

dr hab. inż. Maciej Wilczyński  
Wydział Matematyki  
Politechnika Wroclawska

**Recenzja rozprawy doktorskiej magistra Mateusza Johna zatytułowanej  
„Estymacja i testowanie macierzy kowariancji należących do podprzestrzeni  
kwadratowych”**

## 1 Opis problematyki.

Rozprawa doktorska magistra Mateusza Johna, napisana pod kierunkiem dr hab. inż. Katarzyny Filipiak, liczy 119 stron i składa się z wykazu ważniejszych skrótów i oznaczeń, wstępu, czterech rozdziałów, krótkiego podsumowania, a także bibliografii obejmującej 61 pozycji. W dysertacji przedstawiono wyniki badań, których celem była konstrukcja i analiza własności testów służących do weryfikacji hipotez o strukturach kowariancyjnych w standardowym modelu wielowymiarowym, w modelu krzywych wzrostu oraz w modelu podwójnie wielowymiarowym.

W wielu ważnych zagadnieniach wnioskowania statystycznego pojawia się problem, który w literaturze anglojęzycznej określany jest mianem *high dimensionality problem*. Ten termin oznacza, że liczba estymowanych parametrów jest znacznie większa od rozmiaru analizowanej próby, tzn.  $p \gg n$ . Taka sytuacja występuje w danych dotyczących obrazowania biomedycznego, funkcjonalnego rezonansu magnetycznego, tomografii komputerowej, przetwarzania sygnałów i analizy obrazów. W takich przypadkach klasyczne metody wnioskowania zazwyczaj zawodzą.

Jednym ze sposobów rozwiązania problemu danych wysoko wymiarowych jest regularyzacja. W analizie regresji to podejście pozwala znacznie zmniejszyć liczbę zmiennych objaśniających dołączanych do modelu, a więc i liczbę estymowanych parametrów. Najpopularniejszą z metod regularyzacji jest *Lasso* [Tibshirani (1996)]. Na uwagę zasługują także *Elastic net* [Zou and Hastie (2005)] i *Dantzig selector* [Candes and Tao (2007)] oraz *SLOPE* [Bogdan, Berg, Sabatti, Su and Candes (2015)].

Inny sposób rozwiązania problemu danych wysoko wymiarowych polega na nałożeniu pewnych ograniczeń na rozważany model statystyczny, co także może znacznie zmniejszyć liczbę szacowanych parametrów. W recenzowanej rozprawie doktorant wybrał właśnie ten sposób, wprowadzając pewne ograniczenia dotyczące struktury macierzy kowariancji, odzwierciedlającej zależności pomiędzy zmiennymi pojawiającymi się w analizowanym modelu statystycznym. Za główny cel swojej rozprawy autor uznał „zapropozowanie metod estymacji struktur kowariancyjnych należących do podprzestrzeni kwadratowych oraz konstrukcję testów statystycznych pozwalających na odpowiedni dobór struktury kowariancyjnej” .

## 2 Wyniki rozprawy.

Pierwszy rozdział to wprowadzenie do omawianych zagadnień. Autor podał w nim kluczowe definicje, wykorzystywane dalszej części pracy. Najpierw opisał rozważany w pracy model po-

dwójnie wielowymiarowy, w którym dla każdego z  $n$  obiektów obserwuje się  $m$  cech, ale wartość każdej z tych cech jest mierzona nie jeden raz lecz  $q$ -krotnie (na przykład w różnych lokalizacjach albo w różnych chwilach czasowych). W tym modelu każda z  $n$  niezależnych obserwacji  $\mathbf{y}_1, \dots, \mathbf{y}_n$  jest macierzą wymiaru  $m \times q$ , przy czym zakłada się, że macierz obserwacji  $\mathbf{Y} := (\mathbf{y}_1, \dots, \mathbf{y}_n)'$  ma macierzowy rozkład normalny postaci

$$\mathbf{Y} \stackrel{D}{=} \mathcal{N}_{n,mq}(\mathbf{1}_n \boldsymbol{\mu}', \mathbf{I}_n, \boldsymbol{\Omega}), \quad (1)$$

w którym  $\boldsymbol{\mu}$  jest wektorem  $mq$ -wymiarowym, a  $\boldsymbol{\Omega}$  jest dodatnio określoną macierzą kwadratową stopnia  $mq$ . Dla  $q = 1$  ten model redukuje się do standardowego modelu wielowymiarowego

$$\mathbf{Y} \stackrel{D}{=} \mathcal{N}_{n,m}(\mathbf{1}_n \boldsymbol{\mu}', \mathbf{I}_n, \boldsymbol{\Omega}), \quad (2)$$

w którym  $\boldsymbol{\mu}$  jest wektorem  $m$ -wymiarowym, a  $\boldsymbol{\Omega}$  jest dodatnio określoną macierzą kwadratową stopnia  $m$ . Doktorant zdefiniował także model krzywych wzrostu

$$\mathbf{Y} \stackrel{D}{=} \mathcal{N}_{n,m}(\mathbf{X}\mathbf{B}\mathbf{Z}, \mathbf{I}_n, \boldsymbol{\Omega}), \quad (3)$$

w którym  $\mathbf{B}$  jest macierzą nieznanymi parametrów wymiaru  $p_1 \times p_2$ , przy  $p_1 < n$ , natomiast  $\mathbf{X}$  i  $\mathbf{Z}$  są znanymi macierzami układu, wymiaru  $n \times p_1$  i  $p_2 \times m$ .

Ponieważ głównym celem rozprawy była estymacja i testowanie struktury macierzy kowariancji  $\boldsymbol{\Omega}$ , doktorant opisał następnie typowe struktury kowariancyjne w podprzestrzeniach kwadratowych, między innymi blokową strukturę diagonalną, blokową strukturę kompletnej symetrii oraz blokową strukturę kołowej macierzy Toeplitza. Zdefiniował statystykę ilorazu wiarygodności, statystykę wynikową Rao oraz statystykę Walda, czyli trzy statystyki testowe wykorzystywane do weryfikacji hipotez o strukturze macierzy kowariancji. Zaprezentował również wybrane przekształcenia macierzy wykorzystywane w rozprawie, takie jak operatory śladu blokowego, sumy blokowej i śladu częściowego. Omówił także zasady różniczkowania macierzy i opisał rzeczywisty eksperyment, który posłużył do zilustrowania głównych wyników pracy.

W pierwszej części rozdziału drugiego autor rozważył model komponentów wariacyjnych

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k \mathbf{Z}_i \boldsymbol{\beta}_i + \mathbf{e}. \quad (4)$$

W tej szczególnej wersji liniowego modelu mieszanego,  $\boldsymbol{\beta}$  jest nieznanym wektorem efektów stałych,  $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k$ , są nieznanymi wektorami efektów losowych, a macierz kowariancji  $\mathbf{V}$  wektora  $\mathbf{y}$  należy do pewnej przestrzeni liniowej  $\mathcal{V}$ , tzn.  $\mathbf{V} = \mathbf{V}(\boldsymbol{\sigma}) := \sum_{i=1}^k \sigma_i^2 \mathbf{V}_i$ , przy czym  $\mathbf{V}_1, \dots, \mathbf{V}_k$  są znanymi dodatnio określonymi macierzami symetrycznymi, zaś  $\sigma_1^2, \dots, \sigma_k^2$  są nieznanymi liczbami. Celem wnioskowania statystycznego jest estymacja współczynników  $\sigma_1^2, \dots, \sigma_k^2$ .

Zagadnienie szacowania komponentów wariacyjnych ma istotne znaczenie w recenzowanej rozprawie, gdyż standardowy model wielowymiarowy (??) można przekształcić poprzez wektoryzację do przypadku jednowymiarowego postaci (??). Zatem problem estymacji macierzy kowariancji  $\boldsymbol{\Omega}$  w modelu (??) można rozwiązać w podobny sposób jak w modelu mieszanym (??) pod warunkiem, że macierz  $\boldsymbol{\Omega}$  ma strukturę liniową zależną od wektora współczynników  $\boldsymbol{\sigma}$ .

Do rozwiązania tego problemu autor wykorzystał dwa twierdzenia udowodnione w Rozdziale 2. Z Twierdzenie 2.1 wynika, że jeśli  $\mathbf{y} \stackrel{D}{=} N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ , przy czym macierz kowariancji



$\mathbf{V} = \mathbf{V}(\boldsymbol{\sigma})$  należy do przestrzeni liniowej  $\mathcal{V}$  i komutuje z przestrzenią wartości oczekiwanej ( $\mathbf{P}_X \mathbf{V} = \mathbf{V} \mathbf{P}_X$ ), to estymator największej wiarygodności  $\widehat{\mathbf{V}}$  macierzy  $\mathbf{V}$  ma postać jawną wtedy i tylko wtedy gdy  $\mathcal{V}$  jest podprzestrzenią kwadratową (tzn.  $\mathbf{V} \in \mathcal{V} \implies \mathbf{V}^2 \in \mathcal{V}$ ). Ten wynik jest nietrywialnym rozszerzeniem rezultatu Szatrowskiego na przypadek, gdy wartość oczekiwana  $\mathbb{E}(\mathbf{y})$  nie jest znana. Z Twierdzenia 2.2, mającego podobne założenia co Twierdzenie 2.1, wynika, że jeśli  $\mathcal{V}$  jest podprzestrzenią kwadratową, to estymator  $\widehat{\mathbf{V}}$  macierzy  $\mathbf{V}$  jest rzutem estymatora największej wiarygodności macierzy kowariancji bez struktury na przestrzeń  $\mathcal{V}$ , tzn.  $\widehat{\mathbf{V}} = \mathbf{Q}_X \mathbf{y} \mathbf{y}' \mathbf{Q}_X$ ,  $\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X$ .

W kolejnej części Rozdziału 2. autor wykorzystał oba twierdzenia do znalezienia estymatorów kilku wybranych struktur kowariancyjnych dla standardowego modelu wielowymiarowego (??), który po wektoryzacji przybiera postać

$$\text{vec } \mathbf{Y} \stackrel{D}{=} \mathcal{N}_{qn}(\text{vec}(\mathbf{1}_n \boldsymbol{\mu}'), \boldsymbol{\Omega} \otimes \mathbf{I}_n).$$

Doktorant wykazał także, że jeśli w standardowym modelu wielowymiarowym współczynniki struktury liniowej (skalary) zostaną zastąpione przez macierze, a mnożenia macierzowe zostaną zastąpione przez iloczyny Kroneckera, to wszystkie wyniki uzyskane dla standardowego modelu wielowymiarowego mogą zostać rozszerzone na model podwójnie wielowymiarowy. Przykłady estymacji struktur kowariancyjnych w tym ogólniejszym modelu, a także w modelu krzywych wzrostu pojawiają się w ostatniej części rozdziału drugiego.

W Rozdziale 3. doktorant omówił problem weryfikacji hipotez dla standardowego modelu podwójnie wielowymiarowego (??). W tym problemie testowania zakłada się, że reprezentująca zależność międzyklasowe struktura blokowa dodatnio określonej macierzy kowariancji  $\boldsymbol{\Omega}$  należy do podprzestrzeni kwadratowej. Hipoteza zerowa i alternatywna mają postać:

$$H_0 : \boldsymbol{\Omega} = \sum_{i=1}^v \sum_{j=1}^{\omega} \lambda_{ij} (\mathbf{V}_i \otimes \mathbf{U}_j) \quad \text{vs} \quad H_1 : \boldsymbol{\Omega} = \sum_{i=1}^v \mathbf{V}_i \otimes \boldsymbol{\Delta}_i$$

gdzie  $\{\mathbf{U}_1, \dots, \mathbf{U}_{\omega}\}$  jest bazą komutatywnej podprzestrzeni kwadratowej  $\mathcal{U}$  macierzy stopnia  $q$ ,  $\{\mathbf{V}_1, \dots, \mathbf{V}_v\}$  jest bazą komutatywnej podprzestrzeni kwadratowej  $\mathcal{V}$  macierzy stopnia  $m$ , a  $\boldsymbol{\Delta}_1, \dots, \boldsymbol{\Delta}_v$  są dodatnio określonymi macierzami kowariancji stopnia  $p$ . Hipoteza  $H_0$  stwierdza, że każdy blok macierzy kowariancji ma strukturę należącą do pewnej komutatywnej podprzestrzeni kwadratowej. Do przestrzeni zdefiniowanej w  $H_0$  należy wiele kombinacji znanych struktur kowariancji, na przykład macierze o strukturze BD z blokami posiadającymi strukturę D, CS lub CT.

Dla powyżej zdefiniowanego problemu testowania, autor znalazł estymatory największej wiarygodności macierzy  $\boldsymbol{\Omega}$  dla  $H_0$  i  $H_1$ . Następnie wyznaczył statystyki testu Rao oraz testu ilorazu wiarygodności i dla  $H_0$  znalazł rozkład dokładny drugiego z tych testów. Za pomocą badań symulacyjnych wykazał, że test Rao jest szybciej zbieżny do rozkładu granicznego niż test ilorazu wiarygodności, ale oba testy są konkurencyjne pod względem mocy. Na koniec, wykorzystał oba testy do przeprowadzenia analizy danych rzeczywistych.

W Rozdziale 4. doktorant omówił problem testowania hipotezy o niezależności cech między dowolnymi dwoma powtarzonymi pomiarami w modelu podwójnie wielowymiarowym (??) z blokową strukturą kompletnej symetrii macierzy kowariancji (BCS). Hipoteza zerowa i alternatywna mają postać:

$$H_0 : \boldsymbol{\Omega} = \mathbf{I}_q \otimes \boldsymbol{\Gamma}_0 \quad \text{vs} \quad H_1 : \boldsymbol{\Omega} = \mathbf{I}_q \otimes \boldsymbol{\Gamma}_0 + (\mathbf{J}_q - \mathbf{I}_q) \otimes \boldsymbol{\Gamma}_1 := \boldsymbol{\Gamma}.$$



W tym modelu  $\Gamma_0$  i  $\Gamma_1$  są macierzami stopnia  $m$ , pierwsza z nich jest dodatnio określona, a druga nieujemnie określona, przy czym obie macierze są tak dobrane, by  $\Gamma$  była dodatnio określona. Macierz  $\Gamma_0$  jest macierzą kowariancji  $m$  cech w dowolnym powtarzanym pomiarze, natomiast macierz  $\Gamma_1$  jest macierzą kowariancji  $m$  cech między dowolnymi dwoma powtarzonymi pomiarami.

Doktorant wyznaczył statystyki testu ilorazu wiarygodności, testu wynikowego Rao oraz testu Walda. Za pomocą badań symulacyjnych porównał własności asymptotyczne i moc tych trzech testów oraz testu F i testu największego pierwiastka Roya. Przeanalizował także odporność rozkładów rozważanych statystyk testowych na zaburzenie założenia o normalności rozkładu obserwacji. Na koniec, wykorzystał wymienione testy do analizy danych rzeczywistych.

### 3 Ocena rozprawy.

Rozprawa doktorska mgr Mateusza Johna zawiera nowe i wartościowe rezultaty dotyczące weryfikowania hipotez statystycznych o strukturach kowariancyjnych w standardowym modelu wielowymiarowym, w modelu krzywych wzrostu oraz w modelu podwójnie wielowymiarowym. Zostały one opublikowane w trzech artykułach. Jeden z nich to rozdział w monografii wydawnictwa Springer, a dwa pozostałe ukazały się w renomowanych czasopismach statystycznych:

1. Filipiak, K., John, M., Markiewicz, A. (2020). Comments on Maximum Likelihood Estimation and Projections Under Multivariate Statistical Models. In: Holgersson, T., Singull, M. (eds) Recent Developments in Multivariate and Random Matrix Analysis. Springer.
2. Filipiak, K., John, M. and Klein, D. Testing independence under a block compound symmetry covariance structure. Stat Papers 64, 677–704 (2023).
3. Filipiak, K., John, M. and Liang, Y. Testing covariance structures belonging to a quadratic subspace under a doubly multivariate model. TEST (2024).

Uważam, że zagadnienia rozwiązane w rozprawie wnoszą istotny wkład w rozwój dyscypliny. Udowodnienie najważniejszych wyników wymagało od autora nie tylko dogłębnej znajomości literatury, ale także wykorzystania pomysłowej argumentacji i przeprowadzenia wielu trudnych i żmudnych rachunków. Z każdym z tych problemów doktorant poradził sobie bardzo dobrze, a uzyskane przez niego rezultaty są wartościowe także z powodu ich potencjalnych zastosowań w analizie danych rzeczywistych.

Rozprawa jest starannie zredagowana i ma dobrze przemyślaną strukturę. Jej lekturę ułatwia świetnie napisany Rozdział 1., zawierający wprowadzenie do rozważanej tematyki. Przydatny jest wykaz ważniejszych skrótów i oznaczeń, pojawiający się na początku pracy. Na uwagę zasługuje również krótkie podsumowanie, w którym doktorant streścił główne wyniki rozprawy oraz przedstawił kierunki przyszłych badań. Wszystkie rezultaty formułowane są w sposób przejrzysty i kompletny, a żaden z zamieszczonych w rozprawie dowodów nie wzbudził moich zastrzeżeń. Nie zauważyłem w pracy istotnych błędów językowych, mało jest literówek lub drobnych uchybień stylistycznych. Mam tylko jedną krytyczną uwagę, która dotyczy następującej kwestii. Trzy opublikowane artykuły, zawierające kluczowe wyniki rozprawy, są pracami wieloautorskimi. W związku z tym trudno jednoznacznie ocenić, jaki jest wkład własny autora rozprawy. Jednak ta uwaga nie wpływa na moją wysoką ocenę dysertacji. Po wysłuchaniu znakomitego referatu, wygłoszonego przez mgr. Mateusza Johna podczas seminarium ze statystyki matematycznej

na Wydziale Matematyki i Informatyki Uniwersytetu Wrocławskiego, jestem przekonany, że doktorant odegrał istotną rolę w powstaniu tych trzech artykułów.

## 4 Konkluzja

Rozprawa doktoranta zawiera oryginalne rozwiązania kilku nietrywialnych problemów matematycznych i znacząco przyczynia się do rozwoju wnioskowania statystycznego w modelach wielowymiarowych, poszerzając istniejący stan wiedzy i wpisując się w światowy nurt badań. Nie mam najmniejszych wątpliwości, że dysertacja reprezentuje ponadprzeciętny poziom i z naddatkiem spełnia wszystkie wymagania stawiane pracom doktorskim. Uważam, że doktorant posiada wiedzę i umiejętności niezbędne do samodzielnego prowadzenia badań naukowych. Wnoszę o dopuszczenie mgr. Mateusza Johna do dalszych etapów postępowania o nadanie stopnia doktora w dyscyplinie matematyka.

Maciej Wilczyński



