewnym przybliżeniu złożoność całego układu. Jedną z miar złożoności sygnału chaotycznego jest tzw. **Wymiar Korelacyjny** (d , *Correlation Dimension*). Pojęcie to jest uogólnieniem pojęcia "wymiaru" stosowanego w życiu codziennym. Prosta ma wymiar $d=1$, koło $d=2$ a kula $d=3$. W przypadku obiektów fraktalnych lub nieliniowych wymiar może przyjmować również wartość ułamkową. Definiuje się go jako:

$$d = \lim_{r \rightarrow 0} (\log M / \log r) \quad \text{R.2}$$

gdzie M to "masa" układu a r to jego rozmiar liczony w pojedynczym wymiarze. Przykładowo, masa kuli przyrasta z trzecią potęgą promienia. Pojęcie „masa” zostało umieszczone w cudzysłowie, gdyż jego znaczenie jest płynne. W zależności od kontekstu może to być odległość liniowa, powierzchnia, objętość, masa, a nawet ilość informacji w

bitach. Zastosowanie tej definicji do analizy sygnałów wymusza przyjęcie określonej definicji wartości M i r . Zależność $M(r)$ w tym przypadku określana jest jako liczba par punktów, dla których odległość Euklidesowa jest mniejsza lub równa r . Przez „punkt” nie rozumie się jednak wartości pojedynczej próbki badanego sygnału, lecz punkt w przestrzeni fazowej, posiadający m współrzędnych, które to współrzędne utworzone są przez m kolejnych wartości sygnału oddalonych wzajemnie o określony odstęp L (Lag). Parametr m to tzw. **Wymiar Zanurzenia** (*Embedding Dimension*). Długość fragmentu sygnału tworzącego pojedynczy punkt m -wymiarowy to **szerokość okna** $W=(m-1)L$.

Powyższy atraktor Lorenza ma złożoność równą w przybliżeniu $d=2.02$, gdyż jeśli spojrzymy na niego pod odpowiednim kątem w przestrzeni 3-wymiarowej, to zobaczymy, że punkty układają się w przybliżeniu na płaszczyźnie z niewielkim tylko rozchyleniem "skrzydeł" atraktora w trzecim wymiarze.

Dostępne w literaturze algorytmy (Grassbergera-Proccacia i Takensa-Ellnera) [1-3] wyznaczania wymiaru korelacyjnego dawały błędne wyniki w przypadku sygnałów wysokowymiarowych, czyli o złożoności przekraczającej $d=4$, co związane było z licznymi uproszczeniami i błędami w prezentowanych algorytmach. Publikacje przedstawiały jedynie wyniki działania algorytmów w stosunku do sygnałów o niskiej złożoności, na których działały one w miarę poprawnie.

W swojej rozprawie doktorskiej [4] próbowałem zastosować powyższe algorytmy w stosunku do sygnałów posturograficznych reprezentujących przemieszczanie się środka nacisku ciała człowieka na podłoże w trakcie spokojnego stania. Jednak uzyskane oszacowania wymiaru korelacyjnego nasuwały szereg wątpliwości dotyczących prawidłowości tego oszacowania. Sygnały te są bowiem bardzo złożone a ich złożoność obecnie szacuję na ok. $d=5-7$.

(Wzorcowym sygnałem o złożoności ok. $d=5$, jakkolwiek nie-chaotycznym, może być sygnał utworzony przez dodanie 5 sinusoid o równych amplitudach, zbliżonych częstotliwościach i niewymiernym stosunku częstotliwości, np. $f_i = 1 : \sqrt{2} : \sqrt{3} : \sqrt{5} : \sqrt{7}$)

Kontynuowałem więc swoje badania nad prawidłowością wyznaczania wymiaru korelacyjnego dla sygnałów wysokowymiarowych. Przeprowadziłem dużą liczbę różnych eksperymentów obliczeniowych mających na celu ustalanie przyczyn błędów.

Moja praca habilitacyjna omawia liczne przyczyny niedokładności oszacowania dla sygnałów wysokowymiarowych oraz przedstawia prawidłowe sposoby przeprowadzania obliczeń.

Drugi ważny aspekt mojej pracy to optymalizacja czasu obliczeń oraz oszacowywanie błędu obliczeń. Czas obliczeń rośnie **wykładniczo** w miarę wzrostu złożoności sygnałów. Bazując na wynikach prezentowanych w Tabeli I w publikacji [5] można go oszacować jako $T(d) \approx (2.5 \pm 0.5) \cdot 10^d$. W przypadku konieczności oszacowania złożoności rzędu $d=10$ staje się on niezmiernie wysoki, co zmusza do poszukiwania metod heurystycznych, redukujących czas obliczeń kosztem zmniejszenia dokładności oszacowania. Przeprowadzone analizy zależności między błędem a czasem obliczeń wykazują, że możliwa jest wykładnicza redukcja czasu obliczeń przy liniowym wzroście błędu. Inaczej mówiąc, jeśli przyjmiemy określony,

stały czas obliczeń, jaki chcemy zastosować, to błąd będzie przyrastał w przybliżeniu liniowo w miarę wzrostu wyznaczonej złożoności. Algorytm, który zaproponowałem, umożliwia więc oszacowanie złożoności w przewidywalnym i akceptowalnym czasie obliczeń a dokładność dopasowuje się automatycznie do zadanego czasu.

Większość najważniejszych modyfikacji opisana jest w moich publikacjach [5-8]. Pakiet funkcji w środowisku Matlab HDS-toolkit2 będący wynikiem mojej wieloletniej pracy dostępny jest na mojej stronie: www.drmichalak.pl/chaos/eng/.

Po opracowaniu algorytmu zastosowałem go w stosunku do posiadanych sygnałów posturograficznych. Zbadałem wpływ wzroku na złożoność sygnałów posturograficznych [9]. Można się było spodziewać, że zamknięcie oczu spowoduje wyłączenie części sprzężeń regulacyjnych kontrolujących balans ciała, przez co układ regulujący postawę stanie się prostszy czyli mniej złożony. Po usunięciu dryfu środka masy ciała z sygnału posturograficznego, który w istotny sposób zaburzał analizę, uzyskałem wyniki, w których powyższe założenie się potwierdziło. Dla składowej *przód-tył* wymiar korelacyjny jest o ok. $\Delta d=0.8$ mniejszy przy oczach zamkniętych ($d_{otw}\approx 6.3$, $d_{zamk}\approx 5.5$). Powyższe badanie otwiera drogę do dalszych badań nad złożonością sygnałów posturograficznych w różnych jednostkach chorobowych. Można się spodziewać, że zmniejszanie się wymiaru korelacyjnego sygnałów posturograficznych będzie miernikiem stopnia patologii układu balansu ciała człowieka.

Metoda oszacowywania złożoności sygnałów nieliniowych może mieć zastosowanie w wielu innych dziedzinach nauki, w których mamy do czynienia z układami chaotycznymi: w fizyce, chemii, biologii, ekonomii, meteorologii itp.

Algorytm Grassbergera-Proccacia i Takensa-Ellnera

Gdy przystępowałem do pracy nad wyznaczeniem złożoności sygnałów w 1998 roku, dostępne były w literaturze 2 algorytmy: Grassbergera-Proccacia (GP) [1] oraz Takensa-Ellnera (TE) [2, 3].

W algorytmie GP wydzielić można następujące kroki:

- A. Zanurzenie sygnału w przestrzeni m -wymiarowej (przy określonych wartościach m , L , W);
- B. Wyznaczenie odległości Euklidesowych między parami punktów w przestrzeni m -wymiarowej;
- C. Wyznaczenie histogramu odległości (H) oraz jego sumy kumulatywnej (H_c);
- D. Zobrazowanie H_c w skali logarytmicznej (całka korelacyjna, $C(r)$) i poszukiwanie **odcinka skalowania liniowego** (LSR, Linear Scaling Region) na tym wykresie. Współczynnik kierunkowy tego odcinka to wymiar korelacyjny d_w wyznaczony dla określonej kombinacji L , m i W . Można również wyznaczyć pochodną z $C(r)$ i odczytać wymiar korelacyjny jako wysokość poziomego odcinka na pochodnej $C(r)'$.

E. Powyższa procedura powtarzana jest dla kolejnych kombinacji parametrów zanurzenia L , m , W . Jako ostateczną wartość wymiaru korelacyjnego d przyjmowano wartość plateau na wykresie $d=fn(L)$, jeśli równocześnie obserwowana była saturacja wartości d po przekroczeniu wymiaru zanurzenia $m^* > 2d+1$.

Algorytm Takensa-Ellnera (TE) różni się od powyższego sposobem wyznaczenia wartości d_w . Wszystkie wyznaczone odległości pomiędzy punktami w przestrzeni fazowej są sortowane rosnąco (wektor δ), a następnie odpowiednio wyprowadzony przez autorów algorytm wzór wyznacza wymiar korelacyjny bezpośrednio z tego wektora, z pominięciem punktów C i D algorytmu GP. Algorytm ten jest znacznie wydajniejszy, jednak autorzy zaproponowali błędny przedział indeksów wektora δ , który miałby reprezentować obszar skalowania liniowego, co w przypadku sygnałów wysokowymiarowych generowało bardzo dużą niedokładność.

Powyższe algorytmy zawierały dwa główne punkty, które wymagały odgórnego, intuicyjnego przyjęcia przez badacza wartości parametrów, które w istotny sposób wpływały na wyniki, szczególnie w przypadku analizy sygnałów wysokowymiarowych:

1. określenie parametrów zanurzenia (m , L , W);
2. określenie przedziału skalowania liniowego na całej korelacyjnej (LSR) do odczytu d_w ;

Oba te punkty zostały szczegółowo przeanalizowane. W toku analiz zaproponowane zostały rozwiązania, które doprowadziły do samodzielnego określania przez algorytm optymalnych parametrów, które wcześniej musiały być ustawiane ręcznie przez badacza. Usunięte też zostały błędy metodyczne oraz określone zostały wzory pozwalające oszacować dokładność pojedynczego pomiaru d_w .

2. Predictability of posturographic signals: Influence of window width and embedding dimension,

Michalak KP, Jaśkowski P, Biocybernetics and Biomedical Engineering, 2007, Vol.27, Iss.4, pp.71-83, MNiSZW: 6

Sygnal posturograficzny to 2-wymiarowa ścieżka przemieszczającego się środka nacisku ciała na podłoże w trakcie spokojnego stania lub wykonywania różnych zadań. Posiada on składowe AP (przód-tył) i LAT (lewo-prawo). Jednym z początkowych celów badań w zakresie analizy nieliniowej sygnałów posturograficznych było określenie, w jakim stopniu sygnały te są wysoko złożonymi sygnałami chaotycznymi o bardzo złożonej ale jednak mierzalnej strukturze, a w jakim stopniu są to sygnały stochastyczne, czyli przypadkowe, nie posiadające żadnej wewnętrznej złożonej struktury. Przedstawiona w ww. publikacji analiza przewidywalności (predykcji) sygnałów wskazała, że przybliżona złożoność sygnałów¹

¹ Złożoność wymiarowa to pojęcie opisujące stopień złożoności badanego sygnału. Jedną z miar złożoności wymiarowej jest wymiar korelacyjny (correlation dimension). Najprościej można sobie wyobrazić to pojęcie, jeśli wyobrazimy sobie sygnał otrzymany przez dodanie do siebie 5 sygnałów 1-wymiarowych, np. sinusoid o

posturograficznych wynosi ok. $d \approx 5$, gdyż przy wzroście wymiaru zanurzenia² do $m=5$ następuje poprawa predykcji sygnałów, co oznacza, że w zakresie do $m=5$ kolejny dodany wymiar zanurzenia daje dodatkową informację o strukturze sygnału. Dalszy wzrost wymiaru zanurzenia nie powoduje poprawy predykcji. Dla dużych wartości szerokości okna³ W wzrost wymiaru zanurzenia daje wręcz pogorszenie predykcji, co może oznaczać wysokie prawdopodobieństwo istnienia składowej stochastycznej w sygnałach posturograficznych. Może też taki wynik wynikać z braku korelacji między początkiem i końcem odcinka sygnału o długości W , co pogarsza predykcję. Tym nie mniej w całym zakresie analizowanych parametrów m i W sygnały posturograficzne cechowały się lepszą predykcją niż odpowiadające im strukturą widmową porównawcze sygnały stochastyczne, co wskazuje na istotny element deterministyczny w tych sygnałach.

W publikacji tej po raz pierwszy zastosowana została **metoda** badania predykcji [10-12] z **wykorzystaniem stałej wartości szerokości okna dla różnych wartości wymiaru zanurzenia**, co jest jednym z kluczowych elementów prawidłowego podejścia metodycznego w przypadku sygnałów wysokowymiarowych. Odstęp między próbkami sygnału⁴ (lag, L) był wyznaczany wzorem $L = W/(m-1)$. Wcześniej, przy wzroście wymiaru zanurzenia stosowano wzrastającą szerokość okna zachowując stały odstęp między próbkami: $W = L \cdot (m-1)$. Tego typu podejście było prostsze metodycznie, gdyż nie pojawiał się problem ułamkowych odstępów między próbkami L . Dawało ono w miarę poprawne wyniki jedynie dla sygnałów niskowymiarowych cechujących się długimi czasami predykcji, gdzie wpływ wydłużającej się szerokości okna, który pogarsza predykcję, nie od razu był zauważalny. Uzyskanie praktycznie prawie identycznych wyników dla stałej, określonej szerokości okna i różnych wymiarów zanurzenia wskazuje, że szerokość okna jest bardzo istotnym parametrem przy wszelkich analizach sygnałów wysokowymiarowych. Wszystkie dalsze badania i publikacje bazowały na tej obserwacji. Korekta tego błędu metodycznego wprowadzonego przez wcześniejszych badaczy była jednym z kluczowych elementów do dalszych badań nad właściwościami nieliniowymi sygnałów chaotycznych.

3. Distinguishing Separate Components in High-Dimensional Signals by Using the Modified Embedding Method and Forecasting,

Michalak KP, *Annals of Biomedical Engineering*, 2010, Vol. 38, Iss. 1, pp. 200-207, IF: **2,374**, MNiSW: 30

W dalszym etapie badań nad sygnałami posturograficznymi próbowałem określić, w

niewymiernym stosunku częstotliwości (np. $f_1=1, f_2=\sqrt{2}, f_3=\sqrt{5}, f_4=\sqrt{6}, f_5=\sqrt{7}$) i równych amplitudach. Złożoność wymiarowa takiego sygnału wynosi wówczas ok. $d=5$.

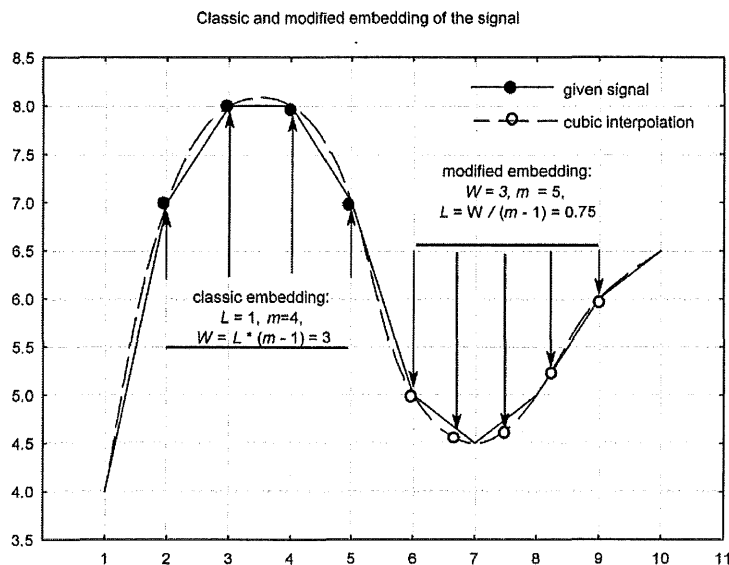
² Wymiar zanurzenia m to ilość kolejnych, równo oddalonych punktów/próbek badanego sygnału, które traktowane są jako kolejne współrzędne pojedynczego m -wymiarowego punktu w przestrzeni m -wymiarowej.

³ Szerokość okna W to długość odcinka sygnału, który tworzy w analizie jeden punkt w przestrzeni m -wymiarowej. $W=(m-1)L$

⁴ Odstęp między próbkami L to odległość między kolejnymi próbkami sygnału tworzącymi jeden punkt w przestrzeni m -wymiarowej. Szerokość okna to iloczyn odstępu i wymiaru zanurzenia $W=(m-1)L$.

jakim stopniu możliwe jest wyróżnienie i rozseparowanie w sygnałach posturograficznych osobnych składowych o prostszej strukturze, które sumują się do wysoko złożonego sygnału końcowego rejestrowanego przez posturograf. Ideą tego podejścia było założenie, że sygnał posturograficzny jest wypadkową prostszych w swej strukturze odruchów działających na różnych piętrach ciała (głowa/klatka/brzuch/biodra/kończyny górne/stopy). Początkowo do tego zadania próbowałem wykorzystać zastosowaną wcześniej metodę predykcji.

Efektom tych badań jest powyższa publikacja, w której wykazałem możliwość zastosowania predykcji do wyróżniania poszczególnych składowych sygnałów. Analizie poddałem sygnały o znanych strukturach, będące sumą 3 oraz 4 niskowymiarowych sygnałów Lorentza o złożoności ok. $d=2.017$, które posiadały różne skale czasowe (czyli strukturze zbliżonej do tej, jaką można by się spodziewać w sygnałach posturograficznych). Tak więc po zsumowaniu 3 lub 4 sygnałów o złożoności ok. $d \approx 2$ uzyskano sygnały o złożoności odpowiednio ok. $d \approx 6$ i $d \approx 8$. W publikacji tej po raz pierwszy zaproponowałem zastosowanie **interpolacji funkcjami sklejanymi**⁵ (*cubic interpolation*) do wyznaczania pośrednich wartości badanego sygnału dla ułamkowych wartości indeksu próbki (ułamkowy odstęp L wynikający ze wzoru $L=W/(m-1)$ dla określonych wartości W i m). Zastosowanie interpolacji rozwiązało problem konieczności stosowania nadmiernie wysokiej częstotliwości próbkowania sygnału w celu uzyskania w przybliżeniu identycznych wartości szerokości okna W dla różnych wymiarów zanurzenia m i odpowiednio dobranej całkowitej wartości L . Wcześniej nie stosowano w prezentowanych algorytmach tej metody matematycznej.



Ryc.2. Prezentacja idei zanurzenia sygnału w przestrzeni m -wymiarowej. Oś OX przedstawia 10 kolejnych próbek sygnału. Po lewej: klasyczne zanurzenie w przestrzeni $m=4$ -wymiarowej, odstęp $L=1$, szerokość okna W

⁵ Interpolacja funkcjami sklejanymi to sposób interpolacji odtwarzający wartości pomiędzy próbkami sygnału poprzez dopasowanie do każdego odstęp między próbkami wielomianu 3 stopnia, tak by przechodził on przez 2 analizowane punkty oraz aby pochodne w tych punktach dla dwóch sąsiednich wielomianów były identyczne. Gwarantuje to ciągłość pochodnej dla takiego dopasowania. Zarówno wartość sygnału jak i szybkość zmiany sygnału zmieniają się w sposób ciągły, nieskokowy.

= 3; po prawej - interpolacja funkcjami sklejanymi $m=5$, $L = 0,75$, $W = 3$. Kolejnych m próbek sygnału tworzy kolejne współrzędne punktu m -wymiarowego. Parametr J procedury zanurzenia będący odległością między początkami dwóch kolejnych punktów m -wymiarowych wynosi w bieżącym przykładzie $J = 6-2 = 4$.

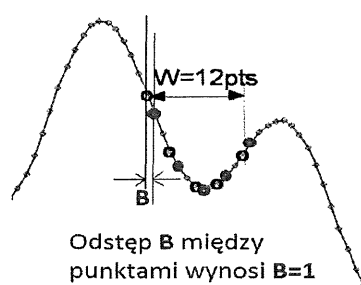
4. Modifications of the Takens-Ellner algorithm for medium- and high-dimensional signals

Michalak KP, Physical Review E, 2011, Vol. 83, Iss. 2, pp. 026206, IF: 2,255, MNiSZW: 35

W powyższej publikacji omówione zostały różne błędy w dotychczas stosowanych algorytmach oraz przedstawione zostały propozycje rozwiązań różnych problemów, które zostały wykryte:

Problem odstępu B

Podstawą wyznaczenia złożoności wymiarowej jest analiza rozkładu odległości Euklidesowych pomiędzy punktami w m -wymiarowej przestrzeni fazowej. Jeśli porównamy pary punktów sygnału (lub punkty z przestrzeni fazowej po zanurzeniu w przestrzeni m -wymiarowej), które leżą bardzo blisko siebie, to będą one silnie skorelowane. Nie będą one reprezentować struktury nieliniowej, a jedynie autokorelację sygnału. Stąd konieczność odrzucenia odległości między takimi parami punktów z algorytmu. Dotychczasowa propozycja w literaturze sugerowała użycie wartości minimalnego odstępu B między punktami, którą reprezentuje wzór: $B = \tau (2/N)^{2/m}$. (N – liczba próbek sygnału, τ – czas autokorelacji sygnału). Wykazałem, że wyliczona w ten sposób wartość jest zbyt mała w przypadku sygnałów wysokowymiarowych, co generuje istotne błędy w oszacowaniu d . Wykazałem, że konieczne jest użycie większych wartości odstępu B , przy czym przyjęcie nadmiernie dużej wartości B nie generuje istotnego błędu metodologicznego. Wydaje się, że wartość drugiego zera funkcji autokorelacji sygnału jest w zupełności wystarczająca jako wartość używanego w algorytmie odstępu B .



Rycina 3. Przykład prezentujący odległość między 2 kolejnymi punktami w przestrzeni fazowej, która jest zbyt mała. Dwa 5-wymiarowe punkty leżą bardzo blisko siebie w 5-wymiarowej przestrzeni. Odległość między takimi punktami mocno zaburza pomiar d i musi być odrzucona z analizy.

Problem optymalnej szerokości okna W (W_{opt}) dla wyznaczenia złożoności sygnału.

W przypadku sygnałów niskowymiarowych problem jest łatwy do oszacowania, ponieważ zbliżona wartość wymiaru korelacyjnego d wyliczana jest poprawnie dla dość szerokiego zakresu W . Stwierdzenie plateau na wykresie $d_w=fn(W)$ ⁶ jest zasadniczo jednoznaczne z określeniem optymalnego zakresu W , dla którego wyznacza się prawidłowo złożoność sygnału.

W miarę wzrostu złożoności sygnału, szerokość plateau w tej funkcji staje się jednak coraz węższa, w końcu zanika, a optymalna szerokość okna reprezentowana jest przez punkt o najmniejszej pochyłości na wykresie $d_w=fn(W)$. Bazując na sygnałach wzorcowych o znanej złożoności wykazałem w bieżącej publikacji, że punkt o najmniejszej pochyłości to punkt, dla którego wyznaczona złożoność jest bardzo bliska wartości teoretycznej.

Zważywszy, że zależność $d_w=fn(W)$ jest zasadniczo monotonicznie rosnąca, zaproponowałem sposób na wyznaczenie punktu (W_{opt} , d_{opt}), który reprezentuje punkt odczytu wymiaru korelacyjnego sygnału. Do wyznaczonych punktów relacji $d_w=fn(W)$ zastosowałem **dopasowanie wielomianu 4 stopnia**. Wielomian taki cechuje się tym, że jego pochodna – wielomian 3 stopnia – posiada co najwyżej jedno minimum lokalne. Stwierdzenie posiadania przez wielomian pochodny minimum lokalnego oznacza istnienie punktu o najmniejszej pochyłości będącego podstawą wyznaczenia złożoności wymiarowej d . Końcowa wersja pakietu HDS-Toolkit umożliwia dopasowanie wielomianów wyższego stopnia oraz innych samodzielnie zdefiniowanych funkcji.

Rozwinięcie problemu interpolacji sygnału.

Zastosowana metoda interpolacji sygnału funkcjami sklejanymi w przypadku konieczności wyznaczenia wartości sygnału dla niecałkowitych wartości indeksu próbki sygnału spotkała się z uwagami recenzentów pracy dotyczącymi możliwości generowania błędów pomiaru złożoności w skutek zastosowania takiej interpolacji. Zaproponowałem więc metodę oszacowania błędu interpolacji dopasowaną do bieżącego problemu. Metoda polega na wyznaczeniu błędu interpolacji dla różnych parzystych wartości współczynnika redukcji częstotliwości próbkowania sygnału s i ekstrapolowaniu tej zależności dla redukcji zerowej ($s=1$). Metoda ta umożliwiła precyzyjne oszacowanie błędu interpolacji, a w połączeniu z analizą wpływu szumu na poprawność wyliczeń d , wykazała że w przypadku zastosowanych sygnałów wzorcowych błąd interpolacji był pomijalny. Zastosowana metoda oszacowania błędu interpolacji może być powszechnie stosowana w tym celu w innych tego typu analizach.

⁶ d_w to wyliczona wartość złożoności wymiarowej przy użyciu określonej wartości szerokości okna W . Jeśli użyta wartość W jest zbyt duża, d_w jest większe niż rzeczywista wartość złożoności wymiarowej sygnału d_{opt} , a jeśli W jest zbyt małe, $d_w < d_{opt}$.

Potwierdzenie efektu saturacji d_w dla $m > m^*$ ⁷ przy użyciu stałej szerokości okna W

W literaturze naukowej podawana jest reguła, że wymiar zanurzenia m użyty do obliczeń musi być ponad dwukrotnie większy niż wyznaczana złożoność sygnału ($m^* = 2d_w + 1$, m^* to minimalny wymiar zanurzenia niezbędny dla prawidłowości obliczeń, d_w – wartość złożoności wymiarowej oszacowana przy użyciu określonej szerokości okna W). W przypadku wyznaczania złożoności d_w dla określonego W i rosnących wartości wymiaru zanurzenia m powinno się obserwować saturację wyznaczanego d_w po przekroczeniu użytego wymiaru zanurzenia $m^* = 2d_w + 1$. Tego efektu jednak nie obserwowano dla sygnałów wysokowymiarowych, gdy d wyznaczone było przy stałej wartości odstepu L , a więc rosnącej wartości W . Podobnie jak przy wyznaczaniu predykcji, również złożoność wymiarowa ulegała saturacji jedynie, gdy stosowano stałą wartość szerokości okna W dopasowując odstęp L wzorem $L = W/(m-1)$. Saturacja pojawiała się w przybliżeniu przy użyciu $m > 2d_w + 1$, co odpowiada wartości spodziewanej.

Stwierdzenie efektu saturacji po przekroczeniu dwukrotności wartości wymiaru zanurzenia $m^* = 2d_w + 1$ pozwoliło uprościć algorytm i wykonywać obliczenia dla jednej, odpowiednio wysokiej wartości wymiaru zanurzenia m , która musi być większa niż dwukrotna spodziewana złożoność sygnału. Wcześniejsze algorytmy nakazywały powtarzać obliczenia dla rosnących wartości m i poszukiwać saturacji. Ta jednak dla sygnałów wysokowymiarowych się nie pojawiała... Konieczność powtórzenia obliczeń przy wyższym wymiarze zanurzenia m może pojawić się jedynie wtedy, gdy oszacowana złożoność okaże się być mimo wszystko większa niż połowa wymiaru zanurzenia (czyli $d_w > (m-1)/2$). Taka sytuacja może się teoretycznie pojawiać w przypadku analizowania sygnałów stochastycznych, czyli przypadkowych.

Problem optymalnego P_{max} ⁸

Parametr P_{max} to względny (czyli zawarty w zakresie $(0,1)$) indeks do posortowanego rosnąco wektora δ przechowującego odległości pomiędzy m -wymiarowymi punktami w przestrzeni fazowej. Określa on koniec obszaru skalowania liniowego w całości korelacyjnej⁹,

⁷ m^* to najmniejszy wymiar zanurzenia, który trzeba użyć, by prawidłowo wyznaczyć złożoność danego sygnału. Musi być ponad dwukrotnie większy niż wyznaczana złożoność. Wyznaczany jest wzorem $m^* = 2d_w + 1$.

⁸ Przykładowo: jeśli długość posortowanego rosnąco wektora δ wynosi $N=10000$ a koniec obszaru skalowania liniowego obserwowany jest przy indeksie $I_{max}=200$, to $P_{max} = I_{max}/N = 0,002$.

⁹ Algorytm wyznaczania złożoności d_w można w uproszczeniu przedstawić następująco:

- Zanurz sygnał w przestrzeni m -wymiarowej (wyznacz punkty m -wymiarowe z fragmentów sygnału)
- Wyznacz odległości Euklidesowe pomiędzy zdefiniowanymi punktami m -wymiarowymi i posortuj go rosnąco (wektor δ)
- Wyznacz histogram rozkładu odległości między punktami
- Wylicz całkę tego histogramu (czyli sumę kumulacyjną wektora przechowującego histogram)
- Przedstaw tę całkę w układzie logarymicznym (=całka korelacyjna/Correlation Integral)
- Poszukaj odcinka prostego na całości korelacyjnej
- Wyznacz współczynnik kierunkowy tego odcinka = złożoność sygnału d .

który to obszar w logarytmicznym układzie współrzędnych jest w przybliżeniu linią prostą, a współczynnik kierunkowy tego odcinka jest szacowanym wymiarem korelacyjnym. Jest to więc ważny parametr przy wyznaczaniu złożoności sygnału. Opublikowany algorytm TE definiował tę wartość tego parametru odgórnie ($P_{max}=0.22$) bez wyjaśnienia takiego przyjęcia. Z kolei algorytm GP wskazywał na konieczność samodzielnego oszacowania obszaru skalowania liniowego, co wiązało się ze znacznymi problemami natury formalnej, gdyż wymagało intuicyjnego przyjęcia granicy przejścia linii w przybliżeniu prostej w łagodny łuk.

W bieżącej pracy wykazałem, że w przypadku sygnałów wysokowymiarowych wartość tego parametru jest znacznie mniejsza niż proponowana wcześniej $P_{max}=0.22$, która jest poprawna jedynie dla sygnałów niskowymiarowych $d < 3$. Mianowicie, dla analizowanych sygnałów wzorcowych o złożonościach $d \approx 4, 6$ i 8 poprawne oszacowanie d uzyskiwane było przy użyciu P_{max} rzędu 10^{-3} - 10^{-2} (czyli o 1-2 rzędy wielkości mniejszej niż proponowana). Przyjęcie do obliczeń proponowanej wartości 0.22 powodowało w przypadku sygnałów wysokowymiarowych bardzo istotne niedoszacowanie estymowanej złożoności i było zasadniczym błędem metodologicznym.

Dokładna analiza numeryczna wąskiego problemu szacowania P_{max} przedstawiona jest w następnej mojej publikacji dołączonej do bieżącego autoreferatu [5].

Problem P_{min}

Parametr P_{min} jest analogiczny do P_{max} i oznacza początek obszaru skalowania liniowego w całości korelacyjnej. Reprezentuje on najkrótsze odległości między m -wymiarowymi punktami w przestrzeni fazowej, które są odrzucane z analizy. Główne czynniki wpływające na optymalną wartość P_{min} to zaszumienie i błąd pojedynczej próbki sygnału. Przedstawione w ww. publikacji wyniki eksperymentów numerycznych wskazują, że w przypadku pomijalnego zaszumienia parametr ten w niedużym stopniu wpływa na poprawność oszacowania złożoności. Jeśli jednak sygnały cechują się istotnym zaszumieniem, P_{min} musi być odpowiednio wysoki, gdyż to właśnie te najkrótsze odległości między punktami w przestrzeni fazowej w największym stopniu ulegają wpływowi zaszumienia i powinny być usunięte z analizy. Niewłaściwy dobór P_{min} może istotnie zaburzać oszacowanie d w przypadku algorytmu Grassbergera-Proccacia, jednak w przypadku zmodyfikowanego przeze mnie algorytmu Takensa-Ellnera, wpływ zbyt małego P_{min} jest niewielki (szczególnie dla sygnałów wysokowymiarowych), gdyż liczba takich bardzo małych odległości jest bardzo niewielka w porównaniu do szybko narastającej liczby odległości większych.

Wpływ zaszumienia sygnału na poprawność oszacowania d

Realnie rejestrowane sygnały są zawsze zaszumione, stąd konieczność określenia zakresu poprawności działania algorytmu w zależności od poziomu zaszumienia sygnału.

Przeprowadzone analizy numeryczne wskazują, że w przypadku sygnałów o złożoności ok. $d = 6-8$ poziom szumu posiadającego moc mniejszą niż 1% mocy sygnału nie zakłóca istotnie mierzonej wartości wymiaru korelacyjnego. Szum może istotnie wpłynąć na kształt powierzchni funkcji $d = \text{fn}(P_{max}, W)$ będącej w ogólności podstawą do wyznaczania złożoności. W obrębie tej funkcji najbardziej wzrasta pod wpływem szumu oszacowanie d dla małych wartości P_{max} . Wynika z tego konieczność zastosowania odpowiednio wysokiej do poziomu szumu wartości P_{min} i odpowiednio większej wartości P_{max} , która jednak nie może wykraczać poza zakres dla poprawnego oszacowania d . (Kolejna prezentowana publikacja pokazuje jednak możliwość ominięcia tego problemu poprzez oszacowanie błędu dla zbyt wysokich P_{max} .) Zastosowanie zaawansowanych metod usuwania szumu istotnie poprawia poprawność metody.

5. How to estimate the correlation dimension of high-dimensional signals

Michalak KP, AIP Chaos, 2014, Vol. 83, Iss. 2, pp. 026206, IF: 1,954, MNiSZW: 40

Powyższa publikacja to kontynuacja analiz numerycznych związanych z poprawnością a także z optymalizacją numeryczną oszacowywania złożoności wymiarowej reprezentowanej przez wymiar korelacyjny.

Przedstawione zostały analizy następujących zagadnień i problemów numerycznych:

Porównanie algorytmu Grassbergera-Proccacia i Takensa-Ellnera.

W pierwszej kolejności przedstawione zostało porównanie dwóch podstawowych algorytmów wyznaczania wymiaru korelacyjnego, jakimi są algorytmy Grassbergera-Proccacia (GP) i Takensa-Ellnera (TE). Pierwszy z nich jest bardziej intuicyjny, gdyż funkcja całki korelacyjnej generowana jest w sposób bezpośredni, tzn. wyznaczany jest histogram odległości między punktami w przestrzeni fazowej, następnie jego suma kumulacyjna przedstawiana jest w skali logarytmicznej - jest to funkcja całki korelacyjnej. W końcu wyznaczany jest współczynnik kierunkowy odcinka skalowania liniowego na tej funkcji. Odcinek skalowania liniowego określony jest przez wartości $P_{min} - P_{max}$. Algorytm ten jest stosunkowo prosty, jednak w przypadku sygnałów wysokowymiarowych staje się całkowicie niewydolny. Przedstawione analizy numeryczne wykazały konieczność wyznaczania olbrzymiej liczby odległości między punktami (rzędu 10^{16}), które byłyby niezbędne do wyznaczenia tą metodą złożoności rzędu $d=8$.

Z drugiej strony - algorytm Takensa-Ellnera, który wyznacza d bardzo mało intuicyjnym wzorem, opisywany był w literaturze jako korzystający z zupełnie innego obszaru całki korelacyjnej niż algorytm GP, wykazywał silną, niewyjaśnioną tendencję do zaniżania estymowanej wartości d .

$$\begin{aligned}
 & \text{for } i = 1..D: \quad y(i) = \rho(I_{\max}) - \rho(I_{\min} + i - 1) \\
 & Y = \text{sum}(y(i)) \\
 & d = \frac{D}{Y + I_{\min} \cdot y(1)}
 \end{aligned}
 \tag{R.3}$$

Równanie 3. Wzór Takensa-Ellnera na wyznaczenie d z posortowanego wektora z odległościami δ . $\rho = \log(\delta)$ – logarytm posortowanego wektora z odległościami, $D = I_{\max} - I_{\min}$ – liczba analizowanych odległości.

Wyprowadzenie tego wzoru było bardzo skąpo omówione przez autorów w ich publikacji, co powodowało problemy ze jego zrozumieniem i interpretacją.

Szczegółowa analiza wyprowadzenia wzoru na algorytm TE wykazała, że korzysta on tak naprawdę z tego samego obszaru całki korelacyjnej, czyli obszaru skalowania liniowego reprezentowanego przez przedział $P_{\min} - P_{\max}$. Autorzy algorytmu błędnie jednak wskazali bez podania wyjaśnienia tego wyboru przedział $P_{\min} - P_{\max} = 0.02$ do 0.22 jako ten, z którego należy korzystać przy wyznaczaniu d .

W swojej publikacji wskazałem, że wykres posortowanego rosnąco wektora z odległościami $\delta(I)$ ¹⁰ (odległość δ na osi OY, indeks I wektora δ na osi OX) w skali logarymicznej reprezentuje odwróconą całkę korelacyjną. Oznacza to, że zamiana miejscami osi OX i OY przedstawia klasyczną całkę korelacyjną. Oś OY przedstawia liczbę I (procent P) odległości mających odległość mniejszą niż $\delta(I)$. Wykres ten jednak istotnie różni się od wykresu uzyskanego metodą klasyczną (GP). W algorytmie GP do wygenerowania pojedynczego punktu wykresu histogramu potrzeba olbrzymiej liczby odległości, natomiast w algorytmie TE każda odległość to 1 punkt na wykresie całki korelacyjnej. Uzyskuje się więc olbrzymi zysk czasu obliczeń i pamięci operacyjnej na wyznaczenie wykresu całki korelacyjnej.

Wyznaczenie wzoru na dokładność oszacowania wartości d_w

Algorytm TE przeznaczony był w swej idei do wyznaczania złożoności sygnału na podstawie małej liczby dostępnych odległości między punktami (D) w przestrzeni fazowej, co ma miejsce w przypadku sygnałów krótkich i niestacjonarnych. Mała liczba odległości będąca podzbiorem wszystkich możliwych odległości może być przyczyną błędu metodycznego oszacowania d . W oryginalnej wersji algorytmu TE proponowano, by oszacowania dokładności dokonywać poprzez powtórzenie pomiaru d j -razy w celu wyznaczenia średniej z tych pomiarów. Podejście to w swej istocie wydawało się niewiele różnić się od jednorazowego oszacowania wymiaru na podstawie j -razy większej liczby punktów. W celu

¹⁰ Zmienna δ opisuje w moich artykułach posortowany rosnąco wektor przechowujący wyznaczone odległości między punktami w m -wymiarowej przestrzeni fazowej. $I = 1, 2, 3, 4, \dots, \text{length}(\delta)$ to indeks do komórek tego wektora. Odpowiada on indeksowi względnemu $P = I / \text{length}(I)$ używanemu do wyznaczenia optymalnego P_{\min} i P_{\max} .

rozstrzygnięcia tego dylematu dokonałem szeregu analiz numerycznych, które doprowadziły do określenia jednoznacznego wzoru na dokładność d , w którym jedynym parametrem zmiennym była liczba odległości D wykorzystana do wyznaczenia d .

Analizy wykazały, że w przypadku losowego wybierania odległości między punktami, rozkład wartości d_j przy powtarzaniu pomiaru j -razy cechuje się rozkładem log-normalnym, a nie normalnym. Oznacza to, że rozkład normalny ma zmienna $\log(d_j)$. Dla różnych sygnałów i różnych wartości W użytych do zanurzania sygnałów wyznaczyłem więc odchylenia standardowe σ rozrzutu wartości $\log(d_j)$ w funkcji liczby użytych odległości $D = l_{max} - l_{min}$. Każdorazowo uzyskiwałem identyczną zależność. Można ją opisać równaniem:

$$\log(\sigma_{\log d_j}) = -0.5 \log(D) \quad \text{R.4}$$

Uwzględniając wzór na odchylenie standardowe zmiennej posiadającej rozkład log-normalny wyprowadziłem wzór na odchylenie standardowe zmiennej d_j :

$$\sigma_{d_j} = \sqrt{(e^{D^{-1}} - 1)e^{2\mu + D^{-1}}} \quad \text{R.5}$$

Powyższy wzór ma charakter uniwersalny, tzn. nie zależy od rodzaju analizowanego sygnału.

Wyznaczenie parametrów P_{max} i δ_{max} określających koniec obszaru skalowania liniowego całki korelacyjnej.

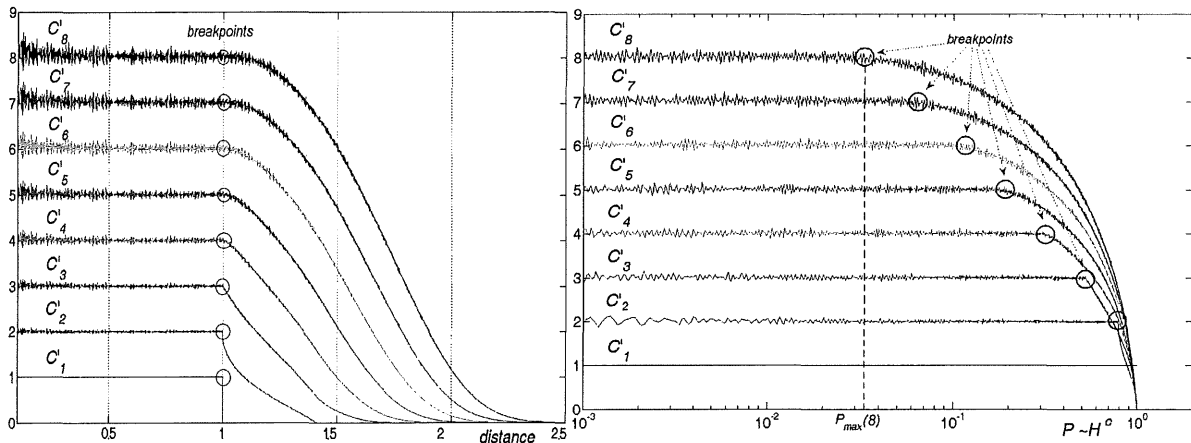
Bieżący problem został już wcześniej zasygnalizowany jako kluczowy dla prawidłowego wyznaczenia wymiaru korelacyjnego sygnału. w szczególności wyznaczenie δ_{max} jest bardzo ważne dla przyspieszenia algorytmu, gdyż umożliwia odrzucanie odległości dłuższych niż δ_{max} w jeszcze trakcie wykonywania obliczeń. Autorzy algorytmu TE zaproponowali ogólnie przyjęty przedział $P_{min}-P_{max} = 0.02 - 0.22$ jako właściwy dla wyznaczenia wymiaru korelacyjnego. We wcześniejszej pracy wykazałem, że użycie tego zakresu w stosunku do sygnałów wysokowymiarowych daje niepoprawne oszacowanie złożoności i że poprawne wartości pojawiają się przy użyciu znacznie niższych wartości P_{max} rzędu $10^{-3}-10^{-2}$. Pojawił się więc problem oszacowania, w jaki sposób zmienia się optymalna wartość P_{max} w miarę wzrostu wymiaru korelacyjnego analizowanego sygnału?

W pierwszej kolejności przeprowadziłem analizy numeryczne na zbiorach liczb losowych. Zanurzenie wektora liczb losowych w przestrzeni m -wymiarowej oznacza, że wyznaczona złożoność wymiarowa takiego ciągu musi być równa m . Zauważyłem, że jeżeli wyznaczymy histogram H_1 różnic między parami liczb losowych w zbiorze Z (co odpowiada wymiarowi zanurzenia $m=1$), to możemy z takiego histogramu łatwo wyznaczyć histogram H_2 odległości w przestrzeni 2-wymiarowej (odległości między losowymi punktami na płaszczyźnie). Można to zrobić wykonując rodzaj splotu histogramu H_1 z samym sobą. Jeśli użyjemy równych szerokości przedziałów histogramu zaczynających się od 0, to indeksy i i j splatanych histogramów reprezentują odległości a wartości pod tymi indeksami - prawdopodobieństwa p_i i p_j pojawienia się określonej odległości. Zgodnie z definicją odległości Euklidesowej, w przestrzeni 2-wymiarowej odległość 2 punktów na płaszczyźnie określona jest wzorem

$d = \sqrt{d_i^2 + d_j^2}$. Dlatego prawdopodobieństwo p_{ij} odległości $k = \sqrt{i^2 + j^2}$ w wynikowym histogramie musi zostać zwiększone o iloczyn $p_i p_j$ dla każdej pary i, j . W podobny sposób można wyznaczyć histogramy dla kolejnych wymiarów zanurzenia dokonując splotu histogramów niższego rzędu: $H_c = H_a \times H_b$, gdzie $c = a + b$. Przykładowo, histogram rzędu $m = 30$ można wyznaczyć przez kolejne sploty: $H_2 = H_1 \times H_1$, $H_4 = H_2 \times H_2$, $H_8 = H_4 \times H_4$, $H_{16} = H_8 \times H_8$, $H_{24} = H_{16} \times H_8$, $H_{28} = H_{24} \times H_4$, $H_{30} = H_{28} \times H_2$.

Wyzaczyłem więc histogramy kolejnych rzędów (H_m) odpowiadające zanurzeniu sygnału składającego się z liczb losowych we wzrastających wymiarach zanurzenia m . Dokonałem analiz dla liczb losowych o rozkładzie stałym (prostokątnym) z zakresu (0,1) i rozkładzie normalnym (0,1).

- Rozkład prostokątny liczb losowych (0,1) reprezentuje obszar skalowania liniowego całki korelacyjnej. Przeprowadzone analizy numeryczne wykazały, że jakkolwiek odległości Euklidesowe między punktami m -wymiarowymi rosną w miarę wzrostu m , to kolejne sploty takiego rozkładu posiadają zawsze koniec obszaru skalowania liniowego przy odległości $\delta_{max} = 1$. Jedynie coraz mniejszy procent wyznaczonych odległości mieści się w przedziale $\delta_{max} \leq 1$. (patrz ryc.3a,b). Podobne zjawisko, tylko w mniejszej skali, pojawia się przy rzucaniu kostkami zawierającymi 0 do 5 oczek. Jeśli największa liczba oczek na kostce to 5, to prawdopodobieństwo, że suma z m rzuconych kostek będzie mniejsza lub równa 5, jest coraz mniejsze w miarę wzrostu liczby rzuconych kostek.



Rycina 4 a) Pochyłości w skali log-log sum kumulatywnych histogramów kolejnych auto-splotów histogramu H_1 liczb losowych (0,1). Koniec obszaru skalowania liniowego obserwuje się zawsze dla odległości $\delta = 1$; b) zmieniona skala OX: wykres obrazuje, jaki ułamek wszystkich odległości odpowiada odległości $\delta = 1$. W miarę wzrostu wymiaru zanurzenia - coraz więcej odległości Euklidesowych między m -wymiarowymi punktami przekracza 1, a coraz mniej mieści się w przedziale (0-1).

Trzeba zauważyć, że dla sygnałów rzeczywistych sploty funkcji muszą być wykonane na rozkładach odległości między punktami (różnic między liczbami losowymi), a nie na rozkładach samego zbioru liczb losowych. Nie jest możliwe uzyskanie takiego zbioru liczb, by rozkład odległości (różnic) między liczbami miał rozkład prostokątny. W zależności od rodzaju rozkładu liczb losowych, w histogramie odległości uzyskuje się jedynie dłuższy lub krótszy

odcinek poziomy po jego lewej stronie odpowiadający małym różnicom. Przeprowadzone analizy numeryczne dla różnych rozkładów liczb wskazały, że stosunkowo najszerszy odcinek poziomy pojawia się dla zbioru liczb losowych posiadających rozkład normalny. Nigdy jednak nie jest to idealny odcinek poziomy. Rozkład ten, w miarę wzrostu odległości między punktami, maleje stopniowo do zera, na początku nieznacznie, następnie coraz szybciej. Podobną relację obserwuje się w rozkładach odległości dla splotów kolejnych rzędów. Główna różnica polega na tym, że w miarę wzrostu m znacznie szybciej maleje wartość P_{max} dla punktu granicznego (patrz Ryc. 5a).

Wyniki powyższych analiz wskazują, że z formalnego punktu widzenia dokładne oszacowanie wymiaru korelacyjnego nie jest możliwe, gdyż trudno jest uzyskać taki rozkład liczb losowych, by rozkład odległości między nimi posiadał idealne plateau po lewej stronie wykresu odpowiadające obszarowi skalowania liniowego. Dlatego też zaproponowałem wprowadzenie pojęcia 'deflection error' - e , czyli błąd odchylenia, reprezentujący akceptowalny poziom odchylenia od idealnego obszaru poziomego. W bieżącej pracy przedstawione zostały wykresy obrazujące powyższe zjawiska oraz równania pozwalające wyznaczyć parametr P_{max} , czyli ułamek odległości między punktami tworzący odcinek poziomy nie przekraczający zadanego odchylenia e . Przeprowadzone analizy wykazały, że logarytm z P_{max} maleje w przybliżeniu liniowo w miarę wzrostu wymiaru zanurzenia (odpowiadającego w przypadku zbioru liczb losowych jego złożoności). Przedstawiona też została zależność odległości δ_{max} między punktami w przestrzeni m -wymiarowej odpowiadającej określonymu odchyleniu $e\%$ jako funkcji wymiaru zanurzenia m . Odległość ta rośnie bardzo nieznacznie w miarę wzrostu m . Dalsze analizy dla rozkładu normalnego (omówione w następnej publikacji) wykazały, że dla wszystkich wartości $e < 30\%$, wartość δ_{max}^2 rośnie liniowo wraz z wymiarem zanurzenia m .

Jako ostatni punkt analizy przydatności wykorzystania wartości P_{max} i δ_{max} uzyskanych przy pomocy splotów rozkładu odległości porównano wyniki oszacowania P_{max} i δ_{max} dla rzeczywistego sygnału będącego sumą 4 sygnałów Lorenza posiadającego złożoność ok. $d=8$ i zanurzonego w przestrzeniach $m=16$, $m=32$ oraz $m=80$ oraz splotów funkcji uzyskanych z rozkładu różnic między próbkami tego sygnału dla tych samych wartości m . Należy powiedzieć, że w przypadku sygnałów chaotycznych histogramy odległości uzyskane metodą splotów nie odpowiadają dokładnie histogramom uzyskanym przez klasyczne zanurzenie sygnału ze względu na korelacje pomiędzy kolejnymi współrzędnymi punktów w klasycznym zanurzeniu, których to korelacji nie ma w przypadku liczb losowych. Jednak uzyskane wyniki wskazują na dużą zbieżność uzyskanych całek korelacyjnych dla niedużych błędów odchylenia e , co pozwala wykorzystać tę metodę do przybliżonego oszacowywania optymalnego P_{max} i δ_{max} .

Uzyskane wyniki eksperymentów obliczeniowych wskazują, że **wartość δ_{max} nie zależy od wartości wyznaczonej złożoności wymiarowej d_w , a jedynie od użytego wymiaru zanurzenia m i przyjętego błędu odchylenia e** . Możliwe jest więc oszacowanie wartości δ_{max} jeszcze przed rozpoczęciem wyznaczania złożoności wymiarowej, a następnie odrzucanie jeszcze w trakcie obliczeń wszystkich odległości większych niż δ_{max} . Zważywszy, że w przypadku

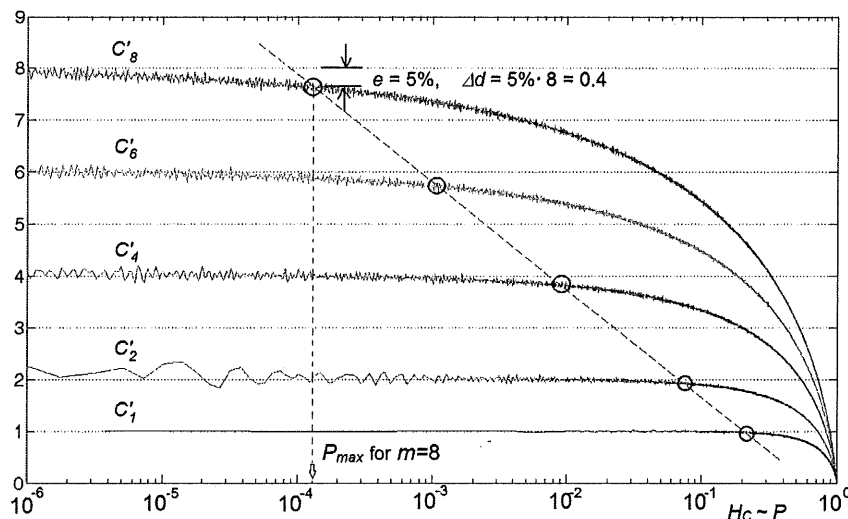
sygnałów o wysokiej złożoności tylko bardzo mały procent odległości spełnia to kryterium, uzyskuje się radykalne przyspieszenie obliczeń i redukcję zajętości pamięci operacyjnej, gdyż nie trzeba przechowywać olbrzymiej liczby odległości między punktami, które nie będą wykorzystywane do wyznaczenia złożoności sygnału.

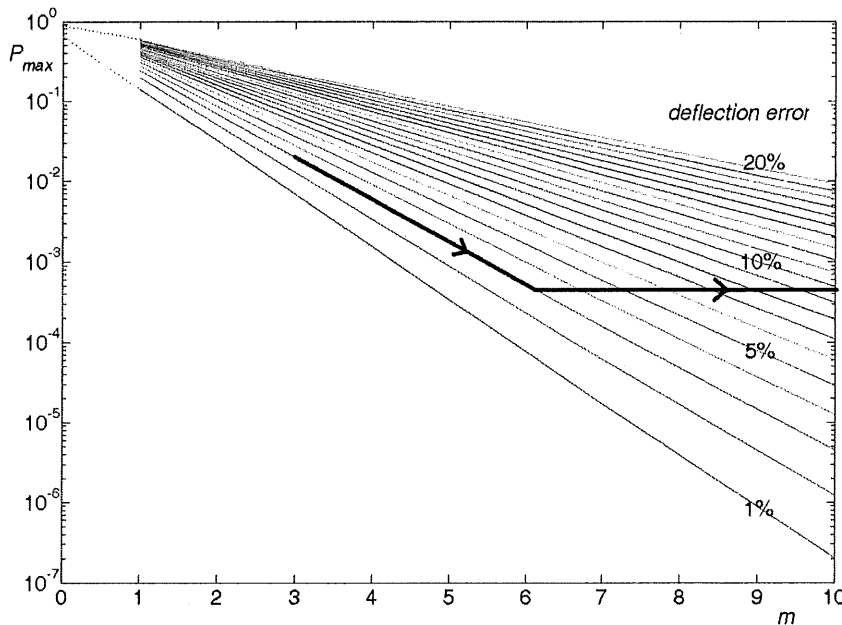
Z kolei parametr P_{max} mówiący, jaki ułamek odległości będzie tworzyć obszar skalowania liniowego, zależy od wyznaczonej złożoności wymiarowej d_w i przyjętego błędu odchylenia e .

Parametr ten jest użyteczny przy określaniu obszaru całki korelacyjnej wykorzystanej do ostatecznego wyznaczenia złożoności sygnału, jednak nie jest on do tego celu niezbędny.

Oszacowanie błędu odchylenia e dla uzyskanej metodą splotów całki korelacyjnej rzędu m umożliwia też skorygowanie uzyskanej metodą bezpośrednią całki korelacyjnej o wartość błędu w odpowiednich punktach. To z kolei umożliwia oparcie obliczeń o fragment całki, w którym błąd odchylenia jest większy, co daje bardzo dużą redukcję liczby odległości między punktami, które trzeba wyliczyć na potrzeby wyznaczenia złożoności. Uzyskuje się to kosztem nieznacznego tylko obniżenia dokładności wyliczeń. Taka metoda wyznaczenia złożoności jest szczególnie przydatna dla wysokich wartości szerokości okna, dla których d_w jest wysokie (często $d_w > 10$), jednak dokładne wyznaczenie nie jest aż tak istotne, gdyż punkty te są potrzebne jedynie do przeprowadzenia dopasowania wielomianu 4-stopnia do relacji $d_w = fn(W)$.

Zwiększanie błędu odchylenia e dla rosnących wartości szerokości okna W





Rycina 5. Sposób oszacowania P_{max} z wykresów histogramów zmiennej losowej o rozkładzie normalnym dla wzrastających wymiarów zanurzenia. (a) przykładowa prosta dla błędzie odchylenia $e=5\%$, (b) proste uzyskane dla innych wartości błędzie odchylenia e . P_{max} maleje wykładniczo w miarę wzrostu wymiaru zanurzenia, który dla zmiennej losowej jest tożsamy z wymiarem korelacyjnym. Czas obliczeń jest w przybliżeniu liniowo zależny od $1/P_{max}$. Tak więc po osiągnięciu maksymalnego akceptowalnego czasu obliczeń w kolejnych iteracjach nie wzrasta wykładniczo P_{max} lecz w przybliżeniu liniowo rośnie e . Strzałki na wykresie obrazują przykładowe działanie algorytmu dla wzrastających wartości W i używanego m : po osiągnięciu czasu obliczeń iteracji odpowiadającej przeanalizowaniu $N = D_a/P_{max}$ odległości, w kolejnych iteracjach P_{max} jest w przybliżeniu stałe, a wzrasta błąd odchylenia, przy którym dokonujemy estymacji d .

Tabela I. Równania przykładowych prostych zobrazowanych na Rycinie 5b.

Defl. err. (%)	dopasowane równanie $y=ax+b$
e=1	$\log_{10}(P_{max}) = -0.6497m - 0.2098$
e=5	$\log_{10}(P_{max}) = -0.4465m - 0.0932$
e=10	$\log_{10}(P_{max}) = -0.3262m - 0.0696$
e=20	$\log_{10}(P_{max}) = -0.1990m - 0.0569$

Jedną z kluczowych poprawek przyspieszających radykalnie czas obliczeń jest zaimplementowanie automatycznego wzrostu wartości dopuszczalnego błędzie odchylenia e przy wyznaczaniu d_w dla rosnących wartości szerokości okna. Dla szerokości okna dążącej do 0, wyznaczany wymiar korelacyjny dąży do $d_w=1$. W miarę wzrostu szerokości okna wyznaczone wartości d_w są coraz większe. Wiąże się to z coraz koniecznością użycia coraz większego wymiaru zanurzenia m oraz coraz mniejszej wartości P_{max} , czyli procentu odległości między punktami w przestrzeni m -wymiarowej, które spełniają kryterium włączenia ich do analizy. Rycina 5 przedstawia liniowe zależności między $\log(P_{max})$ i d dla różnych wartości błędzie odchylenia e wyznaczone dla krzywych reprezentujących pochyłości histogramów kolejnych rzędów uzyskanych metodą auto-splotów dla rozkładu zbioru liczb losowych o rozkładzie normalnym. Tabela I natomiast przedstawia przykładowe równania prostych dopasowanych do ww. punktów.

Ponieważ dla dużych wartości wyznaczanego wymiaru zanurzenia czas obliczeń T zależy głównie od wartości P_{max} ($T \sim 1/P_{max}$), wzrost d_w powoduje wykładniczy przyrost czasu obliczeń potrzebnego do wyznaczenia kolejnych punktów relacji $d_w = \text{fn}(W)$. Przekształcając równania z Tabeli I, dla wysokich wartości wyznaczanego d_w zależność czasu obliczeń od wymiaru korelacyjnego opisuje zależność $T(d) \approx (2.5 \pm 0.5) \cdot 10^d$. Czas ten można również wykładniczo skracać, jeśli zamiast wydłużania czasu obliczeń będziemy zwiększać sukcesywnie (w przybliżeniu liniowo) dopuszczalny błąd odchylenia e . Zaproponowany algorytm działa w następujący sposób:

I. Przed rozpoczęciem obliczeń definiujemy:

- początkową wartość błędu odchylenia e dla określenia δ_{max} (np. 3%)
- żadaną liczbę odległości (assumed D, D_a) krótszych niż δ_{max} , którą chcemy wyznaczyć na potrzeby wyznaczenia d_w .
- maksymalny czas obliczeń T_{MAX} dla znalezienia i wyliczenia żądanej liczby odległości D_a .

IIA. Jeśli znajdziemy D_a odległości w czasie T_{MAX} , przechodzimy do następnej iteracji d_w .

IIB. Jeśli w momencie osiągnięcia czasu T_{MAX} wyznaczonych zostanie więcej odległości niż $50\% \cdot D_a$ – kontynuujemy obliczenia (czas obliczeń nie powinien przekroczyć $2T_{max}$), ale w następnych iteracjach wyznaczenia d_w zwiększamy e o 1 %.

IIC. Jeśli w momencie osiągnięcia czasu T_{MAX} wyznaczonych zostanie mniej odległości niż $50\% \cdot D_a$ – przerywamy obliczenia i rozpoczynamy je od nowa ze zwiększoną o 1 % wartością e . Wzrost dotyczy wszystkich kolejnych iteracji.

III. Jeśli wartość e osiągnie 30% nie zwiększamy dalej e ze względu na zbyt duży błąd obliczeń. Wyliczamy d_w na podstawie mniejszej liczby punktów wyznaczonych w czasie T_{MAX} , przerywamy obliczenia lub wydłużamy czas obliczeń.

W praktyce algorytm działa w ten sposób, że po osiągnięciu maksymalnego czasu na iterację, w kolejnych iteracjach czas obliczeń nie rośnie, a maleje jedynie minimalnie dokładność obliczeń. Umożliwia to wyznaczenie relacji $d_w = \text{fn}(W)$ z wystarczająco dużą dokładnością w relatywnie krótkim czasie rzędu kilku-kilkudziesięciu minut (w zależności od złożoności sygnału, zadanej dokładności i szybkości komputera).

Warto zauważyć, że kilkukrotne nawet skrócenie T_{max} w niedużym stopniu zmniejsza dokładność obliczeń, dzięki czemu algorytm można zapuszczać w trybie szybkim, jeśli w bardzo krótkim czasie chcemy uzyskać orientacyjny przebieg relacji $d_w = \text{fn}(W)$.

Oszacowanie maksymalnej liczby powtórzeń procedury łączenia punktów w pary (*Random Joining Procedure, RJP*)

Następnym punktem publikacji było przedstawienie wzoru na oszacowanie prawdopodobieństwa powtarzania się tych samych par punktów wykorzystywanych do wyznaczenia całki korelacyjnej w trakcie powtarzania procedury łączenia punktów w pary (*Random Joining Porcedure, RJP*). Oryginalny algorytm TE proponuje jednorazowe połączenie n punktów z przestrzeni fazowej w $k=n/2$ par celem wyznaczenia wektora odległości δ , a następnie powtarzanie wyznaczenia wymiaru korelacyjnego dla innych kombinacji łączenia punktów w pary. W swojej wcześniejszej publikacji zaproponowałem powtarzanie większą

liczbę razy procedury łączenia w pary celem wyznaczenia większej liczby odległości między punktami, co poprawia dokładność obliczeń. Wiąże się to jednak z prawdopodobieństwem p_x powtarzania się par w kolejnych RJP. Przy pomocy analizy statystycznej wynikającej z rachunku prawdopodobieństwa wyprowadziłem wzór pozwalający oszacować procent powtórzeń par punktów wziętych do analizy. Trzeba powiedzieć, że jakkolwiek z formalnego punktu widzenia nie jest to poprawne, to jednak w minimalnym stopniu wpływa to na uzyskane wyniki. Wyznaczony wzór pozwala w prosty sposób oszacować skalę zjawiska i w przypadku takiej konieczności przejść na algorytm wyboru par punktów "każdy z każdym", tak jak to proponuje algorytm Grassbergera-Proccacia. Zaimplementowany algorytm domyślnie powtarza procedurę RJP aż do momentu wyznaczenia pożądanej liczby odległości D_a mniejszych niż δ_{max} . Umożliwia również odgórne zadeklarowanie wyznaczania odległości metodą "każdy z każdym" oraz automatyczne przejście na metodę "każdy z każdym" w przypadku przekroczenia akceptowalnej liczby powtórzeń RJP.

Kolejne wyprowadzone i przedstawione w publikacji wzory pozwalają oszacować:

$$R_{max} R_{max} \approx 2p_x / n \quad R. 6a$$

- maksymalną liczbę powtórzeń RJP, przy której procent powtórzonych par nie przekracza p_x

$$\Delta P^* = P_{max} - P_{min} = 2D_{min} / (n \cdot R_{max}) = D_{min} / (p_x \cdot (n - B)^2) \quad R. 6B$$

- minimalna szerokość przedziału, którą można uzyskać przy określonym założeniu p_x .

Powyższe wzory bazują na liczbie n wyznaczonych punktów w przestrzeni fazowej, którą można manipulować zmieniając parametr J w procesie zanurzania sygnału. Parametr ten oznacza odległość między początkami fragmentów sygnałów tworzących kolejne punkty w przestrzeni fazowej. Dzięki zastosowaniu interpolacji funkcjami sklejanymi może on przyjmować wartości ułamkowe. Przykładowo, jeśli użyjemy $J=0.25$ to otrzymamy w przybliżeniu czterokrotnie więcej punktów w przestrzeni fazowej niż jest próbek w analizowanym sygnale. Można więc powyższe wzory wykorzystać również do oszacowania wartości J , by uzyskać wystarczająco dużą liczbę punktów n . Tego rodzaju analizy mogą być szczególnie istotne w przypadku badania sygnałów krótkich, o małej liczbie próbek. W przypadku $J < 0.1$ mogą się jednak pojawiać błędy wynikające z interpolacji, dlatego nie rekomenduje się używania zbyt małych wartości J .

Oszacowanie maksymalnej złożoności wymiarowej d_{max} , którą można wyznaczyć zmodyfikowanym algorytmem

Maksymalny wymiar korelacyjny, który można realnie wyznaczyć, czyli górny limit zastosowania algorytmu zależy od kilku elementów. Wymienić tu należy błąd odchylenia e , minimalną akceptowalną liczbę odległości wziętych do analizy D_{min} i liczbę punktów w przestrzeni fazowej n . Parametr n możemy w rozsądnym zakresie zmieniać poprzez zmianę parametru J , jak opisano wcześniej. Mniejsza wartość D_{min} wpływa na mniejszą dokładność oszacowania d , przy czym dokładność oszacowania opisuje wzór opisany w podpunkcie (b). Użycie $D_{min}=100$ wydaje się być rozsądnym dolnym limitem zastosowania algorytmu.

Największy wpływ na oszacowanie limitu d_{max} ma parametr e . Jego zwiększanie do wartości 20-30% umożliwia oszacowanie złożoności sygnału, co prawda z malejącą dokładnością, nawet bardzo dużych złożoności rzędu $d=20-30$. Trzeba jednak zaznaczyć, że aby prawidłowo oszacować złożoność tak skomplikowanego sygnału, musi on być wystarczająco długi, by przejść przez wszystkie zakamarki 20-30-wymiarowego atraktora przynajmniej kilka razy, co w takim przypadku wymaga bardzo długiego czasu rejestracji. Wpływ długości sygnału na wartość szacowanej złożoności dla sygnałów o ekstremalnie wysokiej złożoności jest osobnym tematem badawczym. Jak pokazuje analiza wyprowadzonego wzoru, opracowany algorytm jest dobrym narzędziem do badania tego problemu.

6. Estimating correlation dimension of high-dimensional signals – quick algorithm

Michalak KP, AIP Advances 8, 105201 (2018), <https://doi.org/10.1063/1.5013255>,
IF 1,653 MNiSZW 25

Powyższa publikacja stanowi kontynuację i dokończenie analizy problemu wyznaczania współczynnika wymiaru korelacyjnego dla sygnałów wysokowymiarowych. Publikacja omawia szybki algorytm, w którym zminimalizowany został czas obliczeń przy zachowaniu zadanej dokładności, prezentuje przyczyny błędów numerycznych podstawowego wzoru Takensa-Elnnera i wprowadza szereg empirycznych poprawek zwiększających dokładność oszacowania. Prezentuje też dodatkowo problem analizy sygnałów o rozkładach mocno odbiegających od rozkładu gaussowskiego oraz przedstawia przykład sygnału posiadającego dwa obszary plateau funkcji $d_w=fn(W)$ dla różnych obszarów skalowania.

Dodanie możliwości dopasowania do relacji $d_w=fn(W)$ innych funkcji niż wielomian 4-tego stopnia

Wielomian 4-tego stopnia nie jest idealną funkcją do wykonywania dopasowania funkcji do wykresu $d_w=fn(W)$, dlatego też przygotowałem funkcje w środowisku Matlab, które umożliwiają dopasowanie dowolnego wielomianu wyższego stopnia, jak również dopasowanie innych własnych bardziej złożonych funkcji dopasowania. Funkcje można zastosować indywidualnie w zależności od specyfiki analizowanego sygnału. Wielomiany wyższego stopnia będą niezbędne, jeśli sygnał ma złożoną specyfikę charakteryzującą się dwoma lub więcej obszarami plateau dla różnych zakresów szerokości okna. Przykład takiego sygnału został wymyślony i przedstawiony w powyższej publikacji.

Wprowadzenie wielostopniowego odrzucania zbyt dużych odległości między punktami, przekraczających wartość δ_{max}

Znaczna część czasu obliczeniowego zużywana jest na wyznaczanie złożoności dla wysokich wartości szerokości okna W , dla których d_w niejednokrotnie przekracza 10. Konieczne jest wtedy użycie bardzo wysokich wymiarów zanurzenia wynoszących nawet $m=40-60$. Wiąże się to z koniecznością wykonywania bardzo dużej liczby niepotrzebnych

obliczeń, gdyż np. 99,99% odległości między punktami m -wymiarowymi jest zbyt duża tzn. większa niż δ_{max} . W celu dokonania wstępnej redukcji punktów zbyt odległych, w pierwszym kroku wyznaczane są odległości w oparciu jedynie o pierwszą i ostatnią współrzędną punktu m -wymiarowego. Taka redukcja wymaga wykonania dwóch operacji mnożenia, a nie odpowiednio 40-60. Następnie dla punktów, które przejdą pierwszą selekcję, wyznaczane są odległości w oparciu o wzrastającą liczbę współrzędnych. Po każdym wzroście liczby współrzędnych dokonywana jest selekcja punktów spełniających kryterium mieszczania się w oszacowanej wartości δ_{max} . W ten sposób liczba wykonywanych operacji mnożenia ulega znacznemu zredukowaniu. W zależności od użytego wymiaru zanurzenia uzyskuje się kilkukrotne przyspieszenie procedury obliczeniowej.

Sygnaly o rozkładzie mocno nie-gaussowskim, wstępna normalizacja sygnału

Osobnym problemem związanym z analizą i wyznaczaniem złożoności sygnałów jest podejście do sygnałów posiadających rozkład mocno różniący się od rozkładu gaussowskiego. W bieżącej pracy przedstawiony jest przykład rozkładu liczb losowych, który swoje maksima gęstości prawdopodobieństwa posiada nie w centrum ale na brzegach histogramu. Sygnał taki nie posiada lewostronnego plateau w rozkładzie różnic między wartościami liczb (histogram H_1) ani w histogramach wyższego rzędu uzyskanych przez sploty histogramu H_1 , więc wyznaczenie złożoności napotyka na istotne problemy. Wyznaczana złożoność jest w takich przypadkach zaniżona względem teoretycznej. Pojawia się więc pytanie o interpretację tego zjawiska i otrzymanej wartości wymiaru korelacyjnego. Jednym ze sposobów rozwiązania tego problemu jest wstępna normalizacja sygnału, która polega na takiej nieliniowej, monotonicznej transformacji sygnału, by sygnał otrzymany miał rozkład normalny.

Dokonywane jest to w 3 krokach:

- wygenerowanie zbioru liczb losowych o rozkładzie normalnym i liczności równej liczbie próbek sygnału;
- posortowanie próbek sygnału oryginalnego i zapamiętanie kolejności indeksów sortowania;
- posortowanie próbek zbioru liczb losowych w użyciu kolejności indeksów uzyskanych w trakcie sortowania zbioru badanego.

Wstępna normalizacja sygnału sprawia, że szerokość plateau na wykresie histogramu odległości między punktami (H_1) jest optymalna, dzięki czemu skraca się czas obliczeń. P_{max} dla przyjętego błędu odchylenia e staje się większe, dzięki czemu wzrasta procent odległości między punktami mieszczących się w zadanym przedziale δ_{max} .

Drugim beneficjentem wynikającym z normalizacji jest możliwość skorzystania z gotowych histogramów odległości między punktami wyższych rzędów (H_n), które zostały wyznaczone przez kolejne sploty histogramu H_1 i zapisane w pliku. Nie ma potrzeby wyznaczania histogramów kolejnych rzędów przed każdą kolejną iteracją dla kolejnej wartości W , w której zwiększa się używany wymiar zanurzenia m .

Otrzymaną wartość wymiaru korelacyjnego proponuje się nazwać *znormalizowanym wymiarem korelacyjnym* (*normalized correlation dimension*) dla odróżnienia od wyznaczenia go bez użycia normalizacji sygnału.

Poprawki do podstawowego wzoru Takensa-Ellnera

Sercem algorytmu wyznaczania wymiaru korelacyjnego jest wzór Takensa-Ellnera, który wyznacza go dla posortowanego wektora odległości między punktami δ i zadanych indeksów granicznych l_{min} , l_{max} określających domniemany obszar skalowania liniowego. Przypomnijmy, wzór ten wygląda następująco:

$$\begin{aligned} \text{for } i = 0 \dots D-1: \quad & y(i) = \ln \delta(l_{max}) - \ln \delta(l_{min} + i) \\ Y = & \text{sum}(y(i)) \\ d = & D / [Y + l_{min} \cdot y(0)] \end{aligned}$$

gdzie $D = l_{max} - l_{min}$ – liczba odległości wziętych do analizy

Wzór ten jest bardzo mało intuicyjny, określa on jednak w przybliżeniu dopasowanie prostej do sumy kumulacyjnej histogramu w układzie logarytmicznym w odwróconym układzie współrzędnych (częstość na osi OY, odległość na osi OX, (patrz strona 17).

Wzór ten był opisywany we wcześniejszych publikacjach jako mający tendencję do zaniżania oszacowania d , nawet w przypadku sygnałów niskowymiarowych. W celu sprawdzenia poprawności wzoru Takensa-Ellnera wygenerowano więc sztuczny wektor odległości odpowiadający idealnemu sygnałowi o złożoności $d=8$, zawierający ciąg liczb: $\delta = c(1, 2, 3, \dots)^{1/8}$ (c – stała skalowania, nie wpływająca na uzyskane wyniki). Wzór ten powinien dać dokładnie wartość $d=8$ dla każdej pary indeksów l_{min} i l_{max} . Empiryczne zastosowanie wzoru Takensa-Ellnera do takiego wektora dało jednak wyniki odbiegające nieco od oczekiwanej wartości $d=8$. Różnice widoczne były w szczególności dla małych wartości indeksów l_{min}/l_{max} . Zauważyłem, że wielkość błędu jest płaszczyzną w przestrzeni 3D: $\log l_{min} \times \log l_{max} \times \log \Delta d$. Dopasowane równanie płaszczyzny pozwoliło wyprowadzić równanie na wielkość błędu:

$$\Delta d \approx 0.531 \cdot d_{real}(l_{min})^{-0.361}(l_{max})^{-0.639} \quad \text{R. 7}$$

Jako jedną z przyczyn tego błędu udało się zidentyfikować niedokładne określenie początku obszaru skalowania związanego z wartością indeksu l_{min} . Obniżenie wartości tego indeksu o 0.5 w mianowniku wzoru Takensa-Ellnera

$$d = D / [Y + (l_{min} - 0.5) \cdot y(0)] \quad \text{R. 8}$$

spowodowało prawie całkowite zredukowanie wartości tego błędu. Bardzo niewielka, pozostała wartość błędu ponownie okazała się być płaszczyzną w logarytmicznym układzie współrzędnych. Dopasowane równanie opisujące wartość błędu przyjęło teraz postać:

$$\Delta d \approx 0.08324 \cdot d_{real}(l_{min})^{-1}(l_{max})^{-1}, \quad \text{R. 9}$$

co pozwoliło wyprowadzić wzór na rzeczywistą wartość wymiaru korelacyjnego d_{real} wyliczoną skorygowanym wzorem (tzn. z użyciem: $(l_{min}-0.5)$) dla wartości estymowanej d_{estim} :

$$d_{real} = d_{estim} \cdot l_{min} \cdot l_{max} / (l_{min} \cdot l_{max} - 0.08324) \quad \text{R. 10}$$

Drugim rodzajem błędu metodycznego, który udało się zidentyfikować i częściowo skorygować jest błąd związany z faktem, że rzeczywisty wektor δ z odległościami między

punktami w przestrzeni fazowej wyznaczany w trakcie obliczeń jest w praktyce zawsze jedynie podzbiorem wszystkich odległości, które utworzyłyby idealny wektor δ , gdyby sygnał był nieskończenie długi i można by wyznaczyć nieskończenie wiele odległości między punktami. Liczne eksperymenty numeryczne z losowym i nielosowym wybieraniem punktów ze sztucznie wygenerowanego idealnego wektora $\delta = c(1, 2, 3, \dots)^{1/8}$ wskazywały, że wynik d wychodzi z reguły zaniżony. Analiza teoretyczna wskazała, że brak generowania błędu uzyskuje się, gdy do analizy weźmie się co r -ty punkt wektora, poczynając od wartości r :

$$\delta_r = \delta[r, 2r, 3r, \dots]^a.$$

$$\text{Wtedy dla wektora } \delta_r = c(r \cdot i)^a \rightarrow \log \delta_r = \log(c(r \cdot i)^a) = a \log i + a \log r + \log c.$$

Skalowanie liniowe w układzie logarytmicznym pozostaje zachowane.

Natomiast w przypadku wektora zredukowanego określonego wzorem:

$$\delta_r = \delta[q, r+q, 2r+q, 3r+q, \dots]^a \text{ dla } q < r$$

(bierzemy do analizy co r -tą próbkę poczynając od q -tej)

W układzie logarytmicznym uzyskujemy:

$$\log \delta_r = a \log(r \cdot i + q).$$

Skalowanie liniowe jest zaburzone.

Zauważmy, że w każdym z przedziałów liczb naturalnych $1..r$, $r+1..2r$, $2r+1..3r$ liczba r , $2r$, $3r$ jest ostatnią liczbą z tego zakresu. W przypadku losowego wybierania średnia wartość indeksu będzie więc niższa niż ta, która nie generuje błędu metodycznego, co skutkuje zaniżeniem szacowanej wartości d_{estim} .

Dokonano więc oszacowania błędu wynikającego z losowego wybrania odległości do analizy powtarzając 10000x oszacowanie błędu dla różnych kombinacji indeksów losowych i wyliczono średni błąd estymacji dla różnych kombinacji współczynnika redukcji r i wartości I_{min}/I_{max} uwzględniając poprawki wynikające z poprzedniego akapitu.

Wielkość oszacowanego błędu opisuje równanie:

$$\Delta d \approx 0.295 \cdot d_{real}(I_{min})^{-0.346}(I_{max})^{-0.633} \quad \text{R. 11}$$

które jest w przybliżeniu dokładne dla $r > 5$. Dla małych wartości r dokładność powyższego wzoru nieco maleje. Głównym problemem związanym z tą korektą jest oszacowanie wartości r dla określonego wyznaczonego wektora δ , co wymaga przeprowadzenia w przyszłości dodatkowych analiz.

Obie korekty zostały zaimplementowane w udostępnionym pakiecie HDS-toolkit.

Sygnał o dwóch obszarach plateau na wykresie $d_w = \text{fn}(W)$

Jednym z błędnych paradygmatów związanych z wyznaczaniem złożoności układów wielowymiarowych jest paradygmat mówiący, że w przypadku układów fizycznych, w których wiele czynników wzajemnie na siebie wpływa, złożoność całego układu

reprezentowana przykładowo przez 6 zmiennych w przestrzeni 6-wymiarowej może być oszacowana poprzez analizę każdej ze zmiennych analizowaną w domenie czasowej. Taką analizą jest właśnie oszacowywanie złożoności pojedynczych sygnałów.

Przedstawiony przykład układu złożonego z 6 zmiennych (ryciny 9-12 w omawianej publikacji [8]) pokazuje, że nie zawsze musi to być prawda. Przedstawiony przykład pokazuje, że poszczególne sygnały układu posiadają różne skale czasowe (reprezentowane przez np. różne czasy autokorelacji), przez co ich złożoności mogą się różnić a uzyskane obszary plateau na wykresie $d_w=fn(W)$ mogą przypadać w różnych obszarach szerokości okna W . W szczególności analiza jednej ze zmiennych przedstawionego przykładu charakteryzuje się dwoma obszarami plateau na wykresie $d_w=fn(W)$, które odpowiadają dwom różnym skalom czasowym oscylacji pojawiającym się w analizowanym układzie. Podobne, choć może nie tak spektakularne wyniki uzyskiwałem dla niektórych innych przykładowych układów chaotycznych (wyniki niepublikowane).

Powyższy przykład ilustruje, że kwestia szacowania złożoności układu (reprezentowanego m.in. przez wymiar korelacyjny pojedynczej zmiennej układu) posiada jeszcze wiele aspektów, które wymagają dalszych analiz.

7. Podsumowanie powyższych publikacji

Efektorem prac nad złożonością sygnałów jest pakiet 'HDS-toolkit' zawierający pakiet procedur i funkcji w języku Matlab, które umożliwiają wyznaczenie wymiaru korelacyjnego badanych sygnałów. Pakiet ten został przeze mnie udostępniony do powszechnego użycia na stronie internetowej <http://www.dr-michalak.pl/chaos/eng/index.htm>. W celu demonstracji działania pakietu na wzorcowym sygnale o złożoności $d \approx 6$ należy wpisać w oknie dialogowym Matlab: dhds [Enter].

8. The influence of the vision on the complexity of the body balance regulation

Michalak KP, Przekoracka-Krawczyk A, Optyka 2019, 1(56), pp. 38-42 *MniSZW*: 2

Przykładowe zastosowanie algorytmu do sygnałów posturograficznych

Powyższa publikacja przedstawia wyniki zastosowania opracowanego algorytmu wyznaczania wymiaru korelacyjnego w stosunku do sygnałów posturograficznych.

Złożoność sygnałów posturograficznych została w dużym przybliżeniu oszacowana na ok. $d=5$ w pierwszej omawianej w bieżącym autoreferacie publikacji. Oszacowanie to było jednak dokonane zupełnie inną metodą cechującą się małą dokładnością.

Jednym z problemów przy wyznaczaniu złożoności sygnału posturograficznego jest charakter widma sygnałów posturograficznych, który zbliżony jest do szumu różowego (szum różowy ma widmo w rozkładzie Fouriera o charakterystyce $P \sim f^{-\alpha}$, inaczej: $\log P \sim -\alpha \log f$,

czyli moc maleje wykładniczo w miarę wzrostu częstotliwości w widmie). Relacja jest więc w układzie logarytmicznym linią prostą malejącą.

Wykazano, że wymiar korelacyjny szumu o takiej charakterystyce jest skończony i dla $\alpha > 1$ cechuje się skończoną złożonością wymiarową określoną wzorem $d_{pinknoise} = 2/(\alpha - 1)$ [13]. Wyznaczenie wartości tego parametru dla sygnałów posturograficznych pozostawało więc dalej niejednoznaczne.

Widma sygnałów posturograficznych posiadają w układzie logarytmicznym dwa obszary liniowe odpowiadające średnio w przybliżeniu wartościom $\alpha \sim 0.8$ dla $f < 1\text{Hz}$ i $\alpha \sim 1.4$ dla $f > 1\text{Hz}$. Wyznaczone wartości współczynnika α dla sygnałów posturograficznych oscylowały łącznie w przedziale $\alpha = 1-1.6$. Oznacza to możliwość, że sygnały posturograficzne mogą ze względu na swoją strukturę widmową, a nie chaotyczną, posiadać skończoną złożoność wymiarową oscylującą pomiędzy 3 a `nieskończoność`.

Przeprowadzone wcześniej eksperymenty obliczeniowe z sygnałami posturograficznymi ze względu na niejednoznaczny charakter wyników nie zostały opublikowane. Problemem pozostawał duży zakres oscylacji sygnału posturograficznego związany z powolnym, ale znacznym dryfowaniem środka nacisku ciała na podłoże. Dryf ten odpowiada sytuacji, gdy przykładowo pacjent przez 10 s utrzymuje środek ciężkości bliżej palców a przez następnie przez 20 s bliżej pięt. Daje to wtedy duże oscylacje w zakresie niskich częstotliwości, zwiększa wartość współczynnika α widma sygnału i obniża wyznaczany wymiar korelacyjny.

Na tle dryfu środka masy obserwuje się niewielkie oscylacje sygnału, które reprezentują właściwe odruchy posturalne, są one jednak maskowane przez wspomniany powolny, znaczny dryf środka masy nad czworokątem stóp. Powolny dryf środka masy może być niewielki, gdy pacjent stoi sztywno na baczność lub znaczny, gdy pacjent stoi „luźno” lub często zmienia obciążenie z lewej nogi na prawą lub z przodu do tyłu. Dalsze badania przebiegały więc w kierunku rozdzielenia składowej dryfu i składowej odruchów posturalnych.

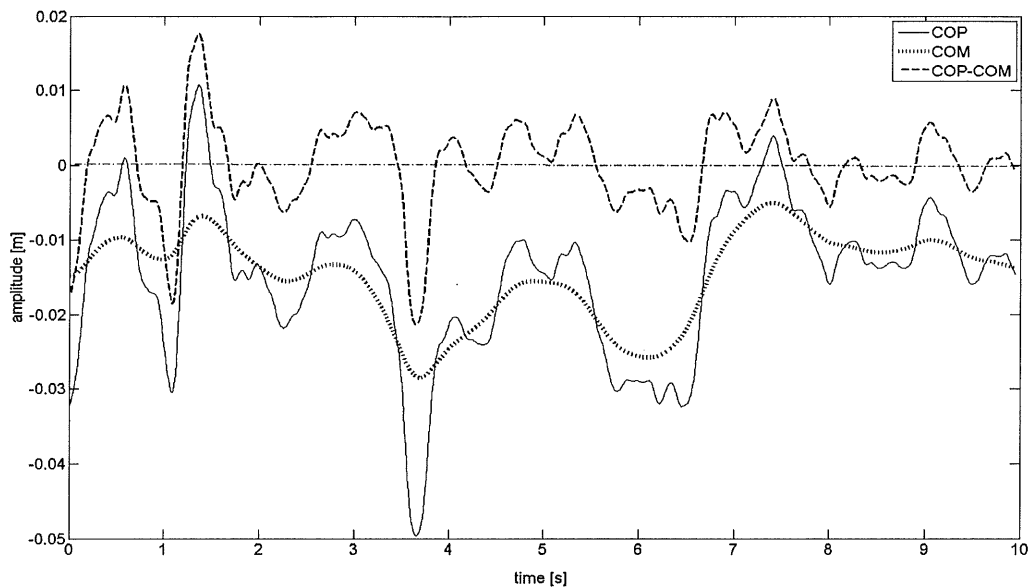
Jedną z metod, która dokonuje rozdziału na te dwie składowe, jest zastosowanie odpowiednio zaprojektowanego filtra cyfrowego. Głównym ograniczeniem metody jest przyjęte założenie, że ciało człowieka w czasie spokojnego stania porusza się jedynie w stawach skokowych, a reszta ciała jest sztywna. W modelu takim ścieżka środka masy ciała (Center of Mass, CoM) może być wyliczona z sygnału posturograficznego (Center of Pressure, CoP) ze wzoru [14]:

$$\frac{CoM(j\omega)}{CoP(j\omega)} = \frac{\omega_0^2}{\omega^2 + \omega_0^2} = \frac{g/h_e}{\omega^2 + g/h_e} \quad R. 12$$

$$\omega_0 = \sqrt{\frac{mgh_e}{I}}, h_e = h \cdot \xi, \quad \xi = 1.15$$

Powyższy wzór jest równaniem filtra cyfrowego, który został wyprowadzony w oparciu o założenie, że chwilowa różnica między CoM a CoP (CoPM) jest proporcjonalna do przyspieszenia CoM. Ciało tym z tym większym przyspieszeniem się przewraca, im bardziej

rzut środka masy jest oddalony od chwilowego środka nacisku ciała na podłoże. Po wyznaczeniu ścieżki CoM możliwe jest wyznaczenie różnicy między CoP a CoM, która reprezentuje właściwe odruchy posturalne zachodzące na poziomie stawów skokowych.



Rycina 6. Zobrazowanie ekstrakcji CoM z CoP i wyznaczenie różnicy między CoP a CoM. Przyspieszenie CoM jest w każdym punkcie proporcjonalne do różnicy CoP-CoM.

Pomimo ograniczeń metody, wobec niejednoznaczności oszacowania wymiaru korelacyjnego dla sygnałów posturograficznych CoP, dokonano pomiaru złożoności sygnałów CoPM, które reprezentują aktywność właściwych odruchów posturalnych. Celem pracy było zbadanie wpływu wejścia wzrokowego na złożoność układu regulacji. Wzrok jest jedną z ważniejszych informacji wejściowych dla układu regulującego balans ciała. Zamknięcie oczu powoduje, że układ regulujący balans ciała traci całkowicie część informacji niezbędnej dla prawidłowego włączania sprzężeń regulujących równowagę. Zważywszy, że w skutek wyłączenia części sprzężeń informacyjnych układ regulujący staje się prostszy, można się spodziewać, że zamknięcie oczu spowoduje zmniejszenie złożoności sygnałów posturograficznych w stosunku do sygnałów rejestrowanych przy oczach otwartych. Przy czym trzeba dodać, że prostsza struktura wcale nie musi oznaczać mniejszych amplitud oscylacji środka nacisku. Wręcz przeciwnie. W skutek tego, że układ staje się prostszy, posiada on mniejszy potencjał regulacyjny i amplitudy oscylacji wzrastają, co też wykazywano wielokrotnie w licznych publikacjach. Prostsza struktura oznacza, że sygnał staje się bardziej regularny, bardziej przewidywalny. Celem prezentowanej analizy było określenie, czy struktura sygnału rzeczywiście staje się prostsza, co będzie widoczne jako niższa wartość wymiaru korelacyjnego sygnałów posturograficznych.

Wyniki

Przebadano składowe CoPM sygnałów posturograficznych zebranych od 30 zdrowych osób w wieku od 18 do 40 lat. Zbadano składowe sygnały AP (anterio-posterior, przód-tył) i

LAT (lateral, lewo-prawo) rejestrowane przy oczach otwartych i zamkniętych. Uzyskane wyniki potwierdzają istotnie niższą złożoność sygnałów CoPM rejestrowanych przy oczach zamkniętych w stosunku do sygnałów rejestrowanych przy oczach otwartych. Silniejszy efekt zaobserwowano dla składowej AP. W przeciwieństwie do niejednoznacznych wyników uzyskiwanych wcześniej dla sygnałów CoP, uzyskano stosunkowo nieduże oscylacje międzyosobnicze relacji $d_w=fn(W)$. Porównanie do sygnałów potasowanych (czyli o identycznych amplitudach w widmie Fouriera ale zrandomizowanych fazach), wykazało niższą złożoność sygnałów oryginalnych, co potwierdza obecność składowej deterministycznej w sygnałach.

Oszacowanie bezwzględnej wartości wymiaru korelacyjnego analizowanych sygnałów w niektórych przypadkach nie było jednak możliwe ze względu na brak obszaru plateau jak również brak punktu mniejszej pochyłości na relacjach $d_w=fn(W)$. W tych przypadkach, w których znaleziono minimum w wielomianie 3-stopnia będącym pochodną wielomianu 4-stopnia dopasowanego do relacji $d_w=fn(W)$, wartość wymiaru korelacyjnego została oszacowana na ok. $d = 5.5-6.4$. Wyniki przedstawia Tabela II.

Tabela II. Wyniki oszacowania średniego wymiaru korelacyjnego składowych CoPM sygnałów posturograficznych. Obserwuje się zmniejszenie złożoności w skutek zamknięcia oczu, większe dla składowej AP (przód-tył). W skutek zamknięcia oczu (eyes closed, E_c) układ regulacyjny staje się prostszy niż przy oczach otwartych (eyes open, E_o).

sygnał	N	d_{opt}	W_{opt}
AP E_o	21	6.34	0.70 s
Ap E_c	22	5.52	0.70 s
LAT E_o	27	5.91	0.73 s
LAT E_c	23	5.64	0.75 s

Podsumowanie

Uzyskane wyniki, jakkolwiek pilotażowe, wskazują na potencjalnie dużą przydatność metody estymacji wymiaru korelacyjnego sygnałów posturograficznych jako wskaźnika stanu zdrowia i funkcjonalności układu posturalnego człowieka. Można się spodziewać, że różne uszkodzenia układu nerwowego będą się objawiać zmniejszeniem wartości współczynnika mierzonego w różnych kombinacjach pomiarowych: oczy otwarte/zamknięte (wzrok), głowa prosto/odchylona (błądźnik), stanie na gąbce lub na palcach (propriocepcja).

9. Podsumowanie prezentowanych publikacji

Głównym osiągnięciem przedstawianym w bieżącym autoreferacie jest opracowanie i udostępnienie algorytmu wyznaczania wymiaru korelacyjnego dla chaotycznych sygnałów

wysokowymiarowych. Algorytm ma charakter uniwersalny i może być zastosowany w wielu dziedzinach nauki, w których mamy do czynienia ze zjawiskami chaotycznymi.

10. Literatura

1. Ben-Mizrachi, A., I. Proccacia, and P. Grassberger, *Characterization of experimental (noisy) strange attractors*. Physical Review A, 1984. **29**: p. 975-977.
2. Ellner, S., *Estimating attractor dimensions from limited data: A new method, with error estimates*. Physics Letters A, 1988. **133**: p. 128-133.
3. Takens, F., *On the numerical determination of the dimension of attractor*. Lecture notes in Mathematics, 1985. **1125**: p. 99-125.
4. Michalak, K.P., *Wykorzystanie teorii dynamiki nieliniowej do badania właściwości sygnału posturograficznego*. Rozprawa Doktorska, 2005.
5. Michalak, K.P., *How to estimate the correlation dimension of high-dimensional signals?* Chaos, 2014. **24**(3): p. 033118.
6. Michalak, K.P., *Distinguishing separate components in high-dimensional signals by using the modified embedding method and forecasting*. Ann Biomed Eng, 2010. **38**(1): p. 200-7.
7. Michalak, K.P., *Modifications of the Takens-Ellner algorithm for medium- and high-dimensional signals*. Phys Rev E, 2011. **83**(2): p. 026206.
8. Michalak, K.P., *Estimating correlation dimension of high-dimensional signals - quick algorithm*. Aip Advances, 2018. **8**(10).
9. Michalak, K.P. and A. Przekoracka-Krawczyk, *Wpływ wzroku na złożoność regulacji balansu ciała*. Optyka, 2019. **1**(56): p. 38-42.
10. Linsay, P.S., *An efficient method of forecasting chaotic time series using linear interpolation*. Physics Letters A, 1991. **153**(6,7): p. 353-356.
11. Sugihara, G. and R.M. May, *Nonlinear forecasting as way of distinguishing chaos from measurement error in time series*. Nature, 1990. **344**: p. 734-741.
12. Wales, D.J., *Calculating the rate of loss of information from chaotic time series by forecasting*. Letters to Nature, 1991. **350**: p. 485-488.
13. Rapp, P.E., et al., *Filtered noise can mimic low-dimensional chaotic attractors*. Physical Review E, 1993. **47**(4): p. 2289-2297.
14. Caron, O., B. Faure, and Y. Brenière, *Estimating the centre of gravity of the body on the basis of the centre of pressure in standing posture*. Journal of Biomechanics, 1997. **30**(11-12): p. 1169-1171.

**11. Funkcje wchodzące w skład pakietu do wyznaczania wymiaru korelacyjnego:
HDS Toolkit 2:**

Nazwa	Wywołanie	Opis
Funkcje podstawowe		
dhds	[<i>dcopt, Wopt, dcW, dcM, dcM_c, WV, PmaxV, other</i>] = dhds (<i>series1, quick_precise_surrog, MaxIterationTime, Da, DeflErr, normalisation, NW_Wmax_WV, Pmin, dcWmethod, fitting_method, plotfigures, signal_number, fighandleset, EachToEach</i>)	Główna funkcja pakietu wyznaczająca wymiar korelacyjny sygnału <i>d</i> . Wykorzystuje pozostałe funkcje pakietu ok. 1400 linii kodu.
dhdsall	[<i>dcopt, Wopt, dcW, dcM, dcM_c, WV, PmaxV, other</i>] = dhdsall (<i>series1, quick_precise, minPmax, DaForMinPmax, Pmin, normalisation, NW_Wmax_WV, fitting_method, plotfigures, signal_number, fighandleset</i>)	Główna funkcja wyznaczająca pełną zależność $d=fn(Pmax, W)$ wg pierwotnego, czasochłonnego algorytmu.
akorel/acorel	[<i>tau, tauV</i>] = akorel (<i>series, FHz</i>)	Znajduje czas autokorelacji sygnału. Dodatkowo zwraca czas, po którym autokorelacja maleje do 1/2, 1/e, 1/4, 1/8, 0. Jeśli podana jest częstotliwość próbkowania, zwraca wartości w sekundach.
createAcubic	[<i>A, n</i>] = createAcubic (<i>series, W, m, J, assumed_n</i>)	Buduje macierz <i>A</i> z punktami <i>m</i> -wymiarowymi dla danego sygnału przy użyciu interpolacji funkcjami sklejanymi.
calcWV	[<i>WV</i>] = calcWV (<i>NW, Wmax, Wmaxdensity</i>)	Wyznacza wektor z szerokościami okna <i>W</i> do użycia dla określonego sygnału w zależności od zadanych wartości wektora: <i>NW</i> – liczba odległości, <i>Wmax</i> – wartość maksymalna <i>W</i> , <i>Wmaxdensity</i> – większe zagęszczenie wartości <i>W</i> w okolicy tej wartości celem poprawy jakości dopasowania w okolicy spodziewanej wartości W_{opt} .
surrog2	[<i>xprim, y, yprim</i>] = surrog2 (<i>X, modif, min_angle, n</i>)	Wyznacza sygnał potasowany z zachowaniem rozkładu sygnału: 1. transformuje na rozkład normalny, 2 - szufluje transformatę znormalizowaną, 3. detransformuje do poprzedniego rozkładu. Zawiera dodatkowe opcje tasowania nieopisywane na razie w literaturze : <i>modif</i> =-1 - różnica kąta dla sąsiednich częstotliwości musi wynosić min. ' <i>minangle</i> ' <i>modif</i> =-2 - wykonuje procedurę szufluj_modif2.m maksymalizującą przesunięcia fazowe między <i>f</i> o największ amplitudach w fft. <i>modif</i> =-3 - maksymalizuje przesunięcia fazowe dla największych częstotliwości, dbając równocześnie by ' <i>n</i> ' sasiednich

		częstotliwości w widmie było przesunięte o $>'minangle'$ (szufluj_modif3.m)
<p>Funkcje wyznaczające histogramy odległości między punktami H, Hc, Hsl oraz kolejne ich sploty oraz funkcje operujące na nich tych histogramach i splotach.</p>		
dists1	$[dd,H1, binnbr, binwidth, Hc1, H1sl] = \text{dists1}(series,J,B, binwidth,binnbr)$	Wyznacza histogram odległości (różnic) między wartościami sygnału ($H1$), jego sumę kumulatywną ($H1c$) oraz pochodną sumy kumulatywnej w skali logarytmicznej ($H1sl$)
dists2	$[hist3, hist3cumsum, hist3slope, dlhist3] = \text{dists2}(H1,H2, norm)$	Elementarny splot 2 histogramów $H1$ i $H2$. Norm - norma matematyczna, domyślnie 2 (wylicza odl. Euklidesowe)
distsm	$[H,Hc,Hsl] = \text{distsm}(H1,to_m)$	Wylicza wszystkie sploty histogramów rzędu od 2 do m
distsm2	$[H,Hc,Hsl] = \text{distsm2}(H, Hc, Hsl, to_m2)$	Dodaje wszystkie kolejne histogramy wyższego rzędu do już wyznaczonych.
distsm3	$[H,Hc,Hsl] = \text{distsm3}(H, Hc, Hsl, to_m2)$	Wylicza nowy wyższy histogram rzędu to_m2 na bazie już wyliczonych. Po drodze wylicza jedynie te pośrednie, które są niezbędne, a nie wszystkie po drodze, szybsza niż distsm2)
find_deltamax	$[deltamaxV,Hslsmooth, deltamaxindV] = \text{find_deltamax}(Hsl,m,binwidth, deflerrV,presmoothHsl)$	Bazując na histogramie splotu Hsl rzędu m wyznacza dla niego wartości $deltamax$ dla określonych wartości błędów odchylenia e ($deflerrV$)
calc_deltamaxM_PmaxM	$[H, Hc, Hsl, deltamaxM, PmaxM, deltamaxindM] = \text{calc_deltamaxM_PmaxM}(x, Mmax, binwidth, presmoothing, deflerrV)$	Dla serii x - wylicza histogramy, a następnie z histogramów macierz $deltamaxM$ dla zadanego wektora z błędami odchylenia $defleErrV$.
calc_deltamaxM_PmaxM2	$[H, Hc, Hsl, deltamaxM, PmaxM, deltamaxindM] = \text{calc_deltamaxM_PmaxM2}(H, Hc, Hsl, deltamaxM, PmaxM, deltamaxindM, newMmax, binwidth, presmoothing, deflerrV);$	Jak wyżej. Dodaje wszystkie nowe wyższe do już wyliczonych (używa distsm2 i find_deltamax).
calc_deltamaxM_PmaxM3	$[H, Hc, Hsl, deltamaxM, PmaxM, deltamaxindM] = \text{calc_deltamaxM_PmaxM3}(H, Hc, Hsl, deltamaxM, PmaxM, deltamaxindM, newMmax, binwidth, presmoothing, deflerrH);$	Jak wyżej. Dodaj nowe ale tylko potrzebne (używa distsm3 i find_deltamax)
deltamax_interpolation	$[deltamax] = \text{deltamax_interpolation}(ED,e, deltamaxM)$	Wyznacza $deltamax$ dla ułamkowych wartości e z sąsiednich całkowitych wartości e zmagazynowanych w macierzy $deltamaxM$
e_interpolation	$[P] = \text{e_interpolation}(e, dc_expected, PmaxM)$	Wyznacza $Pmin/Pmax$ dla ułamkowych wartości d i e
e_rev_interpolation	$[eV] = \text{e_rev_interpolation}(PmaxV, dc_expected, PmaxM)$	Dokonuje interpolacji w drugą stronę niż powyższa: dla $Pmax$ i d wyznacza błędy odchylenia e
<p>Wyznaczanie odległości między punktami w przestrzeni m-wymiarowej (wektor: $deltastar / \delta$)</p>		
Cr_cutoff2	$[deltastar, lcut, Pcut, R, lopt, Popt, DaReached,TotalLength] =$	Funkcja wyznaczająca wektor δ z odległościami dla macierzy A dla różnych parametrów wejściowych.

	Cr_cutoff2 (<i>A, CutoffDist1, CutoffDist2, Da, Imin, Rmax, B, MaxIterationTime, UserAction, EachToEach</i>)	Główna używana.
Cr_PminPmax	[<i>deltastar, lcut, Pcut, R, lopt, Popt, DaReached</i>] = Cr_PminPmax (<i>A, minPmax, maxPmax, Da, Pmin, B, UserAction</i>)	Funkcja wyznacza odległości (wektor δ) dla określonego zakresu <i>Pmin-Pmax</i>
Cr_PminPmaxV	[<i>deltastar, lcut, Pcut, r, lmaxopt, Pmaxopt, deltastarlengthtotal</i>] = Cr_PminPmaxV (<i>A, PmaxV, Pmin, minPmax, DaForMinPmax, B, DistforOptDeflErr</i>)	Funkcja wyznacza odległości δ dla określonego <i>Pmin</i> i określonych wartości <i>PmaxV</i>
Cr_cutoff_minPmax	[<i>deltastar, lcut, Pcut, R, lopt, Popt, DaReached</i>] = Cr_cutoff_minPmax (<i>A, minPmax, CutoffDist1, CutoffDist2, Da, Imin, Rmax, B, MaxIterationTime, UserAction</i>)	Wersja funkcji wyznaczająca odległości δ . <i>minPmax</i> – wartość <i>Pmax</i> , dla której musi być osiągnięta założona liczba odległości <i>Da</i> . Domyślnie <i>minPmax</i> = 10^{-4}
Wyznaczanie wartości wymiaru korelacyjnego d dla określonych kombinacji W, $Pmin$, $Pmax$ na podstawie wyznaczonego wektora δ		
dc_Cr	[<i>dc, sigmadc, D, err</i>] = dc_Cr (<i>Cr, Pcut, PmaxV, PminV, minPmax</i>)	Wylicza d dla określonego wektora δ i określonych wartości <i>Pmin Pmax Pcut</i>
dc_Cr_delta	[<i>dc, sigmadc, D, err</i>] = dc_Cr_delta (<i>deltastar, deltamaxV, deltaminV, Pcut</i>)	Wylicza d dla określonego wektora δ i określonych zakresów <i>Pmin-Pmax</i>
dc_Cr_delta_all	[<i>dc, sigmadc, D</i>] = dc_Cr_delta_all (<i>deltastar, Imin</i>)	Wylicza d dla wszystkich kombinacji indeksów <i>Imin -Imax</i>
Dopasowanie wielomianów i innych funkcji do relacji $d=fn(Pmax,W)$. Wyznaczanie $dopt$, $Wopt$.		
optimal_fitting	[<i>opt_err, opt_fun, opt_param, dc_fit</i>] = optimal_fitting (<i>WV, dcV</i>)	Poszukuje optymalnego dopasowania różnych funkcji zdefiniowanych w nagłówku tej funkcji. Operuje na zmiennych symbolicznych. Umożliwia zdefiniowanie własnych modeli funkcji dopasowywanych do relacji $d=fn(W)$.
evaluate_poly	[<i>Wopt, dcopt, NbrLowestSlopes, PlateauQuality, D0s, D1s, D2s</i>] = evaluate_poly (<i>funs, optparams</i>)	Ewaluuje zoptymalizowaną dowolną funkcję opisaną zmienną symboliczną <i>funs</i> dla współczynników tej funkcji zawartych w wektorze <i>optparams</i> . Zwraca punkty najmniejszej pochyłości.
evaluate_poly_n	[<i>Wopt, dcopt, NbrLowestSlopes, PlateauQuality, D0, D1, D2</i>] = evaluate_poly_n (<i>params</i>)	Ewaluuje optymalny wielomian opisany wektorem <i>params</i> , wyznacza punkty najmniejszej pochyłości i ich parametry.
polyfit4	[<i>dcopt, Wopt, dcM1, iwmaxW, Deriv1, Deriv2, dcM4</i>] = polyfit4 (<i>dcM, PmaxW, WV, czywykresy, minWtoFit</i>)	Dopasowuje wielomiany do relacji $d=fn(W)$ dla poszczególnych wartości <i>Pmax</i> , generuje wykres typu <i>mesh</i> z wielomianami dopasowanymi do relacji $d=fn(Pmax,W)$. Znajduje punkty najmniejszej pochyłości <i>dopt, Wopt</i> .
poly4fit	[<i>dcopt, Wopt, dcMfit, WmaxV, iwmaxV, fittedP, dcoptconfid, Deriv1, Deriv2</i>] = poly4fit (<i>dcM,</i>	Jak powyżej z dodatkowymi opcjami

	$PmaxV, WV, minWtoFit, fig_dir, figname$)	
finddeg	$polydeg=finddeg(stringfun)$	Wyznacza liczbę zmiennych optymalizowanych w zmiennej tekstowej $stringfun$ opisującej optymalizowaną funkcję.
fitting_any	$[params] = fitting_any(fun, xxW, yyW)$	Dokonuje optymalizacji dopasowania dowolnej funkcji zawartej w zmiennej symbolicznej fun dla danych w wektorach xxW i yyW
fitting_err	$[erro, yyfit] = fitting_err (fun, optparams, xx, yy)$	Wylicza sumaryczny błąd dopasowania dokonanego przez $fitting_any$
poly_derivative	$pochodne = poly_derivative(wielomiany)$	wyznacza macierz wielomianów pochodnych dla danej macierzy ze współczynnikami wielomianów
Funkcje rysowania wykresów pojawiających się w trakcie obliczeń		
createfig2	$fighandle = createfig2(X, Y, figtype, descr, labelx, labely, figtitle, figname, fighandle, figsize, figposition)$	Buduje 2-wymiarowy wykres, można wstawić dowolne opisy osi i tytuły. $Figtype$: normal/log-normal/normal-log/loglog, lokuje wykres w odpowiednim miejscu na ekranie.
createfig3	$fighandle = createfig3(X, Y, X2, Y2, X3, Y3, figtype, descr, labelx, labely, figtitle, figname, fighandle, figsize, figposition, Q)$	Buduje wykres z 3 seriami danych o różnych zakresach
tics	$[ticsV] = tics(mintic, maxtic, n, islog)$	Generuje wektor ze znacznikami osi dla wykresów, dopasowując znaczniki do żądanego zakresu, żądanej liczby znaczników oraz czy mają być w skali normalnej czy logarytmicznej.
createmesh	$[fighandle] = createmesh (X, Y, Z, figtype, labelx, labely, labelz, figtitle, sug_nbr_tics, tightaxes, fighandle)$	Buduje wykres siatkowy 3D, dopasowuje skale na osiach w zależności od wartości danych.
createmeshPmaxW	$[fighandle] = createmeshPmaxW (PmaxV, WV, dc, maxdc, figtitle, sug_nbr_tics, figname, fighandle, figsize, figposition)$	Buduje wykres siatkowy 3D, zindywidualizowany do wykresu $d=fn(Pmax, W)$
Pomocnicze funkcje matematyczne		
ndims1	$[ndim, iscol] = ndims1(x)$	Rozszerzenie funkcji $ndim$. Określa liczbę wymiarów zmiennej x . Dodatkowo zwraca 0 dla pojedynczej liczby. Dla wektora zwraca, czy jest kolumną czy rzędem.
max_real	$[Xmax] = max_real(X, dim)$	Wyznacza maksima macierzy X po wymiarze dim z pominięciem wartości Inf i urojonych
min_real	$[Xmin] = min_real(X, dim)$	Wyznacza minima macierzy X po wymiarze dim z pominięciem wartości $-Inf$ i urojonych
min_r	$[w, ind] = min_r(X)$	Wylicza minimum wektora X i zwraca indeks. Pomija Inf NaN .
mean_r	$[sr, ile] = mean_r(X)$	Średnia z elementów rzeczywistych macierzy X . Ignoruje wartości: Inf $-Inf$ NaN .
median_r	$[sr, ile] = median_r(X)$	mediana z elementów rzeczywistych macierzy X . Ignoruje wartości: Inf $-Inf$ NaN .
sum_r	$[suma, ile, sred] = sum_r(X)$	Wylicza sumę i średnią arytmetyczną dla dowolnej macierzy wielowymiarowej pomijając NaN $-Inf$ Inf .
rounds	$[xround] = rounds(x, n, dir)$	Rozszerzone zaokrąglanie. Funkcja zaokrąglania z

		<p>dodatkowymi opcjami dla zaokrąglania w skali logarytmicznej :</p> <p>1 2 lub 5</p> <p>1 lub 3</p> <p>1 2 4 lub 7</p> <p>1 1.4 2 3 5 lub 7</p> <p>1 1.4 2 3 4 5 6 lub 8</p> <p>1 1.3 1.6 2 2.4 3 4 5 6 lub 8</p> <p>dir: floor / round / ceil / fix</p>
meangeom	$[meang, mi, sigma] = meangeom (x, dim, ignore_infs)$	Wylicza średnią geometryczną macierzy x po wymiarze dim . Umożliwia wykluczenie wartości Inf/NaN pojawiające się przy logarytmowaniu liczb niedodatnich.
Funkcje dokonujące monotonicznych nieliniowych przekształceń sygnału		
Ranknorm	$[Y, Xsort, I] = ranknorm (X)$	Przekształca serię danych X monotoniczną nieliniową transformacją do serii o rozkładzie normalnym.
ranknorm2	$[Yred, Y, Xsort, I] = ranknorm2(X, wygladz, gesciej)$	Jak wyżej, z poprawkami zwiększającymi dokładność dla krótkich sygnałów. 'wygladz' - wykonane jest dopasowanie paraboli na podst 'wygladz'-kolejnych punktów. 'gesciej' oznacza ze generowany jest wektor liczb losowych o rozkł. normalnym $n='gesciej'$ x dłuższy, a potem z tego brana jest co n -ta wartość .
Rankcons	$[Y] = rankcons (X)$	Jak ranknorm tylko przekształca do rozkładu stałego (0;1)
rankgiven	$[Y, Xsort, I] = rankgiven (X, distribution)$	przekształca monotonicznie sygnał do rozkładu takiego jak w 'distribution'. 'distribution' - seria zrandomizowana o dowolnym rozkładzie.
Ng	$[seriesng] = ng (series)$	Przekształca serię do serii silnie nie-gaussowskiej opisanej równaniem: $xng = sign(x)*(1-x^2) + 1$ (maksima rozkładu na jego brzegach)

V. Omówienie pozostałych osiągnięć naukowo-badawczych

Moje zainteresowania naukowe znacznie wykraczają poza zakres analizy sygnałów biologicznych i obejmują również inne dziedziny wiedzy. Są nimi w szczególności:

1. Patobiochemia komórki ze szczególnym uwzględnieniem wpływu odżywiania na prawidłowość zachodzących wewnątrzkomórkowych procesów biochemicznych i regulacyjnych.

2. Regulacja produkcji energii w mitochondriach obejmująca glikolizę, cyklu Krebsa, łańcuch cytochromów i stres oksydacyjny ze szczególnym uwzględnieniem zmian w komórkach nowotworowych.

3. Fizyka kwantowa zastosowana do układów biologicznych ze szczególnym uwzględnieniem możliwości wykorzystania pól elektromagnetycznych z zakresu kHz-MHz w diagnostyce chorób człowieka.

4. Pozostały dorobek naukowy i dydaktyczny w obszarze **informatyki**.

Jako lekarz spotykam się z ludźmi chorymi, z których znaczna część zgłasza różnorodne objawy, których klasyczna medycyna nie potrafi do końca sklasyfikować pod względem przyczynowym. Jako inżynier informatyk studiowałem zagadnienia analizy matematycznej, teorię sterowania, sieci neuronalne, fizykę kwantową, programowanie w logice i elektronikę, co istotnie zmieniło moje spojrzenie na problem homeostazy procesów regulacyjnych żywego organizmu. Zdrowie traktuję jako stan prawidłowego funkcjonowania wszystkich mechanizmów autoregulacyjnych zachodzących na poziomach narządowym, tkankowym, komórkowym, biochemicznym i elektromagnetycznym. Ostatni z wymienionych poziomów to bardzo dziewiczy obszar wiedzy i bardzo niewiele ośrodków naukowych w świecie podejmuje badania w tym zakresie. Zdobyte wykształcenie inżynierskie pozwala mi patrzeć na organizm z zupełnie innej perspektywy. Współcześnie bardzo intensywnie rozwija się wiedza z zakresu regulacji ekspresji genów i zmian struktury białek. Jest to bardzo istotna dziedzina wiedzy, gdyż w niedługej przyszłości pozwoli w pełni zrozumieć patobiochemiczne aspekty wielu chorób, których natura pozostaje dziś nie do końca poznana.

Moje zainteresowania kierują się natomiast na regulację produkcji energii w mitochondriach jako podstawowego procesu zasilającego energetycznie wszystkie zjawiska regulacyjne. We wszystkich powyższych trzech dziedzinach posiadam dorobek w postaci artykułów naukowych, popularnonaukowych, a także wykładów i posterów na konferencjach naukowych.

1. Patobiochemia a odżywianie

Odżywianiem interesuję się od początku swojej pracy jako lekarz. Jest powszechnie wiadomym, że prawidłowe odżywianie jest niezmiernie ważne dla prawidłowego funkcjonowania organizmu. Można spotkać się jednak w literaturze z bardzo licznymi koncepcjami prawidłowego odżywiania, co oznacza, że z naukowego punktu widzenia problem nie jest jeszcze należycie zbadany. Wynika to w dużej mierze z faktu, że szczegóły bardzo złożonej regulacji wewnątrzkomórkowej są obecnie dopiero intensywnie badane przez naukę, a odżywianie to w dużej mierze oddziaływanie spożywanymi substancjami na tę regulację. Napisałem kilkanaście artykułów popularnonaukowych przybliżających różne aspekty biochemiczne prawidłowego odżywiania, miałem też kilka wystąpień na konferencjach naukowych. W swoich artykułach starałem się przedstawiać problem odżywiania z punktu widzenia patobiochemii komórki.

Optymalizacja procesu odżywiania człowieka to bardzo złożony problem naukowy. Formalnie podpada on pod problem optymalizacji wielokryterialnej, gdyż optymalizacji podlegać mogą różne przyjęte kryteria: chwilowa sprawność produkcji energii, sprawność procesów regeneracji organizmu, długość życia w sensie spowolnienia fizjologicznych procesów starzenia się i różnorodne zdolności do adaptacji środowiskowych. W przypadku optymalizacji wielokryterialnej jednoznaczne rozwiązanie uzyskuje się jedynie w przypadku określenia współczynników wag, jakie przykłada się do określonych kryteriów podlegających optymalizacji. Dodatkowo sprawę komplikuje różnorodność typów biochemicznych i konstytucyjnych ludzi oraz ich wiek i aktualny stan zdrowia. Określenie optymalnego odżywiania dla poszczególnego człowieka jest więc bardzo skomplikowane, szczególnie, że nie wszystkie zmienne są już przebadane naukowo. W tego typu przypadkach w informatyce stosuje się algorytmy i modele heurystyczne, które dokonując pewnych uproszczeń pozwalają znaleźć przybliżone rozwiązania optymalne.

W przypadku odżywiania wyróżniam 3 główne, w dużej mierze niezależne od siebie osie, które muszą podlegać optymalizacji:

1. Proporcja mieszanki paliwowej: Białka / Tłuszcze / Węglowodany.
2. Ilość i proporcja niezbędnych składników odżywczych: aminokwasów, kwasów tłuszczowych, witamin, makropierwiastków i pierwiastków śladowych.
3. Proporcja stosowanych przypraw posiadających słabsze lub silniejsze właściwości regulujące różne procesy biochemiczne w komórce.

Każda z tych osi wymaga dogłębnej wiedzy dotyczącej wpływu poszczególnych składników odżywczych na równowagę metaboliczną. Jednym ze współczesnych wyzwań dla medycyny jest dokonanie naukowego uporządkowania tej wiedzy w oparciu o dokładną znajomość wszystkich szczegółów biochemii i patobiochemii człowieka. Jest to zadanie tym trudniejsze, że na studiach medycznych brak jest przedmiotu 'Patobiochemia', na którym

studenci uczyliby się biochemicznych podstaw dla dokonywania takiej optymalizacji. Spośród wymienionych 3 osi, trzecia jest najbardziej złożona, najmniej uporządkowana przez naukę i wymagająca najbardziej złożonej, wielowymiarowej analizy.

Wygłosiłem następujące referaty na konferencjach na temat roli pierwiastków śladowych oraz powiązań pomiędzy odżywianiem a metabolizmem:

1. Michalak KP, Jaroszyk F, **Biofizyczne podstawy Diagnostyki Systemowej Mora i jej wykorzystanie do oceny zawartości makro- i mikroelementów w organizmie człowieka**, Konferencja "Pierwiastki w środowisku i życiu człowieka", Poznań 2010, Journal of Elementology, Vol.15, Iss. 3, pp. 66-67
2. Michalak KP, **Rola potencjału błonowego komórki oraz metali dwuwartościowych w etiopatogenezie stresu oksydacyjnego i transformacji nowotworowej**, II Konferencja Medycyny Integracyjnej, Łódź 2013
3. Michalak KP, **Podstawy formalne Medycyny Holistycznej**, V Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź 2016
4. Michalak KP, **Teoria 5 Elementów TCM w świetle patofizjologii i patobiochemii człowieka**, V Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź 2016
5. Michalak KP, **Teoria 5 Elementów TCM w świetle patofizjologii i patobiochemii człowieka - relacje pomiędzy elementami**, VII Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź 2018
6. Michalak KP, **Metabolizm i rola witaminy D w organizmie człowieka** VII Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź 2018
7. Michalak KP, **Biochemiczne uwarunkowania zakwaszenia wewnątrz- i zewnątrzkomórkowego**, VIII Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź 2019

Lista artykułów popularno-naukowych dotyczących odżywiania:

1	Biochemia w pigułce cz.1.*	<i>Wprowadzenie do biochemii</i>
2	Biochemia w pigułce cz.2.*	<i>Podstawy metabolizmu węglowodanów</i>
3	Biochemia w pigułce cz.3.*	<i>Podstawy metabolizmu tłuszczów</i>
4	Biochemia w pigułce cz.4.*	<i>Podstawy metabolizmu aminokwasów</i>
5	To i owo o cholesterolu i miażdżycy*	<i>O roli cholesterolu w rozwoju miażdżycy</i>
6	Fracje lipidowe we krwi*	<i>O przemianach i roli poszczególnych frakcji lipidowych krwi</i>
7	Tajemnice długowieczności*	<i>O nieenzymatycznej glikozylacji białek</i>
8	Mój pierwszy maraton rowerowy.*	<i>O odżywianiu w czasie 6-ciogodzinnego maratonu rowerowego.</i>
9	Borrelia burgorferi – podstępny krętek*	<i>O problemach w diagnozowaniu i leczeniu boreliozy</i>
10	Co należy wiedzieć o oddychaniu?*	<i>O roli płuc w utrzymywaniu równowagi kwasowo-zasadowej</i>

11	Dlaczego warto się przeciągać w łóżku?*	<i>O przyczynach bólów kręgosłupa</i>
12	Skąd się biorą geniusze?	<i>O roli wczesnej stymulacji dziecka oraz nauce czytania i liczenia od 1 r.ż.</i>
13	Dlaczego myję się w zimnej wodzie?	<i>O utajonych zaburzeniach równowagi kwasowo-zasadowej</i>
14	Pij bracie pij.	<i>O przyczynach i skutkach niedostatecznego picia wody</i>
15	Ocena zawartości wód mineralnych – test konsumencki	<i>oceniać ilość minerałów w różnych wodach mineralnych?</i>
16	Test systemowy dr Cornelissena	<i>Przegląd informacji na temat możliwości diagnostyki elektrozonansowej</i>
17	Jak smakują elektrony?	<i>Wstęp do równowagi oksydo-redukcyjnej w organizmie.</i>
18	Detoksykacja w chorobach zwyrodnieniowych	<i>Omówienie niektórych aspektów detoksykacji w chorobach przewlekłych.</i>
19	Len - Polskie Złoto	<i>Rola kwasów tłuszczowych omega-3 dla organizmu.</i>
20	Witamina E – źródło młodości	<i>O roli witaminy E w spowalnianiu procesu starzenia się organizmu</i>

* - artykuły napisane przed uzyskaniem stopnia doktora

2. Regulacja produkcji energii w mitochondriach ze szczególnym uwzględnieniem komórek nowotworowych.

Choroba nowotworowa to poważny problem społeczny, ekonomiczny i psychologiczny zarówno dla pacjenta jak i jego rodziny. Jest to też jeden z trudniejszych tematów, jeśli chodzi o wspomaganie stosowanych terapii nowotworowych. Internet jest pełen różnorodnych metod, które mają leczyć, wspomagać leczenie lub zapobiegać rozwojowi nowotworów. W zdecydowanej większości brak jest odniesień do literatury naukowej, czy choćby do wyjaśnienia w prostych słowach mechanizmu działania na poziomie molekularnym. Problem komplikuje fakt, że różne typy nowotworów posiadają różne profile zaburzeń biochemicznych prowadzących do transformacji nowotworowej komórki, bądź będących jej skutkiem. Gruntowna znajomość podstaw produkcji energii w komórkach nowotworowych musi być podstawą jakiegokolwiek analizy formalnej poszczególnych metod wspomaganie terapii. Jako lekarz interesujący się biochemią, biofizyką i bioelektroniką zgłębiam i analizuję różne doniesienia naukowe na ten niezmiernie złożony temat.

Posiadam w swoim dorobku naukowym kilka artykułów i wystąpień konferencyjnych w tym temacie:

1. Michalak KP, Maćkowska-Kędzióra A, Sobolewski B, Woźniak P. **Key roles of glutamine pathways in reprogramming the cancer metabolism**, Oxidative Stress and Cellular Longevity, 2015, ID964321, doi: 10.1155/2015/964321, IF 4.492, MNiSZW 25

2. Michalak K.P. Nawrocka-Bogusz H. **The changes of the frequency-specific impedance of the human body due the resonance in the kHz range in the cancer diagnostics**, artykuł pokonferencyjny w Journal of Physics - Conference Series, 9th International Frohlich's Symposium "Electrodynamic activity of living cells, including microtubule coherent modes and cancer cell physics. Praga, VII 2011.
3. Michalak K.P. **Procedure for the Krebs cycle metabolism profiling in cancers using MORA-kHz-resonance diagnostics**, EQBS13 - II International Symposium Electromagnetic Fields and Quantum Phenomena in the Biological Systems, Poznań, X 2013
4. Michalak K.P. **Zaburzenia biochemiczne w chorobie nowotworowej w diagnostyce elektrozonansowej MORA**, Polsko-Niemiecki Kongres Integratywnej Terapii Nowotworów, Frankfurt am Main, IV 2016
5. Michalak K.P. **Biochemical aspects of energy production in cancers**, 10th International Froelich's Symposium: Biophysical Aspects of Cancer, Praga, VII 2016
6. Michalak KP, **Stres oksydacyjny jako przyczyna chorób degeneracyjnych**, VI Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź, III 2017
7. Michalak KP, **Rola mitochondriów i fibroblastów w transformacji nowotworowej – 2 typy metaboliczne nowotworów**, VI Międzynarodowa Konferencja Medycyny Integracyjnej, Łódź, III 2017

Regulacja produkcji energii a toksyczność fluorochinolonów

Innym zgłębnym przeze mnie aspektem związanym z zaburzeniem produkcji energii w komórkach jest kwestia wpływu antybiotyków z grupy fluorochinolonów na mitochondria. W ostatnich latach pojawiło się niezmiernie dużo doniesień naukowych na temat opóźnionego, toksycznego oddziaływania tych antybiotyków na organizm, co spowodowało wprowadzenie przez FDA w USA ograniczeń w ich stosowaniu. Mając na uwadze potrzebę znalezienia metod skutecznej pomocy ludziom dotkniętym efektami toksycznymi tych antybiotyków dokonałem obszernego przeglądu literatury na ten temat, który przedstawiłem w publikacji:

Michalak KP, Sobolewska-Włodarczyk A, Włodarczyk M, Sobolewska J, Woźniak P, Sobolewski B, **Treatment of the Fluoroquinolone-Associated Disability: The Pathobiochemical Implications**, *Oxidative Medicine and Cellular Longevity*, 2017, *Vol.3,pp.1-15*, doi: 10.1155/2017/8023935.

IF: 4,563 MNiSZW: 30

Powyższa publikacja jest pierwszą, która próbuje kompleksowo ocenić problem ich toksyczności, wskazuje na dalsze kierunki badań i proponuje metody leczenia bazując na dotychczas znanych mechanizmach oddziaływania na metabolizm. W ciągu 16 miesięcy publikacja osiągnęła poziom 1000 przeczytań na portalu www.researchgate.net, co wskazuje na niezmiernie wysokie zapotrzebowanie społeczne w świecie na wiedzę w tym temacie.

3. **Biologia kwantowa, czyli fizyka kwantowa zastosowana do układów biologicznych**

Jako inżynier informatyk, absolwent studiów magisterskich Wydziału Elektrycznego Politechniki Poznańskiej, a równocześnie lekarz, od początku mojej kariery naukowej interesowałem się zjawiskami elektromagnetycznymi zachodzącymi w układach biologicznych. Jest to bardzo pionierska dziedzina wiedzy, której podwaliny są dopiero tworzone przez nielicznych fizyków kwantowych, których pasją stało się zgłębianie nader złożonych zjawisk kwantowych w układach biologicznych. Występuje też bardzo istotny problem pasażu wiedzy rozwijanej przez fizyków kwantowych do środowisk medycznych, gdyż praktycznie nie ma lekarzy, którzy zajmują się tą tematyką. Już w 2007 nawiązałem kontakt z europejskim środowiskiem naukowym fizyków, którzy zajmują się tą tematyką uczestnicząc w tygodniowym Sommer School organizowanym przez wybitnego pioniera tej dziedziny prof. Fritza Alberta Poppa, autora znanej książki "Biologia światła".

Spotkanie to zaowocowało dalszą współpracą. W kolejnych konferencjach byłem aktywnym uczestnikiem przedstawiając postery lub prezentując wykłady. Osobiście zorganizowałem też dwie konferencje w tym temacie (Electromagnetic Fields and Quantum Phenomena in the Biological Systems, EQBS-2011 i EQBS-2013), na które zaproszenie przyjęli wybitni profesorowie z tej dziedziny: (m.in. prof. Emilio del Giudice, prof. Jiri Pokorny, prof. Jack Tuszynski, prof. Gerard Pollack). Na konferencjach tych rozwijałem swoją wiedzę z zakresu fizyki kwantowej w układach biologicznych, nieodłącznie związanej z tym tematem wiedzę dotyczącą struktury wody, a także przedstawiałem wyniki swoich badań w zakresie pomiarów impedancji elektrycznej pacjentów w zakresie kHz oraz zmian absorpcji w zakresie UV przez wodę pod wpływem oddziaływań mechanicznych. Jako dowód uznania moich badań w tym temacie przez środowisko fizyków kwantowych może posłużyć fakt, że na ostatnią z konferencji (Lugano, X 2016) zostałem zaproszony przez organizatorów z całkowitym pokryciem kosztów przelotu, zakwaterowania i uczestnictwa w konferencji.

Tematyką, jaką się głównie zajmuję w obszarze zjawisk elektromagnetycznych zachodzących w układach biologicznych, jest problematyka zależnej od widma i stanu organizmu zmieniającej się impedancji elektrycznej ludzkiego ciała w obszarze kHz i możliwość wykorzystania tego zjawiska do diagnostyki chorób człowieka. W swoich wystąpieniach konferencyjnych przedstawiałem aktualny stan wiedzy dotyczący tego obszaru nauki oraz przedstawiałem wyniki pomiarów pacjentów.

Wystąpienia konferencyjne i artykuły:

1. Michalak KP, **Physical properties of the structured water (I_E)**, Biophotonics and application of biophotons - Sommer School, Neuss - Germany, VIII 2009
2. Michalak KP, Jaroszyk F., **Biofizyczne podstawy Diagnostyki Systemowej Mora i jej wykorzystanie do oceny zawartości makro- i mikroelementów w organizmie człowieka**, XII Zjazd Polskiego Towarzystwa Magnezologicznego w Poznaniu: Pierwiastki w środowisku i życiu człowieka, Poznań, IX 2010
3. Michalak KP, **The fundamentals for the extra-cellular and inter-organ electromagnetic**

communication in the kHz range, I International Symposium: Electromagnetic Fields and Quantum Phenomena in Biological Systems, Poznań, V 2011. Current Topics in Biophysics, Vol 34A

4. J. Sławiński, K. Michalak, **From mitogenic to necrotic radiation – toward the photon cycling in the biological systems and electromagnetic existence**. Electromagnetic Fields and Quantum Phenomena in Biological Systems, Poznań, V 2011, Current Topics in Biophysics, Vol 34A

5. Nawrocka-Bogusz H, Jaroszyk F, Michalak KP, **Effect of magnetic field, the red light and their combination on the oxygen-dependent mechanism of killing of neutrophils**, EQBS11 - I International Symposium Electromagnetic Fields and Quantum Phenomena in the Biological Systems Poznań, V 2011, Current Topics in Biophysics Vol.34A

6. Michalak KP, **The changes of the frequency-specific impedance of the human body due to the resonance in the kHz range in the diagnostics of the cancer**. Konferencja: 9th International Froelich's Symposium: Electrodynamics Activity of Living Cells including Microtubule Coherent modes and Cancer Cell Physics (EDALC 11) VII/2011, Praga, Book of Abstracts p.21-22, *referat zapraszany*

7. Michalak KP, **The diagnostics of the human by using the spectrum-specific impedance in the range of KHZ**, International Conference on Medical Physics and Engineering on "Physics and Engineering for Health and Wellness, Poznań, IX 2011, Abstrakt: Pol. J. Med. Phys. Eng. 2011 Vol. 17 suppl. 1 s. s29.

8. Michalak KP, **Borreliosis in the electro-resonance MORA diagnostics**. International Conference "Fields of The Cell" Basel, 6-7 Oct 2012, Swiss

9. Michalak KP, **Procedure for the Krebs cycle metabolism profiling in cancers using MORA-kHz-resonance diagnostics**, Electromagnetic Fields and Quantum Phenomena in Biological Systems, Poznań, X 2013

10. Michalak KP, **Autoimmune illnesses in the kHz-resonance diagnostics**, rozdział w monografii Current Topics in Quantum Biology, Poznań, 2013

11. Michalak KP, **Hawthorn Berries caps., Oxalaceticum and other drugs reduce the resonance of 'Lactic acid D2' signal from MORA electro-resonance test in humans**. 4th Internal Symposium Biophysical Aspects of Complexity in Health and Disease, Lugano, X 2016

Podsumowanie

Jednym z celów badań nad strukturą elektromagnetyczną fenomenu życia jest opracowanie **Mapy Elektromagnetycznej Człowieka**. Pod takim szyldem organizowałem sympozja naukowe w Poznaniu w latach 2011 i 2013.

Podobnie jak 50 lat temu postawiono sobie za cel opracowanie mapy całego genomu człowieka, dziś takim celem jest Mapa Elektromagnetyczna obejmująca zbadanie widm emisyjnych i absorpcyjnych wszystkich biomolekuł budujących ciało w stanach prawidłowych i w stanach nieprawidłowych. W dalszym etapie konieczne będzie określenie praw związanych z przekazywaniem kwantów energii wewnątrz molekuł, pomiędzy molekułami, organellami, komórkami, tkankami i organami. Zbadania wymaga struktura i elektroniczne właściwości kolagenu i innych białek przestrzeni międzykomórkowej, jako półprzewodników i przenośników dla kwantowych stanów wzbudzonych. Konieczne są dalsze badania nad strukturą i właściwościami wody budującej nasze ciało oraz jej rolą w zjawiskach kwantowych układów biologicznych.

4. Pozostały dorobek naukowy i dydaktyczny w obszarze informatyki

Na pozostały dorobek naukowy i dydaktyczny należący do dyscypliny „Informatyka” składają się inne publikacje dotyczące przetwarzania sygnałów, wystąpienia konferencyjne, przygotowanie oprogramowania ćwiczeń laboratoryjnych dla studentów w środowisku Matlab prezentujących symulacje zjawisk biofizycznych zachodzących w ludzkim organizmie, a także przygotowanie oprogramowania w środowisku 'Presentation' do eksperymentów badawczych przeprowadzanych w obrębie Pracowni Fizyki Widzenia i Optometrii UAM.

Pozostałe publikacje naukowe i wystąpienia konferencyjne

Artykuły naukowe

Przekoracka-Krawczyk A, Nawrot P, Czaińska M, Michalak KP

Impaired body balance control in adults with strabismus, Vision Reseach, 2014, Vol. 98, pp.35-45, IF: 2,381

Michalak KP, Przekoracka-Krawczyk A, Nawrot P, Woźniak P, Vieregge P

The Filtering of the Posturographic Signals Shows the Age Related Features, The Scientific World Journal, 2014

Michalak KP, Przekoracka-Krawczyk A, Nawrot P, Woźniak P, Vieregge P

The aging process In the principal component analysis of posturographic signals, Applications of Informations Systems in Engineering and Bioscience, 2014, pp.183-192

Konferencje - wykłady

Michalak KP, Jaśkowski P, **The influence of aging process and idiopathic gait disturbance on the complexity of the posturographic signals**, National Conference - physics and engineering in modern medicine and health care, Warszawa, IX 2005.

Michalak KP, Jaśkowski P, **The analysis of posturographic signals using the nonlinear forecasting method**, National Conference - physics and engineering in modern medicine and health care, Warszawa, IX 2005

Michalak KP, Jaroszyk F, Bio-algorithm and Med.-Systems, **The use of the shuffled signals for the estimation of the proper window width for the calculation of the dimensional complexity of the high-dimensional signals**, VII National Conference Cybernetic Modelling of Biological Systems - MCSB 2010, Kraków, 2010 Vol. 6, Iss. 12, pp. 138

Michalak KP, Jaroszyk F, Bio-algorithm and Med.-Systems, **The use of the cubic interpolation for the analysis of the medium- and high-dimensional chaotic signals**, VII National Conference Cybernetic Modelling of Biological Systems - MCSB 2010, Kraków, V 2010, Vol. 6, Iss. 12, pp. 137

Michalak KP, **Decomposition of highdimensional signals by using Empirical Mode Decomposition, capabilities and applications**, II Symposium „Biophysics and Medicine", Sienna-Czarna Góra, II 2013

Michalak KP, **The use of numerical methods for noise reduction in the analysis of highdimensional signals**, II Symposium „Biophysics and Medicine”, Sienna-Czarna Góra, II 2013

Michalak KP, Michalak KP, Przekoracka-Krawczyk A, Nawrot P, Woźniak P, Vieregge P

Applications of Informations Systems in Engineering and Bioscience, **The aging process In the principal component analysis of posturographic signals**, International Conference on Bioscience and Bioinformatics ICBB'14, Gdańsk, V 2014

Konferencje – postery

Michalak KP, Świdziński T, Jaroszyk F, **Combination of posturography and human body movement analysis - initial report**, Conference "Measurement and Control in Biomedicine", Ostrava, IX 2003

Michalak KP, Świdziński T, Jaroszyk F, **Posturography**, II International Symposium on Medical Physics, Ustroń, XI 2003

Michalak KP, Przekoracka-Krawczyk A, Vieregge P, Neuronus 2015, IBRO & IRUN Neuroscience Forum, **Posturography: Estimating the path of 'Center of Body Mass' shows the features of aging process of postural control**, Kraków, Neuronus 2015, Book of abstracts IV 2015, pp. 62

Oprogramowanie do ćwiczeń laboratoryjnych i eksperymentów badawczych w środowiskach Matlab i Presentation

Laboratorium „**Potencjał czynnościowy**” – symulator rzeczywistego eksperymentu w środowisku Matlab

Laboratorium „**Biofizyka mięśnia**” – symulator rzeczywistego eksperymentu w środowisku Matlab

Laboratorium „**Prawa przepływu cieczy**” – symulator rzeczywistego eksperymentu w środowisku Matlab

Laboratorium „**Biokalorymetria**” – symulator rzeczywistego eksperymentu w środowisku Matlab

Laboratorium „**Lepkość cieczy**” – symulator rzeczywistego eksperymentu w środowisku Matlab

Laboratorium „**Posturografia**” – przeprowadzenie analiz numerycznych w środowisku Matlab

Badanie kontrastu w polu centralnym – zaprogramowanie eksperymentu badawczego w środowisku Presentation

Badanie kontrastu w polu peryferyjnym – zaprogramowanie eksperymentu badawczego w środowisku Presentation

Badanie postrzegania peryferyjnego u narciarzy alpejskich – zaprogramowanie eksperymentu badawczego w środowisku Presentation

VI. Podsumowanie dorobku naukowego:**Całość**

Łączna liczba publikacji	19
Łączna liczba publikacji jako pierwszy lub ostatni autor	13
Liczba publikacji z IF	11
Liczba publikacji z punktacją MNiSZW	17
Łączna wartość IF publikacji	23,912
Łączna wartość punktów MNiSZW	342

Dzieło habilitacyjne

Łączna wartość IF publikacji osiągnięcia naukowego	8,238
Łączna wartość punktów MNiSZW osiągnięcia naukowego	138

Cytowania

	WoS	GoogleScholar
Liczba cytowań wg ISI Web of Science bez autocytowań	37	
Liczba wszystkich cytowań	47	97
Indeks Hirsha	4	5

Pozostały dorobek

Liczba publikacji popularno-naukowych	23
Liczba streszczeń ze zjazdów międzynarodowych	22
Liczba streszczeń ze zjazdów krajowych	16

Krzysztof Michałek