

Autoreferat

Paweł Dłotko

30 marca 2021

1. Imię i Nazwisko: Paweł Dłotko
2. Wykształcenie:
 - (a) Tytuł zawodowy magistra matematyki w dyscyplinie informatyka, Uniwersytet Jagielloński w Krakowie, na podstawie pracy “Aproksymacja homologii zbiorów zwartych”, 2007, promotor: prof. dr. hab. Marian Mrozek.
 - (b) Stopień doktora nauk matematycznych w dyscyplinie informatyka, Uniwersytet Jagielloński w Krakowie, na podstawie pracy “Algorytmy redukcji w obliczeniowej teorii homologii i kohomologii z zastosowaniami”, 2012, promotor: prof. dr. hab. Marian Mrozek.
3. Zatrudnienie w jednostkach naukowych:
 - (a) 2011 - 2015 - asystent w Instytucie Informatyki Uniwersytetu Jagiellońskiego (od 2013 przebywałem na urlopie naukowym).
 - (b) 2015-2017 - adiunkt w Instytucie Informatyki Uniwersytetu Jagiellońskiego (przez cały ten okres przebywałem na urlopie naukowym).
 - (c) 2013-2015, Postdoctoral researcher, Departement of Mathematics, University of Pennsylvania, Philadelphia, PA, USA.
 - (d) 2015-2017, Researcher, Inria-Saclay, Francja.
 - (e) 2018-2020, Lecturer and Senior Lecturer, Departement of Mathematics, Swansea University, Swansea, UK.
 - (f) 2020-obecnie, dyrektor Centrum Dioscuri w Topologicznej Analizie Danych, Instytut Matematyczny, Polska Akademia Nauk, Warszawa.

4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule naukowym zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.):

(a) **Tytuł: Algorytmiczna topologia stosowana.**

(b) **Lista publikacji wchodzących w skład osiągnięcia naukowego:**

i. Wektoryzacja *persistence diagrams*:

[V1] P. Bubenik, P. Dłotko, *A persistence landscapes toolbox for topological statistics*, Journal of Symbolic Computations, 78, 91-114, (2017).

[V2] B. Zieliński, M. Lipiński, M. Juda, M. Zeppelzauer, P. Dłotko, *Persistence bag-of-words for topological data analysis*, Artificial Intelligence Review, (2020).

ii. Przyspieszenie obliczeń *persistent homology*:

[S1] P. Dłotko, H. Wagner, *Simplification of complexes of persistent homology computations*, Homology, Homotopy and Applications 16 (1), 49-63, (2013).

iii. Od matematyki ciągłej do dyskretnej:

[CD1] P. Dłotko, T. Wanner, *Rigorous cubical approximation and persistent homology of continuous functions*, Computers & Mathematics with Applications 75 (5), 1648-1666, (2018).

iv. Zastosowania w nauce o materiałach:

[MS1] Y. Lee, S.D. Barthel, P. Dłotko, S.M. Moosavi, K. Hess, B. Smit, *Quantifying similarity of pore-geometry in nanoporous materials*, Nature Communications, 8 (1), 1-8, (2017).

[MS2] P. Dłotko, T. Wanner, *Topological microstructure analysis using persistence landscapes*, Physica D: Nonlinear Phenomena 334, 60-81, (2016).

v. Zastosowania w inżynierii elektrycznej:

[EE1] P. Dłotko, R. Specogna, *Cohomology in 3D magnetoquasistatics modeling*, Communications in Computational Physics, 14 (1), 48-76, (2013).

[EE2] P. Dłotko, B. Kapidani, R. Specogna, *Topoprocessor: An Efficient Computational Topology Toolbox for h-Oriented*

Eddy Current Formulations, IEEE Transactions on Magnetics 53 (6), 1-4, (2017).

[EE3] P. Dłotko, *A fast algorithm to compute cohomology group generators of orientable 2-manifolds*, Pattern Recognition Letters 33 (11), 1468-1476, (2012).

1 Wprowadzenie

1.1 Uwaga o użytej nomenklaturze

W referacie poniżej, gdzie tylko to możliwe, będę używał polskiej terminologii. Gdy dla danego pojęcia nie został dotychczas wypracowany uznany odpowiednik w języku polskim, pozostanę przy terminach anglojęzycznych. Przykładem takich pojęć są *persistent homology*, czy *persistence diagrams*.

1.2 Kilka słów o mojej tematyce

Topologia stosowana i obliczeniowa, w szczególności Topologiczna Analiza Danych, jest stosunkowo młodą dyscypliną. Jeszcze dwadzieścia-trzydzieści lat temu topologia była domeną wyłącznie matematyki czystej, bardzo odległą od bezpośrednich zastosowań. Kilka nurtów badań sprawiło, że pewna jej część wpisała się w kanon nauk informatycznych i stosownych.

Topologia stosowana wywodzi się z bardziej klasycznej *geometrii obliczeniowej*. To tam rozpoczęły się badania nad problemem odtwarzania powierzchni na podstawie skończonej chmury punktów próbkowanej z tejże powierzchni (manifold reconstruction problem). Zaczęto wtedy rozważać pytanie, czy pewne niezmienniki topologiczne, jak na przykład grupy homologii kompleksu uzyskanego z takiej rekonstrukcji, są zgodne z grupami homologii powierzchni, którą chcieliśmy uzyskać. Rozwój topologii stosowanej był również motywowany próbami zastosowania technik geometrycznych w naukach biologicznych; próbowano wtedy wykorzystać algorytmy obliczania grup homologii do detekcji wysokowymiarowych zbiorów punktów, które tworzą nie do końca zamknięty okrąg. Próby te okazały się nieskuteczne ze względu na zniekształcenie szumem i inne niedoskonałości rzeczywistych danych. Zdano sobie wtedy sprawę, że w takich przypadkach pewne charakterystyki topologiczne, które są niezmiennikami przekształceń ciągłych, mogą okazać się skuteczne. To właśnie ta niezmienniczość ze względu na przekształcenia ciągle sprawiła, że metody topologiczne dobrze radzą sobie z niepełnymi oraz zniekształconymi szumem danymi. Te rozważania doprowadziły do powstania persistent homology [1] oraz algorytmu mappera [12].

Rozwój topologii stosowanej nastąpił również dzięki ścisłym metodom dynamiki obliczeniowej, w szczególności pionierskim badaniom prof. Mariana Mrozka nad zjawiskiem chaosu w układzie Lorenza [13]. Metody użyte w tej pracy wykorzystywały ściśle obliczenia numeryczne bazujące na teorii indeksu Conleya. Obliczenia indeksu Conleya, w dużym skrócie, na pewnym etapie sprowadza się do obliczeń homologii relatywnych pewnego zbioru, relatywnie do części jego brzegu. Potrzeba obliczania indeksu Conleya dla coraz bardziej skomplikowanych danych wymagała opracowania oraz implementacji szybkich i efektywnych pamięciowo metod obliczania grup homologii. Prace nad tym problem były źródłem powstania grupy, a dzisiaj można już powiedzieć nawet szkoła szkoły topologii stosowanej rozwiniętej w przez prof. Mariana Mrozka w Instytucie Informatyki Uniwersytetu Jagiellońskiego.

Topologia obliczeniowa jest dyscypliną lokującą się pomiędzy matematyką, informatyką i naukami stosowanymi, która łączy w sobie metody tych trzech dyscyplin. Bazuje ona na adaptacji ścisłych geometrycznych metod topologii do rozwiązywania praktycznych problemów pochodzących z modelowania matematycznego. Zazwyczaj pierwszym etapem badań jest adaptacja matematycznej metody obliczania pewnych niezmienników tak, by mogła być użyta dla dyskretnych/skończonych danych. Następnie, opracowywane są efektywne algorytmy dedykowane obliczeniu tych niezmienników. Na końcu, te algorytmy są implementowane w taki sposób, aby mogły zostać użyte w praktyce. Należy zwrócić uwagę, że dwa z trzech wymienionych powyżej etapów mieszczą się w ramach nauk matematycznych co sprawia, że topologia obliczeniowa jest często uważana za dyscyplinę z pogranicza matematyki stosowanej i informatyki teoretycznej.

Topologia stosowana i obliczeniowa jest intensywnie rozwijana w czołowych instytucjach naukowych, w szczególności w grupie prof. Gunnara Carlsona oraz prof. Leonidasa Guibasa (Stanford University), prof. Herberta Edelsbrunnera (University of Illinois, Duke, IST-Austria), prof. Fredrica Chazala (Inria Scalay), prof. Konstantina Mischaikowa (Rutgers), prof. Yasuaki Hiraoki (Kyoto) czy grupie prof. Ulrike Tillmann (Oxford University) i wielu innych. W Polsce metody topologii obliczeniowej są rozwijane przede wszystkim w grupie prof. Mariana Mrozka. Ponieważ sama dyscyplina jest z pogranicza informatyki i matematyki, wspomniane grupy prowadzą działalność zarówno w instytutach informatyki jak i matematyki.

1.3 Ogólny opis osiągnięć w rozprawie Algorytmiczna Topologia Stosowana.

Poczynając od pracy magisterskiej z 2007 roku jak i rozprawy doktorskiej z 2012 roku, moje zainteresowania naukowe lokują się na pograniczu matematyki (dokładniej: geometrii i topologii obliczeniowej, statystyki i teorii układów dynamicznych), informatyki (efektywnych algorytmów i ich implementacji) oraz zastosowań w różnych dziedzinach nauki. Uważam, że abstrakcyjna matematyka ma bardzo wiele do zaoferowania naukom inżynierskim, naukom o materiałach a w szczególności szeroko rozumianej analizie danych. Jednak, aby umożliwić użycie zaawansowanych metod matematycznych, abstrakcyjne modele matematyczne muszą być przedstawione tak, aby były zrozumiałe dla specjalistów reprezentujących dziedziny, w których mają one znaleźć zastosowanie. Co więcej, egzystencjalne, często niekonstruktywne dowody twierdzeń matematycznych, powinny być stopniowo uzupełniane efektywnymi algorytmami, za którymi idą implementacje komputerowe.

Konieczność algorytmizacji i implementowania różnych koncepcji stanowi pierwszą z trudności, jaką napotykałyśmy w matematycznej informatyce stosowanej. Kolejną trudnością jest konieczność wypracowania wspólnego języka z przedstawicielami nauk, w których stosowana jest matematyka i informatyka (w moim przypadku, z inżynierami, chemikami, fizykami a nawet socjologami i przedstawicielami nauk politycznych, medycznych i innych). Ten wspólny język pozwala im nie tylko zrozumieć modele matematyczne, ale, co bardzo ważne, przedstawiać problemy tamtych nauk w formie możliwie bliskiej precyzyjnego języka używanego w matematyce.

Osiągnięcia naukowe składające się na moją rozprawę habilitacyjną mieszczą się w ramach algorytmicznej topologii stosowanej, będącej dyscypliną z pogranicza matematyki, informatyki i zastosowań. Można je naturalnie podzielić na kilka szczegółowych podgrup.

W pierwszej z nich zajmuję się metodami wektoryzacji *persistence diagrams*. Publikacja [V1] przedstawia pierwszy w literaturze efektywny algorytm do wektoryzacji *persistence diagrams*. Polega on na transformacji *persistence diagrams* do tak zwanych *persistence landscapes*. Częścią składową publikacji jest efektywna implementacja w C++, która jest używana przez liczne zespoły na świecie (jest pierwszą implementacją umożliwiającą użycie *persistence diagrams* w statystyce i nauczaniu maszynowym). Praca [V2] pozwala na wektoryzację *persistence diagrams*, które adaptują się do rozważanego zbioru diagramów. Konsekwentnie, tego typu wektoryzacje są znacznie bardziej efektywne, co znalazło potwierdzenie w licznych ekspery-

mentach przedstawionych w pracy [V2].

Kolejnym punktem jest praca [S1], w której uogólnia się, wprowadzone w grupie Mariana Mrozka, techniki upraszczania kompleksów sympleksyjnych z zachowaniem ich homotopii (homologii) na przypadek kompleksów z filtracją. W tej sytuacji, techniki te mogą skutecznie przyspieszyć obliczenia *persistent homology* kompleksów z filtracją.

W swojej pracy badawczej rozważam również relacje matematyki dyskretnej i ciągłej. Praca [CD1] przedstawia, bazującą na koncepcji arytmetyki przedziałowej, metodę dyskretyzacji funkcji $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tak, by odległość w metryce L^∞ pomiędzy aproksymacją oraz rzeczywistą wartością funkcji była zawsze mniejsza niż zadana na wejściu stała. Metoda ta w szczególności pozwala obliczać *persistent homology* dla funkcji ciągłych.

Prace [MS1] i [MS2] przedstawiają zastosowania metod topologii obliczeniowej do analizy materiałowej. W pierwszej z prac proponujemy schemat porównania materiałów nanoporowatych przy pomocy odpowiednio skonstruowanego filtrowanego kompleksu sympleksyjnego i jego *persistent homology*. W drugiej pracy rozważamy bardziej teoretyczne pytanie - na ile informacje z kostkowych *persistent homology* rozwiązania pewnego równania różniczkowego (w tym przypadku równania Cahna-Hilliarda-Cooka) zachowują informację o parametrach tegoż równania.

Seria prac [E1]-[E2] jest uwieńczeniem długoletniej współpracy z grupą inżynierów elektryków. Celem tej współpracy był, po pierwsze, teoretyczny rozwój Dyskretnego Geometrycznego Podejścia do równań Maxwella. Praca [E1] jest podsumowaniem tego projektu. Następnie, by zaprezentowane przez nas techniki matematyczne mogły zostać użyte w środowisku inżynierów, konieczny był rozwój bardzo szybkich algorytmów implementujących teoretyczne obserwacje z [E1]. Praca [E2] jest ukoronowaniem kilku częściowych publikacji powstałych na przestrzeni ośmiu lat, w których udało nam się przedstawić coraz to szybsze algorytmy obliczające wymagane niezmienniki topologiczne. Przedstawiamy w niej najszybszy obecnie dostępny algorytm oraz implementację umożliwiającą poprawne użycie Dyskretnego Geometrycznego Podejścia do równań Maxwella w praktyce. Na tym etapie należy nadmienić, że implementacja przedstawiona w [E2] została wdrożona przez firmę EMWorks do jej komercyjnego oprogramowania służącego do modelowania różnych komponentów elektrycznych (<https://www.emworks.com/>)

2 Algorytmy i implementacja: czy są istotne?

W przypadku praktycznych zastosowań matematyki i informatyki implementacje proponowanych algorytmów mają kluczowe znaczenie i są nieodłącznym uzupełnieniem wyników teoretycznych. Takie implementacje nie tylko świadczą o tym, że proponowane algorytmy można w praktyce zastosować, ale przede wszystkim są narzędziem używanym przez liczną grupę naukowców z całego świata. To one, w ostatecznym rozrachunku, pozwalają na praktyczne wykorzystanie wyników teoretycznych.

Należy zaznaczyć, że udostępnienie dla środowiska naukowego poprawnej i czytelnej implementacji jest zadaniem trudnym i żmudnym; wymaga nie tylko wyobraźni, głębokiej wiedzy programistycznej i cierpliwości, ale również wiedzy z zakresu inżynierii oprogramowania oraz metod testowania oprogramowania. Poniżej przedstawiam listę bibliotek programistycznych zaimplementowanych przeze mnie w okresie pomiędzy uzyskaniem stopnia doktora a chwilą obecną. Te biblioteki są integralną częścią niniejszej rozprawy habilitacyjnej.

1. BallMapper pakiet CRAN zaimplementowany w języku R, <https://cran.r-project.org/web/packages/BallMapper/index.html> – pakiet ten został stworzony całkowicie przez mnie.
2. Persistence Landscape Toolbox. Toolbox to obliczeń statystycznych i uczenia maszynowego (<https://www2.math.upenn.edu/~dlotko/persistenceLandscape.html>) - również dostępny w rozbudowanej wersji jako część biblioteki Gudhi. Pakiet ten został stworzony całkowicie przez mnie i jest częścią rozprawy habilitacyjnej.
3. ThinIt software, toolbox do obliczania szkieleatów 2,3 oraz 4 wymiarowych modeli (<http://www.comphys.com/thinit.html>). Pakiet ten został stworzony całkowicie przez mnie i jest częścią rozprawy habilitacyjnej.
4. Implementacja kompleksów kostkowych, reprezentacji *persistence diagrams* oraz kilka dodatkowych pakietów w bibliotece Gudhi - w tym przypadku kod przeszedł skrupulatny proces recenzji (dostępne w Gudhi 1.3.0 lub wyższej, <http://gudhi.gforge.inria.fr/>). Pakiet ten został stworzony całkowicie przez mnie i jest częścią rozprawy habilitacyjnej.
5. Implementacja metod przypisywania niezmienników do drzew zanurzonych w \mathbb{R}^3 . Jest ona głównie wykorzystywana to klasyfikacji różnych

typów neuronów. Początkowa wersja implementacji dostępna pod adresem <https://github.com/reemkhalilneurolab/morphology> została zaimplementowana w pakiecie programistycznym Matlab. Pakiet ten został stworzony wspólnie z Ahmadem Farhadem.

3 Szybkie obliczenia *persistent homology*

Persistent homology jest jednym z najbardziej udanych narzędzi topologicznej analizy danych. Idea, by skupiać uwagę na homologiach rosnącej filtracji bardziej niż na homologiach całej przestrzeni, pojawiała się w pracach Vanessy Robins oraz Massimo Ferriego i kierowanej przez niego grupy z Bolonii już w latach dziewięćdziesiątych. Jednak po raz pierwszy pomysł ten został sformalizowany przez Herberta Edelsbrunnera w sposób ogólny w pracy [1].

Przypomnijmy pokrótce definicję *persistent homology*. *Filtracją* kompleksu sympleksialnego \mathcal{K} nazywamy wstępujący ciąg podzbiorów kompleksu \mathcal{K}

$$\emptyset \subset \mathcal{K}_0 \subset \mathcal{K}_1 \subset \dots \subset \mathcal{K}_{n-1} \subset \mathcal{K}_n = \mathcal{K}$$

taki, że każdy z \mathcal{K}_i , dla $i \in \{0, \dots, n\}$, jest kompleksem sympleksialnym. Powyższy ciąg inkluzji indukuje następujący ciąg homomorfizmów w grupach homologii w dowolnym wymiarze k :

$$0 \rightarrow H_k(\mathcal{K}_0) \rightarrow H_k(\mathcal{K}_1) \rightarrow \dots \rightarrow H_k(\mathcal{K}_{n-1}) \rightarrow H_k(\mathcal{K}_n) = H_k(\mathcal{K})$$

Mając dane $0 \leq i \leq j \leq n$ rozważamy homomorfizm $f_k^{i,j} : H_k(\mathcal{K}_i) \rightarrow H_k(\mathcal{K}_j)$ indukowany przez inkluzję $\mathcal{K}_i \hookrightarrow \mathcal{K}_j$. *Persistent homology* jest formalnie zdefiniowane jako obrazy tego homomorfizmu.

Przedstawmy teraz nieformalną intuicję tego pojęcia. Niech będzie dany cykl $0 \neq c \in H_k(\mathcal{K}_j)$. Mówimy, że cykl c *jest utworzony* w $H_k(\mathcal{K}_j)$ o ile c nie jest w obrazie $f_k^{i,j}$ dla $i < j$. Innymi słowy, w tej sytuacji, niezerowy cykl c pojawia się po raz pierwszy w $H_k(\mathcal{K}_j)$. Mówimy, że cykl c *jest zakończony* w $H_k(\mathcal{K}_i)$ o ile $c \in H_k(\mathcal{K}_i)$ oraz albo $0 = c \in H_k(\mathcal{K}_j)$ dla $i < j$ albo $0 \neq [c_1] = [c] \in H_k(\mathcal{K}_j)$, dla $i < j$, oraz cykl c_1 został utworzony wcześniej niż c . Intuicyjnie, cykl c jest zakończony w $H_k(\mathcal{K}_j)$, o ile staje się on trywialny w następnym kompleksie albo w tym kompleksie staje się równy innemu cyklowi, który został utworzony wcześniej.

Dla cyklu c który został utworzony w $H_k(\mathcal{K}_i)$ oraz zakończony w $H_k(\mathcal{K}_j)$, para liczb (i, j) jest zwana *persistence pair* cyklu c . Zbiór wszystkich *persistence pairs* nazywamy *persistence diagram*. Ponieważ dla każdej *persistence pair* (i, j) , mamy $i \leq j$, *persistence diagram* jest zazwyczaj reprezentowany jako zbiór punktów w \mathbb{R}^2 leżących powyżej prostej $y = x$.

Już w pracy [1] zaobserwowano, że podstawowy algorytm do obliczania *persistent homology* jest szczególnym przypadkiem diagonalizacji Gaussa macierzy brzegu kompleksu filtrowanego – zatem jego pesymistyczna złożoność obliczeniowa¹ wynosi $O(n^3)$, gdzie n jest liczbą sympleksów w rozważanym kompleksie. Istnieją przykłady rodzin kompleksów symplecjalnych, dla których algorytm ten osiąga swoją pesymistyczną złożoność.

Jednocześnie, w przypadku obliczeń standardowych grup homologii, grupa krakowska wypracowała liczne algorytmy upraszczające dane, z zachowaniem ich grup homologii (w szczególności [2] oraz [3]), które w praktycznych przypadkach istotnie przyspieszają obliczenia. W tym przypadku, grupy homologii były użyte do zdefiniowania homologicznego Indeksu Conleya.

W pracy [S1] metody redukcji zostały zaadoptowane do kompleksów z filtracją. Główną ideą pracy jest obserwacja, że parę redukcyjną lub ko-redukcyjną (σ, τ) można usunąć z rozważanego kompleksu z filtracją, o ile wartość filtracji na której pojawia się sympleks σ jest taka sama jak wartość filtracji sympleksu τ . Należy zaznaczyć, że takie redukcje są bardzo szybkie, gdyż ich przeprowadzenie nie wymaga żadnych obliczeń algebraicznych i dlatego może być wykonane na poziomie struktury danych reprezentującej kompleks symplecjalny. W szczególności prawdziwe jest następujące twierdzenie:

Twierdzenie (Twierdzenie 4.2 w [S1])

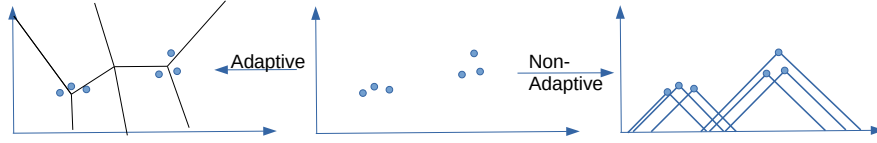
*Niech \mathcal{K} będzie kompleksem z filtracją daną przez funkcję $g : \mathcal{K} \rightarrow \mathbb{R}$. Niech $(\sigma, \tau) \in \mathcal{K} \times \mathcal{K}$ będzie parą (ko)redukcyjną taką, że $g(\sigma) = g(\tau)$. Wtedy *persistent homology* kompleksu \mathcal{K} oraz kompleksu $\mathcal{K} \setminus \{\sigma, \tau\}$ są izomorficzne.*

O ile sama metoda nie została przez nas zaimplementowana, jest ona pośrednio użyta w pakiecie Perseus [4], który do niedawna był wiodącym narzędziem do obliczeń *persistent homology*.

4 *Persistence diagrams*, statystyka i ich wektoryzacja

Od czasów zdefiniowania *persistence diagrams* w 2002 roku w pracy [1], przez ponad 10 lat były one używane do opisu danych wyłącznie w subiektywny sposób - poprzez wizualne porównanie kolekcji dwóch, lub rzadko większej ilości diagramów. Powodem tego stanu rzeczy jest to, że *persistence dia-*

¹Przez pesymistyczną złożoność algorytmu rozumiemy maksymalną teoretycznie możliwą liczbę kroków wykonanych przez algorytm, będącą funkcją rozmiaru danych wejściowych.



Rysunek 1: Adaptacyjne versus nie adaptacyjne metody wektoryzacji.

grams są chmurami punktów w \mathbb{R}^2 o zmiennej wielkości. Przestrzeń *persistence diagrams*, mimo że jest przestrzenią polską, nie jest ani przestrzenią Banacha, ani tym bardziej przestrzenią Hilberta. W związku z tym, nie jest to przestrzeń, na której można stosować tradycyjne metody statystyki i nauczania maszynowego. Dodatkowo nie jest znana jednoznaczna definicja średniej dwóch lub więcej *persistence diagrams*.

Pierwsza teoretyczna metoda oferująca włożenie przestrzeni *persistence diagrams* do przestrzeni funkcji $L^p(\mathbb{R} \times \mathbb{N}, \mathbb{R})$, znana jako *persistence landscapes*, została zaproponowana w pracy [5]. Niedługo potem, w pracy [V1] przedstawiliśmy efektywne obliczeniowo algorytmy konstrukcji *persistence landscapes*. Sam algorytm konstrukcji jest skomplikowany - wykorzystuje on odpowiednio zmodyfikowaną ideę *sweep line*, która mając na wejściu *persistence diagrams*, zwraca indeksowany przez spójny prefiks liczb naturalnych, ciąg funkcji kawałkami liniowych posiadających własność, że n -ta funkcja jest nie mniejsza niż funkcja $(n + 1)$ -sza. Konsekwentnie, wykres wszystkich funkcji przypomina nachodzące na siebie łańcuchy górskie, stąd nazwa.

Dodatkowo, oprócz algorytmów i implementacji, praca [V1] zawiera liczne przykłady użycia proponowanych metod - w szczególności przedstawiamy tam użycie *persistence landscapes* do testu permutacyjnego oraz prostych klasyfikatorów. Porównujemy też prezentowaną przez nas reprezentację dokładną z alternatywną reprezentacją przybliżoną (gdzie funkcje kawałkami liniowe są przybliżane na jednostajnej siatce punktów).

Jak wspomniano wcześniej, algorytmy i implementacja przedstawiona w [V1] dały pierwszy na świecie działający pakiet łączący *persistent homology* z metodami statystyki, nauczania maszynowego oraz sztucznej inteligencji.

W kolejnej pracy [V2] zajmowałem się *adaptacyjną* wektoryzacją *persistence diagrams*. Aby uzmysłowić sobie sens podejścia adaptacyjnego, wyobraźmy sobie sytuację, w której jakiś proces produkuje *persistence diagrams*, których punkty są skoncentrowane w pewnym konkretnym obszarze \mathbb{R}^2 , jak przedstawiono na Rysunku 1.

W tym przypadku, jeżeli punkty z kolejnych *persistence diagrams* różnią się nieznacznie od pozostałych, wtedy *persistence landscapes* skonstruowane na ich podstawie będą bliskie w normie L^p . Jednak, czasami zależy nam na uwypukleniu subtelniejszych różnic pomiędzy diagramami. Należy zaznaczyć, że jest to przypadek często spotykany w praktyce. W tej sytuacji powinniśmy zastosować metody adaptacyjne, które mówiąc nieformalnie, dają lepszą rozdzielczość w rejonach diagramu, w których gęstość punktów jest odpowiednio duża.

Metoda wektoryzacji przedstawiona w pracy [V2] opiera się na wyborze podzbioru punktów L z diagramów D ($L \subset D$), budowie (w najprostszej wersji) komórek Voronoi'a punktów z L^2 . Niech będzie dany dobry porządek na punktach z L . Każdemu punktowi $l \in L$ przypisujemy liczbę punktów z D które są w komórce Voronoi odpowiadającej punktowi l . W ten sposób uzyskujemy wektor w $\mathbb{R}^{|L|}$ opisujący zbiór D .

Przedstawiona powyżej metoda, mimo że intuicyjnie jasna, ma poważne niedociągnięcie - nie jest ona stabilna; infinitesimalna zmiana wejściowego diagramu może istotnie zmienić wyjściowy wektor. W pracy [V2] rozważamy cały szereg metod, które są stabilne i oparte na podobnej idei jak przedstawiona powyżej. Moim oryginalnym wkładem w prace [V2] było przedstawienie dowodów stabilności wszystkich rozważanych tam metod. Jednym z głównych wyników pracy jest twierdzenie o stabilności tak zwanych *persistence VLAD* które podane jest poniżej:

Twierdzenie *Niech B i B' będą persistence diagrams, takimi, że $B, B' \subset [a, b] \times [a, b]$ oraz niech $\lambda = \{w_i, \mu_i, \Sigma_i, i = 1, \dots, N\}$ będzie Gaussian Mixture Model spełniającym warunek, że $w_i \neq 0$. Wtedy stabilny persistence VLAD w \mathbb{R}^N (jest to jedna z rozważanych metod wektoryzacji) jest funkcją Lipschitzowską względem odległości 1-Wassersteina, to znaczy:*

$$\|\mathbf{v}^{\text{SPVLAD}}(B) - \mathbf{v}^{\text{SPVLAD}}(B')\|_{\infty} \leq C \cdot W_1(B, B'),$$

gdzie C jest stałą zależną od $[a, b] \times [a, b]$.

5 Od matematyki ciągłej do dyskretnej

Relacje pomiędzy matematyką ciągłą i dyskretną nie są łatwe ani trywialne. Często, próbując rozwiązać jakiś problem, naukowcy zajmujący się matematyką czystą sięgają po metody komputerowe, szukając inspiracji dla rozwią-

²Komórkę Voronoi'a dla punktu $p \in L$ nazywamy zbiór punktów $x \in \mathbb{R}^2$, które są bliższe p niż innym punktom z L .

zania analitycznego. Jednak, bardzo często, z wyłączeniem problemów matematyki dyskretnej, nie mamy żadnych gwarancji, że rozwiązanie obliczone nam przez komputer jest w jakikolwiek sposób bliskie ciągłemu rozwiązaniu problemu.

Próby użycia komputerów do uzyskiwania ścisłych matematycznie rozwiązań mają obecnie około pięćdziesięcioletnią historię. W Polsce są one aktywnie rozwijane w szczególności przez grupę Mariana Mrozka i Piotra Zgliczyńskiego z Uniwersytetu Jagiellońskiego, z której się wywodzę.

W pracy [CD1] zajmuję się zagadnieniem ścisłej dyskretyzacji funkcji ciągłych. Załóżmy, że mamy daną funkcję $f : Z \rightarrow \mathbb{R}^m$, gdzie $Z \subset \mathbb{R}^n$ jest zbiorem zwartym. W rozważanym przez nas przypadku Z jest n -wymiarowym prostopadłością w \mathbb{R}^n . Dla celów pracy, zbiór Z będzie podzielony na tak zwany kompleks prostopadłościów, to jest kolekcję n -wymiarowych domkniętych prostopadłościów \mathcal{Q} które w całości wypełniają Z i przecinają się wyłącznie na wspólnych krawędziach. We wspomnianej pracy zaproponowaliśmy algorytm częściowy³ zwracający przybliżenie funkcji f , oznaczone przez $\square f$, jako funkcję kawałkami stałą. Jest ona stała na każdym prostopadłościu z rodziny \mathcal{Q} . Co więcej, dla zadanej na wejściu stałej $\epsilon > 0$, o ile algorytm zakończy działanie, mamy gwarancje, że

$$\|(f - \square f)|_Z\|_\infty \leq \epsilon.$$

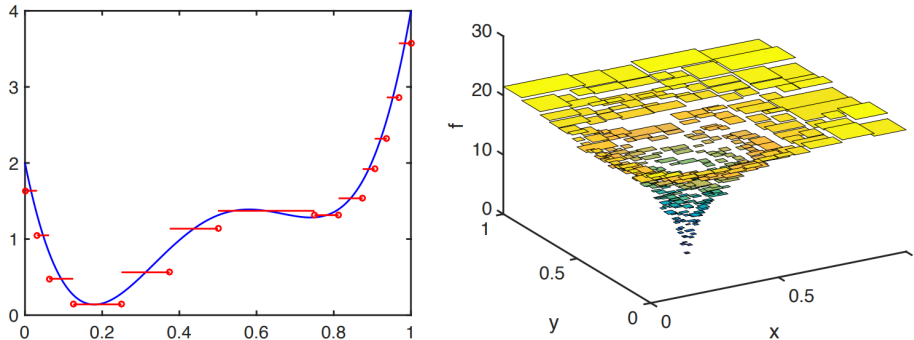
Przykład ilustrujący takie aproksymacje jest pokazany na Rysunku 2.

Głównym narzędziem umożliwiającym nam uzyskanie aproksymacji o tych własnościach jest arytmetyka przedziałowa [6] oraz algorytmy kombinatoryczne konstrukcji pokrycia \mathcal{Q} zbioru Z . Naturalnym jest tu pytanie, czy dla pewnej klasy funkcji, nasz algorytm, przy odpowiednich założeniach, zawsze poda poprawną odpowiedź. O ile, dla dowolnej funkcji f , takie twierdzenie jest fałszywe, o tyle, można je udowodnić dla funkcji Lipschitzowskich przy założeniu, że mamy dostatecznie dokładną reprezentację liczb rzeczywistych będących końcówkami naszych przedziałów. Mówi o tym następujące twierdzenie (Twierdzenie 3.2 w [CD1]):

Twierdzenie (Złożoność obliczeniowa dla funkcji Lipschitzowskich).

Niech $R \subset \mathbb{R}^n$ będzie zwartym prostopadłością oraz funkcja $f : R \rightarrow \mathbb{R}^m$ będzie funkcją Lipschitzowską ze stałą Lipschitza L . Załóżmy, że algorytm zaprezentowany w pracy [DC1] uruchomiony na komputerze, który ma możliwość reprezentowania liczb rzeczywistych z nieskończoną precyzją, dla funk-

³Algorytm częściowy ma własność, że może zakończyć działanie nie zwracając odpowiedzi. Innymi słowy, w przeciwieństwie do standardowego algorytmu, dla pewnych danych wejściowych może on nie przedstawić oczekiwanego wyniku, ale odpowiedzieć 'nie wiem'.



Rysunek 2: Przykłady aproksymacji funkcji: Po lewej, $f(x) = 2 - 25x + 108x^2 - 162x^3 + 81x^4$ na przedziale $[0, 1]$ ze stałą $\epsilon = \frac{1}{2}$. Kolorem niebieskim przedstawiamy wykres funkcji, zaś kolorem czerwonym przedstawiamy elementy $\square f$ aproksymacji funkcji f . Po prawej, widzimy aproksymację funkcji Ackley'a [7] ze stałą $\epsilon = 2$.

cji f oraz stałej ϵ obliczył reprezentacje $\square f$ funkcji f . Wtedy otrzymany podział prostopadłościanu R będzie miał co najwyżej

$$\frac{\alpha(L)\text{diam}(R)}{2\epsilon}$$

n wymiarowych prostopadłościanów. Stała $\alpha(L)$ zależy od stopnia skomplikowania wzoru funkcji g (jak dokładnie opisano w pracy) zaś $\text{diam}(R)$ oznacza średnicę prostopadłościanu R w normie maksimum.

Zaprezentowana w pracy [CD1] aproksymacja ma potencjalnie bardzo szeroki zakres stosowań. Jako jeden z przykładów podaliśmy ścisłą aproksymację *persistent homology* dla funkcji ciągłych. W ten sposób uzyskaliśmy jedyną istniejącą w chwili obecnej metodę ścisłego obliczania *persistent homology* dla funkcji ciągłych.

6 Zastosowania w inżynierii elektrycznej

Historia moich badań we współpracy z inżynierami elektrykami, w szczególności z prof. Rubenem Specogną z Uniwersytetu w Udine sięga początków moich studiów doktoranckich w 2008 roku. Dzisiaj, pomimo upływu prawie piętnastu lat, nasze więzi naukowe wciąż są bardzo silne. To właśnie interakcja z inżynierami przekonała mnie, że wiele dziedzin nauki i techniki potrzebuje współpracy z matematykami i informatykami. Matematycy

są nieocenieni w poprawnym i formalnym zdefiniowaniu i wyabstrahowaniu problemu nad którym pracują inżynierowie. Informatycy zaś pomagają w przygotowaniu najbardziej efektywnych algorytmów i ich implementacji, tak by wypracowana w tym interdyscyplinarnym środowisku metoda była użyteczna jako całość i mogła zostać zastosowana w praktyce.

Jako część mojej rozprawy habilitacyjnej pragnę przedstawić pracę [EE1], która jest przykładem matematycznej formalizacji problemu jak i pracę [EE2], która podsumowuje długą drogę rozwoju szybkich algorytmów do obliczania niezmienników znalezionych w [EE1]. Praca [EE3] przedstawia algorytm umożliwiający obliczenia reprezentantów generatorów kohomologii dla dwuwymiarowych rozmaitości. Jest on ważnym elementem składowym algorytmu przedstawionego w [EE2].

W pracy [EE1] zajmujemy się tak zwanym Dyskretnym Geometrycznym Podejściem (DGP) do Równań Maxwella [8] dla przypadku prądów zmiennych o niskiej częstotliwości. W tej sytuacji można uznać, że prąd przesunięcia w równaniach Maxwell'a jest zerowy. DGP nie bazuje jednak na dyskretyzacji równań różniczkowych. Zakłada ono, że dziedzina $K \subset \mathbb{R}^3$, na której poszukujemy rozwiązania jest zadana przez kompleks sympleksialny \mathcal{K} . Ten kompleks jest sumą dwóch podkompleksów: \mathcal{K}_c – kompleksu modelującego przewodnik oraz \mathcal{K}_a – kompleksu modelującego izolator. Te dwa kompleksy stykają się na sumie dwuwymiarowych rozmaitości będących ich wspólną częścią.

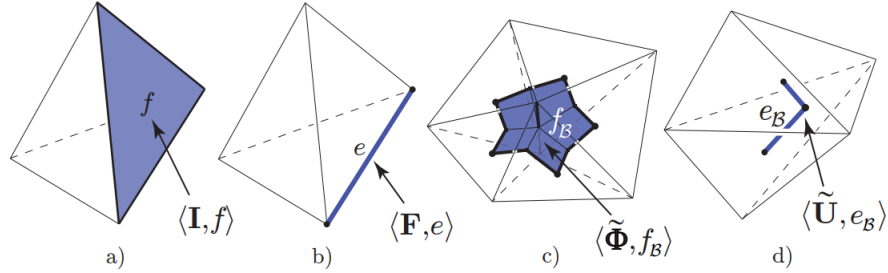
Dla danego kompleksu sympleksialnego \mathcal{K} metoda DGP polega, w pierwszym etapie, na przypisaniu zmiennym fizycznym geometrycznych elementów kompleksu \mathcal{K} , oraz kompleksu do niego dualnego⁴. Poglądowo to przyporządkowanie jest przedstawione na Rysunku 3. I tak, dla przykładu, prąd elektryczny, \mathbf{I} , jest łańcuchem⁵ zbudowanym na dwuwymiarowych sympleksach w \mathcal{K} . Siła Magnetomotoryczna, \mathbf{F} , jest łańcuchem zbudowanym na jednowymiarowych sympleksach (krawędziach) w \mathcal{K} i tak dalej.

Mając związek zmiennych fizycznych z geometrią, możemy wypisać dyskretne odpowiedniki równań Maxwell'a, które w tym przypadku są układem algebraicznych równań liniowych. Dla przykładu;

1. Prawo zachowania ładunku mówi, że $\langle \mathbb{I}, \partial v \rangle = 0$ dla każdego czworoscianu $v \in \mathcal{K}$ – innymi słowy, prąd który wpływa do v , musi z v wypłynąć,

⁴Którego, dla prostoty tekstu, nie będziemy tutaj definiowali, dalsze szczegóły są dostępne w [EE1].

⁵Najprościej rzecz ujmując łańcuch jest funkcją z rodziny sympleksów pewnego wymiaru w \mathbb{R} posiadającą pewne dodatkowe własności.



Rysunek 3: Powiązanie zmiennych fizycznych z geometrycznymi elementami kompleksów rozpinających dziedzinę obliczeń K .

2. Równanie Ampere'a ma postać $\langle \mathbb{I}, f \rangle = \langle \mathbb{F}, \partial f \rangle$ dla każdego dwuwymiarowego sympleksu $f \in \mathcal{K}$ - innymi słowy, prąd przepływający przez dwuwymiarowy sympleks f jest równy sile magnetomotorycznej na brzegu f .

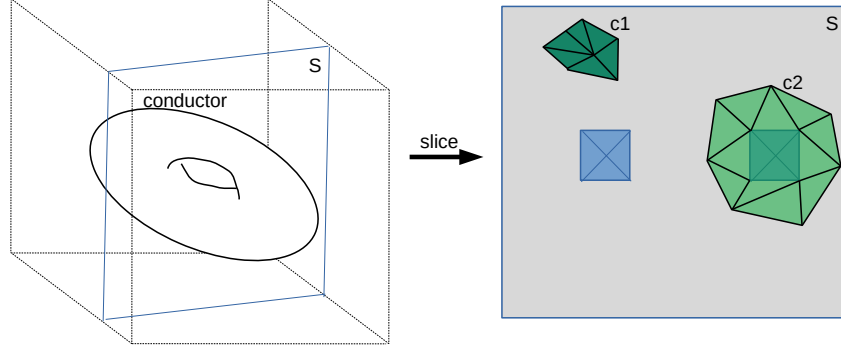
Brakujące równania, będące dyskretnymi odpowiednikami praw Faradaya oraz Gaussa, nie zostały tu przedstawione w celu oszczędności miejsca. Są one również równaniami liniowymi. Otrzymany całosciowy układ równań algebraicznych można rozwiązać przy pomocy deterministycznych lub przybliżonych iteracyjnych metod rozwiązywania układów równań. Jednakże, jest bardzo wiele wewnętrznych zależności pomiędzy niewiadomymi (zmiennymi fizycznymi), które powodują, że wskaźnik uwarunkowania rozważanego układu liniowego jest bardzo wysoki. Ponadto, liczba niewiadomych (czyli rozmiar układu), jest istotnie większa od minimalnej. W związku z tym metody iteracyjne, które są jedyną praktycznie dostępną opcją dla dużych układów równań, albo nie zbiegają do rozwiązania, albo zbiegają do niego bardzo powoli.

By rozwiązać ten problem, wprowadza się dwa potencjały - *Magnetic Scalar Potential* \mathbb{T} , zdefiniowany na krawędziach \mathcal{K}_c oraz *Electric Vector Potential* Ω , zdefiniowany na wierzchołkach \mathcal{K} .

Każda zmienna fizyczna może zostać obliczona przy użyciu tych dwóch potencjałów. W szczególności, siła magnetomotoryczna jest wyrażona wzorem:

$$\mathbb{F} = T + \delta\Omega + \text{poprawka topologiczna} \quad (1)$$

Liczba niewiadomych w układzie równań, w którym obliczamy wyłącznie wartości T oraz Ω jest średnio około 10 razy mniejsza niż dla pełnego układu równań opisanego powyżej. Również wskaźnik uwarunkowania układu ba-



Rysunek 4: Po prawej, szkic przewodnika w kształcie wypełnionego torusa wraz z płaszczyzną S rozcinającą go na pół. Po prawej rzut samej płaszczyzny S . Na niebiesko oznaczono elementy torusa przeciętego przez S . Dwa jednowymiarowe cykle c_1 oraz c_2 wraz z dwuwymiarowymi brzegami b_1 oraz b_2 mającymi własność, że $\partial b_1 = c_1$ oraz $\partial b_2 = c_2$ są widoczne na rysunku po prawej.

zującego na potencjałach jest istotnie lepszy. Te własności sprawiają, że DGP, wraz z poprawką topologiczną z [EE1], jest metodą bardzo efektywną obliczeniowo.

Wyjaśnienie konieczności użycia poprawki topologicznej w równaniu 1, zobrazujemy przy pomocy Rysunku 4.

Załóżmy na moment, że poprawka topologiczna nie jest potrzebna i że siła magnetomotoryczna może być wyrażona prostszym wzorem $\mathbb{F} = T + \delta\Omega$. Wiemy, że T jest zerowy na przedstawionych krawędziach, ponieważ są one zawarte w całości w kompleksie \mathcal{K} , zatem powyższy wzór redukuje się do $\mathbb{F} = \delta\Omega$. Ponadto, skoro Ω jest zerowymiarowym kocyklem, z definicji kocyklu ($\delta\delta\Omega = 0$) otrzymujemy:

$$\langle \mathbb{F}, \partial s \rangle = \langle \delta\Omega, \partial s \rangle = \langle \delta\delta\Omega, s \rangle = 0$$

dla s będącego zarówno 2-łańcuchem, którego brzegiem jest cykl c_1 jak i cykl c_2 . W przypadku cyklu c_1 ten wynik jest poprawny – ponieważ łańcuch s znajduje się w izolatorze, nie przepływa przez niego żaden prąd. Jednak, ten wynik nie jest poprawny w przypadku cyklu c_2 . Zauważmy, że cykl c_2 ma własność, że każdy 2-łańcuch którego jest on brzegiem, musi zawierać elementy przewodnika \mathcal{K}_c . Przez te elementy przewodnika będzie płynął

niezerowy prąd, zatem wartość siły magnetomotorycznej na c_2 powinna być niezerowa.

Podstawowym wynikiem pracy [EE1] jest teoretyczne zdefiniowanie wspomnianej powyżej topologicznej poprawki. Jej sens można nieformalnie wynioskować z Rysunku 4; Weźmy cykl c , który jest nietrywialny w $H_1(\mathcal{K}_a)$ (lub jest całkowitą wielokrotnością takiego nietrywialnego cyklu). Wartość \mathbb{F} na c , $\langle \mathbb{F}, c \rangle$, musi być równa prądowi w przewodniku – który jest parametrem metody – lub odpowiedniej wielokrotności wartości tego prądu dla cyklu c okrążającego przewodnik wielokrotnie.

Jak wykazaliśmy formalnie w pracy [EE1], wzięcie za poprawkę topologiczną kołańców będących generatorami pierwszej grupy kohomologii \mathcal{K}_a , $H^1(\mathcal{K}_a)$, w pełni rozwiązuje problem i czyni z DGP spójny system z jednoznacznym rozwiązaniem. Formalnie, wspomniany wynik jest przedstawiony w poniższym twierdzeniu:

Twierdzenie (Reprezentanci generatorów kohomologii wystarczą do spójności DGP)

Niech $\{c_i\}_{i=1}^{\beta_1(\mathcal{K}_a)}$ będą reprezentancami generatorów $H^1(\mathcal{K}_a)$. Niech $\{\partial\sigma_i\}_{i=1}^{\beta_1(\mathcal{K}_a)}$ będą dualnymi cyklami do nich (w sensie twierdzenia o współczynnikach uniwersalnych dyskutowanego w [EE1]). Jeżeli zdefiniujemy siłę magnetomotoryczną jako $\mathbf{F} = \delta\mathbf{\Omega} + \mathbf{T} + \sum_{i=1}^{\beta_1(\mathcal{K}_a)} j_i c_i$, to prądy przepływające przez 2-łańcuchy σ_i (przecinające przewodnik \mathcal{K}_c) będą równe j_i Amperów. Wtedy Magnetic Scalar Potential oraz Electric Vector Potential dają jednoznaczne rozwiązanie DGP.

Algorytm przedstawiony w pracy [EE1], pomimo że poprawny teoretycznie, był jednak zdecydowanie zbyt powolny, by rozwiązać problemy o wielkościach rozważanych w zastosowaniach. Należy wspomnieć, że był to ogólny algorytm do obliczania grup kohomologii dla dowolnych kompleksów. Jednak, w zastosowaniach inżynierskich ilość trójwymiarowych sympleksów w kompleksie \mathcal{K} często jest liczona w milionach - dlatego też standardowe algorytmy są zbyt wolne. Ponieważ naszym celem było nie tylko teoretyczne rozwiązanie problemu, ale przedstawienie narzędzia, które jest praktycznie użyteczne, więc nie poprzestaliśmy na wyniku teoretycznym z pracy [EE1]. W następnych latach rozwinięliśmy całą serię algorytmów zwieńczoną algorytmem i implementacją opisaną w pracy [EE2]. W ten sposób rozwinięliśmy bardzo szybki algorytm, który sprostał oczekiwaniom inżynierów. Dla przykładu, dla pewnego kompleksu sympleksialnego modelującego induktor, posiadającego 2 miliony sympleksów w wymiarze 3, algorytm przedstawiony w [EE1] nigdy nie zakończył działania. Nasze kolejne algorytmy, uruchomione na tym samym przykładzie, potrzebowały kolejno 60000, 25 sekund

i ostatecznie 1 sekundę aby wykonać obliczenia. Pozostała część obliczeń w tym przypadku zajmuje około 30 sekund. Jest zatem jasne, że algorytm przedstawiony w [EE2] potrzebuje tylko ułamek czasu całościowych obliczeń i może być z powodzeniem stosowany jako przygotowanie do obliczeń.

Algorytm przedstawiony w [EE2] został nie tylko doceniony w środowisku naukowym. Jest on również używany w przemyśle przez firmę EMWorks (<https://www.emworks.com/>)⁶ co jest przykładem przemysłowego wdrożenia wyników mojej pracy naukowej.

W pracy [EE3] rozważam problem obliczania generatorów kohomologii dwuwymiarowych rozmaitości (z brzegiem i bez). O ile problem ten nie jest bezpośrednio związany z zastosowaniami metod topologicznej analizy danych w elektromagnetyzmie, jest on istotnym składnikiem rozwiązania z pracy [EE2], dlatego został omówiony w tym rozdziale. Przedstawiony w pracy [EE3] algorytm, w najprostszym przypadku, działa w czasie $O(n)$, gdzie n jest ilością sympleksów na danej powierzchni. Optymalne (w pewnym sensie) generatory mogą zostać obliczone w czasie $n \log(n)$. O ile kombinatoryczne metody obliczania generatorów homologii dla takich powierzchni były znane, uzyskanie generatorów kohomologii wymagało zastosowania, w sposób konstruktywny, twierdzenia Poincarego o dualności między grupami homologii i kohomologii dla rozmaitości (z brzegiem lub bez), co jest kluczowym wynikiem pracy [EE3].

Kombinatoryczny algorytm przedstawiony w pracy [EE3] znalazł zastosowanie w metodach obliczania poprawki topologicznej dla DGP do równań Maxwella w pracy [EE2]. Uważam, że może on być również przydatny w grafice komputerowej, na przykład by odwzorować dwuwymiarowy obrazek lub teksturę na powierzchnię dwuwymiarową.

Algorytm ten operuje na podziale komórkowym (w najprostszym przypadku triangulacji) \mathcal{K} rozmaitości K . Mając dany kompleks \mathcal{K} oraz kompleks dualny $D(\mathcal{K})$ oblicza on generatory homologii $H_1(\mathcal{K})$ i kohomologii $H^1(\mathcal{K})$ danego kompleksu \mathcal{K} jak opisano poniżej;

Obliczenia bazy $H^1(\mathcal{K})$ zaczepionej w wierzchołku v

Wejście: Kompleks \mathcal{K} , kompleks dualny $D(\mathcal{K})$, wierzchołek $v \in \mathcal{K}$

- Niech $T \subset \mathcal{K}_1$ będzie drzewem najkrótszych ścieżek z v do pozostałych wierzchołków \mathcal{K} .

⁶Mamy również nieformalne informacje, że inne firmy oferujące oprogramowanie do modelowania w elektromagnetyzmie również stosują nasze wyniki, jednak nigdy nie udało nam się uzyskać oficjalnego potwierdzenia. Niestety, kody źródłowe takich platform nie są dostępne, zatem nie jest możliwe sprawdzenie, jakie techniki są używane w obliczeniach.

- Każda krawędź $e \in \mathcal{K}_1 \setminus T$ zamyka, wraz z T , cykl w \mathcal{K}_1 zawierający wierzchołek v . Niech $\sigma(e)$ będzie długością tego cyklu.
- Niech T' będzie maksymalnym drzewem rozpinającym na $D(\mathcal{K}_1) \setminus D(T)$. Waga krawędzi $D(e)$, dla $e \in D(\mathcal{K}_1) \setminus D(T)$, jest dana przez $\sigma(e)$.
- Niech $E = \{e \in \mathcal{K}_1 \mid e \notin T \text{ oraz } D(e) \notin T'\}$.
- Cykle w T zamknięte przez krawędzie z E generują $H_1(\mathcal{K})$.
- Cykle w T' zamknięte przez krawędzie z $D(E)$ generują $H^1(\mathcal{K})$.

Złożoność obliczeniowa powyższego algorytmu wynosi $O(n \log(n))$, gdzie n jest ilością sympleksów w \mathcal{K} .

7 Pozostałe osiągnięcia naukowo-badawcze

Metody geometrii i topologii znajdują zastosowania w bardzo szerokim spektrum nauk technicznych, przyrodniczych, medycznych, socjalnych oraz politycznych. W tym rozdziale pokrótce omówię, w podziale na typy zastosowań, pozostały dorobek naukowy od czasu uzyskania stopnia doktora do chwili obecnej.

7.1 Zastosowania w naukach neurologicznych

Neurologia jest bardzo ciekawym polem dla zastosowań metod topologicznej analizy danych. W swojej pracy naukowej zajmowałem się analizą skierowanego grafu połączeń między neuronami w kolumnie neuronalnej szczura. Model ten został zaproponowany przez naukowców z Blue Brain Project w pracy [9]. Wynik naszej analizy jest przedstawiony w pracy (A), gdzie zaproponowaliśmy techniki uzyskiwania niezmienników topologicznych takiego grafu skierowanego. Następnie te niezmienniki zostały wykorzystane do badania własności zarówno statycznego modelu, jak i podgrafu kolumny neuronalnej, który ulega aktywacji pod wpływem różnych zewnętrznych bodźców.

Innym, nieco mniej abstrakcyjnym przykładem użycia technik topologicznych, jest analiza kształtu poszczególnych komórek neuronalnych występujących w mózgach zwierzęcych oraz ludzkich. Jest to ważny kierunek badań, ponieważ kształt neuronu silnie określa jego funkcje - dlatego neurologi często przy pomocy kształtu przyporządkowują neurony do danych *kategorii*, gdzie każda kategoria neuronów ma określoną funkcję w mózgu.

Dotychczas, od czasu pionierskich badań Santiago Ramóna y Cajal, przyporządkowanie neuronów do ich kategorii, odbywa się na podstawie opinii ekspertów. W pracy (B) przedstawiliśmy jeden z pierwszych uniwersalnych, i w pełni automatycznych, algorytmów umożliwiających klasyfikację neuronów. Techniki takiej klasyfikacji zostały jeszcze dalej rozwinięte w pracy (C), która jest obecnie w fazie recenzji.

7.2 Obliczania grupy podstawowej przestrzeni

Współczesna topologia obliczeniowa bazuje na grupach homologii danej przestrzeni. Jednakże klasyczna topologia algebraiczna poświęca sporo uwagi tak zwanej grupie podstawowej $\pi_1(X)$ dla łukowo spójnej przestrzeni topologicznej X , oraz tak zwanym wyższym grupom homotopii tejże przestrzeni. Jednak ta klasyczna konstrukcja nie jest powszechnie używana w topologii obliczeniowej w związku z trudnościami obliczeniowymi. O ile stosunkowo łatwo, mając reprezentację symplecjalną przestrzeni X , wypisać reprezentację grupy $\pi_1(X)$, to pytanie, czy dwie reprezentacje są równoważne, jest pytaniem o izomorfizm grup, które jest klasycznym problemem nierozstrzygalnym algorytmicznie.

W pracy (D) zauważyliśmy, że o ile problem izomorfizmu jest nierozstrzygalny w ogólności, to, używając heurystyk dostępnych w pakiecie *GAP - Groups, Algorithms, Programming* (gap-system.org), można zaproponować działający w praktyce algorytm częściowy, który dla pewnych reprezentacji pozwala stwierdzić czy są one równoważne. W przypadku, gdy heurystyka nie pozwala stwierdzić równoważności reprezentacji lub jej braku, algorytm zwraca odpowiedź "nie wiem". Nasza intuicja wskazuje, że im bardziej skomplikowana reprezentacja grupy, tym bardziej prawdopodobna jest ostatnia odpowiedź. W związku z tym, przed użyciem pakietu GAP, w pracy (D), używamy Dyskretnej Teorii Morse'a [10], do uproszczenia reprezentacji danej grupy, a dopiero potem, używamy metod dostępnych w pakiecie GAP, by spróbować rozstrzygnąć problem izomorfizmu dla zadanych reprezentacji.

To podejście pozwoliło nam na klasyfikację wszystkich węzłów, do 12 przecięć, przy pomocy grupy podstawowej ich dopełnienia. Uważamy, że przedstawiona w pracy (D) metoda może być używana w praktyce do badań zawężonych białek oraz łańcuchów DNA / RNA.

7.3 Znajdowanie szkieletu obiektów dwu, trój oraz czterowymiarowych

W pracy (E) zajmowaliśmy się ścisłymi oraz szybkimi metodami uzyskiwania szkieletu dwu, trzy oraz czterowymiarowych kompleksów sympleksyjnych oraz kostkowych. Zaproponowana przez nas metoda gwarantuje istnienie dyskretnej homotopii (w sensie Dyskretnej Teorii Morse'a [10]) przekształcającej zbiór wejściowy \mathcal{K} w jego szkielet $S(\mathcal{K})$. Taka homotopia jest dyskretnym odpowiednikiem odwzorowania ciągłego pomiędzy \mathcal{K} oraz $S(\mathcal{K})$, które w pewnym sensie gwarantuje, że kształt \mathcal{K} jest zachowany w $S(\mathcal{K})$.

Metoda przedstawiana w pracy (E) bazuje wyłącznie na komórkach najwyższego wymiaru w kompleksie \mathcal{K} , co gwarantuje jej wysoką efektywność obliczeniową. Jest to możliwe dzięki ponumerowaniu oraz zapamiętaniu wszystkich możliwych lokalnych konfiguracji wokół każdego sympleksu, lub kostki C , które świadczą o tym, że kompleks \mathcal{K} z komórką C jest homotopijnie równoważny kompleksowi \mathcal{K} bez komórki C . Sam proces znajdowania szkieletu \mathcal{K} polega na rekurencyjnym, zachłannym wyszukiwaniu oraz usuwaniu z \mathcal{K} sympleksów lub kostek C , o których mowa powyżej.

7.4 Zastosowania topologicznej analizy danych w naukach ekonomicznych

Metody topologicznej analizy danych znajdują również zastosowania w dziedzinach takich jak nauki polityczne, ekonomiczne i socjologiczne.

W pracy (G) przeprowadzamy analizę szeregów czasowych cen dwudziestu czterech podstawowych towarów na amerykańskim rynku w okresie ostatnich 30 lat. Analiza wzorców korelacji pomiędzy szeregami czasowymi cen wskazuje na istnienie czterech oddzielnych okresów, w których te korelacje były relatywnie niezmiennie. Okresy te są od siebie oddzielone trzema krótkimi przedziałami czasowymi, w których korelacje rozważanych szeregów czasowych wykazują bardzo dużą zmienność w krótkim czasie. Te przedziały zmienności jednoznacznie pokrywają się z największymi kryzysami ostatnich lat - "kryzysem dot-com", krachem finansowym w 2008 oraz w 2013 roku. W pracy (G) pokazaliśmy, że układ korelacji pomiędzy każdymi dwoma kryzysami finansowymi jest unikatowy tylko dla tego okresu a rynek, rozumiany jako układ dynamiczny, nigdy po kryzysie nie wraca do stanu, w którym był poprzednio.

W pracy (F) używamy opracowanego przeze mnie algorytmu *Ball Mapper* [11] do analizy danych, składających się z pięciu wskaźników charakteryzujących amerykańskie firmy pod kątem prawdopodobieństwa, że dana firma

zbankrutuje w nadchodzącym roku. W tym celu przeanalizowaliśmy dane z ostatnich pięćdziesięciu lat. Badaliśmy też standardową metodę regresji logistycznej zwanej *Altman Score*. Jest to metoda oceny ryzyka kredytowego używana przez banki w Stanach Zjednoczonych. Jeżeli dana firma ma wartość *Altman Score* poniżej pewnej minimalnej wartości, banki z reguły nie udzielą jej kredytu w obawie o niewypłacalność. W naszej pracy (F) pokazaliśmy, że biorąc pod uwagę historyczne dane, firmy które faktycznie w niedługim czasie zbankrutowały, są zlokalizowane jedynie w pewnym wąskim obszarze firm z niską wartością *Altman Score*. Ten wynik, potencjalnie może pozwolić otworzyć dyskusję o bardziej racjonalnej, bazującej na danych, polityce kredytowej.

Moje prace w zastosowaniach metod topologicznych do wspomnianych tutaj dziedzin nie ograniczają się tylko do powyższych dwóch pozycji. Jest to jednak stosunkowo nowy front badań, w którym nie mam jeszcze zbyt wielu opublikowanych wyników. Powyższe dwie publikacje zostały wybrane jako przykład tego, co matematyka i informatyka może zaproponować naukom politycznym i społecznym.

7.5 Badania nad Covid-19

W 2020. roku, w trakcie pandemii wirusa Sars-Cov-2, wraz z moją grupą badawczą wziąłem udział w kilku przedsięwzięciach związanych z badaniami nad warunkami, które sprzyjają rozprzestrzenianiu się wirusa. Naszą główną aktywnością, wraz z naukowcami z Oxfordu, było zbudowanie rozległej bazy danych o koronawirusie. Mieści się ona pod adresem <https://covid19.eng.ox.ac.uk/> i zawiera szerokie spektrum informacji epidemiologicznych (liczba zakażonych/zgonów/pacjentów którzy wyzdrowieli), informacji o nefarmaceutycznych interwencjach (takich jak restrykcje, obowiązek noszenia maseczek itp.), mobilności ludzi w trakcie pandemii oraz pogodzie. Dodatkowo baza danych zawiera wyselekcjonowane informacje ekonomiczne z *World Data Bank* oraz dane socjologiczne z *World Value Survey* i *European Value Survey*. Naszym głównym wkładem koncepcyjnym jest odwzorowanie dostępnych danych do ściśle określonych jednostek geograficznych. Cała baza danych jest bazą relacyjną i umożliwia obsługę zapytań w języku SQL. Praca opisująca omawianą bazę danych (H) jest obecnie w fazie recenzji. Baza danych jest wykorzystywana przez *Como Consortium*, como.bmj.com, do modelowania możliwych scenariuszy rozwoju pandemii w różnych regionach świata.

Elementy pozostałego dorobku naukowego

- (A) Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess and Henry Markram, *Cliques of neurons bound into cavities provide a missing link between structure and function*, *Frontiers in Computational Neuroscience* 11, 48, (2017).
- (B) Lida Kanari, Paweł Dłotko, Martina Scolamiero, Ran Levi, Julien Shillcock, Kathryn Hess, Henry Markram. *A topological representation of branching neuronal morphologies*, *Neuroinformatics* 16 (1), 3-13, (2018).
- (C) Reem Khalil, Sadok Kallel, Ahmad Farhat, Paweł Dłotko, *Topological Sholl Descriptors for Neuronal Clustering and Classification*, *bioRxiv*, (2020).
- (D) Piotr Brendel, Paweł Dłotko, Graham Ellis, Mateusz Juda, Marian Mrozek, *Computing fundamental groups from point clouds*, *Applicable Algebra in Engineering, Communication and Computing* 26 (1-2), 27-48, (2015).
- (E) Paweł Dłotko, Ruben Specogna, *Topology preserving thinning of cell complexes*, *IEEE Transactions on Image Processing* 23 (10), 4486-4495, (2014).
- (F) Wanling Qiu, Simon Rudkin, Paweł Dłotko, *Refining understanding of corporate failure through a topological data analysis mapping of Altman's Z-score model*, *Expert Systems with Applications* 156, 113475, (2020).
- (G) Devraj Basu, Paweł Dłotko, *The Four Seasons of Commodity Futures: Insights from Topological Data Analysis*, Available at SSRN 3506780, 2020.
- (H) Adam Mahdi, Piotr Błaszczuk, Paweł Dłotko, Dario Salvi, Tak-Shing Chan, John Harvey, Davide Gurnari, Yue Wu, Ahmad Farhat, Niklas Hellmer, Alexander Zarebski, Bernie Hogan, Lionel Tarassenko, *OxCOVID19 Database: a multimodal data repository for better understanding the global impact of COVID-19*, *medRxiv*, <https://doi.org/10.1101/2020.08.18.20177147>, (2020).

8 Działalność dydaktyczna

Moją działalność dydaktyczną można podzielić na dwa nurty: zajęcia ze studentami studiów magisterskich oraz warsztaty dla pracowników naukowych i osób zajmujących się analizą danych.

Jeżeli chodzi o pierwszą kategorię, podczas mojej pracy na Uniwersytecie Jagiellońskim, uczyłem głównie przedmiotów informatycznych, to jest:

1. Ćwiczenia do przedmiotu *Programowanie* (gdzie prezentowałem języki C/C++, C#, Python, Pearl, Lisp, oraz filozofie programowania obiektowo orientowanego). Zajęcia te prowadziłem co roku, pomiędzy 2007 do 2012.
2. Ćwiczenia do przedmiotu *Topologia klasyczna i obliczeniowa* w roku 2012.
3. Wykład *Wstęp do Informatyki* dla słuchaczy Studiów Podyplomowych dla nauczycieli w latach 2007, 2008.
4. Ćwiczenia do przedmiotu *Wprowadzenie do C#* w 2007 roku.

W Swansea University, z racji tego, że byłem pracownikiem Instytutu Matematyki, prowadziłem głównie zajęcia matematyczne, w szczególności:

1. Wykład *Methods of Algebra and Calculus*, prowadzone w latach 2019 i 2020.
2. Wykład *Numerical Methods with Matlab*, prowadzony w latach 2017 oraz 2018.
3. Wykład *Computational Mathematics*, prowadzony w latach 2017 oraz 2019.

Prowadziłem też wykłady dla pracowników naukowych oraz przedstawicieli biznesu oraz inżynierów danych, w szczególności:

1. 20 godzinny minikurs z topologii obliczeniowej i stosowanej dla pracowników i studentów studiów doktoranckich EPFL (który dał początek stale rozwijanemu podręcznikowi dostępnemu pod adresem <https://arxiv.org/abs/1807.08607>).
2. Dwukrotnie prezentowałem metody topologicznej analizy danych na Data Science Summer School organizowanym przez Ecole Polytechnique w Paryżu - w 2018 (6 godzin) oraz w 2021 (3 godziny, tutorial zdalny – przeniesiony z lata na zimą w związku z pandemią).

3. Podobny tutorial został przeze mnie zaprezentowany w grudniu 2018 dla pracowników Xi'an Jiaotong-Liverpool University w Suzhou w Chinach - w sumie 6 godzin wykładów i zajęć praktycznych.
4. Seria prezentacji dla pracowników *Data Science and Scale Team* w Los Alamos National Laboratory, NM, USA.

9 Prezentacje na konferencjach międzynarodowych

Wyniki swoich badań, po uzyskaniu doktoratu, przedstawiłem na ponad 50 konferencjach międzynarodowych w Austrii, Chinach, Danii, Francji, Hiszpanii, Japonii, Kanadzie, Niemczech, Polsce, Stanach Zjednoczonych, Szwajcarii i Wielkiej Brytanii. Ogromna większość tych referatów miała status "invited lectures".

Ponadto, wygłosiłem 38 wykładów na zaproszenie na seminariach w takich ośrodkach jak (w kolejności chronologicznej: Instytut Studiów Zaawansowanych w Princeton, Tulane University, EPFL, Ohio State University, North Carolina State University, UC Davies, T.U. Gratz, KTH, Nagoya University, Inria, University in Pompadona, QMU London, IST-Austria, Cardiff University, Penn State University, Princeton University, Exeter University, University of Aberdeen, Strathclyde University, University of Liverpool, Queens University Belfast, University of Udine, Child Mind Institute NYC, IBM Watson Centre, Uniwersytet Jagielloński, T.U. Berlin, American University of Sharjah, University Roma "Tor Vergata" oraz Uniwersytet Warszawski.

10 Działalność popularyzatorska

Częścią mojej działalności jest popularyzacja nauki wśród uczniów szkół oraz ogólnie, wśród społeczeństwa. W swojej karierze naukowej po doktoracie dwukrotnie, w latach 2018 oraz 2019 byłem głównym organizatorem, ze strony Department of Mathematics, Festiwalu Nauki w Swansea. Prezentowałem też zagadnienia związane z kombinacją ruchów okresowych w National Science Museum w Londynie. Ostatnio, we wrześniu 2020, brałem udział jako mówca w Warszawskim Festiwalu Nauki. Niestety, od tego czasu wszelka aktywność popularyzatorska nie była możliwa ze względu na stan pandemii.

Byłem również zaangażowany w serie wykładów dla szkół (tak zwanych Masterclasses) dla dzieci w poziomie 9 and 10 w brytyjskim systemie edu-

kacji (odpowiednik 3.-4. klasy liceum w Polsce), gdzie przeprowadziłem pięć czterogodzinnych wykładów z demonstracjami z geometrii, topologii, kryptografii oraz analizy danych (dwukrotnie).

Prowadzę również własny, nie do końca profesjonalny, kanał youtube, <https://www.youtube.com/channel/UC125Yn2MSsxr0B05zLDQtsQ>, na którym umieszczam zarówno prezentacje naukowe jak i krótkie demonstracje popularnonaukowe (obecnie w języku angielskim).

11 Prowadzenie grup badawczych

Podczas pobytu w Swansea University w 2018 roku uzyskałem, wraz z naukowcami z Oxford University oraz University of Liverpool, grant badawczy finansowany przez Engineering and Physical Sciences Research Council (EPSRC) - sumarycznie wysokości 2.5 miliona funtów, z czego 800,000 przypadło Swansea University, gdzie byłem głównym wykonawcą. Ten grant pozwolił mi na utworzenie grupy badawczej składającej się z dwóch naukowców po doktoracie (postdoców): Dr. Tak-Shing Chana oraz Dr. Johna Harveya, oraz dwójki doktorantów - Nicka Sale'a oraz Niklasa Hellmera. Grupa ta jest częścią większego konsorcjum pomiędzy trzema partnerskimi uczelniami: <https://www.maths.ox.ac.uk/groups/topological-data-analysis>

Przez kolejne dwa lata moja grupa zlokalizowana w Swansea zajmowała się użyciem metod geometrii do analizy danych, próbami zrozumienia zachowania sieci neuronowych przy pomocy topologii, modelami przejść fazowych (w szczególności modelem Isinga), a ostatnio badaniami nad Covid19. Obecnie, mimo, że przenieśliśmy się z Swansea University do Instytutu Matematycznego PAN, nadal współpracujemy z pozostawioną tam grupą badawczą i całym konsorcjum pomiędzy trzema uniwersytetami. W ramach tej współpracy kontynuujemy istniejące projekty badawcze i poszukujemy nowych problemów do rozwiązania.

W listopadzie 2019 roku uzyskałem prestiżowy grant Dioscuri, w ramach którego buduję Centrum Dioscuri w Topologicznej Analizie Danych (<https://dioscuri-tda.org/>). Projekt Centrów Doskonałości Dioscuri został zainicjowany przez niemieckie Towarzystwo Maxa Plancka, nadzorowany przez Narodowe Centrum Nauki i realizowany wspólnie przez polskie oraz niemieckie Ministerstwa Nauki. Obecnie w Polsce istnieje pięć Centrów Doskonałości Dioscurii; cztery pozostałe są związane z szeroko rozumianymi naukami biologicznymi. Wstępne finansowanie Centrów Dioscuri jest przewidziane na okres pięciu lat.

Centrum Dioscuri w Topologicznej Analizie Danych, którego jestem kie-

rownikiem, rozpoczęło działalność w Instytucie Matematycznym PAN w lipcu 2020 roku. Obecnie, oprócz mnie, pracuje w nim dwóch doktorantów. Niedługo do grupy dołączy jeden postdoc, zaś w najbliższej przyszłości planuję zatrudnić dwóch kolejnych postdoców. Finansowanie Centrum umożliwia zatrudnienie trzech doktorantów oraz trzech podoców przez cały czas działania Centrum.

Zadaniem mojej grupy badawczej, zarówno tej w Swansea jak i w Centrum Dioscuri w Topologicznej Analizie Danych, jest przygotowywanie matematycznych i informatycznych rezultatów na potrzeby zastosowań wdrażanych przez naszych współpracowników z innych dziedzin nauki.

12 Opieka nad studentami

W trakcie mojej pracy na Uniwersytecie Jagiellońskim jak i staży postdoktorskich w University of Pennsylvania oraz Inrii nie sprawowałem formalnej opieki nad studentami ze względu na brak odpowiednich uprawnień (w Polsce) oraz zbyt krótki okres stażu (za granicą). W Swansea University nadzorowałem około dziesięciu studentów piszących prace licencjackie oraz dwóch doktorantów: Niklasa Hellmera (Magisterium, Bonn, 2019) oraz Nicka Sale'a (Magisterium Oxford, 2019). Po przeniesieniu się do Instytutu Matematycznego PAN, Niklas zdecydował się dołączyć do grupy w Centrum Dioscuri jako doktorant.

Obecnie, w Centrum Dioscuri w Topologicznej Analizie Danych, opiekuję się, jako opiekun pomocniczy, dwoma studentami studiów doktorskich pracującymi na potrzeby Centrum: Niklasem Hellmerem (który uzyskał tytuł magistra na Uniwersytecie w Bonn w 2019 roku) oraz Davide Gurnari (który uzyskał tytuł magistra w 2020 roku na Uniwersytecie w Padwie). W przypadku uzyskania przeze mnie stopnia doktora habilitowanego, zostanę głównym opiekunem zarówno dla Niklasa jak i Davide. Jestem również opiekunem pomocniczym dla Jana Senge, będącego obecnie studentem studiów doktoranckich na Uniwersytecie w Bremie (głównym promotorem jest prof. Dmitry Feichtner-Kozlov).

Literatura

- [1] Herbert Edelsbrunner, David Letscher & Afra Zomorodian *Topological Persistence and Simplification*, Discrete & Computational Geometry vol. 28, pp. 511–533, (2002).

- [2] Marian Mrozek & Bogdan Batko, *Coreduction Homology Algorithm*, Discrete & Computational Geometry vol. 41, pp. 96–118 (2009).
- [3] Piotr Brendel, Paweł Dłotko, Marian Mrozek & Natalia Żelazna *Homology Computations via Acyclic Subspace*, Computational Topology in Image Context pp 117-127.
- [4] Vidit Nanda, *Perseus, the Persistent Homology Software*, <http://www.sas.upenn.edu/~vnanda/perseus>.
- [5] Peter Bubenik, *Statistical topological data analysis using persistence landscapes*, J. Mach. Learn. Res. 16 (1), 77-102.
- [6] Ramon Moore, Interval Analysis. Englewood Cliff, New Jersey, USA: Prentice-Hall. ISBN 0-13-476853-1.
- [7] D. Ackley, *A connectionist machine for genetic hillclimbing*, Kluwer Academic Publishers, Boston MA, (1987).
- [8] Ruben Specogna and Francesco Trevisan, *Discrete Constitutive Equations in $A - \chi$ Geometric Eddy-Current Formulation*, IEEE Transactionso on Magnetics, vol. 41, no. 4, 2005.
- [9] Henry Markram et al., *Reconstruction and simulation of neocortical microcircuitry*, Cell 163 (2), 456-492, (2015).
- [10] Robin Forman, *A users guide to Discrete Morse Theory*, Seminaire Lotharingien de Combinatoire, B48c, 48, (2002).
- [11] Paweł Dłotko, *Ball mapper: a shape summary for topological data analysis*, arXiv:1901.07410.
- [12] Gurjeet Singh, Facundo Memoli and Gunnar Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, Proceedings of Eurographics Symposium on Point-Based Graphics, 2007.
- [13] Konstantin Mischaikow and Marian Mrozek, *Chaos in the Lorenz equations: A computer assisted proof*, Bulletin of the American Mathematical Society 32 (1), 66-72, 1995.

P. Dłotko