

Recenzja rozprawy doktorskiej
mgr. inż. Daniela Andrzeja Piętaka

Metoda oceny jakości wyników eksperymentów wzbudzeń kulombowskich z wykorzystaniem algorytmu genetycznego

1. Tematyka rozprawy

Tematyka rozprawy jest związana z wyznaczaniem parametrów modelu systemu na podstawie pomiarów sygnałów zebranych w trakcie jego obserwacji. Pomimo długiej historii, zagadnienie to sprawia nadal sporo kłopotów w wielu praktycznych przypadkach, w których znajomość modelu jest potrzebna na przykład do optymalizacji, sterowania, czy podejmowania decyzji. Estymacja parametrów modeli jest także używana do wyznaczenia wartości parametrów o znaczeniu fizycznym i może być traktowana jako specyficzna metoda pomiarowa, gdy bezpośredni ich pomiar nie jest możliwy, tak jak w problemie rozważanym w rozprawie. Ze względu na liczne zastosowania tematyka ta jest nadal żywa i obecna w prowadzonych badaniach i publikacjach.

Tematem rozprawy jest estymacja parametrów modeli wzbudzeń kulombowskich oraz oszacowania niepewności uzyskanych ocen parametrów. Pomiary wzbudzeń kulombowskich są eksperymentalną techniką badania elektromagnetycznej struktury jąder atomowych. Nie jestem fizykiem i nie jestem w stanie wypowiadać się w sprawach modeli rozpraszania jąder atomowych wiązki bombardującej tarczę w cyklotronie. W rozprawie są podane tylko podstawowe zależności, z których trudno się zorientować, jak wygląda model i jak mają się do niego mierzone zmienne. Doktorant do obliczania wartości modelu skorzystał z programu GOSIA (program kanałów sprzężonych z minimalizacją najmniejszych kwadratów), którego twórcami byli w latach 1980-tych T. Czosnyka, D. Cline i C. Y. Wu. W rozprawie model jest traktowany właściwie jako czarna skrzynka. W programie GOSIA, jak wynika z pełnej nazwy, przeprowadza się estymację parametrów modelu metodą nieliniowych najmniejszych kwadratów (regresji nieliniowej) i szacuje ich niepewności. Jest to narzędzie powszechnie (i jak odnoszę z lektury rozprawy może nawet jedyne?) stosowane do analizy danych z eksperymentów kulombowskich. W programie GOSIA do optymalizacji używa się metody gradientowej, dosyć typowej dla lat 1980-tych. W rozprawie nie wyjaśniono, jak w programie GOSIA szacuje się niepewności ocen parametrów. Natomiast do określenia obszaru niepewności w rozprawie skorzystano z możliwości wyznaczenia tego obszaru przez odcięcie wartości statystyki chi kwadrat przez odpowiednią hiperpłaszczyznę krytyczną, ortogonalną do wektora osi wartości statystyki (i jego rzut na przestrzeń parametrów).

2. Cel i zakres rozprawy

W rozprawie zastosowano nowszą metodę optymalizacji niż w programie GOSIA. Lepiej radząc sobie z wielomodalnością optymalizowanej funkcji. Lokalne optima to jedna z częściej spotykanych trudności w stosowaniu nieliniowej metody najmniejszych kwadratów. Sygnalizowana już w latach 50-tych, a szerzej w latach 70-tych i 80-tych ubiegłego wieku. Wprowadzenie metaheurystyk spowodowało duży postęp w efektywnej optymalizacji takich funkcji. Doktorant wybrał jako metodę optymalizacji chyba najwcześniej rozpowszechnioną metaheurystykę – algorytmy genetyczne.

Do wyznaczenia przedziałów ufności ocen parametrów uzyskanych w wyniku minimalizacji wskaźnika jakości doktorant zaproponował wykorzystanie repozytorium zbioru punktów uzyskanych w trakcie działania algorytmu genetycznego stosując wspomniany powyżej pomysł oparty na teście chi kwadrat. Jest to metoda nieparametryczna, która bazuje na wyszukaniu punktów najbliższych hiperpowierzchni przecięcia statystyki chi kwadrat przez odpowiednią hiperpowierzchnię.

W literaturze przedstawiano przykłady numerycznego wyznaczania obszarów ufności dla prostych modeli nieliniowych, w których powierzchnie, a raczej na ogół krzywe na płaszczyznach, mają kształty daleko odbiegające od elipsoidy. Długie czasy obliczeń istotnie utrudniają uzyskanie takich wyników dla bardziej skomplikowanych modeli. Opracowanie metody pozwalającej wyznaczyć efektywnie taką hiperpowierzchnię dla skomplikowanych modeli z wieloma parametrami należałoby uznać za istotny wkład w praktyczne rozwiązanie tego zagadnienia. Zaprezentowany w rozprawie pomysł umożliwia przeprowadzenie szybkich obliczeń w celu efektywnego wyznaczania hiperpowierzchni ograniczających obszar ufności. Doktorant był bliski wyznaczenia takich obszarów dla prostszego przypadku testowego (patrz rys.5.6 – 5.17), ale w punkcie 5.3 opisującym estymację niepewności parametrów i w przykładzie rzeczywistym ograniczył się do wyznaczania tylko przedziałów ufności ocen parametrów, a więc tylko do kilkunastu punktów hiperpowierzchni. Jest to z pewnością samo w sobie osiągnięcie, które pozwala na szybkie obliczenie przedziałów ufności i do tego nie wymagające obliczania dodatkowych wartości statystyki chi kwadrat. Ten sposób postępowania jest też bardzo łatwo przenieść na przypadek ogólny wyznaczania dowolnych punktów na hiperpłaszczyźnie ograniczającej obszar ufności.

Doktorant przygotował oprogramowanie do wyznaczania przedziałów ufności i policzył jeden przykład symulacyjny ze znanym rozwiązaniem oraz jeden przykład rzeczywisty, w którym są użyte pomiary z eksperymentu wykonanego w Środowiskowym Laboratorium Ciężkich Jonów Uniwersytetu Warszawskiego.

3. Oryginalność rozprawy

Oryginalnością rozprawy jest użycie nowej metody do szacowania niepewności estymowanych ocen parametrów modelu w modelach wzbudzeń kulombowskich. Postęp w metodologii, który nastąpił w ciągu ostatnich czterdziestu lat, całkowicie motywuje zastosowanie nowych metod estymacji parametrów tego modelu. Szacowanie niepewności ocen parametrów jest kontynuacją pomysłu użycia tej nowej metody – algorytmu genetycznego. Nie jestem w stanie precyzyjnie ocenić rzeczywistego znaczenia tego aspektu rozprawy dla fizyków, niemniej jednak zadanie to wydaje mi się być istotne dla związanych z tym modelem badań eksperymentalnych.

Pomysł estymacji parametrów został właściwie po raz pierwszy sformułowany w szeroko rozumianych badaniach fizycznych, a mianowicie w pomiarach astronomicznych, w których

duże niepewności pomiarów skłoniły Galileusza w traktacie z 1629 roku do sformułowania doktryny o wyborze takiej wartości poszukiwanej wielkości, która minimalizuje błędy pomiarów. Estymacja parametrów jest nadal ważnym narzędziem fizyki empirycznej. Jednak od mniej więcej wieku teoria budowy modeli matematycznych z użyciem danych empirycznych była chyba bardziej rozwijane w statystyce, ekonometrii, a następnie intensywnie w automatyce pod nazwą identyfikacja systemów. Od pewnego czasu podobne badania wchodzi też w zakres informatyki technicznej jako kierunek w obszarze uczenia maszynowego. Jak widać, badania tego typu, a w szczególności związane z praktycznymi zastosowaniami, mają charakter interdyscyplinarny, a między innymi są związane z dyscypliną „informatyka techniczna i telekomunikacja”, w której otwarty jest przewód. Do wyboru tej dyscypliny skłania też użyte narzędzie optymalizacji – algorytmy genetyczne, które są niewątpliwie głównie kojarzone z informatyką techniczną.

Po wielu latach rozwoju badań związanych z budową modeli matematycznych z użyciem danych empirycznych ustaliły się pewne zasady postępowania. Obejmują one takie zagadnienia jak analiza modelu, projektowanie eksperymentu i badanie jakości danych empirycznych (pod kątem dostatecznego pobudzenia modelu czy niezależności od czynników zewnętrznych), a po estymacji walidację uzyskanego modelu. Tezy rozprawy nie dotyczą całości powyższych zagadnień, a tylko jego fragmentu związanego z walidacją modelu, a dokładniej z szacowaniem niepewności ocen estymowanych parametrów. Niemniej jednak etapy budowy modelu są ze sobą mocno zazębiane, w takim sensie, że wyniki uzyskane w trakcie jednego etapu wpływają na postępowanie w innych etapach. Dlatego w rozprawie przydałoby się przynajmniej krótkie omówienie, co wiadomo o takich analizach i wyciągniętych z nich wnioskach dla rozpatrywanego modelu. Więcej uwag na ten temat umieściłem w następnym punkcie recenzji. Mimo że nie są one bezpośrednio związane z celem rozprawy, to być może takie uwagi od osoby patrzącej na zagadnienie od strony teorii systemów przyczynią się w dalszych badaniach do lepszego zrozumienia procesu budowy modelu i do uzyskania lepszej jakości uzyskiwanych ocen parametrów.

Ważną częścią zaproponowanego postępowania jest narzędzie użyte do optymalizacji, w tym przypadku jest to algorytm genetyczny. Algorytmy genetyczne są dobrze znane i tak szczegółowy ich opis, jak przedstawiony w pracy, nie jest potrzebny. Doktorant stosował to narzędzie już w swojej pracy magisterskiej z 2008 r. i pozostał przy nim w omawianej rozprawie. Tymczasem w tej klasie metod podobnie eksplorujących przestrzeń parametrów znacznie większe możliwości dają algorytmy ewolucyjne, w których opracowano wiele dodatkowych specjalizowanych operatorów w stosunku do algorytmów genetycznych. W szczególności rozwinięto w nich algorytmy wielomodalne, jak algorytmy niszowe czy wyspowe, właśnie przeznaczone do zadań takich, jak rozpatrywane w rozprawie. W rozprawie brakuje dyskusji i analizy wyboru narzędzia najbardziej pasującego do rozpatrywanego zagadnienia. Badania idące w tym kierunku byłyby przydatne w dalszym rozwoju omawianej metody.

Jednak, jak wynika z postawionych w rozprawie tez, za podstawowe osiągnięcie Doktorant uważa metodę szacowania niepewności ocen parametrów. Składa się ona z ośmiu etapów nazwanych literami od A do H i stąd zostało nazwane „metodą AH”. Są to następujące etapy. Etap A: utworzenie repozytorium punktów uzyskanych w trakcie optymalizacji; etap B: usunięcie z repozytorium punktów o zbyt dużych wartościach wskaźnika jakości; etap C: rozrzedzenie obszarów o zbyt gęstych punktach; etap D: klasteryzacja punktów w celu wyodrębnienia obszarów atrakcji (zwanych w rozprawie basenami przyciągania) poszczególnych optimów. Dalsze etapy postępowania są prowadzone osobno dla każdego wyodrębnionego obszaru atrakcji. Etap E: statystyka opisowa punktów w obszarze atrakcji; etap F: tworzenie histogramów; etap G: wyodrębnienie punktów najbliższych płaszczyźnie

progowej, odpowiadającej przyjętemu poziomowi ufności w teście chi kwadrat: etap H: wyznaczenie obszaru ufności oraz oszacowanie jego dokładności. Liczba tych etapów jest nazbyt rozbudowana i mogłaby być z powodzeniem zmniejszona przez połączenie niektórych z nich.

Według mojego rozeznania, pomysł zaprezentowanej metody jest nowy w sensie sposobu rozwiązania zadania. Wyznacza się w niej granice obszaru w sposób nieparametryczny poszukując punktów leżących poniżej, ale jak najbliżej płaszczyzny odcięcia. O znaczeniu tego podejścia wspomniałem w poprzednim punkcie recenzji. Jednak Doktorant w zasadzie nie przedstawił przeglądu i porównania proponowanej metody z innymi metodami, jak wyznaczanie punktów przecięcia hiperpłaszczyzny krytycznej z hiperpowierzchnią statystyki chi kwadrat numerycznymi metodami rozwiązywania odpowiednich równań algebraicznych, metody szacowania niepewności ocen parametrów oparte na wyznaczaniu macierzy kowariancji przez aproksymacje optymalizowanej funkcji w pobliżu punktu optymalnego, czy też oparte na gradientach, i nie przedyskutował ich wad i zalet, chociaż wspomina o istnieniu innych metod. Biorąc pod uwagę charakter pracy, szczególnie istotne byłoby porównanie nowej metody z metodą zastosowaną w programie GOSIA pod kątem trudności obliczeniowych oraz uzyskiwanych dokładności. W rezultacie, Czytelnik nie dowiaduje się w sposób bezpośredni z rozprawy, czy opracowana metoda ma zalety w stosunku do innych, typowo stosowanych metod, czy nie.

Warto może też wspomnieć o braku analizy właściwości statystycznych proponowanych estymatorów, czy analizy teoretycznej proponowanej metody. Zdaję sobie sprawę, że rozważania tego typu wymagają zaawansowanej znajomości teorii prawdopodobieństwa i statystyki matematycznej, co może być dużym wyzwaniem dla rozprawy doktorskiej typu technicznego. Ale Doktorant nie wspomina o kierunku prac, reprezentowanym na przykład przez znany artykuł prof. J. Arabasa¹, dotyczących estymacji rozkładu punktów w pobliżu optimum w algorytmach ewolucyjnych, co, jak mi się wydaje, powinno być przynajmniej zauważone w przeglądzie literatury rozprawy z tej tematyki, jeżeli już nie przyjęte jako podstawa głębszej teoretycznej analizy zadania, która mogłaby na przykład pomóc w lepszym wyborze współczynników w algorytmach ewolucyjnych.

Przechodząc do szerszej analizy bibliografii zamieszczonej w rozprawie, to prawie połowa wykazu literatury jest związana z rozdziałem 2 i obejmuje pozycje dotyczące wzbudzeń kulombowskich, a sporo pozycji literaturowych związanych z pozostałymi rozdziałami jest także napisana przez fizyków. Na tym tle literatura dotycząca pozostałych zagadnień jest mniej liczna. Jeżeli chodzi o pozycje związane z procedurą modelowania i walidacji modelu, to są to głównie książki Brandta oraz Boxa i Drapera. Obie te pozycje są dobrze znane i są nadal aktualne, szczególnie po zmianach w późniejszych wydaniach książki Brandta. Nowsza pozycja Tarantoli jest bardziej nastawiona na metody rozwiązywania. Oczywiście literatura w tej tematyce jest obfita, ale ten wybór nie jest zły, pomimo pominięcia pozycji ściśle poświęconych regresji nieliniowej, jak na przykład książki Gallanta² czy Sebera i Wilda³. Ale Doktorant chyba za mało uwagi poświęcił analizie metod walidacji modeli, a to zagadnienie jest dosyć istotne i bliskie tematowi rozprawy. Także literatura dotycząca ewolucyjnych metod optymalizacji zawiera podstawowe pozycje, ale Doktorant przeoczył tu algorytmy ewolucyjne i specjalizowane metody wielomodalne. A użycie metod wielomodalnych być może pozwoliłoby na zrezygnowanie z uciążliwej klasteryzacji w etapie D proponowanej

¹ J. Arabas (2012) Approximating the genetic diversity of populations in the quasi-equilibrium state. *IEEE Transactions on Evolutionary Computation*, 16(5):632-644, DOI: 10.1109/TEVC.2011.2166157

² A.R. Gallant (1987) *Nonlinear statistical models*. Wiley, New York.

³ G.A.F. Seber, C.J. Wild (1989) *Nonlinear regression*. Wiley, New York.

w rozprawie metody. Warto może też ponownie podkreślić sygnalizowany wcześniej brak usystematyzowanych przeglądów niektórych zagadnień istotnych dla tematyki rozprawy.

Ustawa wymaga, aby kandydat na doktora wykazał się ogólną wiedzą teoretyczną w odpowiedniej dyscyplinie naukowej. Sprawdzenie ogólnej wiedzy w dyscyplinie odbywa się w trakcie egzaminu z dyscypliny oraz w czasie publicznej obrony. Ponieważ przypuszczam, że zakres materiału na egzaminie może raczej nie obejmować tematyki związanej z teorią estymacji i algorytmami ewolucyjnymi, to, aby wyjaśnić sygnalizowane w recenzji nie w pełni wyjaśnione kwestie i sformułowania, chciałbym, aby w czasie obrony Doktorant odpowiedział na następujące pytania:

- Co to są obszary ufności estymatorów parametrów modelu oraz jakie metody stosuje się do ich wyznaczania?
- Na jakich pomysłach są oparte wielomodalne algorytmy ewolucyjne?

4. Szczegółowe uwagi krytyczne i dyskusyjne

Praca zawiera sporą liczbę wątpliwych lub nieprecyzyjnych sformułowań. Poniżej przedstawiam ich przykłady oraz inne uwagi dyskusyjne o charakterze bardziej technicznym. Zapis a^b oznacza stronę **a**, **b**-ty wiersz od góry. Analogicznie, a_b oznacza stronę **a**, **b**-ty wiersz od dołu.

13, rozdział 2 W poprzednim punkcie recenzji wspomniałem o etapach postępowania przy budowie modelu z użyciem danych empirycznych. W rozprawie nie opisano modelu pod kątem takiego postępowania i nie podano, z jakim typem modelu mamy do czynienia: z modelem statycznym, równaniami różniczkowymi zwyczajnymi czy cząstkowymi, równaniami stochastycznymi? Odnoszę wrażenie, że model odpowiada dosyć skomplikowanej teorii fizycznej, więc może jest trudno opisać charakter tego modelu w prosty sposób dostępny dla laika w tej dziedzinie. Jednak informacja o typie modelu pozwala na pewną jakościową ocenę charakteru zadania estymacji. W rozprawie wspomina się o trudnościach występujących przy estymacji parametrów tego modelu: ‘uciążliwych parametrach’ i lokalnych minimach. Wydaje mi się, że warto wobec tego może wspomnieć o tym, jak wygląda to od strony teorii tworzenia modeli z udziałem danych eksperymentalnych. Zdaję sobie sprawę, że badania związane z modelowaniem eksperymentów wzbudzeń kulombowskich mają długą historię i związana z tym wiedza jest już na pewno duża. Mam jednak nieśmiałą nadzieję, że przedstawione niżej uwagi mogą nieco pomóc w zrozumieniu możliwych przyczyn występujących w nich trudności.

Zacznę od ‘uciążliwych parametrów’. Są to prawdopodobnie te parametry, które w teorii estymacji nazywa się statystycznie nieistotnymi. Chociaż w rozprawie nie przeprowadzono analizy istotności parametrów, to z oszacowań niepewności ocen parametrów w załączonym przykładzie estymacji z rzeczywistymi pomiarami widać, że niektóre parametry mają bardzo duże przedziały ufności, tak że ich oceny statystyczne są bardzo niepewne. W tradycyjnym podejściu regresji liniowej przyjmuje się, że nieistotne parametry są zerowe, co właściwie oznacza, że związana z nimi zmienne pomiarowe zbyt słabo wpływają na sygnał wyjściowy modelu i mogą być pominięte. Jednak w modelu fizycznym przyjęcie parametru za zerowy może nie być fizycznie uzasadnione, o czym pisze też Doktorant na stronie 106. Jednak z drugiej strony nieistotność parametrów jest wskazówką, że coś w procesie budowy modelu może nie być prawidłowe. Przyczyną nieistotności parametrów jest często to, że model jest nieidentyfikowalny parametrycznie przy zbieranych pomiarach. Nieidentyfikowalność parametryczna modelu oznacza, że nie da się wyznaczyć jednoznacznie (na ogół niektórych) jego parametrów nawet przy bezbłędnych sygnałach pomiarowych. Może to być związane z tym, że niektóre parametry są związane jakąś zależnością funkcyjną, która pojawia się

w modelu jako czynnik wspólny w oddziaływaniu na sygnał wyjściowy, na przykład wyjście modelu zależy od iloczynu dwóch parametrów. Zmiana jednego z nich może wtedy być skompensowana odpowiednią zmianą drugiego, tak aby iloczyn pozostał stały. W rezultacie w przestrzeni parametrów pojawi się podprzestrzeń, w tym przypadku płaszczyzna hiperboliczna, każdy punkt której jednakowo wpływa na wyjście modelu. Takich parametrów nie da się jednoznacznie wyznaczyć z pomiarów sygnałów wyjściowych. Tego typu fatalny przypadek nieidentyfikowalności będzie się objawiał nieistotnością estymowanych parametrów. Można się jej pozbyć jedynie przez przeformułowanie modelu, na przykład przez dołączenie nowych sygnałów pomiarowych, albo przez włączenie dodatkowej wiedzy o parametrach, na przykład ustalając w modelu wartości przynajmniej niektórych nieidentyfikowalnych parametrów korzystając z dostępnej wiedzy o ich wartościach. Usunięcie nieistotnych parametrów ze zbioru parametrów estymowanych z reguły poprawia właściwości optymalizowanej funkcji i zmniejsza, często nawet znacznie, szacowaną niepewność pozostałych parametrów, więc wysiłek włożony w rozwiązanie tego problemu może się opłacać. Oczywiście analizę modelu i danych eksperymentalnych oraz ewentualne założenia o wartościach nieistotnych parametrów musi przeprowadzić fizyk, ale wyniki i wnioski z walidacji modelu może i powinna sformułować osoba, która prowadzi estymację parametrów modelu.

Zagadnienie lokalnych, czy nawet kilku globalnych minimów wskaźnika jakości też może mieć przyczynę w samym modelu, chociaż tutaj sytuacja jest często bardziej skomplikowana. Na przykład, liniowe równanie różniczkowe zwyczajne drugiego rzędu o stałych współczynnikach może mieć rozwiązanie $y(t) = a_1 e^{b_1 t} + a_2 e^{b_2 t}$. Ze względu na symetrię, czy inaczej mówiąc dowolność ustawienia składników powyższej sumy, w idealnych warunkach i bez błędów pomiarowych wskaźnik jakości estymacji będzie miał w tym przypadku dwa optima globalne w przestrzeni parametrów, tzn. (a_1, a_2, b_1, b_2) i (a_2, a_1, b_2, b_1) . Także wtedy model nie jest globalnie identyfikowalny parametrycznie, ale w tym przypadku jest on jednak identyfikowalny lokalnie, to znaczy istnieją obszary w przestrzeni parametrów, w którym jest tylko jedno optimum globalne. W powyższym przypadku może to być na przykład obszar ograniczony nierównością $b_1 < b_2$. Należy się wtedy liczyć z tym, że podobne zbliżone optima globalne pojawiają się też w przypadku błędów w pomiarach. Taka lokalna identyfikowalność na ogół pozwala uzyskać satysfakcjonujące rozwiązanie, chociaż przy wzroście skomplikowania modelu może to nie być takie łatwe. Wielomodalność wskaźnika jakości powoduje też pojawianie się punktów stacjonarnych funkcji, a wśród nich minimów lokalnych. Przykłady mnożących się punktów stacjonarnych ze wzrostem liczby parametrów były przytaczane już na początku lat 1980-tych.

Być może model rozpatrywany w rozprawie jest zbyt skomplikowany, aby można było analitycznie przebadать jego właściwości, ale sporo można sprawdzić numerycznie. Jak wynika z powyższych uwag, właściwości modelu mogą mocno wpływać na kształt optymalizowanego wskaźnika jakości i szersza informacja, co wiadomo na ten temat, byłaby przydatna dla lepszego zrozumienia trudności występujących w estymacji parametrów.

Warto też jeszcze wspomnieć o możliwości przeprowadzenia dodatkowych badań modelu. Przykładowo, w zadaniu z rzeczywistymi pomiarami Doktorant miał dane z kilkunastu powtarzanych eksperymentów. Pozwala to na testowanie powtarzalności eksperymentu oraz, po jej potwierdzeniu, na wykonanie oszacowań dokładności opracowanych modeli, na przykład za pomocą walidacji krzyżowej (one-leave-out). Interesujące byłoby też przetestowanie normalności reszt, gdyż założono rozkład normalny zakłóceń i można by było sprawdzić, czy takie założenie jest uzasadnione. Jest to o tyle ważne, że przy tym

założeniu estymatory najmniejszych kwadratów mają atrakcyjne właściwości asymptotyczne estymatorów największej wiarygodności.

- 20⁴⁻⁶ „Z punktu widzenia informatyki jest to zestaw parametrów (liczb rzeczywistych), których poszukiwanie (sub)optimalnych wartości stanowi zadanie optymalizacyjne” – Teoria optymalizacji i zadania optymalizacyjne z pewnością nie powstały w ramach badań informatycznych; jeżeli już jest potrzebne odniesienie, to chyba trzeba by się było powołać raczej na matematykę.
- 2.2.2b Doktorant miesza tu (i wielokrotnie dalej) pojęcia testu chi kwadrat i kwadratowego wskaźnika jakości identyfikacji. Minimalizowane wyrażenie (2.5) jest wskaźnikiem jakości (funkcją celu), którą minimalizuje się w celu uzyskania ocen parametrów. W takim zadaniu niczego się nie testuje, więc trudno tę funkcję nazywać testem. Tak się jednak składa, że przy pewnych założeniach, dla prawdziwych wartości parametrów statystyka (2.5) ma rozkład chi kwadrat, co pozwala na testowanie pewnych hipotez statystycznych. Jednak są to dwa zupełnie różne zadania.
- 21⁸⁻¹⁰ „W tradycyjnym teście χ^2 , określoną we wzorze (2.5) statystykę najmniejszych kwadratów normalizuje się względem liczby stopni swobody (czyli do liczby danych pomniejszoną o liczbę parametrów)” – W tradycyjnym teście chi kwadrat statystyka ma postać sumy kwadratów, która przy odpowiednich założeniach ma rozkład chi kwadrat z odpowiednimi stopniami swobody. Po podzieleniu przez liczbę stopni swobody dostajemy rozkład chi kwadrat z jednym stopniem swobody, ale nie jest to ‘tradycyjne’ postępowanie.
- 29⁶⁻⁵ Nie rozumiem, dlaczego ważne jest, aby korzystać tylko z punktów zebranych w czasie obliczeń za pomocą algorytmu genetycznego. Stosując algorytm genetyczny, musimy się liczyć z dosyć masowymi obliczeniami wartości funkcji. Czy w takim razie doliczenie jeszcze pewnej liczby wartości jest problemem, szczególnie gdyby się okazało, że w ten sposób można by było znacznie poprawić dokładność szacowania niepewności? Jeżeli z kolei czas obliczeń modelu w programie GOSIA jest na tyle znaczny, że każde jego obliczenie waży, to czy nie lepiej by było użyć innych metod optymalizacji, np. MADS⁴?
- 33, tabl. 3.1, kolumna 2, 4 wiersz „W granicy liczby wywołań automatycznie odnajduje wszystkie minima, w tym globalne” – Czy rzeczywiście tak jest? Wyobraźmy sobie funkcję dwóch zmiennych równą 0 w pewnym punkcie i równą 1 poza tym punktem. Algorytm genetyczny z prawdopodobieństwem 1 nie znajdzie tego minimum, nawet jeżeli liczba wywołań algorytmu będzie dążyć do nieskończoności.
- 36, tabl. 3.2 Podział całego zakresu na równe osiem przedziałów i zakodowanie ich środków powoduje zmniejszenie przedziałów o 25%, a tym samym zwiększenie dokładności.
- 42⁸⁻⁹ „Poszukiwania heurystyczne podejmuje się, gdy znalezienie rozwiązania dokładnego nie jest możliwe – z powodu zbyt długiego czasu obliczeń, zmian zachodzących w środowisku bądź niejednoznaczności funkcji oceny” – Mogę się domyślać, co Doktorant ma na myśli pisząc o zmianach zachodzących w środowisku czy o niejednoznaczności funkcji oceny. I wydaje mi się, że w obu przypadkach można zastosować inne metody optymalizacyjne po odpowiednim sformułowaniu zadania.
- 42₃₋₁ „Do tworzenia poszczególnych osobników w pokoleniu używana jest informacja ze wszystkich rozwiązań z poprzedniego pokolenia. Pozwala to skutecznie unikać przedwczesnej zbieżności algorytmu – problemu lokalnych ekstremów.” – Głównym operatorem związanym z unikaniem zbieżności do lokalnych ekstremów jest operator mutacji, który akurat właśnie nie korzysta z informacji znajdujących się w rozwiązaniach z poprzedniego pokolenia.

⁴ C. Audet, J.E. Dennis Jr. (2006) Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, 17(1):188–217, DOI: 10.1137/040603371.

- 492-1 „W ruletkowym wyborze na rodzica, im słabszy osobnik tym mniejszy kąt wycinka kołowego, czyli mniejsza szansa zostania rodzicem. Tak zdefiniowane „koło ruletki” jest tożsame z dystrybucją prawdopodobieństwa zostania rodzicem” – To nie jest odpowiednik dystrybucji, ale dyskretnego rozkładu prawdopodobieństwa (funkcji masy prawdopodobieństwa).
- 6613 „Z punktu widzenia Metody AH, algorytm genetyczny jest formą sztucznej inteligencji” - ???
- 67, **ostatni akapit** W symulowanym przykładzie Doktorant porównuje wyniki zastosowania swojej metody szacowania niepewności ocen parametrów dla punktów „pomiarowych” generowanych w podanej deterministycznej siatce w przestrzeni parametrów, generowanych przez algorytm genetyczny i generowanych metodą Monte Carlo – niezależnie dla każdego parametru w ostatnim przypadku. Jednak to porównanie nie jest zbyt sprawiedliwe. We wszystkich sposobach generacji przyjęto mniej więcej tyle samo punktów, ale tylko dla algorytmów genetycznych gromadzą się one głównie w pobliżu optimum i ich gęstość w tych obszarach jest z konieczności większa. W pozostałych są one rozmieszczone równomiernie w całej przestrzeni.
- 828 Wymieniono wszystkich autorów z ich imionami, ale nie podano pozycji literatury.
- 92⁶⁻⁸ „Wynika to z faktu, że wartości parametrów cechy elementów populacji są z prawdopodobieństwem 1 zbieżne do ich charakterystyk teoretycznych (Prawa Wielkich Liczb)” – Prawa wielkich liczb mówią o zbieżności rozkładów do rozkładu normalnego, przy pewnych założeniach. Jak się to ma do zbieżności do charakterystyk teoretycznych? I co to znaczy, że wartości parametrów populacji są zbieżne z prawdopodobieństwem 1 do ich charakterystyk teoretycznych? Wydawałoby się, że jedno i drugie pojęcie oznacza to samo. Może raczej chodzi o zbieżność estymatorów parametrów ze wzrostem liczności próby, dążącej do nieskończoności? A to jest właśnie właściwość estymatorów największej wiarygodności.
- 92⁹⁻¹⁰ Rozkłady dla metod ewolucyjnych są właśnie tematem wspomnianego artykułu Arabasa.
- 925 „wartość przeciętna dla i-tego wymiaru informuje, gdzie koncentruje się próbkowanie” – Jeżeli zmienna przyjmuje wartości 0 i 1 z jednakowym prawdopodobieństwem, to jej wartość przeciętna jest równa $\frac{1}{2}$. Czy tu wartości próbkowane koncentrują się przy wartości przeciętnej?
- 93¹⁻² „znormalizowane względem zakresu próbkowania odchylenie standardowe dla i-tego wymiaru umożliwia porównanie jego bezwzględnej wartości między wymiarami” – Odchylenie standardowe jest zawsze dodatnie, więc po co liczyć jego wartość bezwzględną?
- 943 i dalej Do sprawdzenia równości rozkładów lepiej byłoby zastosować odpowiednie testy statystyczne.
- 103, **wzór (5.2)** Brakuje założenia o nieskorelowaniu zmiennych epsilon.
- 104⁶ „Ponieważ y_i są zmiennymi niezależnymi” – Skąd wiadomo, że są one niezależne? Nie założono przecież, że błędy pomiarowe ϵ_j są niezależne. Do tego zdanie to jest niezręcznie sformułowane.
- 1051 Co to są „zależności liniowe”?
- 106⁶⁻¹³ Uzasadnienie niemożności wyznaczenia macierzy kowariancji jest mocno dyskusyjne. Po pierwsze, można przyjąć, że da się zlinearyzować model w pobliżu punktu optymalnego i użyć wzorów dla modelu liniowego, tak jak jest opisane w cytowanej w rozprawie książce Tarantoli. Można też wyznaczyć drugie pochodne znając w tym punkcie gradienty, co jest też dobrze znanym rozwiązaniem.
- Natomiast jeżeli chodzi o argument o ‘uciążliwych parametrach’, które powodują trudności w obliczaniu macierzy kowariancji w modelu zlinearyzowanym spowodowane koniecznością odwracania źle warunkowanej macierzy, to w regresji liniowej są znane dobrze sprawdzone metody radzenia sobie z tymi trudnościami, na przykład w tzw. regresji

grzbietowej (ridge regression), gdzie stosuje się regularyzację. Zresztą obliczeniowo podobne rozwiązanie jest zastosowane w znanej metodzie Levenberga-Marquardta powszechnie stosowanej do optymalizacji w regresji nieliniowej, gdzie są też wyznaczone macierze kowariancji.

Tak więc wyznaczanie macierzy kowariancji tradycyjnie stosowanymi metodami wygląda na realne do wykonania i przedstawione tłumaczenie odrzucające takie podejścia nie jest przekonujące, nawet jeżeli obliczanie macierzy kowariancji może być kłopotliwe. Myślę jednak, że Doktorant miał na myśli to, że przybliżenia liniowe nie są w przypadku rozpatrywanego modelu dostatecznie dokładne i tylko nieprecyzyjnie sformułował ten akapit. Tak czy inaczej, z pewnością przydałaby się szersza dyskusja tych postępowań w rozprawie, bo są to metody konkurencyjne do proponowanej w rozprawie, więc powinno się w niej znaleźć przynajmniej porównanie zalet i wad różnych metod, a może nawet porównanie wyników zastosowania różnych metod w rozważanych przykładach.

106¹⁴⁻¹⁶ Należy tutaj koniecznie dodać, że błędy pomiarowe są nieskorelowane.

106¹⁷⁻²⁰ Przy przyjętych założeniach estymatory parametrów są asymptotycznie nieobciążone i asymptotycznie efektywne oraz mają asymptotycznie rozkład normalny z macierzą kowariancji równą dolnemu ograniczeniu Craméra-Rao, co wynika z tego, że są one estymatorami największej wiarogodności. Potocznie czasami się mówi, że mają one asymptotycznie najmniejsze wariancje. Natomiast o ile estymatory liniowej metody najmniejszych kwadratów są BLUE (czyli nieobciążone i z najmniejszą wariancją wśród liniowych nieobciążonych estymatorów) dla dowolnej długości próby, to estymatory nieliniowej metody najmniejszych kwadratów takiej właściwości nie muszą mieć. Tak więc druga właściwość podana w rozprawie nie musi być prawdziwa, chyba że Doktorant miał na myśli asymptotyczną właściwość i tylko zapomniał to napisać lub zna pracę, gdzie taka właściwość jest udowodniona – wtedy powinien podać odpowiednią pozycję literatury. Dodatkowo, stwierdzenie w przypisie 24 na dole strony jest mylące, gdyż twierdzenie Gaussa-Markova mówi właśnie o tym, że estymatory liniowej metody najmniejszych kwadratów są BLUE, a to się nie odnosi do nieliniowej metody najmniejszych kwadratów.

107¹⁴ Nie słyszałem, aby istniało w statystyce pojęcie standardowego poziomu ufności. Można najwyżej mówić o często używanych poziomach.

108¹²⁻¹¹ „Wartość $q_{reduced}$ w funkcji liczby stopni swobody jest szybko zbieżna do jedności dla każdego poziomu ufności” – Rys. 5.1 tego nie potwierdza. 1000 stopni swobody to wcale nie tak mało.

109¹³ Niedokończone zdanie.

110, rys. 5.3 Wcześniej sugerowano, że liczba stopni swobody równa 1000 to nie jest dużo, a tu jest wykres tylko do dwóch stopni swobody.

121, punkt 5.3 Opis wyznaczania niepewności parametrów jest niezbyt zrozumiały. Przydałyby się rysunki objaśniające przyjęte postępowanie.

121¹² „Współczynnik $spread_{FL}$ stanowi dolne oszacowanie dokładności” – współczynnik $spread_{FL}$ mówi o rozrzucie "pionowych" współrzędnych punktów FL wokół y_{PU} . Mniejsza wartość jest oczywiście lepsza, ale nie rozumiem dlaczego jest oszacowaniem dokładności szacowania niepewności, a tym bardziej dolnym oszacowaniem.

122, punkt 5.3.1 Oszacowanie obszaru ufności dokonano od wewnątrz obszaru ufności, dla punktów znajdujących się poniżej płaszczyzny krytycznej. Nie rozumiem, dlaczego Doktorant nie dokonał podobnego oszacowania obszaru ufności od zewnątrz, dla punktów powyżej płaszczyzny krytycznej. Dałoby to oszacowania z obu stron i pozwoliło na podanie niepewności oszacowania. A do tego, stosując mniej lub bardziej zaawansowane metody geometryczne można by było na pewno uzyskać na tej podstawie znacznie dokładniejsze oszacowania punktów hiperpłaszczyzny ograniczającej obszar ufności. Takie wyniki byłyby ciekawsze i chyba bardziej użyteczne.

122¹³ „Jako liczbę sąsiadów przyjmuje się podwojoną długość wektora parametrów” – czy chodzi o wektor parametrów p_{BEST}? Jaki związek mają te dwie wielkości?

5. Podsumowanie

Uwagi dotyczące etapu formułowania modelu i jego badania mają charakter dyskusji naukowej i nie są oczywiście zasadnicze dla oceny samego osiągnięcia przedstawionego w tezach rozprawy. Może trochę podobna sytuacja jest z uwagami związanymi z etapem minimalizacji wskaźnika jakości, chociaż w tym przypadku przeoczenie możliwości zastosowania innych narzędzi optymalizacyjnych, które wydają się być bardziej odpowiednie w omawianym w rozprawie przypadku, ma większą wagę biorąc pod uwagę, że rozwój metod ewolucyjnych optymalizacji jest powszechnie wiązany z dyscypliną informatyki technicznej, w której jest rozpatrywany doktorat. Zastosowanie specjalizowanych metod ewolucyjnych dla funkcji wielomodalnych mogłoby może także uprościć proponowaną w rozprawie metodę.

Natomiast zasadniczą częścią rozprawy podlegającą ocenie jest opracowanie przez Doktoranta nowej nieparametrycznej metody estymacji ufności parametrów modelu w regresji nieliniowej, korzystającej z punktów uzyskanych w trakcie minimalizacji wskaźnika jakości za pomocą algorytmu genetycznego. Doktorant stworzył algorytm i obszerny program komputerowy oraz przeprowadził obliczenia dla dwóch przypadków, w tym jednego dla rzeczywistych danych. Słabszą stroną rozprawy jest to, że Doktorant nie przedyskutował innych metod wyznaczania obszarów ufności i nie porównał z nimi zaproponowanej metody, ani nie wykazał, jakie korzyści ma zaproponowana metoda w stosunku do znanych metod. Wydaje się też, że metoda mogłaby być lepiej dopracowana i dawać dokładniejsze oceny niepewności, gdyby uwzględnić w niej także oszacowania obszaru ufności od zewnątrz i użyć ich do dokładniejszego wyznaczenia ocen.

Analizując formalnie zapisane w rozprawie tezy:

- Możliwe jest oszacowanie niepewności wyznaczenia wartości elementów macierzowych w analizie danych z pomiarów wzbudzeń kulombowskich w oparciu jedynie o próbkowanie przestrzeni rozwiązań przez algorytm genetyczny.
- Dokładność wyznaczenia niepewności przedstawioną Metodą AH jest porównywalna z innymi metodami wykorzystywanymi do estymacji niepewności wartości parametrów, a czas niezbędny do wykonania obliczeń jest znacząco krótszy.

to pierwsza teza jest w rozprawie wykazana. Podejście przedstawione w rozprawie jest oczywiście nowe i całkiem inne niż prezentowane w innych pracach. Natomiast mam nieco wątpliwości co do wykazania drugiej tezy, chociaż uważam, że jest ona najprawdopodobniej prawdziwa. Doktorant ograniczył się w rozprawie do dyskusji różnych sposobów próbkowania przestrzeni parametrów i porównania jakości uzyskiwanych w ten sposób ocen przedziałów ufności. Ale nie przedyskutował innych możliwych podejść do szacowania niepewności estymacji parametrów i nie porównał zaproponowanej metody z innymi metodami. Tak że ta teza jest sformułowana zbyt ogólnie w stosunku do badań zaprezentowanych w rozprawie.

W ocenie rozprawy trzeba jednak wziąć pod uwagę, że zaprezentowany pomysł jest nowy, a jego implikacje są znacznie dalej idące, niż to przedstawiono w rozprawie. W rozprawie wyznacza się tylko oszacowania niepewności w kierunkach osi poszczególnych parametrów, jednak w analogiczny sposób można wyznaczyć cały obszar ufności przy stosunkowo niewielkim obciążeniu czasowym. Takie oszacowania obszaru ufności można by było też uzyskać rozwiązując numerycznie odpowiednie równania wynikłe z przyjęcia, że wartość statystyki testowej chi kwadrat jest równa wartości krytycznej. Ten sposób wymagałby jednak policzenia wartości wskaźnika jakości zgodnie z wymaganiami metody wyznaczania

rozwiązania równania nieliniowego jednej zmiennej, co wymaga dodatkowych obliczeń statystyki chi kwadrat i znacznie dłuższego czasu w stosunku do korzystania z istniejących punktów. Może daloby się to zrobić dla kilku czy kilkunastu kierunków w przestrzeni parametrów, ale gdyby chciało się oszacować z satysfakcjonującą gęstością punkty na całej hiperpowierzchni ograniczającej obszar ufności, to obliczenia za pomocą numerycznych metod rozwiązywania równań byłyby prawdopodobnie czasowo niewydolne. Natomiast wydaje mi się, że policzenie takiej większej liczby punktów nie powinno trwać tak długo w metodzie zaproponowanej w rozprawie, a uzyskana dokładność może nie być specjalnie gorsza, szczególnie jeżeli dopracuje się zaproponowany w recenzji sposób korzystający z punktów leżących poza obszarem ufności. W ten sposób można by było praktycznie natychmiast po optymalizacji obejrzeć na przykład wykresy przekroju obszarów ufności i uzyskać w ten sposób wstępną jakościową ocenę jakości estymacji i obejrzeć na przykład ewentualne widoczne odchylenia uzyskanej hiperpowierzchni od hiperelipsoidy odpowiadającej przybliżeniu liniowemu.

Podsumowując powyższe uwagi, doceniam przedstawiony w rozprawie pomysł i uważam, że wskazuje on na możliwość efektywnego nieparametrycznego wyznaczania obszarów ufności dla modeli nieliniowych z wieloma parametrami, co jest zbyt trudne obliczeniowo dla innych metod. Dodatkowo należy wziąć pod uwagę, że rozprawa dotyczy praktycznego aspektu ważnych badań fizycznych i że opracowana metoda estymacji parametrów modelu oraz jego statystycznej oceny, być może po modyfikacjach, ma szanse na dołączenie i poprawienie działania programu GOSIA, który jest obecnie podstawowym narzędziem używanym w badaniach związanych z eksperymentami kulombowskimi.

Te względy skłaniają mnie do uznania, że rozprawa spełnia kryteria naukowe stawiane pracom doktorskim zarówno ustawowe jak i zwyczajowe. Wnioskuje więc do Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o dopuszczenie mgr. inż. Andrzeja Piętaka do publicznej obrony.



