

Rafał Stocki

ANALIZA NIEZAWODNOŚCI
I OPTYMALIZACJA ODPORNOŚCIOWA
ZŁOŻONYCH KONSTRUKCJI I PROCESÓW
TECHNOLOGICZNYCH



INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI
POLSKIEJ AKADEMII NAUK

WARSZAWA 2010

ISBN 978-83-89687-58-6

ISSN 0208-5658

Redaktor Naczelny:

prof. dr hab. inż. Zbigniew Kotulski

Recenzent:

prof. dr hab. inż. Zbigniew Zembaty

Praca wpłynęła do redakcji w sierpniu 2010

Instytut Podstawowych Problemów Techniki PAN

Nakład 100 egz. Ark. wyd. 21

Oddano do druku w październiku 2010

Druk i oprawa: Drukarnia Braci Grodzickich, ul. Geodetów 47A, Piaseczno

Spis treści

1. Wstęp	9
1.1. Optymalizacja odpornościowa jako nowoczesna koncepcja projektowania	14
1.2. Cel i zakres pracy	15
1.3. Przegląd literatury	21
1.3.1. Analiza niezawodności w zagadnieniach z niejawną funkcją graniczną	21
1.3.2. Optymalizacja odpornościowa	24
1.3.3. Metody powierzchni odpowiedzi	29
2. Optymalizacja konstrukcji o parametrach losowych	33
2.1. Optymalizacja deterministyczna	34
2.2. Optymalizacja wielokryterialna - rozwiązanie Pareto	35
2.3. Optymalizacja niezawodnościowa	38
2.3.1. Metody rozwiązania	42
2.3.1.1. Metoda dwupoziomowa	42
2.3.1.2. Metoda jednopoziomowa	44
2.3.1.3. Metoda rozprzężona	46
2.4. Optymalizacja odpornościowa	47
2.4.1. Sformułowanie problemu	50
2.4.2. Metody rozwiązania	54
2.4.2.1. Podstawowa strategia aproksymacji	55
2.4.2.2. Strategia aproksymacji statystyk	57
2.4.2.3. Strategia aproksymacji funkcji celu i ograniczeń odpornościowych	59

3. Metody powierzchni odpowiedzi w analizie eksperymentów komputerowych	63
3.1. Liniowa analiza regresji	65
3.2. Kriging	69
3.3. Metoda aproksymacji lokalnej	78
4. Efektywne metody szacowania momentów statystycznych odpowiedzi konstrukcji	83
4.1. Metody symulacyjne	84
4.1.1. Metody <i>descriptive sampling</i> a klasyczna metoda Monte Carlo	85
4.1.2. Podstawowe informacje na temat estymatorów	87
4.1.3. Analiza głównych składowych	92
4.1.4. Przykłady numeryczne: Analityczne funkcje testowe	94
4.1.4.1. Monotoniczna funkcja dwóch zmiennych	94
4.1.4.2. Niemonotoniczna funkcja dwóch zmiennych	101
4.1.4.3. Wielowymiarowa funkcja Rosenbrocka	104
4.1.4.4. Funkcja wykładnicza	107
4.1.5. Przykład: Kratownica Misesa	109
4.1.6. Przykład: Płyta wspornikowa	116
4.1.7. Wnioski dotyczące technik symulacji losowych wykorzystujących łańciskie hiperkostki	121
4.2. Metoda rozwinięcia w szereg Taylora	122
4.2.1. Metoda pierwszego rzędu	122
4.2.2. Metoda drugiego rzędu	123
4.3. Metoda redukcji wymiarów	125
4.4. Metoda rozwinięcia funkcji losowej w chaos wielomianowy	132
5. Metody analizy niezawodności oraz metody poprawy niezawodności konstrukcji w przypadku zaszumionej funkcji granicznej	137
5.1. Dwuetapowa adaptacyjna metoda redukcji wariancji	138
5.1.1. Etap pierwszy: Poszukiwanie punktu projektowego	139
5.1.2. Etap drugi: Budowa wielomodalnej gęstości losującej	142
5.1.3. Szum numeryczny generowany przez losową geometrię konstrukcji. Analiza FORM przy użyciu procedury Z3M	147
5.2. Metoda poprawy niezawodności z wykorzystaniem losowych symulacji OLH	149

5.2.1. Sformułowanie problemu	150
5.2.2. Opis algorytmu	152
5.2.2.1. Wybór punktu startowego	152
5.2.2.2. Kryterium wyboru następnej iteracji zmiennych projek- towych	153
5.2.3. Zmodyfikowany plan OLH	156
5.2.4. Przykład testowy: Minimalizacja funkcji Rosenbrocka z szumem	158
6. Analiza niezawodności i optymalizacja odpornościowa w zagadnie- niach wytrzymałości zderzeniowej	165
6.1. Analiza niezawodności zgniatanej dynamicznie belki cienkościennej z uwzględnieniem losowych wad połączeń zgrzewanych	166
6.2. Algorytm poprawy poziomu niezawodności w zastosowaniu do zagadnień wytrzymałości zderzeniowej	177
6.3. Jakościowa analiza zachowania elementu zgniatanego za pomocą algo- rytmu klasyfikacji wyników symulacji OLH	180
6.4. Optymalizacja odpornościowa zgniatanej dynamicznie belki cienkościen- nej	188
6.5. Optymalizacja odpornościowa w zadaniu kalibracji parametrów modelu numerycznego	193
7. Analiza niezawodności procesu tłoczenia blach	203
7.1. Modelowanie procesów tłoczenia blach	204
7.1.1. Wykresy odkształceń granicznych	208
7.2. Definicja funkcji granicznej zadania analizy niezawodności	211
7.3. Analiza niezawodności dla przykładu testowego konferencji NUMISHEET	213
8. STAND - pakiet obliczeniowy do analizy niezawodności i optymalizacji konstrukcji	227
8.1. Analiza niezawodności konstrukcji	230
8.1.1. Definicja modelu stochastycznego	231
8.1.1.1. Zmienne losowe	231
8.1.1.2. Zmienne zewnętrzne	233
8.1.2. Definicja funkcji granicznej	235
8.1.3. Wybór algorytmu i uruchomienie analizy	235
8.1.4. Wizualizacja wyników	237

8.1.5. Obiektowa architektura programu STAND na przykładzie modu- łu analizy niezawodności	239
8.2. Wybrane zagadnienia implementacji algorytmu optymalizacji odporno- ściowej	243
8.3. Przykłady łączenia z zewnętrznymi komercyjnymi programami MES .	248
8.3.1. Łączenie z programem ABAQUS	248
8.3.2. Łączenie z programem RADIOSS	250
Wnioski końcowe. Elementy oryginalne pracy	255
A. Wybrane zagadnienia komputerowej analizy niezawodności konstruk- cji	261
A.1. Sformułowanie problemu analizy niezawodności	261
A.2. Transformacja zmiennych losowych	263
A.2.1. Zmienne o wielowymiarowym rozkładzie normalnym	265
A.2.2. Zmienne niezależne	266
A.2.3. Zmienne o dowolnym rozkładzie prawdopodobieństwa	266
A.2.4. Zmienne o dowolnym rozkładzie - transformacja przybliżona . .	267
A.3. Algorytmy poszukiwania punktu projektowego	269
A.3.1. Metoda ARF	273
A.3.2. Metoda iHLRF	274
A.4. Metody symulacyjne	274
A.4.1. Klasyczna metoda Monte Carlo	275
A.4.2. Metoda importance sampling	277
A.4.3. Metody adaptacyjne	281
B. Algorytmy tworzenia optymalnych łańciskich hiperkostek	285
B.1. Kryteria optymalizacji	286
B.2. Algorytm CP	288
B.3. Algorytm genetyczny	289
B.4. Warunki zbieżności	291
B.5. Analiza złożoności obliczeniowej algorytmu CP	292
B.6. Dobór parametrów algorytmu genetycznego	295
B.7. Wykorzystanie łańciskich hiperkostek do generowania próbek zmiennych losowych o dowolnych rozkładach prawdopodobieństwa	297
B.8. Porównanie kryteriów optymalizacji w zadaniu estymacji momentów funkcji losowych	298

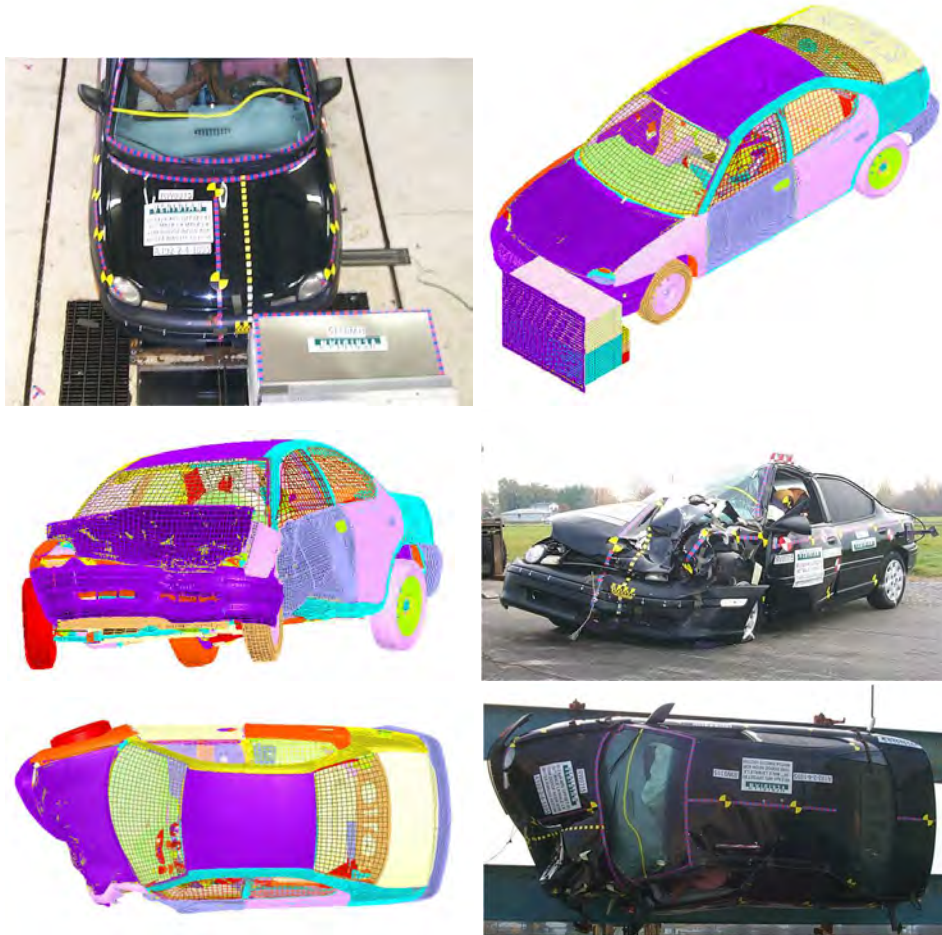
C. Metody dyskretyzacji pól losowych	303
C.1. Podstawowe definicje	303
C.2. Dyskretyzacja punktowa	306
C.3. Metoda średniej przestrzennej	308
C.4. Metoda funkcji kształtu	308
C.5. Metoda optymalnej aproksymacji liniowej	309
C.6. Niegaußowskie pola losowe	312
C.7. Dobór wielkości elementów pola losowego	312
Tłumaczenia angielskich skrótów i terminów występujących w tekście	315
Bibliografia	319

Wstęp

Nieustanny rozwój techniki obliczeniowej pozwala obecnie na tworzenie coraz większych i bardziej dokładnych modeli nieliniowych zjawisk fizycznych. Powszechną praktyką staje się np. komputerowa analiza zderzeń pojazdów, zob. rysunki 1.1 i 1.2. Modele skończenie elementowe mogą zawierać setki tysięcy a nawet miliony elementów skończonych, a do symulacji używa się maszyn wieloprocesorowych o olbrzymiej mocy. Symulacje numeryczne umożliwiają



Rys. 1.1. Chrysler Neon, model MES (około 300 000 elementów skończonych). Rysunek pochodzi z raportu projektu APROSYS, zob. [198].



Rys. 1.2. Porównania deformacji samochodu Chrysler Neon w teście zderzenia czołowego z przesunięciem, uzyskanych doświadczalnie i przy pomocy symulacji MES. Rysunek pochodzi z raportu projektu APROSYS, zob. [46].

dokładne zbadanie wszelkich charakterystyk danego problemu inżynierskiego, zanim jeszcze ostateczny produkt lub nawet jego prototyp zostaną wykonane. Znaczenie tego procesu, zwanego *wirtualnym prototypowaniem*, ciągle wzrasta. Związane jest to z potrzebą skracania czasu trwania cykli projektowych, która podyktowana jest przede wszystkim presją rynku i wymogami konkurencyjności. Wirtualne symulacje pozwalają na szybką analizę wielu alternatywnych rozwiązań, przyczyniając się do podniesienia efektywności procesu projektowego oraz redukując koszty poprzez zmniejszenie liczby niezbędnych testów fizycznych. Ponadto, modelowanie komputerowe ułatwia zastosowanie szeregu metod optymalizacyjnych.

Wynikiem przeprowadzonej optymalizacji, jest konstrukcja/produkt/proces technologiczny, które charakteryzują się optymalnymi cechami ze względu na kryteria przyjęte jako miara ich jakości. O przydatności tak otrzymanego rozwiązania w sposób oczywisty decydują dwa czynniki. Jednym z nich jest adekwatność samego modelu numerycznego, który musi dobrze odzwierciedlać rzeczywiste zjawisko fizyczne. Niespełnienie tego warunku prowadzi do poważnych błędów, a w konsekwencji złych decyzji. Drugi czynnik to właściwe sformułowania zadania optymalizacji. Nietrafny dobór funkcji celu oraz ograniczeń projektowych może łatwo sprawić, że rozwiązanie optymalne będzie zupełnie nieprzydatne.

Z drugiej strony, co wydaje się być już powszechnie podzielaną opinią, nieodzownym składnikiem wirtualnego prototypowania powinna stać się analiza wpływu, jaki na modelowane zjawisko ma losowy charakter opisujących je parametrów. Nie rozpatrując tego zagadnienia, projektant traci niezwykle istotną informację dotyczącą jakości, a co się z tym wiąże, niezawodności tworzonego produktu. Rozwiązania, które spełniają swoją funkcję dla nominalnych wartości parametrów mogą okazać się nie do zaakceptowania jeśli weźmie się pod uwagę imperfekcje parametrów modelu oraz oddziaływań zewnętrznych. Nieunikniony losowy rozrzut wielkości parametrów materiałowych, wymiarów czy też działających obciążeń może mieć znaczący wpływ na zachowanie się konstrukcji czy poprawne działanie procesu technologicznego. Jest to wyjątkowo istotne w przypadku układów wysoce nieliniowych oraz projektów będących wynikiem optymalizacji deterministycznej. To właśnie tzw. konstrukcje optymalne są zazwyczaj szczególnie wrażliwe na imperfekcje parametrów. Umiejscowione na granicy obszaru dopuszczalnego rozwiązania optymalne mogą stosunkowo łatwo okazać się całkowicie nieprzydatnymi jeśli wartości parametrów będą różniły się od założonych wartości nominalnych. Naturalnym wydaje się zatem rozszerzenie sformułowania optymalizacji deterministycznej, w którym uwzględnia się niepewność wartości parametrów. Sformułowania takie oferuje optymalizacja niezawodnościowa oraz optymalizacja odpornościowa [252]. Miarą wpływu losowości parametrów może być natomiast prawdopodobieństwo awarii. Do szacowania wielkości tego prawdopodobieństwa służą metody komputerowej analizy niezawodności [143, 173].

Pomimo dość powszechnego rozumienia znaczenia analizy stochastycznej oraz potrzeby racjonalnego uwzględniania losowego charakteru parametrów konstrukcji, analiza ta jednak nie jest ogólnie przyjętą praktyką projektową. Niestety, dość często dla prostoty lub w celu zredukowania kosztu obliczeń, rezygnuje się z dokładnego badania efektów rozrzutu wartości parametrów. Do obliczeń

przyjmuje się zazwyczaj takie miary statystyczne jak wartość średnia czy kwantyle rozkładów prawdopodobieństwa pomnożone przez odpowiednie współczynniki bezpieczeństwa. Wymienić można szereg przyczyn tego stanu rzeczy. Najistotniejszymi wydają się być:

- trudności w zebraniu odpowiednich danych do przygotowania probabilistycznego opisu zagadnienia,
- ciągle niski poziom wiedzy na temat analizy niedeterministycznej wśród projektantów i inżynierów,
- wysoki koszt takiej analizy w porównaniu do tradycyjnej analizy deterministycznej,
- ograniczony dostęp do wyspecjalizowanych, przyjaznych użytkownikowi programów, które umożliwiają np. analizę niezawodności, optymalizację niezawodnościową lub optymalizację odpornościową.

Z punktu widzenia projektanta ważne jest też aby oprogramowanie do analizy stochastycznej można było łatwo integrować z programami do analizy strukturalnej, w szczególności z komercyjnymi programami metody elementów skończonych.

Efektywne uwzględnienie losowej natury parametrów modelowanych układów w procesie projektowania wymaga zatem ciągłego udoskonalania metod analizy stochastycznej, jak również prac nad oprogramowaniem umożliwiającym taką analizę. Problemy z brakiem dostatecznych danych statystycznych do stworzenia łącznych rozkładów prawdopodobieństwa zmiennych narzucają konieczność użycia dostosowanych sformułowań zadania optymalizacji, np. sformułowania optymalizacji odpornościowej.

Udoskonalanie algorytmów i sformułowań jest niewątpliwie koniecznym elementem poprawy efektywności analizy stochastycznej. Realizacja tego celu staje się jednak dużo łatwiejsza dzięki postępowi, a nawet skokowi jakościowemu, jaki dokonuje się w rozwoju techniki obliczeniowej. Od momentu powstania pierwszych elektronicznych komputerów opartych na układach scalonych aż do chwili obecnej, obserwuje się wykładniczy wzrost ich mocy obliczeniowej. W związku ze zbliżaniem się do granicy możliwości miniaturyzacji, mierzonej umieszczeniem jak największej liczby tranzystorów w układzie scalonym, zob. rys. 1.3, trend całego przemysłu komputerowego skierowany jest teraz na tworzenie układów wieloprocessorowych (lub wielordzeniowych) i na przetwarzanie równoległe, które stosowano do tej pory w wydajnych serwerach i superkomputerach. Maszyny wieloprocessorowe stają się dużo bardziej dostępne. Nikogo obecnie nie dziwi

dziedzin, gdzie jeszcze niedawno, ze względu na olbrzymi koszt pojedynczej symulacji, nie do pomyślenia była jakakolwiek analiza niedeterministyczna. Komputerowa analiza niezawodności jak również optymalizacja odpornościowa zaczynają być stosowane nawet w zagadnieniach dotyczących projektowania konstrukcji samochodowych pod kątem ich wytrzymałości zderzeniowej, czy też w procesach głębokiego tłoczenia blach. Metodom komputerowej realizacji właśnie takich zagadnień poświęcona jest niniejsza praca.

1.1. Optymalizacja odpornościowa jako nowoczesna koncepcja projektowania

Ponieważ rozwiązania optymalne klasycznej-deterministycznej optymalizacji mogą być bardzo wrażliwe na niewielkie zmiany parametrów opisujących je modeli, dlatego bardziej racjonalnym celem, do którego powinno się dążyć w procesie projektowania, jest zapewnienie wystarczająco wysokiego poziomu odporności konstrukcji/procesu na te zmiany, zob. [13]. A zatem, dążąc do minimalizacji lub maksymalizacji wybranych charakterystyk konstrukcji, powinno dążyć się jednocześnie do znalezienia rozwiązania, które jest niewrażliwe na trudne do kontrolowania imperfekcje parametrów modelu oraz zewnętrznych oddziaływań. Największą zaletą optymalizacji odpornościowej jest właśnie to, że prowadzi ona do rozwiązań, które zachowują swoją jakość i założoną funkcjonalność w szerokim zakresie warunków pracy.

Dążenie do rozwoju technik optymalizacji odpornościowej nie jest jednak poddyktowane jedynie potrzebą zmniejszenia wrażliwości funkcji celu oraz funkcji ograniczeń na losowe zmiany parametrów. Jeśli zmienne projektowe opisują jedynie część większej całości - systemu, niezbędne jest poczynienie założeń dotyczących zachowania wszystkich jego elementów dla danych realizacji zmiennych. Wymaga to bądź przeprowadzania symulacji odpowiedzi całego zintegrowanego systemu, bądź założenia pewnych zakresów wielkości oddziaływania pozostałych części na część optymalizowaną, co w sposób naturalny prowadzi do sformułowania optymalizacji odpornościowej. Z podobną sytuacją mamy do czynienia, gdy kilka zespołów projektantów tworzy złożony system optymalizując niezależnie jego poszczególne elementy. Ze względu na stopień skomplikowania systemu oraz ograniczenia czasowe wynikające z narzuconych terminów, zespoły często muszą optymalizować swoje elementy bez pełnej informacji na temat pozostałych elementów, których odpowiedzi stanowią dane wejściowe optymalizowanego podsystemu.

Ciekawą, filozoficzną argumentację wyjaśniającą zalety sformułowania optymalizacji odpornościowej w stosunku do optymalizacji niezawodnościowej przedstawił Hurtado w pracy [102]. Ze względu na możliwość użycia uproszczonego modelu niepewności parametrów, optymalizacja ta w dużo większym stopniu odpowiada oczekiwaniom (i możliwościom) środowiska inżynierów projektantów. Niektórzy z nich cenią bowiem dużo wyżej koncepcję odporności, rozumianej jako bezpieczeństwo w przypadku nieprzewidzianych odstępstw parametrów od założonych wartości nominalnych, niż znajomość prawdopodobieństwa awarii, którego małe wartości nie mają, ich zdaniem, istotnego praktycznego znaczenia. Nie podzielając do końca tej opinii, jako nieuwzględniającej wszystkich wyzwań stawianych przy projektowaniu nowoczesnych konstrukcji, w pracy przedstawione zostaną również nowe metody umożliwiające ocenę niezawodności w złożonych zagadnieniach mechaniki, modelowanych przy pomocy metody elementów skończonych.

1.2. Cel i zakres pracy

Za główny cel pracy należy uznać opracowanie sformułowań teoretycznych oraz algorytmów numerycznych, które umożliwiają analizę stochastyczną złożonych konstrukcji i procesów technologicznych. Przez analizę stochastyczną rozumieć będziemy tu szereg zagadnień, a w szczególności:

- analizę losowego charakteru odpowiedzi układów konstrukcyjnych, wyznaczenie parametrów rozkładu prawdopodobieństwa odpowiedzi,
- analizę niezawodności, tj. oszacowanie prawdopodobieństwa awarii konstrukcji, bądź procesu technologicznego,
- optymalizację odpornościową, gdzie oprócz wartości średnich wybranych kryteriów minimalizuje się wariancje tych kryteriów.

Podkreślenie, iż opisana analiza stochastyczna dotyczyć ma złożonych konstrukcji i procesów służy w istocie zaakcentowaniu konieczności opracowania wyspecjalizowanych, nieklasycznych metod rozwiązania. Jako przykłady zaawansowanych zagadnień mechaniki, które uznać można za reprezentatywne dla złożoności obliczeniowej współcześnie przeprowadzanych analiz numerycznych, wybrano zagadnienia wytrzymałości zderzeniowej elementów konstrukcji pojazdów oraz proces głębokiego tłoczenia blachy. Jakościowy charakter odpowiedzi tego typu konstrukcji modelowany również będzie za pomocą szeregu specjalnie dobranych przykładów analitycznych, jak również mniejszych, testowych zadań mechaniki konstrukcji.

Pomimo nieustannego postępu techniki komputerowej, sam wzrost mocy obliczeniowej nie jest wystarczającym środkiem do zapewnienia "ekspansji" analizy stochastycznej na nowe, niedostępne dotychczas obszary zastosowań. W dalszym ciągu, szczególnie gdy problemy mechaniki reprezentowane są przy pomocy złożonych modeli MES, użycie klasycznych metod symulacji Monte Carlo wiąże się z ogromnym nakładem obliczeniowym. W większości przypadków dodatkowy koszt generowany przez analizę stochastyczną jest niewspółmiernie wysoki, co podważa zasadność przeprowadzania tego typu pogłębionej analizy. Z drugiej strony, np. w analizie niezawodności złożonych konstrukcji i procesów technologicznych, najczęściej nie jest możliwe bezpośrednie zastosowanie bardziej efektywnych metod oszacowania wartości prawdopodobieństwa awarii, takich które wykorzystują koncepcję punktu projektowego. Użycie gradientowych algorytmów lokalizacji punktu projektowego w analizie niezawodności elementów absorbujących energię zderzeń, skazane jest z góry na niepowodzenie. Funkcje graniczne, z którymi ma się tam do czynienia, zazwyczaj nie są różniczkowalne, a obserwowany numeryczny szum skutecznie utrudnia zastosowanie jakichkolwiek metod niesymulacyjnych. Nieróżniczkowalne, a jednocześnie kosztowne obliczeniowo, są także funkcje celu i funkcje ograniczeń zadania optymalizacji odpornościowej. Wszystko to sprawia, że niezbędnym składnikiem efektywnej analizy stochastycznej złożonych konstrukcji są wyspecjalizowane algorytmy, które nie są wrażliwe na silnie nieliniowy charakter odpowiedzi konstrukcji, a jednocześnie są w stanie wykorzystać możliwości przetwarzania równoległego oferowane przez współczesne komputery.

W niniejszej pracy szczególny nacisk położony zostanie na następujące elementy budowanych algorytmów:

- Efektywne symulacyjne metody analizy losowego rozrzutu odpowiedzi konstrukcji. Zastosowane będą metody typu "descriptive sampling", wykorzystujące koncepcję łącińskiej hiperkostki oraz optymalnej łącińskiej hiperkostki. Metody te łączą dobrą efektywność estymacji momentów statystycznych funkcji losowych z małą wrażliwością na charakter zmienności tych funkcji oraz na typy rozkładów prawdopodobieństwa zmiennych losowych. Użyte zostaną efektywne algorytmy tworzenia optymalnych hiperkostek (ang. *optimal Latin hypercube* - OLH).
- Nowoczesne techniki aproksymacji nieliniowych funkcji wielu zmiennych (metody powierzchni odpowiedzi). Wykorzystywane będą przede wszystkim: metoda ważonej liniowej regresji oraz metoda krigingu. To właśnie metoda krigingu, obok efektywnych technik symulacji losowych, stanowić będzie kluczowy element algorytmu rozwiązania zadania optymalizacji

odpornościowej. Jako dominujący plan eksperymentów używany będzie plan punktów generowanych przez optymalne łącińskie hiperkostki.

- Wykorzystanie rozwiązań niewrażliwych na szum numeryczny, charakterystyczny dla jawnego schematu całkowania równań ruchu oraz algorytmów kontaktu stosowanych w modelach MES. Jest to niezbędny warunek zapewnienia zbieżności zarówno algorytmów analizy niezawodności jak też optymalizacji odpornościowej.

Ponadto, dodatkowym celem autora było bliższe przedstawienie koncepcji optymalizacji odpornościowej. Ten typ optymalizacji jest jeszcze ciągle mało znany, szczególnie w polskiej literaturze, i czasami mylony z optymalizacją niezawodnościową. Nawet zaproponowana przez autora nazwa “optymalizacja odpornościowa” nie jest jeszcze powszechnie przyjętym tłumaczeniem angielskiego terminu *robust optimization*. Mając to na uwadze, w niniejszej pracy podjęto próbę usystematyzowania wiedzy na temat alternatywnych sformułowań zadania niedeterministycznej optymalizacji konstrukcji. Przedstawione będą sformułowania oraz zaproponowane zostaną strategie rozwiązania zadania optymalizacji odpornościowej.

Wszystkie rozwijane w pracy metody zaimplementowano w obiektowo zorientowanym programie STAND, który współtworzony jest przez autora w ramach badań prowadzonych w Pracowni Niezawodności i Optymalizacji IPPT PAN. Efektywne tworzenie dużego programu do analizy stochastycznej konstrukcji, przeznaczonego zarówno do analizy niezawodności jak i optymalizacji odpornościowej, a także zadanie jego integracji z zewnętrznymi pakietami obliczeniowymi MES, stanowią ciekawe i nietrywialne problemy informatyczne. Zagadnienia te, o niebagatelnym znaczeniu w praktyce, zostaną w pracy szczegółowo omówione. Zaproponowanych będzie szereg rozwiązań programistycznych dotyczących architektury kodu programu STAND.

Praca składa się z ośmiu rozdziałów i trzech dodatków.

Rozdział 1 zawiera między innymi uzasadnienie decyzji podjęcia rozważanej tematyki badawczej, jej cel i zakres, a także przegląd literatury. Podkreślono korzyści płynące ze stosowania różnych form analizy stochastycznej w procesie projektowania konstrukcji, a jednocześnie ukazano trudności związane z szerszym upowszechnieniem takiego podejścia. Zwrócono szczególną uwagę na renesans metod symulacyjnych wynikający z gwałtownego rozwoju i dostępności rozwiązań wieloprocessorowych oraz z użycia zaawansowanych technik symulacji pseudolosowych. Przegląd literatury dotyczy w szczególności prac związanych z analizą niezawodności złożonych konstrukcji, których zachowanie modelowane jest przy pomocy metody elementów skończonych, jak również zagadnień

optymalizacji odpornościowej oraz technik aproksymacji odpowiedzi układów konstrukcyjnych. Wspomniano najważniejsze publikacje, które charakteryzują rozwój oraz obecny stan wiedzy w wymienionych dziedzinach. Literatura dotycząca bardziej szczegółowych zagadnień jest przedstawiona w poszczególnych rozdziałach pracy.

Rozdział 2 poświęcony jest sformułowaniu zadania optymalizacji konstrukcji o parametrach losowych. Przedstawiono dwa najbardziej znane sformułowania: optymalizację niezawodnościową i optymalizację odpornościową. W przypadku optymalizacji niezawodnościowej, obszernie omówiono najważniejsze strategie rozwiązania, przede wszystkim pod kątem ich efektywności obliczeniowej. Omówiono również problemy związane z szerszym zastosowaniem optymalizacji niezawodnościowej w praktyce. Problemy te spowodowane są głównie koniecznością precyzyjnej znajomości rozkładów prawdopodobieństwa zmiennych losowych. Jako bardziej realistyczne podejście do zagadnienia niedeterministycznej optymalizacji konstrukcji uznano sformułowanie optymalizacji odpornościowej. Przedstawiono trzy strategie aproksymacji funkcji wchodzących w skład tego sformułowania. Po analizie wad i zalet każdej z nich zaproponowano użycie tzw. strategii aproksymacji funkcji celu i ograniczeń odpornościowych.

Tematem rozdziału 3 jest zastosowanie metod powierzchni odpowiedzi w aproksymacji zachowania układów konstrukcyjnych, które modelowane są komputerowo. Omówiono specyfikę eksperymentów komputerowych oraz problemy ze spełnieniem założeń klasycznej metody regresji liniowej. Pokazano, że metodą, która jest najlepiej dostosowana do analizy nieliniowych odpowiedzi konstrukcji jest metoda krigingu. Planując wykorzystanie tej metody do aproksymacji funkcji zawierających składnik szumowy (jak ma to miejsce w optymalizacji odpornościowej) zaprezentowano jak zmodyfikować równania krigingu interpolującego w celu otrzymania techniki aproksymacyjnej. Zbadano wpływ równomierności rozłożenia punktów eksperymentalnych na błąd dopasowania. Przedstawiono również metodę aproksymacji lokalnej wykorzystującą koncepcję tzw. ważonej regresji liniowej. Zarówno kriging, jak też metoda aproksymacji lokalnej stanowią podstawę proponowanych w pracy algorytmów analizy niezawodności oraz optymalizacji odpornościowej. Pokazano, że w przypadku krigingu, wykorzystanie omówionych w dodatku B planów eksperymentów OLH poprawia dopasowanie funkcji aproksymującej.

Rozdział 4 poświęcono analizie metod estymacji momentów statystycznych odpowiedzi konstrukcji. Obok metod aproksymacji, jest to kluczowe zagadnienie, które decyduje o efektywności rozwiązania zadania optymalizacji odpornościowej. W rozdziale skupiono się na bardzo dokładnym omówieniu metod

symulacyjnych, uznając je za najbardziej uniwersalne i wiarygodne. Przedstawiono ideę symulacyjnych metod *descriptive sampling* i porównano je z klasycznymi metodami Monte Carlo. Przeanalizowano właściwości estymatorów wartości średniej i wariancji w przypadku próbek czysto losowych oraz generowanych przez łańciskie hiperkostki. Porównano efektywność różnych metod symulacyjnych dla szeregu analitycznych losowych funkcji testowych, a także dwóch problemów mechaniki. Wyniki testów pozwalają zaproponować regułę doboru wielkości próbki na potrzeby zadania optymalizacji odpornościowej. Dla kompletności prezentacji omówiono wady i zalety alternatywnych metod szacowania momentów statystycznych funkcji losowych.

W rozdziale 5 przedstawiono propozycję dwóch algorytmów wykorzystujących metody symulacji losowych w zadaniach analizy niezawodności z zaszumioną funkcją graniczną. Próbki losowe tworzone są najczęściej przy pomocy łańciskich hiperkostek. Pierwszy, dwuetapowy algorytm Z3M stanowi uniwersalne narzędzie do szacowania prawdopodobieństwa awarii konstrukcji reprezentowanych przy pomocy złożonych modeli skończenie elementowych. Bardzo często w takich przypadkach mamy do czynienia z nieróżniczkowalną funkcją awarii, co wynika z natury samego problemu lub z szumu numerycznego generowanego przez metodę rozwiązania zadania MES. Przeznaczeniem drugiego algorytmu jest szybka poprawa niezawodności konstrukcji, jedynie za pomocą symulacji OLH, bez użycia metod powierzchni odpowiedzi.

Przykłady zastosowania algorytmów omówionych w rozdziale 5 zaprezentowano w rozdziale 6. Dotyczy on analizy niezawodności i optymalizacji odpornościowej w zagadnieniach związanych z wytrzymałością zderzeniową. Zarówno funkcje awarii jak też funkcje tworzące sformułowanie zadania optymalizacji zawierają wpływ szumu. W rozdziale przedstawiono: przykład analizy niezawodności dynamicznie zgniatanej belki cienkościennej, gdzie uwzględnia się losowe awarie połączeń zgrzewanych, zadanie optymalizacji odpornościowej tej belki, jak również przykład zastosowania symulacyjnego algorytmu poprawy jej niezawodności. Ponadto, pokazano przykład jakościowej analizy zachowania zgniatanego elementu tylnej ramy pojazdu. Metoda ta wykorzystuje prosty algorytm klasyfikacji wyników eksperymentów numerycznych. Zaprezentowano również zastosowanie sformułowania optymalizacji odpornościowej w zadaniu kalibracji parametrów modelu numerycznego, używanego w komputerowej symulacji standardowych testów zderzeniowych.

Rozdział 7 poświęcony jest w całości analizie niezawodności procesów tłoczenia blach. Po krótkim wstępie dotyczącym komputerowego modelowania tych procesów, zaprezentowano definicję funkcji granicznej. Funkcja ta, bazująca na

analizie wykresu odkształceń granicznych, jest nieróżniczkowalna. Do oszacowania wartości prawdopodobieństwa awarii zaproponowano wykorzystanie adaptacyjnej metody Monte Carlo. W związku z niemożnością precyzyjnego wyznaczenia krzywej odkształceń granicznych, często nie jest możliwe jednoznaczne określenie stanu awarii procesu tłoczenia. Trudności te były powodem wprowadzenia koncepcji rozmytego prawdopodobieństwa awarii. Proces analizy niezawodności zilustrowano na testowym przykładzie symulacji tłoczenia z konferencji NUMI-SHEET'93.

W rozdziale 8 rozważane są problemy dotyczące efektywnej komputerowej implementacji algorytmów analizy niezawodności oraz optymalizacji konstrukcji. Poszczególne rozwiązania omawiane są na przykładzie programu STAND. Obok obiektowo zorientowanej architektury kodu, szczególnie szeroko przedstawione zostały metody łączenia programu z zewnętrznymi modułami obliczeniowymi. Omówiono przykłady wykorzystania w analizie niezawodności komercyjnych pakietów metody elementów skończonych ABAQUS i RADIOSS. Dużo uwagi poświęcono również rozwiązaniom mającym na celu stworzenie funkcjonalnego interfejsu użytkownika. Program STAND używany był do realizacji większości zaprezentowanych w pracy przykładów obliczeniowych.

Dodatek A stanowi wybór zagadnień dotyczących komputerowej analizy niezawodności konstrukcji. Znajomość tej tematyki jest niezbędna do dobrego rozumienia metod oraz przykładów numerycznych przedstawionych w głównej części niniejszego opracowania. Obok ogólnego sformułowania zadania analizy niezawodności, znaleźć tam można omówienie metod transformacji zmiennych losowych, prezentację klasycznych metod poszukiwania punktu projektowego oraz przegląd popularnych metod symulacji losowych.

Kluczowym elementem wielu omawianych w pracy algorytmów jest plan eksperymentów generowany przez optymalną hiperkostkę łacińską - OLH. W dodatku B obszernie zaprezentowano koncepcję OLH oraz metody znalezienia optymalnego rozkładu punktów łacińskiej hiperkostki. Przedstawiono metodę systematycznego przeglądu możliwych położeń punktów oraz metodę wykorzystującą specjalnie zaprojektowany algorytm genetyczny. Określono złożoność obliczeniową tych metod. Przeprowadzono szereg testów w celu odpowiedniego doboru metody oraz jej parametrów w zależności od wielkości próbki i liczby zmiennych losowych. Zaprezentowano również sposób uwzględnienia korelacji między zmiennymi.

Dodatek C poświęcono wybranym metodom dyskretyzacji pól losowych. Chociaż, za wyjątkiem jednego przykładu w rozdziale 4, modele stochastyczne rozpatrywanych problemów składały się jedynie ze zmiennych losowych, to

znajomość podstaw teoretycznych teorii dyskretyzacji pól losowych jest bardzo pomocna w zrozumieniu wyprowadzenia równań metody krigingu. To właśnie zaprezentowana dyskretyzacja metodą optymalnej aproksymacji liniowej jest podstawą tej techniki aproksymacji.

Dobre zrozumienie treści rozdziałów wymaga znajomości zagadnień dotyczących metod komputerowej analizy niezawodności oraz koncepcji łańciskowej hiperkostki. Jeśli czytelnikowi tematyka ta nie jest dostatecznie znana, sugeruje się rozpoczęcie lektury od dodatków A i B.

1.3. Przegląd literatury

1.3.1. Analiza niezawodności w zagadnieniach z niejawną funkcją graniczną

Za pierwszy ważny krok w kierunku współczesnych metod komputerowej analizy niezawodności konstrukcji należy uznać pracę Hasofera i Linda [86] z 1974 roku. Zawarto w niej koncepcję lokalizacji tzw. punktu projektowego, to jest takiej realizacji zmiennych losowych z obszaru awarii, której odpowiada największa wartość funkcji gęstości prawdopodobieństwa. Linearyzacja funkcji granicznej w punkcie projektowym pozwalała otrzymać miarę niezawodności, która jest niezmiennicza ze względu na równoważne sformułowania warunku granicznego. Jest to tzw. wskaźnik niezawodności Hasofera-Linda. Ideę wskaźnika Hasofera-Linda wykorzystali w 1978 roku Rackwitz i Fiessler w pracy [200]. Zastosowali oni ponadto transformację niezależnych zmiennych losowych o dowolnych rozkładach prawdopodobieństwa do standardowych zmiennych normalnych oraz przedstawili iteracyjny algorytm poszukiwania punktu projektowego. W pracy [93] Hohenbichler i Rackwitz zaproponowali wykorzystanie w analizie niezawodności transformacji Rosenblatta [209] w celu transformacji zależnych zmiennych losowych do przestrzeni standardowej. Transformacja Rosenblatta oraz, wykorzystana po raz pierwszy przez Der Kiureghiana i Liu w pracy [45], transformacja Natafa należą obecnie do najczęściej stosowanych metod transformacji. Aproksymacja funkcji granicznej w punkcie projektowym wielomianem pierwszego lub drugiego stopnia prowadzi odpowiednio do metod analizy niezawodności pierwszego rzędu (FORM) lub drugiego rzędu (SORM), zob. np. [48, 143, 173]. Chociaż w porównaniu z SORM, czas obliczeń numerycznych związany z metodą FORM jest dużo krótszy, to w wielu przypadkach jest on nadal zbyt długi z punktu widzenia praktyki projektowej. Jest tak zazwyczaj, jeśli obliczenie wartości funkcji granicznej wiąże się z czasochłonną, nieliniową analizą za pomocą programu elementów skończonych lub jeśli gradienty tej

funkcji obliczane są różnicowo. Niestety, w wielu przypadkach komputerowej analizy niezawodności złożonych konstrukcji i procesów technologicznych, charakter funkcji granicznej uniemożliwia lokalizację punktu projektowego za pomocą efektywnych metod gradientowych.

Odrębną grupę metod analizy niezawodności stanowią metody symulacyjne. Do ich zalet należą niewątpliwie: łatwość implementacji, możliwość uzyskania wyników z dowolną dokładnością oraz niewrażliwość na nieróżniczkowalność funkcji granicznej czy też na istnienie wielu punktów projektowych. Najlepiej znana, klasyczna metoda Monte Carlo jest jednak bardzo rzadko stosowana w obliczeniach niezawodnościowych. Przy bardzo małych wartościach prawdopodobieństwa awarii rzeczywistych konstrukcji, liczba koniecznych symulacji dyskwalifikuje przydatność tej metody w praktycznych zastosowaniach. Lata osiemdziesiąte ubiegłego wieku przyniosły jednak rozwój metod redukcji wariancji, ang. *importance sampling*. Poprzez odpowiedni dobór funkcji gęstości prawdopodobieństwa, według której generuje się zmienne losowe, można znacznie zmniejszyć obszar próbkowania oraz liczbę symulacji. Metoda ta ma wiele odmian. Stosowana jest też często w połączeniu z FORM lub SORM. Spośród prekursorskich prac jej poświęconych, warto wymienić artykuły Schuëllera i Stixa [217], Hohenbichlera i Rackwitz [94] oraz Dolińskiego [51]. Przegląd oraz omówienie różnych metod symulacyjnych stosowanych w analizie niezawodności znaleźć można w pracach Melchersa [172, 173].

We wczesnych zastosowaniach metod FORM i SORM przyjmowano najczęściej, że funkcja graniczna jest jawną funkcją zmiennych losowych. Niestety, za wyjątkiem akademickich przykładów, takiej zależności funkcyjnej nie można podać, a w większości przypadków dana jest ona za pośrednictwem określonej procedury numerycznej, np. metody elementów skończonych. Poczynając jednak od pierwszych prac Der Kiureghiana oraz współpracowników [42, 43], w połowie lat osiemdziesiątych problematyka wykorzystania metody elementów skończonych w ramach metod analizy niezawodności zaczęła być bardzo intensywnie rozwijana. Jej rozwój wiąże się m.in. z rozwojem metod analizy wrażliwości. Można wymienić wiele prac z ostatniego dziesięciolecia ubiegłego wieku, gdzie bardzo silnie podkreślano konieczność implementacji efektywnych analitycznych metod analizy wrażliwości w zadaniu szacowania prawdopodobieństwa awarii konstrukcji modelowanych z użyciem metody elementów skończonych. Są to m.in. prace amerykańskich grup badawczych z uniwersytetów w San Antonio [205], Berkeley [154, 245, 276], Boulder [49], a także uniwersytetu w Delft w Holandii [83] czy Clermont-Ferrand we Francji [144, 145].

Przy dużej liczbie zmiennych losowych, zagadnienie efektywnego i dokładnego obliczania gradientów funkcji granicznej ma kluczowe znaczenie z punktu widzenia numerycznej efektywności analizy niezawodności. Niestety, podejście bazujące na dostępności pochodnych funkcji granicznej po wszystkich zmiennych losowych, w sposób zasadniczy ogranicza zakres możliwych zastosowań. Przy tworzeniu programu umożliwiającego analizę niezawodności z wykorzystaniem metody elementów skończonych, najkorzystniejszą sytuacją jest posiadanie dostępu zarówno do kodu źródłowego MES, jak i programu niezawodnościowego. Oczywiście, szczególnie w przypadku dużych komercyjnych pakietów analizy skończonej elementowej, jest to warunek nie do spełnienia. Jeśli przyjąć jednak, że taki dostęp jest możliwy, to implementacja analitycznych procedur analizy wrażliwości odpowiedzi konstrukcji względem wszelkich interesujących projektanta parametrów modelu jest zadaniem niezmiernie czasochłonnym i wymagającym bardzo wysokich kompetencji. Szczególnie dużym wyzwaniem jest implementacja metod analizy wrażliwości dla zadań mechaniki nieliniowej oraz gdy rozpatrywane parametry dotyczą kształtu konstrukcji.

Sposobem na przewyżczenie tych trudności jest zastąpienie rzeczywistej funkcji granicznej analitycznym modelem (modelem modelu) zwanym też powierzchnią odpowiedzi. W oparciu o odpowiednio dobrany zbiór eksperymentów numerycznych, funkcja graniczna przybliżana jest za pomocą funkcji wielomianowej lub sieci neuronowej. Skonstruowana w ten sposób zależność stanowi jawną, a przez to tanią w obliczeniach, funkcję zmiennych losowych, co znakomicie ułatwia zastosowanie metod pierwszego lub drugiego rzędu czy też metod symulacyjnych (najczęściej są to metody typu ‘importance sampling’). Jednak, aby niezawodność konstrukcji obliczona dla danej powierzchni odpowiedzi nie była obciążona zbyt dużym błędem, powierzchnia ta musi dobrze aproksymować rzeczywistą funkcję graniczną w okolicy punktu projektowego. Położenie tego punktu niestety nie jest zazwyczaj znane.

Wśród pierwszych prac dotyczących metod tworzenia powierzchni odpowiedzi oraz jej zastosowań w analizie niezawodności należy wymienić prace Faraveli [68], Buchera i Bourgunda [20], Rajashekhara i Ellingwooda [202], El-Tawila i in. [62] czy Gopalakrishny i Donaldsona [80]. W większości tych prac proponowano aproksymację z użyciem wielomianów drugiego stopnia dopasowywanych do punktów eksperymentalnych za pomocą liniowej regresji. W ostatnim dziesięcioleciu obserwować można rozwój metod, które wykorzystują metodę ważonej liniowej regresji [117, 128] oraz sztuczne sieci neuronowe [39, 103] do aproksymacji funkcji granicznej.

Pomimo tak wielu opracowań nie została zaproponowana uniwersalna metoda, a więc i oprogramowanie, które łączyłoby szybkość obliczeń z dokładnością uzyskanych wyników. Często sformułowanie rozważanego zagadnienia determinuje przyjęte w analizie niezawodności rozwiązanie. W pracy Stockiego, Tauzowskiego i Knabla [238], na przykładzie zadania analizy niezawodności zgniatanej dynamicznie belki cienkościennej, przedstawiono koncepcję dwuetapowej metody analizy niezawodności wykorzystującej algorytm Z3M z pracy [278]. Oryginalny algorytm zmodyfikowano używając w etapie poszukiwania punktu projektowego adaptacyjnej metody powierzchni odpowiedzi. Powierzchnia ta budowana była na bazie eksperymentów OLH, przy zastosowaniu techniki aproksymacji lokalnej (ang. *moving least squares*). Zalety przedstawionego podejścia omówione zostaną szeroko w niniejszej pracy.

Na koniec należy wspomnieć, że w Polsce problemami analizy niezawodności konstrukcji zajmowali się m.in. Murzewski [176, 177], Śniady [223, 224], Jendo i Putresza [107, 199] i Biegus [14].

1.3.2. Optymalizacja odpornościowa

Koncepcja projektowania konstrukcji i procesów technologicznych, które są odporne na losowe imperfekcje parametrów, przypisywana jest często Genichi Taguchi'emu - japońskiemu inżynierowi i badaczowi zagadnień poprawy jakości. Chociaż stworzona przez niego w latach pięćdziesiątych, a szeroko rozpropagowana w latach osiemdziesiątych ubiegłego stulecia, metodologia inżynierii jakości (ang. *robust design*) nie jest bezpośrednio stosowana w niniejszej pracy, to ze względu na jej historyczne znaczenie warto tu przytoczyć podstawowe informacje na temat metody Taguchi'ego. W ostatnich dekadach metodologia ta znalazła zastosowanie w wielu dziedzinach projektowania inżynierskiego, prowadząc do redukcji liczby wykonywanych eksperymentów fizycznych, zob. [247].

Metoda Taguchi'ego

Przewodnią ideą klasycznej metody Taguchi'ego jest doprowadzenie do poprawy jakości produktów lub całych procesów przemysłowych nie tylko poprzez zapewnienie osiągnięcia wymaganych wartości nominalnych przez opisujące je charakterystyki, ale także poprzez minimalizację wariacji tych wartości. Minimalizacja wariacji realizowana jest bez eliminowania przyczyn obserwowanych rozrzutów.

Taguchi wyodrębnił trzy grupy parametrów. Są to: **parametry sterowalne** (ang. *control parameters*), których wartości mogą być zmieniane w czasie projektowania lub w procesie optymalizacji, **parametry zakłócające** (ang. *noise parameters*) - niemożliwe do kontroli i będące źródłem wariacji odpowiednich charakterystyk oraz **parametry niesterowalne** - odpowiadające ustalonym parametrom deterministycznym.

Celem zaproponowanej metody jest takie dobranie wartości parametrów sterowalnych, aby uczynić charakterystyki produktu/procesu jak najmniej czułymi na wariacje parametrów zakłócających. Metoda ta bazuje na ustaleniu - zarówno dla parametrów sterowalnych, jak i zakłócających - pewnej liczby poziomów ich wartości, zawartych w dopuszczalnych (w przypadku parametrów sterowalnych) i obserwowanych (w przypadku parametrów zakłócających) zakresach zmienności. Kombinacje poziomów parametrów sterowalnych tworzą tzw. **wewnętrzną tablicę ortogonalną** (ang. *inner array*), natomiast kombinacje poziomów parametrów zakłócających tworzą **zewnętrzną tablicę ortogonalną** (ang. *outer array*). Definiuje się również **funkcję strat jakości**, która określa straty ponoszone przez producenta lub przez społeczeństwo, wywołane odmiennymi od założonych cechami szeroko rozumianego produktu.

Inną wprowadzoną przez Taguchi'ego miarą jakości projektu jest **stosunek sygnału do zakłócenia** S/N (ang. *signal to noise ratio*). Wyraża on przyjętą zasadę mierzenia jakości, jaką jest minimalizacja zmienności działania produktu w odpowiedzi na czynniki zakłócające, przy jednoczesnej maksymalizacji zmienności w odpowiedzi na czynniki sterowalne. Stosunek S/N zdefiniowany jest jako pewna statystyka wartości wybranej charakterystyki, obliczona dla każdej kombinacji parametrów sterowalnych z wewnętrznej tablicy ortogonalnej przy próbkach generowanych przez kombinacje poziomów parametrów zakłócających dane w tablicy zewnętrznej. Rozróżnia się trzy główne typy współczynników S/N : dla charakterystyk o pożądanej najmniejszej wartości, pożądanej największej wartości i pożądanej nominalnej wartości. Maksymalizując współczynnik S/N minimalizuje się jednocześnie funkcję strat. Im większa jest wartość S/N tym większa jest też odporność na zakłócenia.

Podsumowując, procedura inżynierii jakości Taguchi'ego składa się z następujących etapów:

1. Sformułowanie problemu i określenie funkcji kryterium/charakterystyki.
2. Ustalenie parametrów sterowalnych i zakłócających oraz wybór liczby i wartości ich poziomów.
3. Budowa tablic ortogonalnych: wewnętrznej i zewnętrznej.

4. Przeprowadzanie zaplanowanych eksperymentów, wybór optymalnych wartości poziomów parametrów sterowalnych.
5. Weryfikacja otrzymanych wyników.
6. Jeśli projekt optymalny nie spełnia założeń następuje powrót do etapu 2. Jeśli założenia są spełnione, to projekt zostaje przyjęty.

Pomimo, iż metoda Taguchi'ego zyskała dużą popularność wśród projektantów, to jest też dość często przedmiotem krytyki, zob. [180]. Wśród jej głównych niedoskonałości, wymienić należy:

- Ograniczenie możliwych rozwiązań do arbitralnie wybranych poziomów parametrów sterowalnych, gdy tymczasem zmienne projektowe zadania optymalizacji mogą przyjmować wartości z ciągłego zakresu.
- Metoda dostosowana jest do eksperymentów fizycznych, których ze względu na koszt nie można przeprowadzić zbyt wiele. Ponieważ analiza pracy projektowanej konstrukcji wykonywana jest zazwyczaj metodą elementów skończonych, dlatego korzystniejsze jest użycie metod pozwalających na dokładniejsze zbadanie przestrzeni możliwych rozwiązań.
- Ze względu na to, iż odpowiedzi analizowanych układów konstrukcyjnych i procesów technologicznych są najczęściej nieliniowymi funkcjami ich parametrów, zaproponowane przez Taguchi'ego tablice ortogonalne mogą okazać się niewystarczającym narzędziem do zbadania tych zależności.
- Informacja probabilistyczna dotycząca parametrów zakłócających jest często niewykorzystywana. Używa się jej jedynie do ustalenia przedziałów zmienności tych parametrów. Nie jest możliwe uwzględnienie wiedzy na temat wzajemnej korelacji parametrów zakłócających.
- Brak możliwości uwzględnienia funkcji ograniczeń projektowych.

Strategia Taguchi'ego nie jest w istocie strategią optymalizacyjną, mimo to jej elementy wykorzystywane są w wielu pracach dotyczących optymalizacji odpornościowej konstrukcji, zob. [29, 104, 149].

Poniżej przedstawiony zostanie wybór tych prac, w których poprawa jakości/odporności konstrukcji uzyskana zostaje przy użyciu metod optymalizacji. Doskonały przegląd obecnego stanu wiedzy w tej dziedzinie zawarto w pracy Beyera i Sendhoffa [13]. Warto wspomnieć również przeglądową pracę Zanga i in. [272].

Podejście optymalizacyjne

W odróżnieniu od dobrze zdefiniowanego sformułowania zadania optymalizacji niezawodnościowej (zob. podrozdział 2.3), gdzie centralne miejsce zajmuje prawdopodobieństwo niespełnienia ograniczeń, spotykane sformułowania optymalizacji odpornościowej często różnią się stosowaną miarą odporności konstrukcji na imperfekcje. Obok wprowadzonego przez Taguchi'ego stosunku sygnału do zakłócenia S/N , jako miary odporności przyjmuje się wariancje charakterystycznych odpowiedzi konstrukcji bądź różnice wybranych percentyli tych odpowiedzi. Prezentowane w literaturze zadania optymalizacji odpornościowej różnią się także sposobem sformułowania ograniczeń, metodami obliczania wariancji, typem stosowanej aproksymacji i wreszcie samą metodą optymalizacji.

Jeśli chodzi o metody szacowania momentów statystycznych odpowiedzi konstrukcji, występujących zarówno w funkcji celu, jak i ograniczeniach zadania optymalizacji odpornościowej, to w dostępnych autorowi publikacjach dominuje tzw. podejście perturbacyjne. Bazuje ono na rozwinięciu funkcji wokół wartości średnich zmiennych losowych w szereg Taylora. Można wymienić wiele prac, gdzie zastosowano metodę perturbacyjną, zob. np. [28, 54, 108, 129, 138, 139, 203]. Dokładność takiego szacowania wartości oczekiwanych oraz wariancji funkcji celu i ograniczeń zależy zarówno od współczynników zmienności zmiennych losowych, stopnia nieliniowości badanej funkcji, jak i od dokładności obliczenia jej gradientów. Możliwość wykorzystania dokładnych, analitycznych metod różniczkowania odpowiedzi konstrukcji pozwala na bardzo znaczną poprawę efektywności procesu optymalizacji odpornościowej. Doltsinis i Kang w serii publikacji [54–56, 115] zaprezentowali korzyści płynące ze stosowania metody bezpośredniego różniczkowania w zadaniach optymalizacji odpornościowej dotyczących liniowych, jak również fizycznie nieliniowych konstrukcji i procesów technologicznych. Przedstawione przez nich przykłady dotyczyły konstrukcji prętowych, ramowych oraz procesów obróbki plastycznej - wyciskania i kucia metali. Wady i zalety metod szacowania momentów statystycznych bazujących na rozwinięciu funkcji w szereg Taylora omówiono w podrozdziałach 4.2 i 4.3 Wyjaśniono tam również dlaczego postanowiono nie wykorzystywać tych metod w niniejszej pracy.

Alternatywną metodą obliczenia momentów jest metoda symulacji losowych. Chociaż nieumiejętnie stosowane techniki symulacyjne mogą prowadzić do znacznego wzrostu zarówno czasu jak i kosztu rozwiązania, to zalety tych metod sprawiają, iż są one również chętnie używane w ramach optymalizacji odpornościowej, zob. np. [114, 118, 191, 212].

W większości prac dotyczących odpornościowej optymalizacji konstrukcji, w celu redukcji złożoności numerycznej stosuje się aproksymacje tych funkcji wchodzących w skład sformułowania zadania, które wymagają czasochłonnych obliczeń, np. wymagają użycia metody elementów skończonych. Wykorzystuje się zarówno wielomianowe powierzchnie odpowiedzi [28, 29, 133], technikę kriginu [26, 139], jak też sztuczne sieci neuronowe [108, 118, 191]. Doskonałe omówienie różnych technik powierzchni odpowiedzi (różnych metamodeli), szczególnie pod kątem ich przydatności w optymalizacji odpornościowej, znaleźć można w pracy doktorskiej Floriana Jurecki z 2007 roku [113]. Jednoznacznie preferuje on metodę kriginu w stosunku do sieci neuronowych, które bardzo często okazują się być rozwiązaniem nieefektywnym. To właśnie, bazująca na teorii procesów losowych, metoda kriginu w wersji aproksymacyjnej użyta została przez autora niniejszej pracy jako metamodel dla wartości średnich i odchyłeń standardowych funkcji celu i ograniczeń, zob. również [236].

Prace dotyczące optymalizacji odpornościowej konstrukcji podzielić można również w zależności od sposobu wyrażenia tzw. ograniczeń odpornościowych. Ograniczenia te formułowane są zgodnie z podejściem perturbacyjnym

$$\begin{aligned} \mathbb{E}[g(\mathbf{d}, \mathbf{X}, \mathbf{P})] - \sum_{i=1}^{n_X} \left| \frac{\partial g(\mathbf{d}, \mathbf{x}, \mathbf{p})}{\partial x_i} \Big|_{\mathbf{d}, \mu_{\mathbf{X}}, \mu_{\mathbf{P}}} \Delta x_i \right| \\ - \sum_{i=1}^{n_P} \left| \frac{\partial g(\mathbf{d}, \mathbf{x}, \mathbf{p})}{\partial p_i} \Big|_{\mathbf{d}, \mu_{\mathbf{X}}, \mu_{\mathbf{P}}} \Delta p_i \right| \geq 0, \end{aligned} \quad (1.1)$$

lub w ramach “filozofii” DFSS (*design for six sigma*) jako

$$\mathbb{E}[g(\mathbf{d}, \mathbf{X}, \mathbf{P})] - \tilde{\beta} \sigma[g(\mathbf{d}, \mathbf{X}, \mathbf{P})] \geq 0, \quad (1.2)$$

gdzie $\mathbb{E}(\cdot)$ i $\sigma(\cdot)$ są odpowiednio operatorami wartości oczekiwanej i odchylenia standardowego, a wszystkie pozostałe symbole omówiono w rozdziale 2. To pierwsze sformułowanie, które spotkać można w wielu pracach [29, 30, 108, 138], wymaga znajomości gradientów funkcji ograniczeń, natomiast wielkości perturbacji Δx_i i Δp_i ustalane są jako wielokrotności odchyłeń standardowych odpowiednich zmiennych losowych. Sformułowanie (1.2), które stosowane jest także w niniejszej pracy, wydaje się być naturalnym odpowiednikiem ograniczeń probabilistycznych w niezawodnościowej optymalizacji konstrukcji, zob. [129, 133, 212, 218].

Zadanie optymalizacji odpornościowej jest zadaniem optymalizacji wielokryterialnej. O ile zazwyczaj autorzy prac nie poświęcają temu problemowi zbyt

wiele uwagi, koncentrując się na innych aspektach zagadnienia i używając najprostszej techniki skalaryzacji przy pomocy metody współczynników wagowych, o tyle w niektórych pracach temat poprawnego i efektywnego wyznaczenia zbioru rozwiązań kompromisowych analizowany jest dużo dokładniej. Warto tu wymienić prace grupy badawczej kierowanej przez prof. Wei Chen z University of Illinois w Chicago [30,273] dotyczące między innymi tzw. programowania fizycznego, jak również prace [118,218], gdzie zastosowano wielokryterialną wersję algorytmu genetycznego (MOGA).

Na koniec, należy wspomnieć o ciekawym artykule Kocho i in. [132], w którym dokładnie omówiono znaczenie, jakie dla efektywności algorytmów optymalizacji niezawodnościowej oraz odpornościowej ma realizacja zadania na maszynach o architekturze równoległej.

1.3.3. Metody powierzchni odpowiedzi

Termin “powierzchnia odpowiedzi” wykorzystywany jest w niniejszej pracy w szerokim sensie do określenia przybliżonej analitycznej zależności pomiędzy parametrami (zmiennymi) wejściowymi, a wielkościami wyjściowymi (wynikami). Jako powierzchnię odpowiedzi rozumiemy będziemy zależność aproksymującą otrzymaną zarówno przez dopasowywanie metodą najmniejszych kwadratów współczynników równań wielomianowych, jak również technikę krigingu, sztuczną sieć neuronową, metodę aproksymacji lokalnej czy aproksymację z wykorzystaniem radialnych funkcji bazowych. Przyjęta konwencja jest bardziej ogólna w stosunku do angielskiej terminologii, gdzie przez *response surface* rozumie się przede wszystkim wielomianowe funkcje odpowiedzi stworzone za pomocą metod regresji w oparciu o odpowiednio zaprojektowany plan eksperymentów, zob. [58] i [179] - klasyczne już monografie. Określenia *powierzchnia odpowiedzi*, używać będziemy zastępczo z określeniem *metamodel*.

Obszerne omówienie stosowanych w niniejszej pracy metod powierzchni odpowiedzi wraz z odnośnymi pozycjami literatury zamieszczono w rozdziale 3, natomiast publikacje dotyczące zastosowań w analizie niezawodności wymieniono wcześniej, w podrozdziale 1.3.1. Poniżej, w krótkim przeglądzie, wspomniane zostaną jedynie prace, które wydają się być istotne z punktu widzenia rozwoju tej dziedziny oraz jej zastosowań w optymalizacji odpornościowej.

W pracy Barthelemy’ego i Haftki [8] rozróżniono trzy typy aproksymacji. W zależności od obszaru jej ważności wymienia się tam aproksymację punktową, aproksymację w ograniczonym obszarze i aproksymację globalną. O ile

aproksymacja punktowa poprawnie przybliży oryginalną funkcję tylko w bezpośrednim sąsiedztwie badanego punktu, o tyle drugi typ aproksymacji jest dobrze zdefiniowany już w pewnym podobszarze przestrzeni zmiennych projektowych. Aproksymacja globalna ma za zadanie przybliżyć badaną funkcję w całej dziedzinie określoności zmiennych. Metody aproksymacji punktowej wykorzystują informacje na temat wartości i pochodnych funkcji w jednym punkcie, metody aproksymacji w ograniczonym obszarze bazują zaś na informacji o wartościach funkcji w wielu punktach. Obydwie te metody wykorzystywane są w iteracyjnych algorytmach optymalizacji, gdzie obszar ważności danego modelu (obszar zaufania) przesuwana się w obrębie granic wyznaczonych przez ograniczenia proste zmiennych projektowych. W przypadku gdy koszt obliczenia pojedynczej wartości funkcji celu jest zbyt wysoki i nie można pozwolić sobie na rozwiązanie wykorzystujące adaptacyjną powierzchnię odpowiedzi, wtedy stosuje się metody aproksymacji globalnej.

W dalszym ciągu dużą popularnością cieszą się wielomianowe powierzchnie odpowiedzi, których współczynniki ustalane są na podstawie zbioru eksperymentów, za pomocą liniowej regresji. Aby zminimalizować liczbę współczynników, których wartości trzeba wyznaczyć, przyjmuje się zazwyczaj wielomiany pierwszego lub drugiego stopnia. W celu poprawy jakości aproksymacji, stosuje się również transformację zmiennych, np. transformację logarytmiczną lub transformację typu $1/x$. Do wspomnianych w poprzednim podrozdziale prac z zakresu optymalizacji odpornościowej, wykorzystujących wielomianowe funkcje aproksymujące, dodać należy również cały szereg artykułów, gdzie używana jest tzw. strategia aproksymacji statystyk (ang. *dual response surface* - DRS), zob. [149, 248, 265].

W przeciwieństwie do eksperymentów fizycznych, eksperymenty komputerowe (wyniki komputerowych symulacji) nie są obciążone ani błędem pomiarowym, ani też losowym rozrzutem, charakterystycznym dla zmieniających się warunków doświadczenia. Założenie na temat obecności losowego błędu wyników doświadczenia jest podstawą liniowej regresji, która wykorzystuje metodę najmniejszych kwadratów. Dlatego też, metamodel zastępujący deterministyczną analizę komputerową powinien w istocie interpolować wyniki otrzymane w punktach eksperymentalnych. Właśnie takie podejście zaproponowali w 1989 roku Sacks i in. w pracy [210]. Przedstawiona przez nich metoda nazywana jest DACE (ang. *design and analysis of computer experiments*) lub też kriging od Daniela Krige'a, który podobną metodę stworzył na potrzeby analizy geostatystycznej [36].

W analizie porównawczej przeprowadzonej w [110] testowano przydatność poszczególnych rodzajów aproksymacji w optymalizacji odpornościowej. Porównywano klasyczną aproksymację wielomianową, metodę radialnych funkcji bazowych i kriging. Testowano dokładność tak zbudowanych metamodeli oraz ich zdolność do odwzorowywania wariancji oryginalnych funkcji. Technika krigingu dawała najlepsze rezultaty zarówno w przypadku słabo, jak i silnie nieliniowych funkcji. Zauważono jednak także, że jest ona wrażliwa na szum, którym obarczone mogą być dane eksperymentalne. Mając to na uwadze oraz ze względu na stosowane sformułowanie optymalizacji odpornościowej, w niniejszej pracy zastosowano “wygładzoną” wersję krigingu, nie jako funkcji interpolującej lecz aproksymującej.

Warto również jeszcze raz wymienić pracę Jurecki [113], gdzie oprócz analizy różnych typów metamodeli, znaleźć można ciekawy przegląd zastosowań optymalizacji konstrukcji, w których wykorzystywana była koncepcja powierzchni odpowiedzi. Wśród innych interesujących prac przeglądowych dotyczących metamodeli na uwagę zasługują artykuł Jin i in. [109] oraz seria prac Simpsona i współpracowników [220–222].

Optymalizacja konstrukcji

o parametrach losowych

W rozdziale omówione zostaną trzy różne sformułowania oraz wybrane metody rozwiązania zadania optymalizacji konstrukcji. Pierwsze sformułowanie dotyczy klasycznego zadania optymalizacji skalarnej (jednokryterialnej), w którym zakłada się, że wszystkie zmienne i parametry opisujące model analizowanego układu konstrukcyjnego mają charakter deterministyczny. Dwa pozostałe, pozwalające na uwzględnienie niedeterministycznej natury parametrów konstrukcji, opisują zadania tzw. optymalizacji niezawodnościowej oraz optymalizacji odpornościowej. Ze względu na związek ze sformułowaniem problemu optymalizacji odpornościowej, krótki fragment tekstu poświęcony zostanie również deterministycznej optymalizacji wielokryterialnej (wektorowej).

Pomimo iż opis niektórych z wymienionych zadań optymalizacji wymaga wprowadzenia specjalnej terminologii, to wiele pojęć jest wspólnych dla wszystkich sformułowań. I tak, przez *zmienne projektowe/decyzyjne* rozumiemy będziemy te parametry konstrukcji, których wartości ustala się w procesie optymalizacji w celu osiągnięcia oczekiwanej charakterystyki jej pracy. Funkcja, która służy do oceny danego rozwiązania nazywana jest najczęściej *funkcją celu* lub *funkcją kryterium*. Optymalizowana konstrukcja spełniać musi liczne *ograniczenia projektowe*, które definiują *obszar rozwiązań dopuszczalnych* \mathcal{D} . W przypadku większości nieakademickich zadań optymalizacji zarówno funkcje celu, jak i funkcje ograniczeń dane są w sposób niejawny, a obliczenie ich wartości wymaga przeprowadzenia oddzielnej analizy MES. Ograniczeniami, które definiowane są w sposób jawny są tzw. *ograniczenia proste/brzegowe*. Określają one zakres zmienności poszczególnych zmiennych projektowych.

Istnieje obecnie kilka dominujących metod opisu niepewności parametrów układów konstrukcyjnych. Wymienić tu należy przede wszystkim podejście probabilistyczne, gdzie niepewności te modeluje się na gruncie teorii prawdopodobieństwa przy pomocy zmiennych losowych oraz procesów losowych. Jako

alternatywne podejścia, stosowane często do modelowania niepewności w zagadnieniach sztucznej inteligencji, wspomnieć trzeba koncepcję zbioru rozmytego [271] oraz przedziału [27,85]. W niniejszej pracy przyjęto klasyczne, a zarazem najbardziej popularne w analizie konstrukcji inżynierskich, podejście probabilistyczne.

2.1. Optymalizacja deterministyczna

Typowe sformułowanie zadania skalarnej optymalizacji deterministycznej wyrazić można następująco:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \quad (2.1)$$

$$\text{minimalizujące: } f(\mathbf{d}), \quad (2.2)$$

$$\text{przy ograniczeniach: } g_i(\mathbf{d}) \geq 0, \quad i = 1, \dots, k_g, \quad (2.3)$$

$$c_i(\mathbf{d}) = 0, \quad i = 1, \dots, k_c, \quad (2.4)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n, \quad (2.5)$$

gdzie $\mathbf{d} \in \mathbb{R}^n$ jest wektorem zmiennych projektowych, f jest funkcją celu, g_i , $i = 1, \dots, k_g$, są ograniczeniami nierównościami, c_i , $i = 1, \dots, k_c$, są ograniczeniami typu równościowego, a ${}^l d_j$ i ${}^u d_j$, $j = 1, \dots, n$, stanowią odpowiednio dolne i górne ograniczenia proste. W n -wymiarowej przestrzeni zmiennych projektowych ograniczenia przedstawiają hiperpowierzchnie zawierające punkty spełniające te ograniczenia w postaci równości. W powyższym sformułowaniu zarówno zmienne projektowe, jak również wszystkie parametry definiujące model konstrukcji oraz funkcje celu i ograniczeń, mają charakter deterministyczny, tzn. reprezentowane są przez jedną, nominalną wartość.

Istnieje bardzo bogata literatura dotycząca teorii optymalizacji, jak też metod rozwiązywania różnorodnych zagadnień optymalizacji. Wśród znanych monografii poświęconych tej tematyce wspomnieć można prace Arory [6], Haftki i Gürdala [84], Gilla i in. [75] czy Parka [192].

Dominującymi metodami rozwiązania zadania (2.1)–(2.5) są metody programowania matematycznego, liniowego lub nieliniowego – w zależności od postaci funkcji celu i ograniczeń. Rozwiązanie optymalne poszukiwane jest najczęściej w sposób iteracyjny, a spośród najpopularniejszych algorytmów wymienić należy algorytmy gradientowe, takie jak metoda gradientu sprzężonego, metoda sekwencyjnego programowania kwadratowego czy metoda sekwencyjnego programowania liniowego. Porównania różnych metod programowania matematycznego dokonano m.in. w pracy [216].

Ponadto, ze względu na szybko rosnącą moc obliczeniową współczesnych komputerów oraz dostępność rozwiązań bazujących na obliczeniach równoległych, coraz większą popularność zyskują bezgradientowe metody optymalizacji, które inspirowane są zjawiskami z dziedziny fizyki lub biologii. Przykładowo, sposób działania algorytmu symulowanego wyżarzania [121] przypomina proces wyżarzania w metalurgii, metoda roju cząstek [69] naśladuje zachowania społeczne osobników tworzących zorganizowane populacje, np. kolonii owadów lub ławicy ryb, a działanie algorytmu genetycznego [78] wzorowane jest na zjawisku ewolucji biologicznej. Niewątpliwą zaletą tych metod jest możliwość określenia optimum globalnego.

Wspomniane algorytmy bezgradientowe wymagają z reguły obliczenia wartości funkcji celu oraz ograniczeń dla bardzo dużej liczby realizacji zmiennych projektowych. Niestety, w przypadku gdy funkcje wchodzące w skład sformułowania optymalizacji są zdefiniowane w sposób niejawny (np. dane są przez rozwiązanie złożonego zadania MES), stanowi to może istotną przeszkodę w bezpośrednim stosowaniu tych metod do optymalizacji konstrukcji.

W celu przewyciężenia trudności związanych z wysokim kosztem numerycznej realizacji optymalizacji rzeczywistych konstrukcji inżynierskich, oryginalne funkcje celu i ograniczenia projektowe zastępowane są zazwyczaj przez pewne przybliżone modele - powierzchnie odpowiedzi (zob. np. [179] oraz rozdział 3). Dzięki jawnemu sformułowaniu problemu, rozwiązanie tak uproszczonego zadania optymalizacji jest możliwe do przeprowadzenia w krótkim czasie nawet przy użyciu metod bezgradientowych. Stosując powierzchnie odpowiedzi, niesłychanie istotna jest jednak weryfikacja wiarygodności otrzymywanych optimów, co prowadzi najczęściej do iteracyjnej procedury, którą bardzo ogólnie opisać można jako: 1) budowa aproksymacji funkcji celu i funkcji ograniczeń na podstawie ograniczonej liczby eksperymentów numerycznych, 2) rozwiązanie tak uproszczonego zadania optymalizacji, 3) weryfikacja otrzymanego rozwiązania, 4) sprawdzenie warunku zbieżności.

2.2. Optymalizacja wielokryterialna - rozwiązanie Pareto

Zasadniczą cechą odróżniającą optymalizację wielokryterialną, zwaną także optymalizacją wektorową lub polioptymalizacją, od optymalizacji z jedną funkcją celu (skalarnej) jest konieczność uwzględnienia wielu kryteriów optymalności.

Problem optymalizacji wektorowej sformułować można jako

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \quad (2.6)$$

$$\text{minimalizujące: } \mathbf{f}(\mathbf{d}) = \{f_1(\mathbf{d}), \dots, f_m(\mathbf{d})\}, \quad (2.7)$$

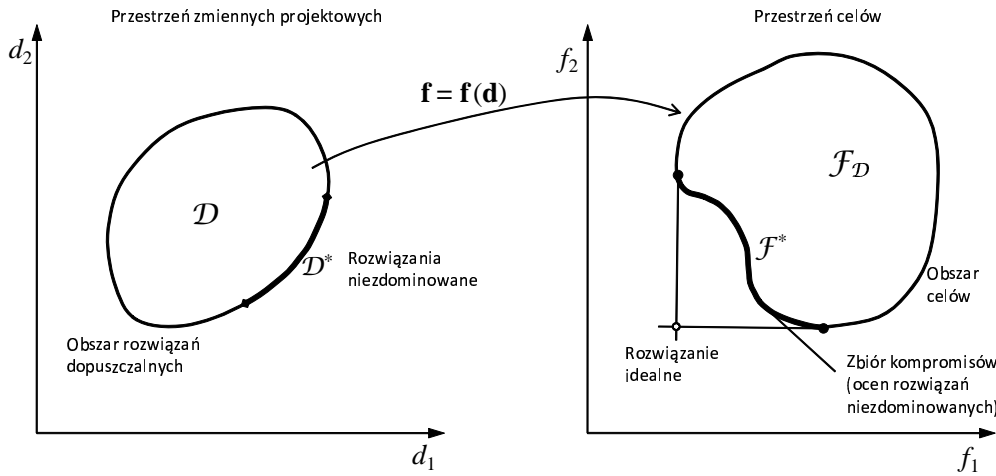
$$\text{przy ograniczeniach: } g_i(\mathbf{d}) \geq 0, \quad i = 1, \dots, k_g, \quad (2.8)$$

$$c_i(\mathbf{d}) = 0, \quad i = 1, \dots, k_c, \quad (2.9)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n. \quad (2.10)$$

Mamy zatem do czynienia z wektorem funkcji celu $\mathbf{f}(\mathbf{d})$, którego realizacje tworzą m -wymiarową przestrzeń celów \mathcal{F} . Rozwiązując tak postawiony problem optymalizacji dąży się więc do zminimalizowania “najlepiej jak to możliwe” wartości różnych funkcji kryteriów. Zazwyczaj kryteria te są wzajemnie konfliktowe, tzn. zmniejszenie wartości jednego z nich prowadzi do wzrostu wartości drugiego. Z tej też przyczyny rozwiązanie problemu optymalizacji wielokryterialnej prawie nigdy nie jest jednoznaczne. Niemożliwym jest znalezienie unikalnej kombinacji wartości zmiennych projektowych minimalizującej wszystkie funkcje celu. Otrzymuje się przeważnie zbiór rozwiązań, które minimalizując pewne kryteria, prowadzą do zwiększenia wartości innych. Rozwiązania takie określane są mianem rozwiązań Pareto, zwanych też rozwiązaniami niezdominowanymi (zob. [35, 166, 187]). Na rys. 2.1 w sposób schematyczny, dla dwóch zmiennych projektowych i dwóch funkcji celu, zademonstrowano wszystkie podstawowe pojęcia optymalizacji wielokryterialnej. Jako $\mathcal{F}_{\mathcal{D}}$ oznaczono obszar celów, będący tą częścią przestrzeni celów, która stanowi odwzorowanie obszaru rozwiązań dopuszczalnych \mathcal{D} z przestrzeni zmiennych projektowych w przestrzeń celów. Na zbiorze dopuszczalnym \mathcal{D} przez \mathcal{D}^* oznaczono rozwiązania niezdominowane. Wektor zmiennych projektowych \mathbf{d}' jest rozwiązaniem niezdominowanym, jeśli nie można znaleźć żadnego wektora $\mathbf{d} \in \mathcal{D}$ i $\mathbf{d} \neq \mathbf{d}'$, który nie prowadzi do zwiększenia wartości chociaż jednego kryterium w stosunku do rozwiązania odpowiadającego \mathbf{d}' . W przestrzeni celów zaznaczono również tzw. rozwiązanie idealne. Odpowiada ono minimalnym wartościom wszystkich funkcji celu, ale zazwyczaj ze względu na konfliktowy charakter kryteriów jest rozwiązaniem czysto hipotetycznym, niemożliwym do otrzymania. W niektórych metodach rozwiązanie idealne służy do wyboru rozwiązania preferowanego z tzw. zbioru kompromisów. Rozwiązania kompromisowe (oznaczone przez $\mathcal{F}_{\mathcal{D}}$ na rys. 2.1) są zbiorem ocen rozwiązań niezdominowanych. Spośród nich, na podstawie dodatkowego kryterium wybiera się rozwiązanie preferowane.

Istnieje wiele metod wyznaczania zbioru rozwiązań niezdominowanych i odpowiadających im kompromisów. Pomimo swoich niedoskonałości, ciągle dużą popularnością (przede wszystkim ze względu na łatwość implementacji) cieszą



Rys. 2.1. Przekształcenie obszaru rozwiązań dopuszczalnych \mathcal{D} z przestrzeni zmiennych projektowych w obszar celów $\mathcal{F}_{\mathcal{D}}$.

się tzw. metody skalaryzacji. Polegają one na zamianie zadania optymalizacji wektorowej na szereg zadań optymalizacji z jedną funkcją celu, które rozwiązywane być mogą przez standardowe algorytmy optymalizacji skalarnej. Wymienić tu można metodę współczynników wagowych (zwaną też metodą ważonych kryteriów), metodę mini-max, metodę programowania celów czy też metodę ε -ograniczeń, zob. [35, 167]. I tak, idea metody współczynników wagowych polega na utworzeniu zastępczej funkcji celu w postaci sumy iloczynu poszczególnych kryteriów i współczynników wagowych

$$\tilde{f}(\mathbf{d}) = \sum_{i=1}^m w_i f_i(\mathbf{d}), \quad (2.11)$$

gdzie

$$\sum_{i=1}^m w_i = 1, \quad w_i \geq 0. \quad (2.12)$$

Rozwiązując zadanie optymalizacji skalarnej z funkcją celu daną równaniem (2.11) przy ograniczeniach (2.8)–(2.10) dla różnych wartości wag w_i , otrzymuje się kolejne punkty zbioru kompromisów. W praktyce stosuje się najczęściej znormalizowaną wersję metody ważonych kryteriów przyjmując

$$\tilde{f}(\mathbf{d}) = \sum_{i=1}^m w_i \frac{f_i(\mathbf{d})}{f_i^0}, \quad (2.13)$$

gdzie f_i^0 , $i = 1, \dots, m$, są wielkościami odniesienia, np. największymi dopuszczalnymi wartościami funkcji celu. Inną możliwością jest normalizacja uwzględniająca zakresy zmian wartości poszczególnych kryteriów (zob. [191])

$$\tilde{f}(\mathbf{d}) = \sum_{i=1}^m w_i \frac{f_i(\mathbf{d}) - \min f_i(\mathbf{d})}{\max f_i(\mathbf{d}) - \min f_i(\mathbf{d})}. \quad (2.14)$$

Wartości ilorazów występujących w równaniu (2.14) zmieniają się w przedziale $[0, 1]$. Wadą metody współczynników wagowych jest niemożność uzyskania pełnego zbioru rozwiązań kompromisowych w przypadku niewypukłych obszarów $\mathcal{F}_{\mathcal{D}}$.

Bardzo szeroką i dynamicznie rozwijają się klasę metod rozwiązywania zadań optymalizacji wielokryterialnej stanowią algorytmy ewolucyjne i heurystyczne, takie jak algorytm genetyczny czy algorytm symulowanego wyżarzania [35, 188].

2.3. Optymalizacja niezawodnościowa

W tradycyjnym ujęciu, losowy charakter zmiennych projektowych oraz innych parametrów wchodzących w skład sformułowania (2.1)–(2.5) uwzględniany był w optymalizacji deterministycznej za pomocą tzw. częściowych współczynników bezpieczeństwa. Współczynniki te, specyfikowane przez odpowiednie normy projektowe, kalibruje się zazwyczaj tak, aby można było je stosować w jak najszerszym zakresie zadań projektowych związanych z danym typem konstrukcji. Niestety, takie podejście prowadzi często do zbyt “konserwatywnych” i zachowawczych rozwiązań, ponieważ częściowe współczynniki bezpieczeństwa nie są w sposób bezpośredni związane z losowym rozrzutem wartości zmiennych projektowych, a co za tym idzie, konstrukcje optymalne nie zapewniają w sposób automatyczny założonego poziomu niezawodności, zob. [230].

W przypadku gdy zagwarantowanie odpowiednio wysokiego poziomu bezpieczeństwa jest jednym z najważniejszych wymagań stawianych projektowanej konstrukcji, problem optymalizacji formułowany jest jako zadanie optymalizacji niezawodnościowej (ang. *reliability-based design optimization* - RBDO). W ramach RBDO, ograniczenia projektowe formułowane są przy pomocy prawdopodobieństw wystąpienia awarii lub innych, równoważnych miar niezawodności. Pod pojęciem prawdopodobieństwa awarii rozumie się prawdopodobieństwo przekroczenia pewnych dopuszczalnych stanów dotyczących wytrzymałości lub użytkowania konstrukcji, które definiowane są przez odpowiednie funkcje graniczne (zob. dodatek A).

Mając swoje początki w latach osiemdziesiątych ubiegłego wieku, niezawodnościowa optymalizacja konstrukcji jest obecnie dobrze ugruntowaną dziedziną nauki z literaturą obejmującą setki publikacji w czasopismach naukowych oraz materiałach konferencyjnych. Szczególnie istotny wpływ na rozwój zarówno sformułowań, jak i metod rozwiązania problemów RBDO wywarły prace badaczy z duńskich uniwersytetów w Aalborgu i Lyngby [64–67, 159, 250], z politechniki w Monachium [134, 135, 239, 240], a ostatnio z uniwersytetu w Iowa [267–270] i Uniwersytetu Illinois w Chicago [59, 60, 266]. Pewne podsumowanie obecnego stanu wiedzy na temat optymalizacji niezawodnościowej stanowi wydana w 2007 roku monografia [252].

Ogólne sformułowanie zadania optymalizacji niezawodnościowej wyrazić można następująco:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \quad (2.15)$$

$$\text{minimalizujące: } f(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{P}}), \quad (2.16)$$

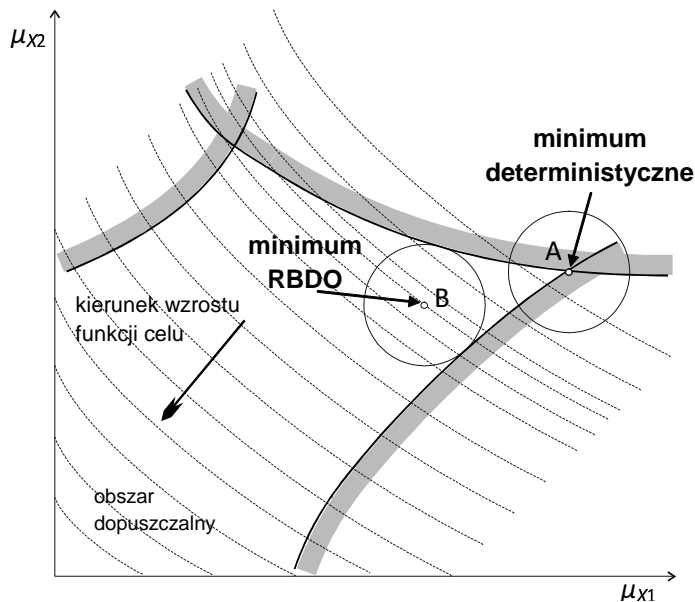
$$\text{przy ograniczeniach: } \mathbb{P}[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P}) \leq 0] - \Phi(-\beta_i^t) \leq 0, \quad i = 1, \dots, k_g, \quad (2.17)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n_d, \quad (2.18)$$

$${}^l \mu_{X_r} \leq \mu_{X_r} \leq {}^u \mu_{X_r}, \quad r = 1, \dots, n_X, \quad (2.19)$$

gdzie $\mathbf{X} = \{X_1, \dots, X_{n_X}\}$ oraz $\mathbf{P} = \{P_1, \dots, P_{n_P}\}$ są wektorami zmiennych losowych o wartościach oczekiwanych odpowiednio $\boldsymbol{\mu}_{\mathbf{X}}$ i $\boldsymbol{\mu}_{\mathbf{P}}$, $P_f^i = \mathbb{P}[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P}) \leq 0]$ jest prawdopodobieństwem awarii odpowiadającym i -tej funkcji granicznej $g_i(\cdot)$, $\Phi(\cdot)$ jest dystrybuantą standardowego rozkładu normalnego, a β_i^t , $i = 1, \dots, k_g$, są minimalnymi, ustalonymi przez projektanta wskaźnikami niezawodności, zob. (A.5). Wyrażenia $\Phi(-\beta_i^t)$ są zatem dopuszczalnymi maksymalnymi wartościami prawdopodobieństwa awarii. Rozróżnienie oznaczeń wektorów losowych wynika z ich odmiennego charakteru. Zmienne \mathbf{X} określić można jako projektowe zmienne losowe, ponieważ ich wartości oczekiwane $\boldsymbol{\mu}_{\mathbf{X}}$ zmieniają się w procesie optymalizacji, prowadząc (w przypadku stałych wartości innych parametrów opisujących rozkład) do przesunięcia funkcji gęstości prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{X})$. Inaczej niż w przypadku zmiennych \mathbf{X} , rozkład prawdopodobieństwa wektora \mathbf{P} nie zmienia się podczas optymalizacji i dlatego, w kontekście zadania optymalizacji niezawodnościowej, zmienne te nazywane są niekiedy parametrami losowymi. W przeciwieństwie do wartości oczekiwanych $\boldsymbol{\mu}_{\mathbf{X}}$, bardzo rzadko w pracach dotyczących RBDO uwzględnia się jako zmienne projektowe inne parametry rozkładów prawdopodobieństwa zmiennych losowych. Rozszerzenie zbioru zmiennych projektowych o odchylenia standardowe zmiennych losowych

(por. [52]) wymaga odpowiedniego sformułowania funkcji celu, które uwzględniłoby wzrost kosztu projektu związany z użyciem np. materiałów o lepszej jakości, bądź zastosowania bardziej precyzyjnych technologii wytworzenia lub montażu elementów optymalizowanej konstrukcji.



Rys. 2.2. Porównanie rozwiązania optymalnego z optymalizacji deterministycznej - punkt A oraz optymalizacji niezawodnościowej - punkt B.

Idea optymalizacji niezawodnościowej przedstawiona została schematycznie na rys. 2.2. W przypadku hipotetycznego problemu optymalizacji z dwiema zmiennymi projektowymi oraz trzema ograniczeniami, rozwiązaniem tego zadania w wersji deterministycznej jest punkt A. Pisząc o optymalizacji deterministycznej mamy na myśli takie sformułowanie, w którym w żaden sposób nie uwzględnia się losowości zmiennych projektowych. Tak jak to przedstawiono na rysunku, w znajdującym się na granicy obszaru dopuszczalnego punkcie optymalnym aktywne są dwa ograniczenia. Załóżmy następnie, że zmienne projektowe nie są wielkościami deterministycznymi, lecz cechują się pewnym rozrzutem, a współrzędne punktu A tworzą wektor wartości oczekiwanych odpowiednich zmiennych losowych. W takim przypadku, większość możliwych realizacji tych zmiennych znajdzie się w pewnym ograniczonym obszarze wokół optimum deterministycznego. Dla prostoty prezentacji oznaczmy ten obszar jako koło o środku w punkcie A. Jak można zobaczyć na rys. 2.2, bardzo duża część realizacji

zmiennych projektowych znajduje się poza obszarem dopuszczalnym, tak więc rozwiązanie A trudno jest uznać za bezpieczne. W celu zapewnienia wymaganego poziomu niezawodności, koło otaczające punkt A należy przesunąć do wnętrza obszaru dopuszczalnego tak, aby jego nowy środek B wyznaczył rozwiązanie gwarantujące wyższą niezawodność. Operacja ta prowadzi oczywiście do zwiększenia wartości funkcji celu, a o tym jak daleko od granicy obszaru dopuszczalnego musi znajdować się rozwiązanie B decyduje założony margines bezpieczeństwa. Jeśli przyjąć, że masa prawdopodobieństwa odpowiadająca oczekiwanej niezawodności rozwiązania w sposób symboliczny ograniczona jest przez koło na rys. 2.2, to w punkcie optymalnym koło to będzie styczne do wybranych powierzchni (krzywych) ograniczeń. Takie ograniczenia nazywa się wtedy probabilistycznie aktywnymi. Proces poszukiwania rozwiązania B o opisanych właściwościach nosi nazwę optymalizacji niezawodnościowej i sprowadza się do rozwiązania problemu (2.15)–(2.19).

Obliczenie wartości prawdopodobieństw P_f^i , które występują w ograniczeniach (2.17), wymaga przeprowadzenia analizy niezawodności przy pomocy odpowiedniej procedury numerycznej. Wybór metody zależy od wielkości $n_X + n_P$ oraz złożoności obliczeniowej rozpatrywanego zadania. Krótki opis kilku najbardziej popularnych metod komputerowej analizy niezawodności przedstawiono w dodatku A. Ponieważ szacowanie prawdopodobieństwa przy pomocy metod symulacyjnych jest zazwyczaj bardzo czasochłonne, dlatego najczęściej, ze względu na swoją prostotę i efektywność, w algorytmach optymalizacji niezawodnościowej stosuje się metody analizy niezawodności pierwszego rzędu (FORM - ang. *first order reliability method*).

Wysoki koszt obliczeń sprawia, że rozpatrując metody rozwiązywania zadania RBDO porównuje się je przede wszystkim ze względu na ich efektywność, dokładność i stabilność. Za doskonałą pracą Aouesa i Chateaufa [4] metody optymalizacji niezawodnościowej podzielić można na trzy kategorie: metody zagnieżdżone lub dwupoziomowe (ang. *nested/two-level*), metody jednopoziomowe (ang. *mono-level/single loop*) i metody rozprężone (ang. *decoupled*). Metody te scharakteryzować można następująco:

- Zagnieżdżony charakter *metody dwupoziomowej* wynika z tego, iż obliczenie wartości ograniczeń niezawodnościowych (2.17) wiąże się w istocie z rozwiązaniem zadania optymalizacji z ograniczeniami, jakim jest poszukiwanie punktu projektowego (zob. dodatek A, a w szczególności podrozdział A.3). Na każdym kroku iteracyjnego procesu wyznaczania optymalnych wartości zmiennych \mathbf{d} i $\boldsymbol{\mu}_X$ (zewnątrznej pętli optymalizacyjnej) realizowanych jest k_g zadań analizy niezawodności (wewnętrznych pętli optymalizacyjnych).

- Celem *metod jednopoziomowych* jest wyeliminowanie wewnętrznej pętli związanej z analizą niezawodności poprzez rozszerzenie zbioru zmiennych decyzyjnych oraz zastąpienie ograniczeń niezawodnościowych poprzez kryteria optymalności zadań poszukiwania punktów projektowych. Wyznaczając optymalny wektor zmiennych decyzyjnych, wyznacza się jednocześnie współrzędne punktów projektowych dla ograniczeń niezawodnościowych odpowiadających projektowi optymalnemu.
- Podejście stosowane w *metodach rozprężonych* sprowadza się do transformacji zadania optymalizacji niezawodnościowej do szeregu zadań optymalizacji deterministycznej z ograniczeniami, które mają zapewnić równoważność rozwiązania tak sformułowanego zadania i zadania oryginalnego.

Literatura dotycząca metod analizy niezawodności z każdej z wymienionych powyżej grup jest bardzo bogata i dlatego warto w ramach tego opracowania omówić dokładniej przykłady metod reprezentujących poszczególne podejścia. Przed dokonaniem takiego przedstawienia należy jednak pokusić się o bardziej ogólną refleksję dotyczącą samej “natury” optymalizacji niezawodnościowej. Niestety, czasem zapomina się, że wiarygodność wyników RBDO uwarunkowana jest precyzyjną znajomością rozkładów prawdopodobieństwa zmiennych losowych, a taki dokładny model stochastyczny obciążeń czy też parametrów materiałowych dostępny jest w praktyce niesłychanie rzadko. Jak to zostało pokazane w książce Ben-Haima i Elishakoffa [12], nieprawidłowe założenia dotyczące typu i parametrów funkcji gęstości prawdopodobieństwa prowadzić mogą do znaczących błędów w analizie niezawodności. Z uwagi na to, przydatność optymalizacji niezawodnościowej, jako bazującej na informacji zgromadzonej w ogonach rozkładów prawdopodobieństwa, jest znacznie ograniczona.

2.3.1. Metody rozwiązania

2.3.1.1. Metoda dwupoziomowa

Metoda dwupoziomowa jest algorytmem bezpośredniego rozwiązania zadania (2.15)–(2.19), gdzie w obrębie pierwszego poziomu (zewnętrznej pętli) poszukiwane są optymalne wartości zmiennych decyzyjnych \mathbf{d} i $\boldsymbol{\mu}_{\mathbf{X}}$, a drugi poziom (pętla wewnętrzna), realizowany w przestrzeni zmiennych losowych, dotyczy obliczania wartości ograniczeń niezawodnościowych (2.17). W ramach metody dwupoziomowej wyróżnić można dwa dominujące podejścia:

Metoda wskaźnika niezawodności. Znana w literaturze naukowej pod nazwą *reliability index approach* (RIA) metoda ta sprowadza się najczęściej do szacowania prawdopodobieństw awarii występującego we wzorze (2.17) za pomocą przybliżenia pierwszego rzędu – FORM, zob. (A.4). Ograniczenia niezawodnościowe wyrazić wtedy można w alternatywnej formie poprzez wskaźniki niezawodności

$$\beta_i^{\text{FORM}}(\mathbf{d}, \mathbf{X}, \mathbf{P}) \geq \beta_i^t, \quad i = 1, \dots, k_g, \quad (2.20)$$

gdzie β_i^{FORM} są wskaźnikami pierwszego rzędu odpowiadającymi punktom projektowym funkcji granicznych $h_i(\mathbf{d}, \mathbf{U}) = g_i(\mathbf{d}, \mathbf{T}(\mathbf{X}, \mathbf{P}))$, $i = 1, \dots, k_g$. Przez $\mathbf{T}(\mathbf{X}, \mathbf{P})$ oznaczono transformację do gaussowskiej przestrzeni standardowej \mathcal{U} , zob. podrozdział A.2.

Zadanie lokalizacji punktu projektowego, który w przestrzeni \mathcal{U} oznaczany jest najczęściej jako \mathbf{u}^* , powtarzane k_g razy na każdym kroku zewnętrznej pętli optymalizacyjnej, stanowi o złożoności obliczeniowej RIA. Zaproponowane przez Nikolaidisa i Burdisso w 1988 roku [185], podejście to podlegało szeregowi modyfikacji mających na celu podniesienie efektywności jego numerycznej realizacji. Modyfikacje dotyczyły przede wszystkim użycia lepiej dostosowanych algorytmów optymalizacji [131, 233], zastosowania aproksymacji metodą powierzchni odpowiedzi (w analizie niezawodności stosowano także metodę Monte Carlo zamiast FORM) [71, 72] czy budowy rozwiązań interaktywnych pozwalających na aproksymację lub rezygnację z obliczania wybranych wskaźników niezawodności w trakcie trwania procesu optymalizacji [126, 225, 230, 235]

Niewątpliwą zaletą podejścia RIA jest względna łatwość implementacji dzięki możliwości bezpośredniego połączenia modułu analizy niezawodności z dowolnym programem optymalizacyjnym ogólnego przeznaczenia.

Metoda minimum funkcji granicznej. Opracowana przez Tu i Choi'a [253] jako *performance measure approach* (PMA) metoda ta bazuje na spostrzeżeniu, że minimalizacja złożonej funkcji przy prostych ograniczeniach jest bardziej efektywna numerycznie niż minimalizacja prostej funkcji na obszarze o złożonym kształcie. Ograniczenia niezawodnościowe (2.17) zastępuje się tu przez

$$G_i = F_{g_i}^{-1}(\Phi(-\beta_i^t)) \geq 0, \quad i = 1, \dots, k_g, \quad (2.21)$$

gdzie F_{g_i} jest dystrybuantą funkcji granicznej g_i , a G_i jest wartością funkcji granicznej odpowiadającą dopuszczalnej niezawodności wyrażonej wskaźnikiem β_i^t . W ramach PMA, G_i wyznacza się jako rozwiązanie odwrotnego zadania analizy niezawodności (por. (A.6)) poszukując w przestrzeni standardowej punktu

położonego na hipersferze $\|\mathbf{u}\| = \beta_i^t$, dla którego funkcja graniczna przyjmuje wartość minimalną

$$G_i = h(\mathbf{u}_{\beta=\beta_i^t}^*),$$

$$h(\mathbf{u}_{\beta=\beta_i^t}^*) = \min h(\mathbf{u}) \quad \text{pod warunkiem: } \|\mathbf{u}\| = \beta_i^t. \quad (2.22)$$

Tak więc, algorytm lokalizacji punktu $\mathbf{u}_{\beta=\beta_i^t}^*$, $i = 1, \dots, k_g$ sprowadza się do przeszukiwania hipersfery o promieniu β_i^t , w odróżnieniu od podejścia RIA, gdzie punkt \mathbf{u}^* poszukiwany jest na granicy obszaru awarii. W celu rozwiązania zadania minimalizacji (2.22) zaproponowano szereg wyspecjalizowanych algorytmów, z których najpopularniejszym wydaje się być algorytm HMV (ang. *hybrid mean value*) [33]. W przeciwieństwie do metod iHLRF lub ARF (zob. podrozdział A.3), HMV nie wymaga procedury minimalizacji kierunkowej (w celu znalezienia wielkości przyrostu). W szeregu prac porównawczych wykazano lepszą efektywność i stabilność metody PMA w stosunku do RIA, zob. np. [4, 263].

Rozwiązanie zadania (2.15)–(2.19) przy użyciu efektywnych algorytmów gradientowych wymaga obliczenia wrażliwości funkcji celu, a także wrażliwości ograniczeń niezawodnościowych względem zmiennych decyzyjnych oraz zmiennych losowych. Ze względu na zagnieżdżony charakter sformułowania, sposób wyznaczania gradientów stanowi o złożoności obliczeniowej przyjętej metody rozwiązania. Niestety, w przypadku modelowania złożonych konstrukcji i procesów technologicznych, gdzie mamy do czynienia z dużym zróżnicowaniem typów zmiennych decyzyjnych i zmiennych losowych, efektywne, analityczne metody obliczenia gradientów nie zawsze są dostępne. W takiej sytuacji odpowiedź badanego układu mechanicznego modelowana jest często za pomocą odpowiednio dobranej funkcji aproksymacyjnej.

2.3.1.2. Metoda jednopoziomowa

Jedną z niedoskonałości metod dwupoziomowych jest brak matematycznego dowodu ich zbieżności. Dowód ten jest niesłuchanie trudny do przeprowadzenia ponieważ obszary awarii określające ograniczenia niezawodnościowe (2.17), a w konsekwencji obszar dopuszczalny zadania optymalizacji, zależą od zmiennych decyzyjnych.

Chęć przezwyciężenia problemów ze zbieżnością była motywacją do poszukiwania alternatywnych sformułowań zadania optymalizacji niezawodnościowej, w których unika się zagnieżdżenia pętli optymalizacyjnych. Podejście takie,

w odróżnieniu od przedstawionego wcześniej, określa się mianem jednopoziomowego. Twórcy tej metody, Madsen i Friis Hansen [159], a następnie Kuschel i Rackwitz [134, 135], zaproponowali zastąpienie ograniczeń niezawodnościowych warunkami Kuhna-Tuckera odpowiednich zadań poszukiwania punktu projektowego (A.33)–(A.35). W ramach podejścia jednopoziomowego rozpatruje się rozszerzony zbiór zmiennych decyzyjnych, gdzie oprócz wartości zmiennych \mathbf{d} i $\boldsymbol{\mu}_{\mathbf{X}}$ w tej samej pętli optymalizacyjnej poszukuje się współrzędnych punktów projektowych \mathbf{u}_i , $i = 1, \dots, k_g$, w standardowej przestrzeni gaussowskiej po transformacji $\mathbf{U} = \mathbf{T}(\mathbf{X}, \mathbf{P})$ oryginalnych zmiennych losowych \mathbf{X} i \mathbf{P} . Zmiany te prowadzą do następującego przeformułowania zadania (2.15)–(2.19):

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \mathbf{u}_1, \dots, \mathbf{u}_{k_g}, \quad (2.23)$$

$$\text{minimalizujące: } f(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{P}}), \quad (2.24)$$

przy ograniczeniach:

$$(\mathbf{u}_i)^T \nabla_{\mathbf{u}} h_i(\mathbf{d}, \mathbf{u}_i) + \|\mathbf{u}_i\| \|\nabla_{\mathbf{u}} h_i(\mathbf{d}, \mathbf{u}_i)\| = 0, \quad i = 1, \dots, k_g, \quad (2.25)$$

$$h_i(\mathbf{d}, \mathbf{u}_i) = g_i(\mathbf{d}, \mathbf{T}^{-1}(\mathbf{u}_i)) = 0, \quad i = 1, \dots, k_g, \quad (2.26)$$

$$\text{sgn}[h_i(\mathbf{d}, \mathbf{0})] \|\mathbf{u}_i\| \geq \beta_i^t, \quad i = 1, \dots, k_g, \quad (2.27)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n_d, \quad (2.28)$$

$${}^l \mu_{X_r} \leq \mu_{X_r} \leq {}^u \mu_{X_r}, \quad r = 1, \dots, n_X, \quad (2.29)$$

$${}^l u_{ik} \leq u_{ik} \leq {}^u u_{ik}, \quad i = 1, \dots, k_g, k = 1, \dots, n_X + n_P, \quad (2.30)$$

gdzie h_i , $i = 1, \dots, k_g$, są funkcjami granicznymi w przestrzeni standardowej, a $\nabla_{\mathbf{u}} h_i$ są ich gradientami względem zmiennych \mathbf{u} . Warunki Kuhna-Tuckera (2.25) i (2.26) mają jasną interpretację geometryczną. Wynika z nich, że $n_X + n_P$ elementowy wektor \mathbf{u}_i jest prostopadły w punkcie projektowym do powierzchni granicznej $h_i(\mathbf{d}, \mathbf{u}_i) = 0$, $i = 1, \dots, k_g$. Ograniczenia (2.27) otrzymać można wprost z ograniczeń (2.20) i definicji wskaźnika niezawodności pierwszego rzędu (A.6).

Niewątpliwą zaletą sformułowania jednopoziomowego jest możliwość zastosowania standardowych algorytmów optymalizacyjnych, dla których istnieją dowody zbieżności. Jednak z drugiej strony, sformułowanie (2.23)–(2.30) pociąga za sobą szereg niedogodności praktycznych, które w znaczący sposób ograniczyć mogą jego szersze zastosowanie. Przyjęta koncepcja wiąże się z koniecznością rozpatrywania zadania optymalizacji z $n_d + n_X + k_g(n_X + n_P)$ zmiennymi. Łatwo sprawdzić, że nawet przy umiarkowanej liczbie zmiennych decyzyjnych,

zmiennych losowych i ograniczeń niezawodnościowych, całkowity wymiar zadania okazać się może bardzo duży. Ponadto, w przypadku stosowania algorytmów gradientowych, postać ograniczeń (2.25) narzuca konieczność obliczania macierzy drugich pochodnych, co najczęściej możliwe jest jedynie za pomocą schematów różnicowych, prowadząc do znacznego wzrostu złożoności numerycznej algorytmu i utrudniając jego zbieżność. Dla przewyciężenia tych trudności stosuje się czasami iteracyjny schemat aproksymacji hesjanu BFGS, zob. [6]. Wiele szczegółów dotyczących implementacji metody jednopoziomowej przedstawiono w pracach Straichera i Rackwitza [239, 240].

Znane są także inne sformułowania, które uznać można jako jednopoziomowe. Dokładna ich analiza wykracza jednak znacznie poza zakres niniejszego opracowania. Takie omówienie oraz porównanie efektywności algorytmów znaleźć można np. w pracach [2, 4, 262].

2.3.1.3. Metoda rozprzężona

Idea rozprzężonego podejścia w optymalizacji niezawodnościowej polega na sformułowaniu problemu RBDO jako sekwencji niezależnych zadań optymalizacji deterministycznej oraz zadań analizy niezawodności. Na danym kroku algorytmu, na podstawie analizy niezawodności przeprowadzanej na kroku poprzednim, modyfikuje się funkcje ograniczeń zadania optymalizacji deterministycznej. Modyfikacje te mają prowadzić do spełnienia ograniczeń niezawodnościowych przez rozwiązanie deterministyczne i do otrzymania zbieżności do rozwiązania optymalnego zadania RBDO po niewielkiej liczbie iteracji.

Wydaje się, iż obecnie najbardziej efektywną metodą z grupy metod rozprzężonych jest metoda SORA (ang. *Sequential Optimization and Reliability Assessment*) zaproponowana przez Du i Chen [59] w 2002 roku. Podejście to zyskało już dość dużą popularność, a przeprowadzone porównania przemawiają często na korzyść tej metody, zob. [261, 262, 266, 275].

Przyjmując, że mamy do czynienia z k_g ograniczeniami niezawodnościowymi, iteracje (cykle) metody SORA składają się z zadania optymalizacji deterministycznej oraz z k_g odwrotnych zadań analizy niezawodności (2.22), gdzie poszukuje się punktów projektowych $\mathbf{u}_{\beta=\beta_i^t}^*$, $i = 1, \dots, k_g$. Obrazami tych punktów w przestrzeni oryginalnej są $\{\mathbf{x}_{\beta=\beta_i^t}^*, \mathbf{p}_{\beta=\beta_i^t}^*\} = \mathbf{T}^{-1}(\mathbf{u}_{\beta=\beta_i^t}^*)$. Wektory $\mathbf{x}_{\beta=\beta_i^t}^*$, $\mathbf{p}_{\beta=\beta_i^t}^*$ i $\mathbf{u}_{\beta=\beta_i^t}^*$ zawierają odpowiednio n_X , n_P i $n_X + n_P$ elementów. Na podstawie znajomości współrzędnych wektora $\mathbf{x}_{\beta=\beta_i^t}^*$ otrzymanego w poprzednim kroku iteracyjnym, ograniczenia zadania optymalizacji deterministycznej przesuwane są

w głąb obszaru dopuszczalnego. Na pierwszym kroku, jako przybliżenie punktów projektowych $\mathbf{x}_{\beta=\beta_i^t}^*$ przyjmuje się wektor wartości oczekiwanych $\boldsymbol{\mu}_X$, natomiast zadanie optymalizacji deterministycznej dane na kroku $k + 1$ ma postać:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_X, \quad (2.31)$$

$$\text{minimalizujące: } f(\mathbf{d}, \boldsymbol{\mu}_X, \boldsymbol{\mu}_P), \quad (2.32)$$

$$\text{przy ograniczeniach: } g_i(\mathbf{d}, \boldsymbol{\mu}_X - \mathbf{s}_i^{(k+1)}, \mathbf{p}_{\beta=\beta_i^t}^{*(k)}) \geq 0, \quad i = 1, \dots, k_g, \quad (2.33)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n_d, \quad (2.34)$$

$${}^l \mu_{X_r} \leq \mu_{X_r} \leq {}^u \mu_{X_r}, \quad r = 1, \dots, n_X, \quad (2.35)$$

gdzie $\mathbf{s}_i^{(k+1)} = \boldsymbol{\mu}_X^{(k)} - \mathbf{x}_{\beta=\beta_i^t}^{*(k)}$, $\boldsymbol{\mu}_X^{(k)}$ jest wektorem optymalnych wartości zmiennych $\boldsymbol{\mu}_X$ na k -tym kroku iteracyjnym, a $\mathbf{x}_{\beta=\beta_i^t}^{*(k)}$ wektorem współrzędnych punktu projektowego odpowiadających zmiennym \mathbf{X} w przestrzeni oryginalnej dla i -tego ograniczenia niezawodnościowego.

Użycie w metodzie SORA zadania odwrotnego analizy niezawodności prowadzi do poprawy efektywności rozwiązania, szczególnie wtedy gdy wiele spośród ograniczeń niezawodnościowych nie jest aktywnych. Podejście to jest również stosunkowo łatwe w implementacji.

2.4. Optymalizacja odpornościowa

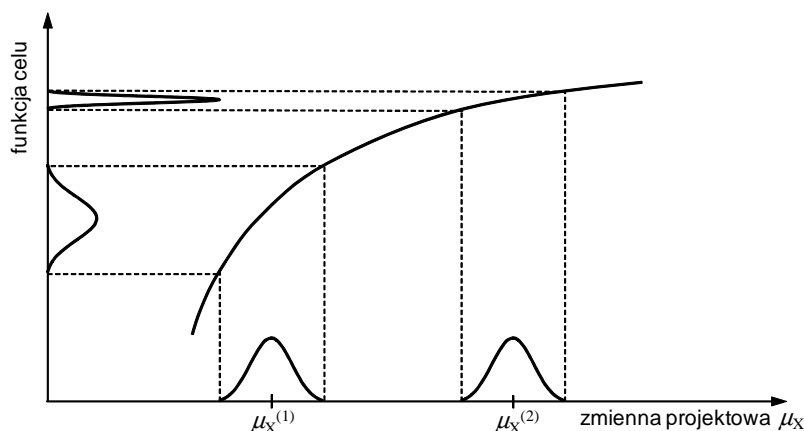
Charakterystyki pracy układów konstrukcyjnych podlegają często istotnym losowym rozrzutom na różnych etapach zaplanowanego okresu eksploatacji. W szczególności odpowiedzi układów nieliniowych cechować się mogą znaczną wariancją. Losowa zmienność odpowiedzi konstrukcji nie tylko obniża jakość samej konstrukcji, powodując odchyłki od założonej funkcjonalności, lecz prowadzi także do zwiększenia kosztów eksploatacji – inspekcji, napraw i utrzymania. Dobrze zaprojektowana konstrukcja powinna zatem minimalizować wymienione koszty poprzez zapewnienie poprawnego funkcjonowania, pomimo losowej natury parametrów determinujących jej pracę. Dla przykładu, w przypadku rozwiązań konstrukcyjnych mających zapewnić bezpieczeństwo pasażerów w czasie zderzenia, duże rozrzuty rejestrowanych kryteriów obrażeń spowodowane niewielkimi zmianami warunków początkowych oznaczać mogą, że “nominalnie” bezpieczny samochód nie gwarantuje dostatecznej ochrony pasażerom podczas rzeczywistego wypadku. Tak więc, nadmierna wariancja określonych odpowiedzi układu konstrukcyjnego jest zazwyczaj oznaką jego niskiej jakości.

Potrzeba zapewnienia wymaganego poziomu jakości konstrukcji była motorem rozwoju odpowiednich metod i procedur projektowania, które powszechnie określa się jako *robust design*. Angielski termin *robust design* odnosi się do szeroko rozumianej metodologii projektowania konstrukcji, urządzeń lub procesów produkcyjnych, w której, zachowując funkcjonalność projektowanego systemu, dąży się do znalezienia rozwiązania jak najmniej wrażliwego (jak najbardziej odpornego) na zmiany wartości jego parametrów.

Jednym ze sposobów na redukcję wariancji wielkości charakteryzujących pracę konstrukcji jest ograniczenie lub wręcz eliminacja rozrzutu wartości parametrów wejściowych. Niestety, postępowanie to jest w praktyce najczęściej niemożliwe do przeprowadzenia lub też generowane dodatkowe koszty przekreślają zasadność podejmowanych działań. Inną metodą jest takie modyfikowanie wybranych zmiennych projektowych, by nie zmieniając wariancji losowych parametrów konstrukcji otrzymać projekt odporny na te wariancje. Częścią tak określonej metodologii jest właśnie optymalizacja odpornościowa.

Jeśli przeanalizować jeszcze raz rysunek 2.2 ilustrujący ideę optymalizacji niezawodnościowej, a w szczególności układ warstwicy funkcji celu, można zauważyć, że wartość tej funkcji w punkcie B (optymalnym rozwiązaniu RBDO) zmienia się dość gwałtownie przy przejściu do punktów położonych w bliskim sąsiedztwie. Punkt B nie jest zatem dobrym rozwiązaniem biorąc pod uwagę odporność na losowe imperfekcje zmiennych. W optymalizacji odpornościowej, jeśli nie wiąże się to ze zbyt znacznym zwiększeniem wartości funkcji celu, dąży się do umiejscowienia rozwiązania w obszarze “spłaszczenia” jej wykresu, tak aby losowe realizacje parametrów konstrukcji wokół tego rozwiązania prowadziły do podobnych wartości funkcji celu. W przypadku zadania optymalizacji z jedną zmienną projektową μ_X ideę tę przedstawiono na rys. 2.3. Zakładając stałość wariancji zmiennej losowej X , można zauważyć, że zaletą rozwiązania w punkcie $\mu_X^{(2)}$ jest dużo mniejsza wariancja funkcji celu w porównaniu z $\mu_X^{(1)}$, gdzie redukcji wartości średniej funkcji celu towarzyszy znaczny wzrost wariancji.

Koncepcja projektowania odpornego na imperfekcje jest bardzo dobrze znana i stanowi ważny element procesu tworzenia wyrobów, np. w przemyśle elektro-nicznym czy przemyśle maszynowym, zob. [247]. Ugruntowana w tych gałęziach przemysłu strategia projektowania, mającego przy możliwie najmniejszych kosztach zapewnić akceptowalną przez klientów jakość produktów, nazywana jest *inżynierią jakości*. Niestety, standardowe procedury inżynierii jakości nie dają się zawsze w bezpośredni sposób przenieść do zagadnień optymalizacji konstrukcji. Dlatego, w przeciwieństwie do precyzyjnie zdefiniowanego zadania optymalizacji niezawodnościowej, istnieje wiele sformułowań odpornościowej optymalizacji



Rys. 2.3. Idea optymalizacji odpornościowej: rozwiązanie $\mu_x^{(2)}$ pomimo iż odpowiada mu większa wartość średnia funkcji celu jest preferowane ze względu na redukcję wariancji.

konstrukcji, w których stosuje się różnorodne postacie funkcji celu, a także różne miary odporności konstrukcji na imperfekcje.

W klasycznym podejściu Taguchi'ego rozróżnia się trzy etapy projektowania inżynierskiego, por. [120]. Są to: projektowanie systemu, projektowanie parametrów i projektowanie tolerancji. Optymalizacja odpornościowa stosowana może być zarówno przy projektowaniu parametrów jak i projektowaniu tolerancji. W bieżącym opracowaniu etapy te nie będą jednak rozróżniane, gdyż z punktu widzenia realizacji numerycznej różnica pomiędzy nimi dotyczy jedynie dopuszczalnych zakresów zmiennych decyzyjnych.

Zarówno niezawodnościowa optymalizacja konstrukcji jak i odpornościowa optymalizacja konstrukcji są przykładami niedeterministycznych sformułowań optymalizacji. Ponieważ w ramach obydwu tych podejść uwzględnia się wpływ losowego charakteru parametrów konstrukcji na rozrzut jej odpowiedzi, dlatego są one niekiedy mylone. Jednak sformułowania te różnią się dość istotnie, pomimo, iż zazwyczaj rozwiązanie optymalizacji odpornościowej prowadzi do zwiększenia niezawodności.

Odporność konstrukcji na losową zmienność jej parametrów szacowana jest przy pomocy wybranej miary rozrzutu odpowiedzi wokół wartości średniej, natomiast niezawodność konstrukcji określana jest na podstawie prawdopodobieństwa awarii (o wartościach rzędu $10^{-7} \div 10^{-3}$). Głównym celem RBDO jest spełnienie ograniczeń niezawodnościowych, nie zaś ograniczenie wariancji funkcji celu, a podstawowym zadaniem optymalizacji odpornościowej jest minimalizacja tej wariancji. Funkcja celu zadania optymalizacji niezawodnościowej, która

wyrażona jest najczęściej przez wartości średnie odpowiedzi konstrukcji, minimalizowana jest pod warunkiem spełnienia ograniczeń nałożonych na dopuszczalne wartości prawdopodobieństwa awarii. Z drugiej strony, typowa funkcja celu w optymalizacji odpornościowej zawiera zazwyczaj człony dotyczące wariancji funkcji odpowiedzi konstrukcji, a ograniczenia mogą mieć charakter deterministyczny lub wyrażone być mogą przez pierwsze dwa momenty statystyczne odpowiedzi. Tak więc, o ile w optymalizacji niezawodnościowej nacisk kładziony jest na zapewnienie bezpieczeństwa konstrukcji w sytuacjach krytycznych, o tyle w optymalizacji odpornościowej większą uwagę zwraca się na odpowiednią pracę konstrukcji poddanych mniejszym fluktuacjom wartości parametrów.

Ponadto, to co stanowi zasadniczą, koncepcyjną, różnicę pomiędzy omawianymi sformułowaniami optymalizacji, to założenia dotyczące modelu stochastycznego. Jak już wspomniano w poprzednim podrozdziale (zob. strona 42), rozwiązanie RBDO silnie zależy od dokładnej znajomości funkcji łącznej gęstości rozkładu prawdopodobieństwa zmiennych losowych, ponieważ to właśnie informacja zawarta w ogonach tych rozkładów determinuje wielkość prawdopodobieństwa awarii. W przypadku optymalizacji odpornościowej problem nieprecyzyjnego określenia typów rozkładów prawdopodobieństwa nie jest jednak bardzo istotny. Jako że wartości pierwszych momentów statystycznych odpowiedzi konstrukcji zależą przede wszystkim od pierwszych momentów zmiennych losowych, dlatego często przy braku odpowiednich danych zakłada się rozkład jednostajny lub normalny zmiennych. Można zatem zaryzykować twierdzenie, że koncepcja optymalizacji odpornościowej jest spójna z zazwyczaj bardzo uproszczonym modelem stochastycznym rozpatrywanego zagadnienia, natomiast w sformułowaniu optymalizacji niezawodnościowej model ten w sposób sztuczny się rozbudowuje.

Ponieważ omówieniu aktualnego stanu prac w dziedzinie optymalizacji odpornościowej poświęcono podrozdział 1.3.2 we wstępie, poniżej przedstawione zostanie jedynie to sformułowanie i algorytmy rozwiązania, które wybrano oraz rozwijano w niniejszym opracowaniu. Zaproponowane metody realizacji zadań optymalizacji odpornościowej użyte zostały w rozdziale 6 w przykładach optymalizacji dotyczących wytrzymałości zderzeniowej elementów konstrukcji pojazdów, a także w pracy autora na temat optymalizacji wałów urządzeń wirnikowych, zob. [236].

2.4.1. Sformułowanie problemu

Zgodnie z filozofią inżynierii jakości, konstrukcja lub proces technologiczny oceniane są nie tylko na podstawie średnich wartości opisujących je charakterystyk, lecz także na podstawie ich rozrzutów. Tak więc, celem optymalizacji

odpornościowej powinna być jednoczesna minimalizacja zarówno wartości średniej jak i losowej zmienności funkcji kryterium. Pewną zgrubną, lecz mimo to często stosowaną, miarą rozrzutu jest odchylenie standardowe.

Zadanie optymalizacji odpornościowej sformułować można zatem jako następujące zadanie optymalizacji wielokryterialnej:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \quad (2.36)$$

$$\text{minimalizujące: } \{\mathbb{E}[f(\mathbf{d}, \mathbf{X}, \mathbf{P})], \sigma[f(\mathbf{d}, \mathbf{X}, \mathbf{P})]\}, \quad (2.37)$$

przy ograniczeniach:

$$\mathbb{E}[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})] - \tilde{\beta}_i \sigma[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})] \geq 0, \quad i = 1, \dots, k_g, \quad (2.38)$$

$$\sigma[c_k(\mathbf{d}, \mathbf{X}, \mathbf{P})] \leq {}^u\sigma_k, \quad k = 1, \dots, k_c, \quad (2.39)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n_d, \quad (2.40)$$

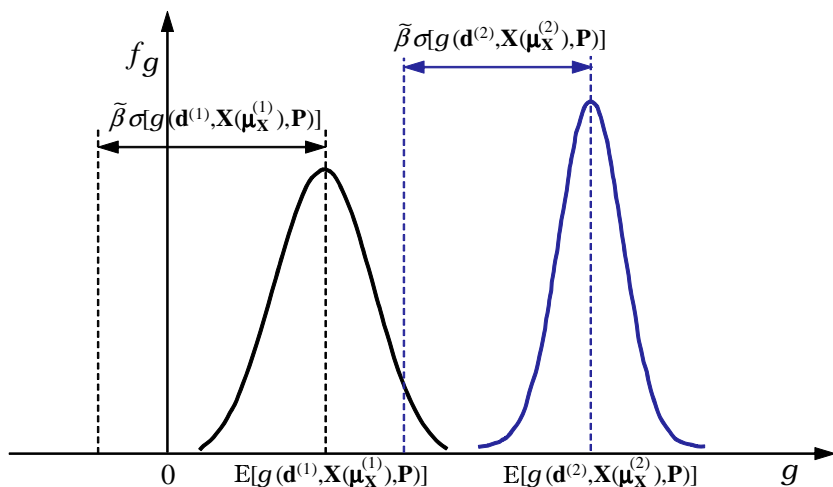
$${}^l \mu_{X_r} \leq \mu_{X_r} \leq {}^u \mu_{X_r}, \quad r = 1, \dots, n_X, \quad (2.41)$$

gdzie, tak jak w zadaniu optymalizacji deterministycznej (2.1)–(2.5), f i g_i , $i = 1, \dots, k_g$, są odpowiednio funkcją celu oraz funkcjami ograniczeń, c_k , $k = 1, \dots, k_c$, są funkcjami, których odchylenia standardowe nie mogą przekroczyć dopuszczalnych wartości ${}^u\sigma_k$, a współczynniki $\tilde{\beta}_i > 0$ odpowiadające ograniczeniom $g_i \geq 0$ reprezentują margines bezpieczeństwa, z jakim spełnione być muszą te ograniczenia. W powyższym sformułowaniu mamy do czynienia z dwuelementowym wektorem kryteriów optymalizacji: wartością średnią $\mathbb{E}[f(\mathbf{d}, \mathbf{X}, \mathbf{P})]$ i odchyleniem standardowym $\sigma[f(\mathbf{d}, \mathbf{X}, \mathbf{P})] = \sqrt{\text{Var}[f(\mathbf{d}, \mathbf{X}, \mathbf{P})]}$, co czyni (2.36)–(2.41) sformulowaniem zadania optymalizacji wielokryterialnej.

Koncepcja ograniczeń (2.38) pokazana została na rys. 2.4. Rozwiązując problem optymalizacji odpornościowej dąży się do tego, aby w punkcie optymalnym wartości średnie funkcji ograniczeń g_i , $i = 1, \dots, k_g$, odsunięte były od zera o co najmniej $\tilde{\beta}_i$ odchyień standardowych. Ograniczenia (2.38) zapisać można w równoważnej postaci

$$\gamma_i \geq \tilde{\beta}_i, \quad i = 1, \dots, k_g, \quad (2.42)$$

gdzie $\gamma_i = \mathbb{E}(g_i)/\sigma(g_i)$. Nierówność (2.42) można potraktować jako odpowiednik ograniczeń niezawodnościowych (2.20), jednak jedynie w przypadku nieskorelowanych zmiennych losowych o rozkładzie normalnym oraz liniowej funkcji granicznej wartość γ_i będzie równa wskaźnikowi niezawodności pierwszego rzędu β_i^{FORM} , zob. [230]. Chociaż zadanie optymalizacji odpornościowej nie obejmuje



Rys. 2.4. Ograniczenia zadania optymalizacji odpornościowej. Punkt $\{\mathbf{d}^{(1)}, \boldsymbol{\mu}_{\mathbf{X}}^{(1)}\}$ nie spełnia ograniczeń, ponieważ margines bezpieczeństwa jest mniejszy od $\tilde{\beta}$ odchyłeń standardowych. W przypadku punktu $\{\mathbf{d}^{(2)}, \boldsymbol{\mu}_{\mathbf{X}}^{(2)}\}$ ograniczenie nie jest aktywne. Zastosowana notacja $\mathbf{X}(\boldsymbol{\mu}_{\mathbf{X}}^{(k)})$ służy podkreśleniu zależności rozkładów prawdopodobieństwa zmiennych \mathbf{X} od zmiennych projektowych.

szacowania prawdopodobieństwa awarii, to naturalną konsekwencją zwiększania wartości $\tilde{\beta}_i$ jest wzrost niezawodności rozwiązania.

Dwa kryteria minimalizowane w sformułowaniu (2.36)–(2.41) są na ogół konfliktowe, dlatego mamy wtedy do czynienia nie z jednym rozwiązaniem, a ze zbiorem rozwiązań niezdominowanych (Pareto). Ponieważ zgodnie z definicją, przy przejściu z jednego punktu zbioru Pareto do drugiego ulega pogorszeniu wartość co najmniej jednej funkcji celu (w naszym przypadku dokładnie jednej), dlatego o ostatecznym wyborze rozwiązania decydują dodatkowe kryteria.

Jak to zostało już omówione w podrozdziale 2.2, pomimo swoich niedoskonałości, bardzo popularną metodą wyznaczenia punktów zbioru Pareto jest metoda skalaryzacji zadania optymalizacji wielokryterialnej, w której jako funkcję celu stosuje się liniową kombinację kryteriów. Zmieniając współczynniki (wagi) przy poszczególnych składowych wektora (2.37) otrzymuje się punkty zbioru Pareto. Rozpatrywane wartości wag zmieniać się mogą w sposób systematyczny, aby dokładnie wyznaczyć zbiór rozwiązań niezdominowanych. Wynikać też mogą z założonych a priori preferencji projektanta co do znaczenia, jakie przykładane jest do minimalizacji wartości średniej i wariancji. Zadanie (2.36)–(2.41)

transformuje się więc do następującego zadania optymalizacji skalarnej:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \quad (2.43)$$

$$\text{minimalizujące: } \tilde{f} = \frac{1-\alpha}{\mu^*} \mathbb{E}[f(\mathbf{d}, \mathbf{X}, \mathbf{P})] + \frac{\alpha}{\sigma^*} \sigma[f(\mathbf{d}, \mathbf{X}, \mathbf{P})], \quad (2.44)$$

przy ograniczeniach:

$$\mathbb{E}[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})] - \tilde{\beta}_i \sigma[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})] \geq 0, \quad i = 1, \dots, k_g, \quad (2.45)$$

$$\sigma[c_k(\mathbf{d}, \mathbf{X}, \mathbf{P})] \leq {}^u\sigma_k, \quad k = 1, \dots, k_c, \quad (2.46)$$

$${}^l d_j \leq d_j \leq {}^u d_j, \quad j = 1, \dots, n_d, \quad (2.47)$$

$${}^l \mu_{X_r} \leq \mu_{X_r} \leq {}^u \mu_{X_r}, \quad r = 1, \dots, n_X. \quad (2.48)$$

Występujący we wzorze (2.44) współczynnik wagowy $\alpha \in [0, 1]$ określa znaczenie każdego z kryteriów, a μ^* and σ^* są stałymi normalizującymi, por. (2.13). Jak już wspomniano na stronie 38, czasami lepsze wyniki daje metoda normalizacji (2.14), gdzie uwzględnia się zakres zmienności wartości kryterium. Przyjmując $\alpha = 0$ problem optymalizacji (2.43)–(2.48) przekształca się w zwykłe zadanie minimalizacji wartości średniej natomiast $\alpha = 1$ prowadzi do zadania minimalizacji wariancji funkcji celu.

Dość często problem optymalizacji odpornościowej formułowany jest jako: *minimalizuj wariancję przy wartości średniej jak najbliższej nominalnej*. Wtedy zlinearyzowana funkcja celu przybiera postać (zob. [133])

$$\tilde{f} = \frac{1-\alpha}{\mu^*} \{\mathbb{E}[f(\mathbf{d}, \mathbf{X}, \mathbf{P})] - M\}^2 + \frac{\alpha}{\sigma^*} \sigma[f(\mathbf{d}, \mathbf{X}, \mathbf{P})], \quad (2.49)$$

gdzie M jest wartością nominalną. Obydwa typy funkcji celu, (2.44) i (2.49), zostały zaimplementowane przez autora w omawianym w rozdziale 8 programie STAND, który posłużył do realizacji wszystkich zamieszczonych w pracy przykładów optymalizacji odpornościowej.

W sformułowaniach odpornościowej optymalizacji konstrukcji, spotkać można również inne miary rozrzutu odpowiedzi analizowanego układu konstrukcyjnego. W pracy [175] do opisu zmienności użyto różnicy wartości wybranych percentyli funkcji celu, które obliczane są przy pomocy metody AMV (ang. advanced mean value) zaproponowanej w [257]. Różnicę tę definiuje się jako $\Delta R_f = f^{R_2} - f^{R_1}$, gdzie f^{R_1} i f^{R_2} są odpowiednio dolnym i górnym percentylem funkcji celu f . Zazwyczaj przyjmuje się percentyle rzędu 5 i 95. Zwolennicy takiego modelowania rozrzutu w zadaniu optymalizacji odpornościowej

podkreślają szereg zalet miary ΔR_f w stosunku do odchylenia standardowego. Stosowane wartości percentyli pozwalają lepiej poznać rozkład prawdopodobieństwa funkcji celu, uwzględnić skośność funkcji gęstości prawdopodobieństwa oraz prawdopodobieństwo zawarte w jej ogonach. Jest to oczywiście dużo więcej niż wielkość rozrzutu wokół wartości średniej oferowana przez odchylenie standardowe. Należy jednak zwrócić uwagę, że takie postępowanie stoi nieco w sprzeczności z filozofią optymalizacji odpornościowej, gdzie zakłada się ograniczoną wiedzę na temat modelu stochastycznego konstrukcji, i może sprawiać wrażenie sztucznego generowania bogatego opisu probabilistycznego odpowiedzi konstrukcji na podstawie ubogich danych wejściowych.

Analizując sformułowanie (2.43)–(2.48) nietrudno zauważyć, że kluczowym elementem algorytmu realizacji zadania optymalizacji odpornościowej jest efektywna metoda szacowania wartości średnich i odchyłeń standardowych funkcji celu oraz ograniczeń. Przegląd takich metod przedstawiony został w rozdziale 4 ze szczególnym uwzględnieniem metod symulacyjnych wykorzystujących koncepcję łańciskowej hiperkostki. W opinii autora, właśnie ten typ technik symulacyjnych stanowi dobry kompromis pomiędzy złożonością obliczeniową, a dokładnością oszacowania oraz jest praktycznie niewrażliwy na charakter rozpatrywanych funkcji losowych. Uniwersalność metod symulacyjnych w połączeniu z nieustannym wzrostem mocy obliczeniowej komputerów oraz dostępnością rozwiązań wieloprocesorowych, a ponadto specyfika analizowanych w niniejszej pracy przykładów optymalizacji odpornościowej, skłoniły autora do wyboru tej metody obliczania momentów statystycznych odpowiedzi konstrukcji.

2.4.2. Metody rozwiązania

W teorii, do rozwiązania zadania optymalizacji odpornościowej zastosować można wiele standardowych algorytmów optymalizacji. Jednak przyjęcie symulacyjnych metod analizy rozrzutu wartości funkcji celu i ograniczeń oraz generowany przez te metody szum numeryczny związany z błędem estymacji, wymagają użycia wyspecjalizowanych algorytmów obliczeniowych. Bezpośrednie zastosowanie metod Monte Carlo w ramach algorytmu optymalizacji, np. algorytmu genetycznego [218], jest możliwe jedynie w przypadku gdy wszystkie funkcje losowe wchodzące w skład sformułowania dane są w sposób jawny.

Ponieważ efektywność estymacji momentów decyduje o powodzeniu optymalizacji odpornościowej, dlatego większość znanych metod realizacji tego zadania wykorzystuje techniki aproksymacji niejawnych funkcji zmiennych projektowych i zmiennych losowych - powierzchni odpowiedzi. Wyróżnić można następujące

strategii zastosowania powierzchni odpowiedzi w zadaniu optymalizacji odpornościowej:

2.4.2.1. Podstawowa strategia aproksymacji

Naturalnym rozwiązaniem jest stworzenie powierzchni odpowiedzi bezpośrednio dla niejawnych funkcji f , g_i , $i = 1, \dots, k_g$, i c_k , $k = 1, \dots, k_c$, w przestrzeni zmiennych \mathbf{d} , \mathbf{X} i \mathbf{P} . Konieczne jest zatem zaplanowanie eksperymentu w przestrzeni $n_{dXP} = n_d + n_X + n_P$ wymiarowej. W zależności od wielkości n_{dXP} oraz specyfiki optymalizowanego zagadnienia istnieje wiele możliwości wyboru funkcji aproksymującej. Omówieniu różnych technik powierzchni odpowiedzi pod kątem ich przydatności do analizy eksperymentów komputerowych poświęcono rozdział 3. Spośród przedstawionych tam rozwiązań najwierniejsze odwzorowanie wydaje się zapewniać metoda krigingu (zob. podrozdział 3.2, a także [110]).

Utworzone powierzchnie odpowiedzi używa się następnie w połączeniu z wybraną metodą szacowania momentów funkcji losowych do obliczania wartości momentów funkcji celu i ograniczeń w (2.44)–(2.46). Główne kroki algorytmu opartego na podstawowej strategii aproksymacji dane są zatem następująco:

1. Zdefiniowanie obszaru dopuszczalnego zgodnie z (2.47) i (2.48). Wybór współczynnika wagowego α .
2. Wybór planu eksperymentów numerycznych w n_{dXP} wymiarowej przestrzeni zmiennych \mathbf{d} , \mathbf{X} i \mathbf{P} .

W większości przypadków dobrym rozwiązaniem jest użycie planu generowanego przez optymalną hiperkostkę łącińską (zob. dodatek B). Obszar pokrywany przez plan eksperymentów odpowiada zakresowi aktualnego obszaru dopuszczalnego w przypadku zmiennych \mathbf{d} oraz zakresom $[\mu_{\mathbf{X}} - 3\sigma_{\mathbf{X}}, \mu_{\mathbf{X}} + 3\sigma_{\mathbf{X}}]$ i $[\mu_{\mathbf{P}} - 3\sigma_{\mathbf{P}}, \mu_{\mathbf{P}} + 3\sigma_{\mathbf{P}}]$ w przypadkach zmiennych \mathbf{X} i \mathbf{P} o obustronnie nieograniczonych funkcjach gęstości prawdopodobieństwa, gdzie $\sigma_{\mathbf{X}}$ i $\sigma_{\mathbf{P}}$ są wektorami odchyłeń standardowych. Jeśli zakresy określoności zmiennych losowych są ograniczone, to granice te muszą być również uwzględnione w definicji obszaru, w którym generowane są eksperymenty.

3. Budowa powierzchni odpowiedzi \hat{f} dla funkcji celu f oraz powierzchni \hat{g}_i , $i = 1, \dots, k_g$, i \hat{c}_k , $k = 1, \dots, k_c$ dla funkcji ograniczeń odpowiednio g_i i c_k .

Na podstawie testów przeprowadzonych w rozdziale 3 sugeruje się wybór aproksymacji metodą krigingu. Jednak w przypadku wielu zmiennych,

wielu punktów eksperymentalnych oraz wielu funkcji ograniczeń proces wyznaczenia optymalnych współczynników równania krigingu może być czasochłonny i wtedy uzasadnione jest przyjęcie innej metody, np. aproksymacji lokalnej (zob. podrozdział 3.3).

4. Przy pomocy wybranego algorytmu optymalizacji deterministycznej rozwiązanie zadania (2.43)–(2.48), gdzie niejawne funkcje f , g_i i c_k zastępuje się odpowiednimi powierzchniami odpowiedzi \hat{f} , \hat{g}_i i \hat{c}_k .

Za każdym razem gdy algorytm wymaga obliczenia wartości wagowej funkcji celu (2.44) lub wartości ograniczeń odpornościowych (2.45) i (2.46) momenty statystyczne szacowane są przy pomocy metod symulacyjnych. Ze względu na analityczną postać funkcji aproksymujących, w oszacowaniu skorzystać można z próbek o dużej liczebności, redukując dzięki temu błąd estymacji. Ponieważ ze względu na niewielki szum numeryczny, związany ze stosowaniem metod symulacyjnych, mogą występować problemy ze zbieżnością algorytmów gradientowych, dlatego dużo lepiej zdają egzamin bezgradientowe algorytmy optymalizacji globalnej, takie jak algorytm genetyczny czy metoda symulowanego wyżarzania.

5. Sprawdzenie warunku zbieżności. Jeśli jest on spełniony, to zakończenie działania algorytmu.

Oprócz typowych kryteriów zbieżności, takich jak wielkość zmiany położenia punktu optymalnego na dwóch ostatnich iteracjach, ze względu na stosowanie aproksymacji funkcji celu i ograniczeń powinno się dokonać weryfikacji ich wartości w punkcie optymalnym, np. za pomocą symulacji losowych. W przypadku znaczącej rozbieżności wyników w stosunku do wartości aproksymowanych proces iteracyjny powinien być kontynuowany.

6. Przesunięcie obszaru dopuszczalnego nad wyznaczony na kroku 4 punkt optymalny. Redukcja obszaru dopuszczalnego. Powrót do 2.

Pomimo prostej koncepcji, realizacja numeryczna powyższej strategii rozwiązania zadania optymalizacji odpornościowej wiąże się z szeregiem trudności. Wymiar n_{dXP} w wielu przypadkach może być bardzo duży, co pociąga konieczność generowania większej liczby punktów eksperymentalnych w celu odpowiedniego dopasowania powierzchni odpowiedzi. W konsekwencji, nieuniknione błędy aproksymacji prowadzić mogą do znacznych niedokładności przy obliczaniu momentów statystycznych utrudniając zbieżność algorytmu. Co za tym idzie, redukcja czasu obliczeń wynikająca z szacowania momentów w oparciu o analityczne powierzchnie odpowiedzi może być zniwelowana poprzez konieczność stosowania bardziej licznych planów eksperymentów oraz większej liczby iteracji.

2.4.2.2. Strategia aproksymacji statystyk

Znana w literaturze angielskojęzycznej jako *dual response surface approach* (DRS) [149,248,256,265], strategia ta polega na tworzeniu powierzchni odpowiedzi nie dla funkcji f , g_i i c_k (jak w przypadku podstawowej strategii aproksymacji), lecz dla ich momentów statystycznych. Przyjmując sformułowanie (2.43)–(2.48), będą to zatem powierzchnie aproksymujące wartości średnie i odchylenia standardowe odpowiednich funkcji. Ze znanych autorowi prac, powierzchnie te przyjmowane były zawsze w następującej postaci wielomianowej:

$$\hat{y}_\mu(\mathbf{z}) = b_0^\mu + \sum_{i=1}^{n_{dX}} b_i^\mu z_i + \sum_{i=1}^{n_{dX}} b_{ii}^\mu z_i^2 + \sum_{i=1}^{n_{dX}} \sum_{j=2, j>i}^{n_{dX}} b_{ij}^\mu z_i z_j, \quad (2.50)$$

$$\hat{y}_\sigma(\mathbf{z}) = b_0^\sigma + \sum_{i=1}^{n_{dX}} b_i^\sigma z_i + \sum_{i=1}^{n_{dX}} b_{ii}^\sigma z_i^2 + \sum_{i=1}^{n_{dX}} \sum_{j=2, j>i}^{n_{dX}} b_{ij}^\sigma z_i z_j, \quad (2.51)$$

gdzie \hat{y}_μ aproksymuje wartości średnie, a \hat{y}_σ odchylenia standardowe funkcji y (może to być funkcja f , g lub c), z_i , $i = 1, \dots, n_{dX}$, są zmiennymi projektowymi (\mathbf{d} i $\boldsymbol{\mu}_X$), a $n_{dX} = n_d + n_X$ jest liczbą tych zmiennych.

W zaimplementowanej przez autora wersji procedury DRS, zamiast typowej powierzchni wielomianowej drugiego stopnia używa się aproksymacji metodą krigingu lub lokalnej aproksymacji metodą ważonej regresji (zob. podrozdziały 3.2 i 3.3). Najważniejsze etapy algorytmu wyrazić można następująco:

1. Zdefiniowanie obszaru dopuszczalnego zgodnie z (2.47) i (2.48). Wybór współczynnika wagowego α .
2. Wygenerowanie N realizacji wektora zmiennych projektowych, równomiernie w aktualnym obszarze dopuszczalnym, zgodnie z planem OLH.
Liczba N punktów planu OLH (zob. dodatek B) powinna być większa od wielokrotności liczby zmiennych n_{dX} . W przypadku umiarkowanej liczby zmiennych ($n_{dX} < 10$), $N = 4n_{dX}$ wydaje się być minimalną wielkością tego planu eksperymentów.
3. Wyznaczenie momentów statystycznych funkcji celu f oraz funkcji ograniczeń g_i , $i = 1, \dots, k_g$ i c_k , $k = 1, \dots, k_c$, dla każdej z N realizacji wektora $\{\mathbf{d}, \boldsymbol{\mu}_X\}$.

Obliczanie wartości średnich i odchyłeń standardowych funkcji losowych przeprowadzić można na wiele sposobów, zob. rozdział 4. Jednak ze względu na specyfikę rozpatrywanych w tej pracy zagadnień optymalizacji odpornościowej, szacowanie momentów statystycznych wykonywane jest

przy pomocy metod symulacyjnych dla próbek generowanych przez hiperkostki łacińskie. Wielkość próbek powinna być wielokrotnością liczby zmiennych losowych $n_{XP} = n_X + n_P$. Obszerne studia nad doбором wielkości hiperkostek przeprowadzono w podrozdziale 4.1.

4. Budowa powierzchni odpowiedzi (techniką krigingu lub ważonej regresji) modelujących nieznaną zależność momentów statystycznych występujących we wzorach (2.44)–(2.46) od zmiennych projektowych \mathbf{d} i $\boldsymbol{\mu}_X$.

Budowanych jest w sumie $2(k_g + 1) + k_c$ powierzchni odpowiedzi dla funkcji $\mu_f(\mathbf{d}, \boldsymbol{\mu}_X)$, $\sigma_f(\mathbf{d}, \boldsymbol{\mu}_X)$, $\mu_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X)$, $\sigma_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X)$, $i = 1, \dots, k_g$ i $\sigma_{c_j}(\mathbf{d}, \boldsymbol{\mu}_X)$, $j = 1, \dots, k_c$. Aproksymacje te dane są odpowiednio jako $\hat{\mu}_f(\mathbf{d}, \boldsymbol{\mu}_X)$, $\hat{\sigma}_f(\mathbf{d}, \boldsymbol{\mu}_X)$, $\hat{\mu}_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X)$, $\hat{\sigma}_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X)$, $i = 1, \dots, k_g$ i $\hat{\sigma}_{c_k}(\mathbf{d}, \boldsymbol{\mu}_X)$, $k = 1, \dots, k_c$.

5. Rozwiązanie następującego zadania optymalizacji deterministycznej:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_X, \quad (2.52)$$

$$\text{minimalizujące: } \tilde{f}^{\text{DRS}} = \frac{1 - \alpha}{\mu^*} \hat{\mu}_f(\mathbf{d}, \boldsymbol{\mu}_X) + \frac{\alpha}{\sigma^*} \hat{\sigma}_f(\mathbf{d}, \boldsymbol{\mu}_X), \quad (2.53)$$

przy ograniczeniach:

$$\hat{\mu}_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X) - \tilde{\beta}_i \hat{\sigma}_{g_i}(\mathbf{d}, \boldsymbol{\mu}_X) \geq 0, \quad i = 1, \dots, k_g, \quad (2.54)$$

$$\hat{\sigma}_{c_k}(\mathbf{d}, \boldsymbol{\mu}_X) \leq {}^u\sigma_k, \quad k = 1, \dots, k_c, \quad (2.55)$$

$${}^l\tilde{d}_j \leq d_j \leq {}^u\tilde{d}_j, \quad j = 1, \dots, n_d, \quad (2.56)$$

$${}^l\tilde{\mu}_{X_r} \leq \mu_{X_r} \leq {}^u\tilde{\mu}_{X_r}, \quad r = 1, \dots, n_X, \quad (2.57)$$

gdzie ${}^l\tilde{d}_j$, ${}^u\tilde{d}_j$, $j = 1, \dots, n_d$, oraz ${}^l\tilde{\mu}_{X_r}$, ${}^u\tilde{\mu}_{X_r}$, $r = 1, \dots, n_X$, są aktualnymi granicami obszaru dopuszczalnego. Współczynniki normujące μ^* i σ^* wyznaczone są na podstawie ekstremalnych wartości odpowiednich momentów otrzymanych w punkcie 3.

Tak jak w przypadku podstawowej strategii aproksymacji, ze względu na jawną postać funkcji występujących w sformułowaniu (2.52)–(2.57), w rozwiązaniu wskazane jest zastosowanie bezgradientowych metod optymalizacji globalnej.

6. Sprawdzenie warunku zbieżności. Jeśli jest on spełniony, to zakończenie działania algorytmu.
7. Przesunięcie obszaru dopuszczalnego nad wyznaczony na kroku 5 punkt optymalny. Redukcja obszaru dopuszczalnego. Powrót do 2.

Niewątpliwą przewagą metody DRS w stosunku do poprzedniej strategii jest budowanie powierzchni odpowiedzi bezpośrednio dla poszczególnych momentów statystycznych. Pominięcie etapu uprzedniego tworzenia aproksymacji funkcji f , g_i i c_k wiąże się z poprawą dokładności oszacowania. Poza tym, metamodele $\hat{\mu}_f$, $\hat{\sigma}_f$, $\hat{\mu}_{g_i}$, $\hat{\sigma}_{g_i}$ i $\hat{\sigma}_{c_k}$ są funkcjami w przestrzeni $n_{dX} \leq n_{dXP}$ wymiarowej, co również sprzyja zapewnieniu jakości aproksymacji. Trzeba jednak zdawać sobie sprawę, że nieunikniony błąd estymacji momentów, charakterystyczny dla metod symulacyjnych, wymusza stosowanie nieinterpolujących technik aproksymacji. Mając to na uwadze, w podrozdziale 3.2 zaprezentowano modyfikację standardowej metody krigingu, która umożliwia otrzymanie jej wersji aproksymującej, zob. (3.39).

Za wadę strategii aproksymacji statystyk uznać należy stosunkowo dużą liczbę powierzchni odpowiedzi ($2(k_g + 1) + k_c$), które należy utworzyć. Szczególnie dla wielowymiarowych wektorów zmiennych projektowych, może wiązać się to ze znacznym wydłużeniem tego etapu działania algorytmu.

2.4.2.3. Strategia aproksymacji funkcji celu i ograniczeń odpornościowych

Sposobem na poprawę efektywności strategii opisanej w poprzednim punkcie jest ograniczenie liczby zadań aproksymacji, poprzez zbudowanie powierzchni odpowiedzi jedynie dla wagowej funkcji celu (2.44) oraz funkcji ograniczeń (2.45) i (2.46). Zadanie doboru parametrów funkcji aproksymującej rozwiązywane jest wtedy $k_g + k_c + 1$ razy, czyli tyle, ile w przypadku strategii podstawowej. Poniżej przedstawiono poszczególne etapy algorytmu:

1. Zdefiniowanie obszaru dopuszczalnego zgodnie z (2.47) i (2.48). Wybór współczynnika wagowego α .
2. Wygenerowanie N realizacji wektora zmiennych projektowych.
Tak jak w przypadku strategii aproksymacji momentów, eksperymenty generowane są zgodnie z planem OLH, równomiernie w aktualnym obszarze dopuszczalnym. Liczba N powinna gwarantować odpowiednie wypełnienie badanego obszaru.
3. Wyznaczenie momentów statystycznych funkcji celu f oraz funkcji ograniczeń g_i , $i = 1, \dots, k_g$ i c_k , $k = 1, \dots, k_c$, dla każdej z N realizacji wektora $\{\mathbf{d}, \boldsymbol{\mu}_X\}$.

Poza tym, na podstawie oszacowanych wartości momentów, dla każdego z punktów eksperymentalnych obliczane są również wartości ważonej

funkcji celu \tilde{f} , danej wzorem (2.44), oraz funkcji

$$\tilde{g}_i(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}) = \mathbb{E}[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})] - \tilde{\beta}_i \sigma[g_i(\mathbf{d}, \mathbf{X}, \mathbf{P})], \quad i = 1, \dots, k_g, \quad (2.58)$$

$$\tilde{c}_k(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}) = {}^u\sigma_k - \sigma[c_k(\mathbf{d}, \mathbf{X}, \mathbf{P})], \quad k = 1, \dots, k_c. \quad (2.59)$$

Uwagi dotyczące metod szacowania momentów zamieszczone w 2.4.2.2 w punkcie 3 dotyczą także bieżącej strategii.

4. Budowa powierzchni odpowiedzi (techniką krigingu lub ważonej regresji) dla funkcji \tilde{f} oraz funkcji ograniczeń $\tilde{g}_i(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}})$, $i = 1, \dots, k_g$, i $\tilde{c}_k(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}})$, $k = 1, \dots, k_c$.

Budowanych jest $1 + k_g + k_c$ powierzchni odpowiedzi, oznaczanych odpowiednio \hat{f} , w przypadku funkcji celu, oraz $\hat{g}_i(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}})$, $i = 1, \dots, k_g$, i $\hat{c}_k(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}})$, $k = 1, \dots, k_c$, w przypadku funkcji ograniczeń.

5. Rozwiązanie następującego zadania optymalizacji deterministycznej:

$$\text{znajdź wartości zmiennych: } \mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}, \quad (2.60)$$

$$\text{minimalizujące: } \hat{f}(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}), \quad (2.61)$$

przy ograniczeniach:

$$\hat{g}_i(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}) \geq 0, \quad i = 1, \dots, k_g, \quad (2.62)$$

$$\hat{c}_k(\mathbf{d}, \boldsymbol{\mu}_{\mathbf{X}}) \geq 0, \quad k = 1, \dots, k_c, \quad (2.63)$$

$${}^l\tilde{d}_j \leq d_j \leq {}^u\tilde{d}_j, \quad j = 1, \dots, n_d, \quad (2.64)$$

$${}^l\tilde{\mu}_{X_r} \leq \mu_{X_r} \leq {}^u\tilde{\mu}_{X_r}, \quad r = 1, \dots, n_X. \quad (2.65)$$

Podobnie jak w strategii aproksymacji momentów, współczynniki normujące μ^* i σ^* , które występują w definicji ważonej funkcji celu wyznacza się na podstawie ekstremalnych wartości odpowiednich momentów obliczonych w punkcie 3.

W przykładach numerycznych optymalizacji odpornościowej prezentowanych w rozdziale 6 oraz innych publikacjach autora [38, 236] do rozwiązania zadania (2.60)–(2.65) używano dwustopniowego algorytmu optymalizacyjnego. Na pierwszym etapie stosowany był algorytm optymalizacji globalnej. W zależności od wielkości zadania był to algorytm symulowanego wyżarzania, algorytm genetyczny bądź algorytm losowego przeszukiwania. Następnie, położenie otrzymanego w ten sposób punktu optymalnego, korygowane było przy pomocy lokalnej metody optymalizacji, zazwyczaj był to simpleksowy algorytm Nedlera Meada [184].

6. Sprawdzenie warunku zbieżności. Jeśli jest on spełniony, to zakończenie działania algorytmu.
7. Przesunięcie obszaru dopuszczalnego nad wyznaczony na kroku 5 punkt optymalny. Redukcja obszaru dopuszczalnego. Powrót do 2.

Chociaż wprowadzone modyfikacje poprawiają efektywność algorytmu w stosunku do DRS, to jednak kluczowymi elementami efektywnej metody rozwiązania zadania są: odpowiednio dobrana metoda aproksymacji funkcji zmiennych projektowych i efektywna metoda szacowania momentów statystycznych. Zagadnieniom tym poświęcono obszerną część niniejszej pracy.

PODSUMOWANIE

W niniejszym rozdziale podjęto próbę usystematyzowania wiedzy na temat alternatywnych sformułowań zadania niedeterministycznej optymalizacji konstrukcji. Rozpatrywano jedynie takie przypadki, gdzie zarówno parametry konstrukcji, jak i działające na nią obciążenia, modelowane są w ramach teorii prawdopodobieństwa za pomocą zmiennych losowych. Po krótkim omówieniu typowego sformułowania zadania optymalizacji deterministycznej, zarówno w wersji jedno jak i wielokryterialnej, oraz najbardziej znanych metod rozwiązania, przedstawiono dwa najistotniejsze sformułowania stochastycznej optymalizacji konstrukcji - optymalizację niezawodnościową i optymalizację odpornościową.

Optymalizacja niezawodnościowa, w sformułowaniu której ograniczenia projektowe wyraża się najczęściej jako funkcje prawdopodobieństwa awarii konstrukcji, jest przedmiotem badań już od ponad trzech dziesięcioleci. Jest to już zatem "niemłoda" dziedzina o bardzo bogatej literaturze, która jednak ciągle się rozwija przyciągając zainteresowanie wielu badaczy. Znaczna część dorobku naukowego autora była również poświęcona tej problematyce, zob. [52, 125, 126, 207, 230, 232, 233, 235]. W rozdziale omówiono trzy główne rodziny metod komputerowej realizacji zadania optymalizacji niezawodnościowej. Są to: metody dwupoziomowe, metody jednopoziomowe i metody rozprężone. Po przeanalizowaniu wad i zalet poszczególnych metod, na podstawie porównań przeprowadzonych w licznych pracach, za najbardziej efektywną numerycznie należy uznać sekwencyjną metodę SORA, z grupy metod rozprężonych.

Możliwość zastosowania optymalizacji niezawodnościowej w praktyce projektowej jest bardzo silnie uwarunkowana dostępnością funkcji łącznej gęstości rozkładu prawdopodobieństwa parametrów konstrukcji i obciążeń. Od precyzyjnego modelu stochastycznego zależy wiarygodność szacowanych wartości

prawdopodobieństwa awarii. Niestety bardzo często, ze względu na brak odpowiednich danych statystycznych, model ten sprowadza się jedynie do pierwszych dwóch momentów zmiennych losowych. Używanie w takim przypadku aparatu optymalizacji niezawodnościowej wydaje się być nieuprawnione.

Sformułowaniem optymalizacji niedeterministycznej, które lepiej dopasowane jest do realiów projektowych, jest optymalizacja odpornościowa. W odróżnieniu od optymalizacji niezawodnościowej, sformułowanie to nie wymaga szacowania prawdopodobieństwa awarii. Losowy charakter odpowiedzi konstrukcji uwzględniany jest poprzez definicje funkcji celu oraz ograniczeń, które zawierają wartości średnie oraz wariancje tych odpowiedzi. Zadanie optymalizacji odpornościowej jest zazwyczaj zadaniem optymalizacji wielokryterialnej ponieważ oprócz nominalnej (średniej) wartości funkcji celu, minimalizacji podlega również jej rozrzut.

Złożoność obliczeniowa zadania optymalizacji odpornościowej wiąże się z koniecznością zastosowania odpowiednich aproksymacji nieznanymi odpowiedziami konstrukcji, a także z użyciem efektywnych metod szacowania momentów statystycznych. Zagadnienia te omówione zostaną w rozdziałach 3 i 4. W niniejszym rozdziale zaprezentowano trzy możliwe strategie aproksymacji: podstawową strategię aproksymacji, strategię aproksymacji statystyk i strategię aproksymacji funkcji celu i ograniczeń odpornościowych. Wszystkie trzy zostały zaimplementowane w tworzonej przez autora bibliotece optymalizacyjnej programu STAND, zob. rozdział 8. Na podstawie licznych testów stwierdzić można, że ostatnia strategia (omówiona w punkcie 2.4.2.3), wydaje się być rozwiązaniem najbardziej efektywnym.

Metody powierzchni odpowiedzi w analizie eksperymentów komputerowych

W przypadku wielu inżynierskich problemów optymalizacji, jak również analizy niezawodności, liczba możliwych do przeprowadzenia obliczeń wartości funkcji (celu, ograniczeń projektowych, awarii) jest w znaczny sposób limitowana bądź to przez koszt analizy, bądź też przez czas jej trwania. Przykładowo, czas potrzebny do przeprowadzenia pojedynczej symulacji MES zderzenia pojazdów albo głębokiego tłoczenia blachy może niejednokrotnie wynosić od kilku do kilkudziesięciu godzin. Jednym ze sposobów na przewyciężenie tych trudności jest zastąpienie rzeczywistych, trudnych obliczeniowo niejawnych funkcji wchodzących w skład analizowanego problemu przez analityczne metamodele - powierzchnie odpowiedzi (ang. *response surface*). Powierzchnie te tworzone są poprzez odpowiednie dopasowanie funkcji aproksymujących do zbioru punktów eksperymentalnych.

Problematyka tworzenia dokładnych, a jednocześnie efektywnych obliczeniowo metamodeli jest współcześnie przedmiotem zainteresowania bardzo szerokiego grona badaczy, reprezentujących różnorodne gałęzie nauki oraz przemysłu. Obecny stan wiedzy oraz główne trendy w tych badaniach opisane zostały w ciekawych pracach przeglądowych [220, 222]. W niniejszym rozdziale zaprezentowane zostaną wybrane techniki powierzchni odpowiedzi, które mogą być przydatne w aproksymacji szerokiej klasy funkcji nieliniowych. Szczególnie dużo miejsca poświęcone zostanie metodzie krigingu. Podstawową ideę tej metody, mającej swoją genezę w geostatystyce, zob. [36, 168], określić można jako modelowanie nieznannej funkcji za pomocą realizacji procesu losowego. Koncepcja ta, z pozoru dość zawiła i nienaturalna, okazuje się być w istocie bardzo ciekawym rozwiązaniem. Dopasowując realizację procesu do danych doświadczalnych ustala się

typowe zachowanie funkcji, takie jak oczekiwany przyrost jej wartości przy różnych przyrostach poszczególnych zmiennych. Inaczej mówiąc, wartość funkcji w nowym punkcie wyznaczana jest tak, aby jak najlepiej pasować do typowej realizacji procesu.

O ile w swych pierwotnych, geologicznych zastosowaniach technika krigingu dostosowana była do zagadnień dwu lub trójwymiarowych oraz do danych obarczonych pewnym szumem (co jest zrozumiałe, jeśli modeluje się zawartość złota w skale jako funkcję jej położenia w analizowanym złożu), o tyle w przypadku symulacji komputerowych mamy zazwyczaj do czynienia z większą liczbą wymiarów, a rozpatrywane funkcje mają charakter deterministyczny.

Jak dotąd, podstawowym narzędziem do tworzenia powierzchni odpowiedzi jest analiza regresji liniowej, zob. [179]. Prezentując dla kompletności wyводу w podrozdziale 3.1 podstawowe równania tej metody, podkreślono jedno z zasadniczych założeń analizy regresji, które nie jest prawdziwe w przypadku eksperymentów numerycznych. Utworzona w celu analizy wyników rzeczywistych (nie wirtualnych) doświadczeń metoda regresji zakłada, że błędy poszczególnych eksperymentów są niezależnymi zmiennymi losowymi o rozkładzie normalnym. W przypadku danych pochodzących z analizy komputerowej, twierdzenie, iż błędy aproksymacji nie są skorelowane jest zazwyczaj nieuprawnione. Problemy te zostaną szczegółowo omówione w dalszej części tekstu.

W podrozdziale 3.2 poświęconym krigingowi pokażemy, że jest to w istocie metoda interpolacji “przeprowadzająca” tworzoną powierzchnię odpowiedzi przez wszystkie punkty eksperymentalne. W przypadku optymalizacji odpornościowej, ze względu na pewien rozrzut wartości funkcji związany ze stosowaną techniką symulacji losowych, zasadne jest jednak modelowanie funkcji celu i ograniczeń za pomocą aproksymujących/wygładzonych powierzchni odpowiedzi. Za pracą [213] zaprezentowana zostanie modyfikacja podstawowych równań metody krigingu mająca na celu otrzymanie wersji aproksymującej. Zachowuje się w ten sposób doskonale odwzorowanie właściwego kształtu funkcji, odfiltrując jednocześnie niepożądany szum.

Na koniec przedstawiona zostanie metoda aproksymacji lokalnej wykorzystująca koncepcję tzw. ważonej regresji liniowej. Podejście to znane jest pod angielską nazwą *moving least squares* - MLS, co przetłumaczyć można jako lokalna metoda najmniejszych kwadratów. Ze względu na łatwość implementacji oraz stosunkowo dużą szybkość działania, MLS zastosowano w zaproponowanym adaptacyjnym algorytmie poszukiwania punktu projektowego (podrozdział 5.1).

3.1. Liniowa analiza regresji

Załóżmy, że dysponujemy wynikami N eksperymentów przeprowadzonych w celu zbadania nieznannej zależności $y = f(\mathbf{x})$, gdzie \mathbf{x} są współrzędnymi punktów w przestrzeni n wymiarowej. Nasze dane tworzy zatem N par $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, N$, gdzie $y_i = f(\mathbf{x}_i)$. Jednym z podstawowych zadań liniowej analizy regresji, zob. [58, 178], jest znalezienie funkcji aproksymującej $\hat{y} = \hat{f}(\mathbf{x})$ (powierzchni regresji) przyjmując, że dane te wygenerowano przy pomocy następującego modelu:

$$f(\mathbf{x}) = \sum_{i=0}^{p-1} B_i q_i(\mathbf{x}) + \varepsilon, \quad (3.1)$$

gdzie B_i , $i = 0, 1, \dots, p-1$, są nieznanymi parametrami (współczynnikami) równania, $q_i(\mathbf{x})$ są funkcjami bazowymi, przy czym $q_0(\mathbf{x}) = 1$, a ε jest błędem uwzględniającym rozrzut wartości y_i poza powierzchnią regresji. Można wobec tego zapisać:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & q_1(\mathbf{x}_1) & \cdots & q_{p-1}(\mathbf{x}_1) \\ 1 & q_1(\mathbf{x}_2) & \cdots & q_{p-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & q_1(\mathbf{x}_N) & \cdots & q_{p-1}(\mathbf{x}_N) \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad (3.2)$$

lub w postaci macierzowej

$$\mathbf{y} = \mathbf{Q}\mathbf{B} + \boldsymbol{\varepsilon}, \quad (3.3)$$

gdzie $\mathbf{B}_{p \times 1}$ jest wektorem współczynników, $\mathbf{y}_{N \times 1}$ jest wektorem wartości funkcji $f(\mathbf{x}_i)$, $i = 1, \dots, N$, odpowiadających N przeprowadzonym eksperymentom, a $\boldsymbol{\varepsilon}_{N \times 1}$ jest losowym wektorem błędów. Macierz $\mathbf{Q}_{N \times p}$ zawiera wartości funkcji bazowych $q_i(\mathbf{x}_j)$, $i = 0, \dots, p-1$, $j = 1, \dots, N$. Pierwsza kolumna \mathbf{Q} odpowiada stałej równania regresji B_0 . W zależności od przyjętego typu aproksymacji oraz stosowanego planu eksperymentów, jako funkcje bazowe wybiera się najczęściej wielomiany pierwszego lub drugiego stopnia. Nazwa "regresja liniowa" świadczy zatem o zależności funkcyjnej względem współczynników \mathbf{B} , a nie o rodzaju funkcji bazowych, które mogą być nieliniowe. Oczywiście, liczba funkcji bazowych (parametrów modelu) p może być nieograniczona, należy jednak pamiętać, że powinna być znacznie mniejsza od liczby punktów eksperymentalnych.

Zakłada się, że $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, a więc, że błędy losowe mają rozkład normalny o zerowych wartościach oczekiwanych i stałej wariancji oraz że są nieskorelowane. Jak to zostanie dokładniej omówione poniżej, założenia te potrzebne

są aby zagwarantować odpowiednie własności estymatorów \mathbf{b} parametrów \mathbf{B} , wyznaczanych poprzez minimalizację reszt z modelu, zdefiniowanych jako

$$\mathbf{r} = \mathbf{y} - \mathbf{Q}\mathbf{b}. \quad (3.4)$$

Minimalizowana jest funkcja celu $S(\mathbf{b})$ równa sumie kwadratów reszt

$$S(\mathbf{b}) = \mathbf{r}^T \mathbf{r} = (\mathbf{y} - \mathbf{Q}\mathbf{b})^T (\mathbf{y} - \mathbf{Q}\mathbf{b}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{Q}^T \mathbf{y} + \mathbf{b}^T \mathbf{Q}^T \mathbf{Q} \mathbf{b}. \quad (3.5)$$

Z warunku pierwszego rzędu dla $\min S(\mathbf{b})$, zerowania się pochodnych $S(\mathbf{b})$ względem b_i , $i = 0, \dots, p-1$, otrzymuje się tzw. układ równań normalnych,

$$(\mathbf{Q}^T \mathbf{Q}) \mathbf{b} = \mathbf{Q}^T \mathbf{y}. \quad (3.6)$$

którego rozwiązaniem są estymatory $\mathbf{b}_{p \times 1} = \{b_0, b_1, \dots, b_{p-1}\}$,

$$\mathbf{b} = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{y}. \quad (3.7)$$

Warunki drugiego rzędu istnienia minimum dotyczą dodatniej określoności macierzy hesjanu funkcji $S(\mathbf{b})$

$$\frac{\partial^2 S(\mathbf{b})}{\partial \mathbf{b} \partial \mathbf{b}^T} = 2\mathbf{Q}^T \mathbf{Q}. \quad (3.8)$$

Ze względu na liniową niezależność równań normalnych, warunek ten jest spełniony dzięki dodatniej określoności macierzy $(\mathbf{Q}^T \mathbf{Q})_{p \times p}$. W przypadku gdy równania normalne są zależne, macierz $\mathbf{Q}^T \mathbf{Q}$ staje się osobliwa. Macierz ta może być również źle uwarunkowana. W tej sytuacji często rozwiązaniem jest użycie metody dekompozycji na wartości singularne (ang. *singular value decomposition* - SVD), zob. [197], i redukcja wyjściowego modelu poprzez eliminację tych wierszy i kolumn macierzy, które powodują jej osobliwość.

Ostatecznie, powierzchnia regresji dopasowana do danych za pomocą metody najmniejszych kwadratów przybiera postać

$$\hat{y} = \hat{f}(\mathbf{x}) = [1 \quad q_1(\mathbf{x}) \quad \dots \quad q_{p-1}(\mathbf{x})] \mathbf{b} = \mathbf{q}^T(\mathbf{x}) \mathbf{b}. \quad (3.9)$$

W praktyce, każda powierzchnia odpowiedzi stanowi jedynie aproksymację rzeczywistej funkcyjnej zależności pomiędzy zmiennymi wejścia i wyjścia (często spotyka się określenia zmienna wyjaśniająca i wyjaśniana). Ponieważ zawsze mamy do czynienia z pewnym niedopasowaniem funkcji (3.9) do danych eksperymentalnych, konieczna jest weryfikacja poprawności tego rozwiązania. Wprowadzając następujące definicje:

całkowita suma kwadratów (Total Sum of Squares)

$$TSS = (\mathbf{y} - \bar{\mathbf{y}})^T (\mathbf{y} - \bar{\mathbf{y}}), \quad (3.10)$$

wyjaśniona suma kwadratów (Explained Sum of Squares)

$$ESS = (\hat{\mathbf{y}} - \bar{\mathbf{y}})^T(\hat{\mathbf{y}} - \bar{\mathbf{y}}), \quad (3.11)$$

suma kwadratów reszt (Residual Sum of Squares)

$$RSS = \mathbf{r}^T \mathbf{r}, \quad (3.12)$$

gdzie $\hat{\mathbf{y}} = \mathbf{Q}\mathbf{b}$, $\bar{\mathbf{y}} = \mathbf{e}\bar{y}$, $\mathbf{e}_{N \times 1} = \{1, \dots, 1\}$, a \bar{y} jest wartością średnią \mathbf{y} , można pokazać [58], że

$$TSS = ESS + RSS. \quad (3.13)$$

Sumy kwadratów (3.10)–(3.12) wykorzystuje się w definicji tzw. współczynnika determinacji R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}, \quad (3.14)$$

przy czym

$$0 \leq R^2 \leq 1. \quad (3.15)$$

Po dokonaniu pewnych podstawowych przekształceń współczynnik ten przedstawić można w alternatywnej postaci jako

$$R^2 = \frac{\mathbf{b}^T \mathbf{Q}^T \mathbf{y} - N\bar{y}^2}{\mathbf{y}^T \mathbf{y} - N\bar{y}^2}. \quad (3.16)$$

R^2 interpretuje się jako procent zmienności wartości funkcji y_i , $i = 1, \dots, N$, która wyjaśniona jest przez model. Współczynnik R^2 jest równy jeden jeśli $\mathbf{r} = \mathbf{0} \Rightarrow \mathbf{y} = \mathbf{Q}\mathbf{b}$. Wartości R^2 bliskie jedności są zazwyczaj oznaką dobrego dopasowania modelu do obserwacji, a wartości małe (znacząco mniejsze od 1) świadczą o złym doborze modelu.

Statystyka R^2 ma jednak istotną wadę. Dodanie kolejnych zmiennych do modelu daje w rezultacie model o wyższym R^2 , nawet jeśli zmienne te są nieistotne, a ich dołączenie prowadzi do zwiększenia wariancji estymatora \mathbf{b} . Z tego powodu R^2 można stosować jedynie jako statystykę opisową. Nie może ona służyć do ustalenia, które zmienne w modelu są istotne statystycznie. Lepszą miarą dopasowania jest skorygowane R^2 , zob. [82, 178],

$$\bar{R}^2 = 1 - \frac{N-1}{N-p}(1 - R^2), \quad (3.17)$$

które zaprojektowano w ten sposób, że wartość \bar{R}^2 spada, gdy wartość bezwzględna statystyki t dla usuniętej zmiennej jest większa od 1.

Na koniec, warto przypomnieć założenia klasycznego modelu regresji liniowej oraz to, co te założenia implikują w przypadku estymatorów uzyskanych przy pomocy metody najmniejszych kwadratów. Przyjmując, że wartość oczekiwana błędu losowego jest równa zero $\mathbb{E}(\varepsilon_i) = 0$, $i = 1, \dots, N$, błędy te są nieskorelowane $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$ oraz, że wariancja błędu pozostaje stała $\text{Var}(\varepsilon_i) = \sigma^2$ twierdzenie Gaussa-Markowa, zob. [81, 178], mówi, że wektor \mathbf{b} dany wzorem (3.7) jest najlepszym (o najmniejszej wariancji) liniowym i nieobciążonym estymatorem parametrów \mathbf{B} w klasycznym modelu regresji liniowej. Pod pojęciem estymatora liniowego rozumie się taki estymator, który można przedstawić w postaci $\mathbf{A}\mathbf{y}$, gdzie \mathbf{A} jest dowolną $p \times N$ elementową macierzą nielosowych współczynników. Zakładając ponadto, że $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ zakładamy rozkład prawdopodobieństwa obserwowanych wartości \mathbf{y} jako $\mathbf{y} \sim N(\mathbf{Q}\mathbf{B}, \sigma^2 \mathbf{I})$.

W przypadkach kiedy istnieją przesłanki ku temu, aby preferować wyniki wybranych eksperymentów lub gdy nie ma się zaufania do pewnych wyników, modyfikuje się założenie dotyczące postaci macierzy kowariancji błędów. Przyjmując nadal, że jest to macierz diagonalna, nie zakłada się już stałości wariancji. Do rozwiązania tak postawionego zadania liniowej regresji stosuje się wtedy ważoną metodę najmniejszych kwadratów. Wyprowadzenie układu równań normalnych tej metody można znaleźć w [58]. Bazuje ono na transformacji wyników eksperymentów na takie zmienne, które spełniają podstawowe założenia metody najmniejszych kwadratów, aby móc następnie zastosować standardowe procedury rozwiązania. Poszukiwany wektor estymatorów \mathbf{b} można wtedy ponownie wyrazić za pomocą pierwotnych danych. Rozwiązanie układu równań normalnych (3.7) przyjmuje następującą postać

$$\mathbf{b} = (\mathbf{Q}^T \mathbf{W} \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{W} \mathbf{y}, \quad (3.18)$$

gdzie \mathbf{W} jest kwadratową macierzą wagową

$$\mathbf{W} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_N \end{bmatrix}. \quad (3.19)$$

a w_i , $i = 1, \dots, N$ są wagami odpowiadającymi poszczególnym eksperymentom. Przyjmując wszystkie wagi równe 1 otrzymujemy oczywiście standardowe rozwiązanie metody najmniejszych kwadratów.

3.2. Kriging

Przed zaprezentowaniem formalnego wyprowadzenia równań krigingu jako metody aproksymacji funkcji przy pomocy realizacji procesów losowych, warto jest dokładniej wyjaśnić dlaczego metoda ta – w przeciwieństwie do przedstawionej w poprzednim podrozdziale liniowej regresji – tak dobrze nadaje się do budowania modeli opartych na deterministycznych eksperymentach numerycznych. Bardzo interesujące i przystępne omówienie tego tematu znaleźć można w pracy [112]. Poniżej przytoczmy jedynie najważniejsze z przedstawionych tam zagadnień.

Podstawowa trudność, jaką napotykamy stosując liniową analizę regresji dotyczy doboru odpowiednich funkcji bazowych $q_i(\mathbf{x})$, $i = 0, \dots, p - 1$, zob. (3.1). Zwiększanie ich liczby w celu lepszego dopasowywania kształtu funkcji do punktów eksperymentalnych pociąga za sobą nieunikniony wzrost liczby parametrów, do wyznaczenia których potrzeba z kolei coraz większej liczby eksperymentów. Ponadto, w przypadku powtarzalnych eksperymentów numerycznych mocno dyskusyjne wydaje się być założenie dotyczące niezależności błędów ε występujących w równaniu (3.3). Błędy, z którymi mamy tu do czynienia wynikają całkowicie z nieodpowiedniego doboru bazowych funkcji aproksymujących, a nie z błędu pomiarowego lub szumu (przykłady symulacji komputerowych, gdzie z pewnych względów korzystne jest uwzględnienie sztucznego “składnika szumowego” oraz zastosowań specyficznych dla optymalizacji odpornościowej omówione zostaną oddzielnie w dalszej części rozdziału). Tak więc można stwierdzić, że błąd niedopasowania powierzchni odpowiedzi jest funkcją położenia punktu, $\varepsilon_i = \varepsilon(\mathbf{x}_i)$. Jeśli funkcja $f(\mathbf{x})$ jest ciągła, wtedy $\varepsilon(\mathbf{x})$ jest również funkcją ciągłą. Wynika z tego, że jeżeli dwa punkty \mathbf{x}_i i \mathbf{x}_j położone są blisko siebie, wtedy także błędy $\varepsilon(\mathbf{x}_i)$ i $\varepsilon(\mathbf{x}_j)$ powinny mieć zbliżone wartości. Zamiast zatem zakładać ich statystyczną niezależność, bardziej racjonalne jest przyjąć, że wielkości błędów są skorelowane, a korelacja ta jest wysoka gdy \mathbf{x}_i i \mathbf{x}_j znajdują się w bliskim sąsiedztwie, zaś niska gdy są od siebie oddalone.

Spostrzeżenie to jest fundamentem technik aproksymacji wykorzystujących teorię procesów losowych. W podejściu tym wprowadza się funkcję korelacji błędów zależną od odległości pomiędzy dwoma punktami eksperymentalnymi. Zazwyczaj nie używa się jednak zwykłej euklidesowej miary odległości, przyjmując następującą odległość ważoną:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^n \theta_k |x_{ik} - x_{jk}|^2, \quad (3.20)$$

gdzie $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ jest wektorem nieujemnych parametrów wagowych, a x_{ik} jest k -tą współrzędną punktu \mathbf{x}_i . Przy pomocy powyższej definicji korelację pomiędzy błędami aproksymacji w punktach \mathbf{x}_i i \mathbf{x}_j wyrazić można jako

$$R[\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j)] = \exp[-d(\mathbf{x}_i, \mathbf{x}_j)] = \exp\left[-\sum_{k=1}^n \theta_k |x_{ik} - x_{jk}|^2\right]. \quad (3.21)$$

Zgodnie z założeniami, wartość funkcji $R[\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j)]$ dąży do jedności gdy \mathbf{x}_i dąży do \mathbf{x}_j oraz maleje do zera gdy odległość między tymi punktami rośnie. Oczywiście można zaproponować wiele innych funkcji korelacji spełniających powyższe wymagania, jednak funkcja (3.21) jest najczęściej spotykaną w literaturze, zob. [219, 221]. Parametry θ_k , $k = 1, \dots, n$, mogą być interpretowane jako miara “aktywności” danej zmiennej x_k . Duża wartość odpowiedniego parametru oznacza, że nawet mała różnica $|x_{ik} - x_{jk}|$ prowadzi do istotnych zmian wartości aproksymowanej funkcji z uwagi na wymuszenie szybkiego spadku korelacji w tym kierunku.

Powyższy sposób opisu korelacji pozwala postulować, iż wyniki eksperymentów $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, N$, można wyjaśnić przy pomocy następującego modelu

$$y_i = f(\mathbf{x}_i) = B + \varepsilon(\mathbf{x}_i), \quad i = 1, \dots, N, \quad (3.22)$$

gdzie B jest stałą, błędy $\varepsilon(\mathbf{x}_i)$ są zmiennymi losowymi o rozkładzie normalnym $N(0, \sigma^2)$, a korelacja między nimi nie jest zerowa, lecz dana za pomocą wzoru (3.21). Zapisując inaczej, poszukujemy parametrów funkcji aproksymującej postaci

$$y = f(\mathbf{x}) = B + H(\mathbf{x}), \quad (3.23)$$

gdzie $H(\mathbf{x})$ jest realizacją gaussowskiego procesu losowego o zerowej wartości średniej i nieznannej wariancji σ^2 oraz funkcji kowariancji. Parametry, które należy wyznaczyć na podstawie przeprowadzonych eksperymentów to B , σ^2 oraz współczynniki θ_k , $k = 1, \dots, n$, w definicji funkcji korelacji (3.21). O ile stałą B można traktować jako pewien najprostszy model globalnej aproksymacji n wymiarowej funkcji, o tyle $H(\mathbf{x})$ zapewnia “dociągnięcie” funkcji do punktów eksperymentalnych.

W metodzie kriginu proces losowy $H(\mathbf{x})$ dyskretyzowany jest przy pomocy jego realizacji w punktach \mathbf{x}_i , $i = 1, \dots, N$, zgodnie z omówioną w dodatku C techniką optymalnej aproksymacji liniowej. Wprowadzając następujące definicje

macierzy korelacji i macierzy kowariancji:

$$\mathbf{R} = \begin{bmatrix} R(\mathbf{x}_1, \mathbf{x}_1) & R(\mathbf{x}_1, \mathbf{x}_2) & \cdots & R(\mathbf{x}_1, \mathbf{x}_N) \\ R(\mathbf{x}_2, \mathbf{x}_1) & R(\mathbf{x}_2, \mathbf{x}_2) & \cdots & R(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \cdots & \vdots \\ R(\mathbf{x}_N, \mathbf{x}_1) & R(\mathbf{x}_N, \mathbf{x}_2) & \cdots & R(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}, \quad (3.24)$$

$$\mathbf{C} = \sigma^2 \mathbf{R}, \quad (3.25)$$

gdzie użyto skróconego zapisu $R(\mathbf{x}_i, \mathbf{x}_j) = R(H(\mathbf{x}_i), H(\mathbf{x}_j)) = R(\varepsilon(\mathbf{x}_i), \varepsilon(\mathbf{x}_j))$, oraz definiując wektor korelacji pomiędzy realizacją pola w punkcie \mathbf{x} a realizacjami w punktach eksperymentalnych jako

$$\mathbf{r}^T(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}_1), \dots, R(\mathbf{x}, \mathbf{x}_N)], \quad (3.26)$$

wykorzystując wzory (C.34), (3.23) i (3.25) oraz pamiętając, że $H(\mathbf{x})$ jest jednorodnym procesem losowym o zerowej wartości oczekiwanej, otrzymuje się wyrażenie na dyskretyzację tego procesu w postaci

$$H(\mathbf{x}) \xrightarrow{\text{dyskretyzacja}} \hat{H}(\mathbf{x}) = \mathbf{r}^T(\mathbf{x}) \mathbf{R}^{-1} (\mathbf{y} - \mathbf{e}\hat{B}), \quad (3.27)$$

gdzie \mathbf{y} jest N elementowym wektorem wartości funkcji $f(\mathbf{x}_i)$, $i = 1, \dots, N$, \mathbf{e} jest wektorem jedynek, a \hat{B} jest estymatorem parametru B . Ostatecznie wzór (3.23) przybiera postać

$$\hat{y}(\mathbf{x}) = \hat{B} + \mathbf{r}^T(\mathbf{x}) \mathbf{R}^{-1} (\mathbf{y} - \mathbf{e}\hat{B}). \quad (3.28)$$

Metodą estymacji nieznanymi parametrów $\boldsymbol{\theta}$, \hat{B} i $\hat{\sigma}^2$ jest opisana między innymi w pracach [70, 76] metoda największej wiarygodności (ang. *maximum likelihood estimation* - MLE). W podejściu tym zakłada się, że wektor $(\mathbf{y} - \mathbf{e}\hat{B})$ jest realizacją gaussowskiego wektora losowego i dlatego dąży się do znalezienia takich wartości parametrów, które maksymalizują następującą funkcję wiarygodności:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{\sqrt{(2\pi\hat{\sigma}^2)^N |\mathbf{R}(\boldsymbol{\theta})|}} \exp \left[-\frac{(\mathbf{y} - \mathbf{e}\hat{B})^T \mathbf{R}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{e}\hat{B})}{2\hat{\sigma}^2} \right]. \quad (3.29)$$

Dla wygody, w wyprowadzeniach korzysta się z logarytmu powyższej funkcji otrzymując

$$l(\boldsymbol{\theta}|\mathbf{y}) = \frac{-(\mathbf{y} - \mathbf{e}\hat{B})^T \mathbf{R}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{e}\hat{B})}{2\hat{\sigma}^2} - \frac{N \ln(2\pi\hat{\sigma}^2) + \ln(|\mathbf{R}(\boldsymbol{\theta})|)}{2}. \quad (3.30)$$

Z warunku pierwszego rzędu, przyrównując pochodne funkcji (3.30) względem \hat{B} oraz $\hat{\sigma}^2$ do zera otrzymuje się wyrażenia na trend \hat{B} i wariancję $\hat{\sigma}^2$ w funkcji parametrów θ

$$\hat{B}(\theta) = [\mathbf{e}^T \mathbf{R}^{-1}(\theta) \mathbf{e}]^{-1} \mathbf{e}^T \mathbf{R}^{-1}(\theta) \mathbf{y}, \quad (3.31)$$

$$\hat{\sigma}^2(\theta) = \frac{1}{N} [(\mathbf{y} - \mathbf{e}\hat{B})^T \mathbf{R}^{-1}(\theta) (\mathbf{y} - \mathbf{e}\hat{B})]. \quad (3.32)$$

Ostatecznie, parametry funkcji korelacji θ znajdowane są jako rozwiązanie następującego problemu maksymalizacji logarymicznej funkcji wiarygodności:

$$\text{znajdź: } \theta, \quad (3.33)$$

$$\text{maksymalizujące: } - \frac{N \ln(\hat{\sigma}(\theta)^2) + \ln(|\mathbf{R}(\theta)|)}{2}, \quad (3.34)$$

$$\text{pod warunkiem: } 0 \leq \theta_i \leq \infty, \quad i = 1, \dots, n. \quad (3.35)$$

Mamy zatem do czynienia z n wymiarowym zadaniem optymalizacji. Należy zauważyć, że model (3.28) można zbudować dla dowolnych wartości współczynników θ spełniających warunki (3.35), rozwiązanie zadania (3.33)–(3.35) zapewnia jednak najlepszy dobór tych parametrów. W praktyce, poszukując optymalnego wektora θ macierz korelacji \mathbf{R} może stać się źle uwarunkowana. W pracach [157, 158] Lophaven i in. opisali szereg technik numerycznych pozwalających przezwyciężyć te trudności.

Problem (3.33)–(3.35) nie jest na ogół łatwy do rozwiązania ze względu na lokalne minima funkcji wiarygodności lub też rozległy obszar parametrów bliższych wartościom optymalnym. W celu znalezienia maximum globalnego bardzo często stosowane są metody losowego przeszukiwania, algorytmy ewolucyjne lub heurystyczne. Na przykład w pracy [219] zastosowano algorytm symulowanego wyżarzania zaproponowany przez Goffe'go i in. [77]. Algorytm ten udostępniony jest również w Internecie na stronie <http://netlib2.cs.utk.edu/opt>. O ile wymienione metody teoretycznie są w stanie wyznaczyć maksimum globalne, o tyle są one zazwyczaj ogromnie kosztowne i wymagają o rzędy wielkości większej liczby obliczeń funkcji (3.34) niż w przypadku większości metod optymalizacji lokalnej, które oczywiście nie gwarantują znalezienia optimum globalnego. Rozważania na temat przydatności różnych algorytmów optymalizacji w wyznaczaniu najlepszych wartości parametrów θ znaleźć można w pracy [70]. W algorytmie metody krigingu zaimplementowanym w programie STAND zastosowano dwuetapową technikę rozwiązywania zadania (3.33)–(3.35). Na początku, za pomocą umiarkowanej liczby iteracji algorytmu symulowanego wyżarzania znajdowany jest punkt startowy dla algorytmu nieliniowego sympleksu Nedlera-Meed'a, zob. [184, 243], który kontynuowany jest następnie aż do osiągnięcia zbieżności.

Analizując wzór (3.28) można zauważyć, że założenie braku korelacji pomiędzy wartością funkcji w punkcie \mathbf{x} , a wartościami w punktach eksperymentalnych ($\mathbf{r} = \mathbf{0}$) prowadzi w konsekwencji do aproksymacji $\hat{y}(\mathbf{x}) = \hat{B}$. Z drugiej strony, aby przekonać się, że założenie istnienia takiej korelacji daje w wyniku funkcję interpolującą, obliczmy zgodnie ze wzorem (3.28) prognozowaną wartość funkcji w jednym z punktów z bazy eksperymentów, np. $\mathbf{x} = \mathbf{x}_i$. W tym przypadku wektor \mathbf{r} będzie taki sam jak i -ta kolumna macierzy \mathbf{R} . Oznaczając ją jako \mathbf{R}_i dostajemy

$$\mathbf{r}^T(\mathbf{x}_i)\mathbf{R}^{-1} = [\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_i)]^T = (\mathbf{R}^{-1}\mathbf{R}_i)^T = \mathbf{e}_i^T, \quad (3.36)$$

gdzie wektor \mathbf{e}_i jest wektorem o zerowych elementach za wyjątkiem elementu i równego 1. Na podstawie równania (3.28) otrzymujemy wtedy

$$\hat{y}(\mathbf{x}_i) = \hat{B} + \mathbf{e}_i^T(\mathbf{y} - \mathbf{e}\hat{B}) = \hat{B} + (y_i - \hat{B}) = y_i, \quad (3.37)$$

co potwierdza, iż mamy do czynienia z interpolacją.

W pracy [210] wyprowadzono wzór na błąd średniokwadratowy (ang. *Mean Squared Error* - MSE) aproksymacji metodą krigingu

$$\text{MSE}[\hat{y}(\mathbf{x})] = \sigma^2 \left[1 - \mathbf{r}^T\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{e}^T\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{e}^T\mathbf{R}^{-1}\mathbf{e}} \right]. \quad (3.38)$$

Postępując podobnie jak w przypadku poprzedniego wyprowadzenia można pokazać, że błąd średniokwadratowy obliczony w punktach zbioru danych wynosi 0.

Do tego miejsca przyjmowaliśmy, że rozpatrywane przez nas dane eksperymentalne nie są obarczone żadną niepewnością. Chociaż takie założenie jest prawdziwe w przypadku większości symulacji numerycznych (wielokrotne powtarzanie obliczeń, na tym samym komputerze i dla takich samych danych wejściowych prowadzi do tych samych wyników) to istnieją zastosowania kiedy wyniki, z którymi mamy do czynienia, obarczone są pewnym szumem i gdy nieracjonalnym jest stosowanie interpolujących powierzchni odpowiedzi. Bardzo dobrym przykładem takiej sytuacji jest optymalizacja odpornościowa, gdzie punkty budujące bazę eksperymentów są wynikami symulacji Monte Carlo, np. momentami statystycznymi funkcji celu i ograniczeń. Ze względu na skończoną wielkość próbek używanych w symulacjach obserwuje się rozrzut tych wyników. Czasem też, wobec braku możliwości właściwego modelowania wpływu wybranych losowych parametrów na wartość analizowanej funkcji, wpływ ten uwzględnia się poprzez dodanie do funkcji składnika szumowego.

Ponieważ zaprezentowana metoda krigingu jest metodą interpolacji, dlatego aby uzyskać kriging w wersji aproksymacyjnej wprowadza się następującą

modyfikację funkcji korelacji, zob. [213],

$$R[\mathbf{x}_i, \mathbf{x}_j] = (1 - \eta) \exp \left[- \sum_{k=1}^n \theta_k |x_{ik} - x_{jk}|^2 \right], \quad (3.39)$$

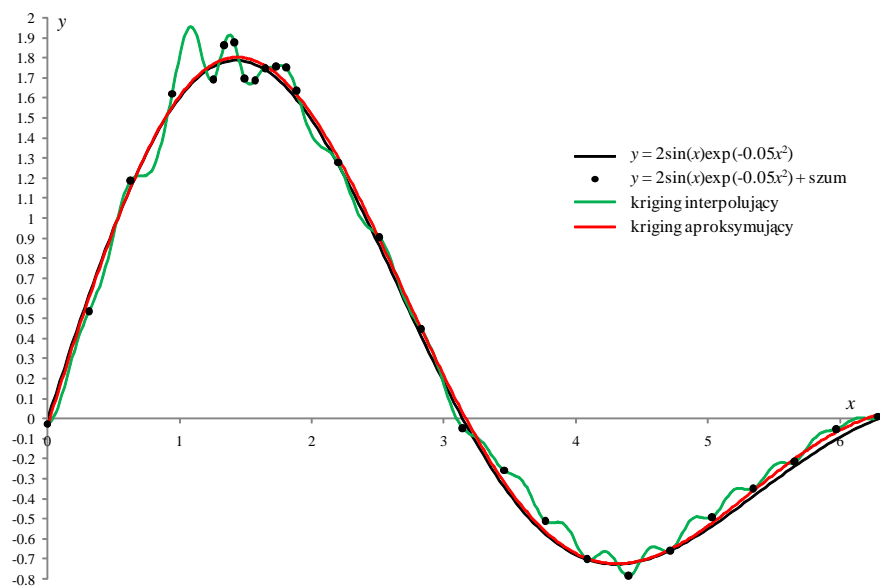
gdzie $0 \leq \eta < 1$ ma w zamyśle modelować, to co w angielskojęzycznej literaturze z zakresu geostatystyki nazywane jest *nugget effect*, czyli efekt grudki złota. Dotyczy on nieciągłości funkcji kowariancji gdy $|\mathbf{x}_i - \mathbf{x}_j|$ dąży do zera. Parametr η dobierany jest w procesie maksymalizacji funkcji wiarygodności razem z parametrami $\boldsymbol{\theta}$. Zerowa wartość η oznacza, że powierzchnia odpowiedzi przechodzi przez wszystkie punkty danych. Zwiększanie η prowadzi do wygładzenia funkcji aproksymującej ponieważ wartości funkcji w punktach położonych bardzo blisko siebie są słabiej skorelowane. Dla $\eta = 1$ funkcja aproksymująca degeneruje się do członu stałego \hat{B} . Należy zaznaczyć, że współczynnik skalujący $(1 - \eta)$ nie jest używany w przypadku wyrazów diagonalnych macierzy \mathbf{R} , ale stosuje się go do wszystkich elementów wektora \mathbf{r} , nawet wtedy gdy punkt, w którym obliczana jest wartość aproksymacji pokrywa się z jednym z punktów z bazy danych. Robi się tak przyjmując, że punkt eksperymentalny jest doskonale skorelowany sam ze sobą, natomiast jego korelacja z innym punktem położonym w tym samym miejscu nie jest równa 1 ze względu na istniejącą niepewność/błąd pomiarowy.

Na rys. 3.1 przedstawiono porównanie działania krigingu w wersji interpolującej oraz aproksymującej, na przykładzie funkcji jednowymiarowej. Wartości funkcji w punktach eksperymentalnych obliczane są według wzoru

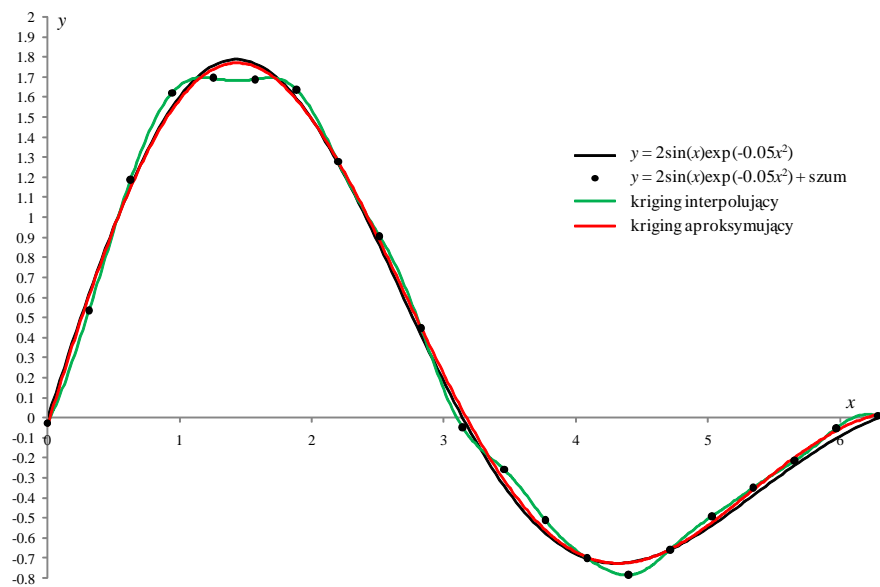
$$y(x) = 2 \sin(x) \exp(-0.05x^2) + a, \quad (3.40)$$

gdzie a jest realizacją zmiennej losowej A o rozkładzie jednostajnym w przedziale $[-0.1, 0.1]$, którą można interpretować jako błąd pomiarowy. 21 punktów wygenerowano w równych odstępach pomiędzy $x = 0$ a $x = 2\pi$, dodano również 6 punktów w pobliżu $x = \pi/2$. Jak można zauważyć, rozrzut wartości funkcji na stosunkowo małym obszarze zagęszczenia punktów dość silnie wpływa na postać funkcji korelacji, prowadząc w rezultacie do oscylacyjnego charakteru funkcji interpolującej (zielony wykres). Wprowadzając, tak jak we wzorze (3.39), parametr η do funkcji korelacji otrzymujemy w wyniku funkcję aproksymującą, która na rys. 3.1 oznaczona jest kolorem czerwonym. Modyfikacja ta prowadzi do niemal idealnego odwzorowania oryginalnej funkcji (czarny wykres).

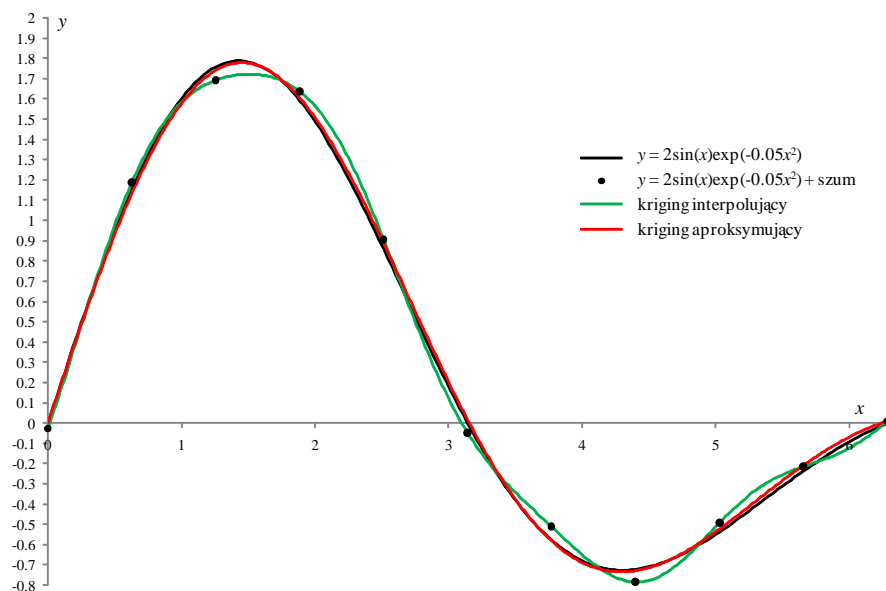
Co ciekawe, odrzucając 6 dodatkowych punktów uzyskujemy dużą poprawę dopasowania. Na rysunkach 3.2 oraz 3.3 pokazano wynik działania dwóch wersji krigingu na, odpowiednio, liczny (21 punktów) i rzadki (11 punktów) zbiorze danych eksperymentalnych.



Rys. 3.1. Różne typu krigingu w analizie “niepewnych” danych. Nierównomierny rozkład punktów eksperymentalnych.



Rys. 3.2. Różne typu krigingu w analizie “niepewnych” danych. Równomierny gęsty rozkład punktów eksperymentalnych.



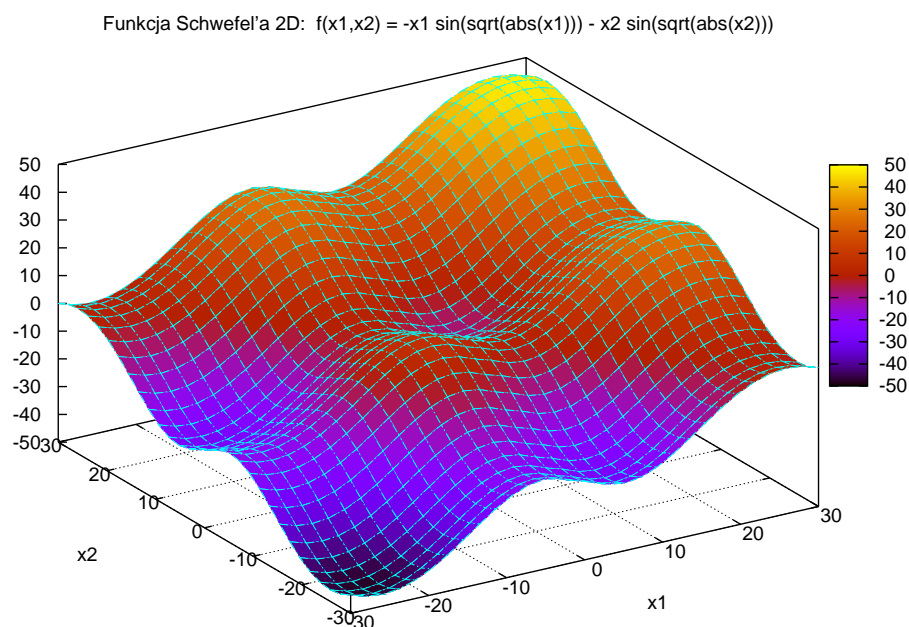
Rys. 3.3. Różne typu krigingu w analizie “niepewnych” danych. Równomierny rzadki rozkład punktów eksperymentalnych.

Zagadnienie doboru planu eksperymentów okazuje się mieć bardzo istotny wpływ na jakość tworzonej powierzchni odpowiedzi. Wielu autorów podziela opinię, zob. np. [74, 222], że tzw. plany wypełniające (ang. *space filling design*), w przypadku deterministycznych, komputerowo generowanych danych eksperymentalnych, prowadzą do minimalizacji średniego błędu kwadratowego aproksymacji metodą krigingu. Zaliczyć do nich można m.in. optymalne łącińskie hiperkostki (zob. dodatek B) - plany, gdzie punkty eksperymentalne rozmieszczone są w sposób jak najbardziej równomierny w n wymiarowej przestrzeni parametrów.

W celu zbadania wpływu optymalizacji ułożenia punktów hiperkostki łącińskiej, służącej jako plan eksperymentów w metodzie krigingu, na jakość aproksymacji, wykonano test porównujący pierwiastek średniego błędu kwadratowego dopasowania (ang. *root mean square error* - RMSE) dla różnych hiperkostek, na przykładzie aproksymacji tzw. funkcji Schwefel'a

$$f(\mathbf{x}) = \sum_{i=1}^n \left[-x_i \sin \left(\sqrt{|x_i|} \right) \right]. \quad (3.41)$$

Jest to funkcja testowa wykorzystywana często do badania algorytmów optymalizacji globalnej bez ograniczeń. Na rys. 3.4 pokazano tę funkcję dla przypadku dwuwymiarowego.



Rys. 3.4. Funkcja Schwefel'a (3.41) dla $n = 2$.

Przyjmując, że zmienne $x_i, i = 1, \dots, n$ określone są na przedziałach $[-30, 30]$, rozpatrywano przypadki $n = 2$ oraz $n = 4$. W przypadku dwóch zmiennych przyjęto dwie wielkości próbki, $N = 20$ i $N = 40$, a dla $n = 4$ próbki $N = 40$ i $N = 80$. Odsyłając do rozdziału 4 w celu szczegółowego omówienia wariantów planów wykorzystujących łacińskie hiperkostki (zob. opis na stronach 95 i 96), zaznaczmy tutaj jedynie, że plany RLH-m odpowiadają konfiguracjom bez optymalizacji rozmieszczenia punktów eksperymentalnych, a plany OLH-m gwarantują większą równomierność wypełnienia zadanej przestrzeni poprzez maksymalizowanie odległości pomiędzy punktami. Oczywiście, na potrzeby budowania planu eksperymentów stanowiącego bazę do aproksymacji funkcji, przyjmuje się, że zmienne \mathbf{x} mają rozkład jednostajny na odpowiednich przedziałach. Wykonując dla każdego rozmiaru i typu hiperkostki po 100 powtórzeń cyklu 1) generacja planu eksperymentów, 2) dopasowanie powierzchni interpolującego

krigingu, 3) obliczenie RMSE na bazie regularnej siatki 20^n punktów testowych, otrzymuje się średnie wartości RMSE, które przedstawiono w tab. 3.1.

Tablica 3.1. Aproksymacja metodą krigingu funkcji Schwefel'a (3.41). Porównanie wartości średnich RMSE dla planów eksperymentów opartych na łańciskich hiperkostkach z (OLH-m) i bez (RLH-m) optymalizacji rozmieszczenia punktów.

	$n = 2$		$n = 4$	
	$N = 20$	$N = 40$	$N = 40$	$N = 80$
RLH-m	6.27	2.08	17.43	15.44
OLH-m	4.10	0.98	14.41	13.96

Tak więc, we wszystkich analizowanych przypadkach kriging wykorzystujący plany OLH daje lepsze oszacowanie badanej funkcji niż w przypadku hiperkostek bez optymalizacji.

Warto wspomnieć jeszcze o pewnych rozwiązaniach mających na celu poprawienie dokładności aproksymacji w przypadku gdy charakter zmienności badanej funkcji nie jest jednorodny w całym obszarze jej określoności. Może zdarzyć się, że funkcja jest gwałtownie zmienna, np. oscyluje w pewnym podobszarze, a poza nim ma ustalony charakter, np. zbliżony do liniowego. W takiej sytuacji założenie dotyczące stacjonarności korelacji (3.21), która jest funkcją jedynie odległości pomiędzy punktami, a nie zależy od położenia tych punktów, może prowadzić do niekorzystnego zaburzenia kształtu poszukiwanej aproksymacji. Niektórzy autorzy, zob. np. [258], postulują wtedy wprowadzenie niestacjonarnego modelu korelacji. Pokazano, że nawet w przypadku zastosowania pewnych założeń uproszczających, model niestacjonarny prowadzi zawsze do aproksymacji nie gorszych niż model stacjonarny. Niestety, nawet w swojej najprostszej postaci, kriging bazujący na niestacjonarnej funkcji korelacji wiąże się z koniecznością wyznaczenia wartości dwa razy większej liczby nieznanymi parametrów (poprzez maksymalizację funkcji wiarygodności) niż w przypadku funkcji stacjonarnej. Jest to szczególnie trudne dla problemów wielowymiarowych i m.in. dlatego metoda ta nie została zastosowana w niniejszej pracy.

3.3. Metoda aproksymacji lokalnej

Przez aproksymację lokalną rozumiemy tu będziemy metodę wykorzystującą zaprezentowaną wcześniej koncepcję ważonej regresji liniowej (3.18). Podejście to określić można również jako lokalną metodę najmniejszych kwadratów, ang. *moving least squares* - MLS (zob. np. [34, 113]).

W metodzie MLS wartości wag w macierzy (3.19) obliczane są jako funkcja odległości pomiędzy punktem \mathbf{x} , w którym poszukuje się wartości aproksymacji, a punktami z bazy eksperymentów. Poprzez zastosowanie odpowiednio dobranej funkcji wagowej, większe znaczenie w zadaniu analizy regresji przypisywane jest tym spośród eksperymentów \mathbf{x}_i , $i = 1, \dots, N$, które znajdują się w pobliżu \mathbf{x} . Za każdym razem kiedy położenie \mathbf{x} zmienia się (np. podczas procesu optymalizacji, gdy funkcja celu przybliżana jest za pomocą powierzchni odpowiedzi), wartości wszystkich wag, w_1, \dots, w_N , są ponownie przeliczane i ponownie rozwiązywany jest układ równań regresji prowadząc do lokalnie lepiej dopasowanej aproksymacji funkcji. Jak to zostanie pokazane na przykładach poniżej, dzięki metodzie MLS możliwa jest dokładna aproksymacja silnie nieliniowych funkcji za pomocą modeli liniowych lub kwadratowych. Analogicznie do równania (3.9) postać funkcji aproksymującej dana jest następująco:

$$\hat{y} = \hat{f}(\mathbf{x}) = [1 \ q_1(\mathbf{x}) \ \cdots \ q_{p-1}(\mathbf{x})] \mathbf{b}(\mathbf{x}) = \mathbf{q}^T(\mathbf{x}) \mathbf{b}(\mathbf{x}), \quad (3.42)$$

gdzie wektor nieznanych parametrów regresji $\mathbf{b}(\mathbf{x})$, którego elementy są tym razem funkcjami położenia punktu, wyraża się wzorem (por. (3.18))

$$\mathbf{b}(\mathbf{x}) = [\mathbf{Q}^T \mathbf{W}(\mathbf{x}) \mathbf{Q}]^{-1} \mathbf{Q}^T \mathbf{W}(\mathbf{x}) \mathbf{y}. \quad (3.43)$$

Podobnie jak w przypadku “zwykłej” analizy regresji, macierz $\mathbf{Q}_{N \times p}$ zawiera wartości liniowo niezależnych funkcji bazowych $q_i(\mathbf{x}_j)$, $i = 0, \dots, p-1$, $j = 1, \dots, N$. Elementy diagonalnej macierzy $\mathbf{W}(\mathbf{x})$ obliczane są najczęściej przy pomocy następującej funkcji wagowej:

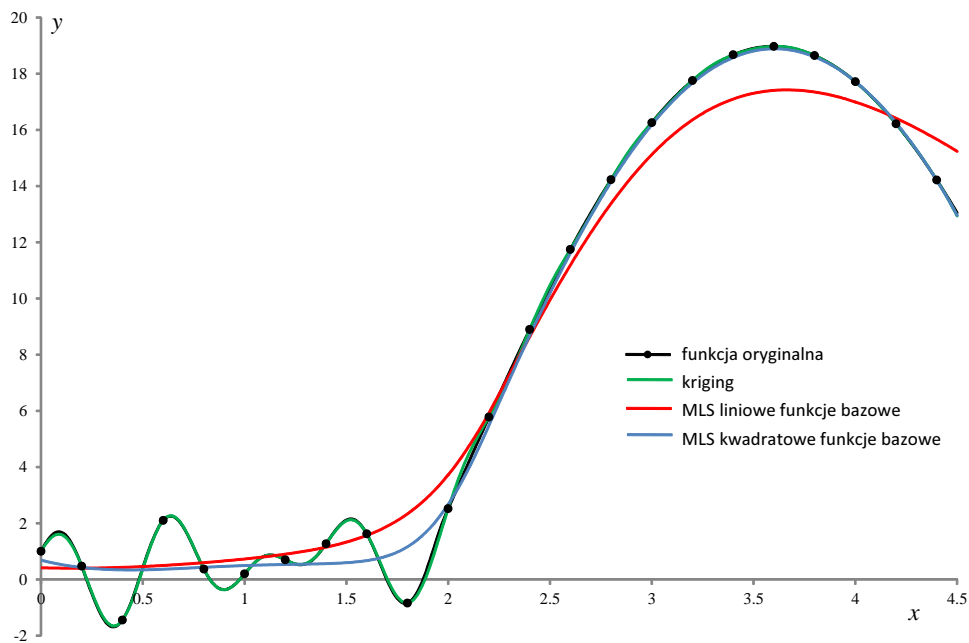
$$w_{ii}(\mathbf{x}) = \exp \left[- \sum_{j=1}^n (x_{ij} - x_j)^2 / (2n\eta^2) \right], \quad \eta > 0, \quad (3.44)$$

gdzie n jest liczbą zmiennych, x_{ij} oznacza j -tą współrzędną i -tego punktu eksperymentalnego, x_j jest j -tą współrzędną punktu \mathbf{x} , a η jest parametrem kontrolującym kształt funkcji.

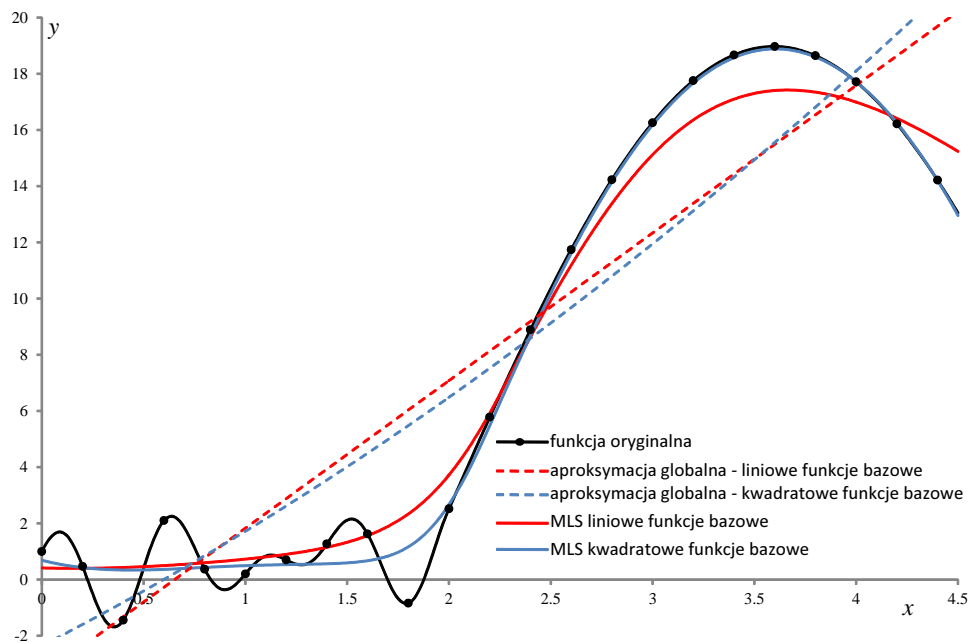
W celu zilustrowania omawianej metody aproksymacji, na rys. 3.5 przedstawiono porównanie wykresów funkcji aproksymujących w przypadku użycia we wzorze (3.42) różnych funkcji bazowych. Kolorem czerwonym oznaczono wykres odpowiadający funkcji liniowej, $\mathbf{q}^T(\mathbf{x}) = [1 \ x]$, a kolorem niebieskim funkcji kwadratowej, $\mathbf{q}^T(\mathbf{x}) = [1 \ x \ x^2]$. Na rysunku umieszczono też wykres aproksymacji metodą krigingu (zielony), a czarnymi punktami oznaczono punkty eksperymentalne. Można zauważyć bardzo dobre dopasowanie aproksymacji kwadratowej do funkcji oryginalnej, szczególnie w prawej części wykresu, natomiast

lewa część wykresu, o zdecydowanie odmiennym, oscylacyjnym charakterze nie jest wiernie odwzorowywana przez metody aproksymacji lokalnej. Prowadzą one w tym obszarze do wygładzenia funkcji oryginalnej. Przy punktach eksperymentalnych danych tak jak na rys. 3.5, jedynie interpolujący kriging prowadzi do doskonałego odtworzenia rzeczywistej zależności $y = f(x)$. Z drugiej strony, kiedy zaburzenia takie jak oscylacje na początku wykresu są wynikiem szumu, to działanie metod aproksymacji lokalnej może mieć korzystny wpływ na zachowanie algorytmów wykorzystujących tak budowane powierzchnie odpowiedzi.

Na koniec, na rys. 3.6 pokazano jak “niebezpieczne” może być stosowanie globalnych technik aproksymacji porównując ważoną (3.42) i standardową metodę liniowej regresji (3.9).



Rys. 3.5. Porównanie metody aproksymacji lokalnej dla liniowych i kwadratowych funkcji bazowych oraz krigingu.



Rys. 3.6. Porównanie metody aproksymacji lokalnej dla liniowych i kwadratowych funkcji bazowych oraz aproksymacji globalnej.

PODSUMOWANIE

Rozpatrywane w niniejszej pracy funkcje zmiennych losowych i zmiennych projektowych, które tworzą sformułowanie zadania analizy niezawodności, jak również optymalizacji odpornościowej konstrukcji, ze względu na swój nieliniowy charakter oraz dużą złożoność obliczeniową reprezentowane są przy pomocy analitycznych metamodeli. Odpowiednia aproksymacja funkcji granicznej w analizie niezawodności oraz funkcji celu i ograniczeń w zadaniu optymalizacji, ma kluczowe znaczenie dla efektywności oraz zbieżności tych analiz.

W rozdziale omówiono zastosowanie metod powierzchni odpowiedzi w aproksymacji zachowania układów konstrukcyjnych, które modelowane są komputerowo. Specyfika eksperymentów komputerowych nie przystaje do założeń klasycznej analizy regresji, gdzie zakłada się statystyczną niezależność błędów eksperymentalnych. W przypadku tych eksperymentów, bardziej racjonalne jest przyjęcie, że wielkości błędów są skorelowane, a korelacja jest wysoka gdy współrzędne dwóch eksperymentów znajdują się w bliskim sąsiedztwie, a niska gdy są do siebie oddalone. Spostrzeżenie to jest fundamentem metod aproksymacyjnych,

które wykorzystują teorię procesów losowych. Przykładem takiej metody jest kriging. Wykorzystując opisaną w dodatku C dyskretyzację pola losowego przy pomocy metody optymalnej aproksymacji liniowej, przedstawiono wyprowadzenie równania krigingu. Bardzo dobre właściwości dopasowywania się powierzchni krigingu do danych eksperymentalnych mogą stanowić jednak pewną wadę, jeśli, tak jak w optymalizacji odpornościowej, mamy do czynienia z szumem generowanym przez wyniki symulacji losowych. Zaprezentowano jak zmodyfikować równania krigingu interpolującego w celu otrzymania techniki aproksymacyjnej. Zbadano wpływ równomierności rozłożenia punktów eksperymentalnych na błąd dopasowania. Pokazano, że w przypadku krigingu, wykorzystanie planów eksperymentów OLH poprawia dopasowanie funkcji aproksymującej. Przedstawiono również metodę aproksymacji lokalnej wykorzystującą koncepcję tzw. ważonej regresji liniowej. Metoda ta, chociaż nie tak dokładna jak kriging, nie wymaga jednak przeprowadzenia zadania minimalizacji funkcji wiarygodności w celu znalezienia parametrów równania. Zarówno kriging jak też metoda aproksymacji lokalnej stanowią podstawę proponowanych w pracy algorytmów analizy niezawodności oraz optymalizacji odpornościowej.

Moduł powierzchni odpowiedzi programu STAND, który powstał w wyniku kierowanych przez autora prac, zawiera zarówno klasyczne metody liniowej regresji dla liniowych i kwadratowych funkcji bazowych, jak i metody liniowej regresji ważonej oraz aproksymację za pomocą krigingu.

Efektywne metody szacowania momentów statystycznych odpowiedzi konstrukcji

Rozwiązanie zadania optymalizacji odpornościowej, które sformułowane zostało w podrozdziale 2.4, wymaga wielokrotnego obliczania momentów statystycznych funkcji celu oraz funkcji ograniczeń. W konsekwencji, wybór metod szacowania wartości średnich oraz odchyłeń standardowych tych funkcji ma decydujący wpływ na efektywność algorytmów poszukiwania optimum.

Ze względu na swój uniwersalny charakter i łatwość implementacji, metody symulacji losowych zostały użyte do obliczania momentów statystycznych w większości przykładów numerycznych prezentowanych w niniejszej pracy. Uniwersalność tych metod polega między innymi na braku jakichkolwiek założeń dotyczących różniczkowalności lub stopnia nieliniowości badanej funkcji, a także braku ograniczeń dotyczących wartości parametrów rozkładów prawdopodobieństwa zmiennych losowych. W szczególności, metody symulacji losowych nadają się doskonale do obliczania momentów statystycznych losowych funkcji ze składnikiem szumowym. Zaobserwowano, że metody symulacyjne wykorzystujące koncepcję łacińskiej hiperkostki, a zwłaszcza optymalnej łacińskiej hiperkostki (ang. *optimal Latin hypercube* - OLH) prowadzą do niezwykle efektywnej estymacji momentów statystycznych, o wiele bardziej efektywnej niż w przypadku klasycznej metody symulacji Monte Carlo. Zagadnieniom budowania planów OLH poświęcony został dodatek B. Poniżej, na licznych przykładach analitycznych funkcji testowych, jak też niejawnych funkcji odpowiedzi układów konstrukcyjnych, pokazano porównanie błędów estymacji otrzymywanych przy użyciu różnych metod symulacyjnych.

W pewnych szczególnych przypadkach, zdecydowanie bardziej efektywne może być użycie innych metod szacowania momentów funkcji losowych. Dla kompletności prezentacji na końcu rozdziału omówione zostaną wybrane metody

niesymulacyjne. I tak, w podrozdziale 4.2 przedstawiono metody bazujące na rozwinięciach funkcji w szereg Taylora wokół wartości średnich zmiennych losowych, w 4.3 omówiono zaproponowane stosunkowo niedawno przez Xu i Rahmana [259] metody redukcji wymiarów, a w podrozdziale 4.4 przedstawiono metody rozwinięcia funkcji w tzw. *chaos wielomianowy*.

4.1. Metody symulacyjne

Komputerowe symulacje losowe, zwane też symulacjami Monte Carlo, są najbardziej znaną metodą badania wpływu losowego charakteru parametrów modelowanego układu konstrukcyjnego na rozrzut wartości parametrów wynikowych (odpowiedzi konstrukcji). Najczęściej przy pomocy metod symulacyjnych estymowane są wartość oczekiwana oraz wariancja lub odchylenie standardowe odpowiedzi. Metody te stanowią też podstawę do analizy głównych składowych, która zostanie krótko omówiona w dalszej części tekstu.

W zależności od estymowanej wielkości, dokładne oszacowanie wartości momentów odpowiedzi konstrukcji przy pomocy metody Monte Carlo wymaga najczęściej od kilkudziesięciu do wielu tysięcy wywołań programu metody elementów skończonych lub innego programu modelującego pracę konstrukcji. Pomimo ciągłego rozwoju dostępnej mocy obliczeniowej komputerów, koszt ten jest często zbyt wysoki. Ma to szczególne znaczenie wtedy gdy, tak jak to jest w przypadku optymalizacji odpornościowej, symulacje Monte Carlo muszą być przeprowadzone wielokrotnie. Dlatego też do szacowania momentów statystycznych odpowiedzi konstrukcji używa się nieraz alternatywnych metod symulacyjnych. Wśród nich należy wymienić metody z grupy metod *descriptive sampling* (DS), co można przetłumaczyć jako próbkowanie opisowe lub zdeterminowane, oraz metody wykorzystujące koncepcję łańciskowej hiperkostki, a w szczególności optymalnej hiperkostki łańciskowej - OLH. Jak to zostanie wyjaśnione w dalszej części rozdziału, metody wykorzystujące łańciskowe hiperkostki można przy pewnych dodatkowych założeniach również zaliczyć do rodziny metod *descriptive sampling*. W odróżnieniu od zwykłej metody Monte Carlo oraz losowej hiperkostki łańciskowej (ang. *random Latin hypercube* - RLH), plany eksperymentów numerycznych OLH są dość trudne do wyznaczenia w przypadku próbek zawierających dużą liczbę punktów w przestrzeniach wielowymiarowych. Jest zatem niezwykle istotnym, aby móc określić wielkość próbki OLH, która dla danej liczby zmiennych losowych może zagwarantować dostateczną jakość oszacowania momentów statystycznych odpowiedzi konstrukcji. O ile w przypadku klasycznej metody Monte Carlo można spotkać pewne wskazówki co do doboru

wielkości próbki w funkcji liczby wymiarów zadania, zob. np. [53], o tyle nie ma wielu takich prac dla symulacji gdzie używa się łącińskich hiperkostek. Problem ten wydaje się być dość trudny przede wszystkim gdy punkty generowane są zgodnie z planem OLH.

Poniżej, głównie za pomocą doświadczeń numerycznych, podjęta zostanie próba określenia zakresu przydatności planów RLH i OLH w estymacji momentów statystycznych odpowiedzi konstrukcji, jak również określenia potrzebnej ilości punktów eksperymentalnych. Na początku jednak omówione zostaną podstawowe różnice pomiędzy klasycznymi metodami Monte Carlo a metodami DS oraz przytoczone zostaną podstawowe informacje na temat stosowanych estymatorów, por. [195].

4.1.1. Metody *descriptive sampling* a klasyczna metoda Monte Carlo

Można pokazać, że techniki symulacji losowych, w których próbki generowane są za pomocą łącińskich hiperkostek, są pewnym szczególnym przypadkiem metod DS. Bardzo przejrzyste i zwięzłe przedstawienie koncepcji *descriptive sampling* oraz wyjaśnienie czym różni się to podejście od “czysto losowych” symulacji Monte Carlo znaleźć można w pracy [211] Eduardo Saliby, który jest prekursorem tej metody.

Descriptive sampling zaproponowano jako metodę redukcji wariancji statystyk obliczanych na podstawie próbki z populacji. Używając klasycznej metody Monte Carlo można mówić o dwóch rodzajach niepewności charakteryzujących losową próbę. Pierwszy dotyczy rozrzutu wartości zmiennych w próbce, a drugi, sekwencji w jakiej wartości te są generowane. W przypadku DS mamy do czynienia jedynie z niepewnością sekwencji, wartości poszczególnych zmiennych są natomiast zdeterminowane. Schematycznie różnice te przedstawić można następująco:

klasyczna metoda Monte Carlo = losowe wartości + losowa sekwencja,

metoda descriptive sampling = ustalone wartości + losowa sekwencja.

Ponadto, w przeciwieństwie do metody Monte Carlo, DS wymaga określenia z góry wielkości próbki. Gdy jest ona znana, wtedy zbiór realizacji $x_k(m)$ każdej zmiennej losowej X_k , $k = 1, \dots, n$, gdzie m jest numerem punktu w N elementowej próbce, $m = 1, \dots, N$, wyznaczany jest jako

$$x_k(m) = F_{X_k}^{-1} \left(\frac{m - 0.5}{N} \right), \quad m = 1, \dots, N, \quad k = 1, \dots, n. \quad (4.1)$$

W powyższym wzorze $F_{X_k}^{-1}$ oznacza funkcję odwrotną do dystrybuanty k -tej zmiennej losowej. Jako realizacje przyjmowane są więc mediany zmiennych losowych w N przedziałach o równym prawdopodobieństwie, na które dzielony jest zakres zmienności każdej zmiennej (por. punkt B.7 w dodatku B). Powtarzając wielokrotnie N punktowe eksperymenty DS używamy za każdym razem tych samych dyskretnych zbiorów wartości dla poszczególnych zmiennych. Losowa jest natomiast sekwencja, w której wartości te pobierane są ze zbiorów w celu utworzenia n wymiarowych punktów eksperymentalnych (realizacji wektora losowego). Idea metody DS opiera się na zaskakującym z pozoru twierdzeniu, iż losowy charakter stosowanych próbek wynika przede wszystkim z losowości sekwencji, a niepewność dotycząca samych wartości nie prowadzi do poprawy efektywności metod symulacji losowych. Twierdzi się, że wartości realizacji zmiennych losowych powinny być precyzyjnie ustalone aby jak najlepiej dopasować się do rozkładów prawdopodobieństwa tych zmiennych.

Pokażemy teraz, że symulacje losowe wykorzystujące koncepcję LH lub OLH mieszczą się w założeniach descriptive sampling. W klasycznej już pracy [170] McKay i in. zaproponowali wykorzystanie hiperkostki łacińskiej jako metody redukcji wariancji, w której pomimo iż pozwala się na pewien losowy rozrzut wartości realizacji zmiennych, to jednak podlegają one ścisłej kontroli. W ramach tego podejścia zakresy zmienności poszczególnych zmiennych losowych dzielone są na przedziały (warstwy) o równym prawdopodobieństwie, a punkty wybierane są w sposób losowy wewnątrz tych warstw. Wzór (4.1) przyjmuje w tym wypadku postać

$$x_k^{\text{LH}}(m) = F_{X_k}^{-1} \left(\frac{m - r}{N} \right), \quad m = 1, \dots, N, \quad k = 1, \dots, n, \quad (4.2)$$

gdzie r jest realizacją zmiennej losowej o rozkładzie jednostajnym w przedziale $[0, 1]$. Istnieją także inne wersje wyboru wartości zmiennej wewnątrz warstwy, które pozwalają zaklasyfikować metody symulacji losowych wykorzystujących LH jako należące do rodziny descriptive sampling. Obecnie, jako wartość reprezentatywną dla danego przedziału najczęściej wybiera się bądź medianę, por. (4.1), bądź wartość oczekiwaną zmiennej w tym przedziale [100, 119]. W tych przypadkach zbiór używanych w symulacjach wartości jest całkowicie zdeterminowany, a losowość uzyskuje się jedynie poprzez przypadkowy charakter “składania” z nich n wymiarowych punktów N elementowej próbki.

Chociaż sposób wyboru wartości wewnątrz warstwy może mieć wpływ na dokładność estymacji (szczególnie przy mało licznych próbkach), to hiperkostka łacińska jest przede wszystkim pewnym sposobem rozmieszczenia N punktów

w przestrzeni n wymiarowej. Dzieląc zakres określoności każdej składowej wektora \mathbf{X} na N przedziałów o jednakowym prawdopodobieństwie równym $1/N$ i wykonując następnie na zbiorach tych przedziałów operację iloczynu kartezjańskiego, dzielimy przestrzeń realizacji zmiennych na N^n obszarów (komórek) o jednakowym prawdopodobieństwie N^{-n} . Każda komórka identyfikowana być może przez wektor n współrzędnych $\mathbf{c}^{(i)} = \{c_1^{(i)}, c_2^{(i)}, \dots, c_n^{(i)}\}$, gdzie $c_j^{(i)}$ jest numerem warstwy zmiennej X_j odpowiadającym komórce i . N elementową hiperkostkę łacińską tworzy się przez wybranie N komórek, $\mathbf{c}^{(1)}, \dots, \mathbf{c}^{(N)}$, spośród N^n komórek w taki sposób, że dla dowolnego indeksu j zbiór $\{c_j^{(i)}\}_{i=1}^N$ jest permutacją liczb $1, \dots, N$. W każdej z wybranych komórek dokonuje się jednej obserwacji. Funkcja łącznej gęstości prawdopodobieństwa wektora losowego \mathbf{X} wewnątrz i -tej komórki wynosi

$$f_{\mathbf{c}^{(i)}}(\mathbf{x}) = \begin{cases} N^n f_{\mathbf{X}}(\mathbf{x}) & \mathbf{x} \text{ jest wewnątrz komórki } \mathbf{c}^{(i)}, \\ 0 & \mathbf{x} \text{ jest poza komórką } \mathbf{c}^{(i)}. \end{cases} \quad (4.3)$$

Przewaga metod symulacyjnych wykorzystujących ideę hiperkostki łacińskiej, jak również i innych metod, w których dokonuje się podziału zakresu poszczególnych zmiennych na warstwy o jednakowym prawdopodobieństwie, manifestuje się szczególnie wtedy gdy funkcja losowa $\xi(\mathbf{X})$, której momenty statystyczne chcemy estymować, jest “zdominowana” jedynie przez kilka składowych wektora \mathbf{X} . Metody symulacji warstwowych gwarantują odpowiednią reprezentację wszystkich zmiennych, co wyklucza pominięcie w próbie realizacji istotnych zmiennych losowych.

4.1.2. Podstawowe informacje na temat estymatorów

Jeśli przez $\xi(\mathbf{X})$ oznaczymy funkcję zmiennych losowych \mathbf{X} o łącznej gęstości prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$, to rozpatrywaną tutaj klasę estymatorów zapisać można w ogólnej postaci jako

$$T(Y_1, \dots, Y_N) = \frac{1}{N} \sum_{i=1}^N h(Y_i), \quad (4.4)$$

gdzie $Y = \xi(\mathbf{X})$, a $h(\cdot)$ jest dowolną funkcją. Przyjmując $h(Y) = Y$ uzyskujemy średnią arytmetyczną z próbki będącą estymatorem wartości oczekiwanej μ_Y . Przez podstawienie $h(Y) = Y^r$ otrzymuje się wzór na estymator momentu r -tego rzędu. Jeśli próbki generowane były w sposób czysto losowy, zgodnie z filozofią podejścia Monte Carlo, to odpowiadające im estymatory oznaczać będziemy

indeksem dolnym R , a w przypadku gdy położenie punktów określano za pomocą łacińskiej hiperkostki, indeksem L .

Estymator wartości oczekiwanej funkcji losowej

Średnia arytmetyczna \bar{Y}_R (empiryczna wartość przeciętna) dana jako

$$\bar{Y}_R = \frac{1}{N} \sum_{i=1}^N Y_i = \frac{1}{N} \sum_{i=1}^N \xi(\mathbf{X}_i), \quad (4.5)$$

gdzie N jest wielkością próbki, a \mathbf{X}_i są niezależnymi wektorami losowymi o funkcji gęstości $f_{\mathbf{X}}(\mathbf{x})$, jest nieobciążonym estymatorem wartości oczekiwanej μ_Y ponieważ

$$\mathbb{E}(\bar{Y}_R) = \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N Y_i\right) = \frac{1}{N} \sum_{i=1}^N \mathbb{E}(Y_i) = \frac{1}{N} N \mathbb{E}(Y) = \mu_Y. \quad (4.6)$$

Wartość oczekiwana estymatora (4.5) równa jest zatem wartości oczekiwanej funkcji, co jest warunkiem nieobciążoności estymatora. Obliczając następnie wariancję średniej arytmetycznej otrzymujemy

$$\mathbb{V}\text{ar}(\bar{Y}_R) = \mathbb{V}\text{ar}\left(\frac{1}{N} \sum_{i=1}^N Y_i\right) = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}\text{ar}(Y_i) = \frac{1}{N^2} N \mathbb{V}\text{ar}(Y) = \frac{\mathbb{V}\text{ar}(Y)}{N}, \quad (4.7)$$

a więc estymator \bar{Y}_R jest także estymatorem zgodnym wartości oczekiwanej populacji generalnej, ponieważ wraz ze wzrostem wielkości próbki N wariancja estymatora zbiega do 0, a wartość \bar{Y}_R dąży do μ_Y . W powyższych przekształceniach wykorzystano fakt niezależności zmiennych Y_i (wynikający z niezależności wektorów \mathbf{X}_i) oraz to, że

$$\forall i \quad \mathbb{E}(Y_i) = \mathbb{E}(Y), \quad \mathbb{V}\text{ar}(Y_i) = \mathbb{V}\text{ar}(Y). \quad (4.8)$$

Jeśli eksperymenty numeryczne generowane są zgodnie z planem określonym przez hiperkostkę łacińską, to uwzględniając charakterystyczny dla tej metody sposób tworzenia próbki losowej, estymator (4.5) wygodnie jest zapisać następująco

$$\bar{Y}_L = \frac{1}{N} \sum_{i=1}^{N^n} w_i Y_i = \frac{1}{N} \sum_{i=1}^{N^n} w_i \xi(\mathbf{X}_i), \quad (4.9)$$

gdzie realizacje wektora \mathbf{X}_i generowane są w komórce i , a w_i jest zmienną o rozkładzie zero-jedynkowym postaci

$$w_i = \begin{cases} 1 & \text{jeśli punkt z komórki } i \text{ wchodzi w skład próbki,} \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (4.10)$$

Z definicji (4.10) oraz z własności łańciskiej hiperkostki w sposób naturalny wynika, że

$$\mathbb{P}(w_i = 1) = \frac{N}{N^n} = \frac{1}{N^{n-1}} = \mathbb{E}(w_i). \quad (4.11)$$

Ponieważ zmienne w_i i Y_i są niezależne, oraz wykorzystując funkcję gęstości prawdopodobieństwa (4.3) wartość oczekiwaną estymatora \bar{Y}_L obliczyć można w następujący sposób:

$$\begin{aligned} \mathbb{E}(\bar{Y}_L) &= \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^{N^n} w_i Y_i\right) = \frac{1}{N} \sum_{i=1}^{N^n} \mathbb{E}(w_i) \mathbb{E}(Y_i) = \frac{1}{N} \sum_{i=1}^{N^n} \frac{1}{N^{n-1}} \mathbb{E}(Y_i) \\ &= \frac{1}{N^n} \sum_{i=1}^{N^n} \int_{\Omega_i} y_i N^n f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}(Y) = \mu_Y, \end{aligned} \quad (4.12)$$

gdzie Ω_i jest obszarem zajmowanym przez i -tą komórkę hiperkostki. Estymator (4.9) jest zatem nieobciążonym estymatorem wartości oczekiwanej. Można wykazać, że w ogólności estymatory (4.4) – w przypadku gdy próbki generowane są za pomocą łańciskich hiperkostek – są nieobciążonymi estymatorami τ , gdzie τ jest wartością oczekiwaną T_R .

W pracy [170] wyprowadzono wzór na wariancję estymatora wartości oczekiwanej \bar{Y}_L w funkcji wariancji \bar{Y}_R

$$\text{Var}(\bar{Y}_L) = \text{Var}(\bar{Y}_R) + \frac{N-1}{N^{n+1}(N-1)^n} \sum_P (\mu_i - \mu_Y)(\mu_j - \mu_Y), \quad (4.13)$$

gdzie $\mu_i = \mathbb{E}(Y | \mathbf{X} \in \text{komórka } i)$, a P jest zbiorem wszystkich par μ_i, μ_j odpowiadających komórkom nie mającym żadnej wspólnej współrzędnej. Wyrażenie (4.13) jest szczególnym przypadkiem wzoru na wariancję dowolnego estymatora typu (4.4) postaci

$$\text{Var}(T_L) = \text{Var}(T_R) + \frac{N-1}{N^{n+1}(N-1)^n} \sum_P (\mu_i - \tau)(\mu_j - \tau), \quad (4.14)$$

gdzie $\mu_i = \mathbb{E}[h(Y) | \mathbf{X} \in \text{komórka } i]$. Można udowodnić następujące twierdzenie (zob. [170]):

Twierdzenie 4.1: *Jeśli funkcja $Y = \xi(X_1, \dots, X_n)$ jest monotoniczna ze względu na każdy ze swoich argumentów, a także $h(Y)$ jest monotoniczną funkcją Y , to $\text{Var}(T_L) \leq \text{Var}(T_R)$.*

Ponieważ powyższe warunki spełnia bardzo wiele funkcji modelujących zachowanie rzeczywistych układów konstrukcyjnych, dlatego jest to niesłychanie ważna właściwość metod symulacyjnych wykorzystujących plany LH, dobrze uzasadniająca celowość ich stosowania. W pracy [96] wymaganie zawarte w twierdzeniu 4.1 zostało nieznacznie złagodzone. Pokazano, że redukcję wariancji uzyskuje się już wtedy gdy funkcja $Y = \xi(\mathbf{X})$ jest monotoniczna względem $n - 1$ zmiennych.

Estymator wariancji funkcji losowej

Do oszacowania wartości wariancji funkcji losowej elementów populacji generalnej $\text{Var}(Y) = \text{Var}[\xi(\mathbf{X})] = \sigma_Y^2$ używana jest zazwyczaj tzw. wariancja empiryczna dana wzorem

$$s^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2, \quad (4.15)$$

gdzie w zależności od stosowanej metody za \bar{Y} podstawia się \bar{Y}_R lub \bar{Y}_L .

W przypadku próbek generowanych w sposób losowy, wartość oczekiwaną tego estymatora oblicza się następująco:

$$\mathbb{E}(s_R^2) = \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y}_R)^2 \right],$$

można zauważyć, że

$$\begin{aligned} \sum_{i=1}^N [Y_i - \bar{Y}_R]^2 &= \sum_{i=1}^N [(Y_i - \mu_Y) - (\bar{Y}_R - \mu_Y)]^2 \\ &= \sum_{i=1}^N [(Y_i - \mu_Y)^2 + (\bar{Y}_R - \mu_Y)^2 - 2(Y_i - \mu_Y)(\bar{Y}_R - \mu_Y)] \\ &= \sum_{i=1}^N (Y_i - \mu_Y)^2 - N(\bar{Y}_R - \mu_Y)^2, \end{aligned} \quad (4.16)$$

ponieważ

$$\sum_{i=1}^N 2(Y_i - \mu_Y)(\bar{Y}_R - \mu_Y) = 2(\bar{Y}_R - \mu_Y) \sum_{i=1}^N (Y_i - \mu_Y) = 2N(\bar{Y}_R - \mu_Y)^2,$$

co daje w rezultacie

$$\mathbb{E}(s_R^2) = \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N (Y_i - \mu_Y)^2 \right] - \mathbb{E} [(\bar{Y}_R - \mu_Y)^2].$$

Ponieważ ostatni człon powyższego równania jest wariancją wartości średniej \bar{Y}_R , dlatego na podstawie (4.7) można zapisać

$$\mathbb{E} [(\bar{Y}_R - \mu_Y)^2] = \frac{\text{Var}(Y)}{N}.$$

Przekształcając następnie pierwszy człon wyrażenia na $\mathbb{E}[s_R^2]$

$$\mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N (Y_i - \mu_Y)^2 \right] = \frac{1}{N} \sum_{i=1}^N \mathbb{E} [(Y_i - \mu_Y)^2] = \frac{1}{N} N \text{Var}[Y] = \text{Var}[Y],$$

otrzymujemy ostatecznie

$$\mathbb{E}(s_R^2) = \frac{N-1}{N} \text{Var}[Y]. \quad (4.17)$$

Estymator (4.15) jest więc obciążonym estymatorem wariancji (jest nieobciążony asymptotycznie), dlatego też częściej używa się nieobciążonego estymatora

$$\tilde{s}_R^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y}_R)^2, \quad (4.18)$$

dla którego

$$\mathbb{E}(\tilde{s}_R^2) = \text{Var}[Y]. \quad (4.19)$$

Można pokazać, że wariancja powyższych estymatorów dąży do zera gdy N wzrasta do nieskończoności. Są to zatem estymatory zgodne.

Gdy plan eksperymentów generowany jest przez łacińską hiperkostkę, to podobnie jak w przypadku estymatora wartości średniej, w celu sprawdzenia obciążenia estymatora wariancji wygodnie jest posłużyć się alternatywną formą estymatora (4.15)

$$s_L^2 = \frac{1}{N} \sum_{i=1}^{N^n} w_i (Y_i - \bar{Y}_L)^2 = \frac{1}{N} \sum_{i=1}^{N^n} w_i [\xi(\mathbf{X}_i) - \bar{Y}_L]^2, \quad (4.20)$$

gdzie realizacje \mathbf{X}_i generowane są w komórce i , a zmienne w_i , $i = 1, \dots, N^n$, dane są wzorem (4.10). Przekształcenia analogiczne do tych w (4.12) prowadzą do wyrażenia

$$\mathbb{E}(s_L^2) = \mathbb{E}[(Y - \bar{Y}_L)^2], \quad (4.21)$$

i dalej, tak jak w (4.16),

$$\mathbb{E}(s_L^2) = \mathbb{E}\left\{[(Y - \mu_Y) - (\bar{Y}_L - \mu_Y)]^2\right\} = \text{Var}(Y) - \text{Var}(\bar{Y}_L). \quad (4.22)$$

Niestety, w przeciwieństwie do estymatora \bar{Y}_R nie jest znane wyrażenie na wariancję \bar{Y}_L . Nie wyprowadzono również wyrażen na wariancję estymatorów wariancji (4.15) i (4.20).

Estymator odchylenia standardowego

Jako estymator odchylenia standardowego funkcji losowej najczęściej używany jest pierwiastek estymatora nieobciążonego wariancji (4.18)

$$s = \sqrt{\frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N - 1}}. \quad (4.23)$$

W odróżnieniu od estymatora wariancji nie jest to estymator nieobciążony. W większości przypadków (4.23) daje zaniżone wyniki w porównaniu z estymowaną wielkością, różnica ta jednak dąży do zera wraz ze wzrostem liczebności próbki, można więc powiedzieć, że estymator ten jest asymptotycznie nieobciążony.

4.1.3. Analiza głównych składowych

Zazwyczaj celem losowych symulacji pracy konstrukcji przeprowadzanych z udziałem modelu komputerowego jest obliczenie pierwszych momentów statystycznych odpowiedzi konstrukcji oraz oszacowanie korelacji pomiędzy losowymi parametrami modelu, a wybranymi wielkościami wynikowymi. Z drugiej strony, bardziej zaawansowana analiza statystyczna wyników symulacji prowadzi może do lepszego zrozumienia głównych przyczyn ich wariancji, co z kolei daje możliwość wprowadzenia odpowiednich zmian w projektowanej konstrukcji.

Jedną z metod służących temu celowi jest analiza głównych składowych (ang. *principal component analysis* - PCA), zob. np. [53]. Identyfikując zmienne losowe, które wnoszą największy wkład w rozrzut wielkości odpowiedzi konstrukcji,

PCA pozwala na redukcję rozmiaru modelu stochastycznego poprzez eliminację mało "istotnych" zmiennych losowych. Szczególnie w przypadkach, gdy analizowany model zawiera wiele zmiennych wejściowych, wygodnie jest móc go uprościć poprzez rozpatrywanie jedynie ograniczonej liczby liniowych kombinacji tych zmiennych, takich kombinacji, które reprezentują znaczną część losowej zmienności reprezentowanej przez macierz kowariancji.

Określmy dalej mianem parametrów zbiór wszystkich zmiennych losowych oraz wybranych odpowiedzi konstrukcji, które brane są pod uwagę w zadaniu symulacji losowych. Ze względu na różnorodną naturę fizyczną tych wielkości, wygodnie jest użyć w analizie głównych składowych macierzy korelacji \mathbf{R} , która jest macierzą kowariancji standaryzowanych parametrów. Analiza PCA bazuje na rozwiązaniu następującego problemu własnego:

$$(\mathbf{R} - \lambda_i \mathbf{I})\mathbf{v}_i = \mathbf{0}, \quad i = 1, \dots, n_p, \quad (4.24)$$

gdzie n_p jest liczbą parametrów, a λ_i i \mathbf{v}_i są odpowiednio wartościami oraz wektorami własnymi. Przyjmijmy następnie, że wartości własne (rzeczywiste i dodatnie) są uporządkowane w kolejności malejącej, zaczynając od największej. Prosta przechodząca przez początek wyznaczonego przez parametry układu współrzędnych i skierowana zgodnie z \mathbf{v}_1 nazywana jest pierwszą osią główną. Analogiczna prosta skierowana zgodnie z wektorem \mathbf{v}_2 nazywana jest drugą osią główną itd. Główne składowe określają pozycję punktów w przestrzeni n_p wymiarowej (zmienne losowe + wybrane odpowiedzi) w układzie głównych osi. Wykorzystując macierz \mathbf{V} , której kolumny utworzone są z wektorów własnych, wartości składowych głównych dla parametrów odpowiadających punktom z losowej próbki wyznacza się za pomocą transformacji

$$\mathbf{F} = \begin{bmatrix} \frac{x_{11} - \bar{X}_1}{s_{X_1}} & \frac{x_{12} - \bar{X}_2}{s_{X_2}} & \dots & \frac{x_{1n_p} - \bar{X}_{n_p}}{s_{X_{n_p}}} \\ \frac{x_{21} - \bar{X}_1}{s_{X_1}} & \frac{x_{22} - \bar{X}_2}{s_{X_2}} & \dots & \frac{x_{2n_p} - \bar{X}_{n_p}}{s_{X_{n_p}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{N1} - \bar{X}_1}{s_{X_1}} & \frac{x_{N2} - \bar{X}_2}{s_{X_2}} & \dots & \frac{x_{Nn_p} - \bar{X}_{n_p}}{s_{X_{n_p}}} \end{bmatrix} \mathbf{V}, \quad (4.25)$$

gdzie N jest liczbą punktów w próbce, \mathbf{F} jest macierzą o wymiarze $N \times n_p$ zawierającą współrzędne parametrów w przestrzeni głównych składowych, a \bar{X}_i i s_{X_i} , $i = 1, \dots, n_p$, są wartościami średnimi oraz odchyleniami standardowymi parametrów, obliczonymi z próbki. Można pokazać, że współczynnik korelacji pomiędzy i -tym parametrem j -tą składową główną dany jest wzorem

$$\rho_{ij} = u_{ij} \sqrt{\lambda_j}, \quad (4.26)$$

gdzie λ_j jest wartością własną odpowiadającą j -tej składowej głównej, a u_{ij} jest i -tym wyrazem związanego z nią wektora własnego.

Jeśli z rozwiązania zadania własnego (4.24) wynika, że jest kilka dominujących wartości własnych λ_i , to silna korelacja pomiędzy parametrem związanym z pewną zmienną losową, a odpowiednią składową główną jest dobrym wskaźnikiem wpływu tej zmiennej na obserwowany rozrzut odpowiedzi analizowanego układu konstrukcyjnego.

W prezentowanym dalej przykładzie płyty sprężystej o losowej grubości (zob. pkt. 4.1.6) przeanalizowana zostanie efektywność poszczególnych metod symulacji losowych w estymacji wartości średniej pierwszej głównej składowej.

4.1.4. Przykłady numeryczne: Analityczne funkcje testowe

4.1.4.1. Monotoniczna funkcja dwóch zmiennych

Sprawdzenie efektywności omówionych metod symulacyjnych rozpoczniemy od funkcji testowej przedstawionej w pracy [89]

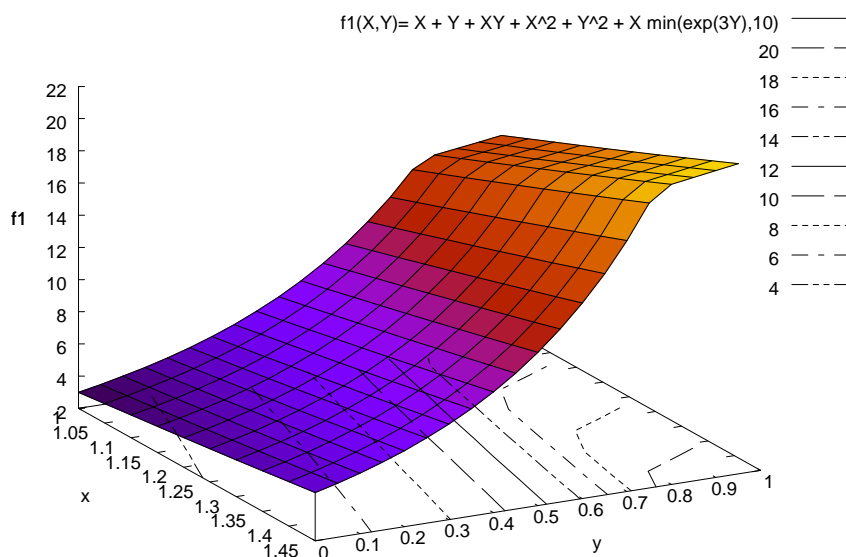
$$f_1(X, Y) = X + Y + XY + X^2 + Y^2 + X \min[\exp(3Y), 10], \quad (4.27)$$

gdzie X i Y są zmiennymi losowymi o rozkładzie jednostajnym określonymi odpowiednio na przedziałach $[1, 1.5]$ i $[0, 1]$. Jak to można zobaczyć na rys. 4.1, funkcja (4.27) jest na tym obszarze monotoniczna. Zgodnie z twierdzeniem 4.1 (strona 90) należy się zatem spodziewać, że zastosowanie symulacji losowych, w których próbki generowane są przez łańciskie hiperkostki, prowadzić będzie do poprawy efektywności w stosunku do zwykłej metody Monte Carlo.

Ze względu na jawną postać funkcji f_1 , jej wartość oczekiwana oraz odchylenie standardowe oszacowane być mogą z dość dużą dokładnością na podstawie symulacji Monte Carlo. Przyjmując wielkość próbki $N = 10^6$ otrzymano

$$\mathbb{E}[f_1(X, Y)] = 10.947, \quad \sigma[f_1(X, Y)] = 5.344. \quad (4.28)$$

Do wartości tych odnosić się będziemy określając średni błąd estymacji wartości oczekiwanej oraz odchylenia standardowego za pomocą szeregu technik symulacji losowych. Brano pod uwagę następujące metody:



Rys. 4.1. Funkcja testowa $f_1(X, Y) = X + Y + XY + X^2 + Y^2 + X \min[\exp(3Y), 10]$

MC Klasyczna metoda symulacji Monte Carlo.

RLH-r Próbkę generowane są przez losowe hiperkostki łącińskie, tzn. takie, w których nie optymalizowano rozmieszczenia komórek zawierających punkty (realizacje wektora losowego), a położenie punktu wewnątrz komórki hiperkostki wybierane jest w sposób losowy zgodnie z rozkładem prawdopodobieństwa zmiennych w komórce.

RLH-m Próbkę generowane są przez losowe hiperkostki łącińskie, a punkty wybierane są w medianach zmiennych losowych wewnątrz komórek.

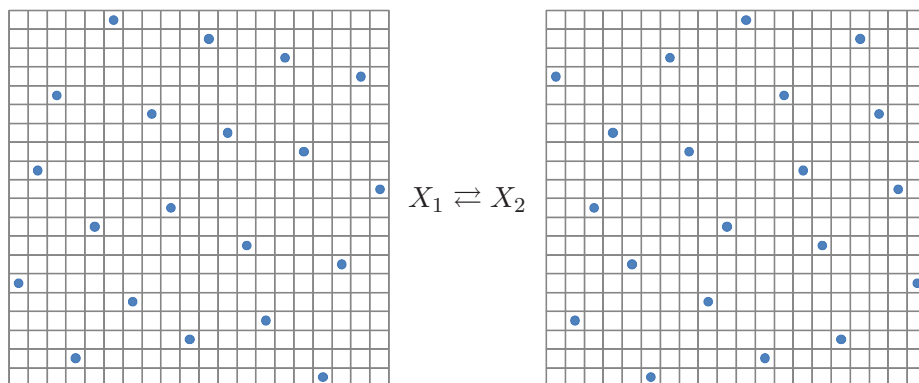
RLH- μ Podobnie jak RLH-m, ale zamiast median przyjmowane są wartości oczekiwane zmiennych. O ile w przypadku zmiennych o obciętych rozkładzie normalnym oraz rozkładzie jednostajnym znane są wzory na wartość oczekiwaną, to dla większości pozostałych rozkładów prawdopodobieństwa w celu znalezienia tej wartości konieczne jest zastosowanie całkowania numerycznego. Oczywiście, gdy wszystkie zmienne losowe mają rozkład jednostajny to metody RLH- μ oraz RLH-m prowadzą do takich samych wyników.

OLH-r W metodzie tej do wygenerowania próbki stosuje się optymalne hiperkostki łącińskie (zob. dodatek B). Pomimo swojej nazwy, w rzeczywistości ten optymalny układ punktów eksperymentalnych nie jest unikalny.

Optymalna hiperkostka przedstawiona może być na $n!$ sposobów przez zwykłą zamianę kolejności współrzędnych punktów z próbki, zachowując stałą wartość kryterium odległości między punktami bądź też kryterium sił (zob. podrozdział B.1). Mechanizm ten zilustrowano na rys. 4.2 na przykładzie planu OLH dla dwóch zmiennych. Inną przyczyną, dla której można mówić o rozrzucie wartości statystyk uzyskiwanych przy pomocy próbek OLH, jest sam proces tworzenia optymalnej hiperkostki. Ponieważ proces optymalizacyjny rozpoczyna się od utworzenia losowej łańciskiej hiperkostki oraz zatrzymywany jest po spełnieniu pewnego arbitralnie przyjętego kryterium zbieżności, dlatego też w rzeczywistości można mówić o pewnej populacji planów OLH. Co jest zrozumiałe, im większa próbka oraz im więcej zmiennych losowych, tym mniejsze jest prawdopodobieństwo uzyskania dwóch identycznych OLH w dwóch niezależnych procesach optymalizacji. Przez OLH-r, analogicznie jak RLH-r, oznaczono metodę, w której realizacja wektora losowego wewnątrz komórki OLH wybierana jest w sposób losowy, zgodnie z funkcją gęstości prawdopodobieństwa (4.3).

OLH-m Próbkę generowane są przez optymalne hiperkostki łańciskie, a punkty planu OLH umiejscowione są w medianach zmiennych losowych wewnątrz wybranych komórek.

OLH- μ Próbkę generowane są przez optymalne hiperkostki łańciskie, a punkty planu OLH określone są przez wartości oczekiwane zmiennych losowych w wybranych komórkach.

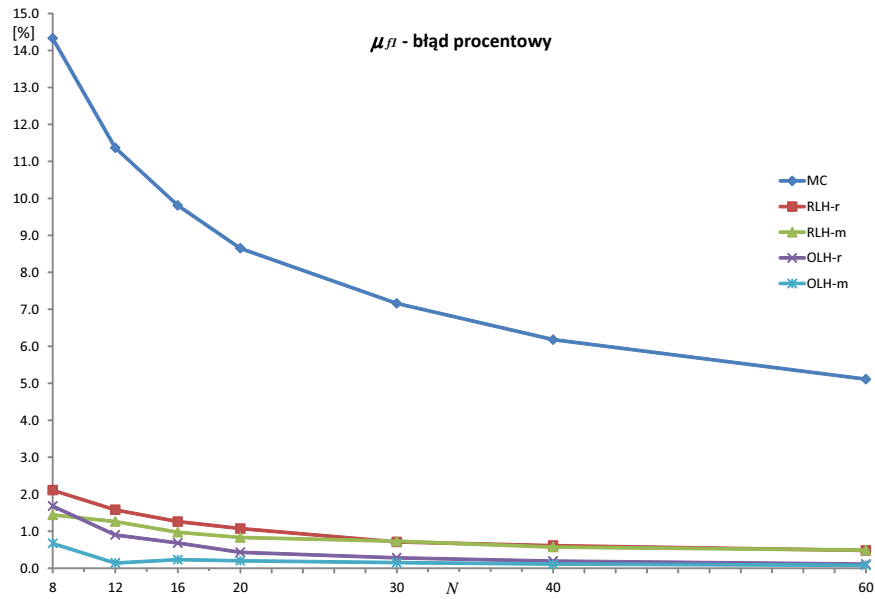


Rys. 4.2. Zmiana układu punktów kwadratu łańciskiego w wyniku zamiany kolejności współrzędnych punktów.

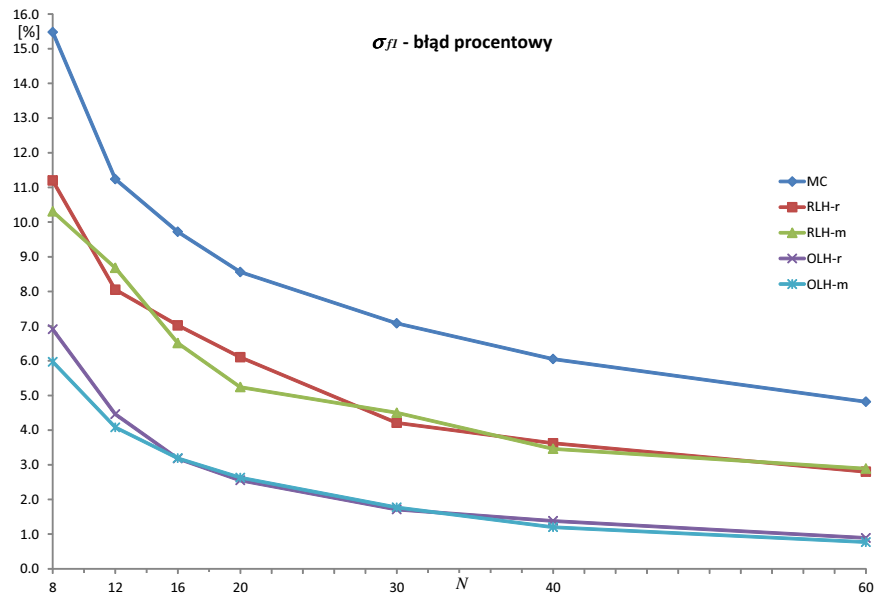
W tabeli 4.1 przedstawiono wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanych oraz odchylen standardowych funkcji (4.1) za pomocą wymienionych powyżej metod symulacyjnych. Rozpatrywano szereg wielkości próbek, od zawierających $N = 4n = 8$ do $N = 30n = 60$ punktów eksperymentalnych. Wartości w tabeli otrzymano na podstawie 100 powtórzeń zadania symulacji losowych każdego typu dla wszystkich założonych liczebności próbek. Te same wyniki przedstawiono również w formie wykresów na rysunkach 4.3 i 4.4. Na ich podstawie łatwo można ocenić efektywność poszczególnych metod. W analizowanym przykładzie, niezależnie od wielkości próbki, techniki wykorzystujące koncepcję łacińskiej hiperkostki prowadzą do znacznie mniejszych błędów estymacji niż standardowa metoda Monte Carlo. Ta dominacja jest szczególnie widoczna przy szacowaniu wartości oczekiwanej funkcji f_1 . Spośród czterech metod generujących próbki za pomocą hiperkostek, najlepszą okazała się metoda OLH-m, która nawet dla małych N oferuje doskonałą jakość estymacji. Można również zauważyć, że wraz ze wzrostem liczebności próbki zacieśnia się różnica między wariantami wyboru realizacji zmiennych wewnątrz komórki hiperkostki.

Tablica 4.1. Wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanej oraz odchylenia standardowego funkcji (4.27) przy użyciu różnych technik symulacji losowych i dla różnych wielkości próbek N . Przyjęto jednostajny rozkład zmiennych losowych. Odchylenia standardowe błędów podane są kursywą.

N	MC err[%]		RLH-r err[%]		RLH-m err[%]		OLH-r err[%]		OLH-m err[%]	
	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}
$4n=8$	14.33	15.48	2.11	11.20	1.44	10.31	1.68	6.91	0.67	5.97
	<i>10.51</i>	<i>12.55</i>	<i>1.61</i>	<i>7.55</i>	<i>1.07</i>	<i>7.17</i>	<i>1.19</i>	<i>4.95</i>	<i>0.20</i>	<i>4.20</i>
$6n=12$	11.37	11.24	1.58	8.05	1.26	8.68	0.90	4.46	0.14	4.08
	<i>8.62</i>	<i>8.98</i>	<i>1.08</i>	<i>5.52</i>	<i>0.84</i>	<i>5.51</i>	<i>0.64</i>	<i>2.06</i>	<i>0.17</i>	<i>1.34</i>
$8n=16$	9.81	9.72	1.26	7.02	0.97	6.51	0.68	3.18	0.23	3.19
	<i>7.14</i>	<i>7.45</i>	<i>0.81</i>	<i>4.54</i>	<i>0.73</i>	<i>4.62</i>	<i>0.47</i>	<i>1.99</i>	<i>0.16</i>	<i>1.73</i>
$10n=20$	8.65	8.56	1.07	6.10	0.83	5.24	0.43	2.55	0.20	2.63
	<i>6.50</i>	<i>6.58</i>	<i>0.73</i>	<i>4.02</i>	<i>0.64</i>	<i>4.00</i>	<i>0.32</i>	<i>1.54</i>	<i>0.13</i>	<i>1.42</i>
$15n=30$	7.16	7.08	0.71	4.21	0.73	4.50	0.28	1.71	0.15	1.77
	<i>5.35</i>	<i>5.21</i>	<i>0.55</i>	<i>3.19</i>	<i>0.53</i>	<i>3.18</i>	<i>0.18</i>	<i>1.03</i>	<i>0.09</i>	<i>0.92</i>
$20n=40$	6.18	6.05	0.61	3.62	0.57	3.46	0.19	1.38	0.11	1.20
	<i>4.89</i>	<i>4.48</i>	<i>0.47</i>	<i>2.74</i>	<i>0.46</i>	<i>2.73</i>	<i>0.13</i>	<i>0.76</i>	<i>0.07</i>	<i>0.65</i>
$30n=60$	5.11	4.82	0.48	2.80	0.49	2.89	0.11	0.89	0.08	0.77
	<i>3.79</i>	<i>3.50</i>	<i>0.39</i>	<i>2.22</i>	<i>0.38</i>	<i>2.30</i>	<i>0.07</i>	<i>0.47</i>	<i>0.06</i>	<i>0.49</i>



Rys. 4.3. Wartość średnia względnego błędu procentowego przy estymacji wartości oczekiwanej funkcji (4.27) dla zmiennych losowych o rozkładzie jednostajnym.



Rys. 4.4. Wartość średnia względnego błędu procentowego przy estymacji odchylenia standardowego funkcji (4.27) dla zmiennych losowych o rozkładzie jednostajnym.

Następnie, to samo zadanie powtórzono zakładając, że zmienne losowe X i Y mają rozkład normalny o identycznych wartościach oczekiwanych jak w przypadku zmiennych jednostajnych oraz tak dobranych odchyleniach standardowych, aby zakres $(-3\sigma, +3\sigma)$ pokrywał się z przedziałami określoności zmiennych jednostajnych. Dostajemy zatem $\mu_X = 1.25$, $\sigma_X = 0.083$, $\mu_Y = 0.5$ i $\sigma_Y = 0.17$. I tym razem, w celu dokładnego oszacowania wartości oczekiwanej oraz odchylenia standardowego funkcji (4.27) przeprowadzono symulacje Monte Carlo przyjmując wielkość próbki losowej $N = 10^6$. Otrzymano następujące wartości:

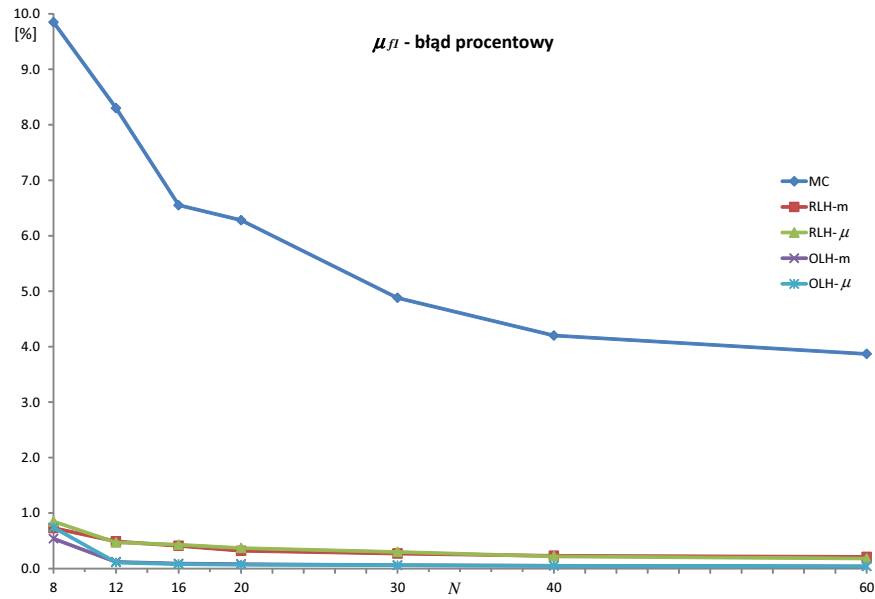
$$\mathbb{E}[f_1(X, Y)] = 10.409, \quad \sigma[f_1(X, Y)] = 3.504. \quad (4.29)$$

W przypadku zmiennych losowych o rozkładzie normalnym wartość średnia zmiennej w komórce łańciskowej hiperkostki nie pokrywa się z jej medianą, dlatego w tabeli 4.2 podano oddzielnie wyniki dla metod LH-m i LH- μ .

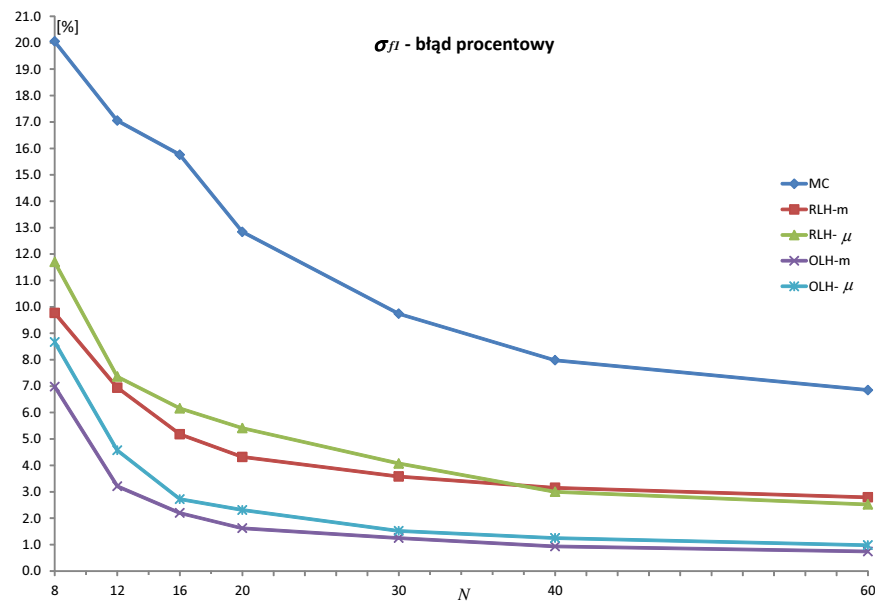
Tablica 4.2. Wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanej oraz odchylenia standardowego funkcji (4.27) przy użyciu różnych technik symulacji losowych i dla różnych wielkości próbek N . Przyjęto gaussowski rozkład zmiennych losowych. Odchylenia standardowe błędów podane są kursywą.

N	MC err[%]		RLH-m err[%]		RLH- μ err[%]		OLH-m err[%]		OLH- μ err[%]	
	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}	μ_{f_1}	σ_{f_1}
$4n = 8$	9.85	20.05	0.73	9.77	0.85	11.70	0.54	6.98	0.74	8.67
	<i>6.74</i>	<i>13.93</i>	<i>0.53</i>	<i>7.18</i>	<i>0.60</i>	<i>8.64</i>	<i>0.26</i>	<i>3.57</i>	<i>0.279</i>	<i>3.64</i>
$6n = 12$	8.30	17.05	0.49	6.94	0.47	7.36	0.12	3.21	0.11	4.57
	<i>6.14</i>	<i>12.95</i>	<i>0.32</i>	<i>4.73</i>	<i>0.31</i>	<i>4.80</i>	<i>0.069</i>	<i>1.45</i>	<i>0.081</i>	<i>1.56</i>
$8n = 16$	6.55	15.76	0.41	5.18	0.43	6.16	0.09	2.20	0.08	2.72
	<i>4.75</i>	<i>12.31</i>	<i>0.27</i>	<i>3.50</i>	<i>0.31</i>	<i>4.38</i>	<i>0.065</i>	<i>1.32</i>	<i>0.061</i>	<i>1.38</i>
$10n = 20$	6.28	12.84	0.32	4.32	0.37	5.41	0.08	1.62	0.07	2.31
	<i>4.66</i>	<i>10.06</i>	<i>0.23</i>	<i>3.19</i>	<i>0.26</i>	<i>3.89</i>	<i>0.063</i>	<i>1.15</i>	<i>0.052</i>	<i>1.21</i>
$15n = 30$	4.88	9.74	0.27	3.58	0.30	4.07	0.06	1.25	0.06	1.52
	<i>3.66</i>	<i>7.71</i>	<i>0.20</i>	<i>2.71</i>	<i>0.21</i>	<i>2.93</i>	<i>0.048</i>	<i>0.87</i>	<i>0.050</i>	<i>0.94</i>
$20n = 40$	4.20	7.98	0.23	3.15	0.22	3.00	0.05	0.93	0.05	1.25
	<i>3.11</i>	<i>5.72</i>	<i>0.16</i>	<i>2.27</i>	<i>0.17</i>	<i>2.24</i>	<i>0.041</i>	<i>0.66</i>	<i>0.038</i>	<i>0.73</i>
$30n = 60$	3.87	6.85	0.21	2.79	0.18	2.52	0.04	0.74	0.04	0.98
	<i>2.88</i>	<i>5.67</i>	<i>0.16</i>	<i>2.32</i>	<i>0.13</i>	<i>1.88</i>	<i>0.031</i>	<i>0.48</i>	<i>0.033</i>	<i>0.58</i>

Podobnie jak poprzednio, wartości z tabeli przedstawiono także na wykresach, zob. rys. 4.5 i rys. 4.6. I w tym przypadku bardzo silnie uwidoczniła się dominacja metod wykorzystujących łańciskie hiperkostki nad zwykłymi symulacjami Monte Carlo. Ciekawe jednak jest, iż inaczej niż to opisywano w pracach [100] oraz [119], zauważalna jest lepsza dokładność estymacji przy wyborze median, a nie wartości średnich zmiennych losowych w komórkach hiperkostek.



Rys. 4.5. Wartość średnia względnego błędu procentowego przy estymacji wartości oczekiwanej funkcji (4.27) dla zmiennych losowych o rozkładzie normalnym.



Rys. 4.6. Wartość średnia względnego błędu procentowego przy estymacji odchylenia standardowego funkcji (4.27) dla zmiennych losowych o rozkładzie normalnym.

4.1.4.2. Niemonotoniczna funkcja dwóch zmiennych

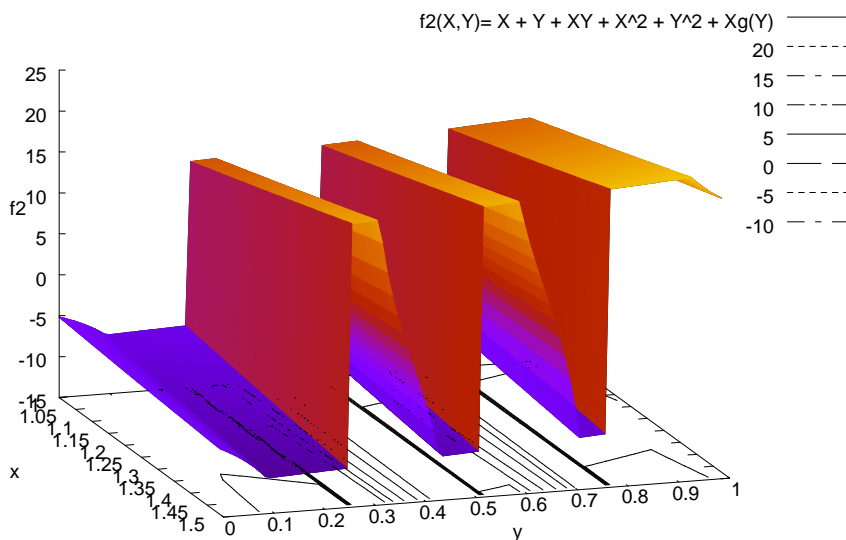
Jak pamiętamy, na podstawie twierdzenia 4.1 można stwierdzić, że metody symulacji losowych wykorzystujące koncepcję łańciskowej hiperkostki zapewniają lepszą jakość estymacji (mniejszą wariancję estymatora) jeżeli rozpatrywana funkcja zmiennych losowych spełnia pewne wymagania dotyczące jej monotoniczności. Interesujące jest zatem porównanie efektywności różnych metod symulacji dla funkcji niemonotonicznej. Ponownie, za pracą [89], przyjęto następującą funkcję dwóch zmiennych losowych X i Y :

$$f_2(X, Y) = X + Y + XY + X^2 + Y^2 + Xg(Y), \quad (4.30)$$

gdzie

$$\begin{aligned} h(Y) &= (Y - 11/43)^{-1} + (Y - 22/43)^{-1} + (Y - 33/43)^{-1}, \\ g(Y) &= h(Y) \quad \text{jeśli } |h(Y)| < 10, \\ g(Y) &= 10 \quad \text{jeśli } h(Y) \geq 10, \\ g(Y) &= -10 \quad \text{jeśli } h(Y) \leq -10. \end{aligned}$$

Tak jak w poprzednim zadaniu, zmienne X i Y mają rozkład jednostajny, odpowiednio na przedziałach $[1, 1.5]$ i $[0, 1]$, Wykres funkcji pokazano na rys. 4.7.



Rys. 4.7. Funkcja testowa $f_2(X, Y) = X + Y + XY + X^2 + Y^2 + Xg(Y)$

Obliczając uprzednio za pomocą symulacji Monte Carlo ($N = 10^7$) wartości statystyk funkcji f_2

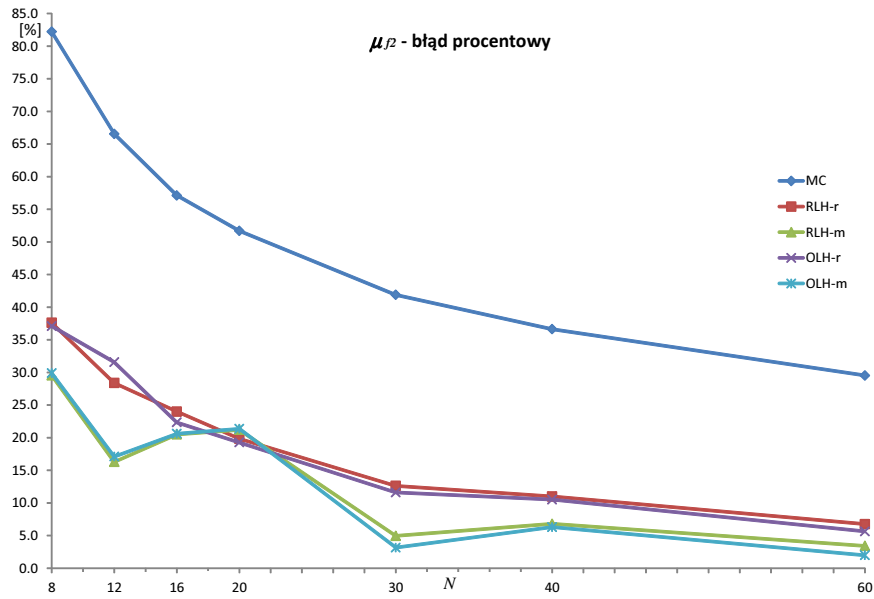
$$\mathbb{E}[f_2(X, Y)] = 4.075, \quad \sigma[f_2(X, Y)] = 11.750, \quad (4.31)$$

w tabeli 4.3 zebrano wyniki dotyczące względnych błędów procentowych popełnianych przy estymacji $\mathbb{E}(f_2)$ i $\sigma(f_2)$, przy użyciu rozpatrywanych technik symulacji losowych. Z uwagi na przyjęty typ rozkładu prawdopodobieństwa zmiennych, nie jest zasadne w tym przypadku rozróżnianie metod LH-m i LH- μ . Na rysunkach 4.8 i 4.9 przedstawiono te same rezultaty w formie wykresów.

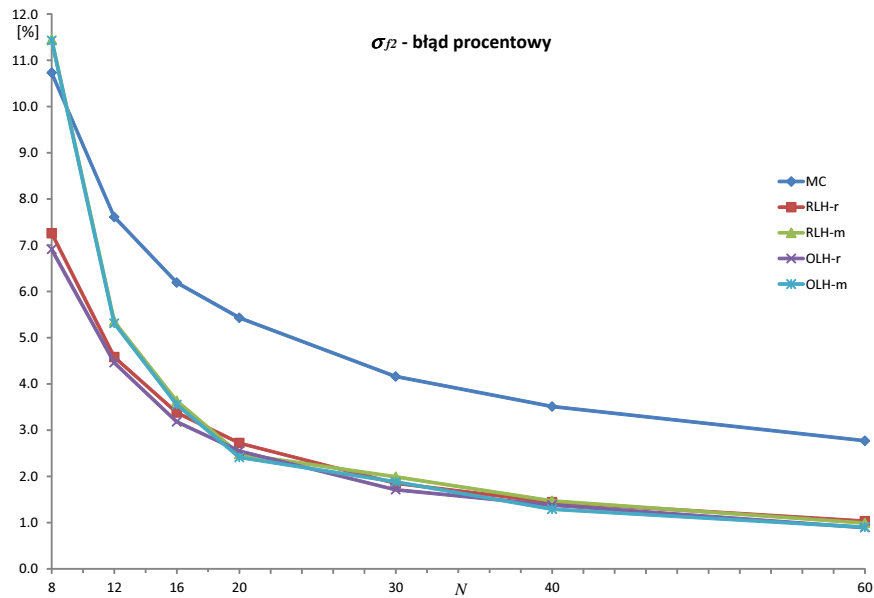
Tablica 4.3. Wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanej oraz odchylenia standardowego funkcji (4.30) przy użyciu różnych technik symulacji losowych i dla różnych wielkości próbek N . Przyjęto jednostajny rozkład zmiennych losowych. Odchylenia standardowe błędów podane są kursywą.

N	MC err[%]		RLH-r err[%]		RLH-m err[%]		OLH-m err[%]		OLH-m err[%]	
	μ_{f_2}	σ_{f_2}	μ_{f_2}	σ_{f_2}	μ_{f_2}	σ_{f_2}	μ_{f_2}	σ_{f_2}	μ_{f_2}	σ_{f_2}
$4n = 8$	82.20	10.73	37.60	7.26	29.55	11.45	37.09	7.37	29.89	11.43
	<i>60.76</i>	<i>9.86</i>	<i>29.21</i>	<i>4.88</i>	<i>12.57</i>	<i>2.42</i>	<i>28.92</i>	<i>4.80</i>	<i>6.22</i>	<i>2.11</i>
$6n = 12$	66.55	7.61	28.38	4.58	16.28	5.35	31.57	4.49	17.10	5.31
	<i>49.28</i>	<i>6.42</i>	<i>20.55</i>	<i>3.01</i>	<i>9.14</i>	<i>2.02</i>	<i>23.37</i>	<i>2.99</i>	<i>10.58</i>	<i>2.00</i>
$8n = 16$	57.12	6.19	24.00	3.38	20.50	3.63	22.34	3.32	20.61	3.55
	<i>42.67</i>	<i>5.02</i>	<i>16.59</i>	<i>2.27</i>	<i>8.22</i>	<i>1.70</i>	<i>15.08</i>	<i>2.48</i>	<i>9.05</i>	<i>1.61</i>
$10n = 20$	51.68	5.43	19.84	2.72	21.21	2.48	19.27	2.55	21.36	2.41
	<i>38.77</i>	<i>4.42</i>	<i>14.07</i>	<i>1.83</i>	<i>7.39</i>	<i>1.41</i>	<i>12.03</i>	<i>1.76</i>	<i>6.52</i>	<i>1.24</i>
$15n = 30$	41.89	4.16	12.62	1.85	4.95	1.99	11.60	1.83	3.15	1.88
	<i>31.72</i>	<i>3.25</i>	<i>9.61</i>	<i>1.26</i>	<i>3.65</i>	<i>1.18</i>	<i>9.34</i>	<i>1.08</i>	<i>2.28</i>	<i>0.83</i>
$20n = 40$	36.62	3.51	10.99	1.44	6.80	1.47	10.51	1.29	6.30	1.29
	<i>27.67</i>	<i>2.76</i>	<i>7.89</i>	<i>0.99</i>	<i>4.43</i>	<i>0.94</i>	<i>7.33</i>	<i>0.79</i>	<i>3.03</i>	<i>0.60</i>
$30n = 60$	29.53	2.77	6.74	1.03	3.41	0.99	5.63	0.91	1.98	0.90
	<i>22.07</i>	<i>2.10</i>	<i>5.09</i>	<i>0.73</i>	<i>2.55</i>	<i>0.70</i>	<i>4.52</i>	<i>0.57</i>	<i>1.47</i>	<i>0.43</i>

Przed omówieniem otrzymanych wyników warto jest przyjrzeć się uważnie wykresowi funkcji f_2 na rys. 4.7. Można spostrzec, iż dla wybranej wartości y zmienność funkcji w kierunku osi x jest nieznacząca, a wykres funkcji składa się w większej części z prawie poziomych segmentów połączonych powierzchniami o gwałtownym spadku. Rzut “stromych” fragmentów wykresu na płaszczyznę xy zajmuje stosunkowo małą powierzchnię w porównaniu do elementów poziomych. Pamiętając także o warstwowej naturze łacińskich hiperkostek, łatwiej jest zinterpretować wyniki symulacji.



Rys. 4.8. Wartość średnia względnego błęd procentowy przy estymacji wartości oczekiwanej funkcji (4.30) dla zmiennych losowych o rozkładzie jednostajnym.



Rys. 4.9. Wartość średnia względnego błęd procentowy przy estymacji odchylenia standardowego funkcji (4.30) dla zmiennych losowych o rozkładzie jednostajnym.

Pomimo niemonotonicznego charakteru funkcji, błąd estymacji zwykłą metodą Monte Carlo jest w większości przypadków wielokrotnie większy od błędów popełnianych przy użyciu pozostałych metod. Jedynym przypadkiem kiedy zaobserwowano odwrotną sytuację jest estymacja odchylenia standardowego funkcji za pomocą próbek generowanych przez hiperkostki z realizacjami wektora losowego w medianach dla $N = 8$. Co ciekawe, ze względu na omówiony powyżej charakterystyczny kształt wykresu funkcji, nie zaobserwowano znaczących różnic pomiędzy podejściem RLH a OLH. Trudno jest też jednoznacznie stwierdzić, czy korzystniejszy jest losowy wybór realizacji zmiennych wewnątrz komórki, czy też wybór median. Pierwsza z tych koncepcji jest bowiem lepsza przy estymacji odchylenia standardowego, a druga przy estymacji wartości średniej. W przypadku realizacji zmiennych losowych w medianach hiperkostki zastanawiająca jest niemonotoniczna zbieżność błędu estymacji wartości średniej wraz ze wzrostem wielkości próbki. Można to tłumaczyć w ten sposób, iż przy pewnych gęstościach podziału łańciskowej hiperkostki, zdeterminowany medianami komórek układ punktów eksperymentalnych systematycznie pomija pewne istotne dla wartości średniej obszary dziedziny funkcji (4.30), co w rezultacie prowadzi do większego błędu estymacji, niż w przypadku mniejszych lecz bardziej “reprezentatywnie” rozłożonych próbek.

4.1.4.3. Wielowymiarowa funkcja Rosenbrocka

W poniższym przykładzie, za pomocą rozpatrywanych technik symulacji losowych, przetestowano dokładność estymacji wartości oczekiwanej oraz odchylenia standardowego uogólnionej wielowymiarowej funkcji Rosenbrocka ([226] strona 49), która dla parzystej liczby zmiennych losowych n ma postać

$$f_3(\mathbf{X}) = \sum_{i=1}^{n/2} [100(X_{2i} - X_{2i-1}^2)^2 + (1 - X_{2i-1})^2]. \quad (4.32)$$

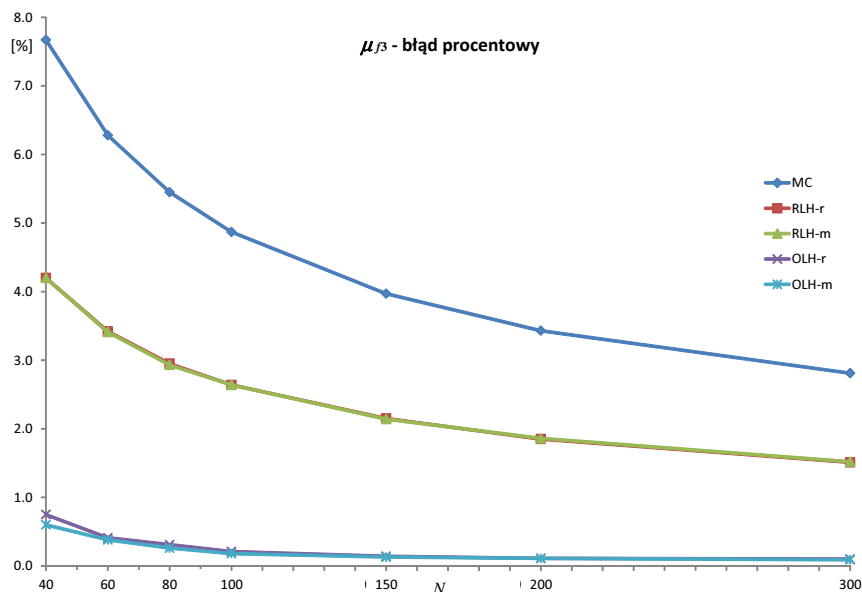
Przyjęto $n = 10$ zmiennych losowych o rozkładzie jednostajnym na przedziale $[0, 2]$. Obliczone na podstawie symulacji Monte Carlo ($N = 10^7$) wartości odniesienia wynoszą

$$\mathbb{E}[f_3(\mathbf{X})] = 935.0, \quad \sigma[f_3(\mathbf{X})] = 571.75. \quad (4.33)$$

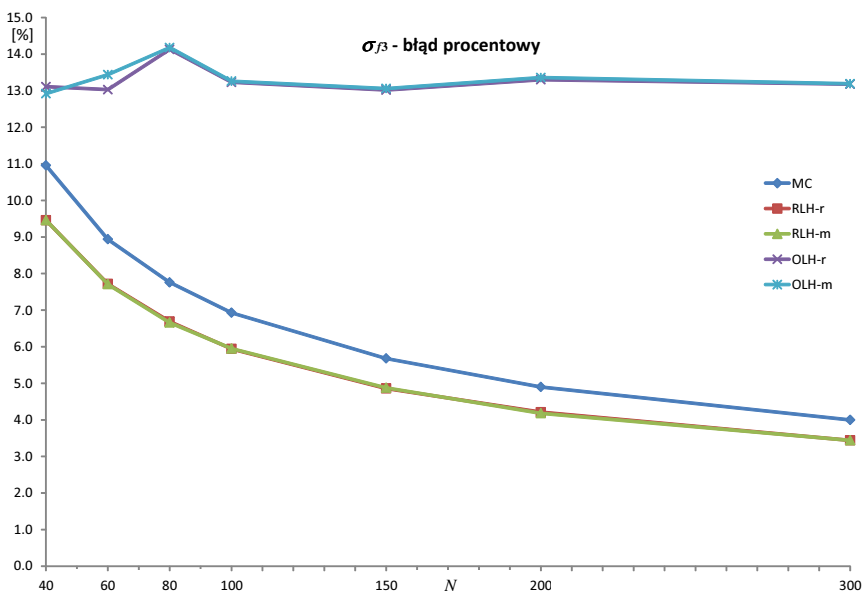
Ze względu na typ rozkładu prawdopodobieństwa zmiennych losowych rozpatrywano jedynie metody LH-r oraz LH-m. W tabeli 4.4 oraz na wykresach 4.10 i 4.11 zebrano wszystkie wyniki symulacji.

Tablica 4.4. Wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanej oraz odchylenia standardowego wielowymiarowej funkcji Rosenbrocka (4.32) przy użyciu różnych technik symulacji losowych i dla różnych wielkości próbek N . Przyjęto jednostajny rozkład zmiennych losowych. Odchylenia standardowe błędów podane są kursywą.

N	MC err[%]		RLH-r err[%]		RLH-m err[%]		OLH-r err[%]		OLH-m err[%]	
	μ_{f_3}	σ_{f_3}	μ_{f_3}	σ_{f_3}	μ_{f_3}	σ_{f_3}	μ_{f_3}	σ_{f_3}	μ_{f_3}	σ_{f_3}
$4n = 40$	7.67	10.96	4.20	9.46	4.20	9.46	0.75	13.11	0.60	12.92
	<i>5.79</i>	<i>8.28</i>	<i>3.18</i>	<i>7.25</i>	<i>3.16</i>	<i>7.21</i>	<i>0.57</i>	<i>6.48</i>	<i>0.45</i>	<i>6.54</i>
$6n = 60$	6.28	8.94	3.42	7.72	3.41	7.71	0.41	13.03	0.38	13.44
	<i>4.74</i>	<i>6.72</i>	<i>2.58</i>	<i>5.87</i>	<i>2.58</i>	<i>5.85</i>	<i>0.31</i>	<i>4.42</i>	<i>0.27</i>	<i>4.54</i>
$8n = 80$	5.45	7.76	2.95	6.69	2.93	6.66	0.31	14.13	0.26	14.18
	<i>4.11</i>	<i>5.85</i>	<i>2.23</i>	<i>5.08</i>	<i>2.22</i>	<i>5.06</i>	<i>0.22</i>	<i>3.44</i>	<i>0.19</i>	<i>3.53</i>
$10n = 100$	4.87	6.93	2.64	5.94	2.64	5.95	0.21	13.23	0.18	13.26
	<i>3.68</i>	<i>5.24</i>	<i>2.00</i>	<i>4.49</i>	<i>1.99</i>	<i>4.51</i>	<i>0.16</i>	<i>2.90</i>	<i>0.13</i>	<i>2.95</i>
$15n = 150$	3.97	5.68	2.15	4.86	2.14	4.88	0.14	13.02	0.13	13.06
	<i>3.00</i>	<i>4.27</i>	<i>1.62</i>	<i>3.67</i>	<i>1.62</i>	<i>3.68</i>	<i>0.10</i>	<i>2.18</i>	<i>0.10</i>	<i>2.23</i>
$20n = 200$	3.43	4.90	1.85	4.21	1.86	4.18	0.11	13.30	0.11	13.36
	<i>2.59</i>	<i>3.71</i>	<i>1.40</i>	<i>3.18</i>	<i>1.40</i>	<i>3.17</i>	<i>0.08</i>	<i>1.60</i>	<i>0.08</i>	<i>1.56</i>
$30n = 300$	2.81	4.00	1.51	3.44	1.52	3.44	0.10	13.18	0.09	13.19
	<i>2.12</i>	<i>3.02</i>	<i>1.14</i>	<i>2.61</i>	<i>1.15</i>	<i>2.60</i>	<i>0.07</i>	<i>1.25</i>	<i>0.07</i>	<i>1.30</i>



Rys. 4.10. Wartość średnia względnego błędu procentowego przy estymacji wartości oczekiwanej wielowymiarowej funkcji Rosenbrocka (4.32) dla zmiennych losowych o rozkładzie jednostajnym.



Rys. 4.11. Wartość średnia względnego błęd procentowy przy estymacji odchylenia standardowego wielowymiarowej funkcji Rosenbrocka (4.32) dla zmiennych losowych o rozkładzie jednostajnym.

Pomimo niemonotonicznego charakteru funkcji można zauważyć, że niezależnie od N , metody wykorzystujące próbki generowane przez RLH prowadzą do mniejszych błędów estymacji niż zwykła metoda Monte Carlo. Zupełnie nieistotny jest też sposób wyboru realizacji wektora \mathbf{X} wewnątrz komórek łańcuchowych hiperkostek – metody LH-r i LH-m dają tu praktycznie takie same wyniki. Symulacje oparte na optymalnych hiperkostkach zapewniają wyśmienite oszacowanie wartości oczekiwanej (średni błąd nawet dla $N = 4n$ nie przekracza 1%). Zaskakująco złe są niestety wyniki estymacji OLH odchylenia standardowego funkcji Rosenbrocka. W przypadku metod MC oraz RLH widoczny jest wyraźny trend malejący funkcji błęd wraz ze wzrostem wielkości próbki, zob. rys. 4.11, natomiast w przypadku OLH błąd ten ustabilizował się na poziomie około 13% bez widocznej tendencji spadkowej. Spostrzeżenie to prowadzić może do wniosku, iż dla pewnych funkcji estymatory odchylenia standardowego przy próbkach generowanych przez optymalne hiperkostki łańcuchowe mogą być w sposób istotny obciążone (w odróżnieniu od estymatora wariancji estymator (4.23) jest obciążony). Ze względu na długotrwały proces tworzenia planów OLH, trudne jest obliczenie wartości błęd estymacji odchylenia standardowego dla bardzo dużych N . Jest to także w sprzeczności z ideą samej metody, która ma za zadanie zapewniać wystarczająco dobrą dokładność statystyk dla małych próbek.

4.1.4.4. Funkcja wykładnicza

Ostatnią funkcją testową, jakiej użyjemy do naszych porównań jest funkcja

$$f_4(\mathbf{X}) = \exp\left(-0.2 \sum_{i=1}^n X_i^2\right). \quad (4.34)$$

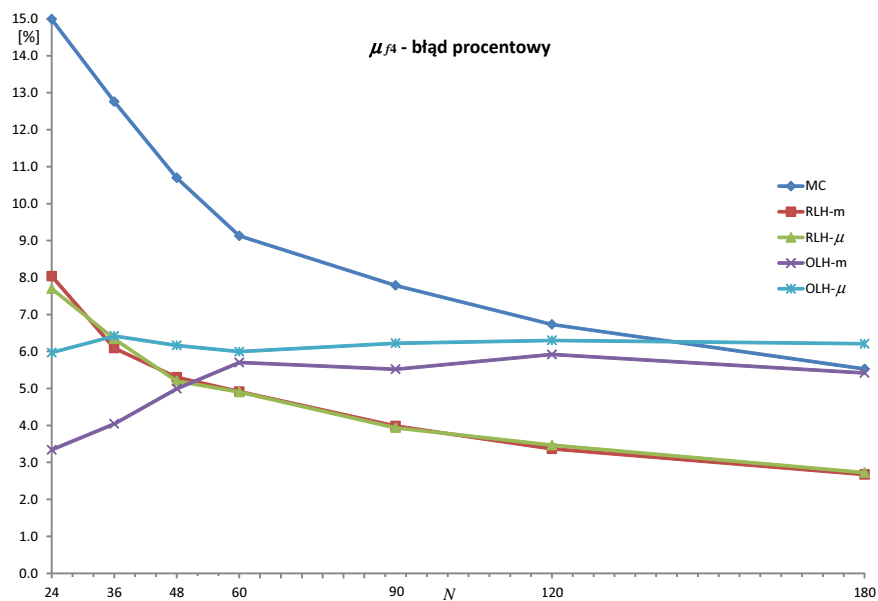
Przyjęto $n = 6$ zmiennych losowych o identycznych gaussowskich rozkładach prawdopodobieństwa danych przez $\mu_{X_i} = 1.0$ i $\sigma_{X_i} = 1.0$, $i = 1, \dots, 6$. Na podstawie symulacji Monte Carlo ($N = 10^7$) ustalono następujące wartości odniesienia:

$$\mathbb{E}[f_4(\mathbf{X})] = 0.1546, \quad \sigma[f_4(\mathbf{X})] = 0.1459. \quad (4.35)$$

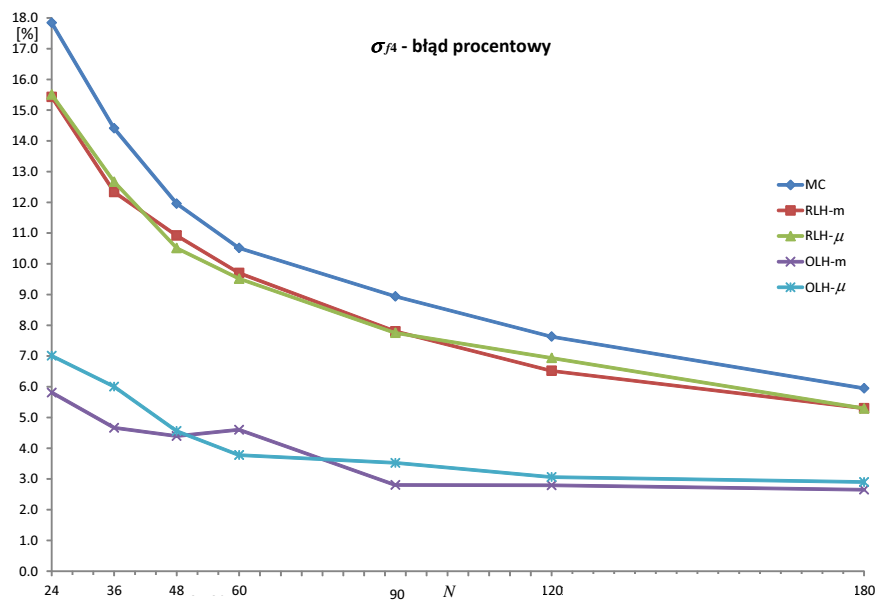
Podobnie jak w przypadku funkcji (4.27), ze względu na założony rozkład zmiennych, zasadne jest rozróżnienie dwóch wariantów wyboru ich realizacji wewnątrz komórek łańciskowej hiperkostki, tj. realizacji w medianach i wartościach oczekiwanych. Rozpatrywane są zatem metody MC, RLH-m, RLH- μ , OLH-m i OLH- μ . Wartości błędów estymacji szacowanych statystyk, odpowiadające wymienionym metodom dla różnych wielkości próbek, otrzymane na podstawie stukrotnego powtórzenia każdego zadania symulacji losowych zebrano w tabeli 4.5 oraz na wykresach 4.12 i 4.13.

Tablica 4.5. Wartości średnie oraz odchylenia standardowe względnych błędów procentowych popełnianych przy estymacji wartości oczekiwanej oraz odchylenia standardowego funkcji (4.34) przy użyciu różnych technik symulacji losowych i dla różnych wielkości próbek N . Przyjęto gaussowski rozkład zmiennych losowych. Odchylenia standardowe błędów podane są kursywą.

N	MC err[%]		RLH-m err[%]		RLH- μ err[%]		OLH-m err[%]		OLH- μ err[%]	
	μ_{f_4}	σ_{f_4}	μ_{f_1}	σ_{f_4}	μ_{f_1}	σ_{f_4}	μ_{f_4}	σ_{f_4}	μ_{f_1}	σ_{f_4}
$4n=24$	14.99	17.84	8.04	15.43	7.70	15.51	3.34	5.81	5.97	7.01
	<i>11.55</i>	<i>12.80</i>	<i>6.13</i>	<i>11.13</i>	<i>5.81</i>	<i>11.09</i>	<i>1.95</i>	<i>4.38</i>	<i>2.48</i>	<i>4.52</i>
$6n=36$	12.76	14.41	6.09	12.34	6.34	12.67	4.04	4.66	6.42	6.00
	<i>9.45</i>	<i>10.28</i>	<i>4.77</i>	<i>9.08</i>	<i>4.78</i>	<i>9.26</i>	<i>1.91</i>	<i>3.32</i>	<i>2.31</i>	<i>3.92</i>
$8n=48$	10.70	11.96	5.30	10.92	5.20	10.51	4.99	4.39	6.16	4.55
	<i>8.29</i>	<i>9.09</i>	<i>3.84</i>	<i>7.62</i>	<i>3.89</i>	<i>7.73</i>	<i>1.72</i>	<i>3.17</i>	<i>1.99</i>	<i>3.66</i>
$10n=60$	9.13	10.51	4.91	9.70	4.90	9.51	5.70	4.60	5.99	3.77
	<i>7.25</i>	<i>8.01</i>	<i>3.58</i>	<i>7.06</i>	<i>3.64</i>	<i>7.12</i>	<i>1.52</i>	<i>3.32</i>	<i>1.46</i>	<i>3.15</i>
$15n=90$	7.79	8.94	3.98	7.80	3.94	7.75	5.52	2.80	6.22	3.52
	<i>6.00</i>	<i>6.56</i>	<i>2.93</i>	<i>5.76</i>	<i>2.86</i>	<i>5.76</i>	<i>0.94</i>	<i>2.59</i>	<i>1.06</i>	<i>2.67</i>
$20n=120$	6.73	7.63	3.37	6.52	3.47	6.94	5.91	2.79	6.30	3.06
	<i>5.16</i>	<i>5.84</i>	<i>2.62</i>	<i>5.01</i>	<i>2.50</i>	<i>5.01</i>	<i>0.87</i>	<i>2.22</i>	<i>0.82</i>	<i>2.44</i>
$30n=180$	5.53	5.95	2.67	5.30	2.73	5.29	5.42	2.65	6.21	2.90
	<i>4.24</i>	<i>4.71</i>	<i>2.03</i>	<i>4.04</i>	<i>2.13</i>	<i>4.15</i>	<i>0.82</i>	<i>2.15</i>	<i>0.75</i>	<i>2.10</i>



Rys. 4.12. Wartość średnia względnego błędu procentowego przy estymacji wartości oczekiwanej wielowymiarowej funkcji (4.34) dla zmiennych losowych o rozkładzie normalnym.

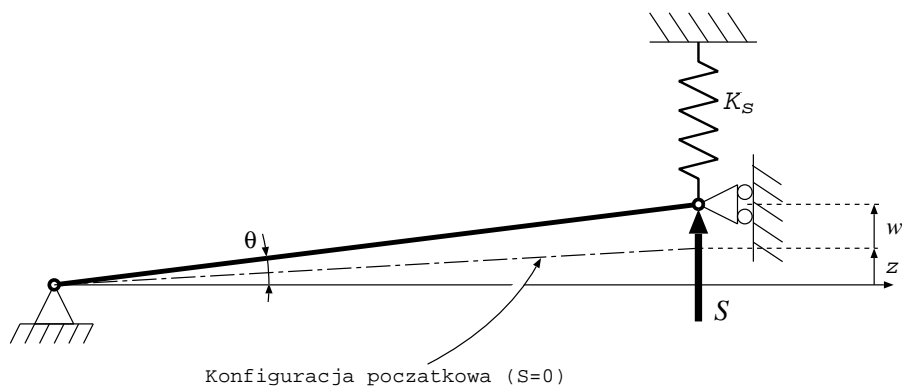


Rys. 4.13. Wartość średnia względnego błędu procentowego przy estymacji odchylenia standardowego wielowymiarowej funkcji (4.34) dla zmiennych losowych o rozkładzie normalnym.

Analizując uzyskane wyniki można spostrzec, że i w tym przypadku metoda RLH góruje dokładnością estymacji nad zwykłą metodą Monte Carlo. Zaskakujące są natomiast oszacowania otrzymane przy użyciu hiperkostek optymalnych. W przeciwieństwie do poprzedniego przykładu, metody OLH prowadzą do bardzo dobrej estymacji odchylenia standardowego, zob. rys. 4.13, a estymator wartości oczekiwanej jest wyraźnie obciążony ze średnim błędem stabilizującym się na około 6%, rys. 4.12. Nie zaobserwowano wyraźnej jakościowej różnicy pomiędzy wersjami LH-m i LH- μ .

4.1.5. Przykład: Kratownica Misesa

Poniższy przykład analizy dokładności oszacowania momentów statystycznych odpowiedzi konstrukcji za pomocą symulacji losowych zaczerpnięty został z pracy [237]. Model konstrukcji stanowi małowyniosła jednoprętowa kratownica Misesa pokazana na rys. 4.14. Jest to układ składający się z odpowiednio zamocowanego pręta obciążonego w węźle siłą pionową oraz ze sprężyny. W przypadku obciążenia skierowanego w dół układ podatny jest na utratę stateczności w postaci przeskoku do innego położenia równowagi.



Rys. 4.14. Kratownica Misesa. Pokazana siła S ma znak dodatni.

Nieliniowe zachowanie kratownicy wynika z założeń kinematycznych (duże przemieszczenia), a nie ze związków konstytutywnych, które przyjęto jako liniowo sprężyste. Poszukiwaną wielkością jest przemieszczenie pionowe prawego końca pręta w będące funkcją modułu Younga E , pola przekroju pręta A i jego długości początkowej l , położenia prawego węzła przed przyłożeniem obciążenia $z > 0$, sztywności sprężyny K_s oraz siły S . Przemieszczenie w wyznaczyć można z następującego równania (por. [37]):

$$S = \frac{EA}{l^3} \left(z^2 w + \frac{3}{2} z w^2 + \frac{1}{2} w^3 \right) + K_s w. \quad (4.36)$$

Wyprowadzając tę zależność założono mały wymiar kąta θ (zob. rys. 4.14), z czego wynika $z, w \ll l$. Równanie (4.36) jest wielomianem trzeciego stopnia zmiennej w . Można pokazać, że w zależności od wartości parametrów (współczynników) równanie to ma jedno albo trzy miejsca zerowe. W przypadku trzech pierwiastków jako rozwiązanie przyjmowany jest ten pierwiastek, który odpowiada najmniejszej wielkości przemieszczenia. Ma to miejsce gdy skierowana do dołu siła S jest mniejsza od krytycznej wartości powodującej przeskok układu pręt-sprężyna do nowego położenia równowagi. Przypadek, gdy równanie (4.36) ma tylko jeden pierwiastek odpowiada bądź konfiguracji konstrukcji po przeskoku, bądź też takiej pracy konstrukcji, w której ten typ utraty stateczności nie występuje ze względu na dużą sztywność sprężyny K_s .

Założmy następnie, że parametry konstrukcji są niezależnymi zmiennymi losowymi o rozkładach wymienionych w tab. 4.6. W przypadku zmiennych o rozkładzie normalnym w tabeli podano ich wartość średnią oraz współczynnik zmienności, a w przypadku zmiennych o rozkładzie jednostajnym lewą i prawą granicę.

Tablica 4.6. Kratownica Misesa: zmienne losowe

Zmienna	Rozkład	$\mathbb{E}(X)$ /Lewa granica	$\nu(X)$ /Prawa granica
E	normalny	$5 \cdot 10^5 \text{ N/mm}^2$	0.02
A	normalny	100 mm^2	0.02
l	normalny	2500 mm	0.03
z	jednostajny	15 mm	27 mm
K_s	normalny	0.9 N/mm	0.2
S	jednostajny	-40 N	-30 N

W celu oceny dokładności oszacowania wybranych losowych charakterystyk przemieszczenia w na podstawie statystyk, których wartości generowane są przez różne metody symulacji losowych, należy najpierw jak najdokładniej określić ich rzeczywistą wartość. Niestety, w tym przypadku jest to możliwe jedynie za pomocą symulacji Monte Carlo, przyjmując jak najliczniejszą próbkę. Na podstawie $N = 100\,000$ losowo wygenerowanych punktów eksperymentalnych obliczono następujące wartości odniesienia dla statystyk:

- wartość oczekiwana wielkości przemieszczenia w

$$\mathbb{E}(|w|) = 40.49 \text{ mm}, \quad (4.37)$$

- odchylenie standardowe wielkości przemieszczenia w

$$\sigma(|w|) = 5.89 \text{ mm}, \quad (4.38)$$

- macierz korelacji

	E	A	l	z	K_s	$ S $	$ w $
E	1.0	0.0	0.0	0.0	0.0	0.0	0.0
A	0.0	1.0	0.0	0.0	0.0	0.0	0.0
l	0.0	0.0	1.0	0.0	0.0	0.0	-0.01
z	0.0	0.0	0.0	1.0	0.0	0.0	0.67
K_s	0.0	0.0	0.0	0.0	1.0	0.0	-0.65
$ S $	0.0	0.0	0.0	0.0	0.0	1.0	0.26
$ w $	0.0	0.0	-0.01	0.67	-0.65	0.26	1.0

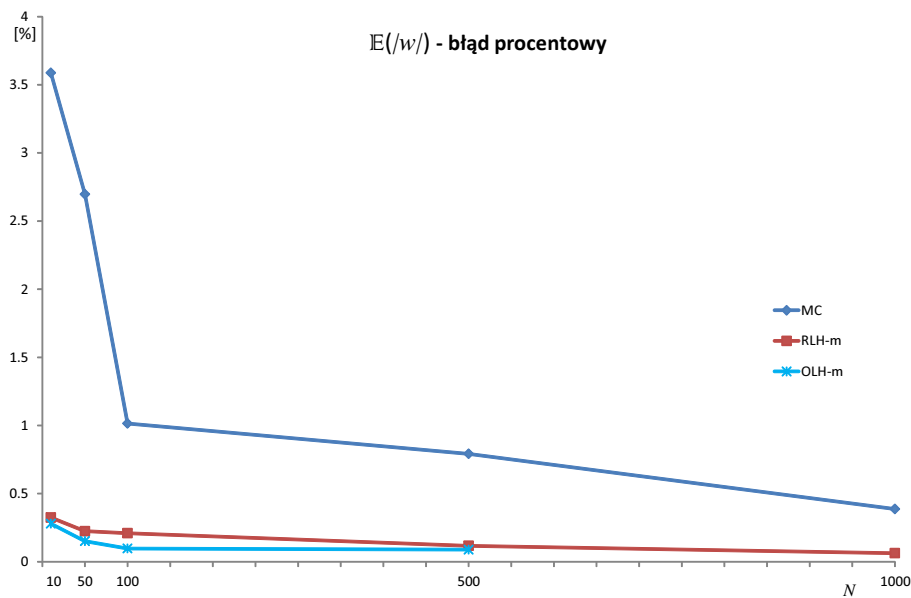
(4.39)

Analizując macierz (4.39) można spostrzec, że dzięki zastosowaniu dostatecznie dużej próbki nie pojawiły się w macierzy nieprawdziwe, niezerowe wartości współczynników korelacji dla zmiennych niezależnych. Ponadto, na podstawie wartości współczynników w ostatnim wierszu/kolumnie wywnioskować można, iż wielkość ugięcia kratownicy jest dodatnio skorelowana z jej wyniosłością z , ujemnie skorelowana ze sztywnością sprężyny K_s oraz prawie zupełnie nieskorelowana z pozostałymi zmiennymi losowymi. Relacje te są oczywiście w pewnym stopniu zależne od przyjętych zakresów zmienności parametrów konstrukcji.

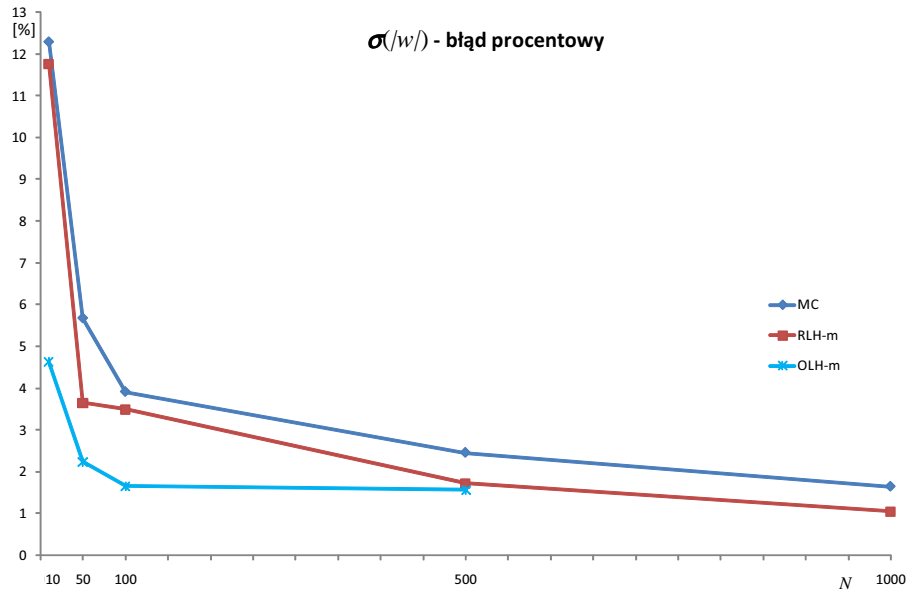
Porównanie różnych metod symulacji losowych przeprowadzimy estymując $\mathbb{E}(|w|)$, $\sigma(|w|)$ oraz dwa “znaczące” współczynniki korelacji $\rho_{|w|K_s}$ i $\rho_{|w|z}$ za pomocą zwykłej metody Monte Carlo, symulacji losową hiperkostką łacińską (wersja RLH-m) oraz symulacji optymalną hiperkostką łacińską (wersja OLH-m). Przyjmując za dokładne wartości odniesienia (4.37)-(4.39), w tabeli 4.7 oraz na wykresach 4.15-4.18 przedstawiono wartości średnich względnych błędów procentowych otrzymanych na podstawie 100 powtórzeń symulacji każdą z wymienionych metod.

Tablica 4.7. Kratownica Misesa. Średni błąd procentowy estymacji przy użyciu różnych metod symulacji losowych oraz dla próbek o różnej wielkości. W przypadku $N > 500$ symulacje OLH-m nie zostały wykonane z uwagi na długi czas tworzenia planów OLH.

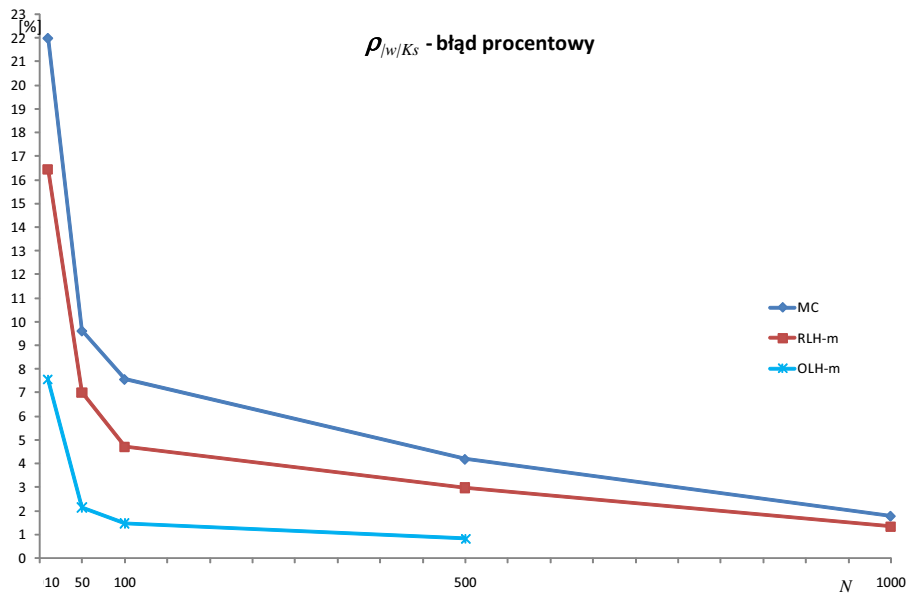
N		10	50	100	500	1000	5000
$\mathbb{E}(w)$ err [%]	MC	3.59	2.70	1.01	0.79	0.39	0.17
	RLH-m	0.32	0.22	0.21	0.12	0.06	0.04
	OLH-m	0.28	0.15	0.1	0.09	-	-
$\sigma_{ w }$ err [%]	MC	12.29	5.68	3.92	2.46	1.65	1.09
	RLH-m	11.76	3.66	3.50	1.73	1.06	0.65
	OLH-m	4.63	2.24	1.65	1.57	-	-
$\rho_{ w K_s}$ err [%]	MC	22.01	9.62	7.58	4.20	1.79	1.07
	RLH-m	16.45	7.02	4.70	2.99	1.35	0.55
	OLH-m	7.56	2.16	1.48	0.84	-	-
$\rho_{ w z}$ err [%]	MC	33.45	14.75	8.09	4.35	3.61	1.32
	RLH-m	15.95	9.51	6.58	3.80	1.78	0.89
	OLH-m	8.07	5.10	4.09	3.70	-	-



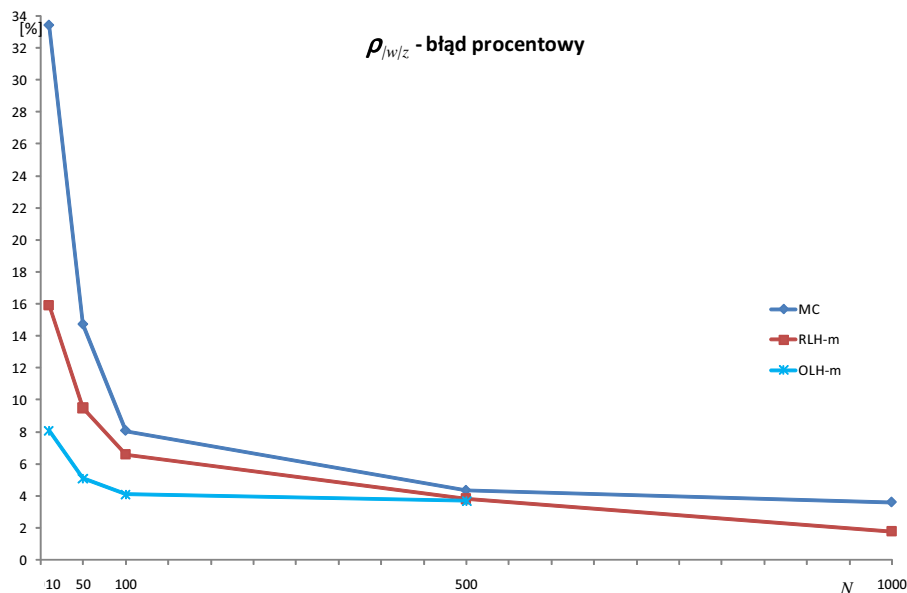
Rys. 4.15. Kratownica Misesa: Estymacja wartości średniej ugięcia. Średni względny błąd procentowy w przypadku różnych metod symulacyjnych i różnych wielkości próbki.



Rys. 4.16. Kratownica Misesa: Estymacja odchylenia standardowego ugięcia. Średni względny błąd procentowy w przypadku różnych metod symulacyjnych i różnych wielkości próbek.



Rys. 4.17. Kratownica Misesa: Estymacja współczynnika $\rho_{w|K_s}$. Średni względny błąd procentowy w przypadku różnych metod symulacyjnych i różnych wielkości próbek.



Rys. 4.18. Kratownica Misesa: Estymacja współczynnika $\rho_{|w/z}$. Średni względny błąd procentowy w przypadku różnych metod symulacyjnych i różnych wielkości próbki.

Porównując otrzymane wartości błędów można stwierdzić, że w rozpatrywanym przypadku:

- Symulacje z wykorzystaniem OLH prowadzą do najdokładniejszej estymacji. W zależności od statystyki oraz wielkości próbki (w szczególności dla małych próbek) błąd estymacji metodą OLH-m jest wielokrotnie mniejszy od błędów MC. Na przykład, przy oszacowaniu wartości średniej ugięcia, próbka $N = 10$ daje trzynastokrotnie mniejszy błąd estymacji.
- Błędy estymacji metodą OLH-m są zawsze mniejsze niż błędy otrzymane przy użyciu hiperkostki łacińskiej bez optymalizacji (nawet trzykrotnie w przypadku $\rho_{|w|K_s}$). O ile zysk ze stosowania OLH przy estymacji wartości średniej jest niewielki to w przypadku momentów statystycznych wyższych rzędów jest już znaczący (szczególnie dla $N < 100$).

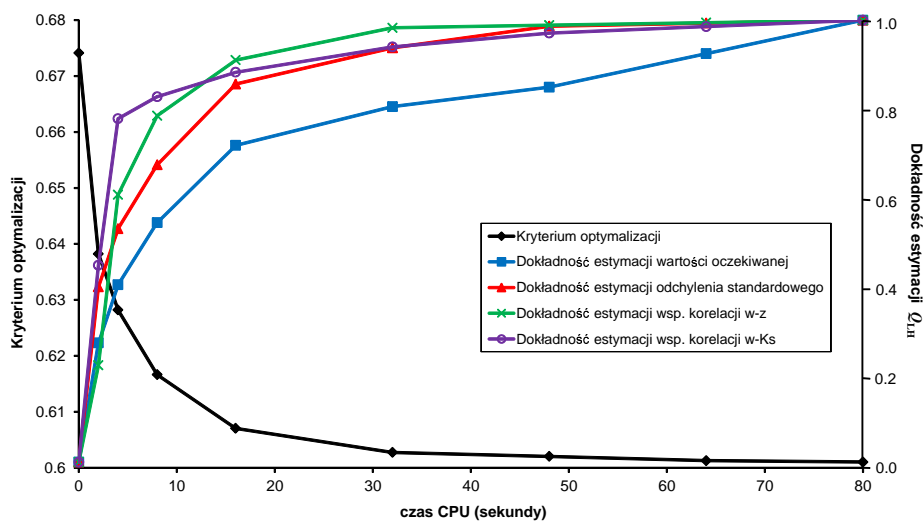
Wydaje się, że warto w tym miejscu zastanowić się nad “poziomem optymalności” planów OLH, który zapewnia metodom korzystającym z optymalnych hiperkostek istotną przewagę (mierzoną błędem estymacji) w stosunku do tych metod symulacyjnych, gdzie próbki generowane są przez RLH. Na podstawie licznych testów numerycznych zauważono, że zarówno w przypadku algorytmu CP jak również algorytmu genetycznego (zob. dodatek B) największa redukcja wartości kryterium optymalizacji (B.3) występuje na pierwszych iteracjach algorytmu CP lub po stosunkowo niedużej liczbie populacji algorytmu genetycznego.

Z drugiej strony, etap dochodzenia do końcowej postaci planu OLH trwa znacznie dłużej niż początkowe iteracje, prowadząc w większości przypadków do niewielkiej poprawy wartości kryterium. Niesłychanie ważny jest zatem taki dobór parametrów zbieżności metod tworzenia planów OLH, aby łańciskie hiperkostki były optymalne również ze względu na czas trwania procesu minimalizacji kryterium (B.3).

Na rys. 4.19 pokazano wykresy zmiany dokładności estymacji wybranych statystyk w funkcji czasu trwania algorytmu CP użytego do optymalizacji położenia punktów łańciskiej hiperkostki dla $N = 100$ i $n = 6$. Dokładność estymacji mierzona jest przy pomocy następującego wskaźnika:

$$Q_{LH}(t) = 1 - \frac{Err_{RLH}(t) - Err_{OLH}}{Err_{RLH} - Err_{OLH}}, \quad (4.40)$$

gdzie Err_{RLH} jest średnim względnym błędem procentowym odpowiadającym metodzie RLH-m, Err_{OLH} jest takim błędem w przypadku zoptymalizowanej hiperkostki OLH-m, a Err_{LH} odpowiada błędowi estymacji gdy jako próbkę użyto “nie w pełni optymalnej” hiperkostki (algorytm CP przerwano po czasie t).

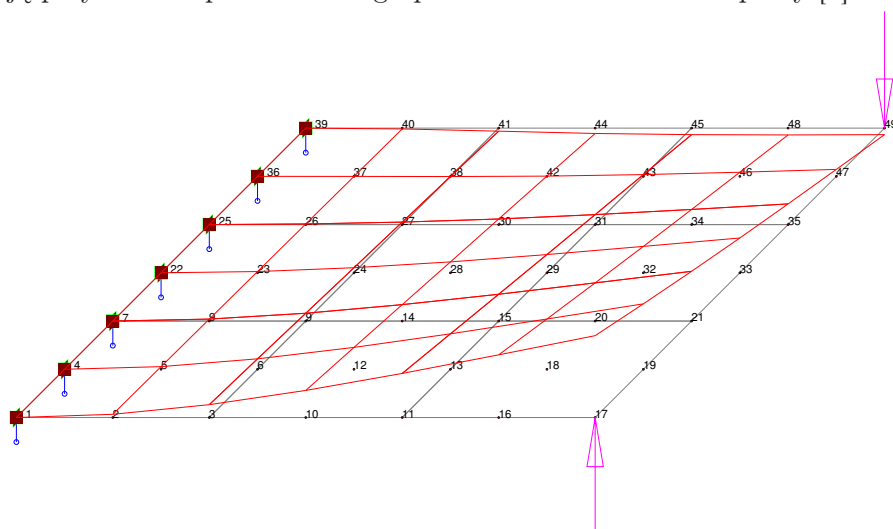


Rys. 4.19. Zmiany dokładności estymacji w funkcji czasu pracy procesora (CPU) użytego do poprawienia przestrzennej jednorodności próbek w celu otrzymania planu OLH. Dolna krzywa przedstawia historię zmian wartości kryterium optymalizacji (B.3) dla hiperkostki 100×6 ($N = 100$, $n = 6$).

Analizując powyższe wykresy łatwo zauważyć, że zasadniczą poprawę dokładności estymacji uzyskuje się już na początku procesu optymalizacji. Niestety, trudno jest zaproponować ogólną regułę uzależniającą dokładność estymacji od wartości kryterium (B.3), a tym bardziej od czasu trwania procesu optymalizacji. Dokładność (4.40) zależy zarówno od rozmiaru zadania i liczebności próbki, jak też od typu estymowanej wielkości. Na pewno jednak można stwierdzić, iż w przypadkach gdy losowanie typu OLH prowadzi do dokładniejszych oszacowań w stosunku do RLH, to nawet ograniczona liczba iteracji np. algorytmu CP, może pomóc w znacznym polepszeniu jakości estymacji.

4.1.6. Przykład: Płyta wspornikowa

Przykładem niejawniej funkcji wielu skorelowanych zmiennych losowych może być ugięcie sprężystej płyty wspornikowej o losowej grubości. Rozpatrywana tutaj płyta przedstawiona jest na rys. 4.20. Podobnie jak w przypadku kratownicy Misesa zadanie to pochodzi z artykułu autora [237] i jest zmodyfikowaną wersją przykładu zaprezentowanego przez Alonso i Collado w pracy [3].



Rys. 4.20. Siatka elementów skończonych oraz postać odkształcona płyty

Poza możliwością dokonania porównania efektywności różnych metod symulacji losowych przykład ten wybrano również w celu zademonstrowania zastosowania analizy głównych składowych (zob. podrozdział 4.1.3) do analizy wyników symulacji oraz do redukcji modelu stochastycznego.

Kwadratowa płyta, $3\text{m} \times 3\text{m}$, dyskretyzowana przy pomocy elementów skończonych, utwierdzona jest wzdłuż jednej krawędzi oraz obciążona dwiema siłami: F_1 - przyłożoną do węzła 17 i skierowaną w górę oraz F_2 - przyłożoną do węzła 49 i skierowaną w dół.

Probabilistyczny opis zadania stanowią następujące zmienne losowe: siła węzłowa F_1 o normalnym rozkładzie prawdopodobieństwa danym przez wartość oczekiwaną $\mu_{F_1} = 80\text{kN}$ i odchylenie standardowe $\sigma_{F_1} = 15\text{kN}$, siła węzłowa F_2 - rozkład normalny, $\mu_{F_2} = 60\text{kN}$, $\sigma_{F_2} = 10\text{kN}$, moduł Younga E - rozkład log-normalny, $\mu_E = 210000\text{MPa}$, $\sigma_E = 21000\text{MPa}$, współczynnik Poissona ν - rozkład log-normalny, $\mu_\nu = 0.3$, $\sigma_\nu = 0.03$. Grubość płyty t modelowana jest jako dwuwymiarowe, jednorodnie gaussowskie pole losowe (zob. dodatek C) o wartości oczekiwanej $\mu_t = 0.06\text{m}$, odchyleniu standardowym $\sigma_t = 0.003\text{m}$ oraz funkcji korelacji $\rho(\mathbf{z}, \mathbf{z}')$ danej wzorem (por. (C.10))

$$\rho(\mathbf{z}, \mathbf{z}') = \exp\left(-\frac{\|\mathbf{z} - \mathbf{z}'\|}{a}\right), \quad (4.41)$$

gdzie \mathbf{z} i \mathbf{z}' są wektorami współrzędnych punktów na powierzchni środkowej płyty, natomiast a jest długością korelacji, zob. (C.12). W niniejszym przykładzie przyjęto $a = 4\text{m}$.

Aby wyznaczyć przemieszczenia punktów płyty za pomocą metody elementów skończonych, konieczna jest uprzednia dyskretyzacja pola losowego grubości płyty. Spośród najbardziej popularnych metod dyskretyzacji pól losowych należy wymienić: metodę punktów środkowych, metodę funkcji kształtu, metodę optymalnej aproksymacji liniowej czy też metody rozwinięcia w szereg, jak np. metoda Karhunenena-Loeve'a. Bliższemu opisowi tych zagadnień poświęcono dodatek C.

Ogólnie rzecz ujmując, dyskretyzacja jest procedurą aproksymacji pola losowego za pomocą skończonej liczby zmiennych losowych, a jednym z najważniejszych celów w doborze metody dyskretyzacji jest uzyskanie jak najlepszej aproksymacji oryginalnego pola przy pomocy jak najmniejszej liczby zmiennych. Ponieważ w analizowanym przykładzie dyskretyzacja pola losowego służy przede wszystkim jako "generator" skorelowanych zmiennych losowych (w celu porównania efektywności metod symulacji losowych dla różnych wielkości modelu stochastycznego) dlatego jako metodę dyskretyzacji wybrano metodę punktów środkowych, zob. C.2. Metoda ta polega na reprezentacji pola losowego wewnątrz każdego elementu skończonego przez jedną zmienną losową zdefiniowaną jako wartość tego pola w środku ciężkości elementu. Nie jest to z pewnością bardzo efektywna i dokładna metoda aproksymacji, ale daje ona możliwość łatwego

generowania wielowymiarowych problemów analizy stochastycznej poprzez zwykle zwiększanie gęstości siatki.

Rozpatrywano trzy przypadki:

- a) siatka 3×3 , 9 elementów, 49 węzłów, $n = 13$ zmiennych losowych,
- b) siatka 6×6 , 36 elementów, 169 węzłów, $n = 40$ zmiennych losowych,
- c) siatka 9×9 , 81 elementów, 361 węzłów, $n = 85$ zmiennych losowych.

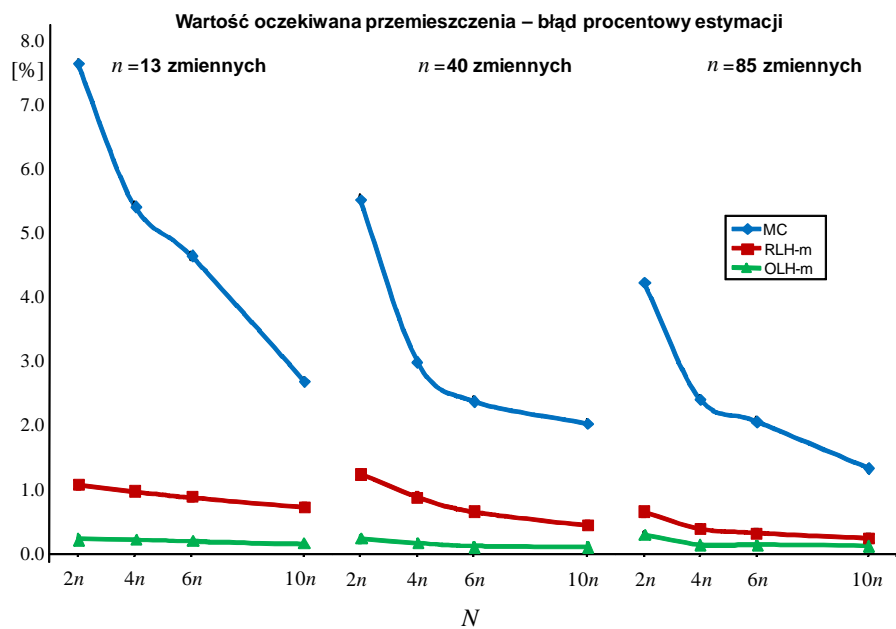
W modelu MES użyto dziewięciowęzłowych elementów skończonych płyty średniej grubości typu Reisnera-Mindlina (zob. np. [98]). Jako miarę odpowiedzi obciążonej siłami skupionymi konstrukcji przyjęto przemieszczenia niepodpartych węzłów.

Na początku, przy pomocy analizy głównych składowych sprawdzono, które z przemieszczeń węzłowych są najbardziej reprezentatywne dla losowego rozrzutu odpowiedzi konstrukcji wyrażanego macierzą kowariancji przemieszczeń. Przyjmując podział na elementy skończone jak w wariancie a) przeprowadzono symulację Monte Carlo dla próbki o $N = 200000$. Wyniki tej symulacji posłużyły do obliczenia macierzy kowariancji, a następnie wartości głównych jako wartości własnych tej macierzy, zob. punkt 4.1.3. Poniżej, wypisano procentowy udział pierwszych dziesięciu wartości głównych w ich sumie:

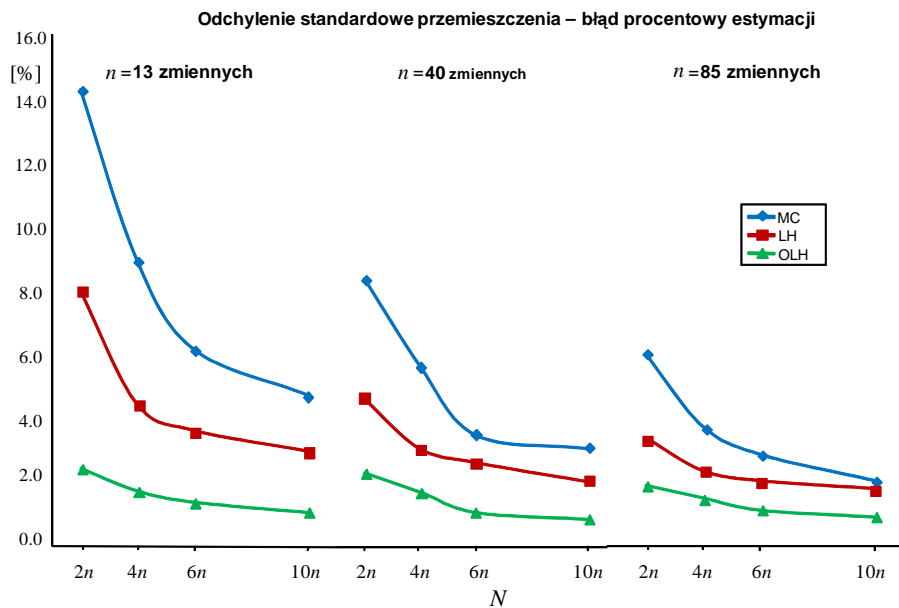
86.4%, 13.5%, 0.033%, 0.024%, 0.0083%, 0.0015%, 0.0006%, 0.0003%, 0.0001%, 0.0000%.

Jak widać, dwie pierwsze wartości główne zdecydowanie dominują i stanowią 99.93% sumy wszystkich wartości. Z tego powodu, jako najbardziej reprezentatywne dla rozrzutu odpowiedzi konstrukcji należy wybrać te losowe odpowiedzi, których rzuty na dwie pierwsze osie główne \mathbf{v}_1 i \mathbf{v}_2 , zob. (4.24), są największe. Zgodnie z intuicją, z analizy tej wynika, że przemieszczenia obciążonych węzłów narożnych cechują się największym rozrzutem i to właśnie one wnoszą najistotniejszy wkład w losowy charakter wyników. W związku z tym, ocena różnych metod symulacji losowych dokonana będzie przez porównanie błędów estymacji wartości oczekiwanej i odchylenia standardowego przemieszczenia węzła oznaczonego numerem 49 na rys. 4.20 oraz estymacji pierwszej wartości głównej.

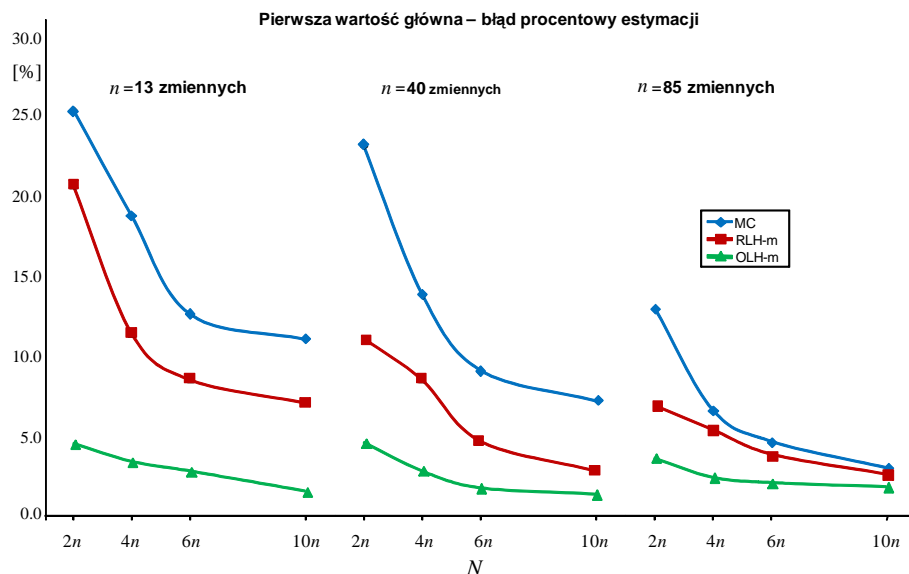
Tak jak w poprzednim przykładzie, porównywane są metody MC, RLH-m i OLH-m. W każdym przypadku rozpatrywano 4 wielkości próbki: $N = 2n, 6n, 8n$ i $10n$. Przyjmując jako wartości odniesienia wyniki symulacji Monte Carlo dla $N = 200000$, na podstawie dziesięciokrotnego powtórzenia każdej symulacji losowej, na rysunkach (4.21)–(4.23) przedstawiono wykresy średnich względnych błędów procentowych estymacji.



Rys. 4.21. Średni względny błąd procentowy estymacji wartości oczekiwanej przemieszczenia węzła narożnego dla różnych wielkości próbki w przypadku trzech różnych siatek elementów skończonych (co pociąga za sobą różną liczbę zmiennych losowych).



Rys. 4.22. Średni względny błąd procentowy estymacji odchylenia standardowego przemieszczenia węzła narożnego.



Rys. 4.23. Średni względny błąd procentowy estymacji pierwszej wartości głównej.

Na podstawie otrzymanych wyników można zauważyć, że:

- Metoda symulacji OLH-m okazała się być najdokładniejszą we wszystkich rozpatrywanych przypadkach. Maksymalny średni błąd estymacji nie przekraczał zazwyczaj 3% i był wielokrotnie niższy od odpowiednich błędów z symulacji Monte Carlo.
- Podobnie jak w przypadku kratownicy Misesa, przewaga metody OLH-m nad RLH-m zmniejsza się wraz ze wzrostem wielkości próbki. Można to przypisywać zarówno lepszemu wypełnieniu przestrzeni realizacji punktami eksperymentalnymi jak również trudnościami w otrzymaniu wystarczająco dobrych planów OLH dla wysokich wartości N oraz n . Natomiast w przeciwieństwie do poprzedniego przykładu, trudno jest na podstawie przebadanych przypadków zaproponować “granicę” przydatności stosowania metod bazujących na zoptymalizowanych łańciskich hiperkostkach w stosunku do metod RLH.
- Niezależnie od liczby zmiennych losowych, wykresy błędu metody OLH-m są bardzo podobne, co – porównując z tendencją zmian wykresów MC i RLH-m – świadczy o problemach ze znalezieniem planu optymalnego dla dużych hiperkostek.

4.1.7. Wnioski dotyczące technik symulacji losowych wykorzystujących łańciskie hiperkostki

Wykonane testy numeryczne potwierdziły wnioski teoretyczne przedstawione w punkcie 4.1.2 dotyczące większej dokładności metod symulacyjnych wykorzystujących próbki generowane przez łańciskie hiperkostki (RLH) w stosunku do zwykłej metody Monte Carlo. Co ciekawe, niezależnie od typu funkcji, w żadnym z siedmiu przykładów nie zaobserwowano przypadku kiedy błąd estymacji metodą RLH byłby większy od błędu metody MC. Wynik ten w świetle wniosków z twierdzenia 4.1 stanowić może pozytywne zaskoczenie.

Dużo trudniejsze do podobnej oceny są metody symulacji losowych wykorzystujące koncepcję optymalnej łańciskiej hiperkostki. Z powodu braku odpowiednich rozważań teoretycznych oraz częstej konieczności stosowania w praktyce hiperkostek sub-optymalnych, wnioski na temat metod OLH opierać można jedynie na przeprowadzonych testach numerycznych. Chociaż w większości rozpatrywanych przypadków metody OLH dają doskonale oszacowanie momentów statystycznych funkcji losowych nawet dla małych próbek, to można wskazać przypadki (zob. rys. 4.11 i rys. 4.12) kiedy użycie próbek generowanych przez optymalne hiperkostki prowadzi do znacznych, trudnych do zredukowania błędów estymacji. Z drugiej strony, problemy te występowały jedynie w dwóch przypadkach estymacji momentów analitycznych funkcji testowych, natomiast metody OLH były bezkonkurencyjne w estymacji momentów funkcji przemieszczeń węzłów konstrukcji. Podsumowując, z punktu widzenia efektywności symulacji losowych metoda RLH wydaje się być zawsze "bezpiecznym" rozwiązaniem, a próbki utworzone przez OLH warto stosować szczególnie wtedy, gdy rozważana funkcja jest kosztowna w obliczeniach i nie można pozwolić sobie na otrzymanie jej wartości dla zbyt wielu realizacji zmiennych losowych.

Na podstawie wykresów błędów estymacji dla analitycznych funkcji testowych, zob. rysunki 4.3-4.6 i 4.8-4.13, nie można jednoznacznie ustalić, która z metod wyboru realizacji zmiennych losowych wewnątrz komórki łańciskiej hiperkostki jest najlepsza. Jednak pomimo braku istotnych jakościowo różnic pomiędzy metodami LH-r, LH-m i LH- μ , po dokładniejszej analizie wyników można skłaniać się ku wyborowi wersji LH-m jako tej, która najczęściej prowadziła do najmniejszego błędu estymacji. Metody wykorzystujące mediany rozkładów prawdopodobieństwa wewnątrz komórek są również znacznie szybsze w porównaniu do tych, które wymagają numerycznego wyznaczenia wartości średnich.

4.2. Metoda rozwinięcia w szereg Taylora

Alternatywnym sposobem oszacowania wartości pierwszych momentów statystycznych odpowiedzi konstrukcji jest rozwinięcie analizowanych funkcji w szereg Taylora wokół wartości oczekiwanych zmiennych losowych. W rozwinięciu pomija się zazwyczaj wyrazy rzędów wyższych niż drugi, a opis probabilistyczny zmiennych ograniczany jest jedynie do ich wartości oczekiwanych oraz macierzy kowariancji. Podejście takie nazywane jest także stochastyczną metodą perturbacyjną. W odróżnieniu od metod symulacyjnych, które sprowadzają się do obliczania wartości funkcji dla wielu realizacji zmiennych losowych wygenerowanych z odpowiednich rozkładów prawdopodobieństwa, istotą metody perturbacyjnej jest analiza wrażliwości, a więc obliczanie gradientów i pochodnych wyższych rzędów funkcji odpowiedzi. Poprawę dokładności oszacowania momentów statystycznych uzyskać można poprzez uwzględnienie większej liczby wyrazów szeregu, co wiąże się jednak z szybkim wzrostem kosztu obliczeń. Poniżej zaprezentowano jedynie wyprowadzenie wzorów metody pierwszego i drugiego rzędu.

4.2.1. Metoda pierwszego rzędu

Pomijając wyrazy wyższego rzędu niż pierwszy, rozwinięcie funkcji zmiennych losowych w szereg Taylora wokół wartości oczekiwanych tych zmiennych ma postać

$$f(\mathbf{X}) \approx \tilde{f}(\mathbf{X}) = f(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^n \left. \frac{\partial f(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} (X_i - \mu_{X_i}), \quad (4.42)$$

gdzie n jest rozmiarem wektora \mathbf{X} , a $\boldsymbol{\mu}_{\mathbf{X}}$ jest wektorem wartości oczekiwanych zmiennych losowych. Obliczając następnie wartość oczekiwaną oraz wariancję funkcji \tilde{f} dostajemy

$$\mathbb{E}[f(\mathbf{X})] \approx \mathbb{E}[\tilde{f}(\mathbf{X})] = f(\boldsymbol{\mu}_{\mathbf{X}}), \quad (4.43)$$

$$\text{Var}[f(\mathbf{X})] \approx \text{Var}[\tilde{f}(\mathbf{X})] = \nabla f^T(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \mathbf{C}_{\mathbf{X}} \nabla f(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}}, \quad (4.44)$$

gdzie $\nabla f(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}}$ jest gradientem funkcji f obliczonym w punkcie $\boldsymbol{\mu}_{\mathbf{X}}$, a $\mathbf{C}_{\mathbf{X}}$ jest macierzą kowariancji zmiennych \mathbf{X} , którą zapisać można następująco:

$$\mathbf{C}_{\mathbf{X}} = \begin{bmatrix} \sigma_{X_1}^2 & \rho_{12} \sigma_{X_1} \sigma_{X_2} & \cdots & \rho_{1n} \sigma_{X_1} \sigma_{X_n} \\ \rho_{21} \sigma_{X_2} \sigma_{X_1} & \sigma_{X_2}^2 & \cdots & \rho_{2n} \sigma_{X_2} \sigma_{X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} \sigma_{X_n} \sigma_{X_1} & \rho_{n2} \sigma_{X_n} \sigma_{X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix}, \quad (4.45)$$

$\rho_{ij} = \rho(X_i, X_j)$, $i, j = 1, \dots, n$, są współczynnikami korelacji. W szczególnym przypadku nieskorelowanych zmiennych losowych wyrażenie (4.44) przyjmuje postać

$$\text{Var}[f(\mathbf{X})] \approx \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \sigma_{X_i} \right)^2. \quad (4.46)$$

Ponieważ we wzorach (4.44) i (4.46) występuje gradient funkcji f , dlatego (korzystając z różnicowego przybliżenia pochodnej) metoda pierwszego rzędu wymaga co najmniej $n + 1$ krotnego obliczenia wartości funkcji. Podejście to jest zatem dużo bardziej efektywne od metod symulacji losowych, prowadzić jednak może do istotnych błędów w przypadku gdy w obszarze określoności zmiennych losowych rozpatrywana funkcja nie jest liniowa lub gdy punkt wartości średnich jest punktem stacjonarnym funkcji f .

4.2.2. Metoda drugiego rzędu

Zachowując w rozwinięciu Taylora również człony drugiego rzędu otrzymuje się następujące przybliżenie funkcji $f(\mathbf{X})$

$$\begin{aligned} f(\mathbf{X}) \approx \hat{f}(\mathbf{X}) &= f(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} (X_i - \mu_{X_i}) \\ &+ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} (X_i - \mu_{X_i})(X_j - \mu_{X_j}). \end{aligned} \quad (4.47)$$

Wykonując na wyrażeniu (4.47) operację wartości oczekiwanej dostajemy

$$\mathbb{E}[f(\mathbf{X})] \approx \mathbb{E}[\hat{f}(\mathbf{X})] = f(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \rho_{ij} \sigma_{X_i} \sigma_{X_j}, \quad (4.48)$$

natomiast wzór na wariancję $f(\mathbf{X})$ ma postać

$$\begin{aligned} \text{Var}[f(\mathbf{X})] \approx \text{Var}[\hat{f}(\mathbf{X})] &= \nabla f^T(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \mathbf{C}_{\mathbf{X}} \nabla f(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \frac{\partial f(\mathbf{x})}{\partial x_j} \Big|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \rho_{ij} \sigma_{X_i} \sigma_{X_j}. \end{aligned} \quad (4.49)$$

Można zauważyć, że wzór (4.49) jest taki sam jak (4.44), i że w istocie w metodzie drugiego rzędu używa się aproksymacji drugiego rzędu wartości oczekiwanej, ale wariancja funkcji f aproksymowana jest tak jak w metodzie rzędu

pierwszego. Jest to tylko pozorna niekonsekwencja gdyż, jak można pokazać (por. [123]), pełne wyrażenie na $\text{Var}[f(\mathbf{X})]$ wymaga znajomości momentów centralnych zmiennych losowych rzędu trzeciego i czwartego, które nie są dane.

Ponieważ w celu wyznaczenia wartości oczekiwanej (4.48) konieczna jest znajomość drugich pochodnych $f(\mathbf{x})$, dlatego minimalny koszt omawianej metody to $N = (n+1)(n+2)/2$ obliczenia wartości funkcji (wywołania programu analizy konstrukcji). Co za tym idzie, metoda drugiego rzędu jest znacznie droższa od metody pierwszego rzędu i o ile dla małych n może być konkurencyjna w stosunku do symulacji losowych, o tyle w przypadku wielu zmiennych losowych staje się mało efektywna. Na przykład, dla 100 zmiennych rozwinięcie Taylora drugiego rzędu wymaga obliczenia wartości funkcji f w 5151 punktach, podczas gdy otrzymanie podobnego oszacowania metodą symulacyjną z użyciem łacińskiej hiperkostki wymaga zazwyczaj znacznie mniejszej liczby eksperymentów numerycznych. Pewną poprawę efektywności obliczeń (kosztem dokładności oszacowania) osiągnąć można pomijając we wzorze (4.48) pochodne mieszane i pozostawiając jedynie wyrazy diagonalne macierzy Hessianu. Liczba punktów, w których obliczyć należy wartość funkcji f redukuje się wtedy do $2n + 1$.

Podsumowując, metodę drugiego rzędu rekomendować można w przypadku gdy krzywizny rozpatrywanej funkcji są znaczne oraz gdy liczba zmiennych losowych nie jest zbyt duża. Z drugiej strony, należy pamiętać, że o rzeczywistej przydatności metod rozwinięcia w szereg Taylora, np. w optymalizacji odpornościowej, decyduje dostępność pochodnych funkcji odpowiedzi układu konstrukcyjnego względem zmiennych losowych. Jeśli funkcja opisująca pracę konstrukcji jest dana w sposób jawny, bądź gdy w programie metody elementów skończonych zaimplementowano efektywne analityczne metody analizy wrażliwości, zob. [115], to stochastyczne metody perturbacyjne mogą okazać się szczególnie korzystnym rozwiązaniem.

W przypadkach komputerowej symulacji silnie nieliniowych zagadnień mechaniki, takich jak zderzenia pojazdów lub głębokie tłoczenie blach, gdy nie dysponujemy analityczną wrażliwością, a numeryczny szum utrudnia dobranie wielkości perturbacji w schemacie różnic skończonych, metody rozwinięcia w szereg nie są alternatywą dla metod typu descriptive sampling. Poza tym, ze względu na przybliżony charakter aproksymacji funkcji obcięty szereg Taylor, nie należy używać metod perturbacyjnych, jeżeli w probabilistycznym modelu zadania występują zmienne losowe charakteryzujące się dużym rozrzutem (współczynnik zmienności większy niż 10%).

4.3. Metoda redukcji wymiarów

Metoda redukcji wymiarów (ang. *dimension reduction method*) zaproponowana w 2004 roku przez Rahmana i Xu [201, 259] jest nową techniką szacowania momentów statystycznych funkcji losowych, która - podobnie jak metoda zaprezentowana w poprzednim podrozdziale - wykorzystuje rozwinięcie funkcji w szereg Taylora wokół wartości średnich. Pomimo oczywistych ograniczeń podejścia perturbacyjnego wynikających z konieczności stosowania skończonej liczby wyrazów rozwinięcia, metoda redukcji wymiarów wnosi istotne usprawnienia zarówno jeśli chodzi o dokładność jak i efektywność obliczania momentów. W ostatnich latach ukazało się szereg prac poświęconych porównaniu metody redukcji wymiarów z innymi metodami [97, 140], a także ocenie jej przydatności do rozwiązywania zagadnień optymalizacji odpornościowej oraz analizy niezawodności konstrukcji [32, 137, 141, 142].

Ideę metody redukcji wymiarów wygodnie jest przedstawić za pracą [259] na przykładzie redukcji dwuwymiarowej, czyli redukcji funkcji n -wymiarowej do sumy funkcji maksymalnie dwuwymiarowych.

Rozwinięcie w szereg Taylora funkcji $f(\mathbf{x})$ wokół punktu $\mathbf{x} = \mathbf{0}$ ma postać

$$\begin{aligned}
 f(\mathbf{x}) = & f(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} \sum_{i=1}^n \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \Big|_{\mathbf{x}=\mathbf{0}} x_i^j \\
 & + \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2!} \sum_{i_1 < i_2} \frac{\partial^{j_1+j_2} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_1}^{j_1} x_{i_2}^{j_2} \\
 & + \sum_{j_3=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2! j_3!} \sum_{i_1 < i_2 < i_3} \frac{\partial^{j_1+j_2+j_3} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2} \partial x_{i_3}^{j_3}} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_1}^{j_1} x_{i_2}^{j_2} x_{i_3}^{j_3} + \dots
 \end{aligned} \tag{4.50}$$

Oznaczając przez

$$I[f(\mathbf{x})] \equiv \underbrace{\int_{-a}^a \dots \int_{-a}^a}_{n \text{ razy}} f(\mathbf{x}) d\mathbf{x}, \tag{4.51}$$

całkę funkcji $f(\mathbf{x})$ po symetrycznym obszarze $[-a, a]^n$, ze wzoru (4.50) dostajemy

$$\begin{aligned}
I[f(\mathbf{x})] &= I[f(\mathbf{0})] + \sum_{j=1}^{\infty} \frac{1}{j!} \sum_{i=1}^n \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \Big|_{\mathbf{x}=\mathbf{0}} I[x_i^j] \\
&+ \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2!} \sum_{i_1 < i_2} \frac{\partial^{j_1+j_2} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} \Big|_{\mathbf{x}=\mathbf{0}} I[x_{i_1}^{j_1} x_{i_2}^{j_2}] \\
&+ \sum_{j_3=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2! j_3!} \sum_{i_1 < i_2 < i_3} \frac{\partial^{j_1+j_2+j_3} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2} \partial x_{i_3}^{j_3}} \Big|_{\mathbf{x}=\mathbf{0}} I[x_{i_1}^{j_1} x_{i_2}^{j_2} x_{i_3}^{j_3}] + \dots
\end{aligned} \tag{4.52}$$

Zapiszmy teraz rozwinięcia Taylora funkcji jednej zmiennej $f(0, \dots, 0, x_i, 0, \dots, 0)$ w $x_i = 0$ i dwuwymiarowej funkcji $f(0, \dots, 0, x_{i_1}, 0, \dots, 0, x_{i_2}, 0, \dots, 0)$ w punkcie $x_{i_1} = x_{i_2} = 0$,

$$f(0, \dots, 0, x_i, 0, \dots, 0) = f(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \Big|_{\mathbf{x}=\mathbf{0}} x_i^j, \tag{4.53}$$

$$\begin{aligned}
f(0, \dots, 0, x_{i_1}, 0, \dots, 0, x_{i_2}, 0, \dots, 0) &= f(\mathbf{0}) + \sum_{j=1}^{\infty} \frac{1}{j!} \frac{\partial^j f(\mathbf{x})}{\partial x_{i_1}^j} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_1}^j \\
&+ \sum_{j=1}^{\infty} \frac{1}{j!} \frac{\partial^j f(\mathbf{x})}{\partial x_{i_2}^j} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_2}^j + \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2!} \frac{\partial^{j_1+j_2} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_1}^{j_1} x_{i_2}^{j_2}.
\end{aligned} \tag{4.54}$$

Sumując powyższe rozwinięcia odpowiednio względem indeksów i oraz par i_1 i i_2 , tak że $i_1 < i_2$, otrzymuje się

$$\sum_{i=1}^n f(0, \dots, 0, x_i, 0, \dots, 0) = n f(\mathbf{0}) + \sum_{i=1}^n \sum_{j=1}^{\infty} \frac{1}{j!} \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \Big|_{\mathbf{x}=\mathbf{0}} x_i^j, \tag{4.55}$$

$$\begin{aligned}
\sum_{i_1 < i_2} f(0, \dots, 0, x_{i_1}, 0, \dots, 0, x_{i_2}, 0, \dots, 0) &= \frac{n(n-1)}{2} f(\mathbf{0}) \\
&+ (n-1) \sum_{i=1}^n \sum_{j=1}^{\infty} \frac{1}{j!} \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \Big|_{\mathbf{x}=\mathbf{0}} x_i^j \\
&+ \sum_{i_1 < i_2} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2!} \frac{\partial^{j_1+j_2} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} \Big|_{\mathbf{x}=\mathbf{0}} x_{i_1}^{j_1} x_{i_2}^{j_2}.
\end{aligned} \tag{4.56}$$

Można zatem wprowadzić następujące dwuwymiarowe przybliżenie funkcji f

$$\begin{aligned} \widehat{f}^{(2)}(\mathbf{x}) \equiv & \sum_{i_1 < i_2} f(0, \dots, 0, x_{i_1}, 0, \dots, 0, x_{i_2}, 0, \dots, 0) \\ & - (n-2) \sum_{i=1}^n f(0, \dots, 0, x_i, 0, \dots, 0) + \binom{n-1}{2} f(\mathbf{0}), \end{aligned} \quad (4.57)$$

które składa się z wyrazów będących funkcjami najwyżej dwóch zmiennych. Po scałkowaniu w granicach $[-a, a]^n$ powyższe wyrażenie przyjmuje postać

$$\begin{aligned} I[\widehat{f}^{(2)}(\mathbf{x})] = & \sum_{i_1 < i_2} I[f(0, \dots, 0, x_{i_1}, 0, \dots, 0, x_{i_2}, 0, \dots, 0)] \\ & - (n-2) \sum_{i=1}^n I[f(0, \dots, 0, x_i, 0, \dots, 0)] + \binom{n-1}{2} I[f(\mathbf{0})], \end{aligned} \quad (4.58)$$

a wykorzystując wzory (4.53) i (4.54) dostajemy

$$\begin{aligned} I[\widehat{f}^{(2)}(\mathbf{x})] = & I[f(\mathbf{0})] + \sum_{i=1}^n \sum_{j=1}^{\infty} \frac{1}{j!} \left. \frac{\partial^j f(\mathbf{x})}{\partial x_i^j} \right|_{\mathbf{x}=\mathbf{0}} I[x_i^j] \\ & + \sum_{i_1 < i_2} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2!} \left. \frac{\partial^{j_1+j_2} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2}} \right|_{\mathbf{x}=\mathbf{0}} I[x_{i_1}^{j_1} x_{i_2}^{j_2}]. \end{aligned} \quad (4.59)$$

Porównując (4.59) z całką (4.52) można zauważyć, że

$$\begin{aligned} I[f(\mathbf{x})] - I[\widehat{f}^{(2)}(\mathbf{x})] = & \sum_{j_3=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \frac{1}{j_1! j_2! j_3!} \sum_{i_1 < i_2 < i_3} \left. \frac{\partial^{j_1+j_2+j_3} f(\mathbf{x})}{\partial x_{i_1}^{j_1} \partial x_{i_2}^{j_2} \partial x_{i_3}^{j_3}} \right|_{\mathbf{x}=\mathbf{0}} I[x_{i_1}^{j_1} x_{i_2}^{j_2} x_{i_3}^{j_3}] + \dots, \end{aligned} \quad (4.60)$$

a zatem błąd popełniany przy aproksymowaniu całki funkcji $f(\mathbf{x})$ całką funkcji $\widehat{f}^{(2)}(\mathbf{x})$ zawiera człony trój- i więcej wymiarowe. Obliczając wartość $I[\widehat{f}^{(2)}(\mathbf{x})]$ konieczne jest wykonanie całkowania jedynie funkcji jednej i dwóch zmiennych, unikając n -wymiarowego całkowania, zob. (4.51), praktycznie niemożliwego do przeprowadzenia dla dużych n . Nie jest również konieczne obliczanie pochodnych cząstkowych, co stanowi niezaprzeczalny atut tej metody.

Założenie symetrycznych przedziałów całkowania $[-a, a]^n$ oraz wybór $\mathbf{x} = \mathbf{0}$ jako punktu rozwinięcia pozwala na uproszczenie wzorów w powyższym wyprowadzeniu. Jednak, w celu obliczenia całki po obszarze niesymetrycznym $\prod_{i=1}^n [a_i, b_i]$

$$I[f(\mathbf{x})] = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \cdots dx_n, \quad (4.61)$$

gdzie $-\infty \leq a_i \leq \infty$ i $-\infty \leq b_i \leq \infty$, dzięki transformacji liniowej

$$x_i = \frac{b_i + a_i}{2} + \frac{b_i - a_i}{2} \xi_i, \quad i = 1, \dots, n, \quad (4.62)$$

całkę (4.61) można zapisać w równoważnej postaci całki po obszarze symetrycznym

$$I[f(\mathbf{x})] = \prod_{i=1}^n \frac{b_i - a_i}{2} \int_{-1}^1 \cdots \int_{-1}^1 \eta(\xi_1, \dots, \xi_n) d\xi_1 \cdots d\xi_n, \quad (4.63)$$

gdzie $\eta(\xi_1, \dots, \xi_n)$ jest funkcją podcałkową po transformacji zmiennych \mathbf{x} na $\boldsymbol{\xi}$. Wzór (4.60) dotyczy zatem całek po obszarze symetrycznym, jak i niesymetrycznym.

Ponieważ, celem metody redukcji wymiarów jest zastosowanie aproksymacji (4.58) do szacowania wartości całek opisujących momenty funkcji losowych $f(\mathbf{X})$, dlatego rozwinięcia w szereg Taylora dokonuje się w punkcie wartości oczekiwanych zmiennych losowych $\boldsymbol{\mu}_{\mathbf{X}} = \{\mu_1, \dots, \mu_n\}$, który w ogólności jest różny od $\mathbf{0}$. Prowadzi to do następującej postaci funkcji (4.57)

$$\begin{aligned} \widehat{f}^{(2)}(\mathbf{X}) &\equiv \sum_{i_1 < i_2} f(\mu_1, \dots, \mu_{i_1-1}, X_{i_1}, \mu_{i_1+1}, \dots, \mu_{i_2-1}, X_{i_2}, \mu_{i_2+1}, \dots, \mu_n) \\ &- (n-2) \sum_{i=1}^n f(\mu_1, \dots, \mu_{i-1}, X_i, \mu_{i+1}, \dots, \mu_n) + \binom{n-1}{2} f(\mu_1, \dots, \mu_n). \end{aligned} \quad (4.64)$$

W pracy [259] wyprowadzono również ogólny wzór na redukcję funkcji n -wymiarowej do sumy funkcji maksymalnie s -wymiarowych, gdzie $s \leq n$

$$f(\mathbf{x}) \approx \widehat{f}^{(s)}(\mathbf{x}) \equiv \sum_{i=0}^s (-1)^i \binom{n-s+i-1}{i} f_{s-i}(\mathbf{x}), \quad (4.65)$$

gdzie

$$f_r(\mathbf{x}) = \sum_{k_1 < k_2 < \dots < k_r} f(0, \dots, 0, x_{k_1}, 0, \dots, 0, x_{k_2}, 0, \dots, 0, x_{k_r}, 0, \dots, 0), \quad 0 \leq r \leq n. \quad (4.66)$$

Całkę tak aproksymowanej funkcji zapisać można zatem jako

$$I[\widehat{f}^{(s)}(\mathbf{x})] = \sum_{i=0}^s (-1)^i \binom{n-s+i-1}{i} I[f_{s-i}(\mathbf{x})]. \quad (4.67)$$

Obliczenie wartości powyższego wyrażenia wymaga obliczenia jedynie całek 1-, 2-, ... oraz s -wymiarowych. W szczególności, w przypadku redukcji jednowymiarowej, najmniej dokładnej ale najtańszej z możliwych redukcji, otrzymujemy wzór (na podstawie rozwinięcia funkcji losowej w wartościach oczekiwanych $\boldsymbol{\mu}_{\mathbf{X}}$)

$$I[\widehat{f}^{(1)}(\mathbf{X})] = \sum_{i=1}^n I[f(\mu_1, \dots, \mu_{i-1}, X_i, \mu_{i+1}, \dots, \mu_n)] - (n-1)f(\mu_1, \dots, \mu_n). \quad (4.68)$$

Jeśli \mathbf{X} jest n -elementowym wektorem zmiennych losowych o łącznej gęstości rozkładu prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$, to zakładając, że zmienne losowe są niezależne, funkcja gęstości zapisana może być w postaci iloczynu gęstości jednowymiarowych $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i)$. Moment zwykły l -tego rzędu funkcji f dany jest wzorem

$$m_l = \mathbb{E}[f^l(\mathbf{X})] = \int_{\mathbb{R}^n} f^l(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}. \quad (4.69)$$

Wprowadzając następnie oznaczenie $z(\mathbf{X}) = f^l(\mathbf{X})$, równanie (4.69) wyrazić można w równorzędnej postaci jako

$$m_l = \mathbb{E}[z(\mathbf{X})] = \int_{\mathbb{R}^n} z(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}. \quad (4.70)$$

Zgodnie z definicjami (4.65)–(4.67), wartość momentu l -tego rzędu oszacować można następująco:

$$m_l \approx \mathbb{E}[\widehat{z}^{(s)}(\mathbf{X})] = \sum_{i=0}^s (-1)^i \binom{n-s+i-1}{i} \mathbb{E}[z_{s-i}(\mathbf{X})], \quad (4.71)$$

gdzie

$$\begin{aligned} & \mathbb{E}[z_{s-i}(\mathbf{X})] \\ &= \sum_{k_1 < k_2 < \dots < k_{s-i}} \mathbb{E}[z(\mu_1, \dots, \mu_{k_1-1}, X_{k_1}, \mu_{k_1+1}, \dots, \mu_{k_2-1}, X_{k_2}, \mu_{k_2+1}, \dots \\ & \quad \dots, \mu_{k_{s-i}-1}, X_{k_{s-i}}, \mu_{k_{s-i}+1}, \dots, \mu_n)] \\ &= \sum_{k_1 < k_2 < \dots < k_{s-i}} \int_{\mathbb{R}^{s-i}} z(\mu_1, \dots, \mu_{k_1-1}, x_{k_1}, \mu_{k_1+1}, \dots, \mu_{k_2-1}, x_{k_2}, \mu_{k_2+1}, \dots \\ & \quad \dots, \mu_{k_{s-i}-1}, x_{k_{s-i}}, \mu_{k_{s-i}+1}, \dots, \mu_n) f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) \, d\tilde{\mathbf{x}}, \quad (4.72) \end{aligned}$$

gdzie $\tilde{\mathbf{x}} = \{x_{k_1}, \dots, x_{k_{s-i}}\} \in \mathbb{R}^{s-i}$ jest realizacją losowego wektora $\tilde{\mathbf{X}} = \{X_{k_1}, \dots, X_{k_{s-i}}\}$ o funkcji łącznej gęstości prawdopodobieństwa

$$f_{\tilde{\mathbf{X}}}(\tilde{\mathbf{x}}) = \prod_{j=1}^{s-i} f_{X_{k_j}}(x_{k_j}). \quad (4.73)$$

W przypadku zmiennych zależnych konieczne jest uprzednie dokonanie ich transformacji do przestrzeni zmiennych losowych niezależnych. Używając transformacji Rosenblatta lub Natafa (zob. podrozdział A.2) zmienne transformowane są do gaussowskiej przestrzeni standardowej.

Ponieważ najczęściej stosuje się redukcję jedno i dwuwymiarową, dlatego warto podać wzór (4.71) w tych dwóch, szczególnych przypadkach

$$m_l \approx \mathbb{E}[\hat{z}^{(1)}(\mathbf{X})] = \sum_{i=1}^n \int_{-\infty}^{\infty} z(\mu_1, \dots, \mu_{i-1}, x_i, \mu_{i+1}, \dots, \mu_n) f_{X_i}(x_i) dx_i - (n-1)z(\mu_1, \dots, \mu_n), \quad (4.74)$$

$$\begin{aligned} m_l \approx \mathbb{E}[\hat{z}^{(2)}(\mathbf{X})] = & \sum_{i_1 < i_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ z(\mu_1, \dots, \mu_{i_1-1}, x_{i_1}, \mu_{i_1+1}, \dots, \mu_{i_2-1}, x_{i_2}, \mu_{i_2+1}, \dots, \mu_n) \right. \\ & \left. \times f_{X_{i_1}}(x_{i_1}) f_{X_{i_2}}(x_{i_2}) \right\} dx_{i_1} dx_{i_2} \\ & - (n-2) \sum_{i=1}^n \int_{-\infty}^{\infty} z(\mu_1, \dots, \mu_{i-1}, x_i, \mu_{i+1}, \dots, \mu_n) f_{X_i}(x_i) dx_i \\ & + \frac{(n-1)(n-2)}{2} z(\mu_1, \dots, \mu_n). \end{aligned} \quad (4.75)$$

Wartość średnia oraz wariancja funkcji losowej wyrażają się odpowiednio jako

$$\mathbb{E}[f(\mathbf{X})] = m_1, \quad \text{Var}[f(\mathbf{X})] = m_2 - m_1^2. \quad (4.76)$$

W celu oszacowania wartości momentu l -tego rzędu zgodnie ze wzorem (4.71) konieczne jest numeryczne obliczenie całek co najwyżej s -wymiarowych, a w przypadku wzorów (4.74) i (4.75) odpowiednio jedno i dwuwymiarowych. Jeśli zmienne losowe mają rozkład normalny, wtedy do całkowania zastosować można kwadratury Gaussa-Hermite'a (całkowanie z wagą e^{-x^2} na przedziale

nieograniczonym), a w przypadku rozkładów jednostajnych kwadratury Gaussa-Legendre'a (waga 1, przedział obustronnie domknięty), zob. [197]. Jeśli mamy do czynienia z innymi rozkładami prawdopodobieństwa, wtedy najbardziej efektywnym rozwiązaniem jest skonstruowanie "dedykowanej" kwadratury Gaussa, gdzie wagi i węzły kwadratury wyliczane są przyjmując funkcję gęstości prawdopodobieństwa jako funkcję wagową. Postępowanie to zostało szczegółowo omówione m.in. w pracach [140, 201, 259].

W porównaniu z nakładem obliczeń koniecznym do przeprowadzenia "pełnego" całkowania numerycznego funkcji n -wymiarowej, szczególnie dla dużych n , metody redukcji jedno i dwuwymiarowej prowadzą do znaczącej poprawy efektywności obliczeń. Przyjmując dla każdego kierunku stałą liczbę węzłów kwadratury równą r , do oszacowania jednowymiarowego potrzeba $nr + 1$, a dwuwymiarowego $(n(n-1)/2)r^2 + nr + 1$ obliczeń wartości funkcji, co wynika bezpośrednio ze wzorów (4.74) i (4.75). Jeśli na przykład, $n = 10$ i $r = 4$, to w porównaniu do całkowania n -wymiarowego, metody redukcji jedno i dwuwymiarowej wymagają odpowiednio 25575 i 1377 razy mniej obliczeń wartości funkcji.

W przeciwieństwie do omówionych w podrozdziale 4.2 metod perturbacyjnych, metoda redukcji wymiarów nie wymaga obliczania pochodnych cząstkowych, co w wielu przypadkach jest bardzo trudne w realizacji. Należy też pamiętać, iż metoda redukcji jednowymiarowej nie jest ekwiwalentem metody perturbacyjnej pierwszego rzędu, a metoda redukcji dwuwymiarowej metody perturbacyjnej drugiego rzędu. Rozwinięcie (4.57) zawiera wszystkie człony mieszane dwóch zmiennych dowolnego rzędu, a więc jest to reprezentacja dużo pełniejsza niż ta dana wzorem (4.47).

Na koniec wspomnieć należy o wpływie transformacji zmiennych na dokładność szacowanych momentów statystycznych funkcji losowej. O ile w przypadku zmiennych skorelowanych nie ma alternatywy dla transformacji, a następnie używania kwadratur Gaussa-Hermite'a, o tyle dla zmiennych niezależnych o dowolnych rozkładach prawdopodobieństwa, jeśli to możliwe powinno się unikać transformacji do przestrzeni standardowej. Korzyść wynikająca z możliwości zastosowania znanych kwadratur Gaussa-Hermite'a dla zmiennych o rozkładzie normalnym, bez konieczności znajdowania kwadratur dopasowanych do oryginalnych rozkładów, okupiona jest w przypadku nieliniowych funkcji losowych słabszą jakością oszacowania, szczególnie momentów wyższych rzędów, zob. [141].

4.4. Metoda rozwinięcia funkcji losowej w chaos wielomianowy

Jeśli zmienna $Y = f(\mathbf{X})$ ma skończoną wariancję, to można ją rozwinąć w tzw. chaos wielomianowy (ang. *polynomial chaos* – PC), zob. [73, 244], w następujący sposób

$$Y = f(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^n} a_{\alpha} \psi_{\alpha}(\mathbf{X}), \quad (4.77)$$

gdzie a_{α} są nieznanymi deterministycznymi współczynnikami, a ψ_{α} są wielowymiarowymi wielomianami, ortogonalnymi względem funkcji łącznej gęstości rozkładu prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$, tzn.:

$$\mathbb{E}[\psi_{\alpha}(\mathbf{X})\psi_{\beta}(\mathbf{X})] = \delta_{\alpha,\beta}, \quad (4.78)$$

gdzie $\delta_{\alpha,\beta} = 1$ jeśli $\alpha = \beta$, w przeciwnym wypadku $\delta_{\alpha,\beta} = 0$.

Załóżmy następnie, że zmienne losowe X_i , $i = 1, \dots, n$, są niezależne (w przypadku zmiennych skorelowanych zastosować najpierw należy transformację Natafa - podrozdział A.2.4). Wtedy funkcję gęstości wektora \mathbf{X} przedstawić można jako iloczyn gęstości jednowymiarowych

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n), \quad (4.79)$$

a każdy wielomian ψ_{α} skonstruowany być może jako iloczyn n jednowymiarowych wielomianów ortogonalnych

$$\psi_{\alpha}(\mathbf{X}) = P_{\alpha_1}^{(1)}(X_1)P_{\alpha_2}^{(2)}(X_2) \cdots P_{\alpha_n}^{(n)}(X_n). \quad (4.80)$$

Oznaczając przez \mathcal{D}_{X_i} przedział określoności zmiennej X_i , wielomiany ortogonalne $\{P_k^{(i)}, k \geq 0\}$, które spełniają warunek

$$\mathbb{E}[P_k^{(i)}(X_i)P_l^{(i)}(X_i)] = \int_{\mathcal{D}_{X_i}} P_k^{(i)}(x)P_l^{(i)}(x)f_{X_i}(x)dx = \delta_{k,l}, \quad \forall (k, l) \in \mathbb{N}^2, \quad (4.81)$$

wyznaczyć można przy użyciu klasycznych algorytmów, zob. [197].

W najistotniejszym z punktu widzenia zastosowań przypadku, gdy zmienne losowe mają standardowy rozkład normalny, wielomianami, które są ortogonalne

względem tej funkcji gęstości są wielomiany Hermite'a $H_k(\mathbf{x})$, $k \geq 0$

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ &\vdots \\ H_{k+1}(x) &= xH_k(x) - kH_{k-1}(x). \end{aligned} \quad (4.82)$$

Ze względów obliczeniowych szereg w równaniu (4.77) obcina się zatrzymując jedynie skończoną liczbę wyrazów. Najczęściej stosowaną metodą jest pominięcie w rozwinięciu tych wielomianów, których stopień $|\boldsymbol{\alpha}|$ przekracza założony stopień p :

$$Y = f(\mathbf{X}) \approx \sum_{|\boldsymbol{\alpha}| \leq p} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{X}), \quad (4.83)$$

gdzie $|\boldsymbol{\alpha}| = \sum_{i=1}^n \alpha_i$. Dla danego p , liczba współczynników $a_{\boldsymbol{\alpha}}$, które należy wyznaczyć wynosi

$$M = \binom{n+p}{p}. \quad (4.84)$$

Przyjmuje się (zob. [15]), że przybliżenie PC przy $p = 2$ jest zazwyczaj wystarczająco dokładne do obliczenia momentów funkcji losowej.

Do wyznaczenia nieznanych współczynników we wzorze (4.83) stosuje się metodę rzutowania bądź metodę liniowej regresji.

Metoda rzutowania wykorzystuje ortogonalność funkcji bazowych. I tak, mnożąc obustronnie równanie (4.77) przez $\psi_{\boldsymbol{\alpha}}(\mathbf{X})$, a następnie wykonując operację wartości oczekiwanej, współczynniki $a_{\boldsymbol{\alpha}}$ wyrażą się następującym wzorem:

$$a_{\boldsymbol{\alpha}} = \mathbb{E}[f(\mathbf{X})\psi_{\boldsymbol{\alpha}}(\mathbf{X})], \quad (4.85)$$

lub inaczej, z definicji wartości oczekiwanej,

$$a_{\boldsymbol{\alpha}} = \int_{\mathcal{D}_{\mathbf{X}}} f(\mathbf{x})\psi_{\boldsymbol{\alpha}}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \quad (4.86)$$

gdzie $\mathcal{D}_{\mathbf{X}}$ oznacza obszar określoności zmiennych \mathbf{X} . Ponieważ oszacowanie wartości wielowymiarowej całki w równaniu (4.86), czy to przy pomocy symulacji losowych, czy też wzorów kwadraturowych, prowadzi zazwyczaj do olbrzymich kosztów obliczeń numerycznych, dlatego metoda rzutowania nie jest często stosowana.

Metoda regresji liniowej. Jak to zostało omówione w podrozdziale 3.1, metoda regresji polega na odpowiednim dopasowaniu założonej a priori powierzchni odpowiedzi (tutaj, obciętego rozwinięcia PC) do rzeczywistej funkcji, której wartości znane są jedynie dla pewnej próbki punktów eksperymentalnych. Równanie (4.83) zapisać można w równoważnej postaci macierzowej

$$f(\mathbf{X}) \approx \mathcal{F}_p(\mathbf{X}) = \sum_{|\alpha| \leq p} a_\alpha \psi_\alpha(\mathbf{X}) = \mathbf{a}^T \boldsymbol{\psi}(\mathbf{X}), \quad (4.87)$$

gdzie \mathbf{a} jest wektorem współczynników $\{a_\alpha, 0 \leq |\alpha| \leq p\}$, a $\boldsymbol{\psi}$ jest wektorem wielomianów bazowych $\{\psi_\alpha, 0 \leq |\alpha| \leq p\}$.

Na podstawie przeprowadzonych N eksperymentów numerycznych $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, N$, gdzie $y_i = f(\mathbf{x}_i)$, zgodnie z metodologią tworzenia powierzchni odpowiedzi, współczynniki w równaniu (4.87) wyznaczone są przez minimalizację pewnej normy residuów $y_i - \mathcal{F}_p(\mathbf{x}_i)$, najczęściej sumy kwadratów reszt

$$S(\mathbf{a}) = \sum_{i=1}^N [f(\mathbf{x}_i) - \mathbf{a}^T \boldsymbol{\psi}(\mathbf{x}_i)]^2. \quad (4.88)$$

Wektor $\hat{\mathbf{a}}$ o wymiarze M , który minimalizuje powyższą sumę wyraża się wzorem, por. (3.7),

$$\hat{\mathbf{a}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{y}, \quad (4.89)$$

gdzie $\mathbf{y} = \{y_1, \dots, y_N\}$ jest wektorem wartości funkcji w punktach eksperymentalnych, a macierz $\boldsymbol{\Psi}_{N \times M}$ ma postać

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_{\alpha_1}(\mathbf{x}_1) & \cdots & \psi_{\alpha_M}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_{\alpha_1}(\mathbf{x}_N) & \cdots & \psi_{\alpha_M}(\mathbf{x}_N) \end{pmatrix}. \quad (4.90)$$

Do otrzymania rozwiązania konieczne jest aby macierz $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ była dobrze warunkowana. Aby to zapewnić, niezbędne jest użycie w zadaniu regresji odpowiednio licznych próbek, przy czym muszą one zawierać więcej punktów niż M . Na podstawie wzoru (4.84) łatwo sprawdzić, że M gwałtownie rośnie wraz z liczbą zmiennych losowych n jak i maksymalnym stopniem wielomianu p . I tak, dla $n = 4$ i $p = 3$, $M = 35$, dla $n = 6$ i $p = 3$, $M = 84$, a już dla $n = 6$ i $p = 5$, $M = 462$. W celu zmniejszenia złożoności numerycznej metody, Blatman i Sudret zaproponowali w [15] adaptacyjny algorytm budowy rzadkiego (ang. *sparse*) rozwinięcia PC, który w sposób iteracyjny znajduje optymalną

bazę wielomianów poprzez identyfikację członów rozwinięcia, które nie są istotne w aproksymacji funkcji $f(\mathbf{X})$. Jak pokazano, dzięki zastosowaniu tej metody uzyskać można wielokrotne zmniejszenie liczby współczynników \mathbf{a} .

Na podstawie założenia o niezależności zmiennych oraz ortogonalności wielomianów bazowych ψ_{α} , zob. równania (4.79)–(4.81), łatwo wykazać, że wartość średnia i wariancja funkcji $f(\mathbf{X})$ dane są wzorami

$$\mathbb{E}[f(\mathbf{X})] \approx \mathbb{E}[\mathcal{F}_p(\mathbf{X})] = a_{\mathbf{0}}, \quad (4.91)$$

$$\text{Var}[f(\mathbf{X})] \approx \text{Var}[\mathcal{F}_p(\mathbf{X})] = \sum_{0 < |\alpha| \leq p} a_{\alpha}^2. \quad (4.92)$$

PODSUMOWANIE

Zasadniczym elementem, który decyduje o efektywności algorytmu odpornościowej optymalizacji konstrukcji jest metoda obliczania wartości średnich oraz wariancji funkcji celu i ograniczeń. Niniejszy rozdział poświęcono w całości analizie metod estymacji momentów statystycznych funkcji losowych. Skupiono się na bardzo dokładnym omówieniu metod symulacyjnych, uznając je za najbardziej uniwersalne i wiarygodne. Pomimo, iż alternatywne podejścia, które omówiono pod koniec rozdziału, w wielu szczególnych przypadkach mogą okazać się bardziej efektywne od metod symulacyjnych, to jednak założenia, jakimi są obwarowane, ograniczają ich uniwersalność i wymagają każdorazowych wstępnych studiów nad ich przydatnością w danym zadaniu optymalizacji odpornościowej. Ograniczenia dotyczące rozkładów prawdopodobieństwa zmiennych, a także różniczkowalności funkcji, bardzo istotne w przypadku metod perturbacyjnych oraz metody redukcji wymiarów, nie mają znaczenia w kontekście symulacji losowych.

Przedstawiono ideę symulacyjnych metod *descriptive sampling* i porównano je z klasycznymi metodami Monte Carlo. Przeanalizowano właściwości estymatorów wartości średniej i wariancji w przypadku próbek czysto losowych oraz generowanych przez łańciskie hiperkostki. Ponieważ nie można przeprowadzić podobnych teoretycznych rozważań dla próbek generowanych przez optymalne hiperkostki łańciskie OLH, dlatego efektywność różnych metod symulacyjnych porównano na podstawie szeregu analitycznych funkcji testowych, a także dwóch problemów mechaniki. W znanych autorowi opracowaniach nie przeprowadzono nigdy tak szczegółowej analizy metod symulacyjnych wykorzystujących plany

OLH. Na podstawie testów stwierdzono, że z punktu widzenia efektywności symulacji losowych, próbki generowane przez hiperkostki łacińskie, są lepszym rozwiązaniem niż czysto losowe próbki w symulacjach Monte Carlo. Próbki utworzone przez OLH warto stosować szczególnie wtedy, gdy ze względu na koszt nie można pozwolić sobie na otrzymanie wartości rozważanej funkcji dla zbyt wielu realizacji zmiennych losowych. Wyniki testów przeprowadzanych dla zadań o różnej wielkości pozwalają również oszacować wielkość próbki na potrzeby zadania analizy optymalizacji odpornościowej.

Dokonano szczegółowej analizy znaczenia różnych sposobów wyboru punktu eksperymentalnego wewnątrz komórki łacińskiej hiperkostki. W przeciwieństwie do wyników prezentowanych w niektórych publikacjach, nie zaobserwowano istotnych jakościowo różnic pomiędzy różnymi metodami generacji realizacji zmiennej losowej w komórce. Wersja z wyborem punktu w medianach zmiennych na przedziale (LH-m) wydaje się jednak być najkorzystniejsza z punktu widzenia czasu tworzenia próbki.

Opisane w dodatku B algorytmy tworzenia optymalnych łacińskich hiperkostek zaimplementowane zostały przez autora w obiektowo-zorientowanym środowisku STAND, a wcześniej w module M-Xplore programu analizy skończonej elementowej RADIOSS.

Metody analizy niezawodności oraz metody poprawy niezawodności konstrukcji w przypadku zaszumionej funkcji granicznej

W zagadnieniach analizy niezawodności złożonych układów konstrukcyjnych funkcje graniczne, z którymi mamy do czynienia, są zazwyczaj niejawne i bardzo często silnie nieliniowe. W większości przypadków samo obliczenie ich wartości dla określonej realizacji zmiennych losowych wymaga przeprowadzenia analizy MES. W zależności od stopnia skomplikowania modelu oraz od rodzaju analizy mechanicznej, czas takiego obliczenia na pojedynczym procesorze wynosić może od kilku minut do kilku dni. W przypadku komputerowej symulacji zderzeń pojazdów czy też symulacji procesów tłoczenia blach, nawet przy użyciu maszyn wieloprocessorowych, analiza skończenie elementowa jest na tyle czasochłonna, że w praktyce można pozwolić sobie jedynie na ograniczoną liczbę takich symulacji. Jest zatem niezmiernie ważne, aby metody używane do analizy niezawodności tego typu zagadnień minimalizowały liczbę niezbędnych wywołań programu MES zapewniając jednocześnie wystarczającą dokładność oszacowania.

Nieliniowy charakter funkcji granicznych opisujących zachowanie konstrukcji wynika nie tylko z natury modelowanego zjawiska fizycznego, ale jest również wywołany szumem numerycznym spowodowanym przez stosowaną metodę rozwiązywania zadania MES. Metoda jawnego całkowania równań ruchu, algorytmy uwzględniania kontaktu lub też automatyczna adaptacja siatki w zadaniach o losowych parametrach opisujących geometrię konstrukcji, potencjalnie mogą być generatorami takiego szumu.

Kłopoty z różniczkowalnością funkcji granicznej na skutek występowania szumu numerycznego, czy też wysoka wrażliwość tej funkcji na perturbacje

wartości zmiennych, stanowią poważną przeszkodę w stosowaniu efektywnych gradientowych metod analizy niezawodności, zob. A.3. Trzeba pamiętać, że szum o którym mówimy w odniesieniu do symulacji komputerowych ma inny charakter niż ten, z którym mamy do czynienia wykonując eksperymenty fizyczne w laboratorium. W przypadku testów laboratoryjnych, powtarzając wielokrotnie identyczny eksperyment otrzymamy zawsze pewien rozrzut wyników, co związane jest z ograniczoną dokładnością przyrządów pomiarowych, jak i niemożnością utrzymania idealnie takich samych ustawień eksperymentu. Z drugiej strony, symulacje komputerowe powtarzane dla takiego samego modelu numerycznego na tym samym komputerze prowadzą zawsze do identycznych wyników. Pomimo to, można również wskazać problemy analizy niezawodności, gdzie wygodnie jest wprowadzić do modelu komputerowego losowy szum znany z doświadczeń fizycznych. Przykładem tego typu podejścia jest, zaprezentowana w podrozdziale 6.1, analiza niezawodności zderzanej belki cienkościennej, w której uwzględnia się losowe defekty połączeń zgrzewanych.

W dalszej części rozdziału przedstawione zostaną takie metody analizy niezawodności, które wydają się odpowiadać specyfice opisanych wyżej problemów. W podrozdziale 5.1 przedstawiono dwuetapową adaptacyjną metodę redukcji wariancji. Bazuje ona na algorytmie zaprezentowanym przez Zou i in. w pracy [278]. W podrozdziale 5.2 opisano metodę szybkiej poprawy poziomu niezawodności konstrukcji w oparciu o symulacje losowe OLH, zob. [231].

5.1. Dwuetapowa adaptacyjna metoda redukcji wariancji

Omawiana metoda jest udoskonaloną wersją algorytmu zaprezentowanego przez Zou, Mahadevana, Mourelatosa i Meernika m.in. w pracach [278, 279]. Od pierwszych liter nazwisk wymienionych autorów określać ją będziemy dalej akronimem Z3M.

Algorytm składa się z etapu lokalizacji punktu projektowego (zob. definicja na stronie 263) oraz etapu symulacji losowych, zwanych metodą redukcji wariancji (importance sampling, A.4.2). W pierwszym etapie, w celu przezwyciężenia trudności związanych z nieliniowym i zaszumionym charakterem funkcji granicznej, używa się adaptacyjnej aproksymacji tej funkcji, tworzonej na bazie eksperymentów OLH, zob. dodatek B. Wykorzystując znajomość położenia punktu projektowego, w trakcie symulacji losowych konstruowana jest wielomodalna funkcja gęstości losującej, która zapewnić ma wymaganą dokładność oszacowania prawdopodobieństwa awarii.

Metoda Z3M łączy zalety analizy niezawodności pierwszego rzędu (FORM), metodologii powierzchni odpowiedzi (metoda ważonej regresji) oraz symulacyjnej techniki importance sampling. Poniższa prezentacja, ukazując podstawowe składniki algorytmu, przedstawia również propozycje rozwiązań mających na celu poprawę jego efektywności. Pewne szczegółowe rozważania dotyczące doboru parametrów sterujących zawarto również przy okazji analizy przykładu w podrozdziale 6.1.

5.1.1. Etap pierwszy: Poszukiwanie punktu projektowego

Głównym zamysłem twórców algorytmu jest zapewnienie jego zbieżności nawet dla silnie nieliniowych funkcji granicznych. Punkt projektowy znajdujący się w wyniku iteracyjnej procedury lokalizacji jego przybliżonego położenia dla bieżącej aproksymacji funkcji granicznej w obrębie tzw. obszaru dopuszczalnego (ang. *trust region*). Wyznaczone przybliżenie punktu projektowego służy następnie do uaktualnienia zarówno położenia obszaru dopuszczalnego, jak i powierzchni odpowiedzi aproksymującej funkcję graniczną. Proces kontynuowany jest aż do spełnienia kryterium zbieżności. Po transformacji oryginalnych zmiennych losowych do przestrzeni standardowej (zob. podrozdział A.2) poszczególne kroki tego etapu metody Z3M zapisać można następująco:

1. *Definicja początkowego obszaru dopuszczalnego – określenie jego położenia i rozmiaru.*

Zazwyczaj jako początkowy obszar dopuszczalny przyjmuje się hiperkostkę o środku w początku układu współrzędnych i boku o wymiarze na tyle dużym, aby obszar ten zawierał realizacje zmiennych odpowiadające zarówno pracy bezpiecznej, jak i awarii. Oczywiście, takie założenie wymaga pewnej uprzedniej wiedzy na temat analizowanego problemu. Jeśli informacje tego typu nie są dostępne, to przyjęcie początkowego obszaru dopuszczalnego jako hiperkostki o boku równym 6 wydaje się być właściwym wyborem w większości praktycznych zadań analizy niezawodności. Bok o długości 6 w przestrzeni standardowej odpowiada najczęściej zakresom $[-3\sigma_{X_i}, 3\sigma_{X_i}]$ zmiennych losowych, gdzie σ_{X_i} , $i = 1, \dots, n$, jest odchyleniem standardowym i -tej zmiennej.

2. *Generacja punktów eksperymentalnych przy pomocy planu OLH.*

Punkty OLH są realizacjami zmiennych losowych o jednostajnym rozkładzie prawdopodobieństwa o wartościach oczekiwanych w środku obszaru dopuszczalnego i granicach określonych przez rozmiary tego obszaru. Po transformacji punktów do przestrzeni oryginalnej obliczane są następnie

odpowiadające im wartości funkcji granicznej. Jeśli liczba zmiennych nie jest zbyt duża, plan OLH można zastąpić np. przez doświadczenie czynnikowe. Jednak plan oparty na koncepcji optymalnej hiperkostki łącińskiej wydaje się być dobrym kompromisem pomiędzy ilością punktów a wymogiem równomiernego wypełnienia obszaru dopuszczalnego.

3. *Budowa lokalnej aproksymacji funkcji granicznej przy pomocy liniowej lub kwadratowej powierzchni odpowiedzi.*

W odróżnieniu od oryginalnej metody Z3M, do budowy aproksymacji funkcji granicznej zastosowano metodę ważonej regresji, która opisana została w podrozdziale 3.3 Zgodnie z przedstawioną tam koncepcją, wartości współczynników odpowiadających poszczególnym punktom planu OLH uzależniane są od odległości tych punktów od punktu \mathbf{u} , w którym poszukujemy aproksymacji funkcji granicznej. Funkcja $w(\mathbf{u} - \mathbf{u}_i)$ definiowana jest w taki sposób, aby największy wpływ na wartość funkcji granicznej w \mathbf{u} miały te eksperymenty \mathbf{u}_i , które znajdują się w pobliżu. W związku z koniecznością uaktualniania wartości wag, współczynniki równania regresji muszą być wyznaczane za każdym razem, kiedy rozpatruje się nowy punkt \mathbf{u} . Zgodnie ze wzorami (3.42) i (3.43) aproksymacja funkcji granicznej w przestrzeni standardowej przyjmuje postać

$$\tilde{h}(\mathbf{u}) = \mathbf{a}^T(\mathbf{u})\mathbf{b}(\mathbf{u}), \quad (5.1)$$

gdzie

$$\hat{\mathbf{b}} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{h} \quad (5.2)$$

jest estymatorem nieznanych parametrów \mathbf{b} , \mathbf{W} jest diagonalną macierzą wag, \mathbf{A} jest macierzą regresji (zawiera wartości liniowo niezależnych funkcji bazowych w punktach OLH), \mathbf{a} jest wektorem funkcji bazowych, zaś \mathbf{h} jest wektorem wartości funkcji granicznej obliczonych dla punktów planu OLH. Ponieważ jako funkcję wagową przyjmuje się najczęściej wielowymiarową wykładniczą funkcję przypominającą funkcję gęstości prawdopodobieństwa rozkładu normalnego, dlatego niezerowe elementy macierzy wag \mathbf{W} mają postać, por. (3.44),

$$w_{ii} = \exp \left[-\frac{1}{2n\eta^2} \sum_{j=1}^n (u_{ij} - u_j)^2 \right], \quad i = 1, \dots, N_e, \quad (5.3)$$

gdzie N_e jest liczbą punktów eksperymentalnych użytych do budowy powierzchni odpowiedzi, u_{ij} oznacza j -tą składową i -tego punktu, $j = 1, \dots, n$, $i = 1, \dots, N_e$, u_j jest j -tą składową punktu \mathbf{u} , a η jest parametrem kształtu funkcji wagowej.

4. Znaleźnienie przybliżonego położenia punktu projektowego jako rozwiązania problemu minimalizacji:

$$\text{znaleźć wektor: } \mathbf{u}, \quad (5.4)$$

$$\text{minimalizujący: } \|\mathbf{u}\|^2 = \mathbf{u}^T \mathbf{u}, \quad (5.5)$$

$$\text{pod warunkiem: } \tilde{h}(\mathbf{u}) = 0, \quad (5.6)$$

gdzie $\tilde{h}(\mathbf{u})$ jest aproksymacją funkcji granicznej daną równaniem (5.1). Do rozwiązania problemu (5.4)–(5.6) stosuje się jeden z algorytmów opisanych w podrozdziale A.3: algorytm ARF lub iHLRF. W znalezionym punkcie projektowym obliczana jest dokładna wartość funkcji granicznej, a wynik ten zostaje następnie dodany do bazy danych eksperymentalnych.

5. Zmiana wielkości obszaru dopuszczalnego

5.1 Jeżeli znaleziony w poprzednim kroku punkt projektowy \mathbf{u}^* znajduje się wewnątrz obszaru dopuszczalnego, to przesuwany jest tam jego środek, a rozmiar (długość boku hiperkostki) jest redukowany. Strategia zmniejszania obszaru dopuszczalnego ma kluczowe znaczenie dla zbieżności metody Z3M. Rozwiązanie zaproponowane w [278] zakłada redukcję boku hiperkostki, np. zmniejszenie jego długości o połowę. Niestety, szczególnie w przypadku zagadnień z dużą liczbą zmiennych losowych oraz zaszumioną funkcją graniczną, podejście to powodować może trudności ze zbieżnością algorytmu. Na podstawie wykonanych testów numerycznych wydaje się, że korzystniejszym rozwiązaniem jest dzielenie w stałym stosunku objętości obszaru dopuszczalnego, a nie jego boku, co prowadzi do mniej gwałtownego zmniejszania obszaru dopuszczalnego. Strategia ta omówiona zostanie szczegółowo przy okazji analizy niezawodności zgniatanej belki cienkościenniej zaprezentowanej w podrozdziale 6.1.

5.2 W przypadku gdy rozwiązanie zadania (5.4)–(5.6) znajduje się poza obszarem dopuszczalnym, wtedy rzutowane jest na brzeg obszaru dopuszczalnego w kierunku jego środka. Rzut punktu projektowego staje się następnie nowym środkiem obszaru dopuszczalnego, a rozmiar obszaru nie ulega zmianie. Czasami wskazane jest jednak, aby obszar dopuszczalny został powiększony, szczególnie wtedy gdy kilka kolejnych iteracji prowadzi do rozwiązania poza jego granicami. Zazwyczaj dzieje się tak jeżeli wpływ szumu numerycznego jest znaczący, a obszar dopuszczalny został już istotnie zmniejszony. W zaimplementowanej wersji algorytmu Z3M obszar dopuszczalny rozszerzany jest w odwrotnej proporcji, niż w przypadku operacji redukcji, za każdym razem gdy

w trzech kolejnych iteracjach punkt projektowy znajdujący się jest poza obszarem dopuszczalnym.

6. *Powtarzanie kroków 2 do 5 aż do spełnienia kryterium zbieżności.*

Sprawdzenie warunku zbieżności wykonywane jest tylko wtedy gdy otrzymana w danym kroku iteracyjnym aproksymacja punktu projektowego znajduje się w obszarze dopuszczalnym (zob. punkt 5.1). W uzupełnieniu warunku dotyczącego odległości pomiędzy kolejnymi położeniami środka obszaru dopuszczalnego, zaimplementowane kryterium zbieżności zawiera sprawdzenie czy punkt projektowy znajduje się na powierzchni granicznej, $h(\mathbf{u}^*) = 0$. W celu zmniejszenia wpływu szumu funkcji granicznej na to kryterium, wartość $h(\mathbf{u}^*)$ obliczana jest kilkakrotnie, a do sprawdzenia kryterium epsilonowej bliskości zera używa się oszacowanej wartości średniej tej wielkości.

Budując lokalną aproksymację funkcji granicznej, w zadaniu analizy regresji wykorzystuje się wszystkie punkty eksperymentalne, które znajdują się w obrębie aktualnego obszaru dopuszczalnego, tzn. oprócz nowo wygenerowanych - także punkty z poprzednich iteracji.

5.1.2. Etap drugi: Budowa wielomodalnej gęstości losującej

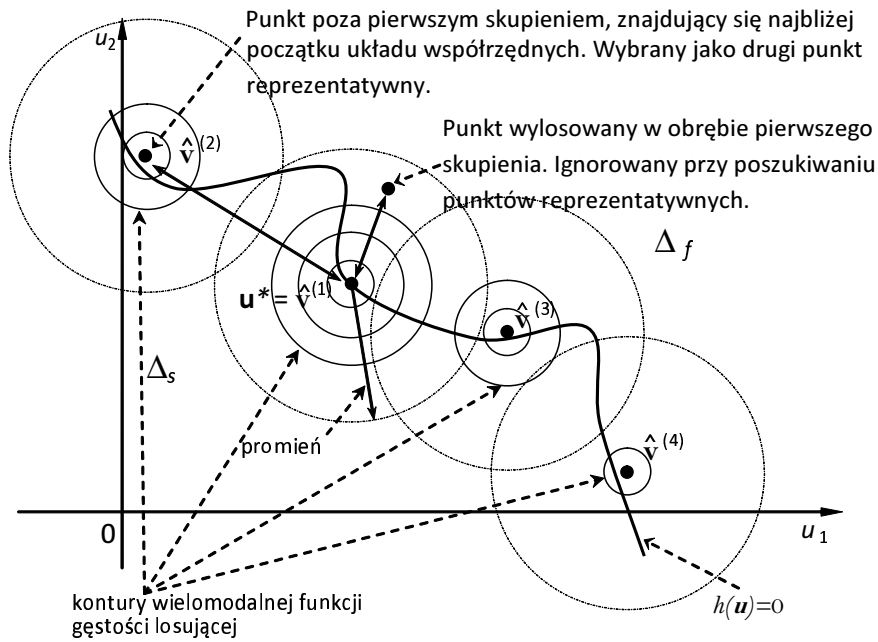
Drugi etap metody Z3M wykorzystuje adaptacyjną procedurę importance sampling zaproponowaną przez Karamchandaniego i in. w pracy [116]. Teoretyczne podstawy symulacji importance sampling omówiono w dodatku A.4.2, a najważniejsze wzory dotyczące technik adaptacyjnych przedstawiono w A.4.3. Poniżej zaprezentowano opis działania metody symulacji losowych zastosowanej w Z3M:

1. *Wygenerowanie N_1 realizacji zmiennych losowych przy pomocy funkcji gęstości zmiennych \mathbf{U} przesuniętej nad znaleziony w pierwszym etapie punkt projektowy. Obliczenie odpowiadających realizacjom wartości funkcji granicznej.*

W pracy [279] zasugerowano przyjęcie $N_1 = 10 \div 50$, jakkolwiek wydaje się bardziej odpowiednim, aby uzależnić liczbę N_1 od wymiaru przestrzeni n , np. $N_1 = 5n \div 10n$. Zgodnie z wprowadzoną w dodatku A.4.2 konwencją oznaczania zmiennych “losujących” symbolem \mathbf{V} oraz pamiętając, że realizacje zmiennych generowane są w gaussowskiej przestrzeni standardowej, gęstość losująca używana do wyznaczenia pierwszej próbki wyraża się wzorem (por. (A.71)) $\varphi_n(\mathbf{v}, \mathbf{u}^*, \mathbf{I})$, gdzie \mathbf{u}^* jest punktem projektowym.

2. Zainicjowanie licznika iteracji $k = 1$.
3. Wybór K punktów reprezentatywnych spośród wszystkich wygenerowanych dotychczas realizacji.

Przez wszystkie realizacje rozumie się punkty wygenerowane zarówno w pierwszym jak i drugim etapie algorytmu Z3M. Pokazany schematycznie na rys. 5.1 algorytm wyboru punktów reprezentatywnych przedstawiać można następująco:



Rys. 5.1. Schemat wyboru punktów reprezentatywnych w procedurze budowania wielomodalnej gęstości losującej.

- 3.1 Pomiń dotychczasowy podział na punkty reprezentatywne i niereprezentatywne.
- 3.2 Wybierz tę realizację zmiennych losowych z obszaru awarii, której odpowiada najwyższa wartość oryginalnej funkcji gęstości prawdopodobieństwa $\varphi_n(\mathbf{v}, \mathbf{0}, \mathbf{I})$. Oznacz ją jako pierwszy punkt reprezentatywny $\hat{v}^{(1)}$. Punkt ten pokrywa się zazwyczaj z punktem projektowym, jeśli jednak iteracyjny proces poszukiwania punktu projektowego zakończył się w obszarze bezpiecznym lub gdy na wartości funkcji granicznej ma

wpływ szum numeryczny, wskazane jest wtedy, aby w algorytmie wyboru punktów reprezentatywnych uwzględnić rozszerzony obszar awarii. Jeśli taka analiza jest możliwa, wielkość epsilonowego rozszerzenia powinna być uzależniona od amplitudy szumu funkcji granicznej, np. punkt klasyfikowany byłby jako należący do obszaru awarii jeśli $h(\mathbf{v}) < \epsilon$, gdzie ϵ jest podwojonym odchyleniem standardowym wartości funkcji granicznej w punkcie projektowym. Jeśli algorytm Z3M zastosowano do zadania bez szumu numerycznego, to w pierwszej iteracji można arbitralnie uznać punkt projektowy za pierwszy punkt reprezentatywny.

- 3.3 Pomiń w dalszej części procedury wyboru wszystkie realizacje znajdujące się wewnątrz skupienia w kształcie n -wymiarowej hipersfery o środku w wybranym ostatnio punkcie reprezentatywnym i promieniu równym połowie odległości punktu projektowego od początku układu współrzędnych, $\|\mathbf{u}^*\|/2$.
- 3.4 Spośród pozostałych realizacji, tzn. punktów, które nie zostały wcześniej pominięte i nie są punktami reprezentatywnymi, znajdź realizację znajdującą się w obszarze awarii, której odpowiada największa wartość funkcji gęstości prawdopodobieństwa. Oznacz tę realizację jako kolejny punkt reprezentatywny.
- 3.5 Powtarzaj kroki 3.3–3.4 aż do “wyczerpania” zbioru wygenerowanych realizacji zmiennych losowych.

4. Obliczenie współczynnika zmienności estymatora prawdopodobieństwa awarii $\nu_{\hat{P}_f}$.

Dla N elementowej próbki, gdzie $N = N_1 + \dots + N_k$, a k jest numerem iteracji współczynnik ten dany jest wzorem, por. (A.63) i (A.75),

$$\nu_{\hat{P}_f}^{(k)} = \frac{\sqrt{\text{Var}[\hat{P}_f^{(k)}]}}{\hat{P}_f^{(k)}}, \quad (5.7)$$

gdzie

$$\begin{aligned} \hat{P}_f^{(k)} = \frac{1}{N} & \left[\sum_{i=1}^{N_1} I_{\Delta_f}(\mathbf{V}_i^{(1)}) \frac{\varphi_n(\mathbf{V}_i^{(1)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(1)}(\mathbf{V}_i^{(1)})} + \sum_{i=1}^{N_2} I_{\Delta_f}(\mathbf{V}_i^{(2)}) \frac{\varphi_n(\mathbf{V}_i^{(2)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(2)}(\mathbf{V}_i^{(2)})} + \right. \\ & \left. + \dots + \sum_{i=1}^{N_k} I_{\Delta_f}(\mathbf{V}_i^{(k)}) \frac{\varphi_n(\mathbf{V}_i^{(k)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(k)}(\mathbf{V}_i^{(k)})} \right], \quad (5.8) \end{aligned}$$

$$\begin{aligned} \text{Var}[\widehat{P}_f^{(k)}] = & \frac{1}{N(N-1)} \left[\sum_{i=1}^{N_1} \left(I_{\Delta_f}(\mathbf{V}_i^{(1)}) \frac{\varphi_n(\mathbf{V}_i^{(1)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(1)}(\mathbf{V}_i^{(1)})} - \widehat{P}_f^{(k)} \right)^2 + \right. \\ & + \sum_{i=1}^{N_2} \left(I_{\Delta_f}(\mathbf{V}_i^{(2)}) \frac{\varphi_n(\mathbf{V}_i^{(2)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(2)}(\mathbf{V}_i^{(2)})} - \widehat{P}_f^{(k)} \right)^2 + \\ & \left. + \dots + \sum_{i=1}^{N_k} \left(I_{\Delta_f}(\mathbf{V}_i^{(k)}) \frac{\varphi_n(\mathbf{V}_i^{(k)}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}^{(k)}(\mathbf{V}_i^{(k)})} - \widehat{P}_f^{(k)} \right)^2 \right], \quad (5.9) \end{aligned}$$

$s_{\mathbf{V}}^{(j)}(\mathbf{v})$ jest wielomodalną funkcją gęstości losującej (5.11) utworzoną w j -tej iteracji (w przypadku pierwszej iteracji $s_{\mathbf{V}}^{(1)}(\mathbf{v}) = \varphi_n(\mathbf{v}, \hat{\mathbf{v}}^{(1)}, \mathbf{I})$), $I_{\Delta_f}(\mathbf{v})$ jest funkcją charakterystyczną obszaru awarii, a przez $\mathbf{v}_i^{(j)}$ oznaczono realizacje zmiennych losowych generowane przez funkcję $s_{\mathbf{V}}^{(j)}(\mathbf{v})$. Jeżeli już dla pierwszej N_1 punktowej próbki obliczona wartość $\nu_{\widehat{P}_f}^{(1)}$ jest mniejsza od pewnej wartości dopuszczalnej, wtedy następuje przejście bezpośrednio do kroku 6.

5. Budowa funkcji gęstości losującej $s_{\mathbf{V}}^{(k)}(\mathbf{v})$ przy użyciu wyznaczonych w kroku 3 punktów reprezentatywnych:

5.1 Dla każdego punktu reprezentatywnego $\hat{\mathbf{v}}^{(i)}$, $i = 1, \dots, K$, oblicz wartość współczynnika wagowego (zob. (A.74))

$$\hat{w}^{(i)} = \frac{\varphi_n(\hat{\mathbf{v}}^{(i)}, \mathbf{0}, \mathbf{I})}{\sum_{j=1}^K \varphi_n(\hat{\mathbf{v}}^{(j)}, \mathbf{0}, \mathbf{I})}. \quad (5.10)$$

5.2 Zgodnie ze wzorem (A.73) zdefiniuj wielomodalną funkcję gęstości $s_{\mathbf{V}}^{(k)}(\mathbf{v})$ jako ważoną sumę funkcji gęstości prawdopodobieństwa odpowiadających poszczególnym punktom reprezentatywnym

$$s_{\mathbf{V}}^{(k)}(\mathbf{v}) = \sum_{i=1}^K \hat{w}^{(i)} \varphi_n(\mathbf{v}, \hat{\mathbf{v}}^{(i)}, \mathbf{I}), \quad (5.11)$$

gdzie $\varphi_n(\mathbf{v}, \hat{\mathbf{v}}^{(i)}, \mathbf{I})$ oznacza standardową funkcję gęstości rozkładu normalnego o wartościach średnich przesuniętych do i -tego punktu reprezentatywnego $\hat{\mathbf{v}}^{(i)}$.

5.3 Zwiększ licznik iteracji, $k = k + 1$,

- 5.4 Wygeneruj N_k dodatkowych realizacji zmiennych losowych używając funkcji (5.11). Wielkość próbki przyjmuje się zazwyczaj jako wielokrotność liczby punktów reprezentatywnych K , np. $N_k = 5K \div 20K$. Oblicz wartości funkcji granicznej odpowiadające wygenerowanym realizacjom.
- 5.5 Oblicz wartości $\widehat{P}_f^{(k)}$, $\text{Var}[\widehat{P}_f^{(k)}]$ oraz współczynnika zmienności estymatora $\nu_{\widehat{P}_f}^{(k)}$ na podstawie wzorów (5.7)–(5.9) używając utworzonej w 5.2 gęstości losującej. Jeżeli współczynnik zmienności jest większy od przyjętej wartości dopuszczalnej, wtedy wyznacz nowe punkty reprezentatywne zgodnie z procedurą opisaną w kroku 3, a następnie powtórz wszystkie działania podane w kroku 5. Jeśli $\nu_{\widehat{P}_f}^{(k)}$ jest mniejsze od wartości dopuszczalnej, to przejdź do kroku 6. W pracy [278] sugerowano, aby wartości dopuszczalne współczynnika zmienności estymatora przyjmować w granicach $0.2 \div 0.3$.
6. *Zwiększenie licznika iteracji, $k = k + 1$.*
7. *Wygenerowanie N_k realizacji zmiennych losowych przy pomocy rozkładu wielomodalnego (5.11) dla końcowej konfiguracji punktów reprezentatywnych.*
8. *Sprawdzenie warunku zbieżności.*
Obliczona ze wzoru (5.8) wartość estymatora prawdopodobieństwa awarii służy do sprawdzenia następującego kryterium zbieżności

$$\frac{|\widehat{P}_f^{(k)} - \widehat{P}_f^{(k-1)}|}{\widehat{P}_f^{(k-1)}} \leq \varepsilon, \quad (5.12)$$

gdzie ε jest założoną tolerancją (np. $\varepsilon = 0.1$). Dodatkowo wymaga się, aby wartość współczynnika zmienności $\nu_{\widehat{P}_f}^{(k)}$ pozostała poniżej wartości dopuszczalnej. W przypadku nie spełnienia powyższych warunków następuje powrót do kroku 6.

Kilka cech wymienionego wyżej algorytmu czyni go szczególnie przydatnym w przypadku nieliniowych i kosztownych obliczeniowo problemów analizy niezawodności. Do problemów tych zaliczyć można analizę niezawodności w zagadnieniach dotyczących wytrzymałości zderzeniowej, jak również zadania, w których zakłada się losowość parametrów opisujących geometrię konstrukcji, gdzie szum numeryczny funkcji granicznej generowany jest w wyniku adaptowania siatki elementów skończonych. W tym kontekście można wymienić następujące zalety metody Z3M:

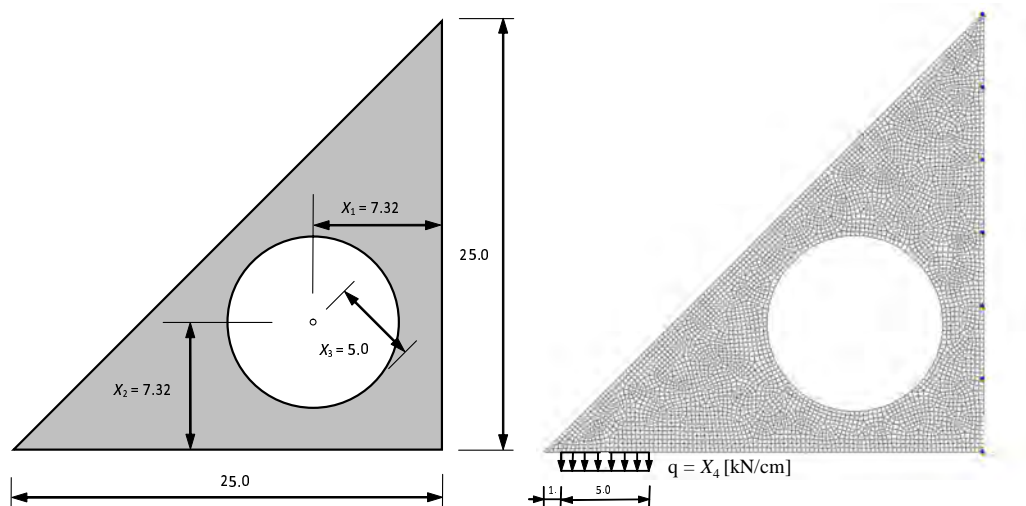
- Ze względu na zastosowanie w algorytmie poszukiwania punktu projektowego adaptacyjnej aproksymacji funkcji granicznej, zbieżność pierwszego etapu metody można stosunkowo łatwo uzyskać nawet w przypadku występowania szumu numerycznego.
- Eksperymenty numeryczne wygenerowane w pierwszym etapie metody są następnie używane w procedurze wyznaczania punktów reprezentatywnych w adaptacyjnym algorytmie importance sampling.
- W sposób elastyczny można kontrolować dokładność metody i związane z nią nakłady obliczeniowe. Jeśli nie wymaga się wysokiej precyzji oszacowania prawdopodobieństwa awarii albo nie można sobie na to pozwolić ze względu na ograniczony budżet analizy niezawodności, wtedy drugi etap algorytmu może być pominięty.
- Dzięki procedurze budowania wielomodalnej gęstości losującej metoda posiada możliwość uwzględniania wpływu wielokrotnych punktów projektowych.

5.1.3. Szum numeryczny generowany przez losową geometrię konstrukcji. Analiza FORM przy użyciu procedury Z3M

Pozostawiając szczegółowe omówienie efektywności algorytmu Z3M do przykładu analizy niezawodności zgniatanej dynamicznie belki cienkościennej (podrozdział 6.1), poniżej przedstawiono użycie pierwszego etapu Z3M w zadaniu gdzie szum numeryczny funkcji granicznej wywołany jest przez zmienne losowe opisujących geometrię konstrukcji. Ponieważ każda realizacja takich zmiennych niesie za sobą konieczność uaktualniania siatki elementów skończonych, dlatego na wartość funkcji granicznej wpływ ma nie tylko nowa geometria konstrukcji lecz także zmieniona dyskretyzacja MES, będąca konsekwencją nowej geometrii.

Rozpatrzmy trójkątną tarczę z otworem kołowym pokazaną na rys. 5.2. Znajdującą się w płaskim stanie naprężenia stalowa tarcza poddana jest obciążeniu jednostajnemu przyłożonemu na fragmencie poziomej krawędzi. Pionowa krawędź jest utwierdzona. Grubość tarczy przyjęto równą 0.3 cm. Model probabilistyczny konstrukcji składa się z pięciu zmiennych losowych o rozkładzie normalnym, których opis (wartości oczekiwane oraz odchylenia standardowe) podano w tab. 5.1.

Awarię konstrukcji zdefiniowano jako przekroczenie w dowolnym punkcie tarczy dopuszczalnego zredukowanego naprężenia Hubera-Misesa $S_a = 21 \text{ kN/cm}^2$,



Rys. 5.2. Rysunek lewy - wymiary tarczy. Rysunek prawy - siatka MES, obciążenie oraz warunki podparcia. Wymiary podano w centymetrach.

Tablica 5.1. Parametry gaussowskich rozkładów prawdopodobieństwa zmiennych losowych z przykładu analizy niezawodności tarczy trójkątnej, rys. 5.2. Ostatnia kolumna - współrzędne punktu projektowego.

Zmienna	Opis	Wartość oczekiwana	Odchylenie standardowe	Wartość w punkcie proj.
X_1	wsp. x środ. otworu	7.32 cm	0.1 cm	7.50 cm
X_2	wsp. y środ. otworu	7.32 cm	0.1 cm	7.45 cm
X_3	promień otworu	5.0 cm	0.1 cm	5.07 cm
X_4	wartość obciążenia	4.0 kN/cm	0.4 kN/cm	4.76 kN/cm
X_5	moduł Younga	21000 kN/cm ²	2100 kN/cm ²	20089 kN/cm ²

co prowadzi do następującego wyrażenia na funkcję graniczną:

$$g(\mathbf{X}) = 1 - \frac{\max[S(\mathbf{X})]}{S_a}, \quad (5.13)$$

gdzie $\max[S(\mathbf{X})]$ oznacza maksymalne naprężenie Hubera-Misesa obliczone w punktach węzłowych siatki. Jak już wspomniano, zmiany położenia oraz wielkości otworu powodują każdorazową rekonfigurację siatki elementów. W związku z tym, odległość najbliższego węzła od punktu gdzie występuje maksimum naprężeń, nawet przy dość gęstym podziale na elementy, jest za każdym razem inna, a obliczane wartości funkcji granicznej obarczone są pewną dodatkową, trudną do oceny, niepewnością.

Zaszumiony charakter funkcji (5.13) uniemożliwia użycie klasycznych metod poszukiwania punktu projektowego, takich jak iHLRF czy ARF (zob. dodatek A.3), dlatego w analizowanym przykładzie zastosowano adaptacyjną technikę powierzchni odpowiedzi zaprezentowaną w pierwszym etapie algorytmu Z3M.

Aproksymacje funkcji granicznej tworzone były na podstawie 50-cio punktowych planów eksperymentów numerycznych OLH. Jako metodę aproksymacji przyjęto regresję ważoną, zob. (5.1)–(5.3), z liniowymi funkcjami bazowymi. Rozpoczynając w punkcie wartości oczekiwanych, zbieżność algorytmu uzyskano po 10 iteracjach. Otrzymany wskaźnik niezawodności pierwszego rzędu β_{FORM} wynosi 2.99, co odpowiada prawdopodobieństwu awarii $P_f^{\text{FORM}} = 1.4 \cdot 10^{-3}$.

Porównując wartości realizacji zmiennych losowych w punkcie projektowym z odpowiednimi wartościami oczekiwanymi (zob. tab. 5.1) można zauważyć jedynie niewielkie różnice w przypadku zmiennych opisujących geometrię oraz moduł Younga. Z drugiej strony, w sposób istotny wzrosła wartość obciążenia, co okazuje się być najbardziej prawdopodobną przyczyną awarii.

W przedstawionym przykładzie analiza stanu naprężenia przeprowadzona została przy pomocy komercyjnego programu elementów skończonych ABAQUS. Sama integracja programu ABAQUS z programem do analizy niezawodności nie jest zadaniem trywialnym. Problematyka ta zostanie szczegółowo omówiona w rozdziale 8 (punkt 8.3.1) na przykładzie integracji z systemem STAND, który rozwijany jest w Pracowni Niezawodności i Optymalizacji IPPT PAN.

5.2. Metoda poprawy niezawodności z wykorzystaniem losowych symulacji OLH

Prezentowana metoda ma na celu szybką poprawę niezawodności konstrukcji przy pomocy symulacji losowych wykorzystujących koncepcję optymalnej łańciskiej hiperkostki. Chociaż zaprojektowana została jako technika rozwiązywania zadania optymalizacji niezawodnościowej (zob. (2.15)–(2.19)), to ze względu na swoją specyfikę oraz planowany zakres zastosowań bardziej właściwe jest mówienie o tej metodzie w kontekście poprawy niezawodności konstrukcji, a nie w kontekście poszukiwania konstrukcji optymalnej. Zaletą algorytmu wydaje się być duża efektywność w przypadku umiarkowanej liczby zmiennych losowych oraz wtedy, gdy konstrukcję wyjściową cechuje stosunkowo niski poziom niezawodności.

W przeciwieństwie do algorytmu Z3M metoda ta nie korzysta z żadnych aproksymacji funkcji granicznej. O ile bowiem umiejętnie zastosowana metoda

powierzchni odpowiedzi stanowić może bardzo przydatne narzędzie zarówno w analizie niezawodności jak też rozwiązywaniu problemów optymalizacji, o tyle nieprzemyślane używanie standardowych metod aproksymacji funkcji wielu zmiennych w przypadku funkcji silnie nieliniowych, prowadzi niekiedy może do zbyt wielu uproszczeń oraz wyników obarczonych znacznym błędem. Z funkcjami takimi mamy często do czynienia gdy składnik szumowy jest znaczący lub gdy opisywana odpowiedź konstrukcji ma charakter nieciągły (tak jak w niektórych przypadkach analizy stateczności konstrukcji).

Z powyższych powodów tworzenie powierzchni odpowiedzi modelującej zachowanie się tych elementów pojazdów, które pochłaniają energię zderzeń, zaliczyć należy do zadań wyjątkowo trudnych i wymagających dużej wiedzy i doświadczenia. Pośród prac naukowych odnaleźć nawet można bardzo negatywne opinie, kwestionujące samą zasadność prowadzenia optymalizacji czy też analizy niezawodności tego typu konstrukcji w oparciu o matematyczną aproksymację ich odpowiedzi, zob. [165].

Dystansując się od tak ostro sformułowanych wniosków, o czym świadczy chociażby tematyka niniejszego opracowania, należy przyznać iż metody rzędu zerowego lub metody pseudolosowego przeszukiwania mają swoje bezsprzeczne zalety, do których należą prostota implementacji i uniknięcie problemu błędów aproksymacji. Mimo, iż nie mają one reputacji efektywnych obliczeniowo, to w pracy [228] pokazano, że w pewnych przypadkach użycie metod losowego przeszukiwania z wykorzystaniem łacińskich hiperkostek w rozwiązywaniu zadań optymalizacji może dać bardzo dobre wyniki w porównaniu do metod opartych na użyciu powierzchni odpowiedzi.

W przedstawionej niżej metodzie zaproponowano wykorzystanie współczynników korelacji pomiędzy funkcjami ograniczeń a zmiennymi losowymi, których wartości oczekiwane są zmiennymi projektowymi w zadaniu optymalizacji. Wartości tych współczynników służą następnie do wyboru kierunku zmian wektora zmiennych projektowych. Zastosowany w algorytmie plan OLH pełni podwójną rolę. Pomaga jak najlepiej zbadać przestrzeń zmiennych projektowych, jak również służy do efektywnego szacowania wartości współczynników korelacji, por. podrozdział 4.1.

5.2.1. Sformułowanie problemu

Przyjmijmy, że $\mathbf{X} = \{X_1, X_2, \dots, X_r\}$ jest wektorem niezależnych zmiennych losowych, których rozkłady prawdopodobieństwa stanowią probabilistyczny

model konstrukcji. Jeśli w rzeczywistości zmienne nie są niezależne, to konieczne jest zastosowanie odpowiedniej transformacji (zob. dodatek A.2). Przyjmijmy następnie, że jako zmienne projektowe wybrano wartości oczekiwane n zmiennych losowych, gdzie $n \leq r$. Oznaczając wektor zmiennych projektowych symbolem \mathbf{y} i zakładając w celu uproszczenia zapisu, że zmienne te odpowiadają n pierwszym zmiennym losowym, mamy

$$\mathbf{y} = \{y_1 = \mu_{X_1}, y_2 = \mu_{X_2}, \dots, y_n = \mu_{X_n}\}, \quad n \leq r. \quad (5.14)$$

W zależności od charakteru zmiennej losowej X_i , $i = 1, \dots, n$, jej odchylenie standardowe pozostaje stałe w procesie optymalizacji lub zmienia się proporcjonalnie do wartości oczekiwanej zgodnie z ustalonym współczynnikiem zmienności $\nu_{X_i} = \sigma_{X_i}/\mu_{X_i}$.

Zadanie optymalizacji niezawodnościowej/poprawy niezawodności dane jest w następujący sposób (por. sformułowanie (2.15)–(2.19)):

$$\text{znajdź wartości zmiennych: } \mathbf{y} \in \mathbb{R}^n, \quad (5.15)$$

$$\text{minimalizujące: } f(\mathbf{y}), \quad (5.16)$$

$$\text{przy ograniczeniach: } \mathbb{P}[g_i(\mathbf{X}) \leq 0] \leq p_i^a, \quad i = 1, \dots, k_g, \quad (5.17)$$

$${}^l y_j \leq y_j \leq {}^u y_j, \quad j = 1, \dots, n, \quad (5.18)$$

gdzie $f(\mathbf{y})$ jest funkcją celu, $g_i(\mathbf{x})$, $i = 1, \dots, k_g$, są funkcjami ograniczeń tworzącymi ograniczenia niezawodnościowe, p_i^a są dopuszczalnymi wartościami prawdopodobieństw awarii, a ${}^l y_j$ i ${}^u y_j$, $j = 1, \dots, n$, są odpowiednio dolnymi i górnymi ograniczeniami prostymi. Ograniczenia (5.17) przedstawiają zbiór warunków $g_i(\mathbf{x}) \leq 0$, których spełnienie dopuszcza się jedynie z bardzo małym prawdopodobieństwem p_i^a , lub zapisując inaczej $P_{f_i} \leq p_i^a$, gdzie P_{f_i} jest prawdopodobieństwem awarii.

Należy zaznaczyć, że metoda Monte Carlo, która używana będzie do obliczania wartości prawdopodobieństwa awarii, prowadzi do bardzo zgrubnych oszacowań w przypadku mało licznych próbek i w przypadku gdy wymagania dotyczące niezawodności konstrukcji są wysokie, zob. dodatek A strona 277. Decydując się, np. na stosowanie próbek nie zawierających więcej niż 100 realizacji zmiennych losowych (co jest dość realistycznym założeniem dla wielu kosztownych obliczeniowo problemów analizy niezawodności) i ustalając $p_i^a = 0.01$, ograniczenia (5.17) powinno się sformułować jako: $g_i(\mathbf{x}) \geq 0$ dla wszystkich realizacji \mathbf{x} wchodzących w skład próbki. W rzeczywistości, aby zagwarantować dobrą estymację (współczynnik zmienności estymatora < 0.1) prawdopodobieństwa awarii rzędu 0.01 próbka musi składać się z około 1000 punktów.

Generowanie realizacji według planów OLH pozwala na uzyskanie pewnego polepszenia efektywności estymacji prawdopodobieństwa awarii, zob. [186, 277], ale nadal w porównaniu do standardowej metody Monte Carlo nie jest to zmiana jakościowa. Pomimo to, ponieważ głównym celem metody jest poprawa niezawodności rozwiązania charakteryzującego się stosunkową wysokim prawdopodobieństwem awarii, a nie poszukiwanie optimum, dlatego akceptuje się ograniczoną dokładność metody symulacji losowych OLH.

5.2.2. Opis algorytmu

Proponowana metoda rozwiązania problemu (5.15)–(5.18) może być zaklasyfikowana jako metoda losowego przeszukiwania. W celu znalezienia początkowych wartości wektora zmiennych projektowych, jak również jego wszystkich kolejnych iteracji, używa się techniki symulacji OLH. Podstawową zasadą algorytmu jest wybieranie następnej iteracji zmiennych projektowych jedynie spośród punktów (realizacji) wygenerowanych do tej pory, tych dla których obliczono wartości zarówno funkcji celu jak i funkcji ograniczeń $g_i(\mathbf{x})$, $i = 1, \dots, k_g$. Postępując w ten sposób, w przeciwieństwie do metod wykorzystujących koncepcję powierzchni odpowiedzi, unika się jakichkolwiek założeń dotyczących postaci funkcji celu oraz ograniczeń. Rezygnacja z aproksymacji oraz poruszanie się jedynie po sprawdzonych rozwiązaniach czynią algorytm podobnym do metody opracowanej przez Marczyka [165] i zaimplementowanej w programie ST-ORM [227].

5.2.2.1. Wybór punktu startowego

W przypadku braku preferencji co do początkowych wartości zmiennych projektowych, punkt startowy wybiera się na podstawie analizy wyników odpowiadających realizacjom wygenerowanym zgodnie z planem OLH. W tym celu oryginalne rozkłady prawdopodobieństwa zmiennych losowych odpowiadających zmiennym projektowym, X_i , $i = 1, \dots, n$, zamieniane są na rozkłady jednostajne o granicach określonych przez l_{y_i} , u_{y_i} , $i = 1, \dots, n$. Następnie generowana jest próbka N realizacji tak zmodyfikowanych zmiennych, dla których obliczane są wartości funkcji f i g_i , $i = 1, \dots, k_g$. Ten punkt, który spełniając ograniczenia $g_i(\mathbf{x}) > 0$ minimalizuje wartość funkcji celu, wybierany jest jako punkt startowy. Oznacza to, że wartości oczekiwane zmiennych X_i , $i = 1, \dots, n$, przyjmują odpowiednie wartości współrzędnych punktu początkowego. Przywraca się oryginalne funkcje gęstości rozkładów prawdopodobieństwa z nowymi wartościami oczekiwanymi i początkowymi wartościami odchyłeń standardowych.

W przypadku założenia stałego współczynnika zmienności, wartość odchylenia standardowego musi być obliczona proporcjonalnie do wartości oczekiwanej.

Powodem zmiany typów rozkładów prawdopodobieństwa jest dążenie do równomiernego wypełnienia punktami testowymi całego zakresu zmiennych projektowych. Poza tym, do poszukiwania punktu startowego używa się specjalnie zmodyfikowanej optymalnej hiperkostki łacińskiej, zob. paragraf 5.2.3.

5.2.2.2. Kryterium wyboru następnej iteracji zmiennych projektowych

Na każdym kroku iteracyjnym algorytmu, zgodnie z przyjętym planem OLH, losowanych jest N realizacji wektora \mathbf{X} . Po obliczeniu odpowiadających poszczególnym realizacjom wartości funkcji f i g_i , $i = 1, \dots, k_g$, spośród wygenerowanych punktów wybierane jest nowe przybliżenie rozwiązania. Zastosowaną strategię wyboru opisać można następująco:

Na początku szacowane są wartości ograniczeń niezawodnościowych (5.17). Oznaczmy przez I zbiór indeksów tych ograniczeń, które nie są spełnione (są aktywne)

$$I = \{i : 1 \leq i \leq k_g, \mathbb{P}[g_i(\mathbf{X}) \leq 0] > p_i^a\}. \quad (5.19)$$

Następnie, biorąc pod uwagę jedynie wyniki dla ostatniej próbki, oblicza się współczynniki korelacji pomiędzy zmiennymi losowymi X_i , $i = 1, \dots, n$, a funkcjami ograniczeń $g_j(\mathbf{x})$, $j \in I$. Przyjmują one postać

$$\rho_{ij} = \frac{\sum_{k=1}^N [x_i^{(k)} - \bar{x}_i][g_j(\mathbf{x}^{(k)}) - \bar{g}_j]}{\sqrt{\sum_{k=1}^N [x_i^{(k)} - \bar{x}_i]^2} \sqrt{\sum_{k=1}^N [g_j(\mathbf{x}^{(k)}) - \bar{g}_j]^2}}, \quad i = 1, \dots, n, \quad j \in I, \quad (5.20)$$

gdzie $\rho_{ij} = \rho(X_i, g_j)$, \bar{x}_i i \bar{g}_j są obliczonymi na podstawie próbki wartościami średnimi, a $x_i^{(k)}$ oznacza i -tą składową k -tej realizacji wektora losowego \mathbf{X} . Otrzymane współczynniki korelacji odgrywają zasadniczą rolę w wyborze kierunku poszukiwania nowych wartości zmiennych projektowych, a dokładniej, w wyodrębnieniu podzbioru wygenerowanych punktów, z którego wybrana będzie następna iteracja zmiennych projektowych. Ze zbioru tego można od razu wyeliminować te realizacje, $\mathbf{x}^{(k)}$, które nie spełniają ograniczeń prostych (5.18) lub warunków $g_i(\mathbf{x}^{(k)}) > 0$, $i = 1, \dots, k_g$. W dalszej części przez \tilde{I} oznaczać będziemy zbiór indeksów punktów spełniających powyższe kryteria

$$\begin{aligned} \tilde{I} = \{k : 1 \leq k \leq N, g_i(\mathbf{x}^{(k)}) \geq 0, i = 1, \dots, k_g, \\ \text{i } {}^l y_j \leq x_j^{(k)} \leq {}^u y_j, j = 1, \dots, n\}. \end{aligned} \quad (5.21)$$

Ze zbioru $\{\mathbf{x}^{(i)} : i \in \tilde{I}\}$ chcemy wybrać taki punkt, który daje największe szanse na poprawę niespełnionych obecnie ograniczeń niezawodnościowych. W tym celu dla każdej zmiennej projektowej obliczane jest następujące wyrażenie

$$b_i = \sum_{j \in I} \rho(X_i, g_j) \left(1 + \mathbb{P}[g_j(\mathbf{X}) < 0]\right), \quad i = 1, \dots, n. \quad (5.22)$$

Jeśli b_i ma wartość dodatnią, wtedy rozpatrywane są jedynie te punkty, których i -ta współrzędna jest większa od wartości oczekiwanej μ_{X_i} , a w przypadku $b_i < 0$ bierze się pod uwagę tylko punkty spełniające warunek $x_i^{(j)} < \mu_{X_i}$, $j \in \tilde{I}$. Przypadek $b_i = 0$ nie eliminuje żadnych realizacji ze zbioru. Pomysł przyświecający wprowadzeniu wyrażenia (5.22) wynika z chęci wykorzystania informacji, które niesie ze sobą współczynnik korelacji. Przyjmując wartości od -1 do 1 , stanowi on miarę liniowości relacji pomiędzy zmiennymi losowymi a funkcjami ograniczeń. Występujący we wzorze współczynnik wagowy (w nawiasach) ma za zadanie uwzględnienie zakresu przekroczenia dopuszczalnej wartości prawdopodobieństwa awarii.

Ze względu na nieliniowy charakter funkcji ograniczeń może się zdarzyć, że opisana powyżej procedura wyboru wykluczy ze zbioru rozpatrywanych realizacji takie, które potencjalnie mogłyby być kandydatami na nową iterację wartości zmiennych projektowych. Aby uniknąć podobnych sytuacji, do wybranych wcześniej punktów z próbki dodaje się (o ile takie istnieją) realizacje, w przypadku których wartości funkcji g_i , $i \in I$, są większe od odpowiednich wartości dla punktu wybranego w poprzednim kroku jako aktualna iteracja zmiennych projektowych.

W celu zilustrowania omawianej procedury rozważmy przykład z trzema zmiennymi projektowymi, gdzie na typowym kroku algorytmu losowanych jest 50 realizacji zmiennych losowych. Wartości dopuszczalnych prawdopodobieństw awarii p^a (zob. (5.17)) są takie same dla wszystkich ograniczeń i wynoszą 0.01. Załóżmy także, że rozpatruje się 4 funkcje ograniczeń, $g_i(\mathbf{x})$, $i = 1, \dots, 4$, i że na danym kroku iteracyjnym funkcje te przyjęły wartości ujemne dla odpowiednio 20, 10, 3 i 15 punktów z próbki OLH. Po obliczeniu współczynników korelacji

(w poniższej tabeli), oblicza się następnie wartości b_i , $i = 1, \dots, 3$,

	g_1	g_2	g_3	g_4
X_1	-0.6	-0.1	0.8	-0.2
X_2	0.01	0.5	-0.9	-0.3
X_3	0.65	0.2	-0.3	0.1

$$b_1 = -0.6\left(1 + \frac{20}{50}\right) - 0.1\left(1 + \frac{10}{50}\right) + 0.8\left(1 + \frac{3}{50}\right) - 0.2\left(1 + \frac{15}{50}\right) = -0.37,$$

$$b_2 = 0.01\left(1 + \frac{20}{50}\right) + 0.5\left(1 + \frac{10}{50}\right) - 0.9\left(1 + \frac{3}{50}\right) - 0.3\left(1 + \frac{15}{50}\right) = -0.73,$$

$$b_3 = 0.65\left(1 + \frac{20}{50}\right) + 0.2\left(1 + \frac{10}{50}\right) - 0.3\left(1 + \frac{3}{50}\right) + 0.1\left(1 + \frac{15}{50}\right) = 0.96.$$

Zgodnie z zaproponowaną strategią, punktami ze zbioru $\{\mathbf{x}^{(i)}, i \in \tilde{I}\}$, spośród których wybrane będzie następne przybliżenie punktu optymalnego, będą punkty spełniające warunki $x_1^{(i)} < \mu_{X_1}$, $x_2^{(i)} < \mu_{X_2}$ i $x_3^{(i)} > \mu_{X_3}$ oraz ewentualnie wspomniane już punkty dodatkowe.

W przypadku dużej liczby zmiennych projektowych jest bardzo prawdopodobne, że żadna z wylosowanych realizacji nie spełni wszystkich kryteriów narzuconych przez znaki wartości b_i , $i = 1, \dots, n$. Wtedy konieczne jest wygenerowanie dodatkowych punktów (co najmniej jednego) w obszarze określonym przez "kierunki" b_i . Oczywiście, nowy punkt rozpatrywany jest jedynie wtedy, jeśli spełnia wszystkie ograniczenia, zob. (5.21).

Po wyodrębnieniu zbioru potencjalnych rozwiązań zadania należy wprowadzić dodatkowe kryterium pozwalające na wybór rozwiązania preferowanego. W zamieszczonych w niniejszej pracy przykładach zastosowań algorytmu przyjęto kryterium postaci

$$\zeta(\mathbf{x}) = w_f \tilde{f}(\mathbf{x}) - \sum_{i=1}^{k_g} w_i \tilde{g}_i(\mathbf{x}), \quad (5.23)$$

gdzie $\tilde{f}(\mathbf{x})$ i $\tilde{g}_i(\mathbf{x})$, $i = 1, \dots, k_g$, są znormalizowanymi wartościami funkcji celu oraz ograniczeń, a w_f i w_i są współczynnikami wagowymi. Realizacja zmiennej losowej $\hat{\mathbf{x}}$, która minimalizuje funkcję ζ przyjmowana jest jako następna iteracja rozwiązania, tzn. $y_i = \hat{x}_i$, $i = 1, \dots, n$.

Istnieje pewna dowolność jeśli chodzi o dobór wartości współczynników wagowych. Przyjmę np. w_f jako

$$w_f = \exp\left(1 - \frac{\mathbb{P}[g_j(\mathbf{X}) < 0]}{1 - p_k^a}\right), \quad (5.24)$$

gdzie j jest indeksem “najbardziej aktywnego” ograniczenia niezawodnościowego, preferuje punkty, które prowadzą przede wszystkim do spełnienia ograniczeń niezawodnościowych. Podobny efekt można osiągnąć wprowadzając następującą definicję współczynników w_i , $i = 1, \dots, k_g$:

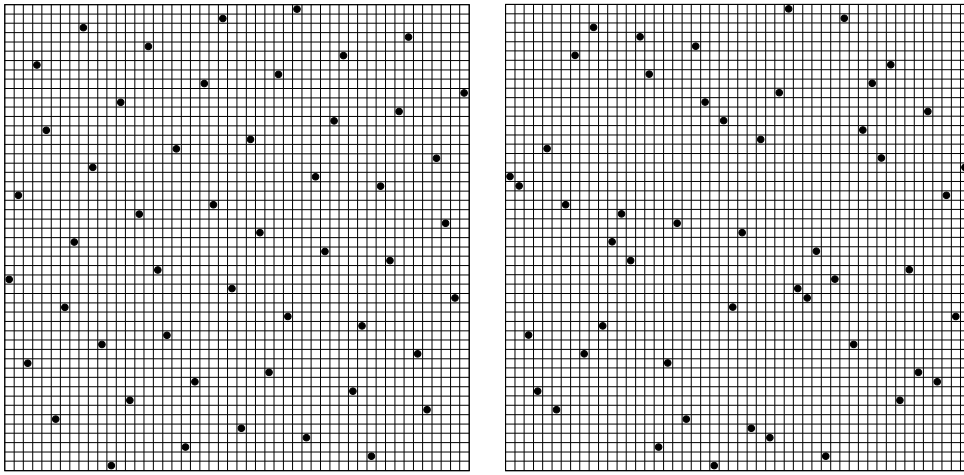
$$w_i = 1 + \mathbb{P}[g_i(\mathbf{X}) < 0]. \quad (5.25)$$

Pominięcie w powyższym wzorze jedynki powoduje uwzględnienie jedynie aktywnych ograniczeń niezawodnościowych. Dobór współczynników wagowych poddyktowany czasem bywa kosztem otrzymania wartości funkcji celu oraz funkcji ograniczeń. Jeśli nie można pozwolić sobie na wiele iteracji algorytmu używając wystarczająco licznych próbek lub gdy głównym zadaniem projektanta jest poprawa niezawodności wyjściowej konstrukcji, wtedy przyjmuje się $w_f = 0$, a proces iteracyjny zatrzymywany jest zaraz po znalezieniu rozwiązania spełniającego ograniczenia niezawodnościowe.

5.2.3. Zmodyfikowany plan OLH

Koncepcja łańciskowej hiperkostki zapewnia równomierne rozłożenie punktów eksperymentalnych przy rzutowaniu ich na dowolną z osi współrzędnych, zob. dodatek B. Ponadto, optymalne hiperkostki - OLH prowadzą do równomiernego rozmieszczenia punktów w hiperkostce r -wymiarowej, gdzie r jest wymiarem wektora losowego. Pozwala to na uniknięcie skupień eksperymentów w pewnych obszarach i pozostawienia innych obszarów niezbadanymi. Pomimo to, nie można uniknąć tworzenia się takich skupień w przypadku rzutowania punktów planu OLH na przestrzeń $n < r$ wymiarowe. Na rys. 5.3 pokazano dwie łańciskowe hiperkostki (kwadraty łańciskowe). Kwadrat po lewej to 50-punktowy 2-wymiarowy plan OLH. Punkty są tu równomiernie rozłożone wewnątrz kwadratu. Po prawej stronie pokazano kwadrat łańciski będący rzutem 50-punktowej 4-wymiarowej hiperkostki OLH na podprzestrzeń dwuwymiarową. W tym przypadku skupienia punktów są wyraźnie widoczne.

Tworzenie się takich skupień wpływa szczególnie niekorzystnie na działanie procedury wyboru punktu startowego, zob. 5.2.2.1. Ponieważ wynik końcowy



Rys. 5.3. Po lewej: plan OLH 50×2 . Po prawej: rzut planu OLH 50×4 na podprzestrzeń dwuwymiarową.

bardzo często zależy od właściwego wyboru punktu rozpoczynającego proces iteracyjny, dlatego jest niesłychanie istotne aby przestrzeń zmiennych projektowych została pokryta w sposób równomierny. Z drugiej strony, ideą prezentowanej metody optymalizacji niezawodnościowej (poprawy niezawodności) jest by zawsze uwzględniać w algorytmie losowy charakter pozostałych zmiennych - tych nie związanych ze zmiennymi projektowymi. W celu znalezienia kompromisu pomiędzy dwoma sprzecznymi oczekiwaniami (optymalności zarówno hiperkostki r -wymiarowej jak i jej n -wymiarowego rzutu) zaproponowano zmodyfikowany plan OLH. Jest on optymalny w podprzestrzeni zmiennych projektowych, natomiast współrzędne punktów odpowiadające pozostałym $r - n$ zmiennym losowym są wybierane przy pomocy algorytmu CP (zob. B.2) lub też algorytmu genetycznego (zob. B.3) przyjmując dodatkowe ograniczenia dotyczące zachowania równomiernego rozkładu rzutów punktów w podprzestrzeni zmiennych projektowych.

Komputerowa implementacja powyższych modyfikacji nie przysparza specjalnych trudności. Na początku budowany jest “projektowy” plan OLH, tzn. znajdowana jest właściwa macierz \mathbf{L}_d o wymiarach $N \times n$, por. (B.1). Następnie w macierzy \mathbf{L} o wymiarach $N \times r$ będącej startowym rozwiązaniem w algorytmie CP (lub w początkowej populacji hiperkostek w algorytmie genetycznym) kolumny odpowiadające zmiennym projektowym zastępowane są przez odpowiednie kolumny z macierzy \mathbf{L}_d . Uwzględnienie optymalnego rozkładu rzutów punktów eksperymentalnych na podprzestrzeń n wymiarową sprowadza się do

pomijania w opisanej w B.2 procedurze *CP-sweep* tych kolumn macierzy \mathbf{L} , które skopiowano z \mathbf{L}_d , a w przypadku algorytmu genetycznego - na niestosowaniu do tych kolumn operatorów mutacji i krzyżowania.

5.2.4. Przykład testowy: Minimalizacja funkcji Rosenbrocka z szumem

Działanie przedstawionego algorytmu zilustrowano na przykładzie minimalizacji z ograniczeniami dwuwymiarowej funkcji Rosenbrocka, por. (B.14) i (4.32). Problem sformułowany jest następująco:

$$\text{znajdź wartości zmiennych: } y_1 = \mu_{X_1}, y_2 = \mu_{X_2}, \quad (5.26)$$

$$\text{minimalizujące: } f(\mathbf{y}) = 100(y_2 - y_1^2)^2 + (1 - y_1)^2, \quad (5.27)$$

$$\text{przy ograniczeniach: } \mathbb{P}[g_1(\mathbf{X}) \leq 0] \leq 0.01, \quad (5.28)$$

$$\mathbb{P}[g_2(\mathbf{X}) \leq 0] \leq 0.01, \quad (5.29)$$

$$\mathbb{P}[g_3(\mathbf{X}) \leq 0] \leq 0.01, \quad (5.30)$$

$$0 \leq y_1 \leq 3, \quad (5.31)$$

$$0 \leq y_2 \leq 3, \quad (5.32)$$

gdzie

$$g_1(\mathbf{X}) = X_7(X_3X_1 + X_2 + X_4), \quad (5.33)$$

$$g_2(\mathbf{X}) = X_7(X_5X_1 - X_2 + X_6), \quad (5.34)$$

$$g_3(\mathbf{X}) = X_7(-2X_1 - X_2 + 7), \quad (5.35)$$

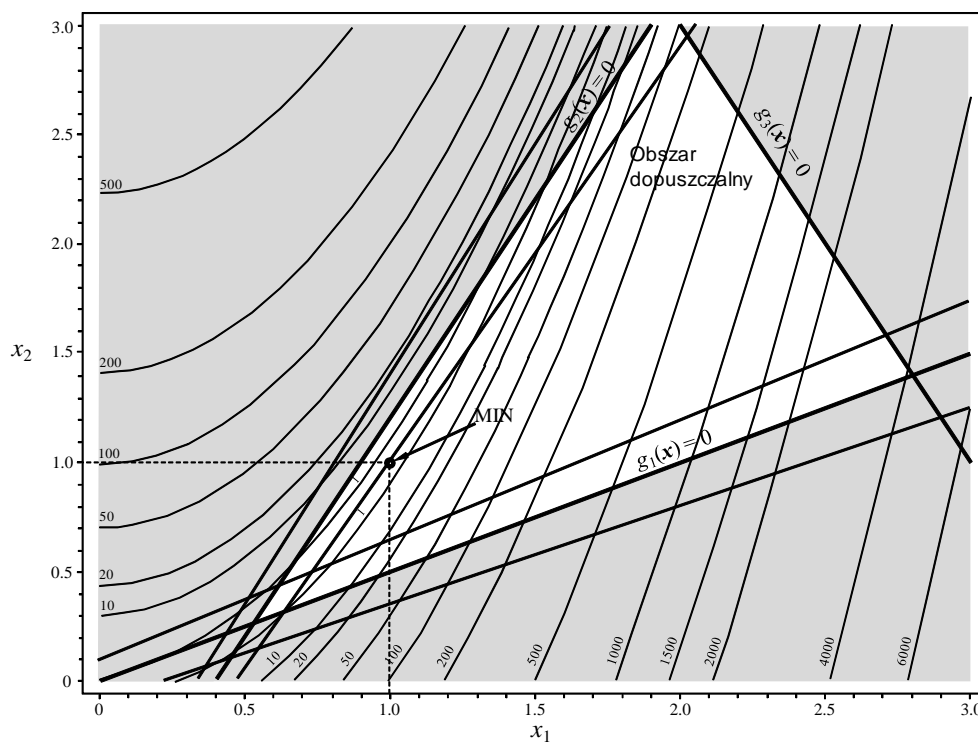
a X_1, \dots, X_7 są zmiennymi losowymi o rozkładach prawdopodobieństwa opisanych w tab. 5.2.

Poszukuje się zatem wartości oczekiwanych zmiennych X_1 i X_2 w zakresie $[0, 3]$, które minimalizowałyby wartość funkcji Rosenbrocka spełniając jednocześnie ograniczenia niezawodnościowe (5.28)–(5.30). Warstwice funkcji celu oraz obszar dopuszczalny ograniczony prostymi: $g_1(\mathbf{x}) = 0$, $g_2(\mathbf{x}) = 0$ i $g_3(\mathbf{x}) = 0$ pokazane są na rys. 5.4. Zmienne X_3, \dots, X_6 są losowymi parametrami występującymi we wzorach dwóch pierwszych funkcji ograniczeń, co prowadzi do niepewności definicji obszaru dopuszczalnego. Losowy parametr X_7 można interpretować jako dodatkowy szum nałożony na obliczone wartości ograniczeń. Analizując rys. 5.4 zauważymy, że punkt o współrzędnych $(1, 1)$, będący rozwiązaniem deterministycznego problemu minimalizacji, jest położony w bliskim sąsiedztwie granicy obszaru dopuszczalnego. Ze względu na losowość granic tego obszaru jest bardzo prawdopodobne, że dla wielu realizacji wektora \mathbf{X} tak

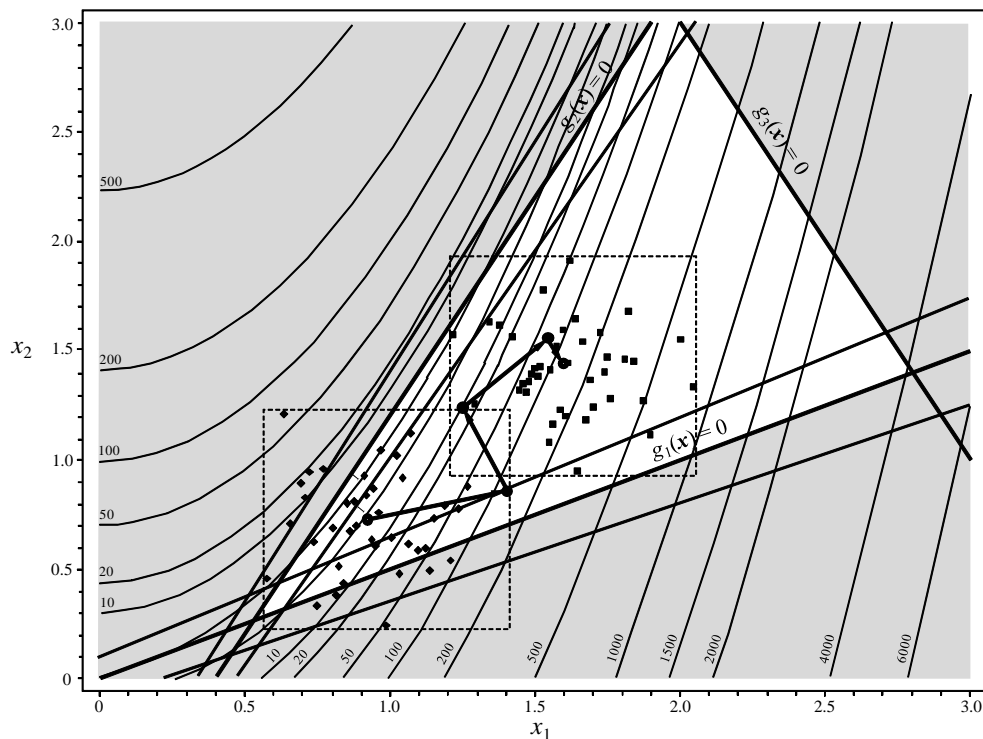
wybrany punkt optymalny znajdzie się poza obszarem dopuszczalnym. Rozwiązanie deterministyczne nie jest zatem rozwiązaniem bezpiecznym.

Tablica 5.2. Zmienne losowe w zadaniu minimalizacji z ograniczeniami funkcji Rosenbrocka.

Zmienna	Rozkład prawdopodobieństwa	Wartość oczekiwana	Odchylenie standardowe
X_1	normalny	y_1	0.2
X_2	normalny	y_2	0.2
X_3	jednostajny	-0.5	0.029
X_4	jednostajny	0.0	0.058
X_5	jednostajny	2.0	0.058
X_6	jednostajny	-0.8	0.058
X_7	normalny	1.0	0.02



Rys. 5.4. Warstwy funkcji Rosenbrocka oraz funkcje ograniczenia zadania (5.26)–(5.32). Losowy charakter współczynników w definicjach funkcji $g_1(\mathbf{X})$ i $g_2(\mathbf{X})$ prowadzi do niepewności położenia i wielkości obszaru dopuszczalnego.



Rys. 5.5. Historia procesu iteracyjnego w przypadku $N = 40$. Zaznaczono rozrzut wynikający z losowego charakteru zmiennych X_1 i X_2 wokół rozwiązania początkowego oraz końcowego.

Historia procesu iteracyjnego pokazana została na rys. 5.5. Rozwiązanie początkowe wybrano z 60 punktowej próbki wygenerowanej przy pomocy zmodyfikowanego planu OLH zakładając rozkład jednostajny zmiennych X_1 i X_2 w przedziale $[0, 3]$ (po rzutowaniu na osie odległość między najbliższymi punktami wynosiła 0.05). W następnych iteracjach algorytmu używano planów OLH z $N = 40$ punktami w celu oszacowania wartości ograniczeń niezawodnościowych oraz obliczenia współczynników korelacji pomiędzy funkcjami g_1 , g_2 i g_3 a zmiennymi losowymi X_1 i X_2 . Widoczne na rys. 5.5 rozrzuty punktów wokół rozwiązania startowego jak również rozwiązania optymalnego/poprawionego wynikające z losowego charakteru zmiennych X_1 i X_2 ilustrują niski poziom niezawodności rozwiązania wyjściowego. Jednak z powodów omówionych w podrozdziale 5.2.1 optymalne rozwiązanie uzyskane przy pomocy zaprezentowanej metody w sposób istotny zależy od wielkości próbki N , której używano do obliczenia prawdopodobieństwa awarii. Ponieważ, dopuszczalną wielkość tego prawdopodobieństwa p^a w równaniach (5.28)–(5.30) przyjęto równą 0.01, to w przypadku niewielkich próbek (tu $N = 40$) można jedynie twierdzić, że otrzymano

rozwiązanie, które jest bardziej niezawodne od początkowego. Nie ma jednak gwarancji zapewnienia założonego poziomu niezawodności.

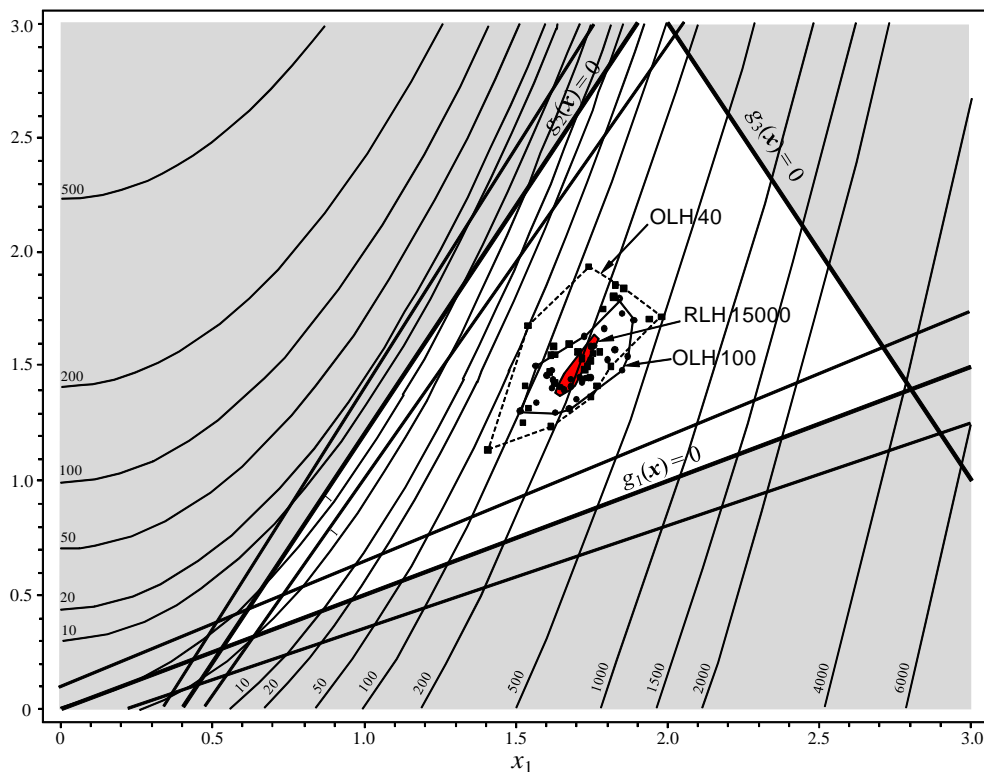
Duży wpływ na wynik działania algorytmu ma wybór punktu początkowego, który z kolei zależy od początkowej wielkości zmodyfikowanego planu OLH. W celu zbadania rozrzutu otrzymywanych rozwiązań proces iteracyjny powtórzono 30 razy, na początku dla $N = 40$, a następnie $N = 100$. W pierwszym przypadku do wyboru punktów startowych używano zmodyfikowanych planów OLH zawierających 60 punktów, a w drugim przypadku 100 punktów. Zdecydowano, że algorytm zostaje przerwany po znalezieniu pierwszego rozwiązania, które spełnia ograniczenia niezawodnościowe. W celu przyspieszenia obliczeń współczynnik wagowy w_f w kryterium (5.23) przyjęto równy zero (taką strategię działania metody omówiono na stronie 156).

Na podstawie wyników z przeprowadzonych 30 powtórzeń działania algorytmu, w tabeli 5.3 przedstawiono wartości średnie oraz odchylenia standardowe rozwiązań zadania (5.26)–(5.32). W kolumnach podano momenty statystyczne współrzędnych punktu optymalnego oraz wartości funkcji Rosenbrocka w tym punkcie, otrzymane przy użyciu różnych metod generowania próbek. Wyniki dla próbek OLH40 ($N = 40$) i OLH100 ($N = 100$) porównano z wynikami otrzymanymi dla próbek generowanych przy pomocy kwadratów łańciskich bez optymalizacji, oznaczonych skrótem RLH (zob. strona 84). Liczące 15000 realizacji plany RLH prowadzą do bardzo niewielkiego rozrzutu wyników i dlatego mogą być traktowane jako rozwiązanie odniesienia. Porównując wartości momentów zamieszczone w tab. 5.3 można zauważyć, że o ile odchylenia standardowe y_1 , y_2 i f są zdecydowanie mniejsze w przypadku OLH100 niż OLH40, o tyle odpowiednie wartości średnie nie różnią się w sposób znaczący.

Wyniki różnych wariantów realizacji omawianego zadania przedstawiono również na rys. 5.6. Oprócz punktów optymalnych znalezionych powtarzając wielokrotnie strategię OLH100 i OLH40 zaznaczono także obwiednie tych rozwiązań. Kolorem czerwonym oznaczono obszar rozwiązań RLH15000.

Tablica 5.3. Oszacowania wartości średnich i odchyłeń standardowych współrzędnych punktu optymalnego oraz wartości funkcji celu w tym punkcie otrzymane na podstawie 30 powtórzeń działania algorytmu.

	\bar{y}_1	$\sigma[y_1]$	\bar{y}_2	$\sigma[y_2]$	\bar{f}	$\sigma[f]$
OLH 40	1.71	0.13	1.50	0.17	213.6	104.2
OLH 100	1.69	0.10	1.47	0.12	200.2	73.2
RLH 15000	1.68	0.03	1.47	0.07	186.6	14.7



Rys. 5.6. Rozrzut (obwiednie) rozwiązań zadania minimalizacji funkcji Rosenbrocka otrzymanych przy pomocy różnych planów losowania punktów eksperymentalnych.

PODSUMOWANIE

W analizie niezawodności złożonych układów konstrukcyjnych, nieliniowy charakter funkcji granicznych wynika nie tylko z natury modelowanego zjawiska fizycznego, ale jest również często wywołany szumem numerycznym spowodowanym przez stosowaną metodę rozwiązania zadania MES. Szum ten uniemożliwia bezpośrednie zastosowanie do oszacowania prawdopodobieństwa awarii klasycznych niesymulacyjnych metod analizy niezawodności, które bazują na gradientowych algorytmach poszukiwania punktu projektowego.

W niniejszym rozdziale zaproponowano metodę, która dzięki zastosowaniu aproksymacji funkcji granicznej oraz adaptacyjnej techniki symulacji losowych jest niewrażliwa na wpływ szumu numerycznego, a jednocześnie jest dużo bardziej efektywna od typowych metod Monte Carlo. Zmodyfikowany przez autora dwuetapowy algorytm Z3M [278] pozwala na kontrolę dokładności, a w

konsekwencji, kosztu obliczeń. Jeśli nie wymaga się wysokiej precyzji obliczanej wartości prawdopodobieństwa awarii, albo nie można sobie na to pozwolić ze względu na ograniczony budżet analizy niezawodności, wtedy drugi, symulacyjny etap algorytmu może zostać pominięty. Wspomniane modyfikacje dotyczą zastosowania na etapie poszukiwania punktu projektowego aproksymacji funkcji granicznej przy pomocy ważonej regresji liniowej oraz generowania eksperymentów zgodnie z planem OLH. Pozwalają one na szybkie określenie sąsiedztwa punktu projektowego, co poprawia zasadniczo zbieżność algorytmu. Działanie metody zilustrowano na przykładzie zadania analizy niezawodności tarczy o losowej geometrii. Będzie on także zastosowany w podrozdziale 6.1 w zadaniu dotyczącym niezawodności zgniatanej dynamicznie belki cienkościennej.

W rozdziale zaprezentowano także nową metodę szybkiej poprawy niezawodności konstrukcji. Działanie algorytmu wykorzystuje analizę korelacji, opierając się jedynie na symulacjach OLH, bez użycia metod powierzchni odpowiedzi. Plan eksperymentów OLH służy zarówno równomiernemu wypełnieniu analizowanego obszaru przez eksperymenty numeryczne, jak również tworzy próbkę, na podstawie której obliczane są współczynniki korelacji projektowych zmiennych losowych i funkcji ograniczeń. Zaprezentowany algorytm jest skuteczny szczególnie wtedy, gdy poziom niezawodności konstrukcji wyjściowej jest niski. Może też stanowić szybką procedurę startową dla algorytmów optymalizacji odpornościowej lub niezawodnościowej.

Analiza niezawodności i optymalizacja odpornościowa w zagadnieniach wytrzymałości zderzeniowej

Jak już wspomniano we wstępie do rozdziału 5, analiza numeryczna MES zagadnień dotyczących dynamicznych deformacji elementów konstrukcji poddanych obciążeniom zderzeniowym wymaga zazwyczaj długotrwałych obliczeń, prowadząc do silnie nieliniowych funkcji odpowiedzi. Funkcje te zawierają często składnik pochodzący od szumu numerycznego, którego wyeliminowanie jest praktycznie niemożliwe.

Ze względu na wymienione trudności, dość powszechną praktyką w analizie niezawodności czy też optymalizacji konstrukcji absorbujących energię zderzeń, jest stosowanie analitycznych metamodeli (zob. rozdział 3) do aproksymacji wyników analizy MES.

W niniejszym rozdziale przedstawionych będzie szereg strategii rozwiązywania problemów analizy niezawodności, jak również optymalizacji odpornościowej elementów konstrukcji ze względu na ich wytrzymałość zderzeniową. Zastosowane metody wykorzystują zarówno rozmaite koncepcje budowania analitycznych powierzchni odpowiedzi, jak też rozwiązania gdzie takich aproksymacji się nie stosuje.

W podrozdziale 6.1 przedstawiono przykład zastosowania omówionego w 5.1 algorytmu Z3M do obliczania prawdopodobieństwa awarii zgniatanej dynamicznie belki cienkościennej (tzw. podłużnicy), por. [238]. W przykładzie zaproponowano sposób uwzględnienia losowych wad połączeń zgrzewanych. Model tej samej belki użyto następnie w podrozdziale 6.2 do przetestowania zaprezentowanej w 5.2 metody poprawy niezawodności, por. [231]. W 6.3 omówiono pewną przybliżoną metodę analizy jakościowej zachowania zgniatanego elementu tylnej ramy pojazdu, wykorzystującą prosty algorytm klasyfikacji wyników

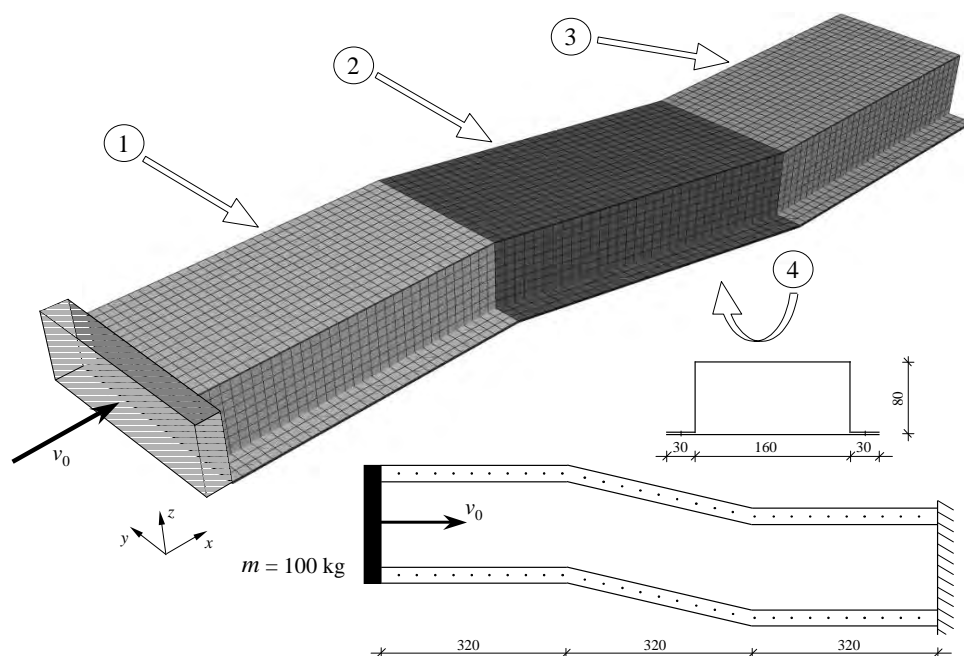
eksperymentów numerycznych generowanych przez plany OLH. Przykład ten zaczerpnięto z pracy [21]. W podrozdziale 6.4 model cienkościennej podłużnicy zastosowano w zadaniu optymalizacji odpornościowej, por. [38]. Poszukiwano takich grubości blach, z których wykonane są segmenty belki, aby nie zwiększając znacząco ciężaru elementu zmaksymalizować wielkość zaabsorbowanej energii zderzenia minimalizując jednocześnie jej rozrzut wywołany wpływem parametrów losowych. W podrozdziale 6.5 przedstawiono przykład użycia techniki optymalizacji odpornościowej do kalibracji parametrów modelu numerycznego. Model ten stosowany był w komputerowej symulacji uderzenia fantomu dolnej kończyny (ang. *legform impactor*) w barierę imitującą przednią część pojazdu osobowego, por. [47]. Proces optymalizacji ma tu na celu takie dobranie wartości oczekiwanych wybranych parametrów materiałowych modelu, aby rozrzuty wyników losowych symulacji komputerowych jak najlepiej odpowiadały obserwowanym rozrzutom wyników testów laboratoryjnych.

6.1. Analiza niezawodności zgniatanej dynamicznie belki cienkościennej z uwzględnieniem losowych wad połączeń zgrzewanych

Ze względu na niedoskonałości procesu produkcyjnego, wady zmęczeniowe, czy też na skutek drobnych wypadków, obserwuje się, że trudna do pominięcia liczba połączeń zgrzewanych występujących w konstrukcjach użytkowanych pojazdów nie funkcjonuje prawidłowo, zob. [57, 164]. Chęć uwzględnienia tego zjawiska w analizie niezawodności wiąże się jednak z istotnymi problemami. Pomijając trudności z uzyskaniem odpowiednich danych statystycznych, modelowanie parametrów materiałowych odpowiadających poszczególnym zgrzewom (elementom skończonym je reprezentującym) przy pomocy zmiennych losowych jest podejściem bardzo nieefektywnym. Rozwiązaniem pośrednim, które pozwala na rozpatrywanie niedoskonałości połączeń, a jednocześnie nie podnosi znacznie kosztu obliczeń, wydaje się być usuwanie pewnej liczby losowo wybranych zgrzewów za każdym razem, gdy w algorytmie analizy niezawodności obliczana jest wartość funkcji granicznej. Strategia ta wymaga jednak zastosowania takiej metody szacowania prawdopodobieństwa awarii, która nie jest wrażliwa na szum generowany przez losową eliminację połączeń. Przedstawiona w podrozdziale 5.1 metoda Z3M wydaje się spełniać to wymaganie.

Pomimo ciągle zwiększającego się zainteresowania stochastyczną analizą zagadnień związanych z wytrzymałością zderzeniową, a w szczególności analizą

niezawodności konstrukcji absorbujących energię zderzeń, nie ma zbyt wiele prac, w których proponuje się algorytmy nakierowane na specyfikę tych problemów. W większości znanych autorowi publikacji analiza niezawodności wytrzymałości zderzeniowej rozpatrywana jest jedynie w kontekście optymalizacji niezawodnościowej konstrukcji samochodowych, zob. np. [137,262,263,267,270]. W pracach tych dąży się przede wszystkim do poprawy efektywności procesu optymalizacji niezawodnościowej. Cel ten osiągany jest najczęściej poprzez stosowanie jak najprostszych metod szacowania niezawodności konstrukcji, które prowadzą jedynie do bardzo zgrubnej oceny wartości prawdopodobieństwa awarii. Wybór algorytmu Z3M do realizacji poniższego zadania ma w założeniu łączyć możliwy do zaakceptowania koszt obliczeń z dobrą dokładnością oszacowania niezawodności.



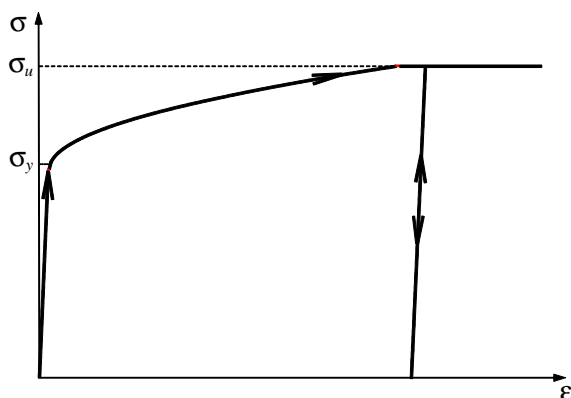
Rys. 6.1. Schemat cienkościennej podłużnicy oraz siatka MES. Wymiary podano w milimetrach. Strzałkami oznaczono numery poszczególnych części belki.

Rozważmy cienkościenną belkę stalową, tzw. podłużnicę, pokazaną na rys. 6.1. Belka utwierdzona jest na jednym końcu, a drugi koniec zostaje uderzony sztywną płytą o masie 100 kg poruszającą się z prędkością początkową $v_0 = 15$ m/s (54 km/h) w kierunku osi poziomej x . Belka składa z trzech części o przekroju w

kształcie litery Ω oraz płyty zamykającej. Części 1–3 połączone są ze sobą spawem ciągłym wzdłuż odpowiednich krawędzi, płyta 4 zamocowana jest do nich przy pomocy 64 punktowych zgrzewów elektrycznych. Jako model materiału przyjęto sprężysto-lepko-plastyczny model Johnsona-Cooka, por. [111],

$$\sigma = (\sigma_y + a_1 \varepsilon_p^n) \left(1 + a_2 \ln \frac{\dot{\varepsilon}}{\dot{\varepsilon}_0} \right), \quad (6.1)$$

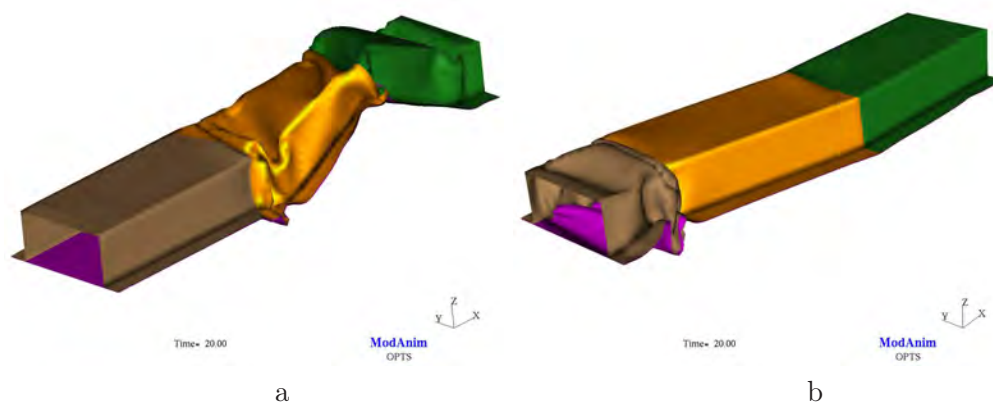
gdzie σ_y jest granicą plastyczności, a_1 , a_2 , n są wyznaczanymi na podstawie doświadczenia parametrami modelu, ε_p jest odkształceniem plastycznym, a $\dot{\varepsilon}_0$ jest wartością odniesienia prędkości odkształcenia. Zależność (6.1) pokazano na wykresie 6.2.



Rys. 6.2. Sprężysto-lepko-plastyczny model materiału Johnsona-Cooka (6.1). Przez σ_u oznaczono naprężenie maksymalne.

Model MES składa się z 5760 czterowęzłowych elementów powłokowych MITC4, zob. [61], oraz 64 specjalnych jednowymiarowych elementów typu belkowego do modelowania punktowych połączeń zgrzewanych. Analiza skończone elementowa przeprowadzana była przy użyciu komercyjnego programu RADIOSS [171], bazującego na jawnym całkowaniu równań ruchu i przeznaczonego do analizy silnie nieliniowych dynamicznych zagadnień mechaniki, w szczególności analizy zderzeń.

Ponieważ belka ma jak najlepiej zaabsorbować energię kinetyczną uderzającej w nią masy, dlatego w projektowaniu dąży się do tego aby podłużnica nie deformowała się tak jak to pokazano na rys. 6.3a (zginając się w wyniku globalnego wyboczenia), a raczej deformowała się poprzez regularne fałdowanie na skutek lokalnej utraty stateczności ścianek, rys. 6.3b. Niestety bywa tak, że



Rys. 6.3. Z lewej: globalna utrata stateczności, mała ilość pochłoniętej energii; z prawej: regularne fałdowanie, dobrze absorbowana energia zderzenia.

konstrukcja, której praca w stanie nominalnym spełnia założenia projektantów, nie jest niezawodna ze względu na wysoką wrażliwość odpowiedzi na nieuniknione losowe imperfekcje parametrów. W rzeczywistości trudno jest zagwarantować, że kierunek w jakim porusza się masa nie będzie odbiegał od założonego, że wszystkie połączenia zgrzewane są poprawnie wykonane, a grubości blach, z których wykonane są poszczególne części belki, nie różnią się od grubości nominalnych. Z tych względów wydaje się być bardzo ważnym aby zbadać wrażliwość odpowiedzi konstrukcji na losowy charakter jej parametrów oraz określić prawdopodobieństwo, iż nie zostanie zaabsorbowana wymagana ilość energii.

Przyjęto model probabilistyczny składający się z 8 zmiennych losowych. Ich opis oraz parametry rozkładów prawdopodobieństwa podano w tab. 6.1. Odpowiadają one grubościom blach poszczególnych elementów, parametrom materiałowym oraz warunkom początkowym zadania. Jak już o tym wcześniej wspomniano, w celu uwzględnienia stochastycznej natury wytrzymałości zgrzewów, 3 losowo wybrane elementy spośród 64 elementów reprezentujących zgrzewy ($\approx 5\%$) usuwane są z modelu za każdym razem, gdy przy pomocy analizy MES obliczana jest wartość funkcji granicznej dla danej realizacji \mathbf{x} wektora \mathbf{X} . W pierwszym etapie algorytmu Z3M strategię taką uznać można za sposób uwzględnienia uśrednionego wpływu awarii połączeń zgrzewanych.

Długość trwania zderzenia wynosi 20 ms. W zadaniu analizy niezawodności jako niepożądane zachowanie konstrukcji (awarię) zdefiniowano niedostateczną absorpcję energii uderzającej masy przyjmując $e_{\min} = 6000 \text{ J}$ jako minimalną wartość dopuszczalną. Wielkość ta stanowi około 80% energii absorbowana-

Tablica 6.1. Zmienne losowe w zadaniu analizy niezawodności zgniatanej podłużnicy.

Zmienna	Opis	Rozkład prawdopod.	Wartość oczekiwana	Odchylenie standardowe
X_1	t_1 - grubość części 1	lognormalny	1.5 [mm]	0.075 [mm]
X_2	t_2 - grubość części 2	lognormalny	1.5 [mm]	0.075 [mm]
X_3	t_3 - grubość części 3	lognormalny	1.5 [mm]	0.075 [mm]
X_4	t_4 - grubość części 4	lognormalny	1.0 [mm]	0.05 [mm]
X_5	σ_0 - granica plastyczności	lognormalny	180 [MPa]	15 [MPa]
X_6	E - moduł Younga	lognormalny	210000 [MPa]	21000 [MPa]
X_7	v_0^y - składowa y wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.5 [m/s]
X_8	v_0^z - składowa z wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.5 [m/s]

nej przez belkę nominalną (odpowiadającą wartościom oczekiwany zmiennych losowych oraz brakowi wadliwych zgrzewów). Postać funkcji granicznej przedstawia się zatem następująco:

$$g(\mathbf{X}, \mathbf{A}) = 1 - \frac{e_{\min}}{e(\mathbf{X}, \mathbf{A})}, \quad (6.2)$$

gdzie $e(\mathbf{X}, \mathbf{A})$ jest energią zaabsorbowaną przez belkę, a \mathbf{A} jest wektorem dyskretnych zmiennych losowych o rozkładzie dwupunktowym modelujących stan (sprawny/wadliwy) połączeń zgrzewanych. Zgodnie z poczynionym założeniem, każda realizacja wektora \mathbf{A} prowadzi do 3 wadliwych połączeń i eliminowania z modelu MES odpowiednich elementów skończonych. Zaobserwowano, że rozrzut wartości funkcji (6.2) wywołany usuwaniem zgrzewów, mierzony jako maksymalne odchylenie od wartości średniej, może wynosić do 0.07. Ponieważ nie jest znana funkcja łącznej gęstości prawdopodobieństwa zmiennych \mathbf{A} , nie jest możliwe aby włączyć je do wektora zmiennych podstawowych \mathbf{X} , a ich obecność w pierwszym etapie Z3M manifestować się będzie jedynie jako szum nałożony na wartości funkcji granicznej.

Zastosowany algorytm realizowany jest zgodnie z opisem przedstawionym w podrozdziale 5.1. Zawiera on jednak pewne usprawnienia mające na celu podniesienie efektywności rozwiązania, które odróżniają go od oryginalnej wersji z pracy [278].

Pierwsza część algorytmu to poszukiwanie punktu projektowego wykorzystujące technikę adaptacyjnej aproksymacji funkcji granicznej. Trzeba tutaj pamiętać, że ze względu na wpływ szumu na obliczane wartości funkcji g , celem tego etapu nie jest "dokładna" lokalizacja punktu projektowego, a jedynie znalezienie

się w jego pobliżu. Nie jest zatem wskazane używanie zbyt rygorystycznych warunków zbieżności ponieważ z dużym prawdopodobieństwem warunki te nie zostaną spełnione. Ponadto, zastosowane kryterium zbieżności (por. opis w punkcie 6 na stronie 142) powinno uwzględniać wymiar zadania n . Kryterium postaci $\delta < \epsilon$, gdzie $\epsilon = 0.1$, a δ jest odległością między dwoma ostatnimi przybliżeniami położenia punktu \mathbf{u}^* (dwoma położeniami środka obszaru dopuszczalnego), może być zbyt restrykcyjne w przypadku problemów z wieloma zmiennymi losowymi. Zmodyfikowane kryterium zapisać można jako $\delta < \kappa\sqrt{n\epsilon^2}$, w którym prawa strona równa się κ krotnej długości przekątnej n -wymiarowej hiperkostki o boku równym ϵ .

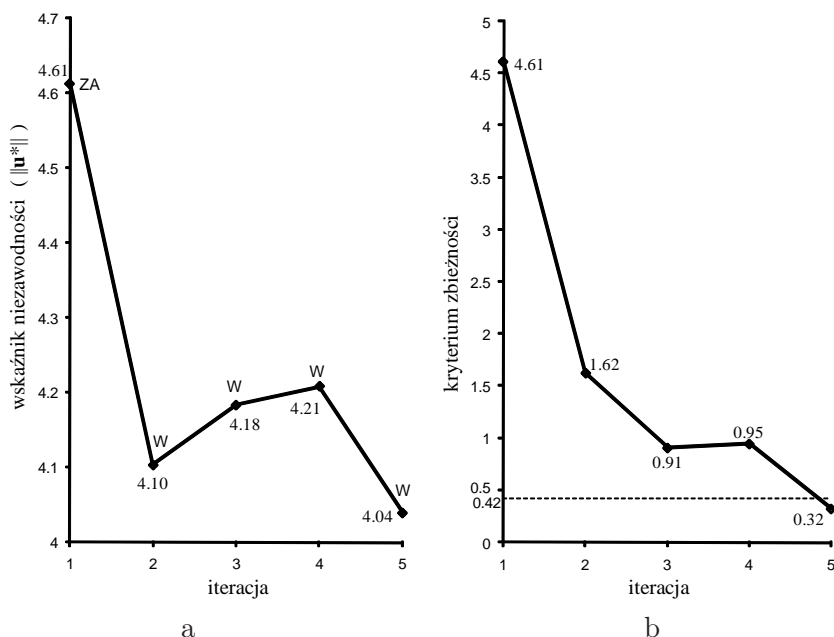
W bieżącym przykładzie dla $n = 8$ przyjęto $\epsilon = 0.15$ i $\kappa = 1$, co prowadzi do warunku $\delta < 0.42$. Drugi warunek zbieżności dotyczy wartości średniej funkcji granicznej w punkcie projektowym i zdefiniowany jest jako $\bar{h}(\mathbf{u}^*) < 0.1$. Na podstawie wcześniejszych doświadczeń ograniczono również liczbę iteracji pierwszego etapu Z3M do 15.

Inna modyfikacja w stosunku do oryginalnego algorytmu polega na odmiennej strategii redukcji wielkości obszaru dopuszczalnego. Chociaż w [278] zaproponowano, aby w każdej iteracji zmniejszać o połowę długość boku hiperkostki tworzącej obszar dopuszczalny, to jednak w przypadku problemów wielowymiarowych podejście takie wydaje się być niekorzystne ze względu na zbyt szybkie zawężanie pola poszukiwań, co prowadzić może do niepożądanego działania algorytmu. Istnieje wysokie prawdopodobieństwo, że aproksymacja położenia punktu projektowego otrzymana z wykorzystaniem powierzchni odpowiedzi zbudowanych w oparciu o zbyt "zlokalizowane" oraz obciążone szumem dane, znajdować się będzie poza obszarem dopuszczalnym. W zaimplementowanej wersji metody Z3M zamiast boku dopuszczalnej n -wymiarowej hiperkostki plan redukcji dotyczy jej objętości. Założono, że w przypadku gdy bieżące przybliżenie punktu projektowego znajduje się wewnątrz obszaru dopuszczalnego, to objętość nowego obszaru zmniejsza się do 25% objętości aktualnej, co zapisać można jako

$$b_i = 0.25^{\frac{1}{n}} b_{i-1}, \quad (6.3)$$

gdzie b_i i b_{i-1} są odpowiednio bieżącą oraz poprzednią długością boku. W przypadku 8 zmiennych losowych oraz początkowej długości boku $b_0 = 6$ wynikiem tej strategii jest następujący ciąg długości 6, 5.05, 4.24, 3.57, 3, 2.52, ..., natomiast zgodnie ze strategią połowienia dostaniemy 6, 3, 1.5, 0.75, 0.325, 0.1625, ...

Do aproksymacji funkcji granicznej przyjęto liniową powierzchnię odpowiedzi oraz metodę regresji ważonej, zob. (5.1)–(5.3). Eksperymenty do budowy powierzchni odpowiedzi generowane były przez plan OLH o $N = 4n = 32$ punktach.



Rys. 6.4. Metoda Z3M, etap lokalizacji punktu projektowego: a) historia zmian oszacowania wskaźnika niezawodności β_{FORM} , b) historia zmian kryterium zbieżności (odległość między kolejnymi iteracjami punktu projektowego)

Wyniki pierwszego etapu algorytmu Z3M pokazano na rys. 6.4. Po lewej zaprezentowano historię zmian wartości wskaźnika niezawodności pierwszego rzędu β_{FORM} , które odpowiadają poszczególnym przybliżonym położeniom punktu projektowego. Oznaczenia “ZA” i “W” przy punktach wykresu informują czy punkt projektowy znaleziony był odpowiednio na zewnątrz czy wewnątrz obszaru dopuszczalnego. Wykres po prawej stronie przedstawia zmiany wartości warunku zbieżności rozumianego jako odległość pomiędzy kolejnymi iteracjami położenia punktu projektowego. Kryterium to zostało spełnione po 5 iteracjach. Otrzymano następujące oszacowania:

$$\beta^I = \beta_{\text{FORM}} = 4.04, \quad P_f^I = 2.7 \cdot 10^{-5}, \quad (6.4)$$

$$\mathbf{u}^* = \{-0.89, -0.29, -1.20, -1.26, -2.19, -0.80, 1.92, -1.81\}, \quad (6.5)$$

$$\bar{h}(\mathbf{u}^*) = 0.0038,$$

liczba obliczeń funkcji granicznej = 213,

gdzie indeks I oznacza pierwszą część algorytmu Z3M. Analizując współrzędne punktu \mathbf{u}^* można zauważyć, że najbardziej prawdopodobna realizacja awarii belki odpowiada zdarzeniu gdy grubości części 3 i 4 jak również granica plastyczności materiału są istotnie poniżej swych wartości oczekiwanych oraz gdy

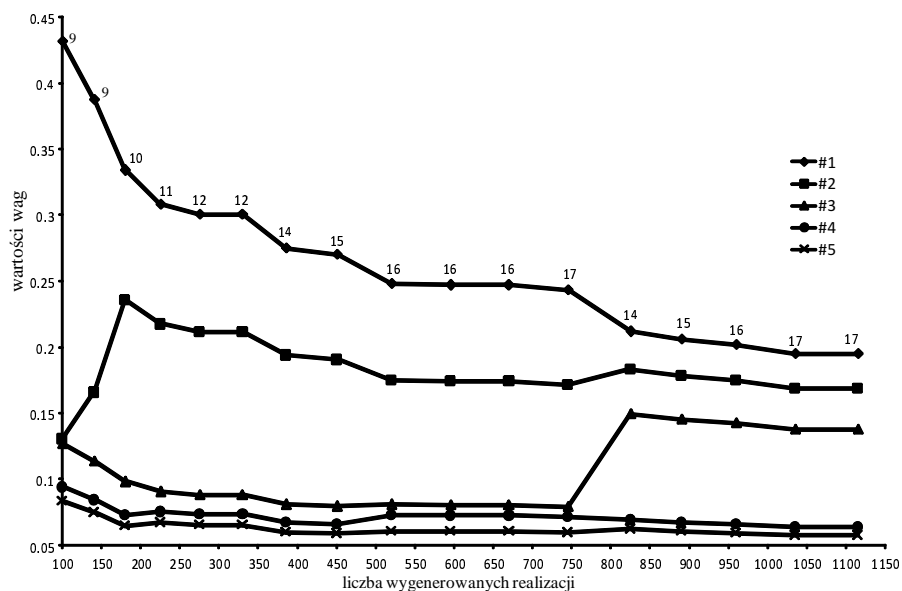
istnieją znaczące składowe poprzeczne wektora prędkości uderzającej masy, który nominalnie pokrywa się z osią x . Trzeba podkreślić, że punkt projektowy odnosi się do najbardziej prawdopodobnej realizacji wektora \mathbf{X} , która prowadzi do awarii, jak również do pewnego uśrednionego wpływu wad połączeń zgrzewanych (zmiennych \mathbf{A}). Nie można zatem traktować punktu \mathbf{u}^* jako rzutu na n -wymiarową przestrzeń zmiennych podstawowych hipotetycznego punktu projektowego, który znaleziony byłby znając funkcję gęstości prawdopodobieństwa zmiennych \mathbf{A} i rozpatrując zadanie w przestrzeni 72-wymiarowej ($72 = 8 + 64$, $n = 8$, 64 jest liczbą zgrzewów).

Pamiętając, że do budowy każdej aproksymacji funkcji granicznej używano 32 punktów eksperymentalnych oraz uwzględniając obliczenia niezbędne do weryfikacji wartości funkcji w poszczególnych iteracjach punktu projektowego, całkowita liczba wywołań programu RADIOSS wynosiła 213.

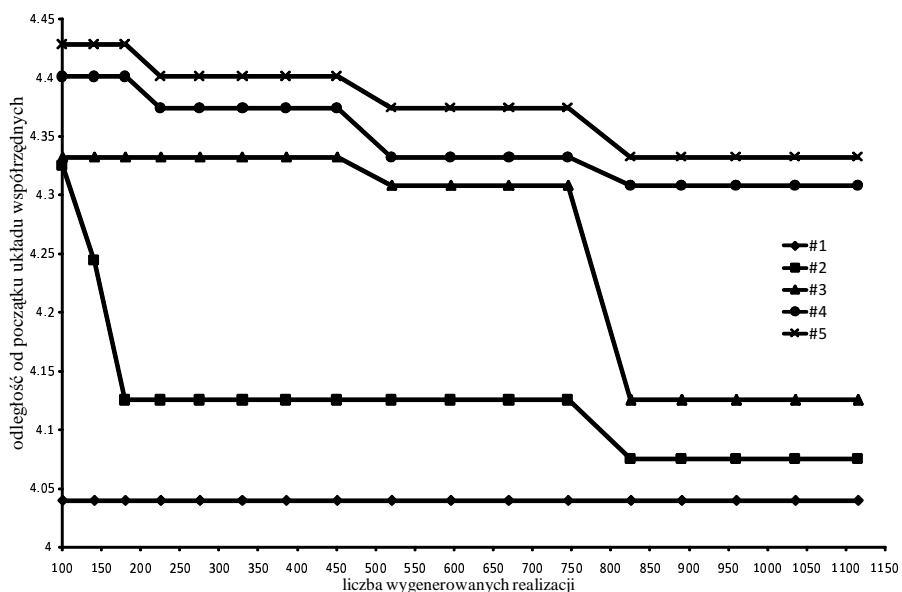
Na początku drugiego etapu algorytmu (adaptacyjnej metody importance sampling) wygenerowano $N_1 = 80$ realizacji zmiennych losowych zgodnie z planem OLH przy użyciu funkcji gęstości prawdopodobieństwa $\varphi_n(\mathbf{u}, \mathbf{u}^*, \mathbf{I})$. Następnie, tak jak to opisano w punktach 1–5 w podrozdziale 5.1.2, zbudowano wielomodalną funkcję gęstości losującej. Jedyną modyfikacją dotyczyła stosowanej wielkości próbki. Zdecydowano, że będzie ona wynosić $5\tilde{K}$, gdzie \tilde{K} jest liczbą “istotnych” punktów reprezentatywnych. Za takie uznaje się minimalną liczbę \tilde{K} pierwszych punktów reprezentatywnych z najwyższymi wagami, których suma przekroczy 0.75. Wartość dopuszczalną współczynnika zmienności estymatora prawdopodobieństwa awarii $\nu_{\hat{P}_f}$ przyjęto równą 0.2. Ostateczną postać wielomodalnej gęstości losującej otrzymano po 18 krokach iteracyjnych (1195 obliczeniach wartości funkcji granicznej). Jest to funkcja typu (5.11) z 17 punktami reprezentatywnymi o wagach odpowiednio równych:

$$\hat{\mathbf{w}} = \{ 0.195, 0.169, 0.138, 0.064, 0.057, 0.051, 0.045, 0.043, 0.035, \\ 0.035, 0.034, 0.031, 0.026, 0.024, 0.019, 0.019, 0.017 \}.$$

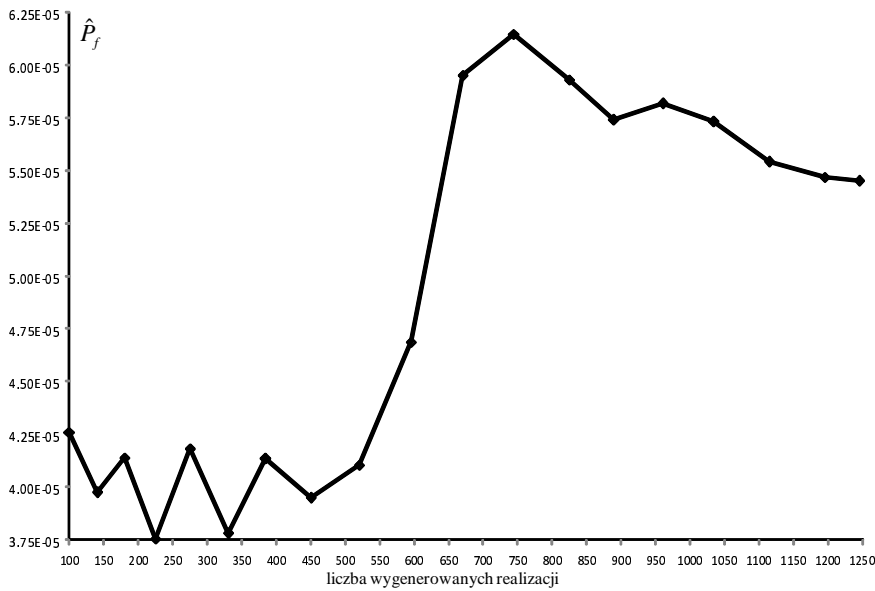
Spośród punktów reprezentatywnych trzy pierwsze są wyraźnie dominujące. Położone są w przybliżeniu w tej samej odległości od punktu $\mathbf{u} = \mathbf{0}$ co punkt projektowy znaleziony w pierwszym etapie Z3M. Na rysunkach 6.5 i 6.6 pokazano wykresy obrazujące, jak wraz z postępem procesu generowania nowych realizacji zmiennych losowych ewoluują wartości wag oraz odległości punktów reprezentatywnych od początku układu współrzędnych. Liczby podane przy punktach wykresu #1 na rys. 6.5 oznaczają całkowitą liczbę punktów reprezentatywnych.



Rys. 6.5. Metoda Z3M, etap importance sampling: Zmiany wartości wag pierwszych pięciu punktów reprezentatywnych. Przy punktach danych wykresu #1 podano całkowitą liczbę punktów reprezentatywnych.



Rys. 6.6. Metoda Z3M, etap importance sampling: Odległości od punktu $\mathbf{u} = \mathbf{0}$ pierwszych pięciu punktów reprezentatywnych.



Rys. 6.7. Metoda Z3M, etap importance sampling: Ewolucja wartości estymatora prawdopodobieństwa awarii.

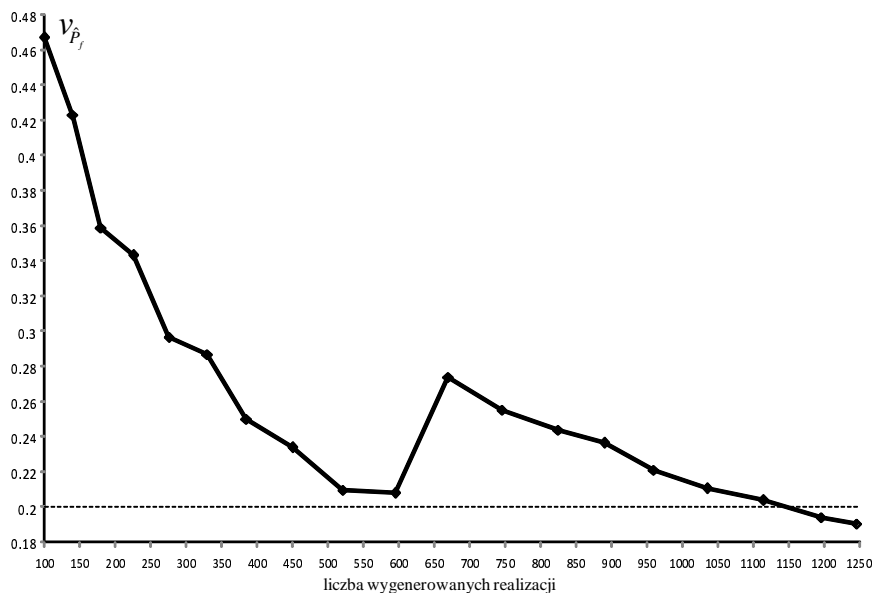
Porównując wykresy na rys. 6.5 i rys. 6.6 z ewolucją estymatora \hat{P}_f na rys. 6.7 można zauważyć znaczące zmiany wartości tego estymatora pomiędzy 500-tą a 900-tą realizacją zmiennych losowych, co niewątpliwie wynika z lepszego dopasowania gęstości losującej do kształtu obszaru awarii. Na rys. 6.8 przedstawiono historię zmian współczynnika zmienności $\nu_{\hat{P}_f}$. Trend ten nie jest monotoniczny. Zaobserwowany skok spowodowany może być rzadkim zdarzeniem polegającym na wygenerowaniu realizacji o wysokim stosunku $\varphi_n(\mathbf{v}_i, \mathbf{0}, \mathbf{I})/s_{\mathbf{V}}(\mathbf{v}_i)$, por. wzór (5.8). Ostateczne oszacowanie prawdopodobieństwa awarii oraz odpowiadający mu wskaźnik niezawodności (A.5) wynoszą:

$$P_f^{\text{II}} = 5.37 \cdot 10^{-5}, \quad \beta^{\text{II}} = -\Phi^{-1}(P_f^{\text{II}}) = 3.87, \quad (6.6)$$

liczba obliczeń funkcji granicznej = 1195,

gdzie indeks II oznacza drugi etap metody Z3M.

Porównując wartości P_f^{I} i P_f^{II} można dojść do wniosku, iż funkcja graniczna nie jest silnie "globalnie" nieliniowa, a zatem jest dyskusyjne, czy poprawka uzyskana dzięki adaptacyjnej metodzie importance sampling wnosi wiele w stosunku do aproksymacji FORM wykorzystującej punkt projektowy z pierwszego etapu. Pomimo, iż skorygowane przez symulacje losowe prawdopodobieństwo awarii jest dwukrotnie większe niż wartość P_f^{I} , to nie jest to jednak zmiana jakościowa i dlatego można kwestionować zasadność poniesienia znaczącego kosztu



Rys. 6.8. Metoda Z3M, etap importance sampling: Ewolucja współczynnika zmienności estymatora \hat{P}_f .

dotychczasowych obliczeń niezbędnych do otrzymania wyniku. Z drugiej strony, rozważania takie są możliwe dopiero po przeprowadzeniu drugiej fazy algorytmu Z3M, która stanowi bardzo dobrą weryfikację poprawności oszacowania FORM.

W celu oceny wpływu wad połączeń zgrzewanych na niezawodność zderzeniową analizowanej belki, powtórzono dwa etapy algorytmu Z3M zakładając tym razem idealny stan wszystkich zgrzewów. Otrzymano następujące rezultaty:

Pierwszy etap, liczba obliczeń funkcji granicznej = 199,

$$\beta^I = \beta_{\text{FORM}} = 4.02, \quad P_f^I = 2.9 \cdot 10^{-5}, \quad (6.7)$$

$$\mathbf{u}^* = \{-1.14, 0.15, -0.99, -1.84, -1.61, -1.30, 1.81, -1.72\}, \quad (6.8)$$

$$\bar{h}(\mathbf{u}^*) = -0.0016,$$

Drugi etap, liczba obliczeń funkcji granicznej = 660,

$$P_f^{II} = 1.9 \cdot 10^{-5}, \quad \beta^{II} = -\Phi^{-1}(P_f^{II}) = 4.12. \quad (6.9)$$

Z porównania prawdopodobieństw awarii odpowiadających przypadkom z uwzględnianiem oraz bez uwzględniania wad połączeń zgrzewanych (odpowiednio równania (6.6) i (6.9)) można stwierdzić, że jakość tych połączeń ma istotne znaczenie w ocenie niezawodności podłużnicy, jako elementu absorbującego energię zderzeń. Jest około 3 razy bardziej prawdopodobne, że belka

z wadliwymi zgrzewami nie zaabsorbuje w czasie zderzenia wymaganej energii niż belka o nieuszkodzonych połączeniach.

Punkty projektowe (6.8) i (6.5) są niewątpliwie różne, a jednak ich współrzędne noszą ze sobą jakościowo podobną informację. Odzwierciedlają one intuicyjny scenariusz awarii polegający na gorszych niż nominalne wartościach parametrów materiałowych, mniejszych niż zakładano grubościach blach oraz znacznych składowych poprzecznych wektora prędkości uderzającej masy.

6.2. Algorytm poprawy poziomu niezawodności w zastosowaniu do zagadnień wytrzymałości zderzeniowej

Omówiony wcześniej model podłużnicy (zob. rys. 6.1) wykorzystano do przetestowania metody poprawy niezawodności zaproponowanej w podrozdziale 5.2. Wybór zmiennych losowych odpowiada temu, który przedstawiono w tab. 6.1. Jednak tym razem wartości oczekiwane zmiennych X_1, \dots, X_4 (grubości blach, z których wykonano poszczególne części belki) traktowane są jako zmienne projektowe, a odpowiadające im odchylenia standardowe obliczane są przy zachowaniu stałego współczynnika zmienności, zob. tab. 6.2. Podobnie jak w poprzednim podrozdziale, wady połączeń zgrzewanych uwzględnia się eliminując za każdym razem z modelu MES 3 losowo wybrane elementy skończone zgrzewów.

Tablica 6.2. Zmienne losowe w zadaniu optymalizacji niezawodnościowej / poprawy niezawodności zgniatanej podłużnicy.

Zmienna	Opis	Rozkład prawdopodob.	Wartość oczekiwana	Odch. std./wsp. zmien.
X_1	t_1 - grubość części 1	lognormalny	y_1 [mm]	$\nu_{X_1} = 5\%$
X_2	t_2 - grubość części 2	lognormalny	y_2 [mm]	$\nu_{X_2} = 5\%$
X_3	t_3 - grubość części 3	lognormalny	y_3 [mm]	$\nu_{X_3} = 5\%$
X_4	t_4 - grubość części 4	lognormalny	y_4 [mm]	$\nu_{X_4} = 5\%$
X_5	σ_0 - granica plastyczności	lognormalny	180 [MPa]	15 [MPa]
X_6	E - moduł Younga	lognormalny	210000 [MPa]	21000 [MPa]
X_7	v_0^y - składowa y wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.5 [m/s]
X_8	v_0^z - składowa z wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.5 [m/s]

Problem optymalizacji sformułowany jest następująco:

znajdź wartości zmiennych:

$$\mathbf{y} = \{y_1 = \mu_{X_1}, y_2 = \mu_{X_2}, y_3 = \mu_{X_3}, y_4 = \mu_{X_4}\}, \quad (6.10)$$

minimalizujące:

$$\text{objętość materiału } V(\mathbf{y}), \quad (6.11)$$

przy ograniczeniach:

dla wszystkich realizacji zmiennych \mathbf{X} :

$$1) \text{ absorbowana energia jest większa niż } 7500 \text{ J}, \quad (6.12)$$

2) deformacja pierwszej części jest większa niż

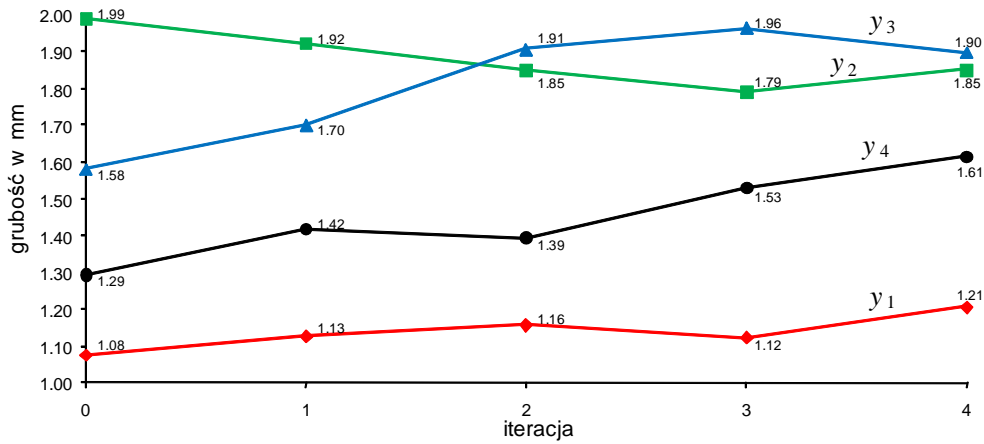
$$50\% \text{ deformacji całkowitej}, \quad (6.13)$$

$$0.8\text{mm} \leq y_i \leq 2.0\text{mm}, \quad i = 1, \dots, 4. \quad (6.14)$$

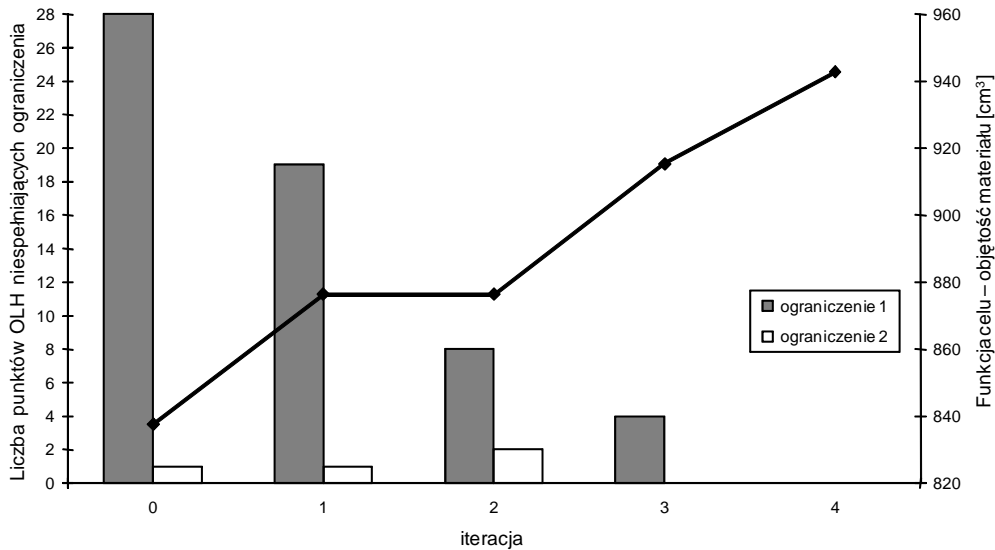
Przyjęty czas trwania zderzenia wynosi 20 ms. Minimalna wartość energii podana w ograniczeniu (6.12) odpowiada nominalnym grubościom blach $t_1 = t_2 = t_3 = 1.5 \text{ mm}$, $t_4 = 1 \text{ mm}$, wartościom średnim zmiennych X_5, \dots, X_8 oraz niewadliwym zgrzewom. Ograniczenie (6.13) wprowadzono aby preferować takie rozwiązania, które prowadzą do deformacji zlokalizowanych w pierwszej części belki (numeracja segmentów belki na rys. 6.1).

Rozwiązanie początkowe wybrano analizując 50 realizacji wektora zmiennych projektowych wygenerowanych przy pomocy zmodyfikowanego planu OLH zgodnie ze strategią opisaną w podrozdziałach 5.2.2.1 i 5.2.3. Na rys. 6.9, gdzie przedstawiono historię zmian wartości parametrów projektowych, punkt startowy odpowiada iteracji 0. Jak to pokazano w formie wykresu słupkowego na rys. 6.10, punkt startowy cechuje bardzo niski poziom niezawodności. Spośród 50 wygenerowanych punktów OLH, 28 nie spełnia ograniczenia (6.12), a 1 - ograniczenia (6.13).

Iteracyjny algorytm opisany w 5.2 realizowany jest aż do znalezienia pierwszego rozwiązania, które spełni ograniczenia. Do szacowania wartości współczynników korelacji pomiędzy funkcjami ograniczeń a zmiennymi projektowymi, a także do eksploracji obszaru rozwiązań dopuszczalnych, zastosowano 50-cio punktowy plan OLH. Rozwiązanie znaleziono po czterech krokach iteracyjnych. Jak widać na rys. 6.9, objętość materiału belki wzrosła w stosunku do rozwiązania nominalnego w wyniku znacznego pogrubienia segmentów 2 i 3 (wartości bliskie górnemu ograniczeniu prostemu). Grubość segmentu 1 jest znacznie mniejsza, co sprzyja realizowaniu się deformacji zlokalizowanych na styku belki z uderzającą masą, takich jak na rys. 6.3b.



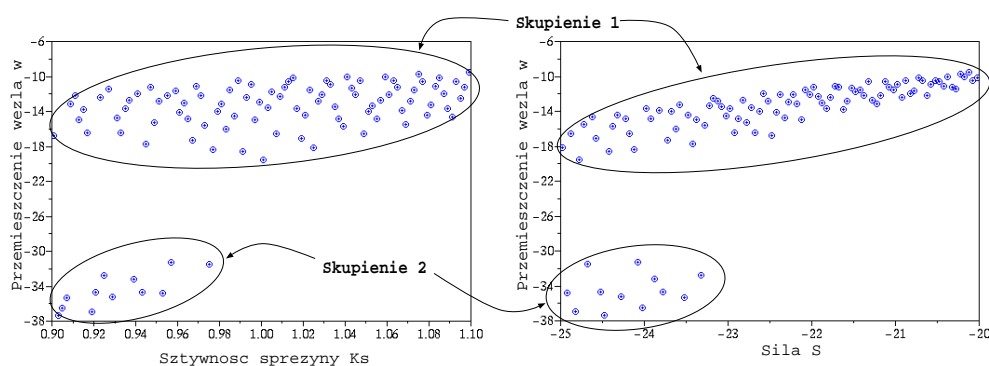
Rys. 6.9. Historia zmian wartości parametrów projektowych.



Rys. 6.10. Historia zmian liczby punktów OLH niespełniających ograniczenia 1 - (6.12) i 2 - (6.13) (wykres słupkowy) oraz historia zmian wartości funkcji celu (linia ciągła).

6.3. Jakościowa analiza zachowania elementu zgniatanego za pomocą algorytmu klasyfikacji wyników symulacji OLH

Podczas analizy wyników losowych symulacji zachowania się konstrukcji, na wykresach, których osiami są zmienne losowe oraz wybrane wielkości odpowiedzi, można często zaobserwować tworzenie się oddalonych od siebie skupień punktów (ang. *clusters*). Obrazy takie niosą ze sobą zazwyczaj istotne informacje dotyczące modelowanego układu konstrukcyjnego, świadczą o istnieniu odmiennych jakościowo postaci zniszczenia lub też o możliwości utraty stateczności. Jeśli przyjrzymy się wykresom punktowym na rys. 6.11, które przedstawiają zależności pomiędzy przemieszczeniem kratownicy Misesa (zob. podrozdział 4.1.5), a sztywnością sprężyny i wielkością przyłożonego obciążenia, to widoczne tam dwa skupienia wyników odpowiadają stanom przed i po przeskoku. Na-



Rys. 6.11. Wyraźne skupienia punktów eksperymentalnych dla przykładu kratownicy Misesa (zob. podrozdział 4.1.5). Grupa punktów w lewym dolnym rogu wykresów K_s - w i S - w (Skupienie 2) odpowiada przeskokowi kratownicy.

leży postawić pytanie, czy można łatwo zidentyfikować obszary w przestrzeni zmiennych, którym odpowiadają skupienia odpowiedzi konstrukcji widoczne na wykresach?

W przypadku jednej lub dwóch zmiennych losowych wystarczy odwzorować punkty poszczególnych skupień w punkty eksperymentalne, aby wizualnie sprawdzić czy można mówić o jakiegokolwiek prawidłowości, która tłumaczyłaby obserwowane skupienia. Jednak jeśli zmiennych jest więcej niż dwie, podejście polegające na analizie ułożenia odpowiednich punktów na płaszczyznach zdefiniowanych przez wszystkie pary zmiennych losowych staje się zbyt złożone i praktycznie bezużyteczne.

Dlatego w celu przeprowadzenia klasyfikacji poszczególnych skupień, poszukuje się powierzchni (w najprostszym przypadku hiperpłaszczyzny), która jak najlepiej rozdziela dwie grupy punktów eksperymentalnych prowadzące do

odmiennych jakościowo zachowań konstrukcji. Powierzchnia taka stanowić może również wygodne narzędzie do szybkiej oceny nowych rozwiązań.

Poniżej przedstawiona zostanie metoda znalezienia równania hiperpłaszczyzny, którą zastosowano w pracach [17, 21] w analizowanych przez autora przykładach dotyczących wytrzymałości zderzeniowej. Zaprezentowane sformułowanie przypomina, znaną z technik sztucznej inteligencji, metodę uczenia nadzorowanego, zwaną maszyną wektorów nośnych (ang. *support vector machine* - SVM) [22, 148]. Co ciekawe, dopiero w ostatnich kilku latach zaczęły ukazywać się pierwsze prace, gdzie proponuje się zastosowanie SVM w analizie niezawodności i optymalizacji niezawodnościowej, zob. [9, 10, 101, 174]. Równanie poszukiwanej hiperpłaszczyzny ma postać

$$\mathbf{a}\mathbf{x} = d, \quad (6.15)$$

gdzie \mathbf{a} jest wektorem normalnym do hiperpłaszczyzny o wymiarze n , a $|d|$ jest odległością od początku układu współrzędnych. Zadanie polega na znalezieniu takich wartości \mathbf{a} i d aby hiperpłaszczyzna (6.15) rozdzielała zbiór eksperymentów (realizacji zmiennych losowych) $S = \{\mathbf{x}^{(i)}, i = 1, \dots, N\}$ na dwa rozłączne podzbiory odpowiadające zaobserwowanym skupieniom. Oznaczmy te podzbiory jako $S_1 = \{\mathbf{x}^{(j)}, j \in I^1\}$ i $S_2 = \{\mathbf{x}^{(k)}, k \in I^2\}$, gdzie I^1 i I^2 są zbiorami indeksów z zakresu $1, \dots, N$ wynikającymi z zaklasyfikowania punktów jako należących do pierwszego bądź do drugiego skupienia. Oczywiście $N_1 + N_2 = N$, gdzie $N_1 = |S_1|$, a $N_2 = |S_2|$. Chcemy tak dobrać współczynniki równania (6.15), aby wszystkie punkty z S_1 znalazły się po jednej stronie hiperpłaszczyzny, a wszystkie punkty z S_2 po drugiej i aby maksymalizować odległość hiperpłaszczyzny od najbliższych punktów z S_1 i S_2 . Wprowadzając dodatkową zmienną r , oznaczającą najmniejszą odległość punktu ze zbioru S od hiperpłaszczyzny, zadanie optymalizacji, którego rozwiązanie daje w wyniku współczynniki \mathbf{a} i d zapisać można jako

$$\text{znajdź wartości zmiennych: } a_1, \dots, a_n, d, r, \quad (6.16)$$

$$\text{maksymalizujące: } r, \quad (6.17)$$

$$\text{przy ograniczeniach: } \sum_{i=1}^n a_i^2 = 1, \quad (6.18)$$

$$\mathbf{a}\mathbf{x}^{(i)} - d \geq r, \quad i \in I_1, \quad (6.19)$$

$$\mathbf{a}\mathbf{x}^{(i)} - d \leq -r, \quad i \in I_2, \quad (6.20)$$

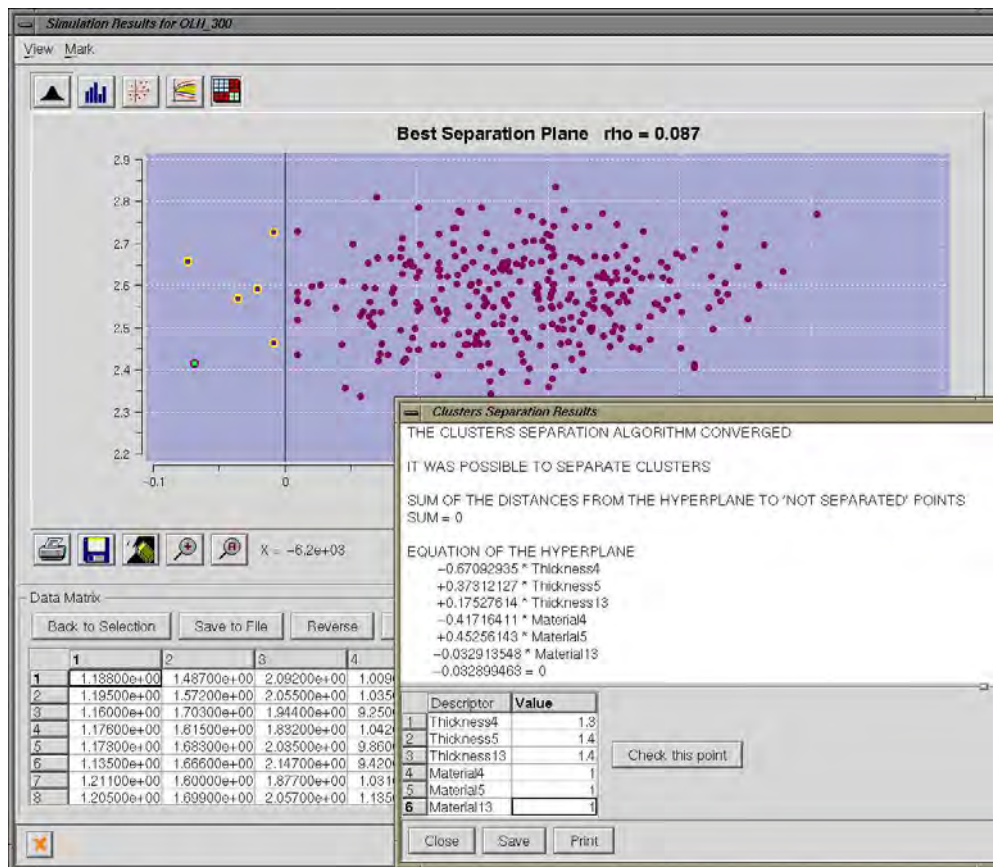
$$-1 \leq a_i \leq 1, \quad i = 1, \dots, n, \quad (6.21)$$

$$-\delta \leq d \leq \delta, \quad (6.22)$$

$$0 \leq r \leq \gamma, \quad (6.23)$$

gdzie a_i jest i -tą składową wektora normalnego \mathbf{a} , wartość δ wybrano jako największą odległość punktu $\mathbf{x}^{(i)} \in S$ od początku układu współrzędnych, a γ to najmniejsza odległość pomiędzy punktami podzbiorów S_1 i S_2 .

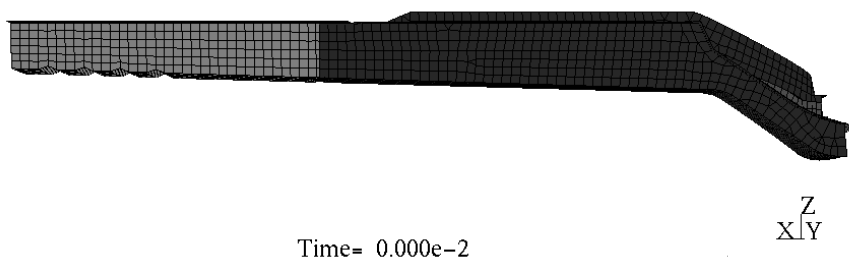
Tak sformułowana metoda klasyfikacji punktów eksperymentalnych zaimplementowana została przez autora w programie M-Xplore, module pakietu RADIOSS, który przeznaczony jest do analizy stochastycznej konstrukcji. Na rys. 6.12 widoczne jest okno dialogowe programu M-Xplore, gdzie oprócz wizualizacji wyników klasyfikacji istnieje możliwość wprowadzenia współrzędnych nowego punktu w celu określenia jego przynależności do jednego ze skupień.



Rys. 6.12. Okno dialogowe programu M-Xplore gdzie pokazano wynik zakończonej powodzeniem klasyfikacji punktów eksperymentalnych przy pomocy hiperpłaszczyzny. Rzutując punkty na płaszczyznę prostopadłą do znalezionej, można wyraźnie zaobserwować rozdział punktów należących do dwóch skupień.

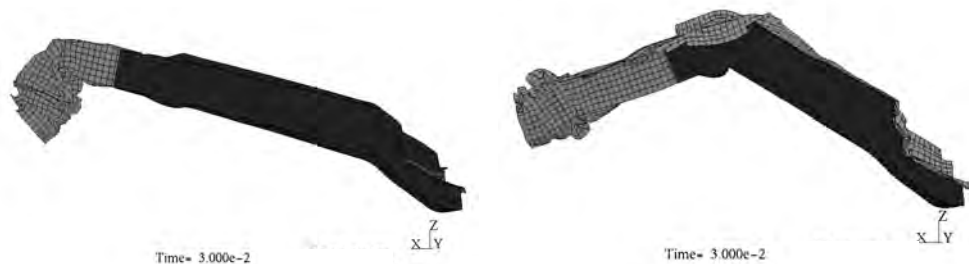
Do rozwiązania zadania (6.16)–(6.23), w prezentowanym poniżej przykładzie analizy wyników symulacji losowych zgniatania elementu tylnej ramy pojazdu, użyto algorytmu sekwencyjnego programowania kwadratowego NLPQL, zob. [215]. Zdarza się jednak, iż ze względu na nieliniowy charakter granicy pomiędzy dwoma skupieniami, algorytm ten nie jest w stanie odseparować podzbiorów hiperpłaszczyzną i nie można uzyskać jego zbieżności. W następnej iteracji, za pomocą metody NLPQL podejmuje się wtedy próbę rozdzielenia podzbiorów S_1 i S_2 hiperpowierzchnią drugiego stopnia.

Na rys. 6.13 zaprezentowano skończenie elementowy model rozpatrywanego fragmentu tylnej ramy zawieszenia samochodu. Model MES składa się z około 6000 elementów powłokowych. Element utwierdzony jest w prawym końcu, natomiast z lewej strony zgniatany jest przez poruszającą się ze stałą prędkością 35 km/h pionową sztywną przegrodę. Czas symulacji przyjęto równy 0.04 s, a więc przegroda przemieszcza się w tym czasie o około 40 cm, co stanowi 1/3 długości całkowitej elementu.



Rys. 6.13. Model MES elementu tylnej ramy zawieszenia pojazdu. Część modelu oznaczona jaśniejszym kolorem z lewej strony ma numer 1, a ciemniejsza część z prawej numer 2. Część znajdująca się wewnątrz części drugiej (widoczny jaśniejszy fragment po prawej stronie) oznaczona jest numerem 3.

Postać deformacji elementu jest kombinacją zgniatania oraz zginania (wybożenia). W zależności od przyjętych grubości poszczególnych części oraz wartości stałych materiałowych jedna z tych form deformacji może być dominująca, zob. rys. 6.14. Ponieważ z pierwszą postacią deformacji wiąże się zazwyczaj dobra absorpcja energii zderzenia, a druga cechuje się niską absorpcją i dużymi przemieszczeniami poprzecznymi, dlatego celem bieżącej analizy jest określenie za pomocą symulacji losowych takich parametrów konstrukcji, które prowadzą do danego typu deformacji.



Rys. 6.14. Dwie dominujące formy deformacji: zgniatanie oraz zginanie (globalne wyboczenie).

Na potrzeby tak postawionego zadania wybrano sześć zmiennych opisujących geometrię i materiał trzech części konstrukcji, które wyszczególnione są w opisie do rys. 6.13. Każda z części reprezentowana jest przez grubość stalowej blachy, z której została wykonana, jak również przez bezwymiarowy mnożnik α w związku konstytutywnym typu (6.1) postaci

$$\sigma = \alpha f(\varepsilon_p, \dot{\varepsilon}). \quad (6.24)$$

Przyjęto, że wszystkie zmienne mają jednostajny rozkład prawdopodobieństwa o następujących granicach:

X_1 - Grubość blachy części 1, $t_1 \in [1.02 \text{ mm}, 1.38 \text{ mm}]$,

X_2 - Grubość blachy części 2, $t_2 \in [1.36 \text{ mm}, 1.84 \text{ mm}]$,

X_3 - Grubość blachy części 3, $t_3 \in [1.7 \text{ mm}, 2.3 \text{ mm}]$,

X_4 - Parametr prawa konstytutywnego materiału części 1, $\alpha_1 \in [0.7, 1.3]$,

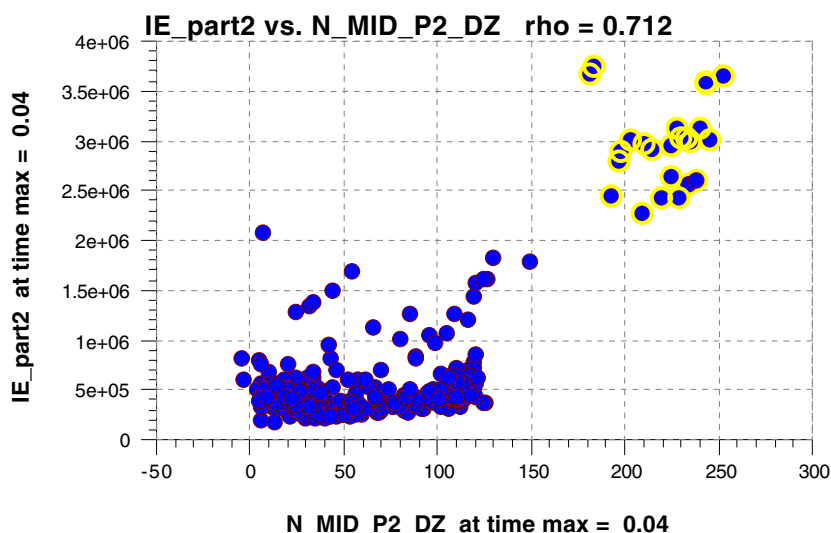
X_5 - Parametr prawa konstytutywnego materiału części 2, $\alpha_2 \in [0.7, 1.3]$,

X_6 - Parametr prawa konstytutywnego materiału części 3, $\alpha_3 \in [0.7, 1.3]$.

Jako charakterystyczne wielkości odpowiedzi deformowanego elementu przyjęto trzy składowe wektorów przemieszczenia oraz prędkości w czterech wybranych węzłach siatki MES. Trzy węzły znajdują się w połowach długości odpowiednich części konstrukcji, a czwarty wybrano na połączeniu części 1 i 2. Ponadto, rejestrowane są składowe “globalnej” prędkości całego elementu, całkowita energia odkształcenia, a także dla każdej z części trzy składowe średnich prędkości węzłów oraz odpowiednie energie odkształcenia. Daje to w sumie 40 monitorowanych odpowiedzi konstrukcji.

Za pomocą modułu M-Xplore przeprowadzono 300 symulacji MES programem RADIOSS generując realizacje zmiennych X_1, \dots, X_6 zgodnie z planem eksperymentów OLH.

Analizę wyników symulacji rozpoczęto od poszukiwania skupień mogących świadczyć o odmiennych postaciach deformacji na dwuwymiarowych wykresach punktowych wybranych odpowiedzi. Na rysunkach 6.15 i 6.16 przedstawiono dwa wykresy punktowe, na których wyraźnie widoczne jest układanie się wyników w dwa skupienia (takie same na obydwu wykresach). Co nie jest zaskakujące, najbardziej reprezentatywnymi dla deformacji typu zgięciowego okazały się przemieszczenie pionowe środkowego węzła części 2 i energia odkształcenia tej części.

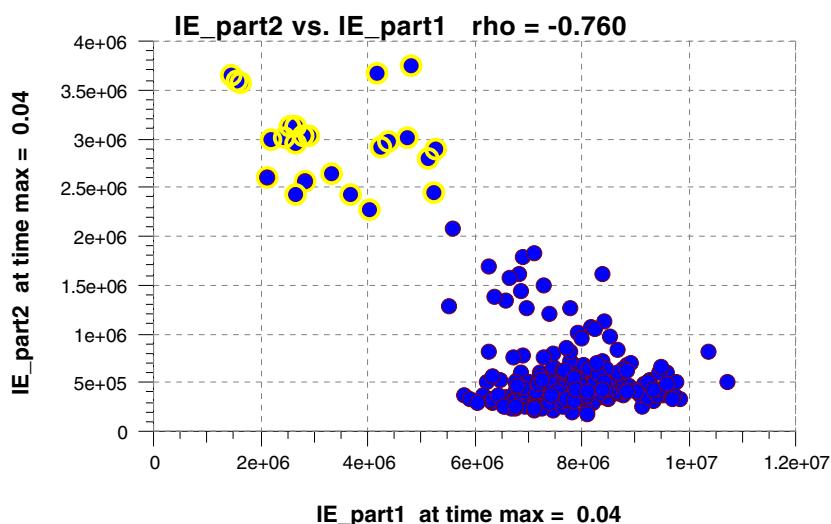


Rys. 6.15. Wykres punktowy sporządzony dla chwili $t = 0.04$ s, tzn. końca symulacji MES. Na osi poziomej, przemieszczenie pionowe węzła znajdującego się w środku części 2 (w milimetrach). Na osi pionowej, energia odkształcenia części 2. Skupienie w prawym górnym rogu odpowiada globalnemu wyboczeniu elementu.

Po dokonaniu klasyfikacji skupień punktów w przestrzeni odpowiedzi, poszukujemy następnie hiperpłaszczyzny rozdzielającej odpowiadające im realizacje w przestrzeni zmiennych losowych. Niestety nie można dokonać takiego prostego podziału oznaczając punkty skupień na piętnastu wykresach punktowych, których osiami są wszystkie możliwe kombinacje dwóch z sześciu zmiennych X_1, \dots, X_6 .

Hiperpłaszczyznę klasyfikującą udało się jednak znaleźć rozwiązując zadanie (6.16)–(6.23). Stosując wprowadzone wcześniej oznaczenia, dana jest ona równaniem

$$-0.70t_1 - 0.38\alpha_1 + 0.43t_2 + 0.42\alpha_2 + 0.02t_3 + 0.05\alpha_3 = 0.21. \quad (6.25)$$



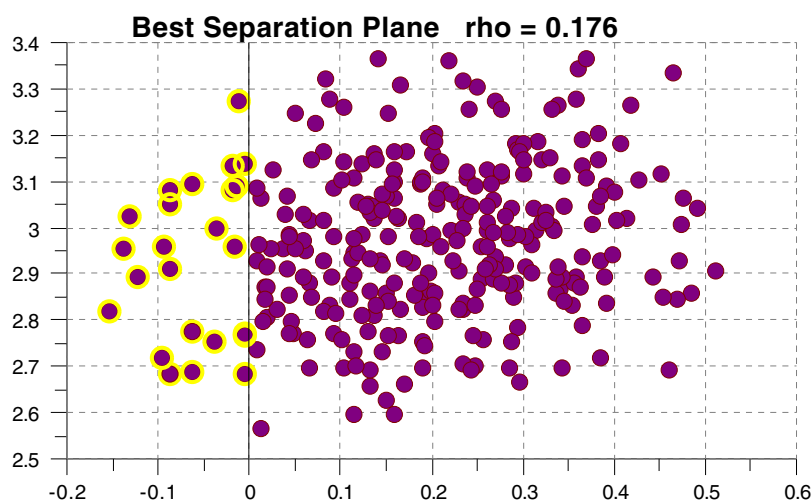
Rys. 6.16. Wykres punktowy sporządzony dla chwili $t = 0.04$ s. Na osi poziomej, energia odkształcenia części 1. Na osi pionowej, energia odkształcenia części 2. Widoczne jest skupienie składające się z tych samych punktów co na wykresie 6.15. Deformacja typu wybozczeniowego charakteryzuje się dużą energią zaabsorbowaną przez część 2 przy stosunkowo małej zaabsorbowanej przez część 1.

Na rys. 6.17 zaprezentowano wykres punktowy powstały z rzutowania punktów planu OLH na płaszczyznę prostopadłą do hiperpłaszczyzny (6.25).

Hiperpłaszczyzna rozdziela przestrzeń realizacji zmiennych losowych na dwa obszary. Po jednej stronie znajdują się te realizacje, które z dużym prawdopodobieństwem prowadzą do deformacji o dominującej postaci zgięciowej - charakterystycznej dla globalnej utraty stateczności. Po drugiej stronie hiperpłaszczyzny mamy do czynienia z regularnym zgniataniem elementu - deformacją, która wiąże się z dobrym absorbowaniem energii zderzenia. Obliczając wartość wyrażenia po lewej stronie równania (6.25) dla nowych wartości grubości blachy oraz nowych współczynników materiałowych określić można przypuszczalny typ deformacji elementu. Jeśli obliczona wartość jest większa od 0.21, to będzie to regularne zgniatanie, a jeśli mniejsza, to element prawdopodobnie ulegnie wybozczeniu.

Można pokusić się także o próbę analizy równania (6.25). Pomińmy na wstępie we wzorze wpływ zmiennych t_3 i α_3 ze względu na małe wartości stojących przy nich współczynników. Dokonując pewnych zaokrągleń zgrupujemy następnie pozostałe 4 zmienne tak, aby otrzymać wzór jakościowo podobny do wyjściowego

$$-0.4(2t_1 + \alpha_1) + 0.4(t_2 + \alpha_2) = 0.2. \quad (6.26)$$



Rys. 6.17. Wykres punktowy będący rzutem eksperymentów OLH na płaszczyznę prostopadłą do hiperpłaszczyzny (6.25). Płaszczyzna, na której zaobserwować można oddzielenie punktów należących do różnych skupień, określona jest przez wektor normalny do hiperpłaszczyzny i dowolny prostopadły do niego wektor.

Występujący w powyższym wzorze składnik $R_1 = 2t_1 + \alpha_1$ można traktować jako reprezentujący “globalną” sztywność części 1. Jeśli R_1 jest duże, to elementy powłokowe, którymi modelowana jest ta część są grubsze, a materiał jest sztywniejszy. W podobny sposób interpretować należy składnik $R_2 = t_2 + \alpha_2$, przy pomocy którego opisywać można sztywność części 2. Po wstawieniu do (6.26) otrzymuje się dalsze uproszczenie znalezionej warunku klasyfikacji

$$R_2 = R_1 + 0.5. \quad (6.27)$$

Ta postać wzoru jest już bardzo prosta do interpretacji. Jeśli $R_2 > R_1 + 0.5$, wtedy część 2 jest na tyle sztywna, że nie ulega zgięciu, a deformacja koncentruje się w części 1. W przeciwnym wypadku zaobserwujemy zginanie w części 2 oraz globalne wyboczenie elementu.

Zaprezentowany sposób analizy wyników komputerowych symulacji losowych pozwala na lepsze zrozumienie zachowania się konstrukcji i okazuje się być przydatny w wielu zastosowaniach. Jednak ograniczeniem w stosowaniu tej metody może być większa liczba obserwowanych skupień, jak również konieczność uwzględnienia zbyt wielu zmiennych.

6.4. Optymalizacja odpornościowa zgniatanej dynamicznie belki cienkościennej

Prezentowany przykład optymalizacji odpornościowej opracowany został w ramach europejskiego projektu badawczego APROSYS. Projekt ten poświęcony był rozwojowi systemów pasywnego bezpieczeństwa w konstrukcjach pojazdów samochodowych oraz wypracowaniu metod i narzędzi komputerowych, które pozwolą ograniczyć liczbę kosztownych testów laboratoryjnych na rzecz testów wirtualnych. Opis przykładu znaleźć można w raporcie [38].

Celem zadania jest takie dobranie grubości poszczególnych części cienkościennej belki podłużnicy, która znana jest z podrozdziałów 6.1 i 6.2, aby nie zwiększając znacznie jej nominalnego ciężaru zmaksymalizować absorbowaną przez belkę energię uderzającej masy i uczynić wielkość tej energii jak najmniej wrażliwą na losowy rozrzut parametrów konstrukcji oraz obciążenia. Jako zmienne projektowe przyjęto wartości oczekiwane losowych grubości czterech części tworzących belkę, zob. rys. 6.1. Założono, że początkowe (nominalne) wartości tych zmiennych wynoszą 1.5 mm, co daje początkową objętość materiału $V_{\text{init}} = 864\,000 \text{ mm}^3$.

Stochastyczny model belki składa się z 8 zmiennych losowych opisujących grubości blach, parametry materiałowe oraz warunki początkowe zadania. Zmienne te przedstawiono w tab. 6.4. Zastosowano, wprowadzoną w podrozdziale 2.3, konwencję podziału zmiennych losowych na zmienne typu X i typu P , w zależności od tego czy odpowiednie wartości oczekiwane są zmiennymi projektowymi czy nie. Założono, że wartości odchyłeń standardowych zmiennych X_1, \dots, X_4 obliczane będą proporcjonalnie do wartości oczekiwanych zgodnie ze stałym współczynnikiem zmienności. Wady zgrzewów uwzględniane są tak jak w poprzednich przykładach, poprzez losowy wybór 3 spośród 64 połączeń, które eliminuje się za każdym razem z modelu MES.

Zadanie optymalizacji odpornościowej sformułowane jest następująco, por. (2.43)–(2.48):

$$\text{znajdź wartości zmiennych: } \boldsymbol{\mu}_{\mathbf{X}} = \{\mu_{X_1}, \mu_{X_2}, \mu_{X_3}, \mu_{X_4}\}, \quad (6.28)$$

$$\text{minimalizujące: } \tilde{f} = \frac{1 - \alpha}{\mu^*} \mathbb{E}[f(\mathbf{X}, \mathbf{P})] + \frac{\alpha}{\sigma^*} \sigma[f(\mathbf{X}, \mathbf{P})], \quad (6.29)$$

$$\text{przy ograniczeniach: } \mathbb{E}[g(\mathbf{X})] - 3\sigma[g(\mathbf{X})] \geq 0, \quad (6.30)$$

$$0.5 \text{ mm} \leq \mu_{X_i} \leq 2.0 \text{ mm}, \quad i = 1, \dots, 4, \quad (6.31)$$

gdzie $f(\mathbf{X}, \mathbf{P})$ jest absorbowaną przez belkę energią wziętą ze znakiem minus, $\alpha \in [0, 1]$ jest współczynnikiem wagowym, μ^* i σ^* są stałymi normalizującymi

Tablica 6.3. Zmienne losowe w zadaniu optymalizacji odpornościowej zgniatanej podłużnicy. Wartości oczekiwane zmiennych \mathbf{X} są zmiennymi projektowymi.

Zmienna typu X	Opis	Rozkład prawdopodob.	Początkowa wartość oczekiwana	Współczynnik zmienności
X_1	t_1 - grubość części 1	lognormalny	1.5 [mm]	5%
X_2	t_2 - grubość części 2	lognormalny	1.5 [mm]	5%
X_3	t_3 - grubość części 3	lognormalny	1.5 [mm]	5%
X_4	t_4 - grubość części 4	lognormalny	1.5 [mm]	5%

Zmienna typu P	Opis	Rozkład prawdopodob.	Wartość oczekiwana	Odchylenie standardowe
P_1	σ_0 - granica plastyczności	lognormalny	180 [MPa]	15 [MPa]
P_2	E - moduł Younga	lognormalny	210000 [MPa]	10500 [MPa]
P_3	v_0^y - składowa y wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.0 [m/s]
P_4	v_0^z - składowa z wektora prędkości początkowej uderzającej masy	normalny	0 [m/s]	1.0 [m/s]

wyznaczonymi na podstawie wyników symulacji losowych przeprowadzonych dla pierwszego zbioru punktów eksperymentalnych, a $g(\mathbf{X}) = 1.1V_{\text{init}} - V(\mathbf{X})$ jest funkcją nakładającą ograniczenie na przyrost objętości belki - nie może ona przekroczyć 110% objętości początkowej. Objętość materiału jest funkcją losową ze względu na losowy charakter grubości blach i dlatego ograniczenie (6.30) sformułowano jako ograniczenie odpornościowe typu (2.45) przyjmując $\beta = 3$. Kryteria minimalizacji wartości średniej absorbowanej energii oraz jej odchylenia standardowego wydają się być wzajemnie konfliktowe. Konstruowanie podłużnicy z grubszych blach niewątpliwie zwiększa zdolność absorbowania energii. Z drugiej strony, rosnąca proporcjonalnie wariancja tych grubości prowadzi do większego losowego rozrzutu obliczanych wartości energii. Trzeba jednak pamiętać, że dla ustalonej objętości materiału wpływ na wyniki ma również decyzja o "dystrybucji" grubości blach w ramach czterech części belki.

Czas pracy procesora konieczny do przeprowadzenia pojedynczej analizy MES za pomocą programu RADIOSS wynosi około dwie godziny. Ze względu na ograniczenia licencyjne posiadanej wersji programu, w obliczeniach można było korzystać z równoległej pracy jedynie 16 procesorów. Z tego powodu zdecydowano się wykonać zaledwie jedną iterację algorytmu optymalizacji odpornościowej, składającą się z:

- Wygenerowania 32 realizacji wektora zmiennych projektowych równomiernie w obszarze wyznaczonym przez zakresy zmienności (6.31), zgodnie z planem OLH.
- Wyznaczenia momentów statystycznych funkcji celu i ograniczeń poprzez symulacje losowe typu OLH na podstawie próbek o wielkości $N = 64$, generowanych dla każdej z 32 realizacji wektora $\boldsymbol{\mu}_{\mathbf{X}}$.
- Budowy aproksymacji metodą krigingu (zob. podrozdział 3.2) funkcji $\mu_f(\boldsymbol{\mu}_{\mathbf{X}})$, $\sigma_f(\boldsymbol{\mu}_{\mathbf{X}})$, $\mu_g(\boldsymbol{\mu}_{\mathbf{X}})$, $\sigma_g(\boldsymbol{\mu}_{\mathbf{X}})$. Aproksymacje te dane są odpowiednio jako $\hat{\mu}_f(\boldsymbol{\mu}_{\mathbf{X}})$, $\hat{\sigma}_f(\boldsymbol{\mu}_{\mathbf{X}})$, $\hat{\mu}_g(\boldsymbol{\mu}_{\mathbf{X}})$, $\hat{\sigma}_g(\boldsymbol{\mu}_{\mathbf{X}})$.
- Dla danego α , przy pomocy wybranego algorytmu optymalizacji deterministycznej, znalezienia rozwiązania zadania minimalizacji

znajdź wartości zmiennych: $\boldsymbol{\mu}_{\mathbf{X}}$, (6.32)

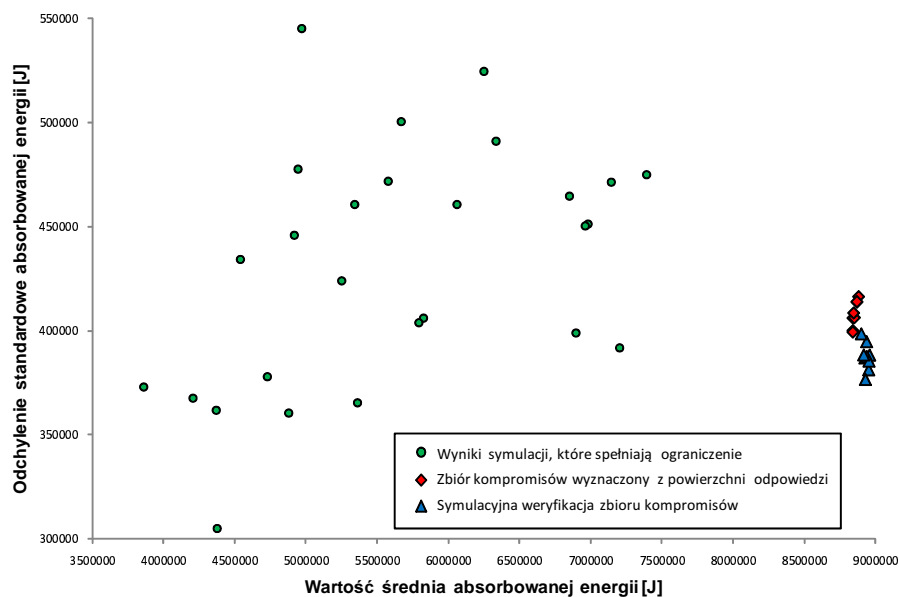
$$\text{minimalizujące: } \frac{1-\alpha}{\mu^*} \hat{\mu}_f(\boldsymbol{\mu}_{\mathbf{X}}) + \frac{\alpha}{\sigma^*} \hat{\sigma}_f(\boldsymbol{\mu}_{\mathbf{X}}), \quad (6.33)$$

$$\text{przy ograniczeniach: } \hat{\mu}_g(\boldsymbol{\mu}_{\mathbf{X}}) - 3 \hat{\sigma}_g(\boldsymbol{\mu}_{\mathbf{X}}) \geq 0, \quad (6.34)$$

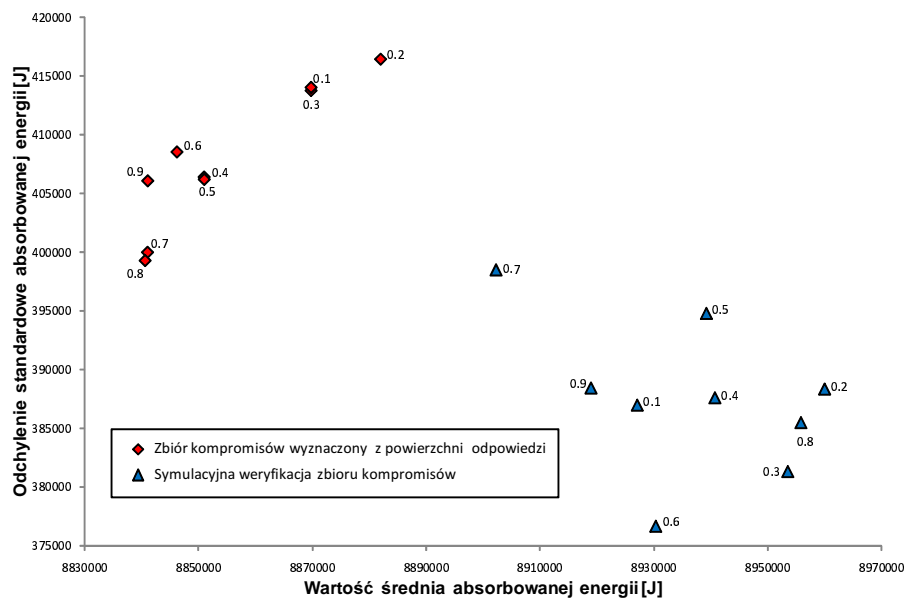
$$0.5 \text{ mm} \leq \mu_{X_i} \leq 2.0 \text{ mm}, \quad i = 1, \dots, 4. \quad (6.35)$$

A zatem, w celu skonstruowania koniecznych powierzchni odpowiedzi należało wykonać 2048 symulacji skończenie elementowych dla różnych wartości parametrów modelu MES. Z uwagi na wspomniane już ograniczenia dotyczące paralelizacji obliczeń, było to zadanie dość czasochłonne. Zastosowane podejście, polegające na tworzeniu niezależnych aproksymacji dla wszystkich występujących w sformułowaniu (6.28)–(6.31) momentów statystycznych, omówione zostało w podrozdziale 2.4.2.2 jako tzw. strategia aproksymacji statystyk.

W celu zbadania wpływu wartości współczynnika wagowego na otrzymywane rozwiązania rozpatrzono 9 przypadków, przyjmując kolejno $\alpha = 0.1, 0.2, \dots, 0.9$. Na wykresie 6.18, którego osiami są kryteria optymalizacji $\mu_f = \mathbb{E}[f(\mathbf{X}, \mathbf{P})]$ i $\sigma_f = \sigma[f(\mathbf{X}, \mathbf{P})]$, pokazano wyniki odpowiadające punktom eksperymentalnym jak również rozwiązania optymalne oraz ich symulacyjne weryfikacje. Punkty oznaczone kółkami reprezentują wartości kryteriów oszacowane za pomocą symulacji losowych i odpowiadają tym spośród 32 realizacji zmiennych projektowych, które spełniają ograniczenie (6.30) dotyczące zmian objętości. Czworokątne symbole oznaczają 9 rozwiązań zadania (6.32)–(6.35) dla różnych współczynników α , a więc są punktami ze zbioru rozwiązań kompromisowych (ocen rozwiązań niezdominowanych). Punkty oznaczone trójkątami są wynikami weryfikacji wartości ocen rozwiązań niezdominowanych za pomocą symulacji losowych, gdzie próbki generowane są przez 64-ro punktowe plany OLH.



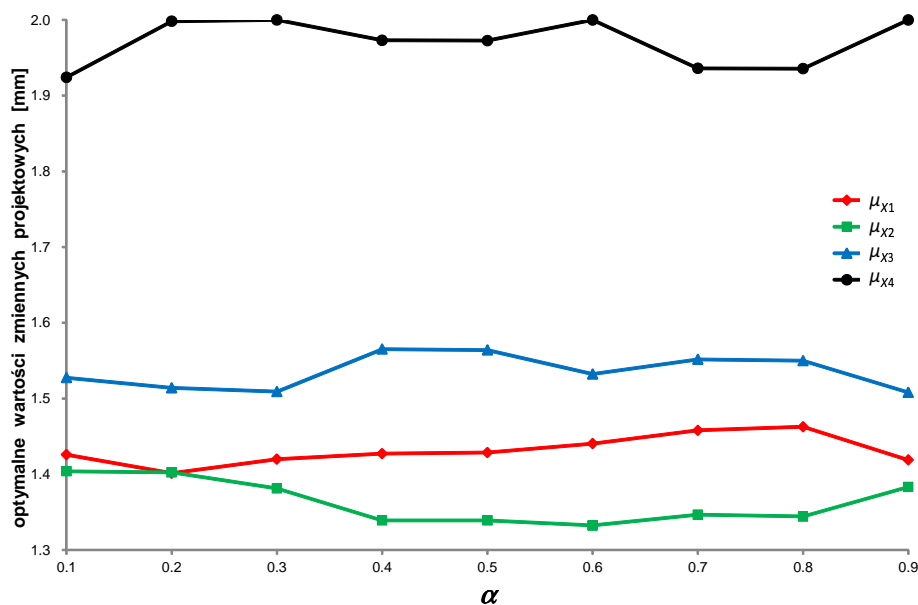
Rys. 6.18. Wartości funkcji kryteriów optymalizacji odpowiadające tym spośród 64 punktów eksperymentalnych, które spełniają ograniczenie (6.30), oraz zbiór rozwiązań kompromisowych i odpowiadających im weryfikacji metodą losowych symulacji OLH.



Rys. 6.19. Punkty zbioru kompromisów oraz ich weryfikacji za pomocą symulacji losowych. Wartości przy symbolach odpowiadają użytym współczynnikom wagowym α .

Zbiór kompromisów oraz odpowiadający mu zbiór weryfikacji “powiększono” na rys. 6.19. Można zauważyć, że w związku z ograniczoną dokładnością powierzchni odpowiedzi budowanych w oparciu o niewielką liczbę punktów eksperymentalnych oraz w związku z błędem, którym obarczone są wyniki symulacji losowych, wzajemne pozycje punktów w zbiorze kompromisów oraz zbiorze weryfikacji nie odpowiadają jakościowo wartościom współczynników α , dla których otrzymano rozwiązania. Ponadto, rozwiązania te wydają się nieznacznie zaniżać wartość średnią absorbowanej energii oraz zawyżać jej odchylenie standardowe.

Z drugiej strony, dla przebadanego zakresu zmienności współczynnika wagiowego uzyskiwane optymalne wartości kryteriów nie różnią się znacznie od siebie, a dokładność aproksymacji wydaje się być zadowalająca z punktu widzenia praktyki projektowej. Analizując rozwiązania niezdominowane (Pareto) pokazane na rys. 6.20 można dojść do wniosku, że rozwiązanie zadania optymalizacji odpornościowej rozpatrywanej belki prowadzi do powiększenia grubości płyty (część 4), nieznacznego zmniejszenia grubości dwóch pierwszych części typu Ω oraz do grubości części trzeciej, która bliska jest nominalnej wartości 1.5 mm. Na podstawie obserwowanego skupienia rozwiązań optymalnych w przestrzeni kryteriów na wykresie 6.18 oraz małego zróżnicowania rozwiązań na rys. 6.20 wnioskować można, iż kryteria $\{\mu_f, \sigma_f\}$ nie są konfliktowe.



Rys. 6.20. Rozwiązania niezdominowane (Pareto) dla rozpatrywanych wartości współczynnika α .

6.5. Optymalizacja odpornościowa w zadaniu kalibracji parametrów modelu numerycznego

Poniższy przykład dotyczy zastosowania techniki optymalizacji odpornościowej w zadaniu doboru wartości parametrów modelu numerycznego przeznaczonego do symulacji zderzenia fantomu dolnej kończyny z barierą AE-MDB (ang. *Advanced European Mobile Deformable Barrier*), zob. [63]. Przykład ten, stanowiący ilustrację możliwości użycia optymalizacji odpornościowej do komputerowego wspomaganie procesu certyfikacji pojazdów pod kątem ich bezpieczeństwa w przypadku kolizji z pieszym, został opublikowany w jednym z raportów z działalności projektu APROSYS [47].

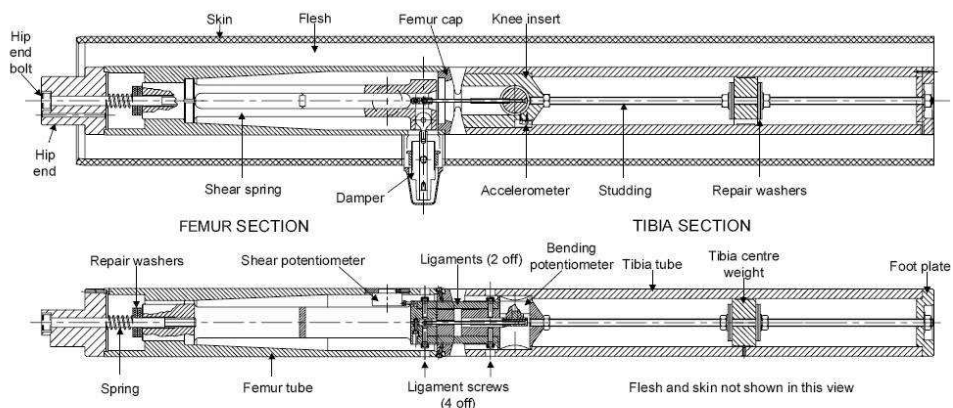
Testy takie jak pokazany na rys. 6.21 służą sprawdzeniu czy konstrukcja przedniej części samochodu, a w szczególności jego zderzaka, nie powoduje zbyt dużych urazów nogi pieszego przy typowym potrąceniu z prędkością 40 km/h.



Rys. 6.21. Próba zderzeniowa na fantomie kończyny dolnej. Zdjęcie z pracy [99], dzięki uprzejmości Iztoka Cigralica z TUG.

W celu kompleksowej oceny jakości stosowanego fantomu, a zwłaszcza powtarzalności uzyskiwanych przy jego pomocy wyników oraz wpływu jakości na rozrzut tych wyników ma losowy charakter warunków początkowych, skonstruowano specjalne stanowisko testowe. Składa się ono z fantomu kończyny dolnej, wystrzeliwanego w kierunku bariery z pneumatycznego działa o wysokiej precyzji pozycjonowanego przez ramię robota, oraz bariery AE-MDB, która modeluje przód pojazdu osobowego.

Schemat fantomu kończyny dolnej przedstawiono na rys. 6.22, a na rys. 6.23 pokazano zdjęcia ukazujące szczegóły jego budowy, m.in. czujniki umożliwiające dokonanie w momencie zderzenia następujących odczytów: przyspieszenia górnej części kości piszczelowej, kąta bocznego zgięcia kolana oraz wzajemnego przemieszczenia ścinającego kości udowej i kości piszczelowej. W uproszczeniu można powiedzieć, że fantom składa się z dwóch rur metalowych o promieniu zewnętrznym 70 mm reprezentujących kość udową i kość piszczelową o odpowiednio dobranej masie, momencie bezwładności i środku ciężkości. Otaczające mięśnie modelowane są za pomocą pianki Confor (CF-45 1) o grubości 25 mm, zob. [24]. Fantom powleczony jest 6 mm warstwą kauczuku syntetycznego reprezentującego ludzką skórę.



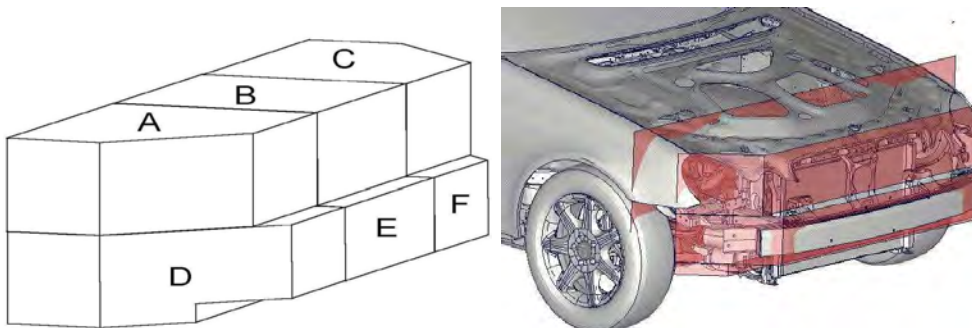
Rys. 6.22. Schemat fantomu kończyny dolnej. Rysunek z pracy [99], dzięki uprzejmości Iztoka Cigralica z TUG.

Zastosowana sztywna bariera odpowiada geometrią połowie przedniej części mobilnej bariery AE-MDB, której kształt modeluje z kolei przód pojazdu osobowego, zob. rys. 6.24. Na rys. 6.25, obok izometrycznego widoku bariery, na przekroju zaznaczono dwie warstwy z pianki polipropylenowej (ang. *Expanded PolyPropylene foam - EPP*) dołączone w celu ochrony fantomu kończyny przed zniszczeniem podczas wielokrotnie powtarzanego doświadczenia. Charakterystyki naprężenie-odkształcenie materiału EPP dla pięciu różnych prędkości odkształcenia przedstawiono na rys. 6.26, por. [16].

Fantom wystrzeliwany jest z pneumatycznego działa, o charakterystyce podanej w [99], które za każdym razem ustawiane jest za pomocą robota firmy KUKA. Na rys. 6.27 pokazano tolerancje położenia oraz prędkości fantomu



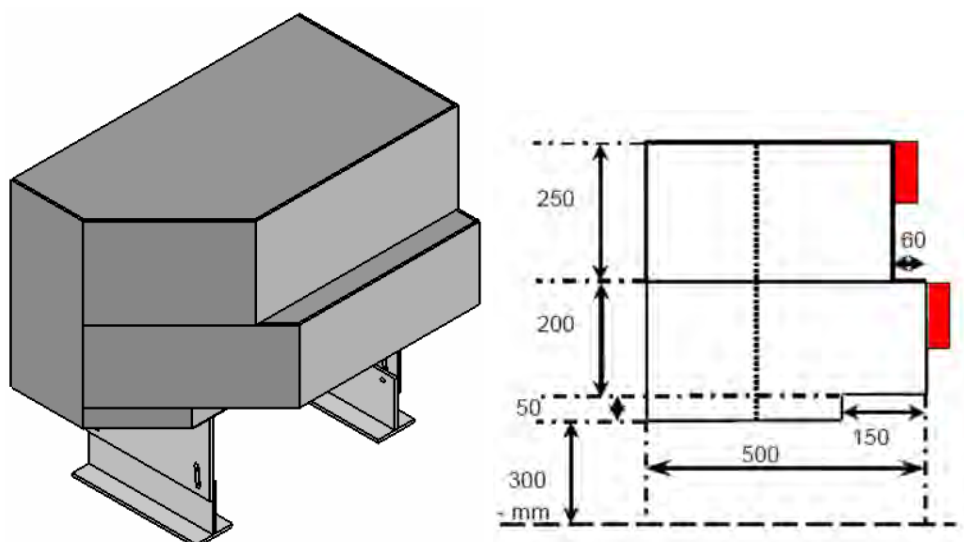
Rys. 6.23. Szczegóły budowy fantomu kończyny dolnej. Zdjęcia z pracy [99], dzięki uprzejmości Iztoka Cigralica z TUG.



Rys. 6.24. Porównanie geometrii bariery AE-MDB i przedniej części współczesnego samochodu osobowego.

w momencie uderzenia w barierę. Są to wielkości wynikające z precyzji urządzenia, a zarazem takie, które są dopuszczalne przez protokół certyfikacyjny.

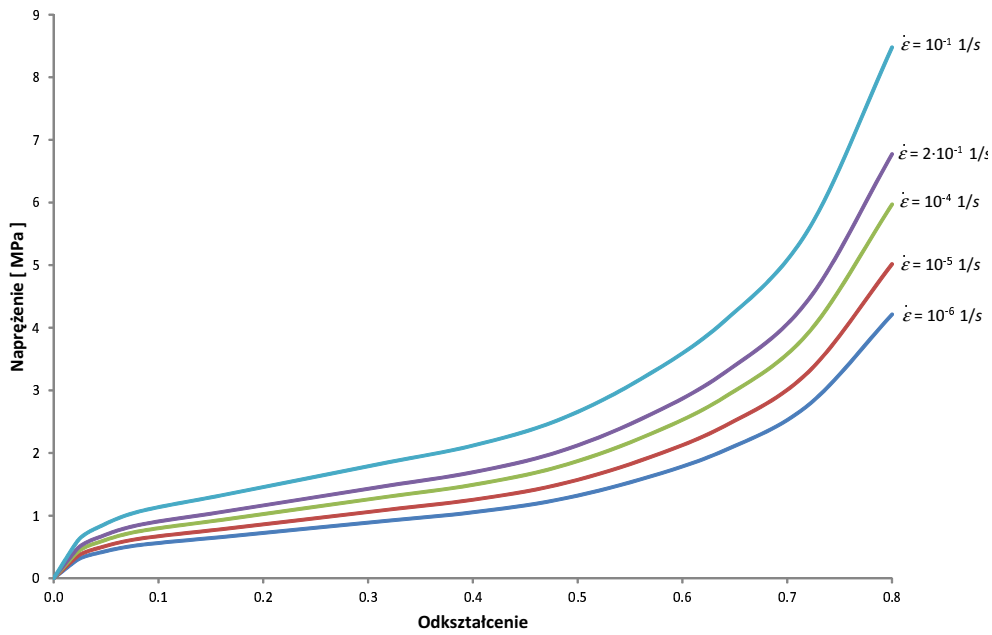
Wykonano 15 powtórzeń eksperymentu otrzymując oszacowanie momentów statystycznych wartości odczytów czujników umieszczonych w fantomie.



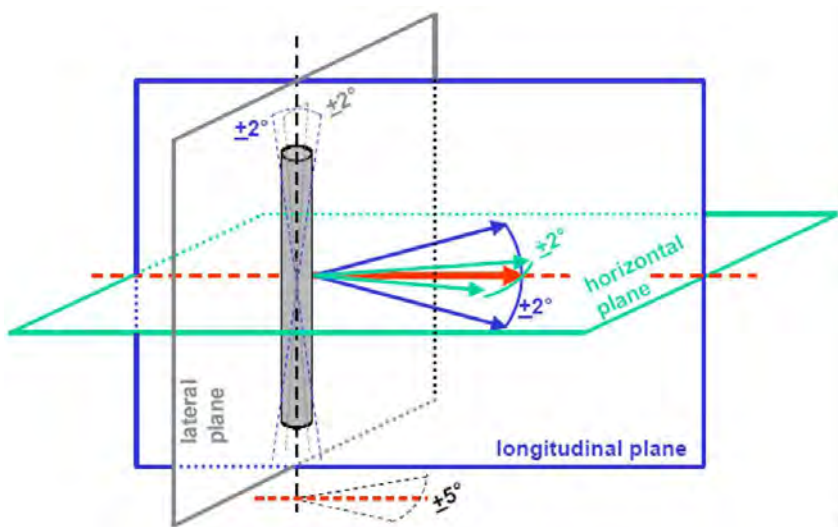
Rys. 6.25. Lewa strona: bariera używana w doświadczeniu (połowa przedniej części AE-MDB), Prawa strona: Przekrój poprzeczny, na czerwono zaznaczono dołączoną piankę polipropylenową. Rysunki z pracy [99], dzięki uprzejmości Izтока Cigralica z TUG.

Celem zaplanowanej optymalizacji odpornościowej jest takie dobranie wybranych parametrów modelu numerycznego, aby rozrzut wyników symulacji MES wynikający z założonych niepewności warunków początkowych odpowiadał jak najlepiej rozrzutom otrzymywanym w doświadczeniu. Model komputerowy złożony z fantomu oraz bariery pokazano na rys. 6.28. Składa się on przede wszystkim z ponad 30 000 przestrzennych elementów 8-węzłowych i 3200 4-węzłowych elementów powłokowych. Czas trwania zderzenia, symulowanego programem RADIOSS, założono równy 33 ms. Jako kryterium uszkodzenia kończyn dolnych przyjęto kąt bocznego zgięcia kolana.

Oprócz warunków początkowych, w modelu stochastycznym uwzględniono losowy charakter parametrów związku konstytutywnego pianki EPP jak również współczynnika tarcia pomiędzy barierą, a syntetycznym kauczukiem pokrywającym fantom. Zadanie optymalizacji polegać będzie na znalezieniu takich wartości oczekiwanych tych dwóch zmiennych, aby statystyki wyników eksperymentalnych odpowiadały statystykom otrzymywanym z testów komputerowych. Opis wszystkich zmiennych losowych przedstawiono w tab. 6.4. W sposób schematyczny zaznaczono je także na rys. 6.28.



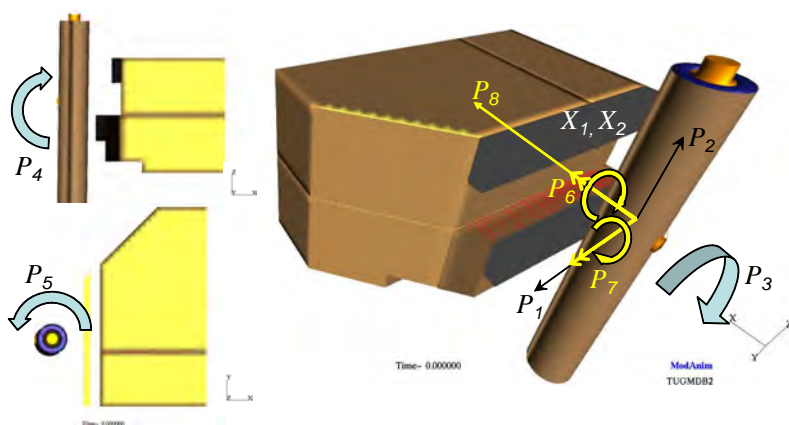
Rys. 6.26. Zależności napężenie-odkształcenie pianki polipropylenowej dla różnych prędkości odkształcenia.



Rys. 6.27. Dopuszczalne tolerancje położenia oraz prędkości fantomu kończyny dolnej w momencie uderzenia w barierę.

Tablica 6.4. Zmienne losowe w zadaniu optymalizacji odpornościowej w problemie kalibracji parametrów modelu numerycznego fantomu kończyny dolnej uderzającego w barierę AE-MDB. Wartości oczekiwane zmiennych \mathbf{X} są zmiennymi projektowymi.

Zmienna typu X	Opis	Rozkład prawdopodob.	Początkowa wartość oczekiwana	Odchylenie standardowe
X_1	mnożnik związków konstytutywnych	normalny	1.0	0.0333
X_2	współczynnik tarcia	normalny	0.3	0.02
Zmienna typu P	Opis	Rozkład prawdopodob.	Wartość oczekiwana	Odchylenie standardowe
P_1	położenie poziome fantomu	normalny	0.0 [mm]	3.333 [mm]
P_2	położenie pionowe fantomu	normalny	0.0 [mm]	3.333 [mm]
P_3	θ_x - początkowy kąt obrotu fantomu wokół osi x	normalny	0.0	0.6667
P_4	θ_y - początkowy kąt obrotu fantomu wokół osi y	normalny	0.0	0.6667
P_5	θ_z - początkowy kąt obrotu fantomu wokół osi z	normalny	0.0	0.6667
P_6	θ_y^v - kąt obrotu wektora prędkości początkowej fantomu wokół osi y	normalny	0.0	0.6667
P_7	θ_z^v - kąt obrotu wektora prędkości początkowej fantomu wokół osi z	normalny	0.0	0.6667
P_8	prędkość początkowa fantomu	normalny	11.11 [m/s]	0.667 [m/s]



Rys. 6.28. Model numeryczny fantomu kończyny dolnej i bariery oraz zmienne losowe zadania optymalizacji odpornościowej.

Podobnie jak w przykładzie dotyczącym optymalizacji podłużnicy (podrozdział 6.4) w oznaczeniach zastosowano podział na zmienne X i P . Do tych pierwszych należą: mnożnik empirycznych związków materiałowych danych za pomocą wykresów na rys. 6.26 oraz współczynnik tarcia. Zmienne typu P , które nie podlegają żadnej kontroli w procesie optymalizacji, reprezentują warunki początkowe zadania.

Problem optymalizacji odpornościowej sformułowano następująco (por. (2.49)):

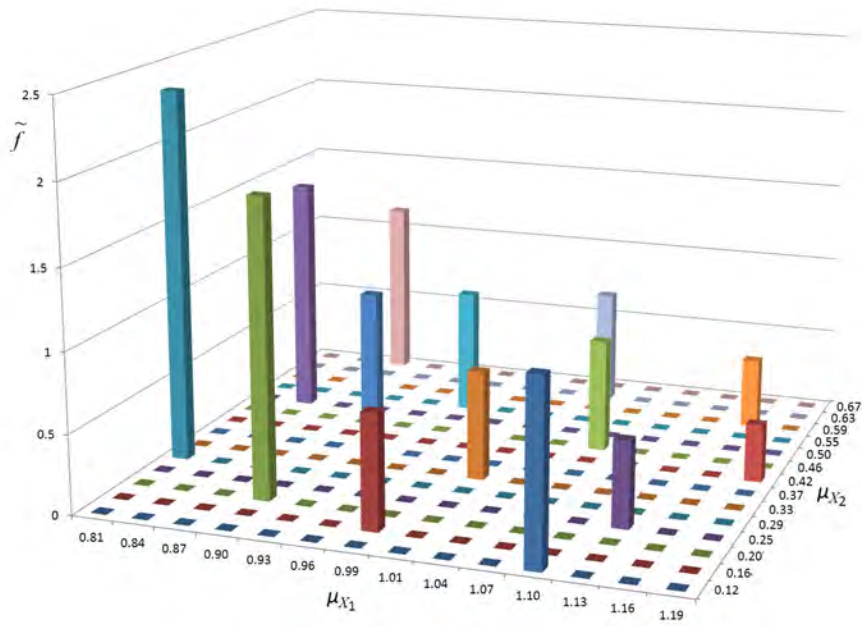
$$\text{znajdź wartości zmiennych: } \boldsymbol{\mu}_{\mathbf{X}} = \{\mu_{X_1}, \mu_{X_2}\}, \quad (6.36)$$

$$\begin{aligned} \text{minimalizujące: } \check{f} = & \frac{1-\alpha}{\mu^*} \left[\mathbb{E}[f(\mathbf{X}, \mathbf{P})] - M^{\mathbb{E}} \right]^2 \\ & + \frac{\alpha}{\sigma^*} \left[\sigma[f(\mathbf{X}, \mathbf{P})] - M^{\sigma} \right]^2, \end{aligned} \quad (6.37)$$

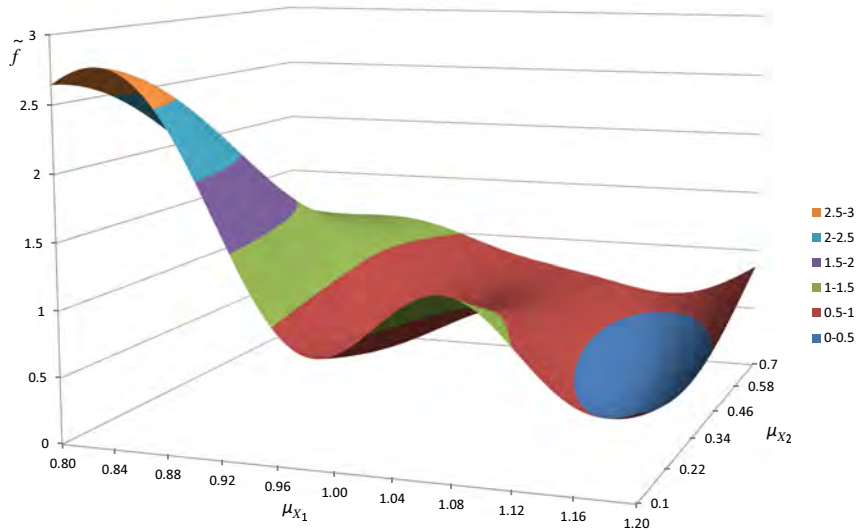
$$\text{przy ograniczeniach: } 0.8 \leq \mu_{X_1} \leq 1.2, \quad 0.1 \leq \mu_{X_2} \leq 0.7, \quad (6.38)$$

gdzie f jest kątem bocznego zgięcia kolana, μ^* i σ^* są stałymi normalizującymi, ustalonymi na podstawie wyników symulacji dla przyjętego zbioru punktów eksperymentalnych, $M^{\mathbb{E}} = 25.77$ i $M^{\sigma} = 1.2$ to wyznaczone eksperymentalnie wartości średnia i odchylenie standardowe funkcji f . Współczynnik wagowy α przyjęto równy 0.5, nie różnicując preferencji osiągnięcia wartości $M^{\mathbb{E}}$ i M^{σ} przez odpowiednie momenty statystyczne otrzymane przy pomocy symulacji numerycznych.

W celu zbudowania aproksymacji funkcji celu (wzór (6.37)), wartości średnie oraz odchylenia standardowe funkcji f oszacowano w 14 punktach eksperymentalnych, wygenerowanych w sposób równomierny na obszarze zdefiniowanym nierównościami (6.38). Punkty generowane były zgodnie z koncepcją OLH. Planów OLH używano także do tworzenia 100-elementowych próbek w 10-wymiarowej przestrzeni zmiennych losowych w celu estymacji momentów $\mathbb{E}[f(\mathbf{X}, \mathbf{P})]$ i $\sigma[f(\mathbf{X}, \mathbf{P})]$. Koszt przygotowania danych do aproksymacji funkcji celu równy jest zatem kosztowi 1400 analiz MES przeprowadzonych za pomocą programu RADIOSS. Ponieważ czas trwania jednej analizy to około 2 godziny, dlatego realizując zadanie na 32 równoległych procesorach, 14 wartości funkcji \check{f} otrzymano po niecałych 4 dobach obliczeń. Wartości te zilustrowane zostały w postaci słupków na rys. 6.29. Powierzchnia odpowiedzi utworzona metodą krigingu "rozpięta" na punktach eksperymentalnych pokazana jest na rys. 6.30. W przeciwieństwie do przykładu przedstawionego w poprzednim podrozdziale, nie stosowano strategii aproksymacji statystyk, tworząc aproksymację metodą krigingu bezpośrednio dla wagowej funkcji celu, zgodnie ze strategią opisaną w podrozdziale 2.4.2.3.



Rys. 6.29. Wartości funkcji celu (6.37) dla 14 punktów eksperymentalnych wygenerowanych zgodnie z planem OLH.



Rys. 6.30. Aproksymacja metodą kringingu funkcji celu (6.37).

Można zauważyć wyraźną tendencję zmniejszania się wartości funkcji celu wraz ze wzrostem mnożnika μ_{X_1} . W przypadku drugiej zmiennej projektowej (współczynnika tarcia) żadna globalna tendencja nie jest widoczna. Skonstruowana aproksymacja posłużyła do znalezienia następującego oszacowania optymalnych wartości zmiennych projektowych:

$$\boldsymbol{\mu}_{\mathbf{X}}^{\text{opt}} = \{\mu_{X_1} = 1.16, \mu_{X_2} = 0.373\}. \quad (6.39)$$

Postępując zgodnie z algorytmem zaproponowanym w podrozdziale 2.4.2.3 można następnie zawęzić obszar poszukiwań, wygenerować eksperymenty numeryczne, przeprowadzić symulacje losowe, zbudować nową aproksymację funkcji f i na tak lokalnie udoskonalonej powierzchni odpowiedzi znaleźć nowe oszacowanie minimum.

Mimo to, ze względu na znaczny koszt analizy MES nie przeprowadzono kolejnych iteracji. Uzyskany wynik może być więc obciążony pewnym błędem w związku z “zaszumionym” charakterem wyników symulacji losowych, co jest nieuniknione zwłaszcza przy stosowaniu niewielkich próbek. Z drugiej strony, można z dużym prawdopodobieństwem założyć, że aby na podstawie symulacji numerycznych otrzymać podobne statystyki wybranego kryterium uszkodzenia kończyny dolnej do statystyk z eksperymentu fizycznego, należy przyjąć współczynnik skalujący w związku konstytutywnym pianki EPP bliski 1.2.

Dokonano weryfikacji wartości średniej oraz odchylenia standardowego funkcji f w punkcie $\boldsymbol{\mu}_{\mathbf{X}}^{\text{opt}}$ otrzymując odpowiednio 24.0 i 1.28, co porównując z wartościami $M^{\mathbb{E}}$ i M^{σ} , wydaje się być bardzo dobrym wynikiem.

PODSUMOWANIE

Czasochłonna analiza numeryczna zagadnień dotyczących wytrzymałości zderzeniowej konstrukcji stanowi często istotną przeszkodę w szerszym stosowaniu analizy stochastycznej, a w szczególności niedeterministycznej optymalizacji tych konstrukcji. W niniejszym rozdziale zaprezentowano pięć przykładów analizy stochastycznej, gdzie dzięki zastosowaniu odpowiednio przygotowanych algorytmów możliwa jest zarówno analiza niezawodności jak i optymalizacja odpornościowa dynamicznie deformujących się elementów konstrukcji. Trzeba podkreślić, że obok dedykowanych metod, niezbędnym składnikiem efektywnej analizy stochastycznej jest nowoczesny sprzęt komputerowy, dający możliwość obliczeń równoległych. Rosnąca nieustannie moc obliczeniowa, jak również dostępność takiego sprzętu pozwala mieć nadzieję na coraz powszechniejsze stosowanie metod analizy stochastycznej w procesie projektowania złożonych konstrukcji.

We wszystkich analizowanych w rozdziale przykładach zarówno funkcje graniczne, jak też funkcje tworzące sformułowanie zadania optymalizacji odpornościowej, zawierały wpływ szumu. Szum ten jest charakterystyczny dla zagadnień zderzeniowych ze względu na metodę całkowania równań równowagi oraz stosowane algorytmy kontaktu. Ponadto, w zadaniu analizy niezawodności zgniatanej cienkościennej podłużnicy dodatkowy szum związany był z modelowaniem losowych awarii zgrzewów. Biorąc pod uwagę ogromną liczbę połączeń zgrzewanych występujących w konstrukcjach samochodowych, zaproponowana metoda usuwania pewnej liczby losowo wybranych zgrzewów wydaje się być praktycznym rozwiązaniem, które pozwala uwzględnić losowe wady tych połączeń. Podejście to wymaga jednak odpowiedniego algorytmu analizy niezawodności. Pokazano, że zastosowanie algorytmu Z3M umożliwia efektywne oszacowanie prawdopodobieństwa awarii pomimo nieróżniczkowalności funkcji granicznej. Model podłużnicy wykorzystano także do ilustracji działania metody poprawy niezawodności (omówionej w podrozdziale 5.2) oraz strategii aproksymacji statystyk w optymalizacji odpornościowej. W obydwu przykładach, podobnie jak w zadaniu analizy niezawodności, stosowano eliminację losowo wybranych zgrzewów. Zaproponowane metody okazały się niewrażliwe na “zaszumiony” charakter rozpatrywanych odpowiedzi konstrukcji, co należy uznać za ich niezwykle korzystną cechę.

W rozdziale przedstawiono również dwa nietypowe zastosowania metod symulacyjnych. Zaprezentowano przykład jakościowej analizy zachowania zgniatanego elementu tylnej ramy pojazdu. Metoda ta wykorzystuje prosty algorytm klasyfikacji wyników eksperymentów numerycznych generowanych przez plany OLH. Otrzymany analityczny wzór pozwala na szybką przybliżoną klasyfikację różnych rozwiązań projektowych pod kątem zdolności elementu do absorbowania energii zderzenia. Przedstawiono również zastosowanie sformułowania optymalizacji odpornościowej w zadaniu kalibracji parametrów modelu numerycznego używanego w komputerowej symulacji standardowych testów zderzeniowych. Odpowiedni dobór parametrów modelu MES pozwala na urealnienie wirtualnych modeli komputerowych poprzez zapewnienie zgodności z wynikami testów laboratoryjnych nie tylko odpowiedzi nominalnych, lecz również ich losowych rozrzutów.

Analiza niezawodności procesu tłoczenia blach

Metalowe elementy wytwarzane w procesie tłoczenia blach znajdują obecnie powszechne zastosowanie w różnych gałęziach przemysłu, przede wszystkim w przemyśle samochodowym, lotniczym oraz innych działach przemysłu środków transportu i przemysłu metalowego. Konsekwencją tak wielkiego zapotrzebowania na wytłaczane wyroby metalowe jest gwałtowny rozwój badań naukowych poświęconych problematyce głębokiego tłoczenia blach, jak również komputerowych technik symulacji tych procesów. Spośród odbywających się cyklicznie międzynarodowych konferencji poświęconych tej tematyce wymienić należy konferencje NUMISHEET [95] i NUMIFORM [25], a za wiodące pakiety obliczeniowe, które zorientowane są na analizę zagadnień tłoczenia, uznać należy AutoFORM, PAM-STAMP i LS-DYNA, zob. [251].

Doskonalone techniki symulacji numerycznych pozwalają na coraz dokładniejsze określenie deformacji oraz stanu naprężenia w każdym punkcie formowanej blachy w dowolnej chwili trwania procesu. Przeprowadzana analiza skończenie elementowa nieliniowego procesu tłoczenia uwzględniać musi duże odkształcenia sprężysto-plastyczne, zjawisko kontaktu, tarcia i nieliniowe związki konstytutywne. Bardzo ważną cechą oprogramowania do symulacji tłoczenia blach jest możliwość przeprowadzenia oceny jakości wytłoczki. Odkształcenia, jakich doznaje materiał podczas procesu tłoczenia, nie mogą bowiem osiągnąć dowolnie dużych wartości, wzrasta wtedy ryzyko wystąpienia takich niepożądanych zjawisk jak: utrata stateczności blachy, pękanie, fałdowanie, zmniejszenie grubości ścianki itp. Ogólnie, jako wadę wytłoczki uznać można każde odstępstwo od zadanych warunków geometrycznych i wytrzymałościowych powodujące, że wytłoczka nie może być użyta do zadań, do których była pierwotnie zaprojektowana. Zdolność do plastycznego kształtowania, która nie powoduje powstania wad określana jest jako tłoczność blachy. Definiowana ona być może przez różne kryteria, z których najczęściej stosowanym są tzw. wykresy odkształceń

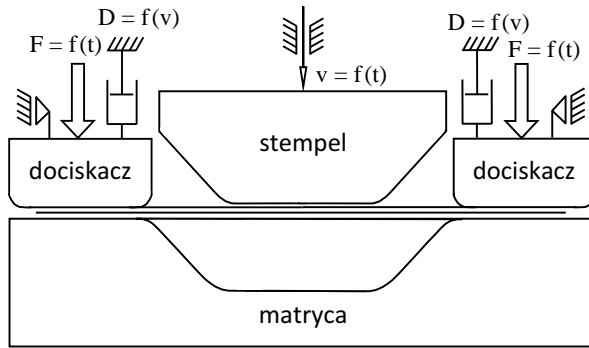
granicznych (WOG), (ang. *forming limit diagrams*) zob. np. [136], wyznaczane doświadczalnie lub teoretycznie.

Inną, bardzo istotną cechą realistycznego modelowania procesów formowania blach jest ocena wpływu, jaki na badane odpowiedzi układu blacha-tłoczniki ma losowy charakter parametrów opisujących to zagadnienie. Rzeczywiste wartości stałych materiałowych i współczynników tarcia, geometria narzędzi, siła docisku czy położenie progów ciągowych cechują się pewnym rozrzutem wokół założonych wielkości nominalnych. Świadomość konsekwencji, jakie nieść mogą odmienne od założonych parametry procesu tłoczenia, wydaje się być zatem niezbędnym składnikiem dobrego projektowania. Szczególnie ważną jest odpowiedź na pytanie, jak losowość wartości parametrów wpływa na możliwość powstania w dowolnym punkcie blachy odkształceń przekraczających dopuszczalne odkształcenia graniczne.

Poniżej, w oparciu o wyniki opublikowane w pracy [125], przedstawiony zostanie sposób oszacowania prawdopodobieństwa awarii procesu głębokiego tłoczenia blachy. Zaproponowana definicja funkcji awarii wykorzystuje wykres odkształceń granicznych. Ze względu na charakterystyczne dla tej funkcji problemy z różniczkowalnością, w analizie niezawodności wykorzystano adaptacyjną metodę Monte Carlo. Szczególnie dokładnie omówiono zagadnienie stochastycznego modelowania współczynników tarcia dla kontaktu blachy z matrycą, dociskaczem i stemplem oraz wpływu wzajemnej korelacji wartości tych współczynników na prawdopodobieństwo awarii. Problematyka analizy niezawodności procesów tłoczenia rozwijana była następnie w pracach [124,207] w kierunku wykorzystania w rozwiązaniu wielomianowej ważonej aproksymacji funkcji granicznej.

7.1. Modelowanie procesów tłoczenia blach

Na typowym schemacie stanowiska do tłoczenia blach, pokazanym na rys. 7.1, obok arkusza blachy wyróżnić można tłoczniki: stempel, matrycę i dociskacz. W modelu komputerowym tłoczniki traktowane są zazwyczaj jak ciała sztywne, wymuszające deformację blachy na skutek oddziaływań kontaktowych dla zadanego ruchu stempla. Poza kontaktem, realistycznie modelowanie procesów tłoczenia wymaga uwzględnienia zjawiska tarcia oraz plastycznej anizotropii blachy wywołanej walcowaniem. Blacha, której materiał podlega zarówno dużym przemieszczeniom jak i dużym odkształceniom, dyskretyzowana jest najczęściej przy pomocy powłokowych elementów skończonych. Elementy te muszą umożliwiać



Rys. 7.1. Schemat tłoczenia z dociskaczem.

odpowiednie odwzorowanie złożonej deformacji tłoczonego arkusza blachy. Powinny być także efektywne obliczeniowo.

Numeryczna symulacja tłoczenia realizowana być może w oparciu o różne sformułowania mechaniczne tego procesu. Do symulacji metodą elementów skończonych najczęściej stosuje się bądź sformułowanie w ramach teorii plastycznego płynięcia (ang. *flow approach*) bądź dynamiczne lub quasi-statyczne sformułowanie bazujące na mechanice ciała stałego, zob. [31, 190, 206]. Model dynamiczny wymaga odpowiedniej kontroli efektów bezwładnościowych.

W przedstawionym w dalszej części rozdziału przykładzie, tłoczenie blach rozpatrywano jako zadanie dynamiki stosując jawny schemat całkowania równań ruchu MES. Ze względu na wysoką efektywność obliczeń, szczególnie w przypadku złożonych modeli, podejście to zyskało dużą popularność w symulacji procesów tłoczenia. W metodzie jawnego całkowania na typowym kroku czasowym rozwiązywane są zdyskretyzowane równania ruchu, które zapisane dla bieżącej konfiguracji układu dane są następująco:

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} = \mathbf{p} - \mathbf{f}, \quad (7.1)$$

gdzie \mathbf{M} i \mathbf{C} są odpowiednio macierzą mas i macierzą tłumienia, $\ddot{\mathbf{q}}$ i $\dot{\mathbf{q}}$ są wektorami węzłowych przyspieszeń i prędkości, a \mathbf{p} i \mathbf{f} są wektorami sił zewnętrznych i wewnętrznych. Elementowy wektor sił wewnętrznych ma postać

$$\mathbf{f}^{(e)} = \int_{V^{(e)}} \mathbf{B}^T \boldsymbol{\sigma} dV, \quad (7.2)$$

gdzie $\boldsymbol{\sigma}$ jest tensorem naprężeń Cauchy'ego, \mathbf{B} jest macierzą operatorów różniczkowych związków geometrycznych (zróżniczkowanych funkcji kształtu), a $V^{(e)}$ jest objętością elementu.

Równania (7.1) całkowane są względem czasu przy użyciu metody różnic centralnych. Jest to jawny schemat całkowania, w którym przemieszczenia \mathbf{q}_{n+1} w chwili t_{n+1} obliczane są na podstawie znanej konfiguracji w chwili t_n :

$$\ddot{\mathbf{q}}_n = \mathbf{M}_D^{-1}(\mathbf{p}_n - \mathbf{f}_n - \mathbf{C}\dot{\mathbf{q}}_n), \quad \text{gdzie } \mathbf{M}_D = \text{diag } \mathbf{M} \quad (7.3)$$

$$\dot{\mathbf{q}}_{n+1/2} = \dot{\mathbf{q}}_{n-1/2} + \ddot{\mathbf{q}}_n \Delta t_{n+1/2}, \quad \text{gdzie } \Delta t_{n+1/2} = \frac{1}{2}(\Delta t_n + \Delta t_{n+1}), \quad (7.4)$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \dot{\mathbf{q}}_{n+1/2} \Delta t_{n+1}. \quad (7.5)$$

W powyższym schemacie, prędkości przemieszczeń węzłowych wyznaczane są w połowie kroku czasowego, są zatem przesunięte w stosunku do punktu na osi czasu, w którym obliczany jest wektor przemieszczeń. Z uwagi na to, do obliczenia sił tłumienia w równaniu (7.3) przyjmuje się $\dot{\mathbf{q}}_n \approx \dot{\mathbf{q}}_{n-1/2}$. Duża efektywność podejścia wykorzystującego jawne całkowanie równań ruchu opiera się na możliwości użycia diagonalnej macierzy mas \mathbf{M}_D , dzięki czemu nie ma potrzeby kosztownego numerycznie odwracania macierzy. Z drugiej strony, ze względu na warunkową stabilność jawnego schematu całkowania, ograniczeniem tej metody jest konieczność stosowania dużo mniejszych kroków czasowych niż np. w przypadku niejawnego schematu Newmarka.

W zaprezentowanym dalej przykładzie trójwymiarowego tłoczenia, do modelowania blachy użyto trójkątnego elementu powłoki cienkiej o jedynie translacyjnych stopniach swobody w węźle. Przy założeniu kinematycznej hipotezy Kirchhoffa dla powłok, prędkość odkształcenia w dowolnym punkcie powłoki można wyrazić poprzez parametry definiujące prędkość odkształcenia powierzchni środkowej, rozłożonego na stan odkształcenia membranowego oraz stan odkształcenia zgięciowego. W sformułowaniu elementu wykorzystano standardowe liniowe funkcje kształtu. Od swojej angielskiej nazwy, *basic shell triangle*, element oznaczony jest akronimem BST. Opracowano go przede wszystkim do zastosowań w symulacji procesów tłoczenia [189]. Brak obrotowych stopni swobody sprawia, że BST jest rozwiązaniem bardzo efektywnym numerycznie, a co za tym idzie szczególnie przydatnym do symulacji dużych przemysłowych problemów głębokiego tłoczenia blach. Więcej informacji na temat sformułowania elementu BST znaleźć można w pracy [206].

Przyjęto sprężysto-plastyczny model materiału zakładając addytywną dekompozycję tensora prędkości odkształcenia na część sprężystą $\dot{\mathbf{e}}^e$ i część plastyczną $\dot{\mathbf{e}}^p$,

$$\dot{\mathbf{e}} = \dot{\mathbf{e}}^e + \dot{\mathbf{e}}^p. \quad (7.6)$$

W celu obliczenia wartości naprężeń $\boldsymbol{\sigma}$ w tłoczonej blasze konieczne jest uwzględnienie odpowiedniego warunku uplastycznienia. Zgodnie z hipotezą kinematyczną Kirchhoffa, tensor $\boldsymbol{\sigma}$ opisuje płaski stan naprężenia, tak więc w kryterium

uplastycznienia brane są pod uwagę jedynie składowe σ_{11} , σ_{22} i σ_{12} . Płaskiemu stanowi naprężenia oraz anizotropii własności plastycznych materiału, która powstaje w procesie walcowania blachy, odpowiada wprowadzony przez Hilla w 1948 roku, zob. [90], warunek

$$\sigma_{11}^2 + \frac{R_0(1+R_{90})}{R_{90}(1+R_0)}\sigma_{22}^2 - \frac{2R_0}{1+R_0}\sigma_{11}\sigma_{22} + \frac{(1+2R_{45})(R_0+R_{90})}{R_{90}(1+R_0)}\sigma_{12}^2 - \bar{\sigma}^2 = 0, \quad (7.7)$$

gdzie współczynniki Lankforda R_0 , R_{45} i R_{90} (dla kierunków określonych kątami 0° , 45° i 90° względem kierunku walcowania blachy) służą opisowi anizotropii materiału, a składowe σ_{11} , σ_{22} i σ_{12} dane są w osiach ortotropii. Dla przypadku materiału izotropowego ($R_0 = R_{45} = R_{90} = 1$) równanie (7.7) redukuje się do kryterium Hubera-Misesa. Występujący we wzorze symbol $\bar{\sigma}$ oznacza granicę plastyczności, której wartość określa się najczęściej ze wzoru

$$\bar{\sigma} = K(a + \bar{\varepsilon})^n, \quad (7.8)$$

gdzie K , a i n są stałymi materiałowymi, a $\bar{\varepsilon}$ jest efektywnym odkształceniem plastycznym. Jeśli przyjąć, że właściwości materiału powłoki w jej płaszczyźnie stycznej są niezależne od kierunku, natomiast są one znacząco różne od właściwości w kierunku poprzecznym (normalnym) do powierzchni powłoki, wtedy kryterium Hilla (7.7) ma następującą postać:

$$\sigma_{11}^2 + \sigma_{22}^2 - \frac{2\bar{R}}{1+\bar{R}}\sigma_{11}\sigma_{22} + \frac{2(1+\bar{R})}{1+\bar{R}}\sigma_{12}^2 - \bar{\sigma}^2 = 0, \quad (7.9)$$

gdzie \bar{R} jest średnim współczynnikiem Lankforda danym wzorem

$$\bar{R} = \frac{R_0 + 2R_{45} + R_{90}}{4}. \quad (7.10)$$

O ile warunek (7.9) dobrze oddaje zachowanie metali charakteryzujących się współczynnikiem $\bar{R} > 1$, o tyle źle reprezentuje powierzchnię płynięcia w przypadku materiałów o małych wartościach współczynnika Lankforda $\bar{R} < 1$, np. dla aluminium. Aby usunąć te wady Hill zaproponował, zob. [91], udoskonaloną postać kryterium uplastycznienia postaci

$$C_1|\sigma_1 + \sigma_2|^M + C_2|\sigma_1 - \sigma_2|^M = \bar{\sigma}^M. \quad (7.11)$$

gdzie σ_1 i σ_2 są naprężeniami głównymi, wykładnik $M > 1$, a C_1 i C_2 są stałymi materiałowymi charakteryzującymi anizotropię. Kryterium to dla przypadku anizotropii normalnej dane jest następująco:

$$\frac{1}{2(1+\bar{R})}|\sigma_1 + \sigma_2|^M + \frac{1+2\bar{R}}{2(1+\bar{R})}|\sigma_1 - \sigma_2|^M = \bar{\sigma}^M. \quad (7.12)$$

Warunki (7.9) i (7.12), jak też warunek uplastycznienia dla przypadku ogólnej anizotropii [92], zaimplementowano w programie STAMPACK, zob. [208], którego użyto w omówionym dalej przykładzie analizy niezawodności do symulacji tłoczenia blachy.

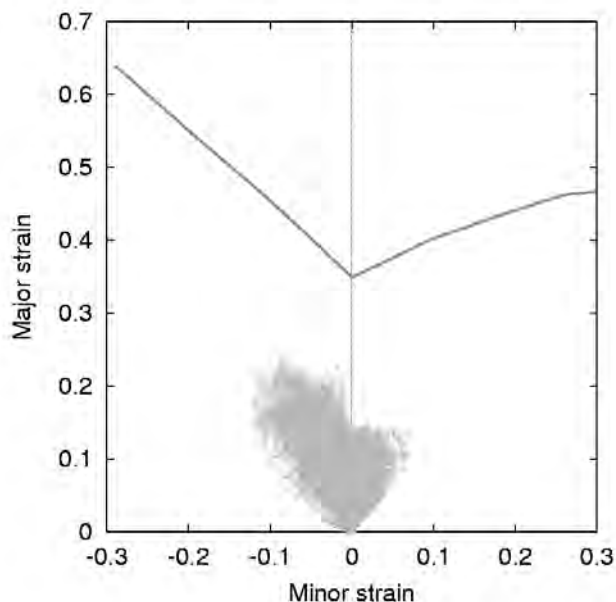
Zjawisko kontaktu odgrywa zasadniczą rolę we właściwym modelowaniu procesów tłoczenia. W trakcie formowania blacha styka się ze stemplem, matrycą oraz dociskaczem. Siły kontaktu uwzględniane są w wektorze sił zewnętrznych \mathbf{p} . Wynikające z kontaktu ograniczenia w kierunku normalnym do powierzchni styku narzucone są stosując metodę funkcji kary, a do obliczenia sił tarcia użyto modelu tarcia Coulomba, zob. [206].

7.1.1. Wykresy odkształceń granicznych

Wykresy odkształceń granicznych, sporządzane na płaszczyźnie odkształceń głównych, są niezwykle użytecznym narzędziem do analizy wyników procesu tłoczenia, a w szczególności do oceny jakości wytłoczki pod kątem możliwości wystąpienia pęknięć lub fałdowania. Dla danego rodzaju blachy krzywa odkształceń granicznych (zwana także graniczną krzywą tłoczności) reprezentuje graniczne wartości odkształceń głównych materiału, których przekroczenie z dużym prawdopodobieństwem prowadzić może do powstania wady wyrobu.

Na typowym WOG (zob. rys. 7.2) oś odciętych odpowiada mniejszemu z odkształceń głównych, a oś rzędnych większemu. Najczęściej stosowaną miarą odkształceń są tu odkształcenia rzeczywiste (logarytmiczne). Widoczne na wykresie punkty oznaczają stany odkształcenia w wybranych miejscach blachy, zazwyczaj w węzłach siatki MES lub w punktach Gaussa elementów. Położenie punktów odnoszone jest do krzywej odkształceń granicznych. Na niektórych WOG, dla łatwiejszej interpretacji stanów deformacji blachy, oprócz granicznej krzywej tłoczności umieszcza się również linie odpowiadające jednoosiowemu ścisnaniu i rozciąganiu, dwuosiowemu rozciąganiu i czystemu ścinaniu. Krzywa odkształceń granicznych rozdziela stany odkształcenia, które prowadzą do niestateczności lub pęknięcia blachy (punkty powyżej krzywej) od stanów dopuszczalnych (poniżej krzywej).

Pomimo, iż opracowano wiele teoretycznych metod wyznaczania krzywych odkształceń granicznych, nadal w powszechnym użyciu są metody eksperymentalne. Najczęściej stosowaną metodą jest opracowana przez Nakazimę [182] i rozwinięta przez Hackera [88] metoda dwuosiowego rozciągania prostokątnych pasów blachy przy pomocy półkolistego stempla. Zmieniając szerokość pasa użyć można różne stany dwuosiowych odkształcenia.

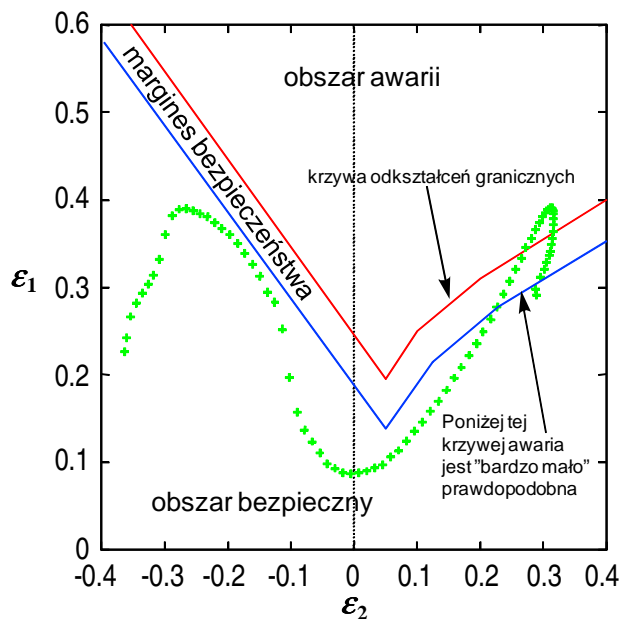


Rys. 7.2. Typowy wykres odkształceń granicznych. Zbiór punktów odpowiadających odkształceniom w węzłach siatki MES w całości znajduje się poniżej granicznej krzywej tłoczności.

Spośród najczęściej stosowanych metod analitycznych służących określeniu granicznej krzywej tłoczności wymienić należy metodę Marciniaka i Kuczyńskiego [163]. Zaproponowana przez nich metoda opiera się na założeniu, że powstawanie bruzd na powierzchni wytłoczki spowodowane jest początkową niejednorodnością grubości blachy.

Ważnym jest, aby zdawać sobie sprawę z pewnych niedoskonałości WOG. Przede wszystkim odkształcenia graniczne są zależne od historii procesu deformacji, zob. [136]. Wytwarzając wytłoczki o złożonej geometrii trudno zagwarantować, że ścieżka zmian odkształceń w danym punkcie blachy odpowiadać będzie tej, którą zaprogramowano w stosunkowo prostym teście laboratoryjnym. Ponadto, na proces tłoczenia wpływ ma szereg parametrów charakteryzujących się pewną niepewnością, tzn. takich, których wartości nie mogą być precyzyjnie określone. Przykładem tego typu parametrów są np. współczynniki tarcia. Nawet niewielka zmiana zastosowanego sposobu smarowania może mieć znaczny wpływ na historię deformacji prowadząc ostatecznie do innego położenia punktu na WOG. Zmienność parametrów materiałowych, w szczególności granicy plastyczności czy parametru wzmocnienia ma również istotne znaczenie na przebieg procesu tłoczenia. Wszystko to sprawia, iż położenie punktu na wykresie odkształceń

granicznych cechuje pewien losowy rozrzut. Graniczną krzywą tłoczności należy zatem traktować jako nieostre ograniczenie zakresu dopuszczalnych kombinacji odkształceń głównych i dlatego, w celu zapewnienia jakości wytłoczki, w projektowaniu rozpatruje się pewną strefę przejściową – “margines bezpieczeństwa” poniżej krzywej, zob. rys. 7.3. Zaznaczony na rysunku obszar bezpieczny należy interpretować jako taki obszar, gdzie wystąpienie wady w postaci pęknięcia blachy jest bardzo mało prawdopodobne, natomiast obszar niepowodzenia (awarii) to ten, gdzie z wysokim prawdopodobieństwem można założyć uszkodzenie tłoczonego wyrobu.



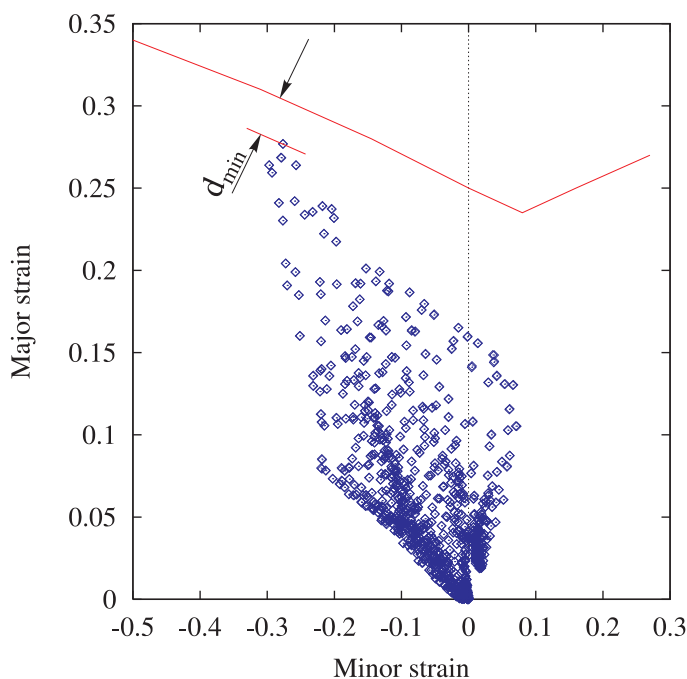
Rys. 7.3. Wykres odkształceń granicznych (przyjęto $\varepsilon_2 < \varepsilon_1$) z zaznaczonym marginesem bezpieczeństwa. Powszechna praktyka polega na wprowadzeniu pewnego arbitralnie dobranego obszaru przejściowego poniżej krzywej odkształceń granicznych.

Zarówno stochastyczny charakter parametrów jak i niepewność definicji krzywej odkształceń granicznych wydają się przemawiać za tym, by standardową komputerową symulację tłoczenia blach uzupełnić o analizę probabilistyczną, a w szczególności o analizę niezawodności. Analiza ta daje możliwość odpowiedzi na zasadnicze pytanie – jakie jest prawdopodobieństwo pojawienia się wyrobów wadliwych w procesie produkcji?

W dalszej części rozdziału zaproponowano sformułowanie oraz algorytm rozwiązania tego zadania.

7.2. Definicja funkcji granicznej zadania analizy niezawodności

Formułując zadanie analizy niezawodności procesu tłoczenia blach, naturalnym wyborem przy definiowaniu funkcji granicznej jest wykorzystanie wykresu odkształceń granicznych. Tak jak pokazano na rys. 7.4, wartość funkcji granicznej przyjmuje się równą najmniejszej odległości (ze znakiem) punktów reprezentujących odkształcenia główne w poszczególnych elementach skończonych od krzywej odkształceń granicznych (ze względu na przyjętą liniową interpolację przemieszczeń, odkształcenia wewnątrz elementu są stałe). Pojęcie odległości ze znakiem rozumieć należy następująco: w przypadku gdy wszystkie rozpatrywane punkty znajdują się poniżej krzywej odkształceń granicznych, wtedy wartość funkcji granicznej jest dodatnia, jeśli jednak część punktów położona jest powyżej krzywej, wtedy największa odległość punktu z tej grupy, wzięta ze znakiem minus, przyjmowana jest jako wartość funkcji granicznej.



Rys. 7.4. Definicja funkcji granicznej jako odległości od krzywej odkształceń granicznych.

W zależności od realizacji wektora zmiennych losowych zmienia się stan odkształcenia w poszczególnych punktach wytłoczki, a w konsekwencji, “chmura”

punktów na wykresie odkształceń granicznych zmienia swój kształt i położenie. Co za tym idzie, różne punkty WOG (odpowiadające różnym elementom skończonym) decydować mogą o wartości funkcji granicznej. Biorąc pod uwagę odcinkowo liniowy charakter granicznej krzywej tłoczności, a także szum numeryczny generowany przez algorytm symulacji tłoczenia (por. dyskusja na str. 137) wydaje się oczywistym, że zdefiniowana przez nas funkcja graniczna nie jest różniczkowalna.

Jak wcześniej wspomniano, kształt i położenie krzywej odkształceń granicznych dane są jedynie w sposób przybliżony. Wynika to m.in. z nieznanomości rzeczywistej ścieżki deformacji, warunków tarcia i stałych materiałowych. Dość powszechną praktyką w analizie procesu tłoczenia jest wprowadzanie tzw. marginesu bezpieczeństwa, zob. rys. 7.3, czyli przejściowej strefy o arbitralnie założonej szerokości pomiędzy obszarem bezpiecznym, a krzywą odkształceń granicznych. Punkty na WOG mogą być wtedy klasyfikowane w zależności od ich położenia względem marginesu bezpieczeństwa jako prowadzące do pęknięcia wytłoczki z dużym lub małym prawdopodobieństwem. Te mało precyzyjne określenia - typowe dla języka naturalnego - skłaniają ku temu, aby w modelowaniu niepewności położenia krzywej odkształceń granicznych użyć koncepcji zbioru rozmytego. Ponieważ granica pomiędzy obszarem bezpiecznym a obszarem awarii nie jest ostra, dlatego zdarzenie niesprzyjające (realizacja zmiennych losowych, w wyniku której proces tłoczenia kończy się niepowodzeniem) należy rozpatrywać jako rozmyte zdarzenie losowe, a prawdopodobieństwo awarii jako zbiór rozmyty.

Stosując standardową notację (zob. dodatek A) wektor zmiennych losowych występujących w stochastycznym opisie procesu tłoczenia blachy oznaczać będziemy jako $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$, natomiast oryginalną przestrzeń zmiennych losowych symbolem \mathcal{X} . Funkcję przynależności rozmytego obszaru awarii $\tilde{\mathcal{D}}_f$, będącą uogólnieniem funkcji charakterystycznej obszaru awarii $I_{\mathcal{D}_f}$ (A.55), można zdefiniować w następujący sposób (zob. WOG na rys. 7.3):

$$m_{\tilde{\mathcal{D}}_f}(\mathbf{x}) = \begin{cases} 1 & \text{jeśli niektóre punkty na WOG znajdują się} \\ & \text{powyżej marginesu bezpieczeństwa,} \\ \in (0, 1) & \text{jeśli wszystkie punkty na WOG są poniżej krzywej} \\ & \text{odkształceń granicznych, a punkt najbliższy} \\ & \text{jest wewnątrz marginesu bezpieczeństwa,} \\ 0 & \text{jeśli wszystkie punkty na WOG znajdują się} \\ & \text{poniżej marginesu bezpieczeństwa.} \end{cases} \quad (7.13)$$

Funkcja $m_{\tilde{\mathcal{D}}_f}$ przyjmuje wartości z przedziału $[0, 1]$ w zależności od odległości

najbliższego punktu od krzywej odkształceń granicznych. Przy założeniu liniowej zależności, sytuacja gdy punkt ten położony jest w środku marginesu bezpieczeństwa prowadzi do $m_{\tilde{\mathcal{D}}_f}(\mathbf{x}) = 0.5$, jeśli zaś położony jest w odległości równej 0.25 szerokości marginesu bezpieczeństwa od krzywej odkształceń granicznych wtedy $m_{\tilde{\mathcal{D}}_f}(\mathbf{x}) = 0.75$ itd. Zgodnie z definicją Yagera, zob. [246, 260], funkcja przynależności rozmytego prawdopodobieństwa awarii m_{P_f} tworzona jest następująco:

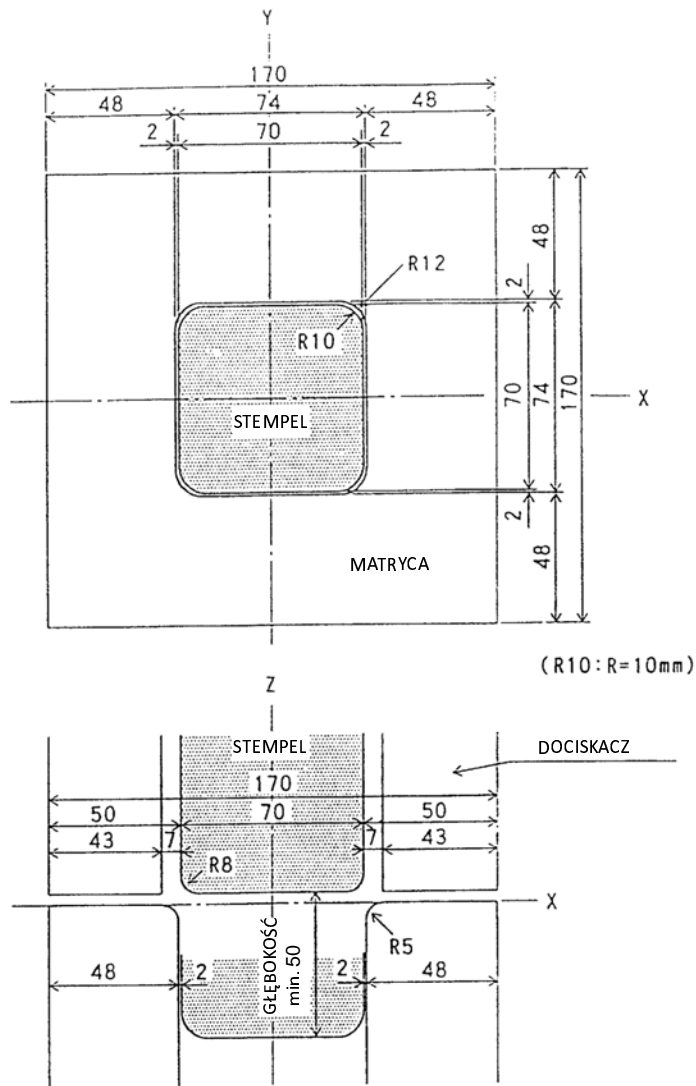
1. Na WOG wybieranych jest kilka krzywych odpowiadających różnym wartościom $m_{\tilde{\mathcal{D}}_f}$, w szczególności tym, które ograniczają margines bezpieczeństwa.
2. Każda z krzywych służy następnie do zdefiniowania funkcji granicznej zadania analizy niezawodności. Obliczane są wartości prawdopodobieństw awarii.
3. Otrzymane prawdopodobieństwa awarii należą do rozmytego prawdopodobieństwa awarii w stopniu określonym przez $m_{\tilde{\mathcal{D}}_f}$.

Tak więc, prawdopodobieństwu awarii obliczonemu dla dolnej krzywej odkształceń granicznych (ograniczającej margines bezpieczeństwa od dołu) odpowiada $m_{P_f} = 0$, a prawdopodobieństwu uzyskanemu przyjmując górną krzywą odkształceń granicznych $m_{P_f} = 1$. Postępowanie to zastosowane zostało w poniższym przykładzie analizy niezawodności tłoczenia.

7.3. Analiza niezawodności dla przykładu testowego konferencji NUMISHEET

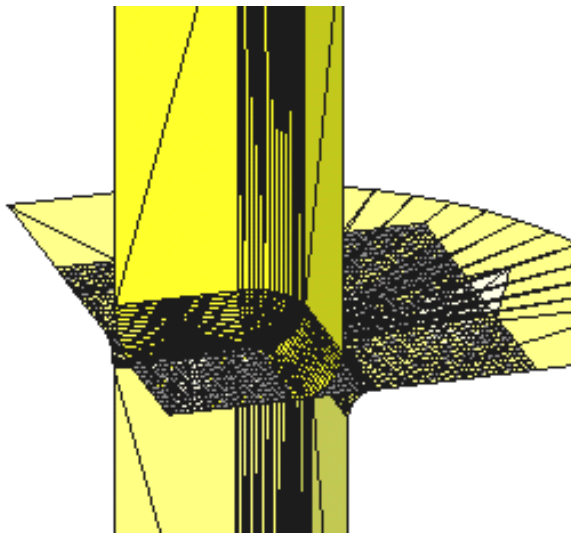
Przedmiotem analizy jest proces tłoczenia aluminiowego naczynia o przekroju kwadratowym. Na schematycznym rysunku 7.5 przedstawiono kształt i wymiary narzędzi: stempla, matrycy i dociskacza. Omawiany przykład był jednym z zadań testowych konferencji NUMISHEET w 1993 roku [162]. Przyjęto następujące dane materiałowe: początkowa grubość blachy 0.81 mm, moduł Younga $E = 71$ GPa, współczynnik Poissona $\nu = 0.33$, krzywa zależności jednoosiowego rzeczywistego naprężenia od rzeczywistego odkształcenia $\sigma = 576.79(0.01658 + \varepsilon^p)^{0.3593}$ MPa, współczynnik tarcia $\mu = 0.162$. Siła dociskacza wynosi 19.6 kN. Założono poprzeczną anizotropię blachy, a w kryterium uplastycznienia Hilla (7.12) przyjęto średni współczynnik Lankforda $\bar{R} = 0.64$. Naczynie tłoczone jest z blachy o wymiarach 150 mm \times 150 mm.

Wykorzystano symetrię modelu i w celu poprawy efektywności obliczeń w symulacji MES rozpatrywano jedynie jego ćwiartkę. Do dyskretyzacji użyto 1800



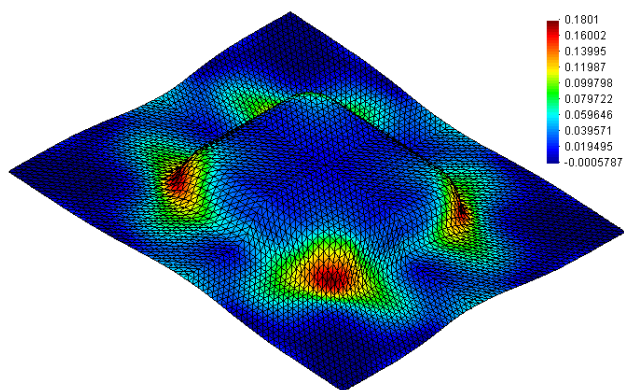
Rys. 7.5. Tłoczenie naczynia o przekroju kwadratowym – geometria tłoczników.

elementów BST, zob. rys. 7.6. Przyjęto prędkość ruchu stempla równą 10 m/s. Aby zredukować efekty dynamiczne, analizę kontynuowano po osiągnięciu założonej głębokości tłoczenia wprowadzając dodatkowe siły tłumienia. W [162] zamieszczono wyniki eksperymentów przeprowadzonych dla głębokości tłoczenia równej 15 mm. Dla takiej właśnie głębokości, na rys. 7.7 pokazano wybrane wyniki symulacji przeprowadzonej programem STAMPACK. Na powierzchni

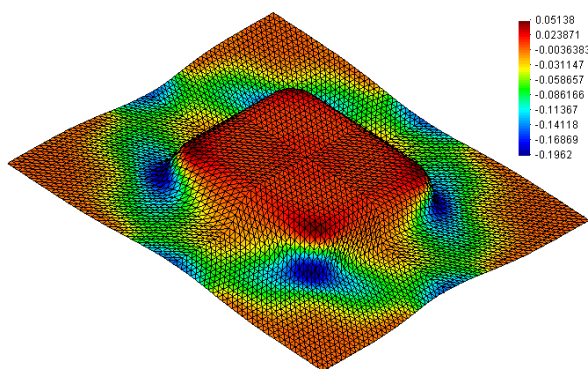


Rys. 7.6. Tłoczenie naczynia o przekroju kwadratowym – dyskretyzacja MES.

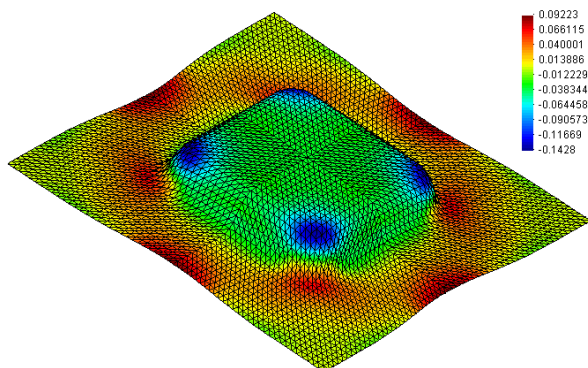
wytłoczki przedstawiono mapę warstwicową głównych odkształceń logarytmicznych oraz odkształceń poprzecznych (w kierunku normalnym do powierzchni). Na rys. 7.8 porównano rozkłady odkształceń poprzecznych wzdłuż dwóch płaszczyzn symetrii wytłoczki otrzymane w wyniku symulacji MES i doświadczalnie (na podstawie [162]). Widoczna jest dobra zgodność tych wyników. Odkształcenia główne w elementach siatki pokazano na WOG na rys. 7.9. Punkty znajdują się w znacznej odległości od krzywej odkształceń granicznych, co może świadczyć, iż jest mało prawdopodobne aby przy tłoczeniu na głębokość 15 mm i dla nominalnych wartości pozostałych parametrów powstała wada wytłoczki. Powstanie pęknięcia zaobserwowano zwiększając głębokość tłoczenia do 20 mm. Odpowiedni WOG oraz mapy odkształceń logarytmicznych przedstawiono na rysunkach 7.10 i 7.11. Jak można zobaczyć na rys. 7.10, jeden z punktów znajduje się bardzo blisko krzywej odkształceń głównych, co oznacza wysokie ryzyko pojawienia się pęknięcia blachy. Spostrzeżenie to potwierdzone jest przez testy laboratoryjne, gdzie wada wytłoczki wystąpiła dla przesunięcia stempla na głębokość 19 mm, zob. [181]. W przypadku tłoczenia na głębokość 20 mm, zwiększając nieznacznie wartość współczynnika tarcia otrzymuje się WOG, który pokazano na rys. 7.12. Porównując z wykresem 7.10 można zauważyć, że więcej punktów znajduje się w sąsiedztwie krzywej odkształceń granicznych, a jeden z nich jest wewnątrz obszaru awarii. Potwierdza to dużą wrażliwość otrzymywanych rozkładów odkształceń na losowe imperfekcje wartości parametrów procesu tłoczenia.



a)

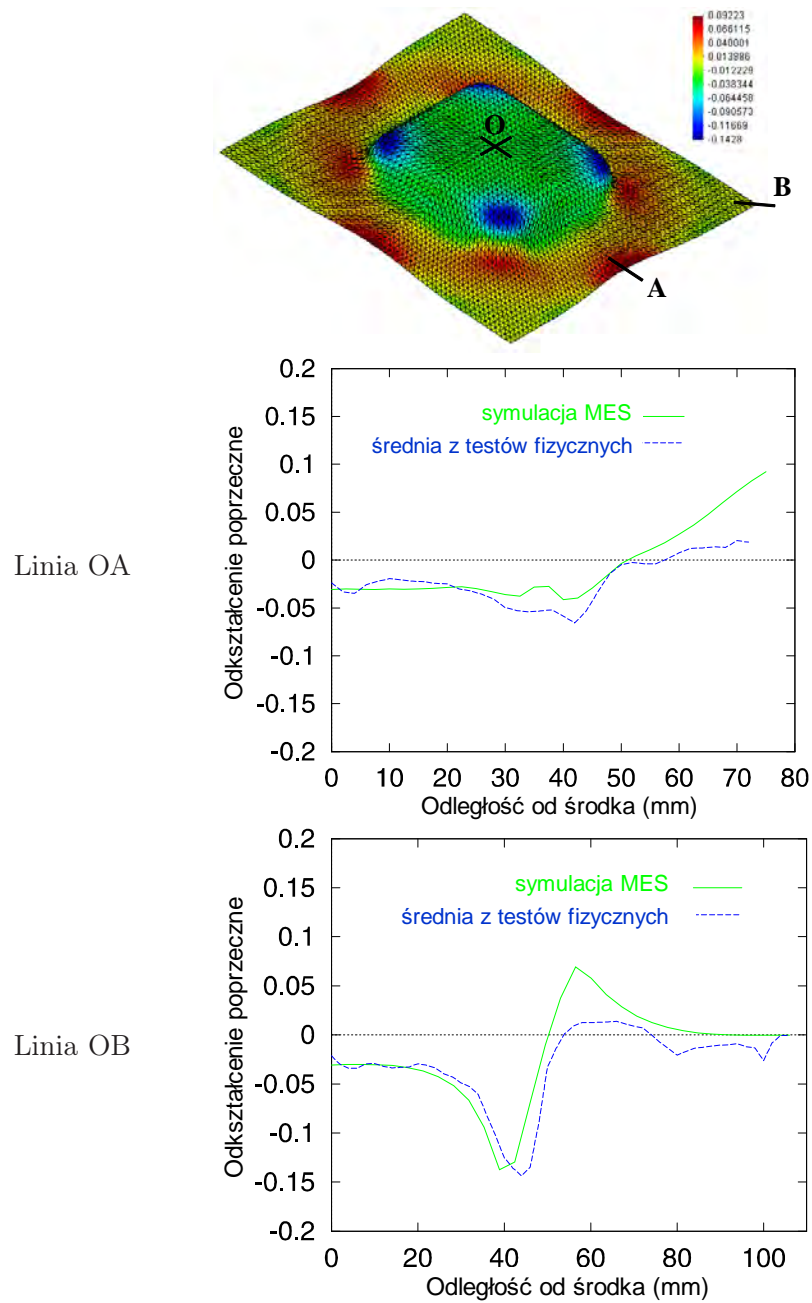


b)

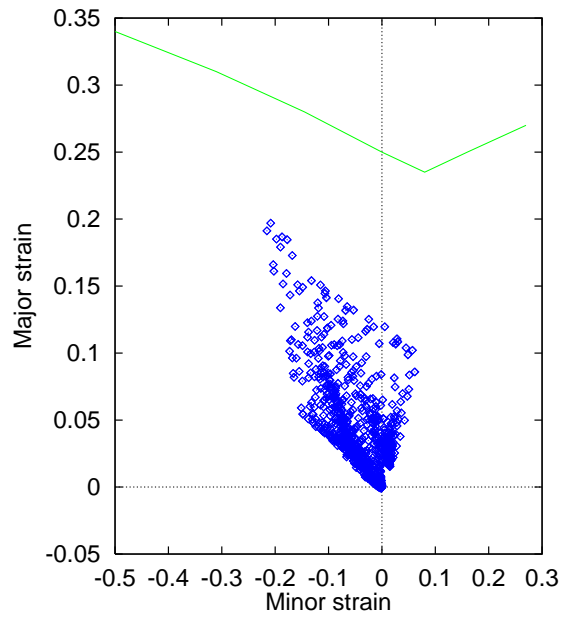


c)

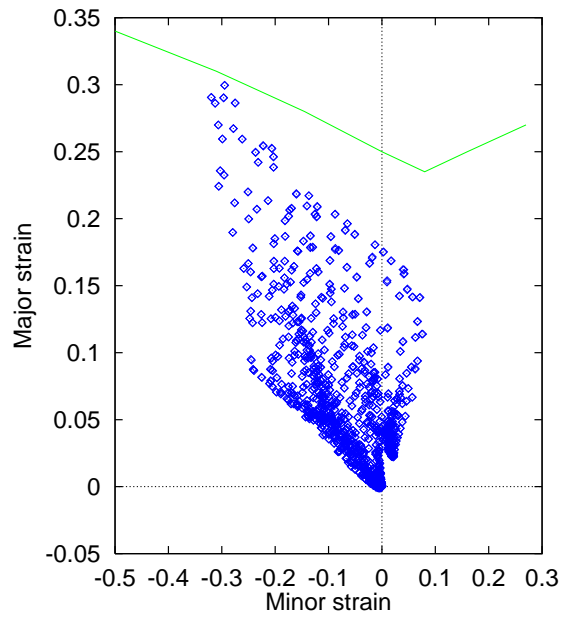
Rys. 7.7. Tłoczenie naczynia o przekroju kwadratowym – rozkład odkształceń logarytmicznych dla głębokości tłoczenia 15 mm: a) większe odkształcenie główne, b) mniejsze odkształcenie główne, c) odkształcenie poprzeczne.



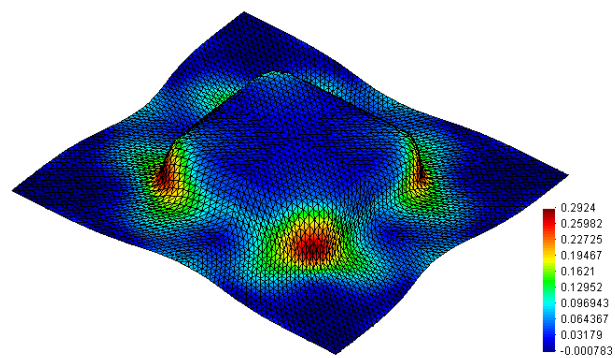
Rys. 7.8. Tłoczenie naczynia o przekroju kwadratowym – porównanie rozkładów odkształceń poprzecznych w wybranych przekrojach wytłoczki, otrzymanych doświadczalnie oraz na podstawie symulacji MES.



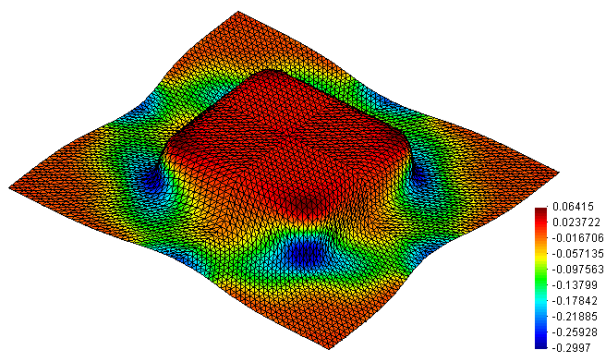
Rys. 7.9. Tłoczenie naczynia o przekroju kwadratowym – wykres odkształceń granicznych dla głębokości tłoczenia 15 mm.



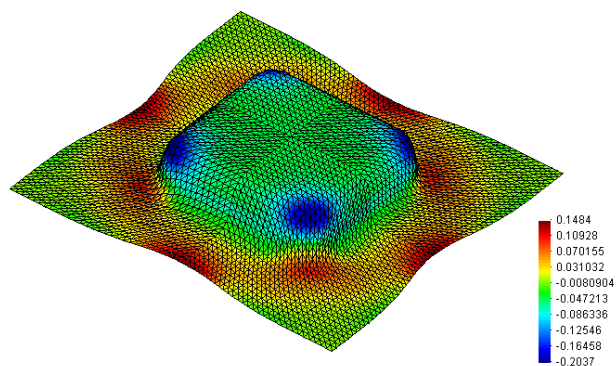
Rys. 7.10. Tłoczenie naczynia o przekroju kwadratowym – wykres odkształceń granicznych dla głębokości tłoczenia 20 mm.



a)

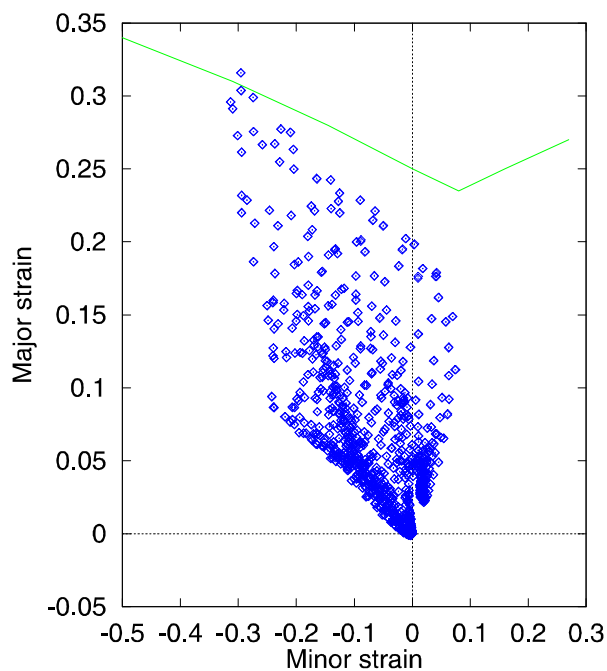


b)



c)

Rys. 7.11. Tłoczenie naczynia o przekroju kwadratowym – rozkład odkształceń logarytmicznych dla głębokości tłoczenia 20 mm: a) większe odkształcenie główne, b) mniejsze odkształcenie główne, c) odkształcenie poprzeczne.



Rys. 7.12. Tłoczenie naczynia o przekroju kwadratowym – wykres odkształceń granicznych dla głębokości tłoczenia 20 mm. Nominalny współczynnik tarcia $\mu = 0.162$ został zwiększony do wartości $\mu = 0.2$.

Stochastyczny model procesu tłoczenia tworzy 6 zmiennych losowych. Są to: początkowa grubość blachy, parametr K wiązków konstytutywnych (7.8), siła dociskacza oraz trzy współczynniki tarcia – pomiędzy blachą a stemplem, blachą a matrycą i blachą a dociskaczem. Założono, że wszystkie zmienne mają rozkład lognormalny, a ich wartości oczekiwane i odchylenia standardowe podano w tab. 7.1.

Tablica 7.1. Tłoczenie naczynia o przekroju kwadratowym – zmienne losowe zadania analizy niezawodności.

Zmienna	Wartość oczekiwana	Odchylenie standardowe	Opis
X_1	0.81 mm	0.04 mm	Początkowa grubość
X_2	$5.7679 \cdot 10^8$ Pa	$3.0 \cdot 10^7$ Pa	Stała materiałowa
X_3	4900 N	50 N	Siła dociskacza
X_4	0.162	0.015	Wsp. tarcia blacha - stempel
X_5	0.162	0.015	Wsp. tarcia blacha - matryca
X_6	0.162	0.015	Wsp. tarcia blacha - dociskacz

Przyjęto, że zmienne losowe X_4, X_5, X_6 reprezentujące współczynniki tarcia są skorelowane, a jednakowy współczynnik korelacji wynosi ρ . Symulacja procesu tłoczenia przeprowadzana jest do głębokości 15 mm. Ze względu na charakter funkcji granicznej, do oszacowania wartości prawdopodobieństwa awarii użyto, omówionej w podrozdziale A.4.3, adaptacyjnej metody importance sampling, zob. (A.75). Wyniki analizy zebrano w tab. 7.2.

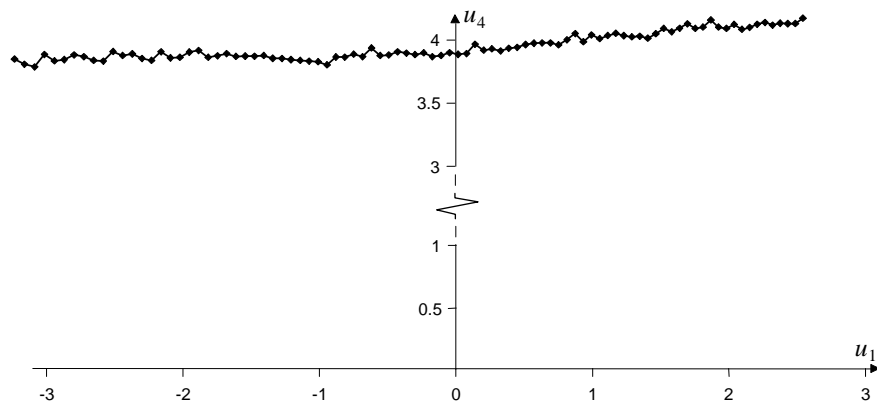
Tablica 7.2. Wpływ korelacji współczynników tarcia na prawdopodobieństwo awarii P_f .

ρ	P_f	$\beta = -\Phi^{-1}(P_f)$	Liczba wywołań funkcji granicznej	$\nu_{\hat{P}_f}$
0	$4.07 \cdot 10^{-7}$	4.93	9264	5%
0.3	$6.51 \cdot 10^{-6}$	4.36	4326	5%
0.9	$4.62 \cdot 10^{-5}$	3.91	3325	5%
1	$5.06 \cdot 10^{-5}$	3.89	2826	5%

Okazuje się, że wielkość korelacji współczynników tarcia ma znaczący wpływ na otrzymywane wartości prawdopodobieństwa awarii. Zmieniając współczynnik korelacji od $\rho = 0$ (zmienne nieskorelowane) do $\rho = 1$ (pełna korelacja), uzyskuje się rozrzut wartości P_f w przedziale, którego granice różnią się o dwa rzędy wielkości, od $4.07 \cdot 10^{-7}$ do $5.06 \cdot 10^{-5}$.

Ponieważ w większości przypadków można uznać, że własności powierzchni tłoczonej blachy mają charakter jednorodny i nie zmieniają się w zależności od miejsca na arkuszu blachy, dlatego bardziej realistycznym wydaje się założenie o silnej korelacji współczynników tarcia. W tym przypadku prawdopodobieństwo awarii w postaci pęknięcia wytłoczki jest równe około 10^{-5} , co oznacza, że dla przyjętych danych należy się spodziewać mniej więcej jednej wady na 10^5 wytłoczek. Tak mała wartość prawdopodobieństwa awarii (o wiele mniejsza od obserwowanych w praktyce) wynikać może ze źle dobranego modelu stochastycznego bądź zbyt "optymistycznych" założeń dotyczących krzywej odkształceń granicznych. Celowe jest zatem zbadanie wpływu nieprecyzyjnego określenia stanu awarii, zgodnie z zaproponowaną wcześniej koncepcją rozmytego obszaru awarii.

Przed omówieniem tego zagadnienia warto przyjrzeć się bliżej własnościom wykorzystywanej w zadaniu analizy niezawodności funkcji granicznej. Aby zilustrować trudności jakie napotyka się stosując standardowe algorytmy poszukiwania punktu projektowego (zob. podrozdział A.3), na rys 7.13 pokazano przecięcie z płaszczyzną u_1, u_4 powierzchni granicznej $h(\mathbf{u}) = 0$ otrzymanej z transformacji oryginalnej powierzchni $g(\mathbf{x}) = 0$ do przestrzeni standardowej \mathcal{U} , zob. podrozdział A.2. Wykres utworzono dla przypadku $\rho = 1$ (tylko jedna



Rys. 7.13. Powierzchnia graniczna - przecięcie z płaszczyzną u_1, u_4 w przestrzeni \mathcal{U} .

zmienna losowa używana jest do opisu wszystkich współczynników tarcia) oraz przyjmując X_2 i X_3 równe ich wartościom oczekiwany. Na podstawie analizy nieregularnego (“zębatego”) charakteru wykresu na rys. 7.13, należy stwierdzić, że nie jest możliwe bezpośrednie zastosowanie jakichkolwiek gradientowych algorytmów poszukiwania punktu projektowego \mathbf{u}^* . Z drugiej strony, wydaje się, że w rozpatrywanym zakresie zmienności, funkcja z dużą dokładnością przybliżana być może liniową powierzchnią odpowiedzi.

Nie korzystając z uprzedniej aproksymacji funkcji granicznej, w celu zminimalizowania wpływu szumu, w zadaniu lokalizacji punktu projektowego postanowiono zastosować gradientowy algorytm ARF (zob. A.3.1), gdzie w różnicowym schemacie obliczania pochodnych funkcji granicznej użyto “dużych” perturbacji (wielokrotnie większych niż w przypadku numerycznego różniczkowania funkcji klasy C^1). Prawdopodobieństwo awarii obliczono dla modelu stochastycznego składającego się ze wszystkich zmiennych wymienionych w tab. 7.1, ustalając współczynnik korelacji zmiennych X_4, X_5, X_6 równy 0.9 i przyjmując rozmiar perturbacji w schemacie różnic centralnych jako równy odchyleniu standardowemu danej zmiennej.

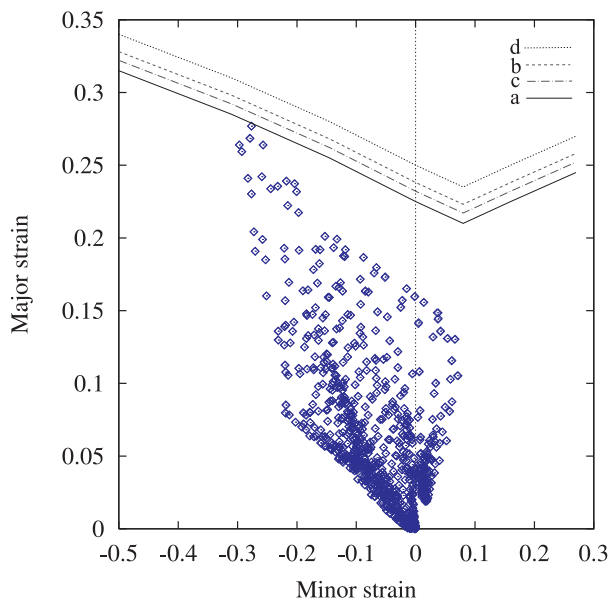
Używając zgrubnych warunków zbieżności, po 82 obliczeniach funkcji granicznej (wywołaniach programu STAMPACK) otrzymano następujące wyniki aproksymacji FORM: $\beta_{\text{FORM}} = 3.87$, $P_f = 5.35 \cdot 10^{-5}$, zob. (A.6). Porównując z odpowiednią wartością z tab. 7.2 rezultat ten można ocenić jako dość dokładny. Niestety, trudno na podstawie tego przykładu zaproponować stosowanie metody dużych perturbacji jako uniwersalnego podejścia do problemów analizy niezawodności procesów tłoczenia. Nieunikniony błąd zaokrągleń, charakterystyczny

dla dużych wartości perturbacji, może przekreślać przydatność tej metody gdy funkcja graniczna $h(\mathbf{u})$ jest silnie nieliniowa. Za każdym razem konieczne jest zatem wstępne studium porównawcze z metodą bardziej dokładną, taką jak adaptacyjna metoda importance sampling.

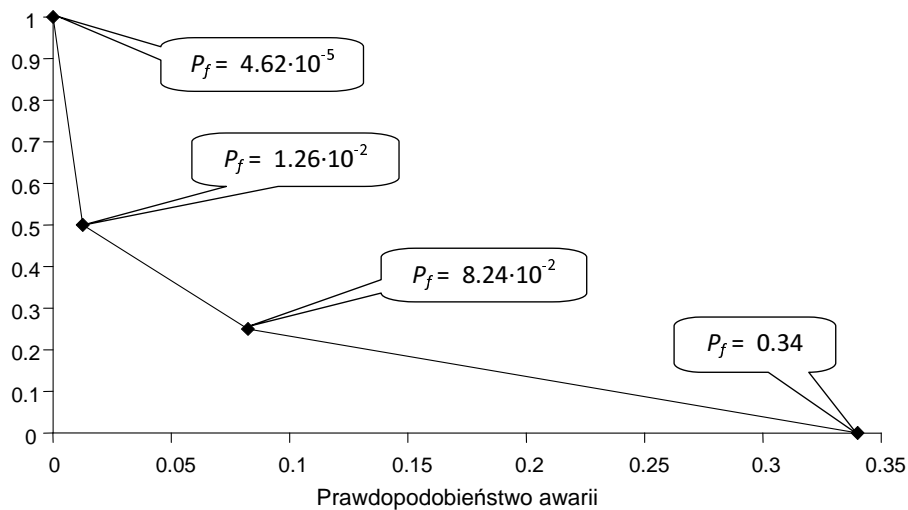
Chcąc sprawdzić jak duży wpływ na niezawodność procesu tłoczenia mogą mieć trudności z precyzyjnym określeniem granicy obszaru awarii, postanowiono oszacować prawdopodobieństwo awarii przyjmując w definicji funkcji granicznej kilka krzywych odkształceń granicznych zawartych w marginesie bezpieczeństwa pokazanym na rys. 7.14. Szerokość tego marginesu ustalono równą około 10% odkształcenia granicznego odpowiadającego płaskiemu stanowi odkształcenia ($\varepsilon_2 = 0$). O ile punkty na wykresie 7.9 znajdowały się daleko od krzywej odkształceń granicznych, o tyle teraz punkty reprezentujące najbardziej odkształcone elementy znajdują się w sąsiedztwie dolnej granicy marginesu bezpieczeństwa. Na rys. 7.14 zaznaczono 4 krzywe znajdujące się w obrębie marginesu bezpieczeństwa powstałe poprzez przesunięcie oryginalnej krzywej odkształceń granicznych. Związane są one z tzw. α -przekrojami rozmytego obszaru awarii (zob. [246]), a odpowiadające im prawdopodobieństwa awarii stanowią będą granice α -przekrojów rozmytego prawdopodobieństwa awarii.

Pokazane na rys. 7.14 krzywe (a), (b), (c) i (d) wybrano w następujący sposób: (a) i (d) stanowią dolne i górne ograniczenie marginesu bezpieczeństwa, (b) znajduje się w jego środku, a (c) w połowie odległości pomiędzy (a) i (b). Analizę niezawodności przeprowadzono dla współczynnika korelacji zmiennych X_4, X_5, X_6 równego 0.9. Aby przedstawić prawdopodobieństwo awarii w postaci zbioru rozmytego tworzymy funkcję przynależności, która dla prawdopodobieństw odpowiadających krzywym (a), (b), (c) i (d) przyjmuje odpowiednio wartości 0, 0.5, 0.25 i 1, zob. rys. 7.15.

Powstaje oczywiście pytanie, jak interpretować wyniki przedstawione w postaci rozmytego prawdopodobieństwa awarii? W przeciwieństwie do standardowej analizy niezawodności, gdzie niepewności parametrów modelu opisywane są językiem teorii prawdopodobieństwa, interpretacja ta nie jest jednoznaczna. Obserwując charakter zmienności funkcji przynależności pokazanej na rys. 7.15 można stwierdzić, że termin “niezawodny” lub “nie prowadzący do wielu wad” lepiej opisuje ten proces tłoczenia niż określenie “cechujący się dużym prawdopodobieństwem niepowodzenia”, którego należałoby użyć wnioskując jedynie na podstawie $P_f = 0.34$, otrzymanego dla krzywej (a). Tak więc, najważniejszą dodatkową informacją jaką niesie ze sobą funkcja przynależności rozmytego prawdopodobieństwa awarii jest ocena ryzyka, które podejmuje się wybierając tą, a nie inną krzywą odkształceń granicznych. Taki dobór parametrów procesu



Rys. 7.14. Tłoczenie naczynia o przekroju kwadratowym – wykres odkształceń granicznych dla głębokości tłoczenia 15 mm. Margines bezpieczeństwa wprowadzony poniżej krzywej odkształceń granicznych (d) podzielono na kilka poziomów odpowiadających α -przekrojom rozmytego obszaru awarii: a - 0, c - 0.25, b - 0.5, d - 1.



Rys. 7.15. Tłoczenie naczynia o przekroju kwadratowym – funkcja przynależności rozmytego prawdopodobieństwa awarii.

tłoczenia aby proces ten był niezawodny względem funkcji granicznej definio-
wanej przy pomocy krzywej (a) może prowadzić do rozwiązań zbyt konserwa-
tywnych i nieuzasadnionych z punktu widzenia ponoszonych kosztów.

PODSUMOWANIE

Rozdział poświęcony jest w całości analizie niezawodności procesów tłoczenia
blach. Prezentowany materiał pochodzi z pracy Kleibera, Rojka i Stockiego
[125], w której po raz pierwszy zaproponowano wykorzystanie komputerowych
metod analizy niezawodności do oceny jakości procesów tłoczenia.

Podstawą do oszacowania prawdopodobieństwa awarii procesu jest znajo-
mość funkcji granicznej, którą zdefiniowano wykorzystując standardowy wykres
odkształceń granicznych. Ponieważ konsekwencją definicji tej funkcji jako odle-
głości zbioru punktów od łamanej jest jej nieróżniczkowalność, dlatego jako me-
todę analizy niezawodności zaproponowano adaptacyjną metodę Monte Carlo.
W związku z niemożnością precyzyjnego wyznaczenia krzywej odkształceń gra-
nicznych, często nie jest możliwe jednoznaczne określenie stanu awarii procesu
tłoczenia. Trudności te były powodem wprowadzenia nowej koncepcji rozmyte-
go prawdopodobieństwa awarii. Proces analizy niezawodności zilustrowano na
testowym przykładzie symulacji tłoczenia z konferencji NUMISHEET'93.

Wykazano, że prawdopodobieństwo awarii w postaci pojawienia się na po-
wierzchni wytłoczki bruzd, które następnie prowadzić mogą do powstania pęk-
nięć, bardzo silnie zależy od wartości korelacji współczynników tarcia na po-
szczególnych powierzchniach kontaktu. Wydaje się, że bezpiecznym rozwiąza-
niem jest przyjęcie w analizie niezawodności pełnej korelacji tych współczynni-
ków.

Biorąc pod uwagę obecne możliwości sprzętu komputerowego, adaptacyjna
metoda Monte Carlo okazała się być skutecznym narzędziem analizy niezawod-
ności, prowadząc do umiarkowanych nakładów obliczeniowych. Z drugiej strony,
ponieważ rozpatrywany przykład zawierał tylko 6 zmiennych losowych, a złożo-
ność kształtów wytłoczek jest zazwyczaj dużo większa, dlatego zrozumiałą jest
dalszy trend rozwoju metod analizy niezawodności procesów tłoczenia blach,
który w dużym stopniu opiera się na wykorzystaniu technik powierzchni odpo-
wiedzi, zob. [106, 207, 274]. Właśnie taki kierunek rozwoju przewidziany został
przez autorów w prekursorskiej pracy [125].

STAND - pakiet obliczeniowy do analizy niezawodności i optymalizacji konstrukcji

Rozwiązywanie zagadnień analizy niezawodności oraz optymalizacji odpornościowej, które przedstawiono w poprzednich rozdziałach, wymaga stworzenia wyspecjalizowanego oprogramowania łączącego zarówno efektywną implementację omówionych algorytmów, jak i walory użytkowe. Oprogramowanie takie musi umożliwiać łatwą komunikację z zewnętrznymi modułami obliczeniowymi (najczęściej z programami MES), a także uruchamianie analiz na komputerach o architekturze wieloprocesorowej bądź klastrach obliczeniowych.

W ciągu ostatnich trzydziestu lat analiza niezawodności konstrukcji była przedmiotem intensywnego rozwoju, przechodząc ewolucję od mało znanej dyscypliny akademickiej do okrzeplej i szeroko stosowanej metodologii wspierającej proces projektowania. Obecnie, dzięki rozwojowi wyspecjalizowanego oprogramowania, analiza niezawodności jest dostępnym narzędziem zarówno dla naukowców, jak też inżynierów. Wśród najpopularniejszych programów do analizy niezawodności konstrukcji wymienić należy programy ANSYS DesignXplorer, NESSUS, PERMAS-RA/STRUREL czy OpenSees. Omówieniu tych oraz innych programów poświęcono specjalny numer czasopisma Structural Safety [194].

Optymalizacja niezawodnościowa, a szczególnie optymalizacja odpornościowa w ujęciu zaprezentowanym w podrozdziale 2.4, nie jest jeszcze powszechnie dostępna w ramach istniejących komercyjnych pakietów optymalizacyjnych. Ze znanych programów, które oferują pewne wersje tej optymalizacji wymienić należy programy LS-OPT [229], Isight [130, 133] i modeFRONTIER [23, 196].

W większości istotnych z punktu widzenia praktyki inżynierskiej przypadków funkcje wchodzące w skład sformułowań optymalizacji odpornościowej oraz

analizy niezawodności nie mają postaci analitycznej lecz zdefiniowane są za pomocą wybranych wyników symulacji skończenie elementowych. Dlatego niezbędną cechą oprogramowania realizującego powyższe zadania jest ścisła integracja z wybranym programem MES lub stworzenie mechanizmów komunikowania się z zewnętrznymi modułami obliczeniowymi.

To pierwsze rozwiązanie, które wymaga znajomości kodu źródłowego, dostępne jest z oczywistych względów jedynie programistom rozwijającym dany pakiet MES, a w szczególności jego moduły pre i post. Przykładami takiego podejścia są programy ANSYS DesignXplorer [204] i PERMAS-RA [79]. Konsolidacja w ramach jednego kodu, która powoduje niezaprzeczalną poprawę efektywności obliczeń, wiąże się jednak z ograniczeniem zakresu możliwych zastosowań analizy niezawodności i optymalizacji do tych, których modelowanie możliwe jest przy pomocy zintegrowanego programu MES.

Dużo bardziej uniwersalną jest strategia zapewnienia komunikacji programu do analizy niezawodności/optymalizacji odpornościowej z wieloma zewnętrznymi modułami obliczeniowymi poprzez wsadowy mechanizm tworzenia danych wejściowych, uruchamiania analizy MES oraz odczytywania wyników. Warto tu wspomnieć o typowym rozwiązaniu, które stosowane jest np. w pakiecie analizy niezawodności STRUREL [242], polegającym na tworzeniu przez użytkownika interfejsu z modułem MES poprzez przygotowanie kodu źródłowego procedury obliczającej wartość funkcji granicznej dla danej realizacji zmiennych losowych. W procedurze tej musi znaleźć się obsługa modyfikacji danych, uruchamiania obliczeń i pobierania wartości odpowiednich odpowiedzi konstrukcji. Przygotowany kod jest następnie kompilowany oraz konsolidowany z programem głównym. Metoda ta jest niewątpliwie dość pracochłonna, a niezbędny wymóg umiejętności programistycznych użytkownika jeszcze dodatkowo ogranicza jej praktyczną użyteczność. Mając to na uwadze należy stwierdzić, że przyjazny użytkownikowi interfejs graficzny, który wspomaga wszystkie etapy realizacji zadania wydaje się być nieodzownym składnikiem dobrego oprogramowania. Jako przykłady doskonale zaprojektowanych interfejsów graficznych należy wymienić interfejsy programów NESSUS [249] (analiza niezawodności) oraz modeFRONTIER (optymalizacja).

Przydatność poszczególnych metod analizy niezawodności oraz optymalizacji konstrukcji zależy zarówno od specyfiki analizowanego zadania jak i wymaganej dokładności rozwiązania. Bogaty wybór metod i algorytmów jest zatem niezbędnym składnikiem dobrego oprogramowania realizującego obliczenia dla szerokiej klasy zagadnień dotyczących projektowania i analizy bezpieczeństwa konstrukcji inżynierskich. Pomimo, iż umiejętne dobranie algorytmu rozwiązującego może

znacząco ograniczyć konieczny nakład obliczeniowy, to w porównaniu do analizy deterministycznej nakład ten jest często o rzędy wielkości wyższy. Ponieważ w przypadku złożonych konstrukcji (modeli komputerowych) czas trwania pojedynczej symulacji skończenie elementowej wynosić może wiele godzin pracy procesora, dlatego zarówno w analizie niezawodności jak i optymalizacji odpornościowej symulacje te zastępuje się przez dużo efektywniejsze analityczne metamodele. W konsekwencji, to właśnie moduł powierzchni odpowiedzi (zob. rozdział 3), który umożliwia odpowiednie dopasowanie funkcji aproksymującej do eksperymentów komputerowych stanowi jeden z podstawowych elementów programu. Dodatkowo, spośród istotnych składników oprogramowania wymienić należy moduły realizujące zadanie automatycznej komunikacji z zewnętrznymi programami MES, a także, wspomniany już wygodny w obsłudze i czytelny graficzny interfejs użytkownika.

Duża liczba oraz zróżnicowanie koniecznych modułów sprawia, że tworzenie zaawansowanego programu do analizy niezawodności i optymalizacji odpornościowej konstrukcji jest poważnym wyzwaniem programistycznym. Architektura kodu zapewniać powinna możliwość jednoczesnego rozwijania go przez wielu programistów, łatwość utrzymania i wprowadzania modyfikacji. Odpowiednią platformą do realizacji tych celów wydaje się być programowanie zorientowane obiektowo. Co jest dość zaskakujące, pomimo oczywistych zalet oraz będąc obecnie dominującym paradygmatem programowania, podejście obiektowe nie jest jednak powszechną praktyką przy pisaniu programów do analizy niezawodności. Spośród najważniejszych, wymienionych w specjalnym numerze *Structural Safety* [194], tylko kilka najnowszych programów (ANSYS/DesignXplorer [204], PHIMECA-SOFT [146], Opensees [41]) zaprojektowano jako kody obiektowo zorientowane.

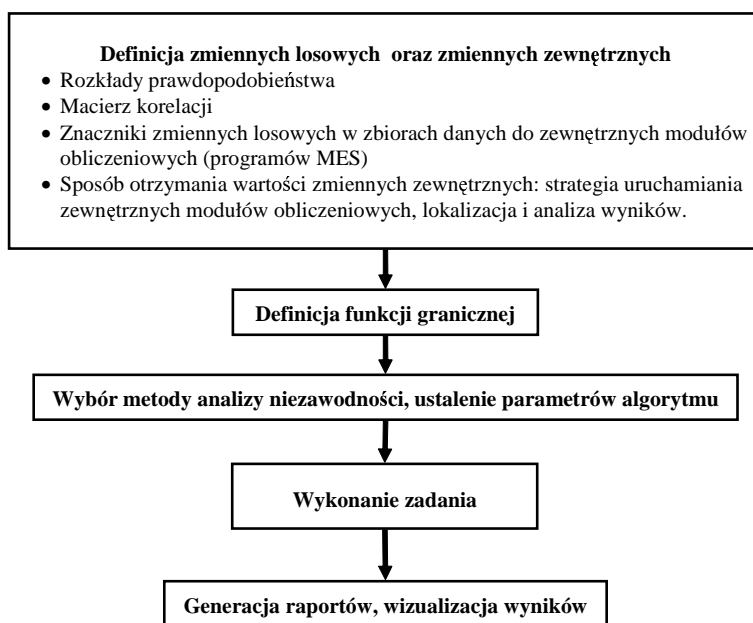
Obiektowo zorientowaną architekturę posiada również program STAND (*Stochastic ANalysis and Design*), który od 2004 roku rozwijany jest w Pracowni Niezawodności i Optymalizacji IPPT PAN. Kod ten, będąc w istocie zbiorem bibliotek w języku C++ grupujących klasy danych oraz algorytmy, zaprojektowany został jako wspólna platforma programowa służąca do implementacji oraz testowania metod szeroko pojętej analizy stochastycznej. W obecnej wersji programu zaimplementowano wiele standardowych jak również oryginalnych rozwiązań analizy niezawodności i optymalizacji odpornościowej konstrukcji. Ponadto, STAND wyposażony jest w mechanizmy interaktywnego definiowania zadania, jak również automatycznej integracji z zewnętrznymi modułami obliczeniowymi. Większość zadań obliczeniowych prezentowanych w niniejszej pracy zrealizowanych zostało właśnie przy pomocy tego oprogramowania.

Poniżej, na przykładzie STANDa omówione zostaną wybrane zagadnienia dotyczące komputerowej implementacji algorytmów analizy niezawodności i optymalizacji odpornościowej, architektury obiektowej programu, a także niezbędnych cech interfejsu użytkownika i sposobów komunikacji z programami MES. Zagadnienia integracji z pakietami skończenie elementowymi ABAQUS i RADIOSS omówiono analizując ponownie przykłady z podrozdziału 5.1.3 i rozdziału 6. Więcej informacji o programie STAND znaleźć można w artykule [234].

8.1. Analiza niezawodności konstrukcji

Najważniejsze etapy realizacji zadania niezawodności konstrukcji zaprezentowano w postaci schematu blokowego na rys. 8.1. Etapy te zostaną szczegółowo omówione w dalszej części rozdziału, w tym miejscu warto podać jedynie ich krótki opis.

Przedstawiona procedura rozpoczyna się od stworzenia modelu stochastycznego analizowanego zagadnienia. Użytkownik programu podaje parametry brzegowych rozkładów prawdopodobieństwa zmiennych losowych, a w przypadku zmiennych skorelowanych także współczynniki korelacji wzajemnej. Jednocześnie definiowane są tzw. *zmienne zewnętrzne*, będące niejawnymi funkcjami zmiennych losowych, których wartości otrzymuje się jako wynik działania innych (zewnętrznych) programów. Aby uzyskać wartość zmiennej zewnętrznej z , która odpowiada danej realizacji wektora zmiennych losowych, realizacja ta musi być zapisana w zbiorze danych wejściowych programu obliczającego wartość z . Wiąże się z tym konieczność podania informacji w jaki sposób realizacja danej zmiennej uaktualnia zbiór danych. Operacja ta realizowana jest przez opisany dalej mechanizm znaczników. Po zdefiniowaniu modelu stochastycznego, przy pomocy graficznego interfejsu programu STAND użytkownik wprowadza wzór funkcji granicznej, która zawierać może zarówno zmienne losowe jak i zmienne zewnętrzne. Następne etapy to wybór metody analizy niezawodności oraz uruchomienie obliczeń. Zadanie kończy się wygenerowaniem raportu zawierającego przede wszystkim wartości prawdopodobieństwa awarii oraz jego wrażliwości na parametry rozkładów prawdopodobieństwa zmiennych losowych. Wybrane wyniki oraz historia działania algorytmu przedstawione mogą być w formie wykresów.



Rys. 8.1. Główne etapy realizacji zadania analizy niezawodności konstrukcji

8.1.1. Definicja modelu stochastycznego

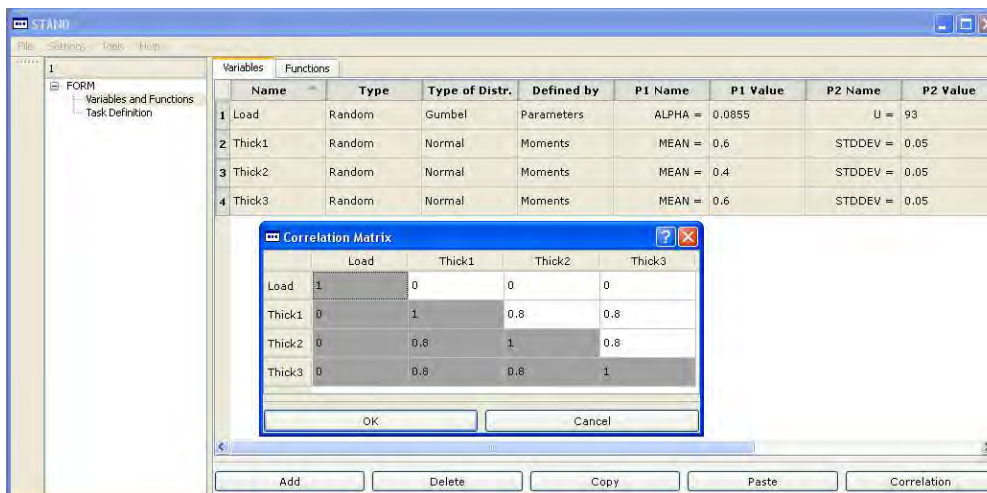
Jak już wspomniano, w definicji modelu stochastycznego uwzględnia się dwa rodzaje zmiennych. Są to (podstawowe) zmienne losowe oraz zmienne zewnętrzne (funkcje zmiennych losowych). Ze względu na ich odmienny charakter informacje konieczne do opisu zmiennych losowych i zmiennych zewnętrznych są też zasadniczo różne.

8.1.1.1. Zmienne losowe

Bardzo często ze względu na brak dostatecznej liczby danych eksperymentalnych trudno jest określić funkcję łącznej gęstości prawdopodobieństwa wektora zmiennych losowych \mathbf{X} , $f_{\mathbf{X}}(\mathbf{x})$. W przypadku zmiennych zależnych, w programie STAND funkcja $f_{\mathbf{X}}(\mathbf{x})$ aproksymowana jest za pomocą tzw. modelu Natafa, definiowanego przez brzegowe rozkłady prawdopodobieństwa zmiennych oraz macierz korelacji (zob. podrozdział A.2.4). Model Natafa pozwala na efektywną transformację, $\mathbf{U} = \mathbf{T}(\mathbf{X})$, oryginalnych zmiennych losowych \mathbf{X} do gaussowskiej

przestrzeni standardowej \mathbf{u} . Większość metod analizy niezawodności zaimplementowanych w programie STAND wykorzystuje własności tej przestrzeni, gdzie miara prawdopodobieństwa określona jest funkcją gęstości $f_{\mathbf{U}}(\mathbf{u}) = \prod_{i=1}^n \varphi(u_i)$, będącą iloczynem n jednowymiarowych standardowych gaussowskich funkcji gęstości prawdopodobieństwa, zob. dodatek A. Jeśli zmienne \mathbf{X} mają wielowymiarowy rozkład normalny, to model Natafa jest dokładny.

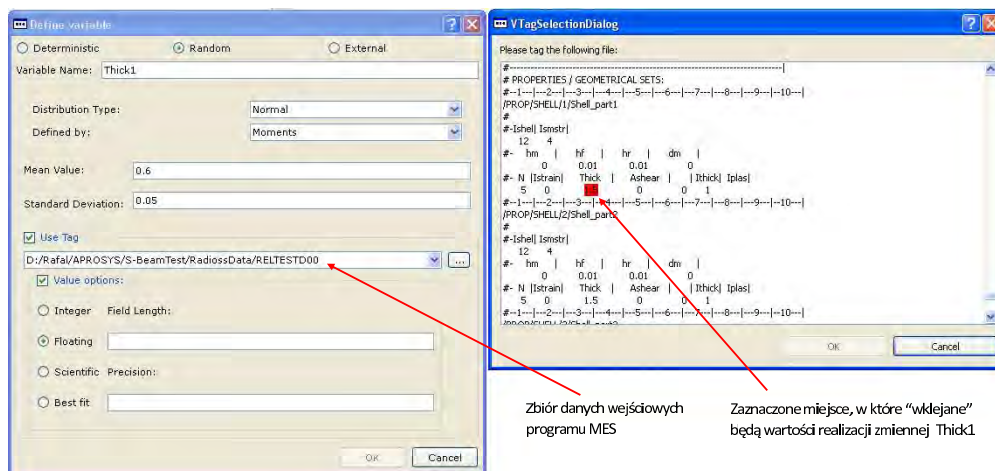
W obecnej wersji programu do opisu zmiennych losowych przyjmowane mogą być następujące rozkłady prawdopodobieństwa: jednostajny, normalny, lognormalny, wykładniczy, Rayleigha, Gumbela, Frecheta i Weibulla. Dany rozkład można zdefiniować za pomocą momentów statystycznych (podając wartość oczekiwaną oraz odchylenie standardowe) bądź też przy pomocy parametrów funkcji gęstości. Domyślnie zakłada się, że zmienne losowe są nieskorelowane. Jeśli tak nie jest, niezerowe wartości współczynników korelacji można wprowadzić w dedykowanym oknie dialogowym, jak to pokazano na rys. 8.2.



Rys. 8.2. Widok graficznego interfejsu służącego do definiowania zmiennych losowych.

Niezależnie od metody, oszacowanie wartości prawdopodobieństwa awarii wymaga wielokrotnego obliczenia wartości funkcji granicznej dla różnych realizacji wektora \mathbf{X} . W większości istotnych z punktu widzenia praktyki przypadków wiąże się to z koniecznością uruchamiania zewnętrznych programów analizy skończenie elementowej. W celu umożliwienia automatycznej komunikacji programu STAND z modułami zewnętrznymi zastosowano prosty mechanizm wymiany danych. Realizacje zmiennych losowych generowane przez algorytmy

analizy niezawodności zastępują odpowiednie wartości w plikach wejściowych programów zewnętrznych. Za pośrednictwem graficznego interfejsu użytkownika, każdej zmiennej można przypisać tzw. znacznik, lokalizujący pole w tekstowym zbiorze wejściowym, gdzie wpisywana będzie wartość realizacji zmiennej, zob. rys. 8.3. Oczywiście połączenie takie możliwe jest wtedy, gdy program MES może być wywoływany w trybie wsadowym oraz gdy zarówno zbiory danych wejściowych jak też zbiory z wynikami mają format tekstowy. Jak to zostanie pokazane w przykładach zaprezentowanych na końcu rozdziału, mechanizm ten okazał się wystarczający do komunikacji z używanymi w niniejszej pracy komercyjnymi pakietami MES.



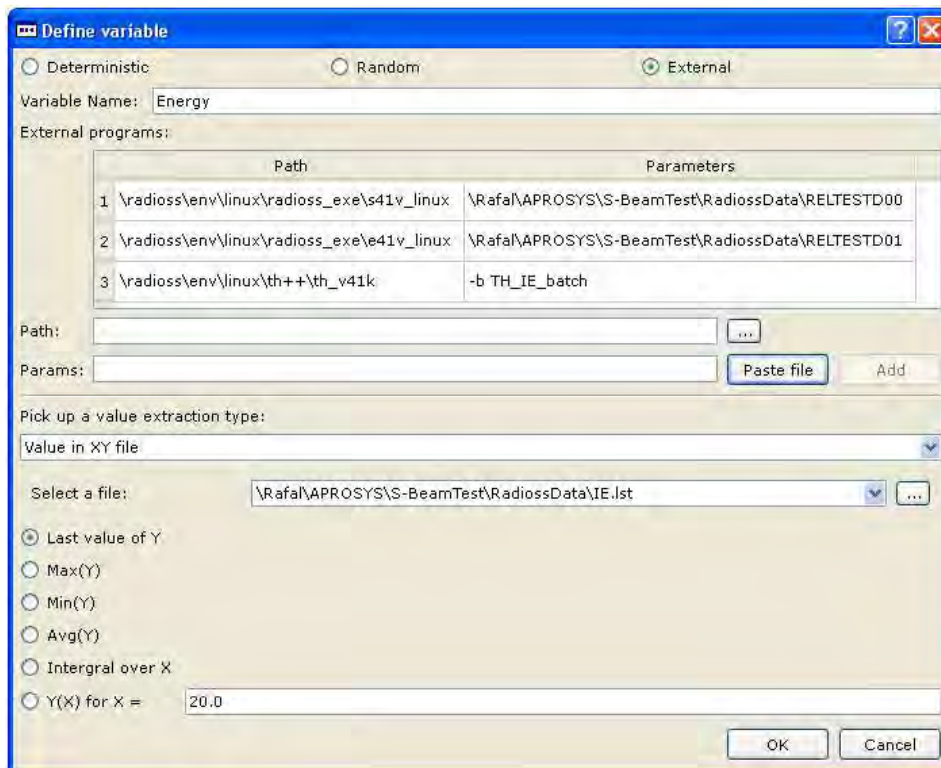
Rys. 8.3. Definicja znacznika - określenie pola w zbiorze danych wejściowych, w które zapisywane będą realizacje zmiennej losowej.

8.1.1.2. Zmienne zewnętrzne

Wartości zmiennych zewnętrznych odczytywane są ze zbiorów tekstowych zawierających wyniki działania zewnętrznych programów obliczeniowych. Dlatego w definicji zmiennej tego typu znaleźć musi się sekwencja wywołań programów, które w rezultacie doprowadzą do otrzymania wartości zmiennej dla danej realizacji x wektora \mathbf{X} . Przykład takiej definicji pokazano na rys. 8.4. Podane tam polecenia wykonywane w trybie wsadowym zawierają listy argumentów, w skład których wchodzi nazwy zbiorów z danymi specyfikowane wcześniej przy okazji definiowania zmiennych losowych. Użytkownik wybiera następnie metodę uzyskania wartości zmiennej ze zbioru wynikowego. Wartość ta lokalizowana być

może przez bezwzględne “współrzędne” pola tekstowego w pliku, przez połączenie względem określonej etykiety lub też jako wartość skalarna otrzymana z analizy historii zmian danej wielkości w czasie.

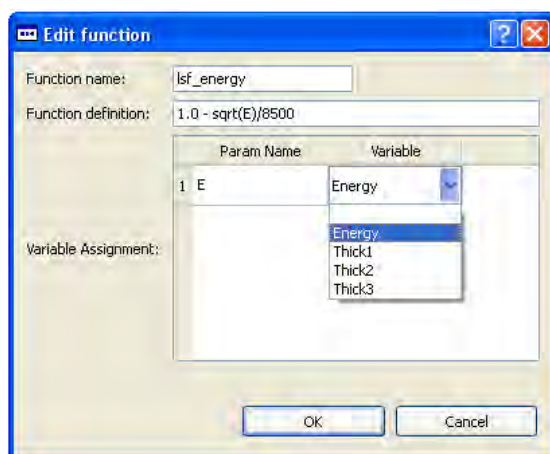
Bardzo często te same programy używane są do obliczenia wartości wielu zmiennych zewnętrznych. Na przykład, wynikiem symulacji MES mogą być przemieszczenia konstrukcji, odkształcenia, naprężenia lub wartości temperatur, które wykorzystuje się następnie w definicji funkcji granicznej. W celu poprawy efektywności analizy, identyfikuje się zmienne zewnętrzne o takiej samej sekwencji wywołań programów, aby uniknąć wielokrotnego uruchamiania obliczeń dla tych samych danych. Strategia ta jest w pełni zautomatyzowana i nie wymaga wcześniejszej deklaracji użytkownika.



Rys. 8.4. Okno interfejsu użytkownika służące do definicji zmiennych zewnętrznych. W celu otrzymania wartości zmiennej **Energy** należy najpierw uruchomić sekwencję trzech wymienionych programów, otrzymując zbiór wyników zawierający punkty do wykresu obrazującego zmianę badanej wielkości w czasie. Następnie jako wartość zmiennej zewnętrznej **Energy** przyjmuje się ostatnią wartość ze stabilizowanej zależności.

8.1.2. Definicja funkcji granicznej

W programie STAND wzór funkcji granicznej wprowadzany jest w standardowym zapisie matematycznym (podobnie jak np. w programie MATLAB) jako zależność od zmiennych losowych i zmiennych zewnętrznych. Definicja zawierać może większość typowych funkcji matematycznych. Przy wpisywaniu wzoru nie jest konieczne używanie takich samych nazw jak te użyte w modelu stochastycznym. Symbole zmiennych występujące we wzorze funkcji lokalizowane są automatycznie przez parser, a użytkownik dokonuje jedynie odpowiedniego przypisania na podstawie listy dostępnych zmiennych losowych i zewnętrznych, zob. rys. 8.5. Jeśli w definicji funkcji granicznej znajdują się zmienne zewnętrzne, to wykonywana jest identyfikacja, od których zmiennych losowych zależą one w sposób niejawny. Odbywa się to poprzez sprawdzenie, które z tych zmiennych modyfikują zbiory danych występujące jako parametry wywołań programów z definicji zmiennych zewnętrznych.

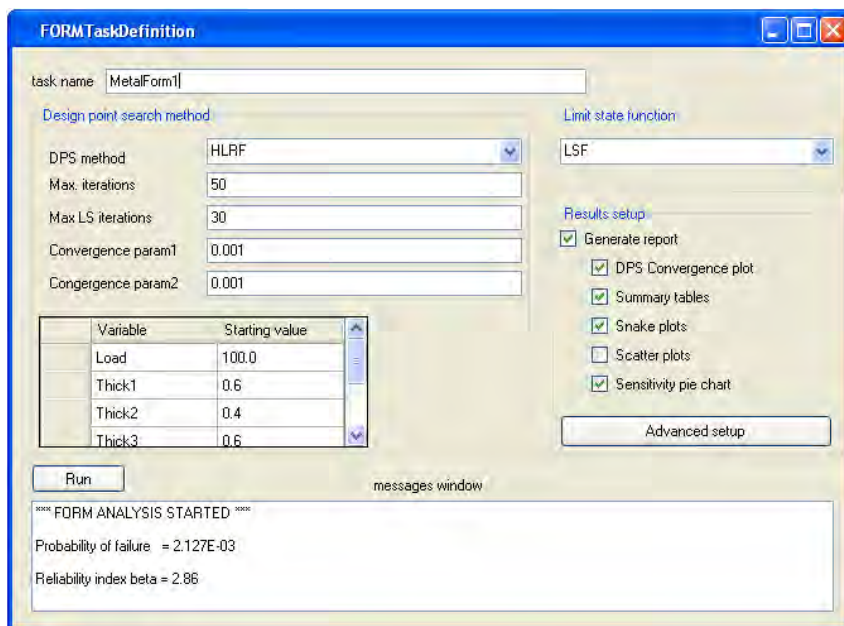


Rys. 8.5. Definicja funkcji granicznej. Przypisanie parametrom funkcji, uprzednio wprowadzonych zmiennych losowych i zewnętrznych.

8.1.3. Wybór algorytmu i uruchomienie analizy

Na rys. 8.6 pokazano przykładowy wygląd okna graficznego interfejsu programu STAND przeznaczonego do wyboru algorytmu analizy niezawodności. Użytkownik specyfikuje tu nazwę zadania, wybiera metodę analizy oraz ustala wartości jej parametrów sterujących, takich jak współczynnik zbieżności, maksymalna liczba iteracji itp. Funkcja graniczna wybierana jest z listy wszystkich

funkcji zdefiniowanych podczas sesji z programem. W przypadku metod, które wymagają lokalizacji punktu projektowego do wyboru są następujące algorytmy: HLRF, ARF i Z3M (zob. dodatek A.3 oraz podrozdział 5.1). Opcje dotyczące post-processingu pozwalają na ustalenie, które wyniki analizy niezawodności oraz jakiego typu wykresy mają znaleźć się w raporcie końcowym. Przycisk Run uruchamia analizę. Poniżej, w polu komunikatów wyświetlane są najważniejsze rezultaty, takie jak wartość prawdopodobieństwa awarii, a także, jeśli to możliwe, diagnozowane są problemy, które mogą wystąpić w czasie rozwiązywania zadania.



Rys. 8.6. Interfejs graficzny przeznaczony do wyboru wartości parametrów algorytmu analizy niezawodności oraz do uruchamiania zadania.

Jak to już wielokrotnie podkreślano, analiza niezawodności z użyciem dużych modeli skończenie elementowych jest z reguły zadaniem kosztownym obliczeniowo. Ponieważ o czasie trwania całej analizy decyduje przede wszystkim czas trwania pojedynczej symulacji MES oraz liczba takich symulacji, dlatego kluczowym elementem efektywnego procesu szacowania prawdopodobieństwa awarii jest odpowiednia strategia realizowania zewnętrznych zadań MES.

Na wielu etapach działania algorytmów analizy niezawodności możliwe jest równoległe wykonywanie obliczeń wartości funkcji granicznej, co w znaczący

sposób podnosi efektywność rozwiązania. Spośród metod, które ze swej “natury” mogą być łatwo przetwarzane na maszynach wieloprocesorowych wymienić należy metody symulacyjne lub wykorzystujące planowane eksperymenty, takie jak Monte Carlo, importance sampling czy Z3M. Ponadto, czas działania metod gradientowych, bazujących na poszukiwaniu punktu projektowego, można w istotny sposób ograniczyć “zrównoleglając” obliczanie gradientów funkcji granicznej. Obecna wersja programu STAND oferuje możliwość realizowania zadań na komputerach o architekturze równoległej oraz na klastrach obliczeniowych.

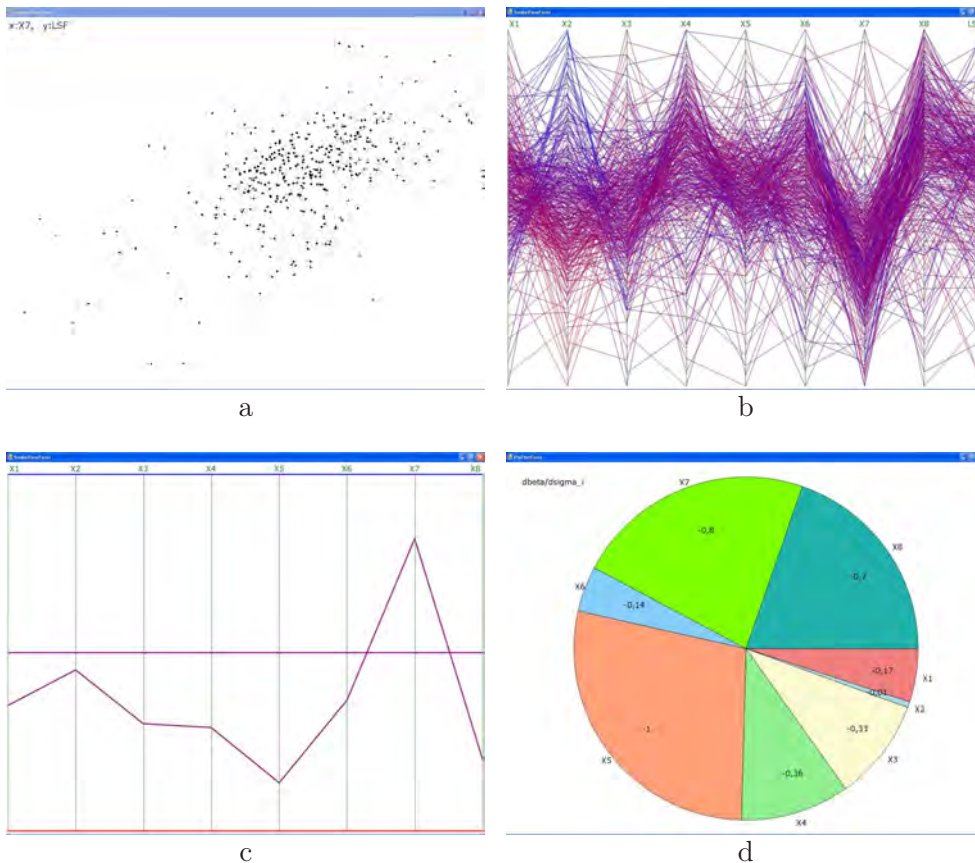
Inną cechą, bardzo ważną z punktu widzenia redukcji czasu trwania analizy, którą powinno się cechować oprogramowanie do obliczania niezawodności konstrukcji, jest możliwość ponownego uruchomienia analizy w miejscu gdzie została ona przedwcześnie przerwana. Aby nie utracić godzin, a nawet dni kosztownych symulacji komputerowych w przypadku awarii zasilania bądź też konieczności zmiany źle dobranej wartości parametru zbieżności, użytkownik powinien mieć możliwość odzyskania jak najwięcej informacji z przeprowadzonych do tej pory obliczeń, a także ich kontynuacji bez potrzeby rozpoczynania analizy od początku.

Program STAND ułatwia obsługę takich awaryjnych sytuacji gromadząc w plikach tymczasowych wartości funkcji granicznej obliczone w czasie trwania analizy. Rozpoczynając proces analizy niezawodności, dokonywane jest najpierw sprawdzenie czy istnieją już zapisane wartości funkcji granicznej odpowiadające generowanym przez algorytm realizacjom zmiennych losowych. Jeśli tak jest, to wartości te są po prostu odczytywane z przygotowanych plików bez przeprowadzania obliczeń MES. Trzeba pamiętać, że wiele metod analizy niezawodności używa generatorów liczb pseudolosowych i dlatego zarodek (ang. seed) takiego generatora musi być także zapisywany w celu zapewnienia powtórzenia generacji tego samego ciągu liczb, co w odtwarzanym procesie.

8.1.4. Wizualizacja wyników

Odpowiednie narzędzia prezentacji pomagają nie tylko w czytelny i przystępny sposób przedstawić główne wyniki analizy niezawodności, ale mogą być również źródłem dodatkowej wiedzy o badanej konstrukcji. Poza standardowymi wynikami, takimi jak prawdopodobieństwo awarii czy wskaźnik niezawodności, do raportu generowanego przez program STAND po zakończeniu obliczeń można dołączyć wykresy wizualizujące np. współrzędne punktu projektowego czy wrażliwości wskaźnika niezawodności na parametry rozkładów prawdopodobieństwa zmiennych losowych.

Realizacje wektora zmiennych \mathbf{X} , jak również odpowiadające im wartości funkcji granicznej obliczane w trakcie symulacji Monte Carlo lub też podczas poszukiwania punktu projektowego, można zaprezentować na wykresie punktowym (ang. *scatter plot*), rys. 8.7a, lub na wykresie wieloosiowym (ang. *parallel coordinate plot*), rys. 8.7b.



Rys. 8.7. a - wykres punktowy, b - wykres wieloosiowy, c - prezentacja współrzędnych punktu projektowego na wykresie wieloosiowym, d - wrażliwości wskaźnika niezawodności na wykresie kołowym.

Wykres punktowy jest standardowym narzędziem wizualizacji danych na płaszczyźnie lub w przestrzeni trójwymiarowej. Pozwala on na analizę rzutów wielowymiarowej “chmury” danych na płaszczyznę o osiach wyznaczonych przez zmienne losowe i funkcję graniczną. Z drugiej strony, na wykresie wieloosiowym

(spotyka się też nazwę “wykres typu sieć rybacka”) pokazać można wielowymiarowe punkty danych “w całości”. Ideą wykresu wieloosiowego jest prezentacja każdego punktu danych w postaci linii łamanej o wierzchołkach umieszczonych na równomiernie rozłożonych, równoległych osiach, reprezentujących wszystkie zmienne oraz funkcję graniczną. Dolne i górne końce osi oznaczają odpowiednio najmniejsze i największe wartości danej zmiennej w zbiorze punktów danych. Wierzchołki poszczególnych łamanych położone są na osiach zgodnie z wartościami współrzędnych punktów, które przedstawiają. Wykresy wieloosiowe zaimplementowane w programie STAND umożliwiają usunięcie pewnej grupy łamanych (punktów) poprzez zawężenie zakresów wybranych współrzędnych. Na przykład, odfiltrowując wszystkie realizacje zmiennych losowych, które prowadzą do dodatnich wartości funkcji granicznej, można otrzymać informację na temat dominującego scenariusza awarii konstrukcji.

Wykres wieloosiowy może być również ciekawym sposobem przedstawienia współrzędnych punktu projektowego, zob. rys. 8.7c. Stosując np. $\mu_{X_i} \pm 3\sigma_{X_i}$ jako zakresy osi odpowiednich zmiennych losowych można natychmiast zauważyć, gdzie w stosunku do wartości oczekiwanych znajduje się najbardziej prawdopodobna realizacja prowadząca do awarii konstrukcji.

Jedną z dużych zalet metody analizy niezawodności pierwszego rzędu jest łatwość otrzymania wrażliwości wskaźnika niezawodności oraz prawdopodobieństwa awarii P_f^{FORM} na parametry rozkładów prawdopodobieństwa zmiennych losowych, zob. [143, 230]. Znajomość wrażliwości pomóc może w identyfikowaniu głównych źródeł losowości odpowiedzi konstrukcji oraz w podejmowaniu decyzji projektowych. W przypadku umiarkowanej liczby zmiennych, wygodnym sposobem wizualizacji obliczonych wrażliwości jest wykres kołowy, rys. 8.7d. Analizując na takim wykresie wrażliwość wskaźnika niezawodności na odchylenia standardowe zmiennych losowych, można wyciągnąć wnioski dotyczące ewentualnych sposobów redukcji prawdopodobieństwa awarii konstrukcji poprzez kontrolę jakości wybranych jej elementów.

8.1.5. Obiektowa architektura programu STAND na przykładzie modułu analizy niezawodności

Wybór paradygmatu programowania zorientowanego obiektowo okazał się bardzo trafną decyzją zarówno jeśli chodzi o organizację kodu programu STAND, jak również o umożliwienie jednoczesnej pracy wielu programistów. Co trzeba również mocno podkreślić, użycie języka C++, który jest obecnie dominującym językiem programowania wykładanym na wyższych uczelniach, pozwala

mieć nadzieję, że młodym członkom zespołu badawczego rozwijającego STAND w IPPT PAN łatwiej będzie w przyszłości włączyć się w pracę nad rozwijaniem kodu.

Przez dekady najważniejszym językiem programowania stosowanym do tworzenia większości programów do obliczeń naukowych i inżynierskich był Fortran. Jest on przykładem tzw. języka strukturalnego, przeznaczonego przede wszystkim do programowania algorytmów obliczeniowych. Jedną z cech programowania strukturalnego jest to, że dane i operacje na danych są rozdzielone, a nacisk kładziony jest właśnie na przetwarzanie danych, implementowanych w sposób jawny. Pomimo, iż dobrze napisany w Fortranie program odznaczać się może dużą efektywnością, to język ten o wiele gorzej nadaje się do budowy dużych projektów informatycznych oraz zarządzania złożonymi strukturami danych. Rosnące możliwości sprzętu komputerowego wiążą się z coraz większym rozmiarem oraz stopniem skomplikowania zadań obliczeniowych, których realizacja wymaga bardziej elastycznych i efektywnych narzędzi programistycznych. Ewolucja obowiązującego paradygmatu programowania była zatem konieczna aby sprostać nowym wyzwaniom, a właśnie programowanie zorientowane obiektowo wydaje się dostarczać potrzebnych ku temu środków.

W ramach podejścia obiektowego, już nie algorytm, a odpowiednia struktura danych jest najważniejszym składnikiem tworzonego oprogramowania. Centralnym elementem każdego obiektowo zorientowanego języka jest *klasa*. Koncepcja klasy łączy w sobie zarówno zmienne (atrybuty) jak i operujące na nich procedury (metody). Dobrze zaprojektowane klasy powinny odzwierciedlać tak wiernie jak to możliwe specyficzne cechy analizowanego zagadnienia.

Innym kluczowym elementem modelowania zorientowanego obiektowo jest możliwość stworzenia struktury klas. Mechanizm *dziedziczenia* pozwala na budowanie hierarchii klas, którą zobrazować można na grafie podobnym do drzewa genealogicznego. Klasy niższego poziomu (podklasy) dziedziczą atrybuty i metody klasy położonej wyżej, tzw. klasy bazowej, umożliwiając przejrzystą oraz efektywną organizację wspólnych cech tych klas. W przypadku obiektowo zorientowanego kodu do analizy niezawodności, wspólnym atrybutem różnych klas odpowiadających algorytmom obliczania prawdopodobieństwa awarii (FORM, SORM, Monte Carlo ...) jest wektor zmiennych losowych, czyli wektor zawierający obiekty typu *Zmienna Losowa* gromadzące wszystkie niezbędne dane oraz metody potrzebne do operacji na tych zmiennych. Ponieważ wymienione algorytmy muszą mieć dostęp do wektora zmiennych losowych, dlatego deklarowany on jest jako atrybut klasy bazowej *Analiza Niezawodności*, zob. rys. 8.8, a dzięki dziedziczeniu należy on także do klas leżących niżej w hierarchii.

Dzięki tzw. *polimorfizmowi* metody odziedziczone z klasy bazowej można ponownie zdefiniować poprzez ich ponowną deklarację w ramach danej klasy oraz napisanie nowej wersji kodu metody. Pozwala to w łatwy sposób oddać różnice między klasami. Taka struktura klas nazywana jest czasem *specjalizacja-generalizacja* ponieważ w klasie bazowej deklaruje się daną metodę w najbardziej ogólnej postaci, natomiast w klasach dziedziczących metody te przybierają dużo bardziej konkretną i wyspecjalizowaną formę.

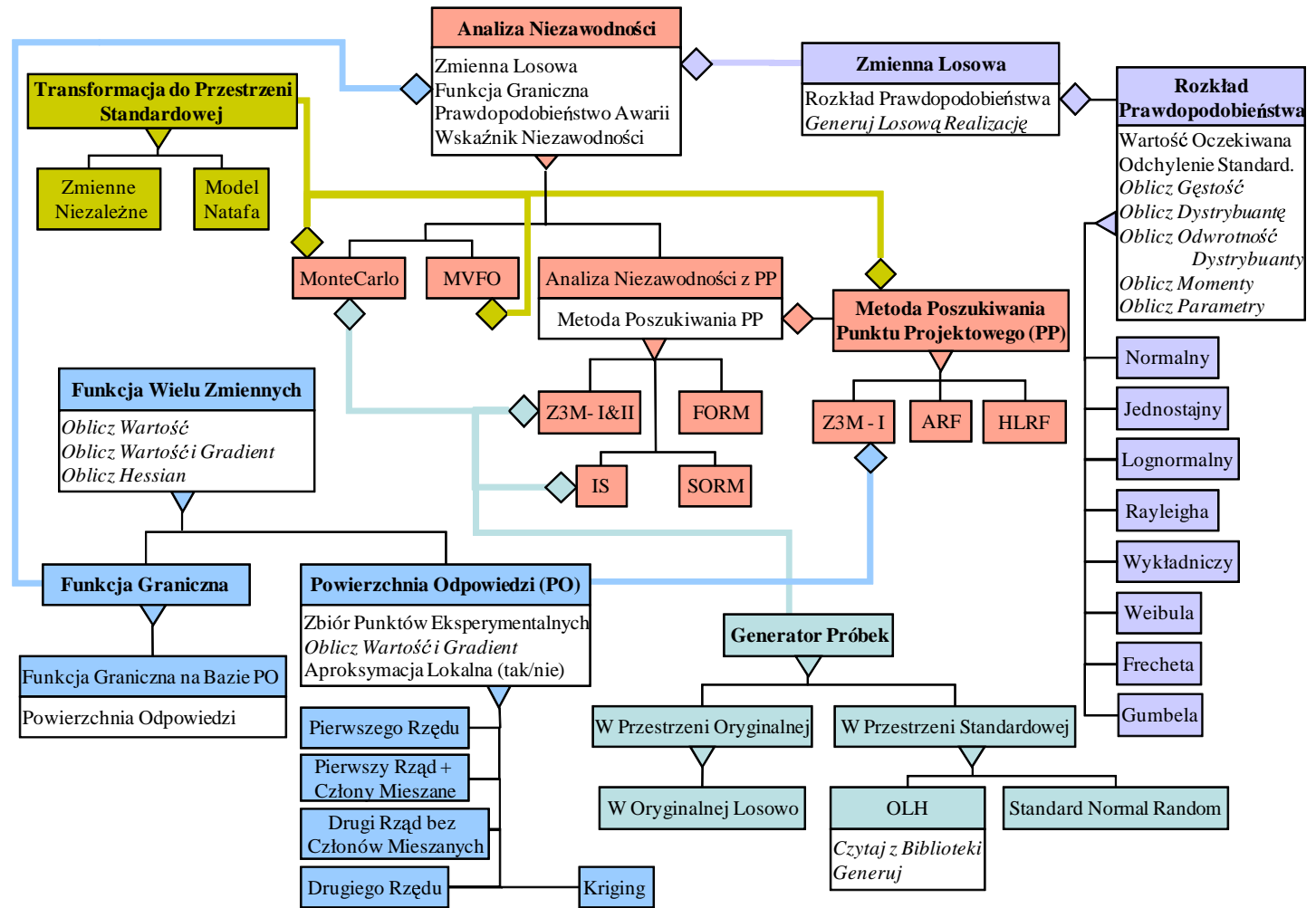
Bardzo często klasy bazowe projektowane są w ten sposób, że jedynie za pomocą kilku wybranych metod możliwa jest ich komunikacja z innymi obiektami. Znając takie “sztywne” interfejsy, nad różnymi częściami oprogramowania mogą pracować oddzielne grupy programistów, wiedząc, że ich fragmenty kodu będą ze sobą współpracowały.

Wszystkie te mechanizmy pomagają bardziej efektywnie tworzyć komputerowe modele rzeczywistych problemów. Cel i zakres niniejszego opracowania nie pozwalają na szczegółową prezentację tematyki programowania zorientowanego obiektowo. Zagadnienia te zostały wyczerpująco omówione np. w klasycznej już książce Bjarne Stroustrupa [241].

Rysunek 8.8 przedstawia (bardzo uproszczony dla przejrzystości) diagram hierarchii klas modułu analizy niezawodności programu STAND. Za kluczowe dla tego modułu należy uznać klasy: Analiza Niezawodności, Zmienna Losowa, Rozkład Prawdopodobieństwa, Funkcja Graniczna, Powierzchnia Odpowiedzi, Metoda Poszukiwania Punktu Projektowego, Transformacja do Przestrzeni Standardowej i Generator Próbek. Klasy te tworzą główny szkielet programu. Klasa bazowa Analiza Niezawodności wprowadza wspólne atrybuty algorytmów rozwiązujących, takie jak zmienne losowe i funkcja graniczna. Klasy reprezentujące symulacje Monte Carlo i metodę MVFO (ang. *mean value first order*) [230] dziedziczą z Analiza Niezawodności bezpośrednio. Klasa Analiza Niezawodności z PP, wyprowadzona z klasy bazowej jest już bardziej wyspecjalizowaną, ale nadal abstrakcyjną klasą¹. Reprezentuje ona te metody analizy niezawodności, które wykorzystują koncepcję punktu projektowego. Obecnie w programie STAND dostępne są cztery takie algorytmy, są to: FORM (zob. dodatek A), SORM (zob. [143, 230]), IS (importance sampling, dodatek A), i Z3M (zob. podrozdział 5.1). Dziedziczą one właśnie z Analiza Niezawodności z PP. Na rys. 8.8 operacja dziedziczenia oznaczona jest łącznikiem z trójkątem.

¹Klasa abstrakcyjna w programowaniu obiektowym jest to klasa, która nie może mieć swoich reprezentantów pod postacią obiektów. Każda klasa, która dziedziczy po klasie abstrakcyjnej i sama nie chce być abstrakcyjną, musi implementować wszystkie odziedziczone metody wirtualne.

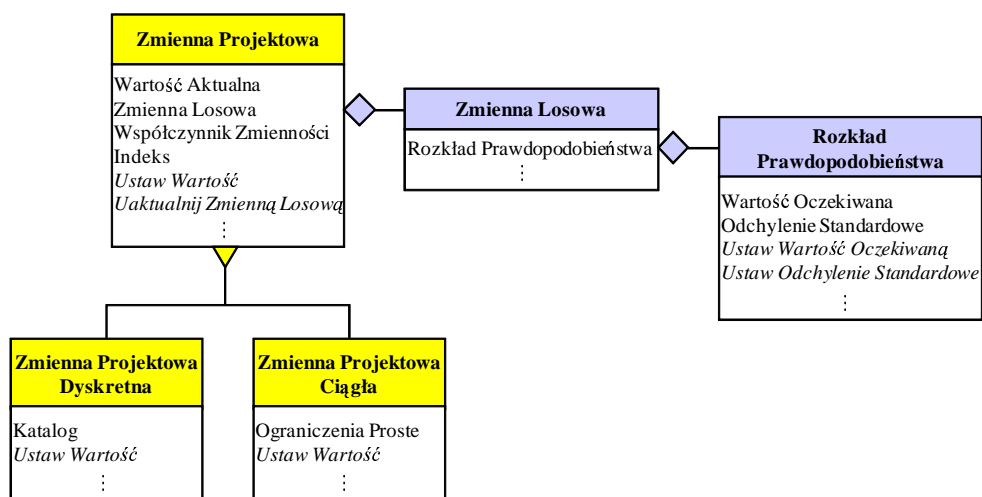
Rys. 8.8. Uproszczona hierarchia klas modułu analizy niezawodności programu STAND



8.2. Wybrane zagadnienia implementacji algorytmu optymalizacji odpornościowej

Analizując sformułowanie optymalizacji odpornościowej konstrukcji (2.43)–(2.48) można zauważyć, że mamy tam do czynienia z różnymi rodzajami zmiennych. Obok deterministycznych zmiennych projektowych $d_i, i = 1, \dots, n_d$, w sformułowaniu występują zmienne losowe $X_i, i = 1, \dots, n_X$, oraz $P_i, i = 1, \dots, n_P$, które tworzą model stochastyczny zadania. Przyjęto, że wartości oczekiwane $\mu_{\mathbf{X}}$ zmiennych \mathbf{X} wchodzi w skład wektora zmiennych projektowych. Zmienne \mathbf{P} przyczyniają się jedynie do losowego rozrzutu wartości funkcji celu i funkcji ograniczeń, a ich charakterystyki probabilistyczne nie zmieniają się w trakcie procesu optymalizacji.

Schemat wzajemnych powiązań klas reprezentujących w programie STAND zmienne projektowe i zmienne losowe pokazano schematycznie na rys. 8.9.



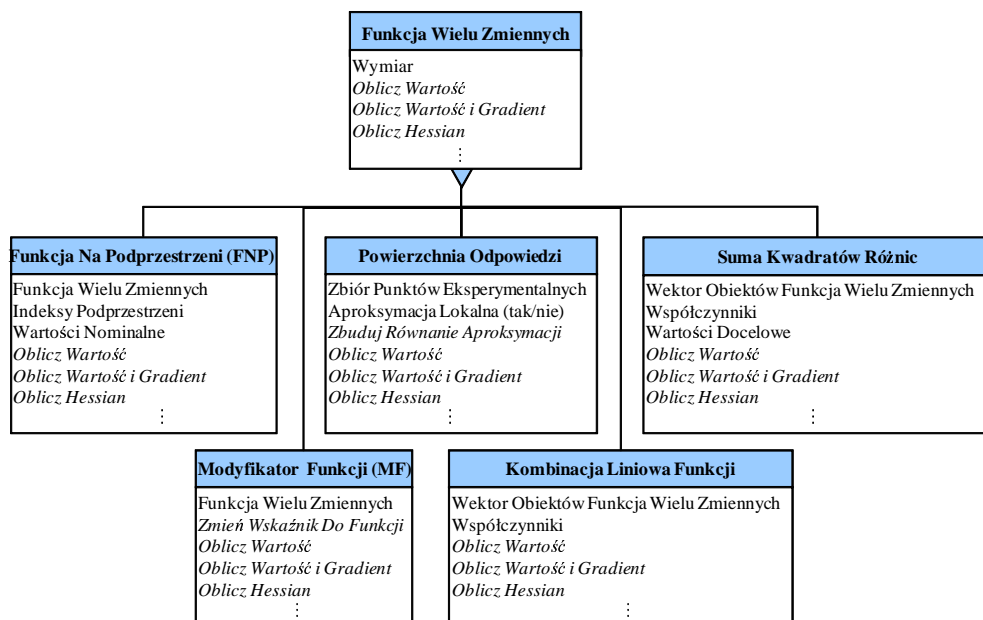
Rys. 8.9. Schemat współpracy klas reprezentujących zmienne projektowe i zmienne losowe zapewniający możliwość modelowania zmiennych projektowych $\mu_{\mathbf{X}}$.

Bazowa klasa abstrakcyjna **Zmienna Projektowa** oprócz takich atrybutów jak *Wartość Aktualna* przechowuje dodatkowo *Współczynnik Zmienności* oraz wskaźnik do obiektu **Zmienna Losowa**. Ma to na celu uwzględnienie w zadaniu optymalizacji odpornościowej zmiennych projektowych $\mu_{\mathbf{X}}$. Deklarując zmienną projektową tego typu podaje się z którą zmienną losową jest ona stowarzyszona.

Jeśli w ramach działania algorytmu optymalizacyjnego zmieniona zostaje wartość aktualna zmiennej projektowej, wtedy wywoływana jest metoda *Uaktualnij Zmienną Losową*, która odpowiednio modyfikuje wartość oczekiwaną stowarzyszonej zmiennej losowej (wywoływana jest metoda *Ustaw Wartość Oczekiwana* obiektu *Rozkład Prawdopodobieństwa* będącego składową stowarzyszonego obiektu *Zmienna Losowa*). Jeśli założono, że odchylenie standardowe zmiennej losowej zmienia się proporcjonalnie do wartości oczekiwanej zgodnie ze stałym współczynnikiem zmienności, wtedy metoda *Uaktualnij Zmienną Losową* dostosowuje również wartość tej charakterystyki. W przypadku deterministycznych zmiennych projektowych stowarzyszenie ze zmienną losową jest puste.

Na algorytm optymalizacji odpornościowej opisany w podrozdziale 2.4.2.3 składa się szereg operacji realizowanych naprzemiennie w przestrzeni zmiennych projektowych oraz przestrzeni probabilistycznej. Zarówno wagowa funkcja celu jak i ograniczenia sformułowania (2.43)–(2.48) wyrażone są jako funkcje momentów statystycznych funkcji zmiennych losowych. Wyznaczenie tych momentów wymaga przeprowadzenia np. symulacji Monte Carlo w $n_X + n_P$ wymiarowej przestrzeni zmiennych losowych. W algorytmie, symulacje takie przeprowadzane są dla wybranych punktów eksperymentalnych w przestrzeni zmiennych projektowych w celu utworzenia odpowiednich aproksymacji (powierzchni odpowiedzi) wagowej funkcji celu i ograniczeń odpornościowych. Tak zbudowane powierzchnie odpowiedzi są $n_{dX} = n_d + n_X$ wymiarowymi funkcjami zmiennych projektowych i właśnie one tworzą zmodyfikowane zadanie optymalizacji deterministycznej, którego rozwiązanie stanowi przybliżone rozwiązanie zadania optymalizacji odpornościowej. W innej wersji algorytmu, przedstawionej w podrozdziale 2.4.2.2 strategii aproksymacji statystyk, powierzchnie odpowiedzi tworzone są oddzielnie dla poszczególnych wartości średnich i odchyłeń standardowych funkcji tworzących zadanie optymalizacji. Implementacja takiego algorytmu wymaga aby przed rozwiązaniem zadania optymalizacji deterministycznej utworzone zostały funkcje scalające te aproksymacje w funkcje celu (2.44) lub (2.49) oraz ograniczenia odpornościowe (2.45) i (2.46).

Aby umożliwić komputerową realizację tak skonstruowanego algorytmu w ramach programowania zorientowanego obiektowo, konieczne było zaprojektowanie całej struktury klas modelujących wszystkie funkcje występujące w kolejnych etapach rozwiązania. Na diagramie przedstawionym na rys. 8.10 przedstawiono najważniejsze atrybuty i metody tych klas. Abstrakcyjna klasa bazowa *Funkcja Wielu Zmiennych* przechowuje przede wszystkim *Wymiar* funkcji oraz wymusza na wszystkich klasach dziedziczących implementację funkcji *Oblicz Wartość*. W przypadku, gdy klasy dziedziczące nie tworzą własnych kopii funkcji *Oblicz*



Rys. 8.10. Struktura klas modelujących funkcje wielu zmiennych używanych w implementacji zadania optymalizacji odpornościowej.

Wartość i Gradient i *Oblicz Hessian*, wtedy obliczane są one numerycznie za pomocą funkcji implementowanych w klasie bazowej.

Funkcje rozpatrywane w zadaniu optymalizacji odpornościowej są tak naprawdę funkcjami $n_d + n_X + n_P$ wymiarowymi. Mówiąc mniej ściśle, w zadaniu symulacji Monte Carlo “występują” one w swojej $n_X + n_P$ wymiarowej wersji (przy ustalonych wartościach zmiennych \mathbf{d} i $\boldsymbol{\mu}_X$), a w zadaniu optymalizacji deterministycznej jako $n_d + n_X$ wymiarowe powierzchnie odpowiedzi. Dlatego, tworząc obiekt algorytmu rozwiązującego, do przekazania niezbędnych danych na temat funkcji celu i ograniczeń używa się obiektów klas *Funkcja na Podprzestrzeni* oraz *Modyfikator Funkcji*. Klasa *Funkcja na Podprzestrzeni* pozwala traktować K wymiarową funkcję jako funkcję $k < K$ wymiarową poprzez ustalenie wartości pozostałych $K - k$ zmiennych. Taka funkcjonalność potrzebna jest na etapie budowania powierzchni odpowiedzi, gdy dla ustalonych zmiennych projektowych dokonujemy oszacowania wartości momentów statystycznych funkcji celu i ograniczeń.

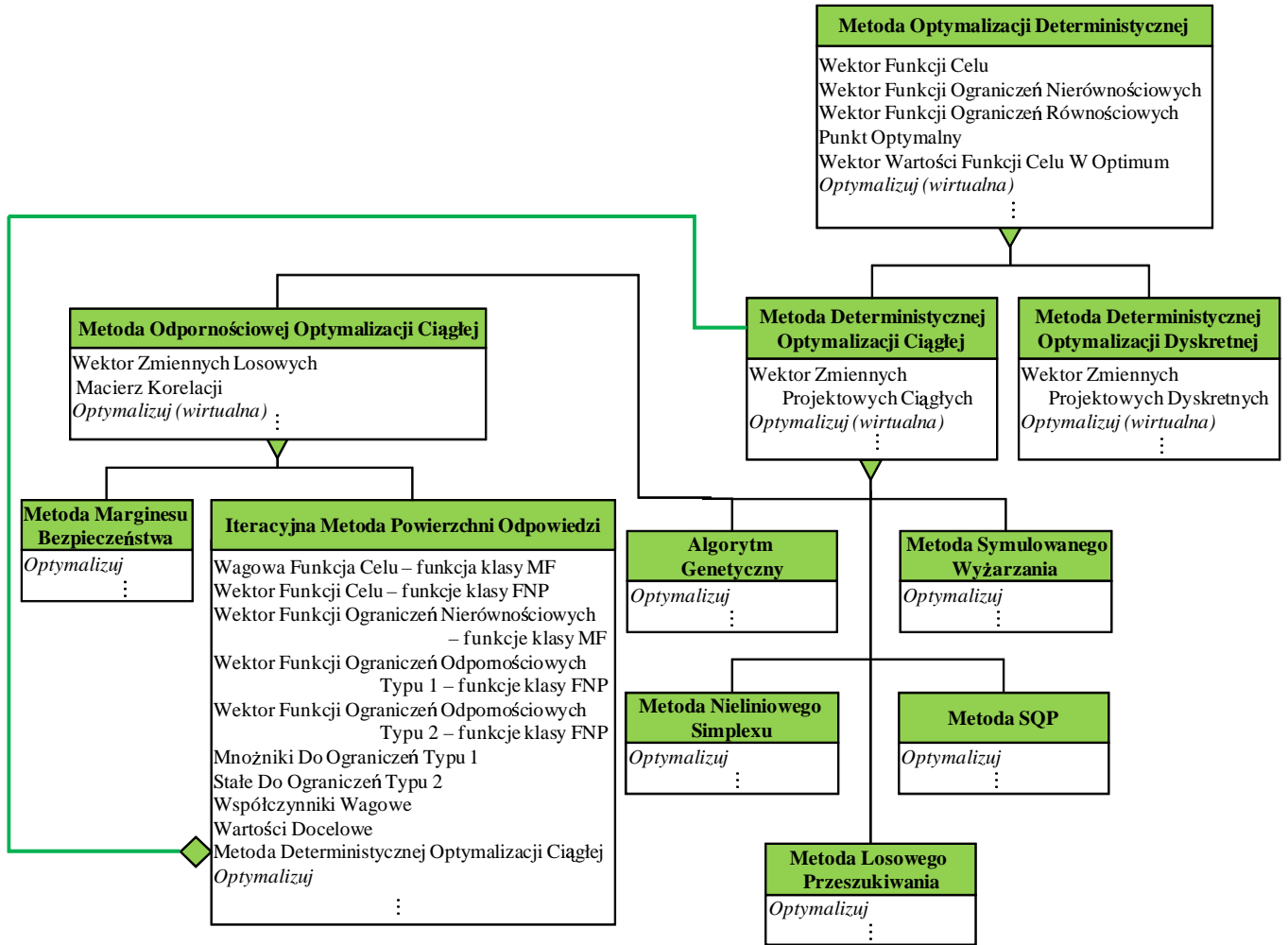
Jak to zostanie pokazane przy okazji omawiania klasy reprezentującej algorytm optymalizacji odpornościowej, tworząc obiekt tej klasy należy podać jako parametr konstruktora obiekt klasy *Metoda Deterministycznej Optymalizacji*

Ciągłej. Z kolei przy konstrukcji tego obiektu niezbędne jest podanie funkcji celu i ograniczeń w przestrzeni $n_d + n_X$ wymiarowej. Ponieważ algorytm optymalizacji deterministycznej operuje na funkcjach aproksymujących, czyli na utworzonych w trakcie procesu optymalizacji odpornościowej obiektach Powierzchnia Odpowiedzi, dlatego funkcje przekazywane obiektowi Metoda Deterministycznej Optymalizacji Ciągłej są obiektami klasy Modyfikator Funkcji. Klasę tę zaprojektowano w ten sposób, żeby pozwalała na wielokrotną zmianę postaci funkcji, co realizowane jest przez zmianę wskazania do zewnętrznego obiektu Funkcja Wielu Zmiennych (metoda *Zmień Wskaźnik Do Funkcji*).

Na rys. 8.11, w bardzo uproszczonej postaci, przedstawiono strukturę tych klas reprezentujących metody optymalizacji, które w programie STAND związane są z realizacją zadania optymalizacji odpornościowej.

Bazowa abstrakcyjna klasa Metoda Optymalizacji Deterministycznej zapewnia interfejs dla szeregu klas z niej dziedziczących, przechowując takie składowe jak funkcje celu i funkcje ograniczeń oraz wyniki procesu optymalizacji. Abstrakcyjna metoda *Optymalizuj* implementowana jest we wszystkich nieabstrakcyjnych klasach pochodnych zgodnie z daną strategią rozwiązania. I tak, dostępne są obecnie następujące metody deterministycznej optymalizacji ciągłej (o ciągłych parametrach projektowych): metoda nieliniowego sympleksu Nedlera-Meed'a, zob. [184,243], metoda losowego przeszukiwania [226], metoda sekwencyjnego programowania kwadratowego [215], metoda symulowanego wyżarzania [121] oraz algorytm genetyczny [78].

Metody optymalizacji odpornościowej, koncepcyjnie dużo bardziej złożone od metod optymalizacji deterministycznej znajdują się - zgodnie z filozofią podejścia obiektowego - niżej w hierarchii klas. Przejmując wspomniane standardowe atrybuty metod optymalizacyjnych, dodają szereg specyficznych składników, niezbędnych do działania algorytmu. Najdokładniej omówiona w niniejszej pracy oraz zilustrowana przykładami iteracyjna metoda powierzchni odpowiedzi (zob. podrozdziały 2.4.2.2 i 2.4.2.3) przechowuje funkcje celu oraz ograniczenia zarówno w wersji Modyfikator Funkcji jak i Funkcja na Podprzestrzeni. Atrybutami klasy są zarówno metoda powierzchni odpowiedzi, jak też metoda optymalizacji deterministycznej. Deklarując te obiekty za pomocą odpowiednich klas bazowych mamy dostęp do szeregu wyspecjalizowanych klas pochodnych, co pozwala na elastyczny i wygodny dobór metody optymalizacji i techniki aproksymacji do specyfiki analizowanego zagadnienia.



Rys. 8.11. Fragment struktury klas związanych z realizacją zadania optymalizacji odpornościowej.

8.3. Przykłady łączenia z zewnętrznymi komercyjnymi programami MES

8.3.1. Łączenie z programem ABAQUS

Poniżej przedstawiono sposób zapewnienia komunikacji programu STAND z pakietem do analizy skończonej elementowej ABAQUS. Program ABAQUS użyty został w przykładzie analizy niezawodności trójkątnej tarczy (rys. 5.2) omówionym w podrozdziale 5.1.3. Zastosowana tam funkcja graniczna (5.13) wyrażona jest przez maksymalne naprężenie Hubera-Misesa obliczane w punktach węzłowych siatki elementów skończonych.

Łączność programu STAND z zewnętrznymi modułami obliczeniowymi zapewniająca jest za pośrednictwem plików tekstowych. Realizacje zmiennych zapisywane są w plikach wejściowych, a wyniki obliczeń MES odczytywane z odpowiednich tekstowych plików wyjściowych. W przypadku programu ABAQUS, dane dotyczące modelu oraz rodzaju analizy znajdują się w zbiorze tekstowym z rozszerzeniem `.inp`. Można również ustalić, aby wyniki zapisywane były w tekstowym pliku z rozszerzeniem `.dat`. Niestety, specyfika analizowanego problemu nie pozwala na bezpośrednie wykorzystanie tych plików. Ze względu na uwzględnienie w zadaniu zmiennych losowych opisujących geometrię tarczy (zmiennne X_1, X_2 i X_3 , tab. 5.1) położenie wszystkich węzłów siatki elementów skończonych podlega zmianie za każdym razem, gdy algorytm analizy niezawodności generuje nową realizację wektora losowego \mathbf{X} . Biorąc pod uwagę liczbę węzłów w modelu (ponad 4700), zmiany położenia są zbyt trudne do uprzedniego zaprogramowania jako funkcje X_1, X_2 i X_3 . Ponadto, standardowa dokładność prezentacji wyników w pliku `.dat` jest zdecydowanie zbyt mała na potrzeby większości algorytmów obliczania prawdopodobieństwa awarii.

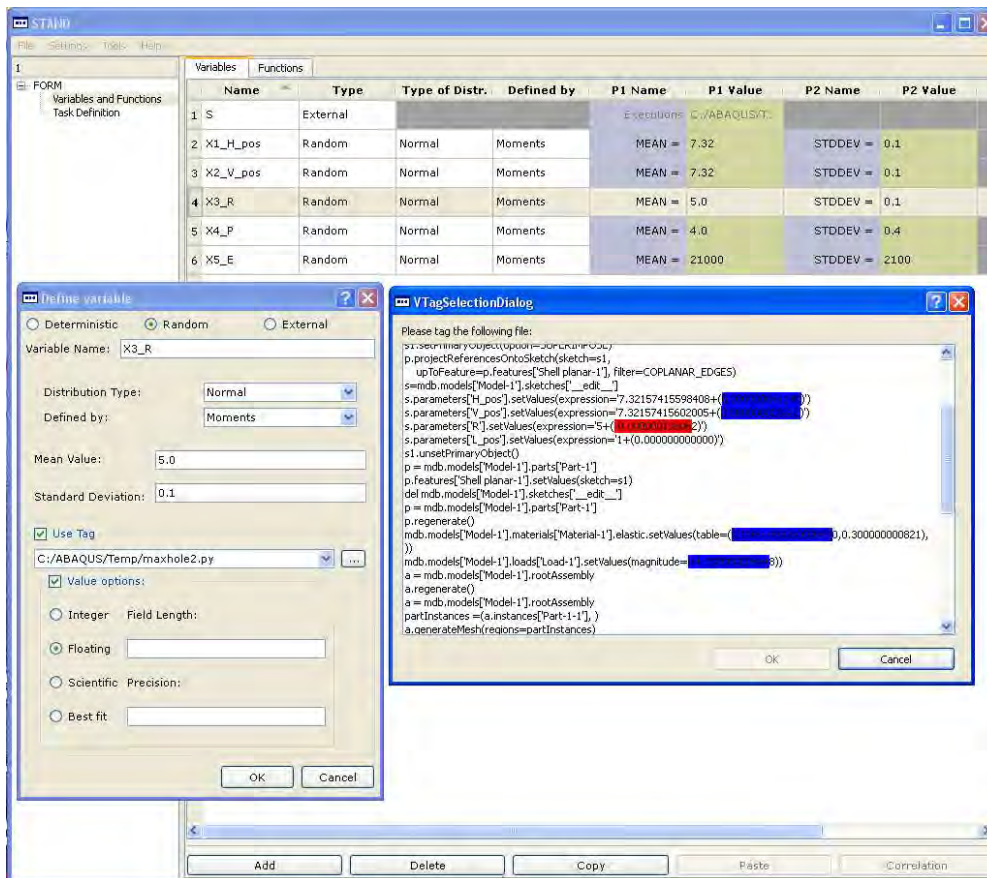
Z uwagi na powyższe trudności postanowiono nie używać plików `.inp` i `.dat`, opierając komunikację STAND-ABAQUS na tzw. ABAQUS Scripting Interface (SI). Jest to interfejs do programowania aplikacji (ang. *application programming interface* - API) służący do dokonywania operacji na modelach oraz danych używanych przez program ABAQUS. ABAQUS SI jest rozszerzeniem obiektowo zorientowanego języka Python i może być używany m.in. do tworzenia i modyfikacji elementów geometrii modelu, materiałów, obciążeń, uruchamiania zadań, a także zapisu oraz odczytu bazy wyników analizy MES.

Model tarczy przygotowany został przy pomocy graficznego środowiska ABAQUS/CAE i właśnie ta sesja pracy z ABAQUS/CAE wywoływana jest za każdym razem gdy trzeba otrzymać wartość funkcji granicznej dla nowej realizacji

zmiennych losowych. Odtworzenie kolejnych kroków budowy modelu MES, od modelu geometrycznego przez definicję obciążeń, warunków brzegowych, danych materiałowych aż po generację siatki, możliwe było dzięki automatycznie tworzonemu zapisowi tych operacji w postaci skryptu programu Python `abaqus.rpy`. Skrypt ten używany jest zazwyczaj w celu powtórzenia sesji pracy z programem ABAQUS/CAE w przypadku nietypowego zakończenia działania programu, np. na skutek awarii zasilania. Po dokonaniu kilku modyfikacji, skrypt `abaqus.rpy` okazał się być doskonałym narzędziem do zapewnienia komunikacji STAND-ABAQUS. To właśnie do tego zbioru zapisywane były wartości realizacji zmiennych losowych. Ponieważ skrypt zawiera polecenie generacji siatki dla zmienionej geometrii tarczy, dlatego nie ma potrzeby bezpośredniej ingerencji w plik `.inp`. Wspomniane modyfikacje dotyczą dodania do skryptu komend wymuszających pracę w trybie “cichym” (bez otwierania okien interfejsu graficznego), a po zakończeniu analizy MES, poleceń utworzenia binarnego pliku z wynikami o rozszerzeniu `.odb`, znalezienia wartości maksymalnego naprężenia Hubera-Misesa i zapisania jej w odpowiednim formacie w zbiorze tekstowym.

Definicje zmiennych losowych tworzących stochastyczny model analizowanego problemu (zob. tab. 5.1) wprowadzono za pomocą graficznego interfejsu programu STAND pokazanego na rys. 8.12. Wybrana w oknie dialogowym opcja “Use Tag” (użyj znacznika) oznacza, iż realizacje danej zmiennej zapisywane będą w odpowiednim tekstowym pliku z danymi. Miejsce, w które wpisywana ma być wartość realizacji zaznacza się myszką podświetlając je na czerwono. Zarezerwowane pola, tzn. te gdzie ustawiono już znaczniki innych zmiennych, podświetlone są kolorem niebieskim.

We wzorze funkcji granicznej (5.13) użyto tylko jednej zmiennej zewnętrznej. Jest to maksymalne naprężenie Hubera-Misesa S obliczane w węzłach siatki. Na rys. 8.13 zaprezentowano okno dialogowe służące do definicji tej zmiennej. Widoczna tam lista “External programs” (programy zewnętrzne) pokazuje sekwencję zawierającą ścieżki do programów i parametry wywołania potrzebne aby otrzymać wartość S . W omawianym przykładzie w tym celu uruchamiany jest program wsadowy `abaqus.bat`. Parametry specyfikowane w linii komend ustalają, że bez otwierania graficznego interfejsu użytkownika program ABAQUS/CAE wykona polecenia zawarte w `maxhole2.py`. Następnie, po zakończeniu działania programu interesujący nas wynik zapisywany jest w pliku `holeMiss.txt`. Można zauważyć, że plik ten, który wygenerowany został przez omawiany wyżej skrypt języka Python, zawiera tylko jedno pole tekstowe z wartością maksymalnego naprężenia Hubera-Misesa.

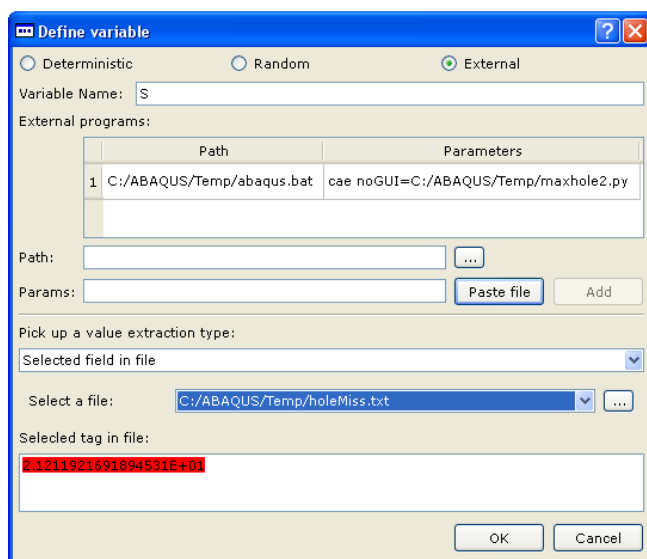


Rys. 8.12. Definicja zmiennych losowych. Ustawianie znacznika zmiennej w skrypcie języka Python, służącym do komunikacji STAND-ABAQUS.

8.3.2. Łączenie z programem RADIOSS

We wszystkich przykładach dotyczących zarówno analizy niezawodności jak i optymalizacji odpornościowej, które zaprezentowano w rozdziale 6, symulacje zderzeń wykonywano przy pomocy programu RADIOSS [171]. Jest to program metody elementów skończonych rozwijany przez firmę Altair, wykorzystujący jawny schemat całkowania równań ruchu. Głównym obszarem jego zastosowań jest symulacja silnie nieliniowych dynamicznych problemów, a w szczególności zderzeń.

Dane do analizy programem RADIOSS zapisane są w dwóch plikach. Są to: plik D00, zawierający dane modelu skończenie elementowego, czytany przez



Rys. 8.13. Definicja zmiennych zewnętrznych. S jest maksymalnym naprężeniem Hubera-Misesa.

program RADIOSS Starter w celu weryfikacji i wstępnego przetworzenia oraz plik D01, gdzie zapisanych jest szereg komend sterujących zarówno algorytmem rozwiązania jak i generacją wyników, czytany przez program RADIOSS Engine, którego uruchomienie rozpoczyna właściwą analizę MES.

W przykładach przedstawionych w podrozdziałach 6.1–6.4 mamy do czynienia jedynie z prostymi przypadkami zmiennych losowych, których realizacje można zapisywać bezpośrednio w pliku D00. Uwzględnienie losowych awarii zgrzewów sprowadza się do automatycznego dopisywania na końcu pliku D01 szeregu komend usuwających wybrane belkowe elementy skończone modelujące zgrzewy. Wartości rozpatrywanych w tych przykładach zmiennych zewnętrznych (energia wewnętrzna, przemieszczenia węzłów) otrzymywane są przy użyciu wchodzącego w skład pakietu RADIOSS programu TH++. Pozwala on otrzymać w formie pliku tekstowego dane do wykresu zmienności w czasie dowolnej odpowiedzi konstrukcji zapisanej w binarnej bazie wyników. Co ważne, program ten może pracować w trybie wsadowym.

Niestety, strategii takiej nie można było zastosować do realizacji zadania optymalizacji odpornościowej zaprezentowanego w podrozdziale 6.5. Optymalizacja jest tam narzędziem do kalibracji parametrów numerycznego modelu fantomu kończyny dolnej stosowanego w testach zderzeniowych. Występujące

w zadaniu zmienne losowe (zob. tab. 6.4) mają złożony charakter, a zmiana ich wartości modyfikuje położenie wszystkich węzłów siatki elementów skończonych fantomu. Nie wchodzi zatem w grę bezpośrednio uaktualnianie pliku D00 przez program STAND. Podobnie jak w przypadku łączenia z programem ABAQUS, wpływ zmiennych opisujących geometrię uwzględniony został przez wprowadzenie ogniwa pośredniego, jakim były skrypty programu HyperStudyDSS wchodzącego w skład rodziny programów Altair. HyperStudyDSS pozwala na całym modelu MES lub na jego części dokonywać takich operacji jak translacja, obrót czy skalowanie, zastępuje też program TH++ w ekstrakcji wyników. Sekwencja otrzymania wartości zmiennej zewnętrznej (tutaj, kąta bocznego zgięcia kolana) dana jest następująco: wpisanie realizacji zmiennych losowych w odpowiednie miejsca skryptu HyperStudyDSS, uruchomienie HyperStudyDSS, który modyfikuje plik D00, uruchomienie Radioss Starter, a następnie Radioss Engine, ponowne uruchomienie HyperStudyDSS aby uzyskać wartość zmiennej zewnętrznej i ostatecznie odczytanie tej wartości z pliku tekstowego.

Jak już o tym wcześniej wspomiano, program STAND oferuje możliwość rozdzielenia przeprowadzanych symulacji na wiele procesorów. Ponieważ czas pojedynczej analizy skończenie elementowej programu RADIOSS wynosił nawet dwie godziny, dlatego we wszystkich przykładach rozdziału 6 korzystano z równoległej pracy 16 lub 32 procesorów (w zależności od ograniczeń posiadanej w danym czasie licencji programu).

PODSUMOWANIE

W niniejszym rozdziale rozważano problemy dotyczące efektywnej komputerowej implementacji algorytmów analizy niezawodności oraz optymalizacji odpornościowej konstrukcji. Poszczególne rozwiązania omawiane były na przykładzie programu STAND rozwijanego w Pracowni Niezawodności i Optymalizacji IPPT PAN. Program ten rozwijany jest pod kierownictwem autora już od szeregu lat w ramach kilku projektów badawczych, zarówno krajowych jak i europejskich.

Niezależnie od przeznaczenia tworzonego kodu, budowa i utrzymanie dużego programu komputerowego jest wyzwaniem samym w sobie. Niezbędne jest zatem zaprojektowanie takiej architektury programu, która umożliwi jednoczesną pracę wielu programistów oraz pozwala na jego dalszy rozwój i modyfikowanie. Pokazano, że tzw. obiektowo-zorientowany paradygmat programowania stanowi bardzo dobre narzędzie do realizacji tych celów. Najważniejsze koncepcje

programowania obiektowego, przedstawiono na przykładach budowy i organizacji klas programu STAND.

Obok obiektowo zorientowanej architektury kodu, szczególnie szeroko przedstawione zostały metody łączenia programu z zewnętrznymi modułami obliczeniowymi. Taka funkcjonalność jest niewątpliwie podstawową cechą programu, który stanowić ma pomocne narzędzie dla projektanta. Dużo uwagi poświęcono również rozwiązaniom mającym na celu stworzenie wygodnego interfejsu użytkownika. Omówiono przykłady wykorzystania w analizie niezawodności i optymalizacji odpornościowej komercyjnych pakietów metody elementów skończonych ABAQUS i RADIOSS.

Program STAND używany był do realizacji większości zaprezentowanych w pracy przykładów obliczeniowych.

Wnioski końcowe. Elementy oryginalne pracy

Jako główne osiągnięcie niniejszej pracy uznać należy teoretyczne opracowanie oraz komputerową implementację szeregu metod umożliwiających szeroko pojętą analizę stochastyczną złożonych problemów mechaniki konstrukcji¹. Podjęto się zadania modyfikacji istniejących, a także zaproponowania nowych algorytmów analizy niezawodności i optymalizacji odpornościowej. Ilustrację tych metod stanowiły przede wszystkim zagadnienia dotyczące wytrzymałości zderzeniowej elementów pojazdów, jak również problemy zapewnienia jakości wytłoczki w procesie głębokiego tłoczenia blachy. W opinii autora, zarówno charakter analizowanych wyników symulacji jak i złożoność numeryczna tych zagadnień, czynią rozpatrywane przykłady reprezentatywnymi dla innych współczesnych problemów modelowania inżynierskiego.

Jedną z najważniejszych składowych większości zaprezentowanych algorytmów jest aproksymacja niejawnych funkcji odpowiedzi konstrukcji. Konstrukcje te modelowane są zazwyczaj za pomocą metody elementów skończonych. Zastąpienie funkcji granicznej w analizie niezawodności oraz funkcji celu i funkcji ograniczeń w zadaniu optymalizacji przez odpowiednio dobrany analityczny metamodel (powierzchnię odpowiedzi) ma kluczowe znaczenie dla efektywności tych analiz. W pracy główny nacisk położono na zastosowanie aproksymacyjnej wersji metody krigingu oraz technik aproksymacji lokalnej, takich jak metoda ważonej liniowej regresji. *Zastosowanie zmodyfikowanej wersji krigingu w metodzie realizacji zadania optymalizacji odpornościowej do aproksymacji wartości momentów statystycznych funkcji celu i ograniczeń można uznać za zupełnie nowe rozwiązanie.* Wszystkie omawiane w pracy metody powierzchni odpowiedzi zaimplementowano w programie STAND.

¹Pochyłą czcionką wyróżniono fragmenty, które zdaniem autora stanowią najważniejsze osiągnięcia pracy.

Wspólną cechą prezentowanych w pracy metod analizy niezawodności i optymalizacji odpornościowej jest wykorzystanie koncepcji łacińskiej hiperkostki, zarówno jako generatora próbek losowych jak również bazy eksperymentów przy tworzeniu powierzchni odpowiedzi. *Niewątpliwym osiągnięciem jest udoskonalenie algorytmów budowy optymalnych łacińskich hiperkostek - OLH, co pozwala na efektywne tworzenie jednorodnych planów eksperymentów, nawet w przestrzeniach wielowymiarowych. Algorytmy te zaimplementowane zostały przez autora zarówno w programie STAND, jak też w module M-Xplore programu analizy skończonej elementowej RADIOSS.*

Hiperkostki łacińskie stanowią fundament opracowanych przez autora metod optymalizacji odpornościowej. Użycie ich do generacji próbek losowych w symulacyjnych metodach estymacji momentów statystycznych funkcji celu i ograniczeń pozwoliło na znaczną poprawę efektywności estymacji. Wybór metod symulacyjnych podyktowany był przede wszystkim ich uniwersalnością. Pomimo iż alternatywne podejścia, takie jak metody perturbacyjne czy też metoda rozwinięcia w chaos wielomianowy, w wielu szczególnych przypadkach mogą okazać się bardziej efektywne od metod symulacyjnych, to jednak założenia wstępne, jakimi są obwarowane, zawężają zakres ich zastosowań. Ponadto, ograniczenia dotyczące rozkładów prawdopodobieństwa zmiennych, a także różniczkowalności funkcji, bardzo istotne w przypadku metod perturbacyjnych oraz metody redukcji wymiarów, nie mają znaczenia w kontekście symulacji losowych.

Na podstawie testów przeprowadzonych dla szeregu funkcji analitycznych oraz dwóch problemów mechaniki konstrukcji, dokonano porównania właściwości estymatorów wartości średniej i wariancji w przypadku próbek czysto losowych oraz generowanych przez losowe łacińskie hiperkostki i hiperkostki z optymalizacją rozmieszczenia punktów. *W znanych autorowi opracowaniach nie przeprowadzono nigdy tak szczegółowej analizy metod symulacyjnych wykorzystujących plany OLH. Wykazano, że próbki generowane przez hiperkostki łacińskie, są lepszym rozwiązaniem niż czysto losowe próbki w symulacjach Monte Carlo. Próbki utworzone przez OLH warto stosować szczególnie wtedy, gdy ze względu na koszt nie można pozwolić sobie na otrzymanie wartości rozważanej funkcji dla zbyt wielu realizacji zmiennych losowych.*

Plany OLH stanowią doskonałą bazę eksperymentów przy budowaniu powierzchni odpowiedzi. W szczególności dotyczy to metody krigingu. W pracy zbadano wpływ równomierności rozłożenia punktów eksperymentalnych na błąd dopasowania. Pokazano, że w przypadku krigingu, wykorzystanie planu eksperymentów OLH poprawia dopasowanie funkcji aproksymującej. *Wykazano również, że tzw. kryterium sił prowadzi do bardziej równomiernego rozłożenia*

punktów hiperkostki niż, dominujące w literaturze, kryterium maksymalizacji minimalnej odległości między punktami.

Bardzo dobre właściwości planów OLH dotyczące równomiernego wypełnienia danego obszaru punktami eksperymentalnymi wykorzystano w oryginalnej metodzie szybkiej poprawy niezawodności. Zaproponowano tam także koncepcję zmodyfikowanego planu OLH - optymalnego jedynie na podprzestrzeni. Przedstawiony algorytm poprawy niezawodności jest skuteczny szczególnie wtedy, gdy poziom niezawodności konstrukcji wyjściowej jest niski. Może też stanowić szybką procedurę startową dla algorytmów optymalizacji odpornościowej lub niezawodnościowej.

W analizie niezawodności złożonych układów konstrukcyjnych, nieliniowy charakter funkcji granicznych wynika nie tylko z natury modelowanego zjawiska fizycznego, ale jest również często wywołany szumem numerycznym spowodowanym przez stosowaną metodę rozwiązania zadania MES. Zarówno algorytm szybkiej poprawy niezawodności, jak też zmodyfikowany przez autora adaptacyjny algorytm symulacji losowych Z3M, nie są wrażliwe na wpływ umiarkowanego szumu.

Niewrażliwość na szum jest także warunkiem koniecznym skutecznej metody odpornościowej optymalizacji konstrukcji. W pracy zaproponowano trzy strategie rozwiązania zadania optymalizacji odpornościowej. Są to: podstawowa strategia aproksymacji, strategia aproksymacji statystyk i strategia aproksymacji funkcji celu i ograniczeń odpornościowych. Wszystkie trzy zostały zaimplementowane w tworzonej przez autora bibliotece optymalizacyjnej programu STAND. Na podstawie licznych testów można stwierdzić, że strategia aproksymacji funkcji celu i ograniczeń odpornościowych jest rozwiązaniem najbardziej efektywnym.

Według autora, sformułowanie optymalizacji odpornościowej wydaje się być lepiej dopasowane do realiów projektowych niż sformułowanie optymalizacji niezawodnościowej. Możliwość zastosowania optymalizacji niezawodnościowej jest w praktyce bardzo silnie uwarunkowana dostępnością funkcji łącznej gęstości rozkładu prawdopodobieństwa parametrów konstrukcji i obciążeń. W odróżnieniu od optymalizacji niezawodnościowej, sformułowanie optymalizacji odpornościowej konstrukcji nie wymaga szacowania prawdopodobieństwa awarii.

W pracy przedstawiono szereg przykładów ilustrujących działanie opracowanych algorytmów. Większość z nich dotyczyła analizy niezawodności oraz optymalizacji odpornościowej dynamicznie deformujących się elementów konstrukcji pojazdów. We wszystkich analizowanych przykładach zarówno funkcje

graniczne jak też funkcje tworzące sformułowanie zadania optymalizacji odpornościowej zawierały wpływ szumu. Szum ten jest charakterystyczny dla zagadnień zderzeniowych ze względu na metodę całkowania równań równowagi oraz stosowane algorytmy kontaktu. Ponadto, w zadaniach dotyczących zgniatanej cienkościennej podłużnicy dodatkowy szum związany był z modelowaniem losowych awarii zgrzewów. *Biorąc pod uwagę ogromną liczbę połączeń zgrzewanych występujących w konstrukcjach samochodowych, zaproponowana metoda usuwania pewnej liczby losowo wybranych zgrzewów wydaje się być praktycznym rozwiązaniem, które pozwala uwzględnić losowe wady tych połączeń.* W przypadku analizy niezawodności podejście to wymaga jednak odpowiednio dostosowanego algorytmu. *Pokazano, że użycie algorytmu Z3M umożliwia efektywne oszacowanie prawdopodobieństwa awarii pomimo nieróżniczkowalności funkcji granicznej.*

Sformułowano zadanie analizy niezawodności procesu głębokiego tłoczenia blachy. Podstawą do oszacowania prawdopodobieństwa awarii procesu jest znajomość funkcji granicznej, którą zdefiniowano wykorzystując standardowy wykres odkształceń granicznych. Jako metodę analizy niezawodności zaproponowano adaptacyjną metodę Monte Carlo. W związku z niemożnością precyzyjnego wyznaczenia krzywej odkształceń granicznych wprowadzono nową koncepcję rozmytego prawdopodobieństwa awarii.

Do realizacji większości zaprezentowanych w pracy przykładów obliczeniowych użyto programu STAND, który rozwijany jest w Pracowni Niezawodności i Optymalizacji IPPT PAN. *Stworzone przez autora liczne moduły programu, a w szczególności moduły: optymalizacji odpornościowej, powierzchni odpowiedzi, zaawansowanych metod analizy niezawodności oraz metod symulacji losowych wykorzystujących koncepcję łacińskich hiperkostek, należy uznać za oryginalne osiągnięcie.*

Przedstawiona tematyka wydaje się mieć bardzo dobre perspektywy rozwoju. Jako obiecujące kierunki dalszych badań należy uznać:

- W optymalizacji odpornościowej procesów głębokiego tłoczenia blachy, - dokładne przetestowanie możliwości użycia alternatywnych metod estymacji momentów statystycznych funkcji losowych, takich jak metoda rozwinięcia w chaos wielomianowy czy metoda redukcji wymiarów.
- Zastosowanie niestacjonarnej funkcji korelacji w aproksymacji metodą krigingu.
- Wykorzystanie w optymalizacji odpornościowej algorytmów optymalizacyjnych przystosowanych do użytej metody aproksymacji, takich jak

algorytm EGO [112] w przypadku krigingu.

- Zastosowanie innych metod optymalizacji wielokryterialnej, np. MOGA (*multi objective genetic algorithm*), które pozwalają wyznaczyć pełen zbiór rozwiązań niezdominowanych.

Trzeba jednak zdawać sobie sprawę z pewnych zagrożeń, które mogą hamować szerszy rozwój zastosowań analizy niezawodności i optymalizacji odpornościowej w praktyce projektowej. O ile wzrost mocy obliczeniowej komputerów jest jednym z głównych motorów upowszechniania się analizy stochastycznej, o tyle skutkiem ubocznym tego wzrostu są coraz dokładniejsze modele skończenie elementowe. Nadmiernie szczegółowe modele redukują przyrost efektywności metod stochastycznych, możliwy do osiągnięcia przy stosowaniu prostszych modeli. W opinii autora, samo zwiększanie gęstości siatki lub złożoności stosowanych elementów skończonych, nie jest w stanie dostarczyć tak jakościowo nowych informacji jak wyniki szeroko rozumianej analizy stochastycznej. Znajomość wpływu, jaki na projektowaną konstrukcję ma losowy charakter jej parametrów oraz obciążeń lub też oszacowanie prawdopodobieństwa awarii, pozwalają dużo lepiej zrozumieć zachowanie badanego układu konstrukcyjnego, niż analiza deterministyczna, wykorzystująca nawet najbardziej złożony model skończenie elementowy.

A

Wybrane zagadnienia komputerowej analizy niezawodności konstrukcji

A.1. Sformułowanie problemu analizy niezawodności

Ze względu na stochastyczny charakter większości parametrów opisujących modele układów mechanicznych oraz działających na nie obciążeń, wielkości te w zadaniu analizy niezawodności konstrukcji traktuje się najczęściej jako zmienne losowe. Oznaczone zazwyczaj wielkimi literami X_1, X_2, \dots, X_n , tworzą wektor podstawowych zmiennych losowych \mathbf{X} , którego realizacje $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ są elementami przestrzeni Euklidesowej. Funkcja łącznej gęstości prawdopodobieństwa wektora \mathbf{X} , $f_{\mathbf{X}}(\mathbf{x})$, definiuje w przestrzeni \mathcal{X} miarę prawdopodobieństwa.

W zależności od realizacji zmiennych podstawowych, praca konstrukcji bądź spełnia wymagania projektowe, bądź nie – co prowadzi do szeroko rozumianego stanu awarii. Kryterium awarii konstrukcji wyraża się zazwyczaj za pomocą równania $g(\mathbf{x}) = 0$ definiującego hiperpowierzchnię zwaną powierzchnią graniczną w przestrzeni \mathcal{X} . Dzieli ona przestrzeń \mathcal{X} na dwa obszary: obszar awarii $\mathcal{D}_f = \{\mathbf{x} : g(\mathbf{x}) \leq 0\}$ oraz obszar bezpieczny $\mathcal{D}_s = \{\mathbf{x} : g(\mathbf{x}) > 0\}$. Prawdopodobieństwo awarii konstrukcji można zatem określić za pomocą następującej całki:

$$P_f = \mathbb{P}[\mathbf{X} \in \mathcal{D}_f] = \mathbb{P}[g(\mathbf{X}) \leq 0] = \int_{\mathcal{D}_f} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \quad (\text{A.1})$$

gdzie $\mathbb{P}[A]$ jest prawdopodobieństwem zdarzenia losowego A .

W praktycznych zastosowaniach, w przypadku gdy model stochastyczny zagadnienia składa się z wielu zmiennych losowych, obszar całkowania \mathcal{D}_f nie jest dany w sposób jawny, a samo obliczenie wartości funkcji granicznej jest kosztowne (np. wymaga rozwiązania złożonego zadania MES), wtedy całkowanie numeryczne równania (A.1) skazane jest z góry na niepowodzenie. Z uwagi

na powyższe, zaproponowano szereg metod przybliżonych służących szacowaniu wartości prawdopodobieństwa awarii. Większość najczęściej stosowanych metod wykorzystuje odpowiednią transformację $\mathbf{U} = \mathbf{T}(\mathbf{X})$ (zob. np. [50, 160, 173]) zadania analizy niezawodności z przestrzeni \mathcal{X} do przestrzeni \mathcal{U} , gdzie miara prawdopodobieństwa określona jest przez funkcję gęstości $f_{\mathbf{U}}(\mathbf{u}) = \prod_{i=1}^n \varphi(u_i)$ będącą iloczynem n jednowymiarowych standardowych normalnych funkcji gęstości prawdopodobieństwa zmiennych $U_i = T_i(\mathbf{X})$, (zob. podrozdział A.2). Ponieważ transformacja zmiennych pociąga za sobą również transformację warunku granicznego do przestrzeni \mathcal{U} , $g(\mathbf{x}) = 0 \rightarrow h(\mathbf{u}) = g[\mathbf{T}^{-1}(\mathbf{u})] = 0$, prawdopodobieństwo awarii wyraża się jako

$$P_f = \mathbb{P}[\mathbf{U} \in \Delta_f] = \mathbb{P}[h(\mathbf{U}) \leq 0] = \int_{\{\mathbf{U}:h(\mathbf{U}) \leq 0\}} f_{\mathbf{U}}(\mathbf{u}) \, d\mathbf{u}, \quad (\text{A.2})$$

gdzie Δ_f jest obszarem awarii w przestrzeni \mathcal{U} . Dzięki osiowej symetrii funkcji $f_{\mathbf{U}}(\mathbf{u})$, dla dowolnej funkcji liniowej $l(\mathbf{u}) = \beta - \boldsymbol{\alpha}^T \mathbf{u}$ prawdziwa jest następująca równość:

$$\mathbb{P}[l(\mathbf{U}) \leq 0] = \int_{\{\mathbf{U}:l(\mathbf{U}) \leq 0\}} f_{\mathbf{U}}(\mathbf{u}) \, d\mathbf{u} = \Phi(-\beta), \quad (\text{A.3})$$

gdzie współczynniki, $-\alpha_i, i = 1, 2, \dots, n$, są składowymi znormalizowanego gradientu funkcji $l(\mathbf{u})$, tzn. $\boldsymbol{\alpha}^T \boldsymbol{\alpha} = 1$, $\beta = \text{sign}[l(\mathbf{0})]\delta$ jest równe (z dokładnością do znaku) odległości δ pomiędzy hiperpłaszczyzną $l(\mathbf{u}) = 0$ a początkiem układu współrzędnych w \mathcal{U} , zaś $\Phi(\cdot)$ jest dystrybuantą standardowego rozkładu normalnego. Wynika z tego, że liniowa aproksymacja transformowanej funkcji granicznej $h(\mathbf{u})$ w punkcie znajdującym się najbliżej początku układu współrzędnych stanowić może proste oszacowanie prawdopodobieństwa awarii konstrukcji

$$P_f = \mathbb{P}[h(\mathbf{U}) \leq 0] \approx \mathbb{P}[l(\mathbf{U}) \leq 0] = \Phi(-\beta), \quad (\text{A.4})$$

gdzie przez β przyjęto oznaczać tzw. wskaźnik niezawodności konstrukcji. Podejście bazujące na liniowej aproksymacji funkcji granicznej nazywa się metodą analizy niezawodności pierwszego rzędu (ang. *first order reliability method* - FORM). Ponieważ wskaźnik niezawodności występujący we wzorze (A.4) jest jedynie przybliżeniem wskaźnika “dokładnego” definiowanego jako

$$\beta = -\Phi^{-1}(P_f), \quad (\text{A.5})$$

dlatego, w celu odróżnienia, oznacza się go jako β_{FORM} lub β_{I} . Oczywiście, aby skorzystać ze wzoru (A.5) konieczna jest uprzednia znajomość dokładnej wartości prawdopodobieństwa awarii, co w większości rzeczywistych (nieakademickich) zadań analizy niezawodności nie jest możliwe. Stosuje się go jednak np.

w metodzie niezawodności drugiego rzędu czy też metodach rodziny Monte Carlo w celu obliczenia ekwiwalentnego wskaźnika niezawodności na podstawie obliczonej wcześniej przybliżonej wartości prawdopodobieństwa awarii. Wskaźnik β_{FORM} wyznaczany jest jako rozwiązanie następującego zadania minimalizacji z ograniczeniami:

$$\beta_{\text{FORM}} = \text{sign}[h(\mathbf{0})]\delta^*,$$

$$\delta^* \equiv \|\mathbf{u}^*\| = \min\|\mathbf{u}\| \quad \text{pod warunkiem: } h(\mathbf{u}) \leq 0. \quad (\text{A.6})$$

Punkt \mathbf{u}^* położony na powierzchni granicznej $h(\mathbf{u}) = 0$ i znajdujący się najbliżej początku układu współrzędnych $\mathbf{u} = \mathbf{0}$ nazywany jest punktem projektowym lub najbardziej prawdopodobnym punktem awarii (ang. *design point, most probable failure point*). Zadanie (A.6) może być teoretycznie rozwiązane przy pomocy dowolnego algorytmu optymalizacji nieliniowej (zob. [153] i [143], gdzie porównano różne metody optymalizacji) jednak w większości typowych przypadków stosuje się algorytmy oparte na metodzie Rackwitz-Fiesslera [1] lub na algorytmie sekwencyjnego programowania kwadratowego (zob. np. [215]).

W dalszej części tekstu, w celu lepszego wyjaśnienia metod analizy niezawodności prezentowanych w zasadniczej części pracy, omówiono najważniejsze zagadnienia dotyczące transformacji zmiennych losowych do przestrzeni standardowej, gradientowych algorytmów poszukiwania punktu projektowego oraz metod analizy niezawodności wykorzystujących symulacje losowe.

A.2. Transformacja zmiennych losowych

Jako gaussowską przestrzeń standardową \mathcal{U} zwykle określać się przestrzeń probabilistyczną, w której miara prawdopodobieństwa zdefiniowana jest przez następującą funkcję gęstości:

$$\varphi_n(\mathbf{u}, \mathbf{0}, \mathbf{I}) = \prod_{i=1}^n \varphi(u_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u_i^2\right] = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\mathbf{u}^T \mathbf{u}\right]. \quad (\text{A.7})$$

Funkcja ta jest iloczynem n niezależnych, standardowych zmiennych o rozkładzie normalnym U_1, \dots, U_n . Za Der Kiureghianem [40] można wymienić szereg własności przestrzeni standardowej, które są szczególnie istotne w analizie niezawodności konstrukcji:

1. Funkcja (A.7) jest osiowo symetryczna - jej warstwie tworzą współśrodkowe hipersfery (dla dwu zmiennych, okręgi) wokół początku układu współrzędnych.

2. Funkcja gęstości prawdopodobieństwa maleje wykładniczo z kwadratem odległości od początku układu.
3. Dla zadanego $u_i = b$, warunkowa gęstość rozkładu prawdopodobieństwa pozostałych zmiennych jest również gęstością standardowego rozkładu normalnego. Oznacza to, że funkcja gęstości prawdopodobieństwa w hiperpłaszczyźnie wyznaczonej przez $u_i = b$ maleje wykładniczo wraz z kwadratem odległości od osi u_i . Z własności osiowej symetrii wynika, że dla dowolnej hiperpłaszczyzny funkcja gęstości maleje wykładniczo wraz z kwadratem odległości od punktu będącego rzutem początku układu współrzędnych na tę hiperpłaszczyznę.
4. Prawdopodobieństwo zdarzenia polegającego na tym, że realizacja \mathbf{u} wektora \mathbf{U} należy do podprzestrzeni $u_i \geq b$ wydzielonej przez hiperpłaszczyznę $u_i = b$ wynosi $\Phi(-b)$. Podobnie jak w punkcie 3, z własności symetrii osiowej wynika, iż w przypadku podprzestrzeni wydzielonej przez dowolną hiperpłaszczyznę (niekoniecznie prostopadłą do jednej z osi), prawdopodobieństwo, że realizacja \mathbf{u} należy do podprzestrzeni nie zawierającej $\mathbf{u} = \mathbf{0}$ wynosi $\Phi(-b)$, gdzie b jest odległością hiperpłaszczyzny od początku układu współrzędnych.

Aby można było wykorzystać własności przestrzeni \mathcal{U} konieczna jest transformacja oryginalnych zmiennych losowych \mathbf{X} do standardowych zmiennych normalnych \mathbf{U} o funkcji gęstości danej przez (A.7). Transformacja $\mathbf{U} = \mathbf{T}(\mathbf{X})$ musi zapewniać równoważność sformułowania problemu analizy niezawodności, tzn. prawdopodobieństwo awarii zdefiniowane w przestrzeni zmiennych oryginalnych \mathcal{X} musi być równe prawdopodobieństwu awarii w przestrzeni \mathcal{U}

$$P_f = \int_{\mathcal{D}_f} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = \int_{\Delta_f} \prod_{i=1}^n \varphi(u_i) \, du_1 du_2 \dots du_n. \quad (\text{A.8})$$

Transformację obszarów awarii można zapisać symbolicznie jako

$$\mathcal{D}_f = \{\mathbf{x} : g(\mathbf{x}) \leq 0\} \longrightarrow \Delta_f = \{\mathbf{u} : h(\mathbf{u}) \leq 0\}, \quad (\text{A.9})$$

gdzie $h(\mathbf{u}) = 0$ jest powierzchnią graniczną po transformacji

$$g(\mathbf{x}) = 0 \longrightarrow g[\mathbf{T}^{-1}(\mathbf{u})] = h(\mathbf{u}) = 0. \quad (\text{A.10})$$

W celu zachowania miary prawdopodobieństwa (A.8), transformacja $\mathbf{U} = \mathbf{T}(\mathbf{X})$ musi być odwzorowaniem wzajemnie jednoznacznym (bijekcją) przestrzeni \mathcal{X} na

\mathbf{U} oraz musi spełniać następujący warunek:

$$f_{\mathbf{X}}(x_1^0, x_2^0, \dots, x_n^0) = \prod_{i=1}^n \varphi(u_i^0) |\mathbf{J}_{u,x}|, \quad (\text{A.11})$$

gdzie $\mathbf{u}^0 = \mathbf{T}(\mathbf{x}^0)$ dla każdego $\mathbf{x}^0 \in \mathcal{X}$, a $|\mathbf{J}_{u,x}|$ jest jacobianem transformacji $\mathbf{T}(\cdot)$. Rozważmy następnie kilka najczęściej spotykanych przypadków rozkładów prawdopodobieństwa wektora podstawowych zmiennych \mathbf{X} .

A.2.1. Zmienne o wielowymiarowym rozkładzie normalnym

Wielowymiarowy rozkład normalny zdefiniowany jest w sposób jednoznaczny przez wektor wartości oczekiwanych $\boldsymbol{\mu}_{\mathbf{X}}$ i macierz kowariancji $\mathbf{C}_{\mathbf{X}}$. Poszukuje się transformacji liniowej postaci $\mathbf{u} = \mathbf{a} + \mathbf{A}\mathbf{x}$, dla której $\boldsymbol{\mu}_{\mathbf{U}} = \mathbf{0}$ i $\mathbf{C}_{\mathbf{U}} = \mathbf{I}$. Obliczając wartości oczekiwane oraz macierz kowariancji funkcji tworzących liniową transformację otrzymuje się

$$\boldsymbol{\mu}_{\mathbf{U}} = \mathbf{a} + \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}, \quad (\text{A.12})$$

$$\mathbf{C}_{\mathbf{U}} = \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T = \mathbf{I}. \quad (\text{A.13})$$

Ponieważ $\mathbf{C}_{\mathbf{X}}$ jest z definicji macierzą dodatnio określoną można ją zatem przedstawić jako $\mathbf{C}_{\mathbf{X}} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^T$ gdzie $\widehat{\mathbf{L}}$ jest macierzą dolną trójkątną z dekompozycji Choleskiego $\mathbf{C}_{\mathbf{X}}$. Dokonując następnie podstawienia do równania (A.13) dostajemy

$$\mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T = (\mathbf{A}\widehat{\mathbf{L}})(\widehat{\mathbf{L}}^T\mathbf{A}^T) = (\mathbf{A}\widehat{\mathbf{L}})(\mathbf{A}\widehat{\mathbf{L}})^T = \mathbf{I}. \quad (\text{A.14})$$

Powyższa równość spełniona jest jeśli

$$\mathbf{A} = \widehat{\mathbf{L}}^{-1}. \quad (\text{A.15})$$

W konsekwencji, wektor \mathbf{a} wyznaczony z równania (A.12) wynosi

$$\mathbf{a} = -\widehat{\mathbf{L}}^{-1}\boldsymbol{\mu}_{\mathbf{X}}. \quad (\text{A.16})$$

Ostatecznie, transformacja zmiennych o dowolnym wielowymiarowym rozkładzie Gaussa do przestrzeni niezależnych standardowych zmiennych normalnych przybiera postać

$$\mathbf{U} = \widehat{\mathbf{L}}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}). \quad (\text{A.17})$$

Jeśli macierz kowariancji wyrazić za pomocą macierzy korelacji $\mathbf{R}_{\mathbf{X}} = [\rho_{ij}]$ jako $\mathbf{C}_{\mathbf{X}} = \mathbf{D}\mathbf{R}_{\mathbf{X}}\mathbf{D}$, gdzie $\mathbf{D} = [\sigma_{X_i}]$ jest diagonalną macierzą odchyłeń standardowych, otrzymuje się transformację (A.17) w równoważnej formie

$$\mathbf{U} = \mathbf{L}^{-1}\mathbf{D}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}), \quad (\text{A.18})$$

gdzie \mathbf{L} jest macierzą dolną trójkątną z dekompozycji Choleskiego macierzy korelacji $\mathbf{R}_{\mathbf{X}}$. Jakobian tego przekształcenia zapiszemy następująco:

$$\mathbf{J}_{u,x} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{L}^{-1}\mathbf{D}^{-1} = \widehat{\mathbf{L}}^{-1}. \quad (\text{A.19})$$

A.2.2. Zmienne niezależne

Jeżeli zmienne podstawowe są statystycznie niezależne, to transformacja o opisanych wcześniej własnościach sprowadza się do wzoru

$$\Phi(u_i) = F_{X_i}(x_i) \quad \Rightarrow \quad u_i = \Phi^{-1}[F_{X_i}(x_i)], \quad i = 1, \dots, n, \quad (\text{A.20})$$

gdzie F_{X_i} jest dystrybuantą zmiennej X_i . Jakobian transformacji (A.20) obliczamy różniczkując wyrażenie po lewej stronie względem x_i . Zgodnie z regułą różniczkowania funkcji złożonych, $\frac{\partial}{\partial x} = \frac{\partial}{\partial u} \frac{\partial u}{\partial x}$, dostaje się ostatecznie

$$\mathbf{J}_{u,x} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{f_{X_i}(x_i)}{\varphi(u_i)} \end{bmatrix}. \quad (\text{A.21})$$

A.2.3. Zmienne o dowolnym rozkładzie prawdopodobieństwa

W ogólnym przypadku, gdy nie zakłada się żadnego szczególnego rozkładu zmiennych \mathbf{X} , ani też nie postuluje się ich niezależności, na podstawie znajomości łącznej gęstości prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$ Hohenbichler i Rackwitz [93] zaproponowali użycie tzw. transformacji Rosenblatta, zob. [209], następującej postaci:

$$\Phi(u_1) = H_1(x_1) = F_1(x_1) = \int_{-\infty}^{x_1} f_1(t) dt, \quad (\text{A.22a})$$

$$\Phi(u_2) = H_2(x_2|x_1) = \int_{-\infty}^{x_2} \frac{f_2(x_1, t)}{f_1(x_1)} dt, \quad (\text{A.22b})$$

⋮

$$\Phi(u_i) = H_i(x_i|x_1, x_2, \dots, x_{i-1}) = \int_{-\infty}^{x_i} \frac{f_i(x_1, x_2, \dots, x_{i-1}, t)}{f_{i-1}(x_1, x_2, \dots, x_{i-1})} dt, \quad i = 1, \dots, n, \quad (\text{A.22c})$$

gdzie $f_i(x_1, x_2, \dots, x_i)$ jest gęstością brzegową wektora \mathbf{X} dla i pierwszych jego składowych

$$f_i(x_1, x_2, \dots, x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_{i+1} dx_{i+2} \dots dx_n. \quad (\text{A.23})$$

Ponieważ dystrybuanty $\Phi(\cdot)$, $F_1(\cdot)$ oraz dystrybuanty warunkowe $H_i(\cdot|\cdot)$ są monotonicznie rosnącymi funkcjami względem odpowiednich argumentów X_i , przekształcenie (A.22) jest wzajemnie jednoznaczne. W celu sprawdzenia warunku (A.11) należy najpierw wyznaczyć jakobian tej transformacji

$$\mathbf{J}_{u,x} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{f_1(x_1)}{\varphi(u_1)} \\ \frac{\partial u_2}{\partial x_1} & \frac{f_2(x_1, x_2)}{\varphi(u_2)f_1(x_1)} \\ \vdots & \vdots & \cdots \\ \frac{\partial u_n}{\partial x_1} & \frac{\partial u_n}{\partial x_2} & \cdots & \frac{f_n(x_1, x_2, \dots, x_n)}{\varphi(u_n)f_{n-1}(x_1, \dots, x_{n-1})} \end{bmatrix}. \quad (\text{A.24})$$

Obliczając wyznacznik powyższej macierzy dostajemy

$$|\mathbf{J}_{u,x}| = \frac{f_n(x_1, x_2, \dots, x_n)}{\prod_{i=1}^n \varphi(u_i)}, \quad (\text{A.25})$$

a co za tym idzie, warunek (A.11) jest spełniony dla każdego $\mathbf{u} = \mathbf{T}(\mathbf{x})$.

Transformacji Rosenblatta używa się jeśli znana jest łączna gęstość prawdopodobieństwa zmiennych \mathbf{X} lub jeśli dostępne są dystrybuanty warunkowe H_i . Łatwo sprawdzić, że w przypadku niezależnych zmiennych losowych transformacja (A.22) redukuje się do (A.20). Trzeba jednak zaznaczyć, że przedstawiona transformacja może zależeć od uporządkowania zmiennych w wektorze \mathbf{X} . Tak jak to zostało pokazane w pracy [50], konsekwencją tej niejednoznaczności mogą być różne postacie powierzchni granicznej $h(\mathbf{u}) = 0$, a co za tym idzie różne lokalizacje punktu projektowego i w konsekwencji różne oszacowania wartości prawdopodobieństwa awarii metodą pierwszego rzędu. W takiej sytuacji powinno się przeanalizować wszystkie $n!$ permutacji wektora \mathbf{X} i wybrać ten punkt projektowy, który leży najbliżej początku układu. Dla dużych n jest to oczywiście zadanie bardzo czasochłonne obliczeniowo.

A.2.4. Zmienne o dowolnym rozkładzie - transformacja przybliżona

Bardzo często ze względu na brak odpowiednich danych statystycznych znane są jedynie brzegowe gęstości prawdopodobieństwa zmiennych losowych oraz

ich wzajemne korelacje, natomiast nieznana jest funkcja łącznej gęstości prawdopodobieństwa. W takiej sytuacji stosowana jest przybliżona transformacja Natafa [183], użyta po raz pierwszy w analizie niezawodności konstrukcji przez Der Kiureghiana i Liu [151].

Na początku, za pomocą brzegowej transformacji zmiennych \mathbf{X} (A.20) definiuje się zbiór zmiennych \mathbf{Z} o rozkładzie standardowym normalnym.

$$Z_i = \Phi^{-1}[F_{X_i}(x_i)], \quad i = 1, \dots, n. \quad (\text{A.26})$$

Następnie zakłada się, że również funkcja łącznej gęstości prawdopodobieństwa zmiennych \mathbf{Z} jest funkcją gaussowską (oczywiście nie jest to zagwarantowane przez normalne rozkłady brzegowe zmiennych). Co za tym idzie, wyrażenie na gęstość prawdopodobieństwa zmiennych \mathbf{X} otrzymuje się z elementarnej zasady transformacji funkcji gęstości $f_{\mathbf{X}}(\mathbf{x}) dx_1 dx_2 \cdots dx_n = \varphi(\mathbf{z}, \mathbf{R}_{\mathbf{Z}}) dz_1 dz_2 \cdots dz_n$

$$f_{\mathbf{X}}(\mathbf{x}) = \varphi(\mathbf{z}, \mathbf{R}_{\mathbf{Z}}) \frac{dz_1 dz_2 \cdots dz_n}{dx_1 dx_2 \cdots dx_n} = \varphi(\mathbf{z}, \mathbf{R}_{\mathbf{Z}}) \frac{f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)}{\varphi(z_1)\varphi(z_2) \cdots \varphi(z_n)}, \quad (\text{A.27})$$

gdzie $\varphi(\mathbf{z}, \mathbf{R}_{\mathbf{Z}})$ jest funkcją łącznej gęstości rozkładu normalnego zmiennych \mathbf{Z} o macierzy korelacji $\mathbf{R}_{\mathbf{Z}}$, którą należy wyznaczyć. Ostateczna postać równania (A.27) wynika z różniczkowania wyrażenia (A.26) względem x_i . Za pomocą przekształcenia (A.18), wykorzystując wzór (A.26) oraz ze względu na własności zmiennych standardowych, skorelowane zmienne losowe \mathbf{Z} może przetransformować na nieskorelowane zmienne \mathbf{U} w następujący sposób:

$$\mathbf{U} = \mathbf{L}_{\mathbf{Z}}^{-1} \mathbf{Z} = \mathbf{L}_{\mathbf{Z}}^{-1} \begin{bmatrix} \Phi^{-1}[F_{X_1}(x_1)] \\ \Phi^{-1}[F_{X_2}(x_2)] \\ \vdots \\ \Phi^{-1}[F_{X_n}(x_n)] \end{bmatrix}, \quad (\text{A.28})$$

gdzie $\mathbf{L}_{\mathbf{Z}}$ jest macierzą dolną trójkątną utworzoną przez dekompozycję Choleskiego macierzy $\mathbf{R}_{\mathbf{Z}}$. Łatwo jest sprawdzić (zob. (A.19) i (A.21)), że jacobian takiego przekształcenia wynosi

$$\mathbf{J}_{u,x} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{L}_{\mathbf{Z}}^{-1} \begin{bmatrix} f_{X_i}(x_i) \\ \varphi(z_i) \end{bmatrix}. \quad (\text{A.29})$$

Do wyliczenia pozostają współczynniki korelacji zmiennych \mathbf{Z} zebrane w macierzy $\mathbf{R}_{\mathbf{Z}} = [\rho'_{ij}]$ w funkcji elementów macierzy $\mathbf{R}_{\mathbf{X}} = [\rho_{ij}]$. Współczynnik korelacji pomiędzy dwoma ciągłymi zmiennymi losowymi X_i i X_j zapisać można jako

$$\rho_{ij} \equiv \rho_{X_i X_j} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_i - \mu_{X_i}}{\sigma_{X_i}} \right) \left(\frac{x_j - \mu_{X_j}}{\sigma_{X_j}} \right) f_{X_i X_j}(x_i, x_j) dx_i dx_j, \quad (\text{A.30})$$

gdzie $f_{X_i X_j}(x_i, x_j)$ jest dwuwymiarową funkcją gęstości prawdopodobieństwa zmiennych X_i and X_j . Na podstawie (A.27) dostaje się równanie całkowe uzależniające ρ_{ij} od ρ'_{ij}

$$\begin{aligned} \rho_{ij} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{x_i - \mu_{X_i}}{\sigma_{X_i}} \right) \left(\frac{x_j - \mu_{X_j}}{\sigma_{X_j}} \right) \varphi_2(z_i, z_j, \rho'_{ij}) \frac{f_{X_i}(x_i) f_{X_j}(x_j)}{\varphi(z_i)\varphi(z_j)} dx_i dx_j \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{F_{X_i}^{-1}(\Phi(z_i)) - \mu_{X_i}}{\sigma_{X_i}} \right] \left[\frac{F_{X_j}^{-1}(\Phi(z_j)) - \mu_{X_j}}{\sigma_{X_j}} \right] \varphi_2(z_i, z_j, \rho'_{ij}) dz_i dz_j, \end{aligned} \quad (\text{A.31})$$

gdzie $F^{-1}(\cdot)$ jest funkcją odwrotną do dystrybuanty, a gaussowska funkcja łącznej gęstości prawdopodobieństwa dwóch zmiennych standardowych ma postać

$$\varphi_2(z_i, z_j, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[-\frac{z_i^2 - 2z_i z_j \rho + z_j^2}{2(1-\rho^2)} \right]. \quad (\text{A.32})$$

W celu obliczenia poszukiwanego współczynnika korelacji równanie (A.31) należy rozwiązać ze względu na ρ'_{ij} . Dla wybranych par rozkładów prawdopodobieństwa w pracy [151] znaleziono analityczne rozwiązania tego równania. Wykazano, że za wyjątkiem rozkładów odbiegających bardzo od normalnego, wartości ρ_{ij} i ρ'_{ij} nie różnią się znacznie.

A.3. Algorytmy poszukiwania punktu projektowego

Precyzyjne wyznaczenie punktu projektowego \mathbf{u}^* ma niezwykle istotne znaczenie w wielu metodach analizy niezawodności, przede wszystkim w metodach pierwszego i drugiego rzędu (FORM, SORM) oraz w niektórych metodach redukcji wariancji, takich jak importance sampling. Położony na powierzchni granicznej $h(\mathbf{u}) = 0$ najbliższej początku układu współrzędnych w gaussowskiej przestrzeni standardowej punkt projektowy zdefiniowany jest jako rozwiązanie zadania minimalizacji (A.6). Poniżej omówione zostaną wybrane najpopularniejsze algorytmy lokalizacji \mathbf{u}^* . Chociaż ze względu na specyfikę funkcji granicznych rozpatrywanych w tej pracy (są drogie obliczeniowo, zawierają szum) klasyczne metody znajdowania punktu projektowego nie mogą być stosowane bezpośrednio, to jednak w połączeniu z odpowiednią metodą aproksymacji funkcji granicznej, stanowią kluczowy element proponowanych algorytmów.

Można wymienić wiele prac mających na celu porównanie efektywności różnych algorytmów poszukiwania punktu projektowego. Większość z nich pochodzi od badaczy związanych z pracownikami naukowymi Rackwiza oraz Der Kiureghiana [1, 87, 153, 276]. Obecnie, ze względu na swą efektywność oraz łatwość implementacji najczęściej stosowane algorytmy bazują na standardowej metodzie Rackwiza-Fiesslera (używany jest czasem również skrót HLRF od nazwisk Hasofer, Lind, Rackwitz, Fiessler).

Po krótkim omówieniu iteracyjnej metody HLRF w dalszej części podrozdziału przedstawione zostaną wybrane usprawnienia tego algorytmu, mające na celu eliminację występujących w niektórych sytuacjach problemów ze zbieżnością. Większość usprawnień dotyczy dodania do oryginalnego algorytmu procedury minimalizacji kierunkowej.

Najważniejsze operacje gradientowego algorytmu poszukiwania punktu projektowego dane są następująco:

1. Transformacja współrzędnych punktu początkowego $\mathbf{x}^{(0)}$ w przestrzeni \mathcal{X} na odpowiadający mu punkt $\mathbf{u}^{(0)}$ w przestrzeni \mathcal{U} .
2. Obliczenie oraz zapamiętanie (jako stałej normalizującej) wartości funkcji granicznej $g(\mathbf{x}^{(0)})$.
3. Dopóki nie spełniony jest warunek zbieżności, na typowej k -tej iteracji:
 - 3.1 Transformacja odwrotna wektora $\mathbf{u}^{(k)}$ z przestrzeni standardowej na $\mathbf{x}^{(k)}$ w przestrzeni oryginalnej. Operacja ta pomijana jest w iteracji początkowej, kiedy wektor $\mathbf{x}^{(0)}$ jest już znany.
 - 3.2 Obliczenie wartości $g(\mathbf{x}^{(k)})$. W większości przypadków konieczne jest przeprowadzenia analizy metodą elementów skończonych, co z kolei wymaga wygenerowania (przy pomocy odpowiedniego interfejsu) zbioru z danymi do programu MES, który zawierać będzie aktualną realizację wektora zmiennych losowych.
 - 3.3 Obliczenie gradientu funkcji granicznej $\nabla g(\mathbf{x}^{(k)}) = \partial g / \partial \mathbf{x} |_{\mathbf{x}=\mathbf{x}^{(k)}}$ albo za pomocą metody bezpośredniego różniczkowania (zob. np. [122]) lub też za pomocą schematu różnic skończonych. Gradient transformowany jest następnie do standardowej przestrzeni normalnej za pomocą macierzy jacobianu transformacji $\mathbf{J}_{u,x}$, $\nabla h(\mathbf{u}^{(k)}) = (\mathbf{J}_{u,x}^{-1})^T \nabla g(\mathbf{x}^{(k)})$, (zob. podrozdział A.2).
 - 3.4 Sprawdzenie warunku zbieżności.
 - 3.5 Jeśli warunek zbieżności nie jest spełniony, wybranie nowego przybliżenia punktu projektowego jako $\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}$, gdzie $\Delta \mathbf{u}^{(k)}$ jest kierunkiem poszukiwań, a $\xi^{(k)}$ jest długością kroku.

W celu łatwiejszego wyprowadzenia wzorów metody HLRF wygodnie jest zapisać zadanie optymalizacji (A.6) w równoważnej postaci jako:

$$\text{znaleźć wektor: } \mathbf{u}, \quad (\text{A.33})$$

$$\text{minimalizujący: } Q(\mathbf{u}) = \|\mathbf{u}\|^2 = \mathbf{u}^T \mathbf{u}, \quad (\text{A.34})$$

$$\text{pod warunkiem: } h(\mathbf{u}) = 0. \quad (\text{A.35})$$

Rozwinięcie kwadratowej funkcji celu $Q(\mathbf{u})$ w punkcie $\mathbf{u}^{(k)}$ (k -ta iteracja wektora \mathbf{u}) w szereg Taylora i linearyzacja funkcji ograniczenia $h(\mathbf{u})$ prowadzą do następującego zadania minimalizacji, którego rozwiązaniem jest optymalny przyrost $\Delta \mathbf{u}^{(k)}$:

$$\text{znaleźć wektor: } \Delta \mathbf{u}^{(k)}, \quad (\text{A.36})$$

$$\begin{aligned} \text{minimalizujący: } \tilde{Q}(\Delta \mathbf{u}^{(k)}) &= Q(\mathbf{u}^{(k)}) + \nabla Q^T(\mathbf{u}^{(k)}) \Delta \mathbf{u}^{(k)} \\ &\quad + \frac{1}{2} \Delta \mathbf{u}^{(k)T} \nabla^2 Q(\mathbf{u}^{(k)}) \Delta \mathbf{u}^{(k)} \\ &= \mathbf{u}^{(k)T} \mathbf{u}^{(k)} + 2\mathbf{u}^{(k)T} \Delta \mathbf{u}^{(k)} + \Delta \mathbf{u}^{(k)T} \Delta \mathbf{u}^{(k)}, \end{aligned} \quad (\text{A.37})$$

$$\text{pod warunkiem: } \bar{h}(\Delta \mathbf{u}^{(k)}) = h(\mathbf{u}^{(k)}) + \nabla h^T(\mathbf{u}^{(k)}) \Delta \mathbf{u}^{(k)} = 0. \quad (\text{A.38})$$

Funkcja Lagrange'a dla zadania (A.36)–(A.38) ma postać

$$\begin{aligned} L(\Delta \mathbf{u}^{(k)}, \lambda^{(k)}) &= \mathbf{u}^{(k)T} \mathbf{u}^{(k)} + 2\mathbf{u}^{(k)T} \Delta \mathbf{u}^{(k)} + \Delta \mathbf{u}^{(k)T} \Delta \mathbf{u}^{(k)} \\ &\quad - \lambda^{(k)} \left(h(\mathbf{u}^{(k)}) + \nabla h^T(\mathbf{u}^{(k)}) \Delta \mathbf{u}^{(k)} \right), \end{aligned} \quad (\text{A.39})$$

natomiast warunki Kuhna-Tuckera dane są równaniami:

$$\nabla L = 2\mathbf{u}^{(k)} + 2\Delta \mathbf{u}^{(k)} - \lambda^{(k)} \nabla h(\mathbf{u}^{(k)}) = 2\mathbf{u}^{(k+1)} - \lambda^{(k)} \nabla h(\mathbf{u}^{(k)}) = 0, \quad (\text{A.40})$$

$$\begin{aligned} h(\mathbf{u}^{(k)}) + \nabla h^T(\mathbf{u}^{(k)}) \Delta \mathbf{u}^{(k)} &= h(\mathbf{u}^{(k)}) + \nabla h^T(\mathbf{u}^{(k)}) \mathbf{u}^{(k+1)} \\ &\quad - \nabla h^T(\mathbf{u}^{(k)}) \mathbf{u}^{(k)} = 0. \end{aligned} \quad (\text{A.41})$$

W punkcie optymalnym mnożnik Lagrange'a spełniać musi warunek (zob. [6])

$$\nabla \tilde{Q} = \lambda^{(k)} \nabla \bar{h}, \quad (\text{A.42})$$

a więc, $\lambda^{(k)}$ wyraża się przez

$$\lambda^{(k)} = \frac{2\nabla h^T(\mathbf{u}^{(k)}) \mathbf{u}^{(k+1)}}{\nabla h^T(\mathbf{u}^{(k)}) \nabla h(\mathbf{u}^{(k)})}. \quad (\text{A.43})$$

Następnie, na podstawie (A.41) można zastąpić wyrażenie $\nabla h^T(\mathbf{u}^{(k)})\mathbf{u}^{(k+1)}$ w (A.43) przez $\nabla h^T(\mathbf{u}^{(k)})\mathbf{u}^{(k)} - h(\mathbf{u}^{(k)})$. Podstawiając tak przekształcony wzór na $\lambda^{(k)}$ do (A.40) otrzymuje się podstawową formułę iteracyjną metody HLRF

$$\mathbf{u}^{(k+1)} = \frac{1}{\|\nabla h(\mathbf{u}^{(k)})\|^2} \left(\nabla h^T(\mathbf{u}^{(k)})\mathbf{u}^{(k)} - h(\mathbf{u}^{(k)}) \right) \nabla h(\mathbf{u}^{(k)}). \quad (\text{A.44})$$

Optymalny przyrost dany jest zatem jako

$$\Delta \mathbf{u}^{(k)} = \left(\frac{h(\mathbf{u}^{(k)})}{\|\nabla h(\mathbf{u}^{(k)})\|} + \boldsymbol{\alpha}^{(k)T} \mathbf{u}^{(k)} \right) \boldsymbol{\alpha}^{(k)} - \mathbf{u}^{(k)}, \quad (\text{A.45})$$

gdzie $\boldsymbol{\alpha}^{(k)} = -\nabla h(\mathbf{u}^{(k)})/\|\nabla h(\mathbf{u}^{(k)})\|$ jest jednostkowym wektorem, przeciwnym do znormalizowanego gradientu funkcji granicznej w punkcie $\mathbf{u}^{(k)}$.

Proces iteracyjny kontynuuje się, aż do spełnienia warunków zbieżności. Po pierwsze, znaleziony punkt powinien znajdować się na powierzchni granicznej $h(\mathbf{u}) = 0$, po drugie, powinien leżeć jak najbliżej początku układu współrzędnych. Pierwszy warunek można zapisać następująco:

$$\left| \frac{h(\mathbf{u}^{(k)})}{H_0} \right| < \epsilon, \quad (\text{A.46})$$

gdzie wartość w mianowniku jest stałą skalującą, a ϵ jest dopuszczalną tolerancją. Zazwyczaj przyjmuje się $H_0 = h(\mathbf{u}^{(0)})$ lub $H_0 = 1$ jeśli $h(\mathbf{u}^{(0)}) = 0$. Drugi warunek zbieżności równoważny jest wymaganiu, aby w punkcie projektowym wektor gradientu skierowany był do początku układu współrzędnych w przestrzeni \mathcal{U} . Można to wyrazić za pomocą wektorów $\mathbf{u}^{(k)}$ i $\boldsymbol{\alpha}^{(k)}$. W punkcie projektowym kąt θ pomiędzy tymi dwoma wektorami równy jest zero, a ponieważ $\cos(\theta) = \frac{\boldsymbol{\alpha}^{(k)T} \mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|}$, drugie kryterium zbieżności przybiera postać

$$1 - \frac{\boldsymbol{\alpha}^{(k)T} \mathbf{u}^{(k)}}{\|\mathbf{u}^{(k)}\|} < \epsilon. \quad (\text{A.47})$$

Ponieważ wykazano, zob. [152], że oryginalny algorytm HLRF bez kontroli przyrostu (A.45) może w przypadku niektórych funkcji granicznych zachowywać się niestabilnie, oscylując wokół rozwiązania, dlatego jego modyfikacje dotyczyły przede wszystkim metod redukcji długości kroku. Dokonuje się tego minimalizując w kierunku $\Delta \mathbf{u}^{(k)}$ funkcję celu rozszerzoną o odpowiednią funkcję kary, która

uwzględnia niespełnienie ograniczenia $h(\mathbf{u}) = 0$. Ponieważ dokładne wyznaczenie minimum w optymalizacji kierunkowej może być zbyt kosztowne, dlatego często używa się mniej dokładnej, ale bardzo efektywnej reguły Armijo [5]

$$\Psi^{(k)}(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}) - \Psi^{(k)}(\mathbf{u}^{(k)}) \leq a \xi^{(k)} (\nabla \Psi(\mathbf{u}^{(k)})^T \Delta \mathbf{u}^{(k)}), \quad (\text{A.48})$$

gdzie $\Psi^{(k)}(\cdot)$ jest wartością rozszerzonej funkcji celu w k -tym kroku iteracyjnym, $a > 0$ jest stałą (przeważnie przyjmuje się $a = 0.5$), długość kroku $\xi^{(k)}$ wyznacza się jako $\xi^{(k)} = b^j$, gdzie $0.0 < b < 1.0$, a j jest liczbą całkowitą o wartości początkowej równej 0. Wykładnik j zwiększany jest o 1 aż do znalezienia zadowalającej długości przyrostu, która spełnia warunek (A.48). Przyjmując $b = 0.5$ krok dzielony jest na pół za każdym razem, gdy poprzednia jego długość nie spełnia kryterium Armijo.

Poniżej, w punktach A.3.1 oraz A.3.2, zaprezentowano dwie najczęściej stosowane rozszerzone funkcje celu.

A.3.1. Metoda ARF

W pracy [1] Abdo i Rackwitz zaproponowali aby w procedurze poszukiwania optymalnej długości kroku algorytmu HLRF wykorzystać funkcję znaną z metody sekwencyjnego programowania kwadratowego, zob. [214]. Od nazwisk autorów podejście to oznaczać będziemy skrótem ARF (Abdo, Rackwitz, Fiessler).

Uwzględniając definicję (A.34), rozszerzona funkcja celu dana jest jako

$$\begin{aligned} \Psi^{(k)}(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}) &= \|\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}\|^2 + \lambda^{(k)} h(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}) \\ &\quad + \frac{1}{2} r^{(k)} h(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)})^2, \end{aligned} \quad (\text{A.49})$$

gdzie $r^{(k)}$ jest parametrem kary postaci

$$r^{(k)} = \max \left\{ |\lambda^{(k)}|, \frac{1}{2} (r^{(k-1)} + |\lambda^{(k)}|) \right\}. \quad (\text{A.50})$$

Długość kroku $\xi^{(k)}$ minimalizująca funkcję (A.49) jest następnie przyjmowana do określenia kolejnej iteracji punktu projektowego

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}. \quad (\text{A.51})$$

A.3.2. Metoda iHLRF

Inna rozszerzona funkcja celu użyta została przez Zhanga i Der Kiureghiana [276]. Pokazano, że zapewnia ona zbieżność algorytmu HLRF i tak rozbudowaną metodę oznaczono jako iHLRF (i - od angielskiego słowa *improved* - udoskonalony). Funkcja zdefiniowana jest następująco:

$$\Psi^{(k)}(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}) = \frac{1}{2} \|\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)}\|^2 + c |h(\mathbf{u}^{(k)} + \xi^{(k)} \Delta \mathbf{u}^{(k)})|, \quad (\text{A.52})$$

gdzie $c > 0$ jest parametrem kary. W [276] wykazano, że aby kierunek poszukiwań był również kierunkiem spadku funkcji (A.52), parametr kary musi spełniać warunek

$$c \geq \frac{\|\mathbf{u}^{(k)}\|}{\|\nabla h(\mathbf{u}^{(k)})\|}, \quad (\text{A.53})$$

W programie OpenSees [169], rozwijanym w ramach otwartego projektu badawczego poświęconemu modelowaniu odpowiedzi konstrukcji inżynierskich na wymuszenia sejsmiczne, zaimplementowano algorytm iHLRF, w którym parametr c obliczany jest ze wzoru

$$c = \gamma \frac{\|\mathbf{u}^{(k)}\|}{\|\nabla h(\mathbf{u}^{(k)})\|} + \eta, \quad (\text{A.54})$$

gdzie $\gamma > 1$ i $\eta \geq 0$ są specyfikowane przez użytkownika, tak by spełnić ograniczenie (A.53). Jako wartości domyślne przyjmuje się, $\gamma = 2$ i $\eta = 10$, zob. [87].

A.4. Metody symulacyjne

Pod nazwą *metody symulacyjne* lub też *symulacje Monte Carlo* w odniesieniu do analizy niezawodności konstrukcji rozumieć będziemy grupę technik wykorzystujących losowe generowanie realizacji wektora \mathbf{X} w celu oszacowania wartości prawdopodobieństwa awarii. Metody te bardzo często stanowią interesującą alternatywę dla metod FORM/SORM, a w przypadku wielu problemów analizy niezawodności okazują się być jedynym możliwym sposobem rozwiązania.

Pomimo, że metody symulacyjne zyskują coraz większą popularność, między innymi ze względu na rozwój mocy obliczeniowej komputerów, to wciąż potrzeba bardzo wielu (od dziesiątek tysięcy do milionów) obliczeń funkcji granicznej, aby otrzymać wiarygodną estymację wyników przy pomocy klasycznej

metody Monte Carlo. Ma to miejsce szczególnie wtedy, gdy szacowana wartość prawdopodobieństwa awarii jest mała. W związku z tym, w praktyce stosuje się szereg metod redukcji wariancji, takich jak importance sampling, gdzie realizacje zmiennych losowych generowane są przede wszystkim w pobliżu powierzchni granicznej.

Metody symulacyjne są niezastąpione w przypadku, gdy funkcja graniczna jest nieróżniczkowalna. Zastosowanie mają wtedy różne adaptacyjne metody importance sampling, w których funkcja gęstości losującej modyfikowana jest na podstawie wyników z dotychczas wygenerowanych próbek.

W następnych podrozdziałach przedstawione zostaną podstawowe informacje na temat klasycznej metody Monte Carlo oraz kilku odmian metody importance sampling, ważnych z punktu widzenia prezentowanych w zasadniczej części pracy metod analizy niezawodności.

A.4.1. Klasyczna metoda Monte Carlo

Istota klasycznej metody Monte Carlo (dla podkreślenia jej “pierwotnego” charakteru używa się czasami określeń: standardowa, zgrubna - ang. *crude*) polega na generowaniu realizacji \mathbf{x} losowego wektora \mathbf{X} zgodnie z funkcją gęstości prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$ oraz sprawdzaniu czy realizacje te prowadzą do awarii, zob. rys. A.1. Stosunek liczby realizacji, które znalazły się w obszarze awarii do liczby wszystkich wygenerowanych realizacji stanowi prosty estymator prawdopodobieństwa awarii. W celu zapisania podstawowych własności tego estymatora wygodnie jest posłużyć się funkcją charakterystyczną obszaru awarii \mathcal{D}_f

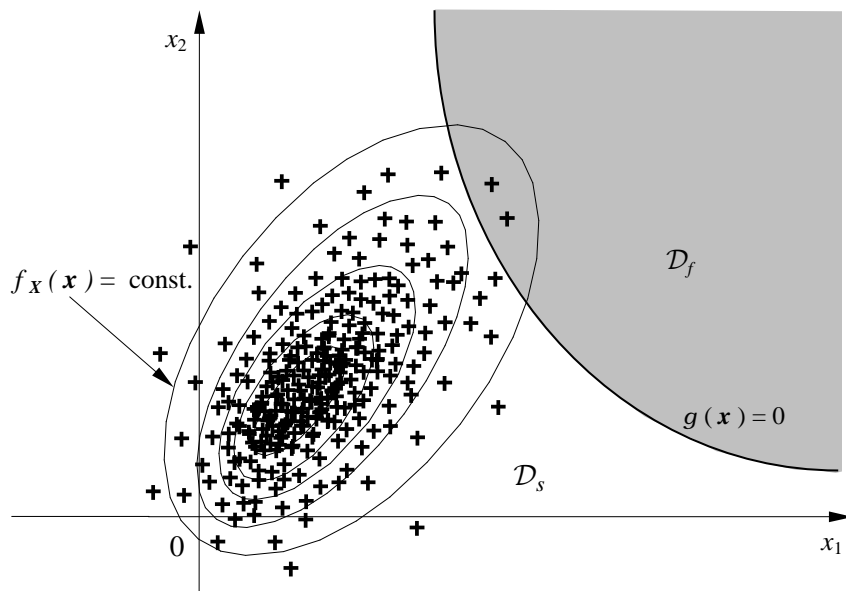
$$I_{\mathcal{D}_f}(\mathbf{x}) = \begin{cases} 1 & \text{gdy } \mathbf{x} \in \mathcal{D}_f, \\ 0 & \text{gdy } \mathbf{x} \notin \mathcal{D}_f, \end{cases} \quad (\text{A.55})$$

i wyrazić prawdopodobieństwo awarii (A.1) jako

$$P_f = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_n I_{\mathcal{D}_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (\text{A.56})$$

Funkcja $I_{\mathcal{D}_f}(\mathbf{X})$ może być zatem traktowana jak zmienna losowa o rozkładzie dwupunktowym

$$\mathbb{P}[I_{\mathcal{D}_f}(\mathbf{X}) = 1] = P_f, \quad \mathbb{P}[I_{\mathcal{D}_f}(\mathbf{X}) = 0] = 1 - P_f, \quad (\text{A.57})$$



Rys. A.1. Klasyczna metoda Monte Carlo. Symulacje losowe w przestrzeni oryginalnej, $n = 2$.

gdzie $P_f = \mathbb{P}[\mathbf{X} \in \mathcal{D}_f]$. Wartość oczekiwana oraz wariancja $I_{\mathcal{D}_f}(\mathbf{X})$ wynoszą

$$\mathbb{E}[I_{\mathcal{D}_f}(\mathbf{X})] = 1 \cdot P_f + 0 \cdot (1 - P_f) = P_f, \quad (\text{A.58})$$

$$\text{Var}[I_{\mathcal{D}_f}(\mathbf{X})] = \mathbb{E}[I_{\mathcal{D}_f}^2(\mathbf{X})] - (\mathbb{E}[I_{\mathcal{D}_f}(\mathbf{X})])^2 = P_f - P_f^2 = P_f(1 - P_f). \quad (\text{A.59})$$

W celu obliczenia prawdopodobieństwa awarii w metodzie Monte Carlo wprowadza się następujący estymator wartości oczekiwanej funkcji charakterystycznej:

$$\widehat{\mathbb{E}}[I_{\mathcal{D}_f}(\mathbf{X})] = \frac{1}{N} \sum_{k=1}^N I_{\mathcal{D}_f}(\mathbf{X}^{(k)}) = \widehat{P}_f, \quad (\text{A.60})$$

gdzie $\mathbf{X}^{(k)}$, $k = 1, \dots, N$, są niezależnymi wektorami losowymi o gęstości prawdopodobieństwa $f_{\mathbf{X}}(\mathbf{x})$. Jak łatwo pokazać, wartość oczekiwana i wariancja estymatora (A.60) dane są wzorami

$$\mathbb{E}(\widehat{P}_f) = \frac{1}{N} \sum_{k=1}^N \mathbb{E}[I_{\mathcal{D}_f}(\mathbf{X}^{(k)})] = \frac{1}{N} N P_f = P_f, \quad (\text{A.61})$$

$$\text{Var}(\widehat{P}_f) = \frac{1}{N^2} \sum_{k=1}^N \text{Var}[I_{\mathcal{D}_f}(\mathbf{X}^{(k)})] = \frac{1}{N^2} N P_f(1 - P_f) = \frac{1}{N} P_f(1 - P_f). \quad (\text{A.62})$$

Estymator (A.60) jest więc nieobciążony, a odpowiadający mu współczynnik zmienności ma postać

$$\nu_{\hat{P}_f} = \frac{\sqrt{\text{Var}(\hat{P}_f)}}{\mathbb{E}(\hat{P}_f)} = \sqrt{\frac{1 - P_f}{NP_f}}. \quad (\text{A.63})$$

Przekształcając równanie (A.63) dostaje się następujący wzór na liczbę symulacji (wielkość próbki) N :

$$N = \frac{1 - P_f}{\nu_{\hat{P}_f}^2 P_f}. \quad (\text{A.64})$$

Co ciekawe, liczba symulacji N konieczna dla zapewnienia określonej “jakości” wyników, mierzona współczynnikiem zmienności $\nu_{\hat{P}_f}$, nie zależy od liczby zmiennych losowych n . Z drugiej strony, równanie (A.64) ukazuje największą słabość klasycznej metody Monte Carlo. Chcąc uzyskać dobre oszacowanie wartości P_f , np. odpowiadające współczynnikowi $\nu_{\hat{P}_f} = 0.1$, oraz spodziewając się wartości prawdopodobieństwa awarii w przedziale $10^{-6} \div 10^{-3}$ należy obliczyć wartość funkcji granicznej dla $N = 10^5 \div 10^8$ realizacji zmiennych losowych. Czas takiej analizy, szczególnie wtedy gdy funkcja graniczna nie jest dana w sposób jawny i wymaga przeprowadzenia długotrwałych obliczeń MES, jest w większości przypadków trudny do zaakceptowania.

Chociaż olbrzymi koszt obliczeniowy związany z klasyczną metodą Monte Carlo wyklucza korzystanie z tej metody w analizie niezawodności złożonych konstrukcji, które modelowane są za pomocą tysięcy elementów skończonych, to w przypadku niewielkich zadań stanowić może ona alternatywę dla innych metod. Niezaprzeczalnymi zaletami klasycznej metody Monte Carlo jest jej intuicyjny charakter, łatwość implementacji, niewrażliwość na kształt obszaru awarii oraz możliwość przeprowadzenia symulacji w przestrzeni oryginalnej, bez konieczności transformacji do przestrzeni \mathcal{U} .

A.4.2. Metoda importance sampling

Metodą, która pozwala na znaczną redukcję wariancji estymatora prawdopodobieństwa awarii, a co za tym idzie na ograniczenie liczby niezbędnych symulacji losowych N , jest importance sampling. Istnieje wiele odmian tej metody. Przegląd tych, które znalazły największe zastosowanie w analizie niezawodności konstrukcji można znaleźć w pracy [172]. Idea metody importance sampling

opiera się na spostrzeżeniu, iż wartość oczekiwaną (A.58) można w sposób równoważny zapisać jako

$$\begin{aligned} \mathbb{E}_{\mathbf{X}}[I_{\mathcal{D}_f}(\mathbf{X})] &= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_n I_{\mathcal{D}_f}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_n I_{\mathcal{D}_f}(\mathbf{v}) \frac{f_{\mathbf{X}}(\mathbf{v})}{s_{\mathbf{V}}(\mathbf{v})} s_{\mathbf{V}}(\mathbf{v}) d\mathbf{v} = \mathbb{E}_{\mathbf{V}} \left[I_{\mathcal{D}_f}(\mathbf{V}) \frac{f_{\mathbf{X}}(\mathbf{V})}{s_{\mathbf{V}}(\mathbf{V})} \right] \end{aligned} \quad (\text{A.65})$$

gdzie $s_{\mathbf{V}}(\cdot)$ jest łączną gęstością prawdopodobieństwa n elementowego wektora losowego \mathbf{V} , a $\mathbb{E}_{\mathbf{X}}[\cdot]$ i $\mathbb{E}_{\mathbf{V}}[\cdot]$ oznaczają operację wartości oczekiwanej odpowiednio względem funkcji $f_{\mathbf{X}}(\cdot)$ i $s_{\mathbf{V}}(\cdot)$. Ponieważ wartość oczekiwana funkcji charakterystycznej równa jest prawdopodobieństwu awarii (zob. (A.58)), dlatego można zapisać

$$P_f = \mathbb{E}_{\mathbf{X}}[I_{\mathcal{D}_f}(\mathbf{X})] = \mathbb{E}_{\mathbf{V}} \left[I_{\mathcal{D}_f}(\mathbf{V}) \frac{f_{\mathbf{X}}(\mathbf{V})}{s_{\mathbf{V}}(\mathbf{V})} \right], \quad (\text{A.66})$$

lub po transformacji zmiennych podstawowych \mathbf{X} na standardowe zmienne normalne \mathbf{U}

$$P_f = \mathbb{E}_{\mathbf{U}}[I_{\Delta_f}(\mathbf{U})] = \mathbb{E}_{\mathbf{V}} \left[I_{\Delta_f}(\mathbf{V}) \frac{\varphi_n(\mathbf{V}, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}(\mathbf{V})} \right], \quad (\text{A.67})$$

gdzie I_{Δ_f} jest funkcją charakterystyczną obszaru awarii Δ_f w przestrzeni \mathbf{U} , a $\varphi_n(\mathbf{v}, \mathbf{0}, \mathbf{I})$ jest n -wymiarową łączną gęstością prawdopodobieństwa zmiennych normalnych o zerowych wartościach średnich i jednostkowej macierzy kowariancji. W takim przypadku estymator prawdopodobieństwa awarii \hat{P}_f przyjmuje postać (por. (A.60))

$$\hat{P}_f = \frac{1}{N} \sum_{k=1}^N I_{\Delta_f}(\mathbf{V}_k) \frac{\varphi_n(\mathbf{V}_k, \mathbf{0}, \mathbf{I})}{s_{\mathbf{V}}(\mathbf{V}_k)}. \quad (\text{A.68})$$

Wartość estymatora oblicza się podstawiając do powyższego wzoru realizacje \mathbf{v}_k wektora losowego \mathbf{V}_k wygenerowane zgodnie z funkcją gęstości $s_{\mathbf{V}}(\cdot)$, która nazywana jest gęstością losującą. Odpowiedni wybór gęstości losującej decyduje o efektywności metody importance sampling. Funkcja ta powinna minimalizować wariancję estymatora (A.68). Można pokazać (zob. np. [173] strona 74), że jeżeli $s_{\mathbf{V}}(\cdot)$ będzie funkcją proporcjonalną do $\varphi_n(\mathbf{v}, \mathbf{0}, \mathbf{I})$ w obszarze całkowania (obszarze awarii), to wariancja \hat{P}_f redukuje się do zera. Warunek ten zapiszemy

jako

$$s_{\mathbf{v}}(\mathbf{v}) = \frac{I_{\Delta_f}(\mathbf{v}) \varphi_n(\mathbf{v}, \mathbf{0}, \mathbf{I})}{P_f}, \quad (\text{A.69})$$

lub w sposób równoważny

$$\int_{\Delta_f} s_{\mathbf{v}}(\mathbf{v}) d\mathbf{v} = 1, \quad \text{ i } \quad s_{\mathbf{v}}(\mathbf{v}) = \zeta \varphi_n(\mathbf{v}, \mathbf{0}, \mathbf{I}) I_{\Delta_f}(\mathbf{v}) \quad \text{ prawie wszędzie w } \Delta_f, \quad (\text{A.70})$$

gdzie ζ jest stałą. W przypadku optymalnej gęstości losującej wystarczy tylko jedna realizacja (w obszarze Δ_f) aby otrzymać dokładną wartość prawdopodobieństwa awarii. Niestety, w celu znalezienia takiej gęstości niezbędna jest uprzednia znajomość wartości P_f . Pomimo pozornej niepraktyczności, kryterium (A.69) stanowi cenną wskazówkę przy konstruowaniu gęstości losujących bliskich optymalnej.

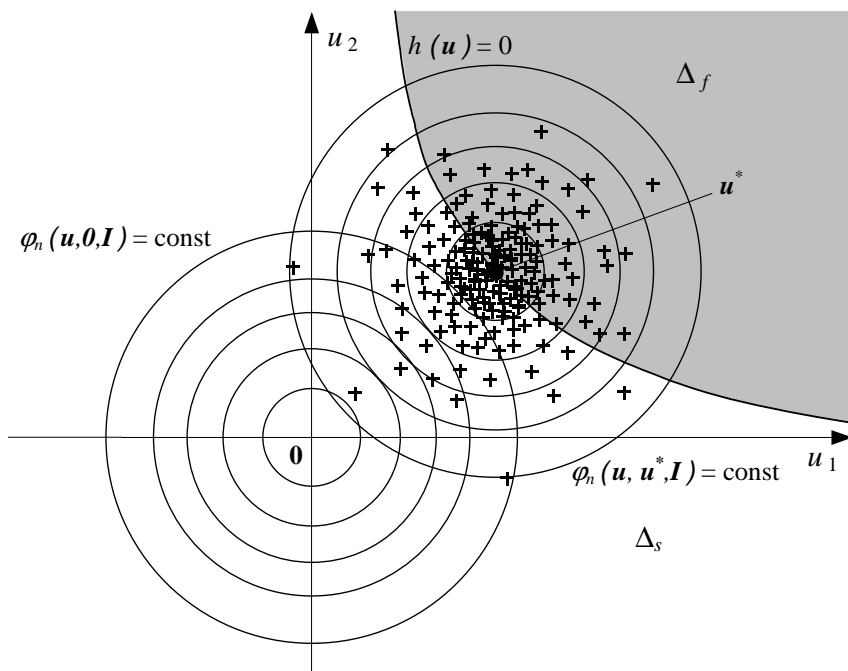
Często zdecydowaną poprawę efektywności obliczeniowej osiąga się wybierając jako gęstość $s_{\mathbf{v}}(\mathbf{v})$ n -wymiarową gaussowską funkcję gęstości o wartościach oczekiwanych w punkcie projektowym \mathbf{u}^* , a więc skupioną w obszarze, który ma największy wpływ na wartość P_f (zob. rys. A.2)

$$s_{\mathbf{v}}(\mathbf{v}) = \varphi_n(\mathbf{v}, \mathbf{u}^*, \mathbf{I}) = \prod_{i=1}^n \varphi(v_i - u_i^*). \quad (\text{A.71})$$

W takim przypadku estymator prawdopodobieństwa awarii przybiera postać

$$\hat{P}_f = \frac{1}{N} \sum_{k=1}^N I_{\Delta_f}(\mathbf{V}_k) \frac{\varphi_n(\mathbf{V}_k, \mathbf{0}, \mathbf{I})}{\varphi_n(\mathbf{V}_k, \mathbf{u}^*, \mathbf{I})}. \quad (\text{A.72})$$

Wybierając gaussowską gęstość losującą (A.71) oszacowanie wartości P_f nie jest bardzo wrażliwe na kształt obszaru Δ_f i – za wyjątkiem przypadków silnie nieliniowych powierzchni granicznych – prawdopodobieństwo, że realizacja zmiennej losowej wygenerowanej zgodnie z gęstością $s_{\mathbf{v}}(\cdot)$ znajdzie się w obszarze awarii wynosi około 50%. Kontrastuje to z klasyczną metodą Monte Carlo gdzie prawdopodobieństwo “trafienia” realizacji w obszar awarii było mniej więcej równe obliczanemu prawdopodobieństwu awarii. Liczne doświadczenia w stosowaniu metody importance sampling (w wersji (A.72)) pozwalają stwierdzić, że w większości przypadków zadowalającą estymację P_f można otrzymać już po wygenerowaniu od kilkuset do kilku tysięcy realizacji zmiennych. Jest to o wiele rzędów wielkości mniej niż w przypadku klasycznego Monte Carlo.



Rys. A.2. Metoda importance sampling. Symulacje losowe wokół punktu projektowego \mathbf{u}^* w przestrzeni standardowej.

Mimo niewątpliwych zalet metody, istnieją pewne ważne ograniczenia, które mogą czasem utrudnić analizę niezawodności przy pomocy zaprezentowanej techniki importance sampling (wyliczając za [172] te, które dotyczą symulacji w przestrzeni \mathbf{u}).

1. Gęstość losująca $s_{\mathbf{V}}(\cdot)$ może nie być dobrze dobrana.
2. Powierzchnia graniczna może posiadać wiele punktów projektowych (minimów lokalnych funkcji $\|\mathbf{u}\|$ pod warunkiem, że $h(\mathbf{u}) = 0$).
3. Jeśli należy jednocześnie rozpatrywać wiele funkcji granicznych, wtedy również mamy do czynienia z wieloma punktami projektowymi.
4. Funkcje graniczne mogą charakteryzować się znaczną ujemną krzywizną w punkcie projektowym.

Sposobem przewyciężenia wymienionych trudności wydają się być adaptacyjne metody importance sampling, których najważniejsze elementy omówione zostały w następnym podrozdziale.

A.4.3. Metody adaptacyjne

Zgodnie z warunkiem (A.69), metoda importance sampling jest tym bardziej efektywna im bardziej gęstość losująca $s_{\mathbf{V}}(\cdot)$ zbliżona jest do oryginalnej funkcji łącznej gęstości prawdopodobieństwa zmiennych losowych w obszarze awarii \mathcal{D}_f (lub Δ_f jeśli próbki generowane są w przestrzeni \mathcal{U}). Ponieważ nie istnieje żadna absolutnie skuteczna metoda doboru takiej funkcji, dlatego zaproponowano szereg metod dopasowania funkcji $s_{\mathbf{V}}(\cdot)$ do hipotetycznej gęstości idealnej.

Jedną z propozycji jest założenie pewniej gęstości losującej, wygenerowanie początkowej próby, a następnie obliczenie wartości średnich oraz macierzy kowariancji na podstawie tych realizacji, które znalazły się w obszarze awarii. W pracy [19] jako funkcję losującą przyjęto wielowymiarowy rozkład normalny zdefiniowany przez tak oszacowane momenty statystyczne. Poprzez generowanie kolejnych realizacji zmiennych losowych funkcja $s_{\mathbf{V}}(\cdot)$ jest ciągle udoskonalana, gdyż parametry rozkładu estymuje się na podstawie coraz liczniejszej próbki. Przyjmuje się, że gęstość losująca o tak uaktualnianym położeniu oraz kształcie zbiega ostatecznie do gęstości idealnej. Niestety, zbieżności tej nie można zagwarantować, szczególnie w przypadku złego początkowego wyboru gęstości losującej.

Innym pomysłem jest przyjęcie funkcji $s_{\mathbf{V}}(\cdot)$ jako sumy K składowych gęstości $s_{\mathbf{V}}^{(i)}(\mathbf{v})$

$$s_{\mathbf{V}}(\mathbf{v}) = \sum_{i=1}^K w^{(i)} s_{\mathbf{V}}^{(i)}(\mathbf{v}), \quad (\text{A.73})$$

gdzie $w^{(i)}$, $i = 1, \dots, K$, są wagami, które dobiera się tak aby $s_{\mathbf{V}}(\cdot)$ dopasować jak najlepiej do $f_{\mathbf{X}}(\cdot)$ w obszarze \mathcal{D}_f (lub do $\varphi_n(\cdot, \mathbf{0}, \mathbf{I})$ w Δ_f).

Najczęściej każda z elementarnych gęstości $s_{\mathbf{V}}^{(i)}(\cdot)$ we wzorze (A.73) uaktualniana jest wraz z postępem procesu symulacji losowych. W artykule [116] Karamchandani i in. przyjęli gęstości $s_{\mathbf{V}}^{(i)}(\cdot)$, $i = 1, \dots, K$, jako równe oryginalnej łącznej gęstości prawdopodobieństwa zmiennych losowych, których wartości oczekiwane przesunięto do wybranych punktów $\hat{\mathbf{v}}^{(i)}$, zwanych punktami reprezentatywnymi. Odpowiednie współczynniki wagowe zdefiniowane są wtedy jako (w przypadku symulacji w przestrzeni standardowej)

$$w^{(i)} = \frac{\varphi_n(\hat{\mathbf{v}}^{(i)}, \mathbf{0}, \mathbf{I})}{\sum_{j=1}^K \varphi_n(\hat{\mathbf{v}}^{(j)}, \mathbf{0}, \mathbf{I})}. \quad (\text{A.74})$$

Oznacza to, że wartości wag są proporcjonalne do “wkładu” jaki dany punkt $\hat{\mathbf{v}}^{(i)}$ wnosi do oszacowania wartości prawdopodobieństwa awarii. Trzeba tu również podkreślić, iż efektywność dochodzenia do idealnej gęstości losującej zależy w znacznym stopniu od wyboru K początkowych punktów reprezentatywnych oraz rozkładu prawdopodobieństwa, z którego są generowane. W podrozdziale 5.1 omówiono pewną wersję powyższego adaptacyjnego algorytmu zaproponowaną w pracy [278]. Po wprowadzeniu pewnych modyfikacji algorytm ten wykorzystano w podrozdziale 6.1, poświęconym analizie niezawodności w zagadnieniach dotyczących wytrzymałości zderzeniowej, do obliczenia prawdopodobieństwa awarii dynamicznie zgniatanej belki cienkościennej.

W celu obliczenia wartości estymatora (A.72) niezbędna jest znajomość położenia punktu projektowego. Niestety, bardzo często specyfika funkcji granicznej uniemożliwia wykorzystanie efektywnych gradientowych algorytmów takich jak iHLRF lub ARF w celu szybkiego i precyzyjnego znalezienia punktu \mathbf{u}^* . Jednym ze sposobów przewyciężenia tego problemu jest aproksymacja oryginalnej funkcji granicznej przy pomocy odpowiednio dobranej powierzchni odpowiedzi.

Odmianą metodą jest poszukiwanie punktu projektowego w trakcie samych symulacji losowych oraz uaktualnianie funkcji gęstości losującej za każdym razem gdy znajdowane jest jego lepsze oszacowanie. Zazwyczaj, w przestrzeni \mathbf{u} jako wyjściową gęstość losującą wybiera się funkcję łącznej gęstości standardowych zmiennych normalnych (A.7) lub funkcję typu (A.71), gdzie punkt \mathbf{u}^* specyfikowany jest na podstawie pewnej wiedzy na temat analizowanego zagadnienia. Pomijając szczegóły dotyczące pierwszych kroków algorytmu i przyjmując, że aktualnie gęstość losująca “skoncentrowana” jest nad punktem w obszarze awarii, zasada zmiany położenia gęstości losującej sprowadza się do sprawdzenia, czy wygenerowana realizacja wektora losowego prowadzi do ujemnej wartości funkcji granicznej i czy norma tego wektora $\|\mathbf{u}\|$ jest mniejsza od normy wektora przyjmowanego dotychczas jako wektor wartości średnich. Jeśli tak jest, to punkt ten przyjmuje się jako nowy wektor wartości średnich w definicji gęstości losującej, a estymator prawdopodobieństwa awarii przybiera postać

$$\begin{aligned} \hat{P}_f = \frac{1}{N} & \left[\sum_{i=1}^{N_1} I_{\Delta_f}(\mathbf{V}_i^{(1)}) \frac{\varphi_n(\mathbf{V}_i^{(1)}, \mathbf{0}, \mathbf{I})}{\varphi_n(\mathbf{V}_i^{(1)}, \mathbf{u}^{(1)}, \mathbf{I})} + \sum_{i=1}^{N_2} I_{\Delta_f}(\mathbf{V}_i^{(2)}) \frac{\varphi_n(\mathbf{V}_i^{(2)}, \mathbf{0}, \mathbf{I})}{\varphi_n(\mathbf{V}_i^{(2)}, \mathbf{u}^{(2)}, \mathbf{I})} + \right. \\ & \left. + \dots + \sum_{i=1}^{N_m} I_{\Delta_f}(\mathbf{V}_i^{(m)}) \frac{\varphi_n(\mathbf{V}_i^{(m)}, \mathbf{0}, \mathbf{I})}{\varphi_n(\mathbf{V}_i^{(m)}, \mathbf{u}^{(m)}, \mathbf{I})} \right], \end{aligned} \quad (\text{A.75})$$

gdzie $N = N_1 + N_2 + \dots + N_m$ jest całkowitą liczbą wygenerowanych realizacji, m jest liczbą zmian funkcji gęstości losującej, $\varphi_n(\mathbf{v}^{(1)}, \mathbf{u}^{(1)}, \mathbf{I})$ jest początkową

funkcją gęstości losującej (umieszczoną nad punktem $\mathbf{u}^{(1)}$), zgodnie z którą wygenerowano N_1 realizacji, $\varphi_n(\mathbf{v}^{(2)}, \mathbf{u}^{(2)}, \mathbf{I})$ jest zmodyfikowaną gęstością losującą służącą do wygenerowania N_2 realizacji itd.

Nie ulega wątpliwości, że adaptacyjne techniki importance sampling, w których punkt projektowy lokalizowany jest zgodnie z przedstawioną powyżej procedurą, są znacznie mniej efektywne od metod gradientowych w przypadku różniczkowalnych funkcji granicznych. Stanowią jednak interesującą alternatywę, gdy warunek ten nie jest spełniony.

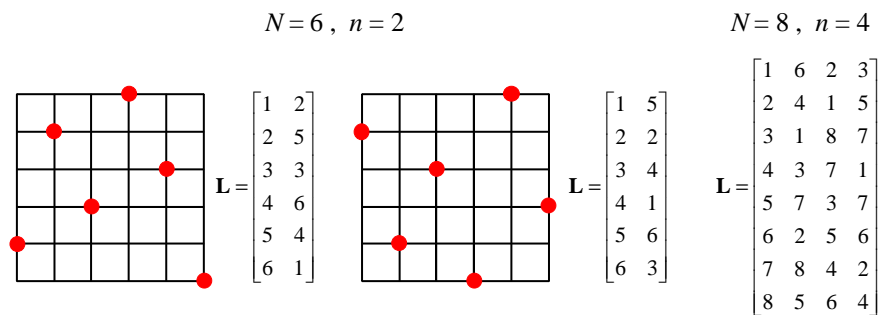
B

Algorytmy tworzenia optymalnych łańciskich hiperkostek

Plan eksperymentów zwany łańciską hiperkostką (ang. *Latin hypercube* - LH) lub też n -wymiarowym kwadratem łańciskim definiuje się za pomocą macierzy \mathbf{L} o N wierszach i n kolumnach, gdzie każda z kolumn jest permutacją liczb całkowitych od 1 do N (N to ilość punktów). Wiersze macierzy \mathbf{L} zawierają dyskretne współrzędne poszczególnych punktów. Możemy zatem zapisać

$$\mathbf{L} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{N1} & \cdots & x_{Nn} \end{bmatrix}, \quad (\text{B.1})$$

gdzie \mathbf{x}_j jest j -tym punktem próbki/planu eksperymentów. Charakterystyczną cechą LH jest to, że dowolnie wybrane dwa punkty nie mają żadnej współrzędnej takiej samej. Na rys. B.1 pokazano przykładowe macierze \mathbf{L} oraz odpowiadające im rozkłady punktów kwadratów łańciskich.



Rys. B.1. Przykładowe macierze oraz układy punktów 2 i 4 wymiarowych kwadratów/hiperkostek łańciskich.

Szczególny przypadek LH stanowią wielowymiarowe optymalne kwadraty łańciskie, czy też optymalne łańciskie hiperkostki (ang. *optimal Latin hypercube* - OLH). Są one niezwykle ważnym elementem omawianych w niniejszej pracy

metod analizy niezawodności oraz optymalizacji odpornościowej. W rozdziale 4 pokazano, że estymacja momentów statystycznych funkcji losowych za pomocą metod symulacyjnych wykorzystujących OLH jest dużo bardziej efektywna w porównaniu do klasycznych, “czysto losowych” metod symulacyjnych. Plan OLH posiada również właściwość równomiernego wypełniania analizowanego obszaru, co jest niesłychanie istotne przy konstruowaniu powierzchni odpowiedzi, jak również w metodach losowego przeszukiwania.

Niestety, w przypadku dużej liczby zmiennych oraz punktów wchodzących w skład planowanego eksperymentu numerycznego, znalezienie takiej konfiguracji łańciskiej hiperkostki, która jest optymalna (w rozumieniu pewnego kryterium) nie jest łatwe i wymaga użycia wyspecjalizowanych algorytmów. Omówione zostaną dwie metody budowania OLH.

Poniżej, w podrozdziale B.2, przedstawiona została tzw. metoda CP służąca do konstruowania optymalnych hiperkostek, a w podrozdziale B.3 omówiony jest wyspecjalizowany algorytm genetyczny, który w przypadku dużych wartości N i n okazuje się być bardziej efektywny od CP. Prezentacja algorytmów poprzedzona jest przeglądem kryteriów optymalności oraz sformułowaniem problemu optymalizacji. W podrozdziale B.4 omówione są stosowane kryteria zbieżności algorytmów.

Dużą zaletą planów OLH jest ich uniwersalność. Raz wygenerowana optymalna n -wymiarowa hiperkostka o N punktach może być zapisana i używana wielokrotnie do różnych zadań, w których występuje n -zmiennych losowych.

Ostatecznym celem przedstawianych algorytmów jest otrzymanie współrzędnych N punktów w przestrzeni \mathbb{R}^n (N realizacji n -wymiarowego wektora losowego). Wygenerowana próbka musi oczywiście uwzględniać rozkłady prawdopodobieństwa zmiennych losowych. W podrozdziale B.7 wyjaśniono w jaki sposób wykorzystać hiperkostki łańciskie do generowania próbek ze zmiennych losowych o rozkładach ciągłych.

Prezentowany w niniejszym dodatku materiał opiera się w znacznym stopniu na rozważaniach zamieszczonych w pracy [150].

B.1. Kryteria optymalizacji

W celu zdefiniowania kryteriów, względem których optymalizowany będzie układ punktów łańciskiej hiperkostki, wprowadźmy trzy funkcje określone na zbiorze dopuszczalnych konfiguracji hiperkostek o wartościach w \mathbb{R} .

Funkcja d zdefiniowana jest jako kwadrat najmniejszej (euklidesowej) odległości pomiędzy dwoma punktami planu eksperymentów

$$d(\mathbf{L}) := \min_{1 \leq i, j \leq N, i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|^2. \quad (\text{B.2})$$

Przyjmijmy następnie, że n_d jest liczbą wystąpień minimalnej odległości (B.2). Dla danej realizacji łańciskiej hiperkostki będzie to $n_d(\mathbf{L})$.

Ostatnia funkcja, G , dana jest przez analogię do sumy wzajemnych sił działających na zbiór elektrycznie naładowanych cząstek (zob. [7] i [11])

$$G(\mathbf{L}) := \sum_{i=1}^N \sum_{j=i+1}^N \frac{1}{\|\mathbf{x}_i - \mathbf{x}_j\|^2}. \quad (\text{B.3})$$

Funkcje te posłużą teraz do zaproponowania dwóch kryteriów optymalizacji łańciskich hiperkostek.

Kryterium najmniejszej odległości:

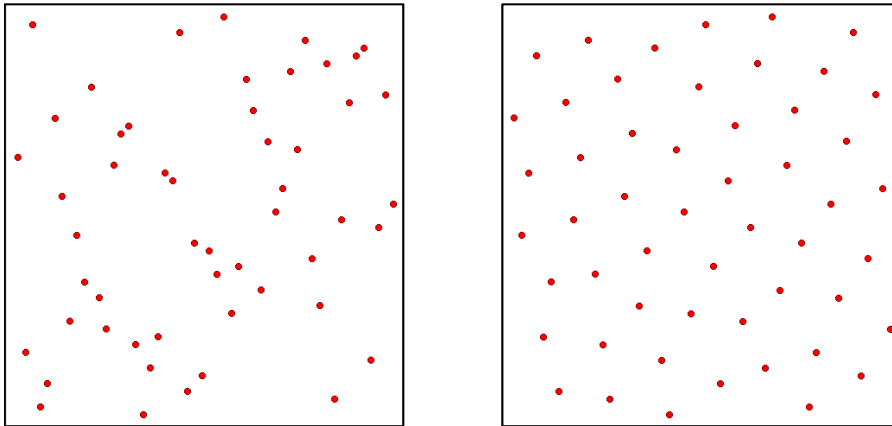
Przyjmuje się, że hiperkostka łańciska \mathbf{L}_1 jest lepsza od \mathbf{L}_2 jeśli $d(\mathbf{L}_1) > d(\mathbf{L}_2)$. W przypadku gdy $d(\mathbf{L}_1) = d(\mathbf{L}_2)$, \mathbf{L}_1 jest lepsza od \mathbf{L}_2 jeśli $n_d(\mathbf{L}_1) < n_d(\mathbf{L}_2)$.

Kryterium sił:

\mathbf{L}_1 jest lepsza od \mathbf{L}_2 jeśli $G(\mathbf{L}_1) < G(\mathbf{L}_2)$.

Należy zauważyć, że sprawdzenie warunku równości występujące w pierwszym kryterium implementuje się również jako sprawdzenie równości nawet w ramach arytmetyki zmiennoprzecinkowej. Wynika to z tego, iż porównywane są kwadraty odległości punktów o współrzędnych całkowitoliczbowych, a więc liczby całkowite.

Zmieniając konfigurację punktów w łańciskich hiperkostkach tak, by polepszyć wartość danego kryterium (maksymalizować pierwsze lub minimalizować drugie) otrzymuje się w wyniku taki plan eksperymentów, który wypełnia równomiernie objętość n -wymiarowej hiperkostki (ang. *space filling design*). Unika się w ten sposób koncentracji punktów w niektórych obszarach (ang. *clustering*), a w konsekwencji, pozostawienia innych obszarów niezbadanymi. Na rys. B.2 pokazano przykładowe konfiguracje kwadratu łańciskiego przed i po optymalizacji.



Rys. B.2. Kwadrat łaciński ($N = 50, n = 2$). Po lewej układ losowy, po prawej optymalny.

B.2. Algorytm CP

Metoda CP została zaproponowana przez Parka w [193], a następnie zmodyfikowana w pracy [264]. Skrót CP pochodzi od pierwszych liter angielskiej nazwy “columnwise-pairwise”, co można przetłumaczyć jako “kolumnami i parami”. Nazwa ta wynika wprost z dokonywanych w algorytmie przekształceń macierzy \mathbf{L} .

Na potrzeby prezentacji okreśmy mianem *prostej zamiany* taką modyfikację kolumny macierzy \mathbf{L} , która polega na zamianie miejscami dwóch elementów tej kolumny. Poniżej, w postaci pseudo-kodu, przedstawiona jest operacja tzw. przeczesywania macierzy \mathbf{L} (ang. *CP-sweep*)

```

for  $i = 1 \dots n$ 
  for  $j = 1 \dots$  liczba możliwych zamian
    Dokonaj prostej zamiany elementów pary  $j$ .
    Oblicz wartość kryterium.
    Powróć do stanu sprzed zamiany.
  end of for-loop
  Zamień kolumnę  $i$  na tę, która odpowiada najlepszej prostej zamianie.
end of for-loop

```


Oczywiście, przez “lepszą” bądź też “najlepszą” prostą zamianę rozumie się taką zamianę, która prowadzi do odpowiedniej poprawy wartości wybranego kryterium optymalizacji. Przechwywanie jest więc strategią systematycznego sprawdzania wszystkich kolumn macierzy \mathbf{L} pod kątem znalezienia najlepszej możliwej prostej zamiany.

W pewnym uproszczeniu, algorytm CP polega na wielokrotnym wykonaniu operacji przechwywania aż do spełnienia warunku zbieżności (zostanie on omówiony w punkcie B.4). Wynikiem działania algorytmu jest macierz \mathbf{L} o optymalnej konfiguracji, determinująca plan eksperymentów OLH.

Czasem dla odróżnienia od OLH oraz dla podkreślenia, że mamy do czynienia z łańciską hiperkostką o losowym układzie punktów, używać będziemy skrótu RLH, od angielskiej nazwy *random Latin hypercube*. Zbudowanie RLH jest zadaniem bardzo prostym w implementacji i polega na wygenerowaniu n losowych permutacji liczb 1 do N , bez konieczności zmiany konfiguracji punktów.

B.3. Algorytm genetyczny

Algorytmy genetyczne są obecnie bardzo powszechnie stosowanymi metodami optymalizacji globalnej. Znajdują one zastosowanie w szerokiej klasie zagadnień nauki i techniki. Najważniejsze operacje wchodzące w skład algorytmu genetycznego wyszczególnione są w poniższym pseudo-kodzie:

Losowo wygeneruj *populację początkową*.

Dokonaj *oceny* osobników *populacji początkowej*.

while (nie jest spełniony warunek zbieżności)

 Dokonaj *selekcji*. Zachowaj jedynie najlepiej przystosowane osobniki.

 Na zachowanych osobnikach przeprowadź operację *krzyżowania*.

 Dokonaj losowych *mutacji* osobników.

 Dokonaj *oceny* nowej populacji.

 Sprawdź warunek zbieżności.

end of while-loop

W przypadku zadania optymalizacji łańciskich hiperkostek istnieją różne wersje algorytmu w zależności od sposobu zdefiniowania operacji selekcji, krzyżowania, mutacji oraz wyboru populacji początkowej. Na przykład, w pracy [11]

użyty został standardowy algorytm genetyczny ogólnego przeznaczenia do znajdowania OLH według kryterium (B.3). To podejście wymaga jednak specjalnego kodowania zmiennych projektowych, którymi są współrzędne punktów, oraz wprowadzenia do sformułowania funkcji celu specjalnych członów wymuszających spełnienie przez dany układ punktów założeń łacińskiej hiperkostki (żadna współrzędna punktu nie powtarza się).

W odróżnieniu od tego podejścia, zaimplementowany przez autora algorytm genetyczny został zaprojektowany specjalnie pod kątem specyficznego problemu tworzenia OLH. W zainspirowanej pracą Simpsona [219] metodzie wszystkie operacje genetyczne zdefiniowane są bezpośrednio na macierzach \mathbf{L} , zob. (B.1), w taki sposób aby zachować ich charakterystyczną budowę. Wymienione powyżej podstawowe elementy algorytmu genetycznego określone są następująco:

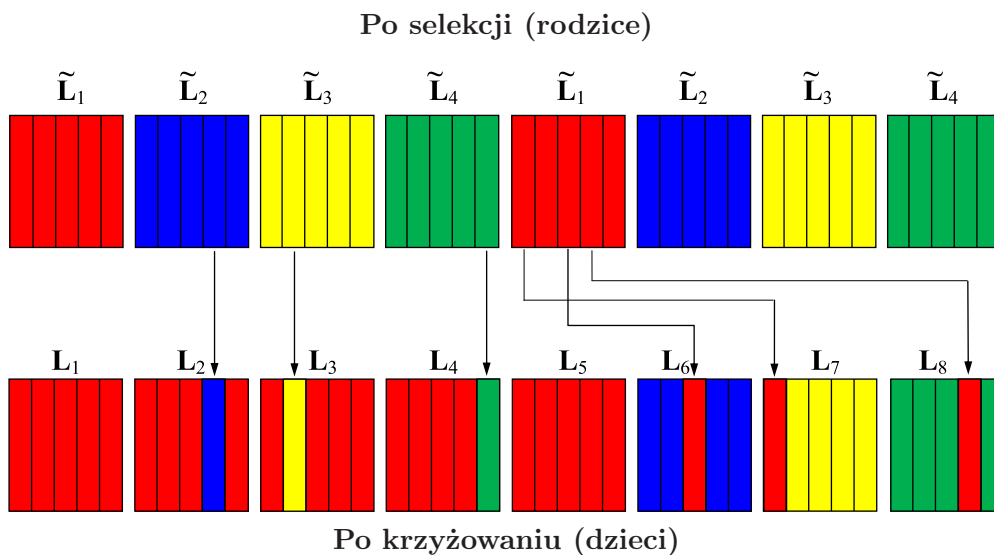
Populacja początkowa. Działanie algorytmu rozpoczyna się od losowego wygenerowania N_{pop} macierzy \mathbf{L} reprezentujących początkową populację hiperkostek łacińskich. Ze względu na zdefiniowaną dalej operację selekcji, liczba N_{pop} musi być parzysta.

Selekcja. Uwzględniając jedno z kryteriów optymalizacji opisanych w B.1, zatrzymuje się $N_{\text{pop}}/2$ najlepszych hiperkostek. Pozostałe są pomijane. Hiperkostki indeksowane są od najlepszej do najgorszej. Oznaczmy je jako $\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2, \dots, \tilde{\mathbf{L}}_{N_{\text{pop}}/2}$.

Krzyżowanie. W trakcie tej operacji, przedstawionej schematycznie na rys. B.3, z $N_{\text{pop}}/2$ hiperkostek wybranych na etapie selekcji (“rodziców”) generowanych jest N_{pop} nowych hiperkostek (“dzieci”).

Macierz $\tilde{\mathbf{L}}_1$ odpowiadająca najlepszej łacińskiej hiperkostce umieszczana jest na początku nowej populacji. Oznaczamy ją jako \mathbf{L}_1 . Ta sama macierz umieszczana jest również na miejscu $N_{\text{pop}}/2 + 1$. W wyniku krzyżowania najlepszej łacińskiej hiperkostki z k -tą, $k \in [2, N_{\text{pop}}/2]$, powstają dwie nowe hiperkostki. Pierwszą hiperkostkę (w nowej populacji o indeksie k) tworzy się przez zastąpienie wybranej losowo kolumny macierzy $\tilde{\mathbf{L}}_1$ odpowiednią kolumną macierzy $\tilde{\mathbf{L}}_k$. Macierz drugiej hiperkostki, o indeksie $N_{\text{pop}}/2 + k$, powstaje przez zastąpienie losowo wybranej kolumny $\tilde{\mathbf{L}}_k$ przez odpowiednią kolumnę z $\tilde{\mathbf{L}}_1$.

Mutacja. Ta niesłychanie istotna operacja genetyczna wykonywana jest na wszystkich hiperkostkach z populacji za wyjątkiem najlepszej, o macierzy \mathbf{L}_1 . Mutację przeprowadza się na każdej kolumnie macierzy $\mathbf{L}_2, \dots, \mathbf{L}_{N_{\text{pop}}}$. Polega ona na zamianie miejscami dwóch losowo wybranych wyrazów kolumny, jeśli wygenerowana z rozkładu jednostajnego $[0, 1]$ liczba ma wartość mniejszą niż założona wartość progowa p_{mut} .



Rys. B.3. Ilustracja operacji krzyżowania algorytmu genetycznego w przypadku $N_{\text{pop}} = 8$. Należy zauważyć, że po krzyżowaniu L_1 i L_5 są takie same jak $L_{\tilde{1}}$ - macierz najlepszej łącińskiej hiperkostki przed krzyżowaniem. Na etapie mutacji L_1 nie jest jednak modyfikowana, co sprawia, że jedna kopia najlepszej hiperkostki zostaje zawsze zachowana.

Występujący w pętli **while** pseudo-kodu warunek zbieżności algorytmu genetycznego przedstawiony jest w podrozdziale B.4.

B.4. Warunki zbieżności

Warunki zbieżności zaprezentowanych algorytmów omówione zostaną na przykładzie kryterium sił wykorzystującego funkcję G daną równaniem (B.3). Zarówno w przypadku metody CP, jak też algorytmu genetycznego, warunek zbieżności algorytmu sprowadza się do porównania aktualnej zmiany wartości kryterium do zmiany początkowej. Ze względu na specyfikę obydwu metod istnieją istotne różnice w implementacji tego warunku i dlatego wymaga on osobnego omówienia dla każdego z algorytmów.

Algorytm CP

Oznaczmy przez ΔG_1 zmianę wartości funkcji G po pierwszej iteracji, tzn. po pierwszej operacji przeczesywania. Warunek zbieżności, który sprawdzany jest po iteracji k ma następującą postać:

$$\Delta G_k < \epsilon \Delta G_1. \quad (\text{B.4})$$

gdzie ΔG_k jest zmianą wartości funkcji G na k -tej iteracji, a ϵ jest parametrem. W przykładach numerycznych przedstawionych w dalszej części tekstu przyjęto $\epsilon = 10^{-6}$. Spełnienie nierówności (B.4) powoduje zakończenie działania algorytmu. Oczywiście, proces optymalizacji trwa tym dłużej, im większe są wartości N i n oraz im mniejsza jest wartość ϵ . Na przykład, utworzenie optymalnej łacińskiej hiperkostki w przypadku $N = 30$ i $n = 6$ wymagało 31 iteracji. Wartość funkcji kryterium odpowiadająca OLH wynosiła $G(\mathbf{L}) = 0.5331$.

Algorytm genetyczny

W odróżnieniu od operacji przeczyszczenia z metody CP, wygenerowanie nowej populacji macierzy \mathbf{L} w zaproponowanym algorytmie genetycznym jest zadaniem stosunkowo tanim i nie wymagającym długiego czasu pracy procesora. Z drugiej strony, prawdopodobieństwo, że następna populacja nie przyniesie poprawy kryterium optymalizacji nie jest pomijalne, a co za tym idzie zamiast rozpatrywać zmianę wartości funkcji G w czasie “życia” jednej populacji należy porównywać łączne zmiany generowane przez p populacji (zazwyczaj zakłada się się $p = 50$ lub 100). Jako wartość odniesienia przyjmuje się więc

$$\Delta \tilde{G}_p = G(\mathbf{L}_p) - G(\mathbf{L}_0), \quad (\text{B.5})$$

gdzie \mathbf{L}_p odpowiada najlepszej hiperkostce z p -tej populacji, a \mathbf{L}_0 odpowiada najlepszej hiperkostce w losowo wybranej populacji początkowej. Jeśli numer populacji k jest wielokrotnością p , sprawdzany jest następujący warunek zbieżności:

$$G(\mathbf{L}_k) - G(\mathbf{L}_{k-p}) < \epsilon \Delta \tilde{G}_p. \quad (\text{B.6})$$

W obliczeniach przyjęto $\epsilon = 10^{-7}$. Stosując algorytm genetyczny do znalezienia konfiguracji OLH w przypadku $N = 30$ i $n = 6$, zbieżność rozwiązania została osiągnięta po około 3300 pokoleniach, dając w wyniku $G(\mathbf{L}) = 0.5326$.

B.5. Analiza złożoności obliczeniowej algorytmu CP

Dokładna analiza złożoności obliczeniowej zarówno algorytmu CP jak i algorytmu genetycznego wydaje się być zadaniem bardzo trudnym, a wręcz niemożliwym. Jednakże, jak to zostanie pokazane poniżej, w przypadku metody CP jesteśmy w stanie podać oszacowanie czasu rozwiązania, które powinno okazać się przydatnym z punktu widzenia praktyki.

Poprzez złożoność obliczeniową rozumiemy asymptotyczny wzrost czasu działania algorytmu wraz ze wzrostem liczby punktów N oraz liczby zmiennych n .

W przypadku funkcji $T(N)$ zapis $T = \mathcal{O}(N^q)$ oznacza, że istnieją takie stałe \tilde{c}_1 i \tilde{c}_2 , dla których

$$\tilde{c}_1 N^q < T < \tilde{c}_2 N^q, \quad (\text{B.7})$$

jeżeli N jest dostatecznie dużą liczbą. Zależność tą wyrazić można również jako $T \sim N^q$.

W celu oszacowania złożoności obliczeniowej operacji przeczesywania algorytmu CP (zob. pseudo kod na stronie 288) należy pamiętać, że:

- Zewnętrzna pętla przebiega n kolumn macierzy \mathbf{L} .
- W każdej kolumnie dokonuje się sprawdzenia wszystkich możliwych prostych zamian. Jest ich $N(N-1)/2$, a więc $\mathcal{O}(N^2)$.
- Każda prosta zamiana pociąga za sobą konieczność uaktualnienia odległości pomiędzy punktami planu LH. Odległości te tworzą macierz $N \times N$. Należy jednak zauważyć, że za wyjątkiem pierwszej iteracji, nie jest konieczne obliczenie wszystkich odległości międzypunktowych. Zamiana miejscami dwóch elementów kolumny macierzy \mathbf{L} powoduje jedynie konieczność uaktualnienia wartości w dwóch wierszach oraz dwóch kolumnach macierzy odległości. Ze względu na symetrię tej macierzy, pociąga to za sobą wykonanie $2N-3$ obliczeń odległości punktów na jedną operację prostej zamiany, a więc wymaga $\mathcal{O}(N)$ operacji. Po uaktualnieniu macierzy odległości należy następnie obliczyć wartość kryterium optymalizacji, co wiąże się z przeprowadzeniem $\mathcal{O}(N^2)$ dodatkowych operacji. Ostatecznie, oszacowanie nakładu obliczeń wykonywanych przy okazji jednej prostej zamiany T_{swap} wynosi

$$T_{\text{swap}} \sim c_1 N + c_2 N^2. \quad (\text{B.8})$$

Podsumowując powyższe spostrzeżenia można napisać, że złożoność obliczeniowa operacji przeczesywania wyraża się wzorem

$$T_{\text{CP}} \sim n N^2 (c_1 N + c_2 N^2). \quad (\text{B.9})$$

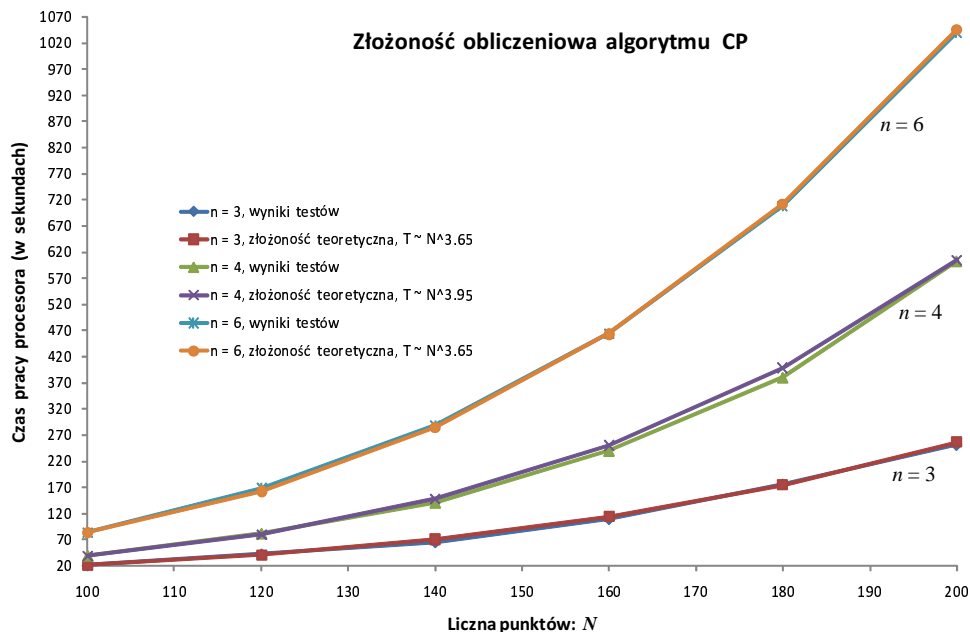
Jedynym elementem, który należy jeszcze uwzględnić w naszej analizie jest (nieznana) liczba powtórzeń operacji przeczesania. Oznaczmy ją przez $\gamma(N, n)$. Na podstawie wykonanych testów wydaje się, że uprawnione jest następujące założenie:

$$\gamma(N, n) \sim N. \quad (\text{B.10})$$

Ostatecznie, równania (B.9) oraz (B.10) prowadzą do oszacowania całkowitej złożoności obliczeniowej algorytmu CP jako

$$T_{\text{tot}} \sim nN^3(c_1N + c_2N^2). \quad (\text{B.11})$$

Na podstawie powyższego wzoru należy podkreślić bardzo silną nieliniową zależność czasu poszukiwania optymalnej konfiguracji hiperkostki łacińskiej od liczby punktów N . Bardzo zgrubnie można założyć, że $T \sim N^q$ gdzie $q \in (3, 5)$. W celu weryfikacji tego oszacowania przeprowadzono szereg testów numerycznych przyjmując $n = 2, 3, 6$ oraz $N = 100, 120, \dots, 200$. Na rys. B.4 pokazane jest porównanie złożoności obliczeniowej uzyskanej z testów z zależnością potęgową typu N^q . Dopasowując wartość wykładnika q do otrzymanych czasów optymalizacji (obliczonych jako wartość średnią z 10 testów dla każdej wartości N) dostajemy $q = 3.65, 3.95, 3.65$ dla, odpowiednio, $n = 3, 4, 6$. Konserwatywnym podejściem wydaje się przyjęcie np. $q = 4$.



Rys. B.4. Weryfikacja teoretycznie złożoności obliczeniowej $T \sim N^q$ algorytmu CP.

Aby uświadomić sobie ogromny nakład obliczeń związany z tworzeniem dużych planów OLH, można oszacować czas utworzenia łacińskiej hiperkostki o $N = 1000$ i $n = 3$ na podstawie wyników dla $N = 100$. Używając jednego

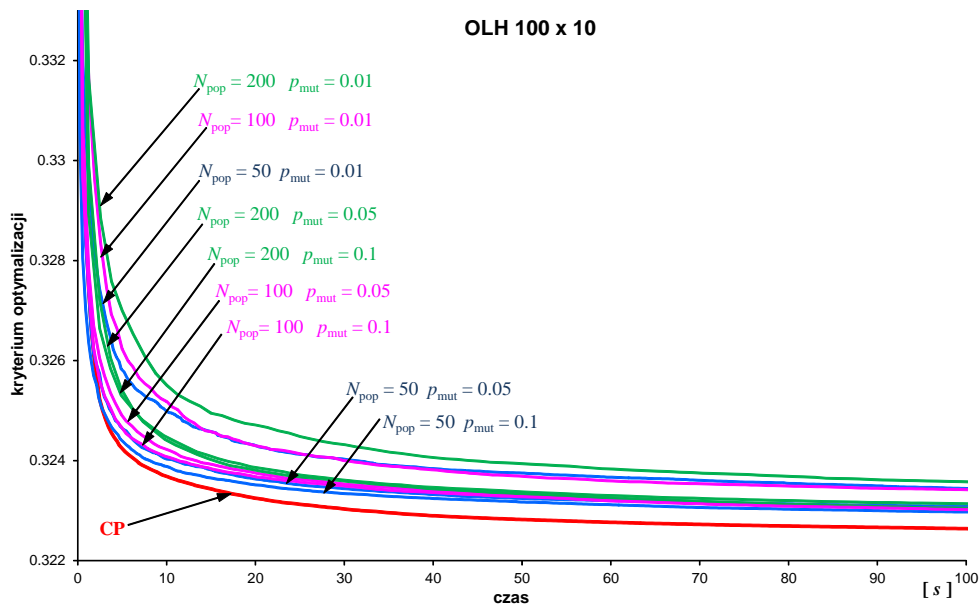
procesora Pentium 4 3.20GHz na komputerze typu PC wygenerowanie takiej hiperkostki zajęłoby ponad 10 dni!

B.6. Dobór parametrów algorytmu genetycznego

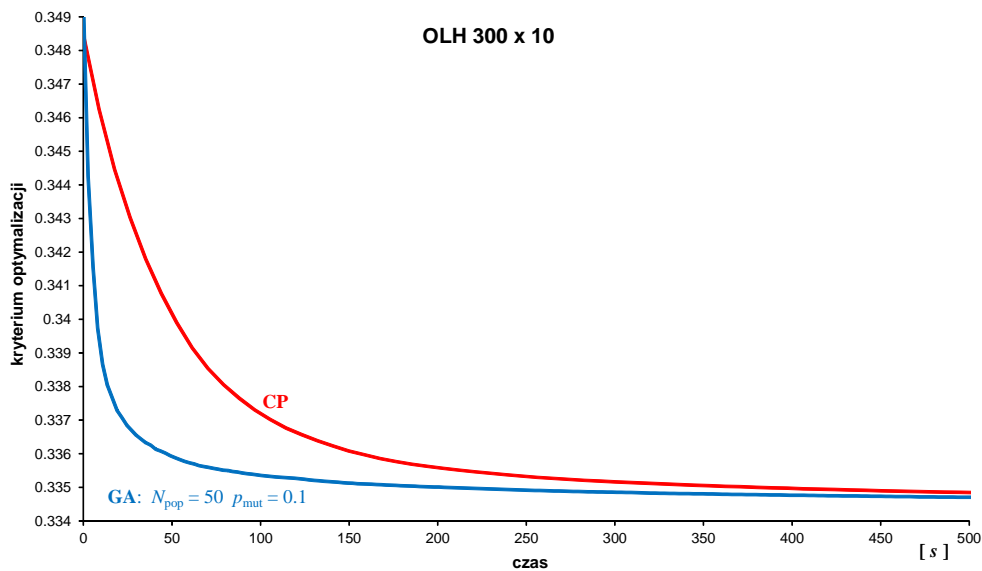
Na podstawie przeprowadzonych licznych testów numerycznych ustalono, że parametrami, które mają decydujący wpływ na efektywność działania algorytmu genetycznego są liczebność populacji N_{pop} oraz prawdopodobieństwo mutacji p_{mut} .

Optymalne wartości tych parametrów zależą zarówno od liczby punktów N jak i liczby zmiennych n . Na rys. B.5 pokazano jak zmienia się kryterium optymalizacji w funkcji czasu pracy procesora w przypadku łańciskiej hiperkostki zawierającej 100 punktów w 10 wymiarach. Można zauważyć, że przyjęcie zbyt małego prawdopodobieństwa mutacji $p_{\text{mut}} = 0.01$ prowadzi do bardzo wolnej zbieżności algorytmu niezależnie od wielkości populacji N_{pop} . Powiększanie populacji hiperkostek zwiększa prawdopodobieństwo wystąpienia korzystnej modyfikacji w trakcie operacji krzyżowania, ale równocześnie wydłuża czas trwania tej operacji. Po wykonaniu wielu testów dla różnych hiperkostek wydaje się, że wartości $N_{\text{pop}} = 50$ i $p_{\text{mut}} = 0.1$ można przyjąć jako “bezpieczne” rozwiązanie dla szerokiej klasy łańciskich hiperkostek.

Warto w tym miejscu zająć się również porównaniem efektywności algorytmu genetycznego oraz algorytmu CP. O ile w zadaniu optymalizacji hiperkostki o $N = 100$ punktach i $n = 10$ wymiarach szybkość zbieżności algorytmu CP przewyższa algorytm genetyczny (zob. rys. B.5), o tyle dla $N = 300$ i $n = 10$ zbieżność algorytmu genetycznego jest znacznie lepsza w porównaniu do CP, rys. B.6. Na podstawie testów numerycznych sugeruje się korzystanie z algorytmu CP w przypadku $N < 150$, a z algorytmu genetycznego w przypadku większych hiperkostek łańciskich. Oczywiście granica ta jest dosyć rozmyta i zależy również w pewnym stopniu od parametru n , ale ogólny trend jest bardzo czytelny - algorytm genetyczny jest lepszym wyborem dla dużych planów LH. Znaczną zaletą tego algorytmu jest również jego stabilność. Krzywe zbieżności zawsze są podobne do tych z rysunków B.5 i B.6, a w czasie testów nigdy nie obserwowano zbieżności do lokalnego minimum ani przypadku, gdzie algorytm CP dawałby zasadniczo lepsze rozwiązanie.



Rys. B.5. Krzywe zbieżności algorytmu CP oraz algorytmu genetycznego dla wielu kombinacji parametrów N_{pop} i p_{mut} w zadaniu optymalizacji LH o $N = 100$ i $n = 10$. Użyto kryterium sił dane funkcją (B.3).



Rys. B.6. Porównanie zbieżności algorytmu CP i algorytmu genetycznego ($N_{pop} = 50$ i $p_{mut} = 0.1$) w przypadku hiperkostki łańciskiej o $N = 300$ punktach i $n = 10$ wymiarach.

B.7. Wykorzystanie łańciskich hiperkostek do generowania próbek zmiennych losowych o dowolnych rozkładach prawdopodobieństwa

Opiszemy tu, jak wykorzystać otrzymaną w procesie optymalizacji macierz \mathbf{L} , opisującą przy pomocy współrzędnych całkowitoliczbowych układ punktów OLH, do generowania próbek zmiennych losowych \mathbf{X} o dowolnych funkcjach gęstości rozkładu prawdopodobieństwa. Metoda ta oczywiście będzie wszystkich łańciskich hiperkostek, których szczególnym przypadkiem są hiperkostki optymalne - OLH.

Przypomnijmy, że każda kolumna macierzy \mathbf{L} odpowiada jednej zmiennej losowej o zadanym rozkładzie prawdopodobieństwa. Kolumna ta jest permutacją liczb od 1 do N . Realizacja $x_k(m)$ zmiennej losowej X_k , $1 \leq k \leq n$, odpowiadającej liczbie m , $1 \leq m \leq N$, w k -tej kolumnie macierzy \mathbf{L} wyznaczana jest przy pomocy dystrybuanty zmiennej X_k w następujący sposób:

$$x_k(m) = F_{X_k}^{-1}(\tilde{x}_m), \quad (\text{B.12})$$

gdzie

$$\tilde{x}_m = \frac{m}{N} - \frac{1}{2N}. \quad (\text{B.13})$$

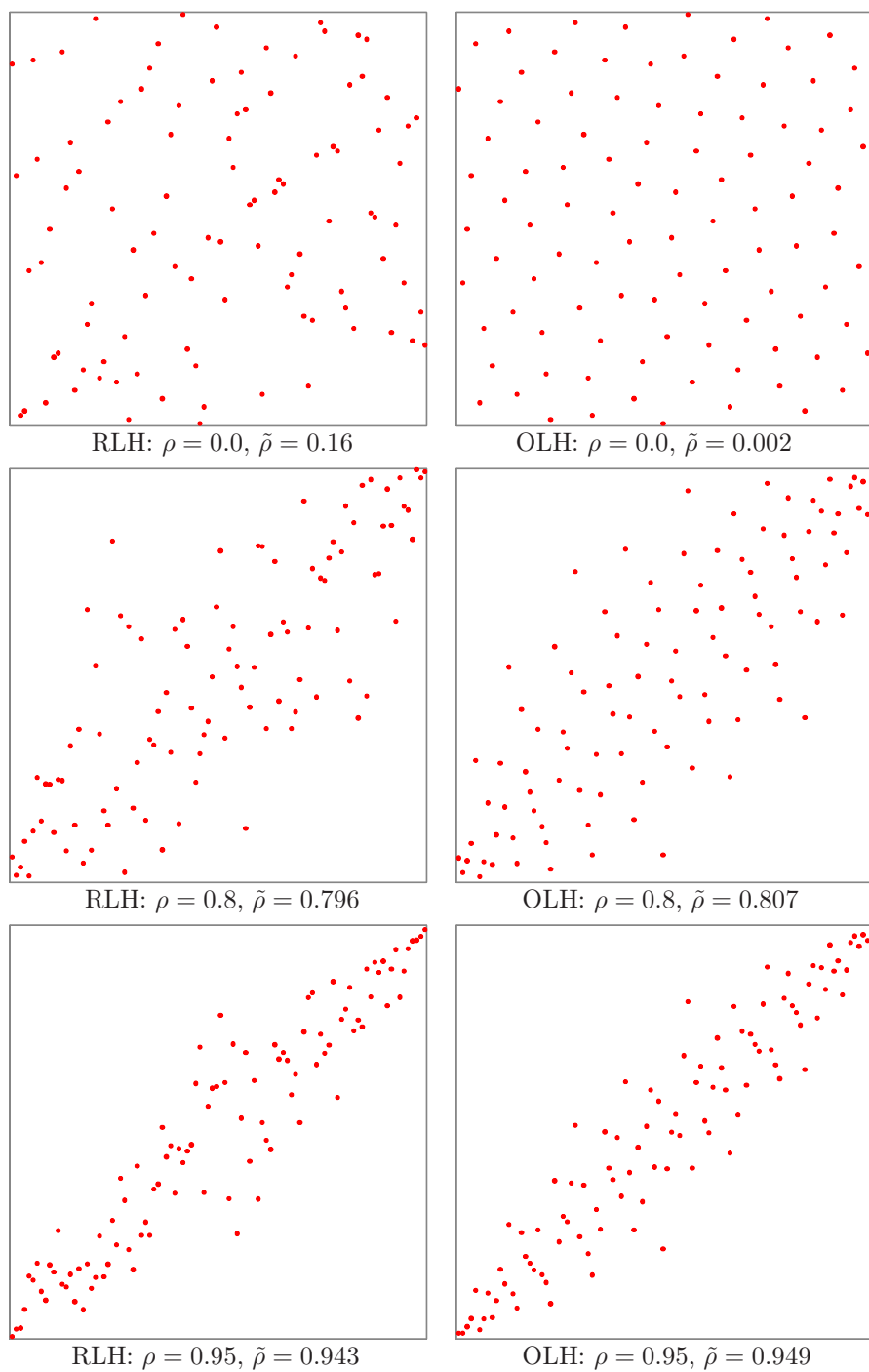
Metoda kryjąca się za powyższymi wzorami sprowadza się do podzielenia zakresu zmienności każdej zmiennej losowej na N przedziałów o równym prawdopodobieństwie, a realizacje $x_k(i)$, $i = 1, \dots, N$, odpowiadają probabilistycznym środkom (medianom) zmiennej X_k w tych przedziałach. Innym możliwym rozwiązaniem jest wybór $x_k(i)$ w sposób losowy wewnątrz i -tego przedziału. W pracach [119] i [100] pokazano jednak, że wybór median bądź też wartości średnich (tutaj niezbędne jest całkowanie numeryczne) zmiennych losowych wewnątrz przedziałów prowadzi do bardziej precyzyjnej estymacji (mniejszej wariancji estymatorów). Zagadnieniom wyboru położenia punktów wewnątrz komórki łańciskiej hiperkostki poświęcono więcej uwagi w rozdziale 4.

Należy pamiętać, że w ogólnym przypadku zmienne losowe mogą posiadać dowolne rozkłady prawdopodobieństwa oraz być skorelowane. Z drugiej strony, opisana metoda generowania realizacji zmiennych losowych na podstawie planu eksperymentów danego przez hiperkostkę łańciską nie uwzględnia żadnej wzajemnej zależności zmiennych. Aby zatem móc wykorzystać LH w symulacjach losowych, oryginalne zmienne muszą być uprzednio przetransformowane na zbiór niezależnych zmiennych losowych. W przypadku, gdy znana jest funkcja łącznej gęstości prawdopodobieństwa zmiennych możliwe jest zastosowanie

transformacji Rosenblatta [209], a jeżeli znane są jedynie brzegowe rozkłady prawdopodobieństwa zmiennych losowych oraz macierz współczynników korelacji, można użyć transformacji Natafa [183]. Zachowując miarę prawdopodobieństwa, obydwie te transformacje przekształcają oryginalne zmienne losowe na niezależne standardowe zmienne gaussowskie, zob. również podrozdział A.2 w dodatku A. Realizacje zmiennych wygenerowane w przestrzeni standardowej zgodnie z regułą (B.12) są następnie przy pomocy transformacji odwrotnej przekształcane na zmienne oryginalne \mathbf{X} . Na rys. B.7 przedstawiono przykładowe plany eksperymentów generowanych przy pomocy OLH i RLH w przypadku dwóch jednakowych zmiennych o rozkładach jednostajnych i dla różnych wartości współczynnika korelacji wzajemnej $\rho_{XY} = \text{Cov}(X, Y) / [\sigma(X)\sigma(Y)]$ (współczynnika Pearsona). Porównując otrzymane wykresy można zauważyć, że niezależnie od wartości ρ , układ punktów OLH charakteryzuje się pewną regularnością, symetrią i jednorodnością. Trzeba jednak zaznaczyć, że wspomniane transformacje nie gwarantują równomiernego rozłożenia punktów przy rzutowaniu na poszczególne osie. Inne znane autorowi metody “narzucenia” korelacji pomiędzy zmiennymi losowymi, które zachowują jednocześnie równomierność rozłożenia punktów przy rzutowaniu, polegają na odpowiedniej zmianie kolejności kolumn macierzy \mathbf{L} , zob. [89, 100, 105]. Algorytmy te są w większości przypadków bardziej złożone obliczeniowo od metod transformacyjnych i przeznaczone są zazwyczaj do narzucenia wartości tzw. współczynnika korelacji Spearmana (jest to współczynnik korelacji Pearsona, liczony dla rang zmiennych). Wydaje się jednak, że podejście wykorzystujące transformację do zmiennych niezależnych jest bardziej uniwersalne i nie wymaga tworzenia nowej łacińskiej hiperkostki dla każdej macierzy korelacji. Pozwala to na wykorzystanie tej samej, uprzednio przygotowanej optymalnej łacińskiej hiperkostki, do przypadku dowolnej korelacji zmiennych losowych.

B.8. Porównanie kryteriów optymalizacji w zadaniu estymacji momentów funkcji losowych

W podrozdziale B.1 wprowadzono dwa kryteria optymalizacji używane do znalezienia optymalnej konfiguracji łacińskiej hiperkostki. Są to: kryterium najmniejszej odległości oraz kryterium sił. W celu porównania tych kryteriów oraz aby oszacować poprawność estymacji momentów statystycznych funkcji losowych przy użyciu metod symulacyjnych wykorzystujących plany OLH, przeprowadzono szereg testów z tzw. funkcją Rosenbrocka (zwaną też funkcją bananową)



Rys. B.7. Porównanie rozrzutu 100 punktów eksperymentalnych planów OLH i RLH w przypadku dwóch jednakowych zmiennych o rozkładzie jednostajnym. ρ oznacza zakładany współczynnik korelacji, a $\tilde{\rho}$ jest współczynnikiem obliczonym z wygenerowanej próbki.

$$b(X_1, X_2) = 100(X_2 - X_1^2)^2 + (1 - X_1)^2. \quad (\text{B.14})$$

Założono, że dwie zmienne losowe X_1 i X_2 mają rozkład jednostajny w przedziale $[0, 2]$. Funkcja łącznej gęstości rozkładu prawdopodobieństwa jest w tym przypadku stała i wynosi $f_{X_1, X_2}(x_1, x_2) = 1/4$. Celem symulacji losowych jest oszacowanie wartości średniej funkcji $b(X_1, X_2)$. Wartość ta może być wyznaczona analitycznie, co umożliwi lepszą ocenę wyników symulacji

$$\mathbb{E}[b(X_1, X_2)] = \int_0^2 \int_0^2 b(x_1, x_2) \frac{1}{4} dx_1 dx_2 = 187. \quad (\text{B.15})$$

Wartość średnia (B.15) oszacowana będzie za pomocą następujących metod symulacyjnych:

- Symulacje z użyciem OLH, kryterium sił - OLH1.
- Symulacje z użyciem OLH, kryterium najmniejszej odległości - OLH2.
- Łacińska hiperkostka bez optymalizacji - RLH.
- Standardowa metoda Monte Carlo - MC.

Wyniki testów zebrane zostały w tabeli B.1, w której przedstawiono wartości średnich błędów procentowych estymacji $\mathbb{E}[b(X_1, X_2)]$ w funkcji wielkości próbki N . Dla wybranej metody, wielkości błędów obliczane są przez porównanie wartości średniej oszacowania z M powtórzeń symulacji losowych z wartością dokładną

$$\frac{1}{187M} \sum_{k=1}^M |\bar{b}^{(k)} - 187|, \quad (\text{B.16})$$

gdzie $\bar{b}^{(k)}$ jest k -tą realizacją wartości średniej, a M przyjęto równe 10.

Analizując otrzymane wyniki łatwo jest dokonać klasyfikacji metod. Najmniejszy błąd estymacji przy ustalonym N otrzymuje się używając pierwszego rodzaju symulacji OLH (według kryterium sił), nieco gorszy jest OLH z kryterium minimalnej odległości, następnie RLH, a największy błąd daje metoda MC.

Jak to zostało pokazane w punktach B.5 i B.6, koszt wygenerowania dużych planów OLH ($N > 500$) jest bardzo wysoki. Jednak mając na uwadze analizę niezawodności bądź optymalizację odpornościową złożonych i kosztownych obliczeniowo konstrukcji oraz procesów technologicznych, trudno wyobrazić sobie możliwość wykonania tysięcy eksperymentów numerycznych. Siłą optymalnych łacińskich hiperkostek jest zapewnienie wystarczająco dokładnej estymacji przy użyciu niewielkich próbek.

N	OLH1	OLH2	RLH	MC
10	9.1	12.5	20.7	37.8
20	3.5	7.9	14.3	17.8
50	1.5	3.9	10.2	12.5
100	1.1	2.7	6.6	9.4
200	0.6	1.8	5.6	6.0
500	0.3	0.8	2.4	4.8
1000	-	-	1.5	3.2
2000	-	-	1.1	2.2
5000	-	-	0.7	1.2

Tablica B.1. Błąd procentowy oszacowania wartości średniej funkcji Rosenbrocka przy pomocy różnych metod symulacyjnych. Plany OLH o N większym niż 500 nie były generowane ze względu na zbyt długi czas obliczeń.

Metody dyskretyzacji pól losowych

Niniejszy dodatek poświęcony jest zagadnieniom dotyczącym modelowania przestrzennej losowej zmienności parametrów układów konstrukcyjnych przy pomocy pól losowych, a w szczególności metodom ich dyskretyzacji. Problematykę tę uznano za istotną i wymagającą dokładniejszego przedstawienia przede wszystkim ze względu na związek metody optymalnej aproksymacji liniowej z metodą krigingu (dominującą techniką powierzchni odpowiedzi stosowaną w optymalizacji odpornościowej) oraz z używanymi w przykładach numerycznych metodami dyskretyzowania pól losowych.

C.1. Podstawowe definicje

Wartości niektórych parametrów opisujących materiał, geometrię, wielkość i rozkład uszkodzeń czy też obciążenia działające na układ konstrukcyjny charakteryzują się losową zmiennością nie tylko w obrębie populacji, ale również przestrzennymi losowymi fluktuacjami w ramach samej konstrukcji. Takie parametry modelowane mogą być przy pomocy *pól losowych*.

Skalarne pole losowe o wartościach rzeczywistych zdefiniować można jako rodzinę jednowymiarowych zmiennych losowych $H(\mathbf{z})$ indeksowanych wektorem współrzędnych \mathbf{z} określającym położenie w k wymiarowym obszarze $\Omega \subset \mathbb{R}^k$

$$H(\mathbf{z}) \in \mathbb{R}, \quad \mathbf{z} = \{z_1, \dots, z_k\} \in \Omega \subset \mathbb{R}^k. \quad (\text{C.1})$$

Zmienna losowa $H(\mathbf{z})$, nazywana jest *wartością pola losowego* w punkcie \mathbf{z} . W literaturze pola losowe określane są również jako procesy losowe (stochastyczne). Najczęściej jednak termin pole losowe używany jest gdy $k = 2$ lub 3 , natomiast proces losowy odnosi się zazwyczaj do przypadku jednowymiarowego, gdzie parametr położenia utożsamia się najczęściej z czasem. Zbiór realizacji zmiennych $H(\mathbf{z})$ we wszystkich punktach $\mathbf{z} \in \Omega$ nazywamy *realizacją pola losowego*.

Pole losowe $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, nazywamy polem losowym drugiego rzędu, jeżeli spełniony jest następujący warunek:

$$\mathbb{E}[|H(\mathbf{z})|^2] < \infty. \quad (\text{C.2})$$

W przypadku takiego pola określić można *funkcję wartości oczekiwanej*

$$\mu_H(\mathbf{z}) = \mathbb{E}[H(\mathbf{z})], \quad (\text{C.3})$$

oraz *funkcję kowariancji* pola losowego (ang. *auto-covariance function*) daną jako

$$K_H(\mathbf{z}, \mathbf{z}') = \mathbb{E}[(H(\mathbf{z}) - \mu_H(\mathbf{z})) (H(\mathbf{z}') - \mu_H(\mathbf{z}'))], \quad (\text{C.4})$$

będącą w istocie kowariancją między zmiennymi losowymi $H(\mathbf{z})$ i $H(\mathbf{z}')$ określonymi w punktach \mathbf{z} i \mathbf{z}' . Funkcja kowariancji zawiera informację na temat przestrzennej zmienności pola losowego.

Pole losowe $H(\mathbf{z})$ nazywamy *ślabo stacjonarnym* (ang. *weakly homogenous*) jeżeli

$$\forall \mathbf{z} \in \Omega \quad \mu_H(\mathbf{z}) = \text{const} \quad \text{oraz} \quad \forall \mathbf{z} \in \Omega, \mathbf{z} + \Delta \mathbf{z} \in \Omega \quad K_H(\mathbf{z}, \mathbf{z} + \Delta \mathbf{z}) = K_H(\Delta \mathbf{z}). \quad (\text{C.5})$$

Stacjonarne pole losowe jest *izotropowe*, jeśli jego funkcja kowariancji zależy jedynie od odległości pomiędzy dwoma punktami.

$$\forall \mathbf{z} \in \Omega \quad K_H(\mathbf{z}, \mathbf{z} + \Delta \mathbf{z}) = K_H(\|\Delta \mathbf{z}\|). \quad (\text{C.6})$$

Uzupełniając, warto wprowadzić również definicję *funkcji wariancji* pola losowego

$$\text{Var}[H(\mathbf{z})] = K_H(\mathbf{z}, \mathbf{z}) = \sigma_H^2(\mathbf{z}) = \mathbb{E}[(H(\mathbf{z}) - \mu_H(\mathbf{z}))^2], \quad (\text{C.7})$$

gdzie $\sigma_H^2(\mathbf{z})$ jest funkcją odchylenia standardowego, oraz przy pomocy (C.4) i (C.7) definicję *funkcji korelacji*

$$\rho_H(\mathbf{z}, \mathbf{z}') = \frac{K_H(\mathbf{z}, \mathbf{z}')}{\sigma_H(\mathbf{z})\sigma_H(\mathbf{z}')} \quad (\text{C.8})$$

W większości istotnych z praktycznego punktu widzenia przypadków korelacja między zmiennymi $H(\mathbf{z})$ i $H(\mathbf{z}')$ zanika wraz z odległością punktów \mathbf{z} i \mathbf{z}' . Dlatego też, spośród funkcji korelacji najczęściej używanych w modelowaniu pól losowych wymienić należy funkcje:

- trójkątną

$$\rho(\Delta \mathbf{z}) = \begin{cases} 1 - \frac{\|\Delta \mathbf{z}\|}{a} & \text{gdy } \|\Delta \mathbf{z}\| \leq a \\ 0 & \text{gdy } \|\Delta \mathbf{z}\| > a \end{cases}, \quad (\text{C.9})$$

gdzie parametr a jest pewną miarą “zasiegu” fluktuacji pola losowego zwaną *długością korelacji*;

- wykładniczą

$$\rho(\Delta \mathbf{z}) = \exp\left(-\frac{\|\Delta \mathbf{z}\|}{a}\right); \quad (\text{C.10})$$

- gaussowską

$$\rho(\Delta \mathbf{z}) = \exp\left(-\frac{\|\Delta \mathbf{z}\|^2}{a^2}\right). \quad (\text{C.11})$$

Długość korelacji a definiuje się formalnie jako

$$a = \frac{\int_0^\infty \delta |K_H(\delta)| d\delta}{\int_0^\infty |K_H(\delta)| d\delta}, \quad (\text{C.12})$$

gdzie $\delta = \|\Delta \mathbf{z}\|$.

Pole losowe nazywamy gaussowskim, jeśli każdy wektor $\{H(\mathbf{z}_1), \dots, H(\mathbf{z}_n)\}$, dla dowolnego $n \geq 1$, ma rozkład normalny. Pole to jest w sposób jednoznaczny określone przez jego wartość średnią $\mu_H(\mathbf{z})$, wariancję $\sigma_H^2(\mathbf{z})$ oraz funkcję korelacji $\rho(\Delta \mathbf{z})$.

Reprezentowanie pola losowego jako niepoliczalnego zbioru zmiennych losowych nie jest wygodne w obliczeniach numerycznych, dlatego też zazwyczaj stosuje się przybliżenie ciągłego pola losowego przy pomocy skończonej (dyskretnej) liczby zmiennych losowych, tzw. dyskretyzację pola losowego, cf. [254].

Dyskretyzacja skalarne pola losowego $H(\mathbf{z})$ zdefiniowana jest przez następującą funkcję aproksymującą:

$$H(\mathbf{z}) \xrightarrow{\text{Dyskretyzacja}} \hat{H}(\mathbf{z}) = \mathcal{H}(\mathbf{z}, \mathbf{X}), \quad (\text{C.13})$$

gdzie $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ jest n -wymiarowym wektorem losowym. W ogólności funkcję $\mathcal{H}(\mathbf{z}, \mathbf{X})$ przedstawić można jako

$$\mathcal{H}(\mathbf{z}, \mathbf{X}) = c_0(\mathbf{z}) + \sum_{i=1}^n c_i(\mathbf{z}) X_i, \quad (\text{C.14})$$

gdzie $c_i, i = 0, 1, \dots, n$, są funkcjami $\mathbf{z} \in \Omega$, a $X_i, i = 1, \dots, n$, są zmiennymi losowymi, na których opiera się dyskretyzacja. Istnieje wiele metod dyskretyzacji. Większość z nich wykorzystuje podział obszaru Ω , na którym dane jest pole losowe, na rozłączne, przylegające podobszary (zwane elementami pola losowego) $\Omega_i^e, i = 1, 2, \dots, N_e$. W przypadku kiedy analiza stochastyczna, która wymaga użycia modeli pól losowych, stosowana jest w połączeniu z analizą metodą elementów skończonych, warto jest tak dobrać elementy pola losowego, aby ich siatka odpowiadała w pewien sposób siatce MES. Za [245] można podać następującą klasyfikację metod dyskretyzacji pól losowych:

- *Metody dyskretyzacji punktowej* (ang. *point discretization*), w których zmienne losowe $\{X_1, X_2, \dots, X_n\}$ są wartościami pola losowego $H(\mathbf{z}_i)$ w wybranych punktach $\mathbf{z}_i, i = 1, \dots, n$, obszaru Ω , np. charakterystycznych punktach elementów $\Omega_i^e, i = 1, 2, \dots, N_e$.
- *Metody uśredniania*, w których zmienne \mathbf{X} są ważonymi całkami pola losowego $H(\mathbf{z})$ wewnątrz elementów Ω_i^e

$$X_i = \int_{\Omega_i^e} H(\mathbf{z})w(\mathbf{z}) d\Omega. \quad (\text{C.15})$$

Metody uśredniania, jak również metody dyskretyzacji punktowej, pomimo heurystycznego charakteru, są chętnie stosowane w dyskretyzacji pól losowych przede wszystkim ze względu na swoją prostotę.

- *Metody aproksymacji funkcjami kształtu*, wykorzystujące znane z metody elementów skończonych funkcje kształtu oraz skończony zbiór (węzłowych) wartości pola losowego. Metody te można traktować jako szczególny przypadek metod dyskretyzacji punktowej.
- *Metody rozwinięcia w szereg*, wykorzystujące reprezentacje pola losowego za pomocą odpowiednio obciętego rozwinięcia w szereg funkcji deterministycznych ze współczynnikami losowymi, zob. [73].

W kolejnych podrozdziałach omówionych zostanie jedynie kilka wybranych metod dyskretyzacji. Jak już wspomniano we wstępie, szczególnie dużo uwagi poświęcone zostanie metodzie optymalnej aproksymacji liniowej ze względu na jej związek z używaną w pracy metodą krigingu, zob. podrozdział 3.2.

C.2. Dyskretyzacja punktowa

Istotą metody dyskretyzacji punktowej jest reprezentowanie pola losowego w każdym z elementów $\Omega_i^e, i = 1, 2, \dots, N_e$, przez zmienną losową będącą warto-

ścią tego pola w wybranym punkcie elementu. Najpopularniejszym wariantem metody dyskretyzacji punktowej jest tzw. metoda *punktów środkowych* (ang. *midpoint method* - MP), w której pole losowe wewnątrz elementu Ω_i^e przybliżane jest wartością tego pola w środku ciężkości elementu, zob. [44].

Wprowadza się następujący wektor losowy:

$$\mathbf{X}_{\text{MP}} = \{H(\mathbf{z}_1^*), H(\mathbf{z}_2^*), \dots, H(\mathbf{z}_{N_e}^*)\}, \quad (\text{C.16})$$

gdzie $X_i = H(\mathbf{z}_i^*)$, $i = 1, 2, \dots, N_e$, są wartościami pola losowego w środkach ciężkości elementów. Zakłada się, że własności wektora \mathbf{X}_{MP} wynikają bezpośrednio z własności pola losowego $H(\mathbf{z})$, na przykład,

$$\text{Cov}(X_m, X_n) = K_H(\mathbf{z}_m^*, \mathbf{z}_n^*), \quad m, n \in \{1, 2, \dots, N_e\}. \quad (\text{C.17})$$

Ostatecznie, aproksymację pola losowego $H(\mathbf{z})$ metodą punktów środkowych zapisać można jako

$$\hat{H}_{\text{MP}}(\mathbf{z}) = H(\mathbf{z}_i^*), \quad \mathbf{z} \in \Omega_i^e, \quad i = 1, 2, \dots, N_e, \quad (\text{C.18})$$

lub inaczej

$$\hat{H}_{\text{MP}}(\mathbf{z}) = \sum_{i=1}^{N_e} I_{\Omega_i^e}(\mathbf{z}) H(\mathbf{z}_i^*), \quad (\text{C.19})$$

gdzie $I_{\Omega_i^e}$, $i = 1, \dots, N_e$, są funkcjami charakterystycznymi elementów pola losowego. Realizacje $\hat{H}_{\text{MP}}(\mathbf{z})$ są funkcjami schodkowymi o wartościach stałych wewnątrz elementów pola losowego Ω_i^e , $i = 1, 2, \dots, N_e$, z nieciągłościami na brzegach elementów. Można pokazać, zob. [44], że dyskretyzacja metodą punktów środkowych prowadzi do większej losowej zmienności pola $\hat{H}_{\text{MP}}(\mathbf{z})$ wewnątrz elementów w porównaniu z oryginalnym polem losowym $H(\mathbf{z})$.

Inna z wersji dyskretyzacji punktowej jest ściśle związana z tzw. stochastyczną metodą elementów skończonych (ang. *Stochastic Finite Element Method* - SFEM) [123]. Ponieważ wszystkie operacje całkowania wykonywane w trakcie rozwiązania zadania MES przeprowadza się obliczając wartości funkcji podcałkowych w punktach Gaussa elementów skończonych, dyskretyzacji pola losowego dokonuje się właśnie w punktach całkowania elementów, zob. [18]. Metoda ta pozwala otrzymać bardzo dobrą aproksymację pola losowego w przypadku małych wartości długości korelacji, lecz jej podstawową wadą jest gwałtowny wzrost liczby zmiennych losowych wraz z wielkością siatki MES.

C.3. Metoda średniej przestrzennej

Do grupy metod uśredniania zaliczyć można dyskretyzację pola losowego metodą średniej przestrzennej (ang. *spatial average method* - SA), zob. [255] i [127]. W przeciwieństwie do metody dyskretyzacji punktowej, gdzie pole losowe w elemencie reprezentowane jest przez wartości tego pola w arbitralnie wybranych punktach, w metodzie średniej przestrzennej pole losowe uśredniane jest wewnątrz elementu, co prowadzi do następującej definicji:

$$\hat{H}_{\text{SA}}(\mathbf{z}) = \frac{\int_{\Omega_i^e} H(\mathbf{z}) d\Omega_i^e}{|\Omega_i^e|} = \bar{H}_i, \quad \mathbf{z} \in \Omega_i^e, \quad i = 1, 2, \dots, N_e. \quad (\text{C.20})$$

Konsekwentnie, wektor $\mathbf{X}_{\text{SA}} = \{\bar{H}_1, \bar{H}_2, \dots, \bar{H}_{N_e}\}$ przyjmuje się jako wektor zmiennych losowych stanowiących podstawę dyskretyzacji. Ostatecznie, podobnie jak w metodzie punktów środkowych, dyskretyzację pola losowego można zapisać w postaci następującej sumy:

$$\hat{H}_{\text{SA}}(\mathbf{z}) = \sum_{i=1}^{N_e} I_{\Omega_i^e}(\mathbf{z}) \bar{H}_i. \quad (\text{C.21})$$

Niestety podejście to ma zastosowanie jedynie w przypadku gaussowskich pól losowych. Tylko wtedy możliwe jest otrzymanie funkcji gęstości wektora \mathbf{X}_{SA} . Wartości średnie oraz kowariancje zmiennych \mathbf{X}_{SA} dane są poprzez odpowiednie całki funkcji wartości średnich oraz kowariancji pola losowego w elementach dyskretyzacji.

Aproksymacja metodą średniej przestrzennej prowadzi do pola o większej regularności w porównaniu z oryginalnym polem losowym. Można zatem traktować tę metodę oraz metodę punktów środkowych jako, odpowiednio, dolne i górne ograniczenia losowej zmienności dyskretyzowanego pola.

C.4. Metoda funkcji kształtu

Dyskretyzacja pola losowego metodą funkcji kształtu (ang. *shape function method* - SF), zob. [155, 156], wykorzystuje koncepcję funkcji aproksymujących znaną z metody elementów skończonych. Obszar Ω , w którym określone jest pole losowe, tak jak w przypadku poprzednich metod, dzielony jest na elementy Ω_i^e , $i = 1, 2, \dots, N_e$. Aproksymacja pola wewnątrz tych elementów dana jest

jako

$$\hat{H}_{\text{SF}}(\mathbf{z}) = \sum_{j=1}^{q_i} N_j(\mathbf{z})H(\mathbf{z}_j), \quad \mathbf{z} \in \Omega_i^e, \quad i = 1, 2, \dots, N_e, \quad (\text{C.22})$$

gdzie q_i jest liczbą węzłów elementu Ω_i^e , $N_j(\mathbf{z})$, $j = 1, \dots, q_i$, są wartościami wielomianowych funkcji kształtu w punkcie \mathbf{z} , a \mathbf{z}_j są współrzędnymi węzłów. Co za tym idzie, funkcja wartości oczekiwanej oraz funkcja kowariancji aproksymacji typu SF mają następującą postać:

$$\mu_{\hat{H}_{\text{SF}}}(\mathbf{z}) = \sum_{i=1}^{q_i} N_i(\mathbf{z})\mu_H(\mathbf{z}_i), \quad (\text{C.23})$$

$$K_{\hat{H}_{\text{SF}}}(\mathbf{z}, \mathbf{z}') = \sum_{i=1}^{q_i} \sum_{j=1}^{q_j} N_i(\mathbf{z})N_j(\mathbf{z}')K_H(\mathbf{z}_i, \mathbf{z}_j), \quad (\text{C.24})$$

gdzie q_i jest liczbą węzłów w elemencie zawierającym punkt \mathbf{z} , a q_j jest liczbą węzłów w elemencie zawierającym punkt \mathbf{z}' . Realizacje pola losowego \hat{H}_{SF} są funkcjami ciągłymi na Ω , co stanowi niewątpliwą zaletę tej metody dyskretyzacji w porównaniu z metodą punktów środkowych czy też metodą uśredniania.

C.5. Metoda optymalnej aproksymacji liniowej

Metoda optymalnej aproksymacji liniowej (ang. *optimal linear estimation method* - OLE) opisana została przez Li i Der Kiureghian'a w pracy [147]. Można ją rozpatrywać jako szczególny przypadek metody funkcji kształtu, ponieważ w dyskretyzacji metodą OLE również wykorzystuje się wektor wartości pola losowego w punktach węzłowych

$$\mathbf{X} = \{H(\mathbf{z}_1), H(\mathbf{z}_2), \dots, H(\mathbf{z}_q)\}, \quad (\text{C.25})$$

gdzie \mathbf{z}_i , $i = 1, \dots, q$ są współrzędnymi węzłów, a aproksymacja pola losowego zdefiniowana jest wzorem (por. (C.22)):

$$\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X}) = a(\mathbf{z}) + \sum_{i=1}^q b_i(\mathbf{z})X_i. \quad (\text{C.26})$$

Dla ustalonego $\mathbf{z} \in \Omega$ wartości współczynników $a(\mathbf{z})$ i $b_i(\mathbf{z})$, $i = 1, \dots, q$, wyznaczane są przez rozwiązanie następującego zadania optymalizacji:

$$\text{znajdź minimum: } \text{Var}[H(\mathbf{z}) - \hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})], \quad (\text{C.27})$$

$$\text{przy ograniczeniu: } \mathbb{E}[H(\mathbf{z}) - \hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})] = 0. \quad (\text{C.28})$$

Poszukuje się zatem takich wartości $a(\mathbf{z})$ i $b_i(\mathbf{z})$, które minimalizują wariancję błędu aproksymacji przy jednoczesnym zapewnieniu, że aproksymacja \hat{H}_{OLE} w punkcie \mathbf{z} jest nieobciążona, tzn. $\mathbb{E}[H(\mathbf{z})] = \mathbb{E}[\hat{H}_{OLE}(\mathbf{z}, \mathbf{X})]$.

Podstawiając (C.26) do warunku (C.28) współczynnik $a(\mathbf{z})$ wyznaczyć można jako

$$a(\mathbf{z}) = \mu_H(\mathbf{z}) - \sum_{i=1}^q b_i(\mathbf{z})\mathbb{E}[X_i]. \quad (\text{C.29})$$

Podstawiając następnie powyższe wyrażenie do (C.26) otrzymuje się

$$\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) = \mu_H(\mathbf{z}) + \sum_{i=1}^q b_i(\mathbf{z}) (X_i - \mathbb{E}[X_i]). \quad (\text{C.30})$$

Uwzględnienie (C.30) w wyrażeniu na wariancję przy jednoczesnym wykorzystaniu warunku (C.28), prowadzi do następującej postaci funkcji celu (C.27):

$$\begin{aligned} \text{Var} [H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z})] &= \mathbb{E} \left[\left(H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}) \right)^2 \right] = \\ &= \sigma_H^2(\mathbf{z}) - 2 \sum_{i=1}^q b_i(\mathbf{z}) \text{Cov}[H(\mathbf{z}), X_i] + \sum_{i=1}^q \sum_{j=1}^q b_i(\mathbf{z}) b_j(\mathbf{z}) \text{Cov}[X_i, X_j]. \end{aligned} \quad (\text{C.31})$$

Wartości funkcji $b_i(\mathbf{z})$ powinny minimalizować funkcję (C.31) w każdym punkcie obszaru Ω . Warunek konieczny istnienia minimum, zerowanie się pierwszych pochodnych (C.31) ze względu na $b_i(\mathbf{z})$, dany jest jako

$$\forall i = 1, 2, \dots, q \quad -\text{Cov}[H(\mathbf{z}), X_i] + \sum_{j=1}^q b_j(\mathbf{z}) \text{Cov}[X_i, X_j] = 0. \quad (\text{C.32})$$

Ostatecznie, na podstawie (C.32), wartości b_i w punkcie \mathbf{z} oblicza się rozwiązując następujący układ równań liniowych:

$$\mathbf{b}(\mathbf{z}) = \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z}), \quad (\text{C.33})$$

gdzie $\mathbf{b}(\mathbf{z}) = \{b_1(\mathbf{z}), b_2(\mathbf{z}), \dots, b_q(\mathbf{z})\}$, $\mathbf{C}_{\mathbf{X}}$ jest macierzą kowariancji zmiennych losowych \mathbf{X} , a $\mathbf{C}_{HX}(\mathbf{z}) = \{\text{Cov}[H(\mathbf{z}), X_1], \text{Cov}[H(\mathbf{z}), X_2], \dots, \text{Cov}[H(\mathbf{z}), X_q]\}$. Wykorzystując powyższy wynik w wyrażeniu (C.30) otrzymuje się macierzową postać wzoru na dyskretyzację pola losowego metodą OLE

$$\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) = \mu_H(\mathbf{z}) + [\mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z})]^T (\mathbf{X} - \boldsymbol{\mu}_X), \quad (\text{C.34})$$

gdzie $\boldsymbol{\mu}_X = \{\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_q]\}$. Wyodrębniając we wzorze (C.34) część deterministyczną można przedstawić go w następującej równoważnej postaci:

$$\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X}) = [\boldsymbol{\mu}_H(\mathbf{z}) - \mathbf{C}_{HX}(\mathbf{z})\mathbf{C}_X^{-1}\boldsymbol{\mu}_X] + \sum_{i=1}^q [\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z})]_i X_i. \quad (\text{C.35})$$

Łatwo zauważyć, że metoda optymalnej aproksymacji liniowej bardzo podobna jest do dyskretyzacji pola losowego metodą funkcji kształtu (C.22). Pomijając wspólny składnik w wyrażeniu (C.35), funkcje kształtu odpowiadające punktom węzłowym w metodzie OLE zapisać można jako

$$N_i^{\text{OLE}}(\mathbf{z}) = [\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z})]_i = \sum_{j=1}^q (\mathbf{C}_X^{-1})_{ij} \sigma_H(\mathbf{z}) \sigma_H(\mathbf{z}_j) \rho_H[H(\mathbf{z}), X_j], \quad (\text{C.36})$$

$$i = 1, \dots, q.$$

Wariancja dyskretyzacji OLE w punkcie \mathbf{z} wyraża się następująco:

$$\begin{aligned} \text{Var} [\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})] &= \mathbb{E} \left[\left(\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X}) - \mathbb{E} [\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})] \right)^2 \right] = \\ &= \mathbb{E} \left[\left(\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X}) - \boldsymbol{\mu}_H(\mathbf{z}) \right)^2 \right] \\ &= \mathbb{E} [\mathbf{C}_{HX}^T(\mathbf{z})\mathbf{C}_X^{-1}(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{X} - \boldsymbol{\mu}_X)^T\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z})] \\ &= \mathbf{C}_{HX}^T(\mathbf{z})\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z}). \end{aligned} \quad (\text{C.37})$$

Wykorzystując (C.33) wariancja błędu aproksymacji (C.31) dana jest w postaci macierzowej jako

$$\begin{aligned} \text{Var} [H(\mathbf{z}) - \hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})] &= \sigma_H^2(\mathbf{z}) - 2\mathbf{C}_{HX}^T(\mathbf{z})\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z}) + \mathbf{C}_{HX}^T(\mathbf{z})\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z}) \\ &= \sigma_H^2(\mathbf{z}) - \mathbf{C}_{HX}^T(\mathbf{z})\mathbf{C}_X^{-1}\mathbf{C}_{HX}(\mathbf{z}). \end{aligned} \quad (\text{C.38})$$

Można zauważyć, że drugi wyraz w wyrażeniu (C.38) równa się wariancji dyskretyzacji metodą OLE (C.37). Ponieważ wariancja z definicji przyjmuje jedynie wartości nieujemne, z równania (C.38) wynika zależność

$$\text{Var} [\hat{H}_{\text{OLE}}(\mathbf{z}, \mathbf{X})] \leq \sigma_H^2(\mathbf{z}), \quad (\text{C.39})$$

co oznacza, że dyskretyzacja metodą OLE prowadzi do pola losowego o większej regularności (mniejszej losowości) niż oryginalne pole losowe.

C.6. Niegaussowskie pola losowe

W większości praktycznych zadań analizy niezawodności konstrukcji otrzymywane wartości prawdopodobieństwa awarii są bardzo małe, a co za tym idzie bardzo wrażliwe na założenia dotyczące typów rozkładów prawdopodobieństwa zmiennych losowych, w szczególności kształtu tzw. ogonów funkcji gęstości prawdopodobieństwa. Z tego względu, jest niesłychanie istotne żeby wybrana metoda dyskretyzacji była w stanie jak najdokładniej oddać rzeczywisty kształt rozkładu prawdopodobieństwa zmiennych reprezentujących pole losowe, a nie jedynie ich dwóch pierwszych momentów statystycznych.

Przedstawione w tym dodatku metody dyskretyzacji spełniają ten warunek tylko w przypadku gaussowskich pól losowych. Z drugiej strony, mogą być one stosowane w przypadku niegaussowskich pól losowych H_{NG} , jeżeli istnieje następująca nieliniowa transformacja:

$$H_{NG}(\mathbf{z}) = \mathbf{T}[H(\mathbf{z})], \quad (\text{C.40})$$

gdzie H jest polem gaussowskim. H_{NG} można przybliżyć przez transformację dyskretyzacji pola losowego.

$$\hat{H}_{NG}(\mathbf{z}) = \mathbf{T}[\hat{H}(\mathbf{z})], \quad (\text{C.41})$$

Jednak, jak to podkreślono w doskonałej przeglądowej pracy Sudreta i Der Kiureghiana [245], w przypadku transformacji dyskretyzacji metodą OLE otrzymane przybliżenie może utracić własność minimalizacji błędu aproksymacji (C.27).

Do klasy transformacji, które wykorzystać można w równaniu (C.40) zaliczyć należy transformację Natafa, zob. [151, 183], która znajduje zastosowanie przy dyskretyzacji logarytmiczno-normalnych pól losowych. Pola te są szczególnie ważne w modelowaniu losowych własności materiału ze względu na przyjmowanie jedynie dodatnich realizacji.

C.7. Dobór wielkości elementów pola losowego

Niektóre z metod dyskretyzacji, np. MP, SA i SF, wymagają podziału obszaru Ω na elementy pola losowego. Ponadto, pomimo że do zbudowania dyskretyzacji metodą OLE wystarczy podać dowolny zbiór punktów $\{\mathbf{z}_1, \dots, \mathbf{z}_q\} \in \Omega$, to wybierane są najczęściej węzły regularnych siatek opisujących geometrię analizowanego zagadnienia.

Ponieważ w większości przypadków, które są istotne z punktu widzenia praktyki projektowej, do rozwiązania problemu mechanicznego wykorzystywana jest metoda elementów skończonych, ważne jest ze względu na prostotę implementacji, aby siatka elementów skończonych oraz siatka elementów pola losowego były w pewien sposób ze sobą związane. Z drugiej strony, należy pamiętać, że dobór wielkości tych dwóch rodzajów elementów wynika z innych przesłanek. Zazwyczaj nie powinno się stosować tej samej siatki do dyskretyzacji pola losowego w zagadnieniu analizy stochastycznej oraz dyskretyzacji pola przemieszczeń lub temperatur w zadaniu MES. Siatka elementów skończonych uwzględnia charakter pola naprężeń i zagęszczana jest w miejscach występowania dużych gradientów tych naprężeń, natomiast najważniejszą wielkością, która steruje doбором wymiaru elementów pola losowego, jest długość korelacji (C.12) funkcji kowariancji tego pola.

Jak wspomniano, w praktyce najczęściej stosuje się jednak siatki pola losowego powiązane z siatkami MES. Ułatwia to odwzorowanie realizacji tego pola na parametry elementów skończonych. Podstawową techniką jest tu “składanie” elementów pola losowego z kilku elementów MES. W zależności od zastosowanej metody dyskretyzacji, w literaturze można znaleźć szereg sugestii dotyczących doboru wielkości elementów pola losowego, za [245]:

- W przypadku metod uśredniania, na podstawie zadań dotyczących analizy belek oraz jednowymiarowych pól losowych w [161] zaleca się, aby odległości pomiędzy środkami ciężkości elementów były większe od tzw. skali fluktuacji (ang. *scale of fluctuation*) zdefiniowanej jako (zob. [255])

$$\theta = \int_{-\infty}^{\infty} \rho(\Delta z) d\Delta z. \quad (\text{C.42})$$

- Analizując belkę o losowej sztywności z trójkątną funkcją korelacji (C.9), w pracy [44] zaproponowano przyjęcie długości elementów losowych w granicach $a/4$ do $a/2$.
- Używanie zbyt gęstej siatki elementów pola losowego prowadzić może do bardzo wysokiej korelacji zmiennych losowych stosowanych w dyskretyzacji pola. Powodować to może problemy numeryczne, np. przy odwracaniu macierzy korelacji we wzorze (C.34). Zatem, konstruując dyskretyzację pola losowego powinno się unikać silnego skorelowania zmiennych stanowiących podstawę dyskretyzacji.

Tłumaczenia angielskich skrótów i terminów występujących w tekście

- AE-MDB** *Advanced European Mobile Deformable Barrier* – europejska zaawansowana ruchoma odkształcalna bariera.
- AMV** *Advanced Mean Value* – udoskonalona metoda wartości średniej, przybliżona metoda wyznaczania dystrybuanty funkcji losowej (odpowiedzi konstrukcji), a w szczególności jej kwantyli.
- API** *Application Programming Interface* – interfejs do programowania aplikacji.
- ARF** *Abdo-Rackwitz-Fiessler algorithm* – algorytm Abdo-Rackwitza-Fiesslera poszukiwania punktu projektowego.
- BFGS** *Broyden-Fletcher-Goldfarb-Shanno* – skrót od nazwisk twórców, quasi-newtonowska metoda aproksymacji hesjanu.
- BST** *Basic Shell Triangle* – podstawowy trójkątny element powłokowy.
- Control parameters** – parametry sterowalne, określenie używane w metodzie Taguchi’ego.
- CA** *Cluster Analysis* – analiza skupień.
- CP** *Columnwise-Parwise algorithm* – algorytm “kolumnami i parami”, metoda optymalizacji rozmieszczenia punktów łacińskiej hiperkostki.
- CPU** *Central Processing Unit* – procesor.
- DACE** *Design and Analysis of Computer Experiments* – metoda analizy eksperymentów komputerowych oparta na technice krigingu.
- DFSS** *Design For Six Sigma* – nowoczesna strategia projektowania procesów i konstrukcji gdzie główny nacisk kładziony jest na spełnienie wysokich wymagań dotyczących jakości.
- DRM** *Dimension Reduction Method* – metoda redukcji wymiarów, jedna z metod szacowania statystyk funkcji losowej.

- DRS** *Dual Response Surface method* – strategia aproksymacji statystyk, jedna ze strategii rozwiązania zadania optymalizacji odpornościowej.
- DS** *Descriptive Sampling* – próbkowanie opisowe lub zdeterminowane, do tej klasy można zaliczyć także metody symulacyjne wykorzystujące koncepcję łacińskiej hiperkostki.
- EPP** *Expanded PolyPropylene foam* – pianka polipropylenowa.
- ESS** *Explained Sum of Squares* – wyjaśniona suma kwadratów, pojęcie używane w liniowej analizie regresji.
- FORM** *First Order Reliability Method* – metoda analizy niezawodności pierwszego rzędu.
- GPU** *Graphics Processing Unit* – procesor kart graficznej.
- HMV** *Hybrid Mean Value algorithm* – hybrydowy algorytm wartości średniej, metoda rozwiązywania zadania odwrotnego analizy niezawodności w ramach podejścia PMA.
- iHLRF** *improved Hasofer-Lind-Rackwitz-Fiessler algorithm* – udoskonalony algorytm Hasofera-Linda-Rackwitza-Fiesslera poszukiwania punktu projektowego.
- Inner array** – wewnętrzna tablica ortogonalna, określenie używane w metodzie Taguchi’ego.
- IS** *Importance Sampling* – metody redukcji wariancji, rodzina symulacyjnych metod analizy niezawodności.
- LH** *Latin Hypercube* – łacińska hiperkostka.
- MLE** *Maximum Likelihood Estimation* – metoda największej wiarygodności, używana do wyznaczania współczynników równania krigingu.
- MLS** *Moving Least Squares method* – lokalna metoda najmniejszych kwadratów, metoda aproksymacji lokalnej oparta na metodzie ważonej regresji liniowej.
- MC** *Monte Carlo simulation* – podstawowa symulacyjna metoda Monte Carlo, zwana czasem klasyczną (*classical*) lub zgrubną (*crude*).
- MOGA** *Multi Objective Genetic Algorithm* – wielokryterialny algorytm genetyczny.
- MP** *Midpoint method* – metoda punktów środkowych, metoda dyskretyzacji pola losowego.
- MPP** *Most Probable Point* lub *Most Probable failure Point* – punkt projektowy zwany też punktem beta.

- MSE** *Mean Squared Error* – błąd średniokwadratowy.
- MVFO** *Mean Value First Order* – zgrubna metoda szacowania prawdopodobieństwa awarii polegająca na liniowej aproksymacji funkcji granicznej w wartościach oczekiwanych zmiennych losowych.
- Noise parameters** – parametry zakłócające, określenie używane w metodzie Taguchi’ego.
- Nugget effect** – efekt grudki złota, metoda umożliwiająca uzyskanie równań krigingu w wersji aproksymacyjnej.
- OLE** *Optimal Linear Estimation method* – metoda optymalnej aproksymacji liniowej, metoda dyskretyzacji pola losowego.
- OLH** *Optimal Latin Hypercube* – optymalna hiperkostka łacińska, łacińska hiperkostka o równomiernym rozmieszczeniu punktów, np. taka, która maksymalizuje minimalną odległość między punktami.
- Outer array** – zewnętrzna tablica ortogonalna, określenie używane w metodzie Taguchi’ego.
- Parallel coordinate plot** – wykres wieloosiowy do wizualizacji danych wielowymiarowych.
- PCA** *Principal Component Analysis* – analiza głównych składowych.
- PCE** *Polynomial Chaos Expansion* – metoda rozwinięcia funkcji losowej w chaos wielomianowy.
- PMA** *Performance Measure Approach* – metoda minimum funkcji granicznej, jedno z alternatywnych sformułowań zadania optymalizacji odpornościowej.
- RDO** *Robust Design Optimization* – optymalizacja odpornościowa konstrukcji.
- RBDO** *Reliability Based Design Optimization* – optymalizacja niezawodnościowa konstrukcji.
- Response surface** – powierzchnia odpowiedzi, metamodel (model modelu).
- RIA** *Reliability Index Approach* – metoda wskaźnika niezawodności, jedno z alternatywnych sformułowań zadania optymalizacji odpornościowej.
- RLH** *Random Latin Hypercube* – losowa hiperkostka łacińska, nazwa ma podkreślać, że układ punktów hiperkostki jest przypadkowy (choć zgodny z koncepcją LH) i nie podlegał żadnym modyfikacjom.
- RMSE** *Root Mean Squared Error* – pierwiastek błędu średniokwadratowego.
- Robust design** – inżynieria jakości, metodologia projektowania konstrukcji, urządzeń lub procesów produkcyjnych, w której, zachowując funkcjonalność projektowanego systemu, dąży się do znalezienia rozwiązania jak

najmniej wrażliwego (jak najbardziej odpornego) na zmiany wartości jego parametrów.

RSS *Residual Sum of Squares* – suma kwadratów reszt, pojęcie używane w liniowej analizie regresji.

SA *Spatial Average method* – metoda średniej przestrzennej, metoda dyskretyzacji pola losowego.

Scatter plot – wykres punktowy.

SF *Shape Function method* – metoda funkcji kształtu, metoda dyskretyzacji pola losowego.

SFEM *Stochastic Finite Element Method* – metoda stochastycznych elementów skończonych.

S/N *Signal to Noise ratio* – stosunek sygnału do zakłócenia, określenie używane w metodzie Taguchi’ego.

SORA *Sequential Optimization and Reliability Assessment* – metoda sekwencyjnej optymalizacji i analizy niezawodności, strategia realizacji zadania optymalizacji niezawodnościowej z grupy metod rozprzężonych.

SORM *Second Order Reliability Method* – metoda analizy niezawodności drugiego rzędu.

Space filling design – wypełniający plan eksperymentów (taki jak OLH albo ciąg Sobola), pseudolosowy eksperyment, który równomiernie wypełnia dany obszar.

SVD *Singular Value Decomposition method* – metoda dekompozycji na wartości singularne.

SVM *Support Vector Machine method* – metoda maszyny wektorów nośnych, metoda klasyfikacji.

Trust region – obszar dopuszczalny.

TSS *Total Sum of Squares* – całkowita suma kwadratów, pojęcie używane w liniowej analizie regresji.

Bibliografia

1. T. Abdo, R. Rackwitz. *A new beta-point algorithm for large time-invariant and time-variant reliability problems*. w: A. Der Kiureghian, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '90, Proc. 3rd WG 7.5 IFIP Conf., Berkeley, 26–28 March 1990*, 1–12, Berlin, 1991.
2. H. Agarwal. *Reliability based design optimization: formulations and methodologies*. Praca Dokt., University of Notre Dame, 2004.
3. C.A. Alonso, S.A. Collado. *Principal component analysis in data reduction – An application to structural mechanics*. w: I. Doltsinis (ed.), *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*. CIMNE, Barcelona, Spain, 1999.
4. Y. Aoues, A. Chateaufneuf. *Benchmark study of numerical methods for reliability-based design optimization*. *Structural and Multidisciplinary Optimization*, 41:277–294, 2010.
5. L. Armijo. *Minimization of functions having Lipschitz – continuous first partial derivatives*. *Pacific J. Math.* 16:1–3, 1966.
6. J.S. Arora. *Introduction to Optimum Design*. McGraw-Hill, 1989.
7. P. Audze, V. Eglais. *New approach to planning out of experiments*. w: *Problems of dynamics and strength*, volume 35, 104–107, 1977. (in Russian).
8. J.F.M. Barthelemy, R. Haftka. *Approximations concepts for optimal structural design - a review*. *Structural Optimization*, 5:129–144, 1993.
9. A. Basudhar, S. Missoum. *Adaptive explicit decision functions for probabilistic design and optimization using support vector machines*. *Computers and Structures*, 86:1904–1917, 2008.
10. A. Basudhar, S. Missoum, A. Harrison Sanchez. *Limit state function identification using Support Vector Machines for discontinuous responses and disjoint failure domains*. *Probabilistic Engineering Mechanics*, 23:1–11, 2008.
11. S.J. Bates, J. Siens, D.S. Langley. *Formulation of the Audze-Eglais uniform Latin hypercube design of experiments*. *Advances of Engineering Software*, 34:493–506, 2003.
12. Y. Ben-Haim, I. Elishakoff. *Convex Models of Uncertainty in Applied Mechanics*. Elsevier, Amsterdam, 1990.

13. H.-G. Beyer, B. Sendhoff. *Robust optimization - A comprehensive survey*. Computer Methods in Applied Mechanics and Engineering, 196:3190–3218, 2007.
14. A. Biegus. *Probabilistyczna analiza konstrukcji stalowych*. Wydawnictwo Naukowe PWN, 1999.
15. G. Blatman, B. Sudret. *An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis*. Probabilistic Engineering Mechanics, 25:183–197, 2010.
16. R. Bouix, P. Viot, J.-L. Lataillade. *Polypropylene foam behaviour under dynamic loadings: Strain rate, density and microstructure effects*. International Journal of Impact Engineering, 36:329–342, 2009.
17. V. Braibant, M. Bulik, M. Liefvendahl, S. Molinier, R. Stocki, C. Wauquiez. *Stochastic simulation of highly nonlinear dynamic systems using the M-XPLORE extension of the RADIOSS software*. w: *Workshop on Optimal Design of Materials and Structures*. Ecole Polytechnique, 2002. On CD.
18. C.E. Brenner, C. Bucher. *A contribution to the SFE-based reliability assessment of nonlinear structures under dynamic loading*. Probabilistic Engineering Mechanics, 10(4):265–273, 1995.
19. C.G. Bucher. *Adaptive sampling – an iterative fast Monte Carlo procedure*. Structural Safety, 5:119–126, 1988.
20. C.G. Bucher, U. Bourgund. *A fast and efficient response surface approach for structural reliability problems*. Structural Safety, 7:57–66, 1990.
21. M. Bulik, M. Liefvendahl, R. Stocki, C. Wauquiez. *Stochastic simulation for crashworthiness*. Advances in Engineering Software, 35:791–803, 2004.
22. C.J.C. Burges. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, 2(2):121–167, 1998.
23. F. Campos, P. Geremia, S. Weston, M. Islam. *Fast robust design optimisation of a vehicle cooling duct using CFD*. w: *Proc. of NAFEMS World Congress, Vancouver, 22nd May, 2007*.
24. N. Cappetti, A. Donnarumma, A. Naddeo, L. Russo. *Design of experiment about foam CF45 for pedestrian safety in car design*. Journal of Materials Processing Technology, 175:77–82, 2006.
25. J.M.A. César de Sá, A.D. Santos (eds.). *Materials Processing and Design: Modeling, Simulation and Applications NUMIFORM 2007 Proceedings of the 9th International Conference on Numerical Methods in Industrial Forming Processes*, volume 908 of *AIP Conference Proceedings, Materials Physics and Applications*. Springer, 2007.
26. P.B. Chang, B.J. Williams, T.J. Santner, W.I. Notz, Bartel D.L. *Robust optimization of total joint replacements incorporating environmental variables*. Journal of Biomechanical Engineering, 121:304–310, 1999.
27. S.H. Chen, J. Wu, Y.D. Chen. *Interval optimization for uncertain structures*. Finite Elements in Analysis and Design, 40:1379–1398, 2004.

28. W. Chen, J.K. Allen, K.L. Tsui, F. Mistree. *A procedure for robust design: minimizing variations caused by noise factors and control factors*. ASME Journal of Mechanical Design, 118:478–485, 1996.
29. W. Chen, W. Fu, S.B. Biggers, R.A. Latour. *An affordable approach for robust design of thick laminated composite structure*. Optimization and Engineering, 1:305–322, 2000.
30. W. Chen, A. Sahai, A. Messac, G.J. Sundararaj. *Exploration of the effectiveness of physical programming in robust design*. ASME Journal of Mechanical Design, 122:155–163, 2000.
31. J.-L. Chenot, F. Bay. *An overview of numerical modelling techniques*. Journal of Materials Processing Technology, 80-81:8–15, 1998.
32. J.H. Choi, W.H. Lee, J.J. Park, B.D. Youn. *A study on robust design optimization of layered plate bonding process considering uncertainties*. Structural and Multidisciplinary Optimization, 35:531–540, 2008.
33. K.K. Choi, B.D. Youn. *Hybrid analysis method for reliability-based design optimization*. w: *Proc. of ASME Design Engineering Technical Conference*, Pittsburgh, Pennsylvania, September 2001.
34. K.K. Choi, B.D. Youn, R.-J. Yang. *Moving least square method for reliability-based design optimization*. w: *Proc. of the 4th World Congress on Structural and Multidisciplinary Optimization*, Dalian, 2001.
35. Y. Collette, P. Siarry. *Multi-objective Optimization: Principles and Case Studies*. Springer, 2003.
36. N. Cressie. *The origins of kriging*. Mathematical Geology, 22:239–252, 1990.
37. M.A. Crisfield. *Non-linear Finite Element Analysis of Solids and Structures, Volume 1: Essentials*. Wiley, 1991.
38. F. Delcroix, T. Bekkour, R. Stocki. *Report on robust optimization and ADVISER extensions*. Report no. D734A, EC FP6 Project APROSYS, 2008.
39. J. Deng, D. Gu, X Li, Z.Q. Yue. *Structural reliability analysis for implicit performance functions using artificial neural network*. Structural Safety, 27:25–48, 2005.
40. A. Der Kiureghian. *Safety Engineering in Structural Design*. Lecture notes for seminar on safety engineering, Press Center Hall, Tokyo, Japan, July 11-12, 1989.
41. A. Der Kiureghian, T. Haukaas, K. Fujimura. *Structural reliability software at the University of California, Berkeley*. Structural Safety, 28:44–67, 2006.
42. A. Der Kiureghian, J.-B. Ke. *Finite-element based reliability analysis of frame structures*. w: *Proc. 4th International Conference on Structural Safety and Reliability, Kobe, Japan, Part I*, 395–404, 1985.
43. A. Der Kiureghian, J.-B. Ke. *The stochastic finite element method in structural reliability*. Journal of Engineering Mechanics, ASCE, 112:85–104, 1986.
44. A. Der Kiureghian, J.-B. Ke. *The stochastic finite element method in structural reliability*. Probabilistic Engineering Mechanics, 3(2):83–91, 1988.

45. A. Der Kiureghian, P.-L. Liu. *Structural reliability under incomplete probability information*. Journal of Engineering Mechanics, ASCE, 112:85–104, 1986.
46. M. Diez, F. Delcroix, R. Puppini, H. Cappon, R. Meijer, M. Avalle. *Sources of dispersion and ADVISER templates for stochastic analysis*. Report no. D721, EC FP6 Project APROSYS, 2006.
47. M. Diez, R. Pupini, F. Delcroix, R. Stocki, I. Ciglaric. *Complete demonstrators in various numerical codes along with procedural report for virtual testing implementation in pedestrian impacts*. Report no. D742B, EC FP6 Project APROSYS, 2009.
48. O. Ditlevsen, H.O. Madsen. *Structural Reliability Methods*. Wiley, 1996.
49. Frangopol D.M., Y.-H. Lee, K. Willam. *Nonlinear finite element reliability analysis of concrete*. Journal of Engineering Mechanics, 122:1174–1182, 1996.
50. K. Doliński. *First-order second-moment approximation in reliability of structural systems: critical review and alternative approach*. Structural Safety, 1:211–231, 1983.
51. K. Doliński. *Importance sampling techniques in reliability calculations*. Prace IPPT, 37, 1988.
52. K. Doliński, R. Stocki. *Reliability-based structural optimization accounting for manufacturing and material quality*. Engineering Transactions, 4:623–636, 2001.
53. I. Doltsinis. *Stochastic Analysis of Multivariate Systems in Computational Mechanics and Engineering*. CIMNE, Barcelona, Spain, 1999.
54. I. Doltsinis. *Design optimization and robustness of structures against uncertainties based on Taylor series expansion*. w: Y. Tsompanakis, N.D. Lagaros, M. Papadrakakis (eds.), *Structural Design Optimization Considering Uncertainties*, Structures and Infrastructures Series. Taylor and Francis, 2007.
55. I. Doltsinis, Z. Kang. *Robust design of structures using optimization methods*. Computer Methods in Applied Mechanics and Engineering, 193:2221–2237, 2004.
56. I. Doltsinis, Z. Kang. *Robust design of non-linear structures using optimization methods*. Computer Methods in Applied Mechanics and Engineering, 194:1179–1195, 2005.
57. S. Donders, M. Brughmans, L. Hermans, C. Liefoghe, W. Van der Auweraer, H. Desmet. *The robustness of dynamic vehicle performance to spot weld failures*. Finite Elements in Analysis and Design, 42:670–682, 2006.
58. N. Draper, H. Smith. *Applied Regression Analysis*. Wiley, New York, 1998.
59. X. Du, W. Chen. *Sequential optimization and reliability assessment method for efficient probabilistic design*. w: *Proc. of ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada, September 29–October 2, 2002.
60. X. Du, A. Sudjianto, W. Chen. *An integrated framework for optimization using inverse reliability strategy*. Journal of ASME Journal of Mechanical Design, 126(4):562–570, 2004.

61. E.N. Dvorkin, K.J. Bathe. *A continuum mechanics based four-node shell element for general nonlinear analysis*. Engineering Computations, 1:77–88, 1984.
62. K. El-Tawil, M. Lemaire, J.-P. Muzeau. *Reliability method to solve mechanical problems with implicit limit states*. w: R. Rackwitz, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11–13 September 1991*, 181–190. Springer-Verlag, 1992.
63. J. Ellway, W. Donaldson, M. Edwards, T. Versmissen, M. Bosch-Rekveltdt. *Development and evaluation of the advanced European mobile deformable barrier (AE-MDB) Test procedure*. Report no. D1.1.1B, EC FP6 Project APROSYS, 2006.
64. I. Enevoldsen. *Reliability-based optimization as an information tool*. Mechanics of Structures and Machines, 22:117–135, 1994.
65. I. Enevoldsen, J.D. Sørensen. *Strategies for optimal design of structural systems*. w: R. Rackwitz, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11–13 September 1991*, 191–206. Springer-Verlag, 1992.
66. I. Enevoldsen, J.D. Sørensen. *Reliability-based optimization of series systems of parallel systems*. Journal of Structural Engineering, ASCE, 119:1069–1084, 1993.
67. I. Enevoldsen, J.D. Sørensen. *Decomposition techniques and effective algorithms in reliability-based optimization*. w: P.D. Spanos (ed.), *Computational Stochastic Mechanics*, 149–156. Balkema, 1995.
68. L. Faravelli. *Response surface approach for reliability analysis*. Journal of Engineering Mechanics, ASCE, 115:2763–2781, 1989.
69. P. Forys. *Nowy algorytm optymalizacji rojem cząstek i jego zastosowanie w kształtowaniu elementów konstrukcji*. Praca Dokt., Politechnika Krakowska, Instytut Mechaniki Stosowanej, www.optalg.com, 2007.
70. S.E. Gano, J.E. Renaud, J.D. Martin, T.W. Simpson. *Update strategies for kriging models used in variable fidelity optimization*. Structural and Multidisciplinary Optimization, 32:287–298, 2006.
71. M. Gasser, G.I. Schuëller. *Reliability-based optimization of structural systems*. Mathematical Methods of Operations Research, 46:287–307, 1997.
72. M. Gasser, G.I. Schuëller. *Some basic principles of reliability-based optimization (RBO) of structures and mechanical components*. w: K. Marti, P. Kall (eds.), *Stochastic Programming Methods and Technical Applications, Lecture Notes in Economics and Mathematical Systems 458*, 80–103. Springer Verlag, 1998.
73. R.G. Ghanem, P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, 1991.
74. R. Gijs. *Subset selection from large datasets for Kriging modeling*. Structural and Multidisciplinary Optimization, 38:545–569, 2009.
75. P.E. Gill, W. Murray, M.H. Wright. *Practical Optimization*. Academic Press, 1981.

76. A.S. Giunta, L.T. Watson. *A comparison of approximation modeling techniques: polynomial versus interpolating models.* w: *Proc. of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 392–404, St. Louis, MO, 1998.
77. W. L. Goffe, G. D. Ferrier, J. Rogers. *Global optimization of statistical functions with simulated annealing.* *Journal of Econometrics*, 60(1–2):65–100, 1994.
78. D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, 1989.
79. S. Gollwitzer, B. Kirchgässner, R. Fischer, R. Rackwitz. *PERMAS-RA/STRUREL system of programs for probabilistic reliability analysis.* *Structural Safety*, 28:108–129, 2006.
80. H.S. Gopalakrishna, E. Donaldson. *Practical reliability analysis using a general purpose finite element program.* *Finite Elements in Analysis and Design*, 10:75–87, 1991.
81. W.H. Green. *Econometric Analysis.* Prentice-Hall, 1997.
82. L. Gu, R.J. Yang. *Recent applications on reliability-based optimization of automotive structures.* w: *Reliability & Robust Design in Automotive Engineering*, 77–88. SAE, 2003.
83. M.A. Gutiérrez, J. Carmeliet, R. De Borst. *Finite element reliability methods using DIANA.* w: G.M.A. Kusters, M.A.N. Hendriks (eds.), *Proc. of the First International DIANA Conference on Computational Mechanics, Delft, The Netherlands, 24–25 October 1994*, 255–263. Kluwer Academic Publishers, 1994.
84. R.T. Haftka, Z. Gürdal. *Elements of Structural Optimization.* Kluwer, 1992.
85. E. Hansen, G.W. Walster. *Global Optimization Using Interval Analysis.* Dekker, New York, NY, 2003.
86. A.M. Hasofer, N.C. Lind. *Exact and invariant second moment code format.* *Journal of the Engineering Mechanics Division, ASCE*, 100:111–121, 1974.
87. T. Haukaas. *Finite Element Reliability and Sensitivity Methods for Performance-Based Engineering.* Praca Dokt., University of California, Berkeley, 2003.
88. S.S. Hecker. *A simple FLD technique and results on aluminium alloys.* w: *7th Congress of IDDRG*, Amsterdam, 1972.
89. J.C. Helton, F.J. Davis. *Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems.* *Reliability Engineering and System Safety*, 81:23–69, 2003.
90. R. Hill. *A theory of the yielding and plastic flow of anisotropic metals.* *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 193:281–297, 1948.
91. R. Hill. *Theoretical plasticity of textured aggregates.* *Mathematical Proceedings of the Cambridge Philosophical Society*, 85:179–191, 1979.
92. R. Hill. *Constitutive modelling of orthotropic plasticity in sheet metals.* *Journal of the Mechanics and Physics of Solids*, 38:405–417, 1990.

93. M. Hohenbichler, R. Rackwitz. *Non-normal dependent vectors in structural safety*. Journal of the Engineering Mechanics Division, ASCE, 107:1227–1238, 1981.
94. M. Hohenbichler, R. Rackwitz. *Improvement of second-order reliability estimates by importance sampling*. Journal of the Engineering Mechanics, ASCE, 114:2195–2199, 1988.
95. Pavel Hora (ed.). *Numisheet 2008, 7th International Conference and Workshop on Numerical Simulation of 3D Sheet Metal Forming Processes*, Interlaken, Switzerland, September 1–5, 2008.
96. N. Hoshino, A. Takemura. *On reduction of finite-sample variance by extended Latin hypercube sampling*. Bernoulli, 6:1035–1050, 2000.
97. B. Hu, X. Du. *Analytical robustness assessment for robust design*. Structural and Multidisciplinary Optimization, 34:123–137, 2007.
98. H.-C. Huang. *Static and Dynamic Analyses of Plates and Shells: Theory, Software and Applications*. Springer-Verlag, 1989.
99. C.G. Huijskens et al. *Common road map VT in regulations & Feasibility report of VT in a selected regulated or consumer testing procedure*. Report no. D7.4.1 & D7.4.3, EC FP6 Project APROSYS, 2008.
100. D.E. Huntington, C.S. Lyrantzis. *Improvements to and limitations of Latin hypercube sampling*. Probabilistic Engineering Mechanics, 13:245–253, 1998.
101. J.E. Hurtado. *An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory*. Structural Safety, 26:271–293, 2004.
102. J.E. Hurtado. *Structural robustness and its relationship to reliability*. w: Y. Tsompanakis, N.D. Lagaros, M. Papadrakakis (eds.), *Structural Design Optimization Considering Uncertainties*, Structures and Infrastructures Series. Taylor and Francis, 2007.
103. J.E. Hurtado, D.A. Alvarez. *Neural-network-based reliability analysis: a comparative study*. Computer Methods in Applied Mechanics and Engineering, 191:113–132, 2001.
104. K.-H. Hwang, K.-W. Lee, G.-J. Park. *Robust optimization of an automobile rearview mirror for vibration reduction*. Structural and Multidisciplinary Optimization, 21:300–308, 2001.
105. R.L. Iman, W.J. Conover. *A distribution-free approach to inducing rank correlation among input variables*. Communications in Statistics, 11:311–334, 1982.
106. T. Jansson, L. Nilsson, R. Moshfegh. *Reliability analysis of a sheet metal forming process using Monte Carlo analysis and metamodels*. Journal of Materials Processing Technology, 202:255–268, 2008.
107. S. Jendo, J. Putresza. *Multicriteria optimization of elasto-plastic bar structures*. w: R. Rackwitz, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11–13 September 1991*, 240–249. Springer-Verlag, 1992.

108. C. Jianjiang, X. Renbin, Z. Yifang. *A response surface based hierarchical approach to multidisciplinary robust optimization design*. The International Journal of Advanced Manufacturing Technology, 26:301–309, 2005.
109. R. Jin, W. Chen, T. Simpson. *Comparative studies of metamodeling techniques under multiple modeling criteria*. Structural and Multidisciplinary Optimization, 23:1–13, 2001.
110. R. Jin, X. Du, W. Chen. *The use of metamodeling techniques for optimization under uncertainty*. Structural and Multidisciplinary Optimization, 25:99–116, 2003.
111. G. Johnson, W. Cook. *A constitutive model and data for metals subjected to large strains, high strain rates and high temperatures*. w: *7th Symposium on Ballistics, The Hague*, 1983.
112. D.R. Jones, M. Schonlau, W.J. Welch. *Efficient global optimization of expensive black-box functions*. Journal of Global Optimization, 13:455–492, 1998.
113. F. Jurecka. *Robust Design Optimization Based on Metamodeling Techniques*. Shaker Verlag, Aachen, 2007.
114. F. Jurecka, M. Ganser, Bletzinger. K.-U. *Update scheme for sequential spatial correlation approximations in robust design optimization*. Computers and Structures, 85:606–614, 2007.
115. Z. Kang. *Robust Design Optimization of Structures under Uncertainty*. Praca Dokt., Institut für Statik und Dynamik der Luft- und Raumfahrtkonstruktionen Universität Stuttgart, 2005.
116. A. Karamchandani, P. Bjerager, C.A. Cornell. *Adaptive importance sampling*. w: A.H.-S. Ang, M. Shinozuka, G.I. Schuëller (eds.), *Proceedings 5th International Conference on Structural Safety and Reliability, San Francisco*. ASCE, 1989.
117. I. Kaymaz, C.A. McMahon. *A response surface method based on weighted regression for structural reliability analysis*. Probabilistic Engineering Mechanics, 20:11–17, 2005.
118. E. Kazancioglu, G. Wu, J. Ko, S. Bohac, Z. Filipi, S.J. Hu, D. Assanis, K. Saitou. *Robust optimization of an automobile valvetrain using a multiobjective genetic algorithm*. w: *Proc. of DETC'03 ASME 2003 Design Engineering Technical Conferences*, Chicago, Illinois, September 2-6, 2003. DETC2003/DAC-48714.
119. M. Keramat, R. Kielbasa. *Modified Latin hypercube sampling Monte Carlo (MLH-SMC) estimation for average quality index*. Analog Integrated Circuits and Signal Processing, 19:87–98, 1999.
120. A. Kielbus. *Wielokryterialna ocena jakości wyrobów przemysłu maszynowego*. Praca Dokt., Politechnika Krakowska, Wydział Mechaniczny, 2007.
121. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. *Optimization by Simulated Annealing*. Science, New Series, 220(4598):671–680, 1983.
122. M. Kleiber, H. Antúnez, T.D. Hien, P. Kowalczyk. *Parameter Sensitivity in Non-linear Mechanics; Theory and Finite Element Computations*. Wiley, 1997.

123. M. Kleiber, T.D. Hien. *The Stochastic Finite Element Method*. Wiley, 1992.
124. M. Kleiber, J. Knabel, J. Rojek. *Response surface method for probabilistic assessment of metal forming failures*. International Journal for Numerical Methods in Engineering, 60:51–67, 2004.
125. M. Kleiber, J. Rojek, R. Stocki. *Reliability assessment of sheet metal forming operations*. Computer Methods in Applied Mechanics and Engineering, 191:4511–4532, 2002.
126. M. Kleiber, A. Siemaszko, R. Stocki. *Interactive stability-oriented reliability-based design optimization*. Computer Methods in Applied Mechanics and Engineering, 168(1–4):243–253, 1999.
127. W. Knabe, J. Przewłócki, G. Różyński. *Spatial averages for linear elements for two-parameter random fields*. Probabilistic Engineering Mechanics, 13(3):147–167, 1998.
128. J. Knabel. *Analiza niezawodności konstrukcji sprężysto - plastycznych przy użyciu powierzchni odpowiedzi*. Praca Dokt., IPPT PAN, 2004.
129. P.N. Koch. *Probabilistic design: optimization for six sigma quality*. w: *Proc. of 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, number AIAA-2002-1471, Denver, Colorado, USA, 22-25 April, 2002.
130. P.N. Koch, J.P. Evans, D. Powell. *Interdigitation for effective design space exploration using iSIGHT*. Structural and Multidisciplinary Optimization, 23:111–126, 2002.
131. P.N. Koch, S. Kodiyalam. *Variable complexity structural reliability analysis for efficient reliability-based design optimization*. w: *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, 40th, St. Louis, MO, April 12-15, 1999, Collection of Technical Papers*, volume 1, 1999.
132. P.N. Koch, B. Wujek, O. Golovidov. *A multi-stage, parallel implementation of probabilistic design optimization in an MDO framework*. w: *Proc. of the 8-th AIAA/USAF/NAS/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, USA, 6-8 Sept., 2000.
133. P.N. Koch, R.-J. Yang, L. Gu. *Design for six sigma through robust optimization*. Structural and Multidisciplinary Optimization, 26:235–248, 2004.
134. N. Kuschel, R. Rackwitz. *Two basic problems in reliability-based structural optimization*. Mathematical Methods of Operations Research, 46(3):309–333, 1997.
135. N. Kuschel, R. Rackwitz. *Optimal design under time-variant reliability constraints*. Structural Safety, 22(2):113–128, 2000.
136. K. Lange (ed.). *Handbook of Metal Forming*. McGraw-Hill, 1985.
137. I. Lee, K.K. Choi, L. Du, D. Gorsich. *Dimension reduction method for reliability-based robust design optimization*. Computers and Structures, 86:1550–1562, 2008.
138. K.-H. Lee, G.-J. Park. *Robust optimization considering tolerances of design variables*. Computers and Structures, 79:77–86, 2001.

139. K.-H. Lee, G.-J. Park, W.-S. Joo. *A global robust optimization using the kriging based approximation model*. w: *Proc. of the 6th World Congresses on Structural and Multidisciplinary Optimization*, Rio de Janeiro, Brazil, 30 May - 3 June, 2005.
140. S.H. Lee, W. Chen. *A comparative study of uncertainty propagation methods for black-box-type problems*. *Structural and Multidisciplinary Optimization*, 37:239–253, 2009.
141. S.H. Lee, W. Chen, B.M. Kwak. *Robust design with arbitrary distributions using Gauss-type quadrature formula*. *Structural and Multidisciplinary Optimization*, 39:227–243, 2009.
142. S.H. Lee, H.S. Choi, B.M. Kwak. *Multilevel design of experiments for statistical moment and probability calculation*. *Structural and Multidisciplinary Optimization*, 37:57–70, 2008.
143. M. Lemaire, A. Chateaneuf, J-C. Mitteau. *Structural Reliability*. Wiley, 2009.
144. M. Lemaire, A. Mohamed. *Finite element and reliability : A happy marriage? w: A.S. Nowak, M.M. Szerszen (eds.), Proc. 9th IFIP WG 7.5 Working Conference on Reliability and Optimization of Structural Systems, University of Michigan, Ann Arbor, MI, USA Sept. 25-27, 2000*, 3–14, 2000.
145. M. Lemaire, O. Mohamed, A. Florès Macias. *The use of finite element codes for the reliability of structural systems*. w: D.M. Frangopol, R.B. Corotis, R. Rackwitz (eds.), *Proc. 7th IFIP WG 7.5 Working Conference on Reliability and Optimization of Structural Systems, Boulder, CO, USA 2-4 April, 1996*, 223–230. Pergamon, 1997.
146. M. Lemaire, M. Pendola. *Phimeca-Soft*. *Structural Safety*, 28:130–149, 2006.
147. C.-C. Li, A. Der Kiureghian. *Optimal discretization of random fields*. *Journal of Engineering Mechanics ASCE*, 119(6):1136–1154, 1993.
148. H. Li, Z. Lü, Z. Yue. *Support vector machine for structural reliability analysis*. *Applied Mathematics and Mechanics*, 27(10):1295–1303, 2006.
149. Y.Q. Li, Z.S. Cui, X.Y. Ruan, D.J. Zhang. *CAE-based six sigma robust optimization for deep-drawing sheet metal process*. *The International Journal of Advanced Manufacturing Technology*, 30:631–637, 2006.
150. M. Liefvendahl, R. Stocki. *A study on algorithms for optimization of Latin hypercubes*. *Journal of Statistical Planning and Inference*, 136(9):3231–3247, 2006.
151. P.-L. Liu, A. Der Kiureghian. *Multivariate distribution models with prescribed marginals and covariances*. *Probabilistic Engineering Mechanics*, 1(2):105–112, 1986.
152. P.-L. Liu, A. Der Kiureghian. *Optimization algorithms for structural reliability analysis*. Report No. UCB/SEMM-86/09, Department of Civil Engineering, University of California, Berkeley, CA, 1986.
153. P.-L. Liu, A. Der Kiureghian. *Optimization algorithms for structural reliability*. *Structural Safety*, 9:161–177, 1991.

154. P.-L. Liu, A. Der Kiureghian. *Finite element reliability of geometrically nonlinear uncertain structures*. Journal of Engineering Mechanics, ASCE, 117:1806–1825, 1992.
155. W.K. Liu, T. Belytschko, A. Mani. *Probabilistic finite elements for non linear structural dynamics*. Computer Methods in Applied Mechanics and Engineering, 56:61–86, 1986.
156. W.K. Liu, T. Belytschko, A. Mani. *Random field finite elements*. International Journal for Numerical Methods in Engineering, 26(10):1831–1845, 1986.
157. S.N. Lophaven, H.B. Nielsen, J. Søndergaard. *Aspects of Matlab toolbox DACE*. Technical Report IMM-REP-2002-13, Technical University of Denmark, 2002.
158. S.N. Lophaven, H.B. Nielsen, J. Søndergaard. *DACE a Matlab kriging toolbox. Version 2.0*. Technical Report IMM-REP-2002-12, Technical University of Denmark, 2002.
159. H.O. Madsen, P. Friis Hansen. *A comparison of some algorithms for reliability based structural optimization and sensitivity analysis*. w: R. Rackwitz, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11–13 September 1991*, 443–451. Springer-Verlag, 1992.
160. H.O. Madsen, S. Krenk, N.C. Lind. *Methods of Structural Safety*. Prentice-Hall, 1986.
161. S. Mahadevan, A. Haldar. *Practical random field discretization in stochastic finite element analysis*. Structural Safety, 9(4):283–304, 1991.
162. A. Makinouchi, E. Nakamachi, E. Oñate, R.H. Wagoner (eds.). *Proceedings of the 2nd International Conference NUMISHEET 93: Numerical Simulation of 3-D Sheet Metal Forming Processes – Verification of Simulation with Experiment*. Isehara, Japan, 1993.
163. Z. Marciniak, K. Kuczyński. *Limit strains in the processes of stretch-forming sheet metal*. International Journal of Mechanical Sciences, 9:609–621, 1967.
164. J. Marczyk. *Principles of Simulation-Based Computer-Aided Engineering*. FIM Publications, Madrid, 1999.
165. J. Marczyk. *Stochastic multidisciplinary improvement: beyond optimization*. w: *Proc. 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, September, 2000*.
166. W. Marks, S. Owczarek [pod red.]. *Optymalizacja Wielokryterialna Budynków Energooszczędnych*. Number 46. Studia z Zakresu Inżynierii, IPPT, KILW PAN, Warszawa, 1999.
167. R.T. Marler, J.S. Arora. *Survey of multi-objective optimization methods for engineering*. Structural and Multidisciplinary Optimization, 26:369–395, 2004.
168. G. Matheron. *Principles of geostatistics*. Economic Geology, 58:1246–1266, 1963.

169. S. Mazzoni, F. McKenna, G.L. Fenves. *Open System for Earthquake Engineering Simulation User Manual*. Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA, <http://opensees.berkeley.edu>.
170. M.D. McKay, R.J. Beckman, W.J. Conover. *A comparison of three methods for selecting values of input variables from a computer code*. *Technometrics*, 21:239–245, 1979.
171. Mecalog SARL, 2 Rue de la Renaissance 92160 Antony, France. *RADIOSS Input Manual, Version 4.2*, 2000.
172. R.E. Melchers. *Simulation in time-invariant and time-variant reliability problems*. w: R. Rackwitz, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11–13 September 1991*, 39–82. Springer-Verlag, 1992.
173. R.E. Melchers. *Structural Reliability Analysis and Predictions, 2nd Ed.* Wiley, 1999.
174. S. Missoum, P. Ramu, R.T. Haftka. *A convex hull approach for the reliability-based design optimization of nonlinear transient dynamic problems*. *Computer Methods in Applied Mechanics and Engineering*, 196:2895–2906, 2007.
175. Z. Mourelatos, J. Liang. *An efficient unified approach for reliability and robustness in engineering design*. w: *Proc. NSF Workshop on Reliable Engineering Computing, Savannah, Georgia*, 2004.
176. J.M. Murzewski. *Niezawodność konstrukcji inżynierskich*. Arkady, 1989.
177. J.M. Murzewski. *Podstawy projektowania i niezawodności konstrukcji: podręcznik dla studentów wyższych szkół technicznych*. Kraków: Politechnika Krakowska, 1999.
178. J. Mycielski. *Materiały do nauki ekonometrii*. <http://inflacja.net/jmyc/>.
179. R.H. Myers, D.C. Montgomery. *Response Surface Methodology, Process and Product Optimization Using Designed Experiments, Second Edition*. Wiley, 2002.
180. V.N. Nair. *Taguchi's parameter design: a panel discussion*. *Technometrics*, 34:127–161, 1992.
181. E. Nakamachi. *Sheet forming process characterization by static-explicit anisotropic elastic-plastic finite element simulation*. w: A. Makinouchi, E. Nakamachi, E. Oñate, R.H. Wagoner (eds.), *Proceedings of the 2nd International Conference NUMISHEET 93: Numerical Simulation of 3-D Sheet Metal Forming Processes – Verification of Simulation with Experiment*, Isehara, Japan, 1993.
182. K. Nakazima, T. Kikuma, K. Hasuka. *Study on the formability of steel sheets*. Yawata Technical Report, 264:141–154, 1968.
183. A. Nataf. *Determination des distribution dont les marges sont donnees*. *Comptes Rendus de l'Academie des Sciences*, 225:42–43, 1962.
184. J.A. Nedler, R. Mead. *Simplex method for function minimization*. *Computer Journal*, 7:308–313, 1965.

185. E. Nikolaidis, R. Burdisso. *Reliability based optimization: a safety index approach*. Computers and Structures, 28(6):781–788, 1988.
186. A. Olsson, G. Sandberg, O. Dahlblom. *On Latin hypercube sampling for structural reliability analysis*. Structural Safety, 25:47–68, 2003.
187. A. Osyczka. *Multicriterion Optimization in Engineering*. Ellis Horwood, 1984.
188. A. Osyczka. *Evolutionary Algorithms for Single and Multicriteria Design Optimization*. Phisica-Verlag Heidelberg, 2003.
189. E. Oñate, P. Cendoya, J. Rojek, J. Miquel. *A simple thin shell triangle with translational degrees of freedom only*. w: S.N. Atluri, G. Yagawa (eds.), *Advances in Computational Engineering Science*, 308–313. Tech Science Press, USA, 1997.
190. E. Oñate, J. Rojek, C. García Garino. *NUMISTAMP: a research project for assessment of finite element models for stamping processes*. Journal of Materials Processing Technology, 50(1-4):17–38, 1995.
191. M. Papadrakakis, N.D. Lagaros, V. Plevris. *Design optimization of steel structures considering uncertainties*. Engineering Structures, 27:1408–1418, 2005.
192. G.-J. Park. *Analytic Methods for Design Practise*. Springer-Verlag London Limited, 2007.
193. J.-S. Park. *Optimal Latin-hypercube designs for computer experiments*. Journal of Statistical Planning and Inference, 39:95–111, 1994.
194. M.F. Pellissetti, G.I. Schueller. *On general purpose software in structural reliability – An overview*. Structural Safety, 28:3–16, 2006.
195. A. Plucińska, E. Pluciński. *Probabilistyka. Rachunek prawdopodobieństwa, statystyka matematyczna, procesy stochastyczne*. Wydawnictwa Naukowo-Techniczne, 2000.
196. C. Poloni, A. Clarich. *Mode frontier framework and its uncertainty capabilities in aeronautics*. w: *Proc. of 8th World Congress on Computational Mechanics (WCCM8), 5th. European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2008), June 30 - July 5, 2008, Venice, Italy*, 2008.
197. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
198. R. Puppini, M. Diez, A. Ibba, M. Avalle, I. Ciglaric, F. Feist. *Generic car (FE) models for categories super minis, small family cars, large family/executive cars, MPV, heavy vehicle*. Report no. D714A, EC FP6 Project APROSYS, 2006.
199. J. Putresza. *Optymalizacja konstrukcji z uwzględnieniem jej niezawodności*. Praca Dokt., IPPT PAN, 1992.
200. R. Rackwitz, B. Fiessler. *Structural reliability under combined random load sequences*. Computers and Structures, 9:484–494, 1978.
201. S. Rahman, H. Xu. *A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics*. Probabilistic Engineering Mechanics, 19:393–408, 2004.

202. M.R. Rajashekhar, B.R. Ellingwood. *A new look at the response surface approach for reliability analysis*. Structural Safety, 12:205–220, 1993.
203. S. Rangavajhala, A. Mullur, A. Messac. *The challenge of equality constraints in robust design optimization: examination and new approach*. Structural and Multidisciplinary Optimization, 34:381–401, 2007.
204. S. Reh, J.D. Beley, S. Mukherjee, E.H. Khor. *Probabilistic finite element analysis using ANSYS*. Structural Safety, 28:17–43, 2006.
205. D.S. Riha, H.R. Millwater, B.H. Thacker. *Probabilistic structural analysis using a general purpose finite element program*. Finite Elements in Analysis and Design, 11:201–211, 1992.
206. J. Rojek. *Modelowanie i symulacja komputerowa złożonych zagadnień mechaniki nieliniowej metodami elementów skończonych i dyskretnych (rozprawa habilitacyjna)*. Prace IPPT, 4, 2007.
207. J. Rojek, M. Kleiber, A. Piela, R. Stocki, J. Knabel. *Deterministic and stochastic analysis of failure in sheet metal forming operations*. Steel Grips - Journal of Steel and Related Materials, 2:29–34, 2004.
208. J. Rojek, E. Oñate, E. Postek. *Application of explicit FE codes to simulation of sheet and bulk metal forming processes*. Journal of Materials Processing Technology, 80-81:620–627, 1998.
209. M. Rosenblatt. *Remarks on multivariate transformation*. The Annals of Mathematical Statistics, 23:470–472, 1952.
210. J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn. *Design and analysis of computer experiments*. Statistical Science, 4(4):409–435, 1989.
211. E. Saliby. *Descriptive sampling: an improvement over Latin hypercube sampling*. w: S. Andradóttir, K.J. Healy, D.H. Withers, B.L. Nelson (eds.), *Proceedings of the Winter Simulation Conference*, 230–233, 1997.
212. E. Sandgren, T.M. Cameron. *Robust design optimization of structures through consideration of variation*. Computers and Structures, 80:1605–1613, 2002.
213. M.J. Sasena, M. Parkinson, M.P. Reed, P.Y. Papalambros, P. Goovaerts. *Improving an ergonomic testing procedure via approximation-based adaptive experimental design*. ASME Journal of Mechanical Design, 127:1006–1013, 2005.
214. K. Schittkowski. *The nonlinear programming method of Wilson, Han and Powell with an augmented lagrangian type line search function, Part 1: Convergence analysis, Part 2: An efficient implementation with linear least squares subproblems*. Numerische Mathematik, 38:83–127, 1981.
215. K. Schittkowski. *User's guide for the nonlinear programming code NLPQL*. Handbook to optimization program package NLPQL, University of Stuttgart - Institute for Informatics, Germany, 1985.
216. K. Schittkowski, C. Zillober, R. Zotemantel. *Numerical comparison of nonlinear programming algorithms for structural optimization*. Structural Optimization, 7:1–19, 1994.

217. G.I. Schuëller, R. Stix. *A critical appraisal of methods to determine failure probabilities*. Structural Safety, 4:293–309, 1987.
218. K. Shimoyama, A. Oyama, K. Fujii. *A new efficient and useful robust optimization approach - design for multi-objective six sigma*. w: *The 2005 IEEE Congress on Evolutionary Computation*, volume 1, 950–957, 2005.
219. T.W. Simpson. *A concept exploration method for product family design*. Praca Dokt., Georgia Institute of Technology, 1998.
220. T.W. Simpson, A.J. Booker, D. Ghosh, A.A. Giunta, P.N. Koch, R.-J. Yang. *Approximation methods in multidisciplinary analysis and optimization: a panel discussion*. Structural and Multidisciplinary Optimization, 27:302–313, 2004.
221. T.W. Simpson, T.M. Mauery, J.J. Korte, F. Mistree. *Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization*. AIAA Journal, 39(12):2233–2241, 2001.
222. T.W. Simpson, J.D. Peplinski, P.N. Koch, J.K. Allen. *Metamodels for computer-based engineering design: survey and recommendations*. Engineering with computers, 17:129–150, 2001.
223. P. Śniady. *Random vibrations and reliability of structures under various loads*. w: *Proc. of OMAE1996, Safety and Reliability, ASME, vol.II*, 143–150, Florence, 1996.
224. P. Śniady, R. Sieniawska, S. Żukowski. *Reliability of bridge beams with hybrid cross-sections*. Archives of Civil and Mechanical Engineering, 3:13–23, 2003.
225. J.D. Sørensen, P. Thoft-Christensen, A. Siemaszko, J.M.B. Cardoso, J.L.T. Santos. *Interactive reliability-based optimal design*. w: *Reliability and Optimization of Structural Systems, Proc. 6th IFIP WG 7.5 Conf., Assisi, Sept. 7–9, 1994*, 243–253. Chapman & Hall, 1995.
226. J.C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley-Interscience, 2003.
227. *ST-ORM, A Meta-Computing System for Stochastic Optimization and Robustness Management*. EASi Engineering GmbH, Germany, 2000.
228. N. Stander, W. Roux, M. Giger, M. Redhe, N. Fedorova, J. Haarhoff. *Crashworthiness optimization in LS-OPT: case studies in metamodeling and random search techniques*. w: *Proc. 4th European LS-DYNA Users Conference, Ulm, 22–23 May, 2003*.
229. N. Stander, W. Roux, T. Goel, T. Eggleston, K. Craig. *LS-OPT User's Manual. A Design Optimization and Probabilistic Analysis Tool for the Engineering Analyst. Ver. 4.0*. Livermore Software Technology Corporation, Livermore, California 94551, 2009.
230. R. Stocki. *Niezawodnościowa optymalizacja konstrukcji prętowych w zakresie dużych przemieszczeń - teoria i program komputerowy*. Praca Dokt., IPPT PAN, 1999.

231. R. Stocki. *A method to improve design reliability using optimal Latin hypercube sampling*. Computer Assisted Mechanics and Engineering Sciences, 12:393–411, 2005.
232. R. Stocki, K. Kolanek, S. Jendo, M. Kleiber. *Computational aspects of discrete reliability-based optimization of structural systems*. w: *Proc. ECCM-2001 European Conference on Computational Mechanics, Cracow, Poland, June 26-29, 2001*. na CD.
233. R. Stocki, K. Kolanek, S. Jendo, M. Kleiber. *Study on discrete optimization techniques in reliability-based optimization of truss structures*. Computers and Structures, 79(22–25):2235–2247, 2001.
234. R. Stocki, K. Kolanek, J. Knabel, P. Tazowski. *FE based structural reliability analysis using STAND environment*. Computer Assisted Mechanics and Engineering Sciences, 16:35–58, 2009.
235. R. Stocki, A. Siemaszko, M. Kleiber. *Interactive methodology for reliability-based structural design and optimization*. Computer Assisted Mechanics and Engineering Sciences, 6:39–62, 1999.
236. R. Stocki, T. Szolc, P. Tazowski, J. Knabel. *Robust design optimization of the vibrating rotor shaft system subjected to rubbing constraints*. w: *Proc. 8th IFToMM International Conference on Rotordynamics, Seoul, Korea, 12–15 September, 2010*.
237. R. Stocki, P. Tazowski, M. Kleiber. *Efficient sampling techniques for stochastic simulation of structural systems*. Computer Assisted Mechanics and Engineering Sciences, 14:127–140, 2007.
238. R. Stocki, P. Tazowski, J. Knabel. *Reliability analysis of a crashed thin-walled s-rail accounting for random spot weld failures*. International Journal of Crashworthiness, 13:693–706, 2008.
239. H. Streicher, R. Rackwitz. *Reliability-oriented optimization for time-invariant problems with optimization algorithm JOINT 5, research report*. Technical Report Project 28159, EU-Project ASRA-HPC - Advanced Structural Reliability Analysis on High Performance Computers, 2001.
240. H. Streicher, R. Rackwitz. *Structural optimization - a one level approach*. w: S. Jendo, K. Doliński, M. Kleiber (eds.), *AMAS Workshop on Reliability-Based Design and Optimization - RBO'02*, 2002.
241. B. Stroustrup. *C++ Programming Language (special Edition)*. Addison Wesley, 2000.
242. *STRUREL: Structural Reliability Analysis Program Package, User's manual*. RCP GmbH, Barrer Strasse 48, Munich, Germany, 1999.
243. M.B. Subrahmanyam. *An extension of the simplex method to constrained nonlinear optimization*. Journal of Optimization Theory and Applications, 62:311–319, 1989.
244. B. Sudret. *Uncertainty propagation and sensitivity analysis in mechanical models - Contributions to structural reliability and stochastic spectral methods*. Clermont-

- Ferrand (France): Habilitation à diriger des recherches. Université Blaise Pascal, 2007.
245. B. Sudret, A. Der Kiureghian. *Stochastic Finite Element and Reliability. A State-of-the-Art Report*. Technical report, Department of Civil and Environmental Engineering, University of California, Berkeley, 2000.
246. E. Szeliga, M. Witkowski. *Teoria zbiorów rozmytych i jej zastosowania w wybranych zagadnieniach technicznych*. w: *Sieci Neuronowe, Algorytmy Genetyczne, Zbiory Rozmyte*. Rzeszów 28-29 maja, 141–164. Studio BEL s.c., 1999.
247. G. Taguchi, S. Chowdhury, Y. Wu. *Taguchi's Quality Engineering Handbook*. Wiley, 2005.
248. L.C. Tang, K. ; Xu. *A unified approach for dual response surface optimization*. Journal of Quality Technology, 34:437–447, 2002.
249. B.H. Thacker, D.S. Riha, S.H.K. Fitch, L.J. Huyse, J.B. Pleming. *Probabilistic engineering analysis using the NESSUS software*. Structural Safety, 28:83–107, 2006.
250. P. Thoft-Christensen. *On reliability-based structural optimization*. w: A. Der Kiureghian, P. Thoft-Christensen (eds.), *Reliability and Optimization of Structural Systems, Proc. 3rd IFIP WG 7.5 Conf., Berkeley, 26–27 March 1990*, 387–402. Springer-Verlag, 1990.
251. M. Tisza. *Numerical modelling and simulation in sheet metal forming*. Journal of Materials Processing Technology, 151:58–62, 2004.
252. Y. Tsompanakis, N.D. Lagaros, M. Papadrakakis (eds.). *Structural Design Optimization Considering Uncertainties*, Structures and Infrastructures Series. Taylor and Francis, 2007.
253. J. Tu, K.K. Choi. *A new study on reliability based design optimization*. ASME Journal of Mechanical Design, 121(4):557–564, 1999.
254. E. Vanmarcke. *Random Fields: Analysis and Synthesis*. The MIT Press, Cambridge, Massachusetts, 1983.
255. E. Vanmarcke, M. Grigoriu. *Stochastic finite element analysis of simple beams*. Journal of Engineering Mechanics ASCE, 109(5):1203–1214, 1983.
256. G.G. Vining, R.H. Myers. *Combining Taguchi and response surface philosophies: a dual response approach*. Journal of Quality Technology, 22:38–45, 1990.
257. Y.-T. Wu, H.R. Millwater, T.A. Cruse. *Advanced probabilistic structural analysis method for implicit performance functions*. AIAA Journal, 28:1663–1669, 1990.
258. Y. Xiong, W. Chen, D. Apley, X. Ding. *A non-stationary covariance-based Kriging method for metamodelling in engineering design*. International Journal for Numerical Methods in Engineering, 71:733–756, 2006.
259. H. Xu, S. Rahman. *A generalized dimension-reduction method for multidisciplinary integration in stochastic mechanics*. International Journal for Numerical Methods in Engineering, 61:1992–2019, 2004.

260. R.R. Yager. *A note on probabilities of fuzzy events*. Information Sciences, 18:113–129, 1979.
261. R.-J. Yang, C. Chuang, L. Gu, G. Li. *Experience with approximate reliability-based optimization methods II: an exhaust system problem*. Structural and Multidisciplinary Optimization, 29:488–497, 2005.
262. R.-J. Yang, L. Gu. *Experience with approximate reliability-based optimization methods*. Structural and Multidisciplinary Optimization, 26(1–2):152–159, 2004.
263. R.-J. Yang, L. Gu, C.H Tho, K.K. Choi, B.D Youn. *Reliability-Based Multidisciplinary Design Optimization of Vehicle Structures*. w: *International Conference on Statistics and Analytical Methods in Automotive Engineering*, 187–196. IMechE, 2002.
264. K.Q. Ye, W. Li, A. Sudjianto. *Algorithmic construction of optimal symmetric Latin hypercubes*. Journal of Statistical Planning and Inference, 90:145–159, 2000.
265. O. Yeniay, R. Unal, R.A. Lepsch. *Using dual response surfaces to reduce variability in launch vehicle design: A case study*. Reliability Engineering and System Safety, 91:407–412, 2006.
266. X. Yin, W. Chen. *Enhanced sequential optimization and reliability assessment method for probabilistic optimization with varying design variance*. Structure and Infrastructure Engineering, 2(3–4):261–275, 2006.
267. B.D. Youn, K.K. Choi. *A new response surface methodology for reliability-based design optimization*. Computers and Structures, 82(2–3):241–256, 2004.
268. B.D. Youn, K.K. Choi, L. Du. *Enriched performance measure approach for reliability-based design optimization*. AIAA Journal, 43(4):874–884, 2005.
269. B.D. Youn, K.K. Choi, Y.H. Park. *Hybrid analysis method for reliability-based design optimization*. ASME Journal of Mechanical Design, 125(2):221–232, 2003.
270. B.D. Youn, K.K. Choi, R.-J. Yang, L. Gu. *Reliability-based design optimization for crashworthiness of vehicle side impact*. Structural and Multidisciplinary Optimization, 26(3–4):272–283, 2004.
271. L.A. Zadeh. *Fuzzy sets*. Information and control, 8:338–353, 1965.
272. C. Zang, M.I. Friswell, J.E. Mottershead. *A review of robust optimal design and its application in dynamics*. Computers and Structures, 83:315–326, 2005.
273. J. Zhang, M.M. Wiecek, W. Chen. *Local approximation of the efficient frontier in robust design*. Transactions of ASME, Journal of Mechanical Design, 122:232–236, 1999.
274. W. Zhang, R. Shivpuri. *Probabilistic design of aluminium sheet drawing for reduced risk of wrinkling and fracture*. Reliability Engineering and System Safety, 94:152–161, 2009.
275. X. Zhang, H.-Z. Huang. *Sequential optimization and reliability assessment for multidisciplinary design optimization under aleatory and epistemic uncertainties*. Structural and Multidisciplinary Optimization, 40:165–175, 2010.

276. Y. Zhang, A. Der Kiureghian. *Finite element reliability methods for inelastic structures*. Report No. UCB/SEMM-97/05, Department of Civil & Environmental Engineering, University of California, Berkeley, CA, 1997.
277. K. Ziha. *Descriptive sampling in structural safety*. *Structural Safety*, 17:33–41, 1995.
278. T. Zou, S. Mahadevan, Z.P. Mourelatos, P. Meernik. *Reliability analysis of automotive body-door subsystem*. *Reliability Engineering and System Safety*, 78(3):315–24, 2002.
279. T. Zou, S. Mahadevan, Z.P. Mourelatos, P. Meernik. *Reliability Analysis of systems with nonlinear limit states; application to automotive door closing effort*. w: *Reliability & Robust Design in Automotive Engineering*, 19–29. SAE International, 2003.