

Krzysztof Kolanek

**ANALIZA I OPTYMALIZACJA
NIEZAWODNOŚCIOWA KONSTRUKCJI
ZA POMOCĄ ADAPTACYJNYCH METOD
SYMULACYJNYCH**



INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI
POLSKIEJ AKADEMII NAUK
WARSZAWA 2007

ISSN 0208-5658
ISBN 978-83-89687-20-3

Redaktor Naczelny:
doc. dr hab. Zbigniew Kotulski

Promotor:
prof. dr hab. Stefan Jendo

Recenzenci:
doc. dr hab. Krzysztof Doliński
prof. dr hab. Paweł Śniady

Praca wpłynęła do Redakcji 8 czerwca 2007 r.

Praca doktorska

Instytut Podstawowych Problemów Techniki PAN

Nakład 100 egz. Ark. druk. 13,12

Oddano do druku w czerwcu 2007 r.

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

Spis treści

1. Przedmiot, cel i zakres pracy	7
1.1. Przedmiot rozważań	7
1.2. Cel i zakres pracy	8
1.3. Przegląd literatury	14
2. Analiza niezawodności konstrukcji	19
2.1. Wstęp	19
2.2. Pojęcia podstawowe	20
2.3. Metody numeryczne analizy niezawodności	26
2.3.1. Gaussowska przestrzeń standardowa	27
2.3.2. Transformacja wektora losowego do gaussowskiej przestrzeni standardowej	28
2.3.3. Metoda analizy niezawodności pierwszego rzędu (FORM)	30
2.3.4. Metoda analizy niezawodności drugiego rzędu (SORM)	32
2.3.5. Metoda Monte Carlo	35
2.3.6. Klasyczne całkowanie Monte Carlo	35
2.3.7. Metoda importance sampling	37
2.3.8. Adaptacyjne metody Monte Carlo	40
3. Optymalizacja niezawodnościowa	43
3.1. Podstawowe sformułowanie optymalizacji niezawodnościowej	43
3.2. Optymalizacja niezawodnościowa mieszana i dyskretna	45
3.2.1. Transformacja do przestrzeni parametrów ciągłych	46
3.2.2. Metoda kontrolowanego przeglądu	48
4. Pola losowe w analizie niezawodności	61
4.1. Wstęp	61
4.2. Wybrane własności skalarnych pól losowych	63
4.3. Wybrane analityczne własności realizacji pól losowych	68
4.4. Dyskretyzacja pól losowych	73
4.4.1. Metody lokalnego uśredniania	75

4.4.2.	Rozwinięcia w szereg	77
4.4.3.	Metoda funkcji kształtu	79
4.4.4.	Metoda optymalnej aproksymacji liniowej	80
4.4.5.	Dyskretyzacja niegaussowskich pól losowych	83
4.4.6.	Przykłady dyskretyzacji pól losowych	83
5.	Niezawodność zależna od czasu	95
5.1.	Wstęp	95
5.2.	Sformułowanie zagadnienia niezawodności zależnej od czasu	95
5.2.1.	Transformacja do problemu niezależnego od czasu	96
5.2.2.	Klasyczna teoria niezawodności	97
5.2.3.	Prawdopodobieństwo pierwszego przekroczenia	98
5.3.	Częstość przekroczeń ciągłych procesów losowych	102
5.4.	Częstość wyjść procesu odnowy o prostokątnych impulsach	105
6.	Niezawodność systemów konstrukcyjnych	111
6.1.	Wstęp	111
6.2.	Modelowanie konstrukcji za pomocą systemów	111
6.3.	Analiza niezawodności systemów	115
6.4.	Zastosowanie FORM w analizie niezawodności systemów	116
6.4.1.	Systemy szeregowe	116
6.4.2.	Systemy równoległe	118
6.4.3.	Systemy mieszane	122
6.5.	Przedziały prawdopodobieństwa awarii systemu	123
6.6.	Metody Monte Carlo w analizie niezawodności systemów	128
7.	Metoda wzajemnej entropii w analizie niezawodności konstrukcji	131
7.1.	Adaptacyjne metody symulacyjne	131
7.2.	Metoda wzajemnej entropii	132
7.3.	Algorytm metody wzajemnej entropii	135
7.4.	Ocena efektywności metody wzajemnej entropii	139
8.	Analiza niezawodności za pomocą metod wykorzystujących łańcuchy Markowa	147
8.1.	Markov chain Monte Carlo (MCMC)	147
8.2.	Wybrane informacje o łańcuchach Markowa	148
8.3.	Algorytm Metropolis-Hastings (M-H)	150
8.4.	Rodzaje algorytmu Metropolis-Hastings	155
8.4.1.	Przypadkowe błędzenie	155
8.4.2.	Łańcuch niezależny	156
8.5.	Efektywna implementacja algorytmu Metropolis-Hastings	157
8.5.1.	Ocena efektywności algorytmu Metropolis-Hastings	157
8.5.2.	Zalecana wartość wskaźnika przejść przypadkowego błędzenia	160

8.6. Adaptacyjne algorytmy Metropolis-Hastings	161
8.6.1. Kryteria adaptacji	162
8.6.2. Dostosowanie rozkładu pomocniczego	165
8.7. Wyznaczanie stałych normalizujących	167
8.7.1. Estymator wykorzystujący przybliżenie Laplace'a	167
8.7.2. Odwrotna metoda importance sampling	168
8.7.3. Bridge sampling	169
8.7.4. Annealed Importance Sampling (AIS) oraz Linked Importance Sampling (LIS)	171
8.8. Oszacowanie prawdopodobieństwa awarii za pomocą metod MCMC	175
8.9. Ocena błędu oszacowania za pomocą metody bootstrap	178
8.10. Optymalizacja algorytmu Metropolis-Hastings do zastosowań w analizie niezawodności	180
9. Analiza niezawodności ściskanej półki blachownicy	189
10. Wnioski, spostrzeżenia i kierunki dalszych badań	197

Przedmiot, cel i zakres pracy

1.1. Przedmiot rozważań

Projektowanie bezpiecznych konstrukcji inżynierskich wymaga uwzględnienia niepewności parametrów projektowych. W praktyce, najczęściej w tym celu wykorzystuje się współczynniki bezpieczeństwa, które zmieniają wartości parametrów zgodnie z zaleceniami podanymi w normach. Od około trzydziestu lat intensywnie rozwijane są metody pozwalające wykorzystać model probabilistyczny do oszacowania prawdopodobieństwa awarii konstrukcji. Ze względu na znaczną złożoność metody te nie upowszechniły się w praktyce inżynierskiej, pozwalają one jednak na bardziej racjonalną ocenę bezpieczeństwa niż podejście tradycyjne.

W probabilistycznej analizie niezawodności przyjmuje się, że parametry konstrukcyjne obciążone niepewnością modelowane są za pomocą zmiennych losowych. Podobnie jak w tradycyjnym podejściu deterministycznym analiza stochastyczna wykorzystuje pojęcie stanu granicznego do określenia, czy reakcja konstrukcji na obciążenie jest dopuszczalna, czy też występuje zagrożenie jej bezpieczeństwa. Zbiór wartości niepewnych parametrów projektowych prowadzących do przekroczenia stanu granicznego określa zdarzenie losowe (tzw. obszar awarii), które jest utożsamiane z awarią konstrukcji. Prawdopodobieństwo tego zdarzenia (tzw. prawdopodobieństwo awarii), a także jego funkcje są miarami niezawodności konstrukcji. Zauważmy, że ten sposób oceny bezpieczeństwa zasadniczo składa się z dwóch zagadnień: określenia realistycznego modelu stochastycznego oraz oszacowania prawdopodobieństwa awarii.

Określenie rozkładów prawdopodobieństwa parametrów konstrukcyjnych wymaga uwzględnienia, często wieloletnich, obserwacji atmosferycznych oraz badań jakości materiałów i elementów konstrukcyjnych. Modelowanie własności, które obok zmienności losowej wykazują zależność od czasu lub położenia w przestrzeni wymaga zastosowania funkcji losowych. Przykładem może być tu przedstawienie obciążeń jako procesów losowych lub imperfekcji geometrycznych jako pól losowych. Wykorzystanie funkcji losowych, obok trudności związanych z opracowaniem danych statystycznych, jest również

związane z zastosowaniem skomplikowanych procedur numerycznych. Należy także zaznaczyć, że wiele problemów związanych z zastosowaniem funkcji losowych w analizie niezawodności wciąż nie zostało rozwiązanych.

Wyznaczenie prawdopodobieństwa awarii wymaga obliczenia całki funkcji gęstości prawdopodobieństwa losowych parametrów konstrukcyjnych w obszarze awarii. Ponieważ, w celu zdefiniowania realistycznego modelu konstrukcji konieczne jest użycie co najmniej kilku zmiennych losowych, to całki spotykane w analizie niezawodności są zwykle wielowymiarowe. Okazuje się, że kwadratury są efektywne jedynie w przypadku numerycznego całkowania funkcji określonych w mniej niż pięciu wymiarach. Analiza niezawodności wymaga więc zastosowania alternatywnych algorytmów wyznaczania całek. Najbardziej efektywne numerycznie, a także zazwyczaj zapewniające wystarczającą dokładność są tzw. metody FORM i SORM. Metody te wykorzystują liniową bądź kwadratową aproksymację obszaru awarii. Ich zastosowanie jest jednak ograniczone do przypadków, w których funkcje określające stany graniczne są różniczkowalne. Jeżeli warunek ten nie jest spełniony stosowane jest całkowanie Monte Carlo, najczęściej wraz z metodami redukcji wariancji takimi jak importance sampling. Skuteczne zastosowanie importance sampling wymaga właściwego doboru rozkładu prawdopodobieństwa, z którego generowana jest próba losowa. Do rozwiązania zadań, w przypadku których określenie takiego rozkładu jest trudne, można wykorzystać tzw. adaptacyjne metody symulacyjne. W trakcie wykonania tych algorytmów rozkład prawdopodobieństwa służący do generowania próby losowej podlega systematycznej adaptacji na podstawie wykonanych wcześniej symulacji.

Jeden z rozdziałów tej pracy poświęcono optymalizacji niezawodnościowej. Korzystając ze stochastycznego opisu konstrukcji, podejście to pozwala w sposób jawny wprowadzić do zadania optymalizacji ograniczenia na wielkość prawdopodobieństwa awarii. Wydaje się, że w ten sposób można rozwiązać najistotniejszy problem związany z optymalizacją deterministyczną, jakim jest wrażliwość konstrukcji optymalnych na zmiany parametrów projektowych. Przewiduje się, że optymalizacja niezawodnościowa może znaleźć istotne zastosowanie przy projektowaniu nietypowych i drogich konstrukcji. W tej pracy szczególną uwagę poświęcono zagadnieniu dyskretnej optymalizacji niezawodnościowej. Przyjęcie dodatkowych dyskretnych ograniczeń na wartości parametrów projektowych pozwala uwzględnić problem dostępności odpowiednich elementów konstrukcyjnych.

1.2. Cel i zakres pracy

Jak wspomniano w poprzednim podrozdziale, do oszacowania prawdopodobieństwa awarii można wykorzystać adaptacyjne metody symulacyjne. Algorytmy tego rodzaju

nie wymagają, tak jak FORM/SORM, różniczkowalności funkcji granicznej, a zarazem są bardziej efektywne obliczeniowo niż klasyczna metoda Monte Carlo. W literaturze można znaleźć wiele przykładów zastosowania algorytmów adaptacyjnych do rozwiązania zadania analizy niezawodności konstrukcji. Podstawą tych metod zwykle są skomplikowane procedury iteracyjne, dla których analiza zbieżności jest bardzo trudna. Dlatego też w publikacjach dotyczących tego zagadnienia najczęściej przedstawiane są przykłady efektywnego zastosowania proponowanych algorytmów ale bez dowodów ich zbieżności. Oczywiście znajomość warunków zbieżności jest cechą ułatwiającą efektywne wykorzystanie jakichkolwiek metod iteracyjnych. Podjęto więc próbę wyszukania i zbadania możliwości zastosowania w analizie niezawodności konstrukcji adaptacyjnych metod symulacyjnych, które mają dobrze opracowane podstawy teoretyczne. Ostatecznie analizie poddano metodę wzajemnej entropii (ang. *Cross-Entropy Method*) oraz metody generujące łańcuchy Markowa, które w literaturze angielskiej są określane jako *Markov chain Monte Carlo*.

Metoda wzajemnej entropii została zaproponowana w roku 1997 jako efektywny algorytm służący do oszacowania prawdopodobieństwa rzadkich zdarzeń. Od tego czasu metoda ta była intensywnie rozwijana i obecnie jest dobrze opracowanym teoretycznie algorytmem symulacji rzadkich zdarzeń i optymalizacji kombinatorycznej. Zastosowanie metody wzajemnej entropii w analizie niezawodności konstrukcji wymaga jedynie uwzględnienia innej postaci nierówności, która definiuje przekroczenie stanu granicznego. Autorowi nie są jednak znane wcześniejsze przykłady wykorzystania tego podejścia w analizie niezawodności konstrukcji. W jednym z rozdziałów omówiono podstawy teoretyczne oraz szczegółowo przedstawiono implementację metody wzajemnej entropii. Wykorzystano przy tym notację charakterystyczną w przypadku analizy niezawodności konstrukcji. Ponadto, na podstawie przykładów testowych reprezentujących trudności związane z oszacowaniem prawdopodobieństwa awarii zbadano efektywność numeryczną tego algorytmu. Otrzymane wyniki mogą być wykorzystane jako podstawa doboru wartości jego parametrów, zależnie od cech rozwiązywanego zadania.

Metody Markov chain Monte Carlo są drugą grupą algorytmów symulacyjnych, których zastosowanie w analizie niezawodności jest przedmiotem tej pracy. Jak wiadomo efektywność metody importance sampling zależy od właściwego doboru rozkładu prawdopodobieństwa, z którego generowana jest próba losowa. Gęstość rozkładu prawdopodobieństwa minimalizującego wariancję estymatora importance sampling jest niezerowa jedynie w obszarze awarii, gdzie jest proporcjonalna do gęstości prawdopodobieństwa losowych parametrów konstrukcji (estymator określony w ten sposób ma rozkład punktowy). Zastosowanie metod bezpośredniego generowania liczb losowych do tak zdefiniowanej gęstości prawdopodobieństwa wymaga wyznaczenia jej stałej normalizującej, którą w tym przypadku jest prawdopodobieństwo awarii. Często można spotkać się ze stwierdzeniem, że w celu dokładnego oszacowania prawdopodobieństwa

awarii za pomocą importance sampling wystarczyłaby jedna symulacja z rozkładu optymalnego, ale do uzyskania oszacowania w ten sposób konieczna jest znajomość rozwiązania zadania. Zdaniem autora stwierdzenie takie jest mylące, gdyż znając stałą normalizującą znany rozwiązanie zadania i przeprowadzanie symulacji jest bezcelowe. Zasadniczym problemem nie jest w tym przypadku generowanie próby losowej, a jest nim oszacowanie stałej normalizującej równej prawdopodobieństwu awarii. Algorytmy Markov chain Monte Carlo pozwalają otrzymać próbę losową z rozkładu o danej gęstości prawdopodobieństwa, jeżeli jest znana funkcja do niej proporcjonalna. Za pomocą tych metod można więc wygenerować próbę losową o rozkładzie optymalizującym metodę importance sampling. Przybliżenia prawdopodobieństwa awarii na podstawie takiej próby losowej wymaga oszacowania stałej normalizującej jej gęstość, do czego konieczne jest posłużenie się odpowiednim algorytmem.

Rozdział poświęcony zastosowaniu Markov chain Monte Carlo w analizie niezawodności zawiera omówienie metod wykorzystywanych do oszacowania prawdopodobieństwa awarii za pomocą przedstawionego podejścia. Poruszane są zagadnienia kluczowe w przypadku numerycznej implementacji algorytmów, takie jak efektywność oszacowania czy modyfikacja rozkładu prawdopodobieństwa pozwalająca przyjąć dowolny punkt startowy. Rozdział ten zawiera również przykłady zastosowania proponowanej metody do rozwiązania zadań testowych charakterystycznych w przypadku analizy niezawodności.

W niniejszej pracy również przedstawiono wykorzystanie pól losowych do modelowania imperfekcji geometrycznych konstrukcji jedno i dwuwymiarowych. Rozważane imperfekcje mają postać wstępnych odchyień od idealnej osi pręta bądź powierzchni środkowej płyty. Szczególną uwagę poświęcono uzyskaniu pól losowych, których realizacje spełniają warunki brzegowe utwierdzenia. W tym celu posłużono się metodą warunkowania pól losowych. To podejście było już wcześniej stosowane do definiowania pól losowych, których realizacje spełniają warunki brzegowe swobodnego podparcia. Uwzględnienie warunków utwierdzenia wymagało teoretycznego opracowania zagadnienia warunkowania pól losowych na ich pochodnych. Zaproponowaną metodę przedstawiono na przykładach numerycznych, które pokazały, że uzyskane realizacje pól losowych spełniają warunki brzegowe z zadowalającą dokładnością. Ponadto omówiono warunki ciągłości i różniczkowości realizacji pól losowych.

Przedmiotem jednego z rozdziałów jest przykład, w którym analizuje się niezawodność ściskanej półki blachownicy z uwzględnieniem jej wstępnej deformacji. Jako model tego elementu konstrukcyjnego przyjęto płytę z imperfekcjami w postaci gaussowskiego pola losowego. Spełnienie warunków podparcia przez realizacje pola losowego zostało zapewnione za pomocą metody warunkowania. Rozważanym stanem granicznym jest przekroczenie dopuszczalnych wartości przemieszczeń normalnych do idealnej powierzchni płyty, w wyniku jej ściskania. Do wyznaczenia prawdopodobieństwa awarii

zastosowano metodę wzajemnej entropii i metody oparte na Markov chain Monte Carlo. Zwrócono także uwagę, że symetrie konstrukcji skutkują symetrami powierzchni awarii, co wiąże się z występowaniem kilku istotnych punktów projektowych. Intrygujący jest fakt, że pomimo powszechnego wykorzystywania symetrycznych modeli konstrukcji, zagadnienie to nie było wcześniej poruszane w literaturze. Do rozwiązania tego problemu zastosowano podejście oparte na przedstawieniu stanu granicznego w postaci systemu awarii.

Tematem poruszonym w pracy jest także optymalizacja niezawodnościowa konstrukcji. Oryginalnym elementem w tym przypadku jest przyjęcie założenia, że pewne parametry projektowe przyjmują wartości w zbiorach dyskretnych. Optymalizacja niezawodnościowa jest problemem, którego rozwiązanie wymaga długotrwałych obliczeń komputerowych. Uwzględnienie dyskretnego charakteru parametrów projektowych stanowi dodatkowe utrudnienie, gdyż uniemożliwia zastosowanie algorytmów gradientowych. Do rozwiązania tak sformułowanego zadania zaproponowano dwa algorytmy. Pierwszy z nich wykorzystuje transformację zmiennych dyskretnych do rozszerzonej przestrzeni parametrów ciągłych. Drugi, to algorytm kontrolowanego przeglądu, w którym wprowadzono modyfikacje zwiększające jego efektywność. Praktyczne zastosowanie tych metod zostało przedstawione na przykładach optymalizacji konstrukcji kratownicowych, gdzie dyskretnymi parametrami projektowymi były pola przekrojów prętów. Takie sformułowanie zagadnienia odzwierciedla ograniczone możliwości doboru elementów konstrukcyjnych przy wykorzystaniu standardowych wyrobów hutniczych.

W pracy omówiono również podstawowe zagadnienia związane z analizą niezawodności. Obok klasycznych metod oszacowania prawdopodobieństwa awarii przedstawione jest sformułowanie niezawodności zależnej od czasu oraz metody analizy niezawodności systemów awarii.

Rozprawa składa się z 10 rozdziałów. Poza zawierającym wstęp pierwszym rozdziałem, opisują one następujące zagadnienia:

Rozdział 2 zawiera podstawy analizy niezawodności. Na wstępie wprowadzane są pojęcia, takie jak: awaria konstrukcji, funkcja graniczna, obszar awarii, prawdopodobieństwo awarii oraz wskaźnik niezawodności. Wykorzystanie tych pojęć zilustrowano za pomocą prostego przykładu. Dalej, rozdział 2 zawiera omówienie podstawowych algorytmów numerycznych wykorzystywanych do wyznaczenia prawdopodobieństwa awarii; są to: transformacja zmiennych losowych do gausowskiej przestrzeni standardowej, metody FORM i SORM, a także całkowania Monte Carlo z uwzględnieniem metod redukcji wariancji.

Rozdział 3 poświęcono zagadnieniu optymalizacji niezawodnościowej konstrukcji. Przedstawiono sformułowanie minimalizacji kosztu początkowego przy ograniczeniach nałożonych na wielkość prawdopodobieństwa awarii. Następnie sformuło-

wanie to rozszerzono o ograniczenia pozwalające uwzględnić dyskretny charakter zmiennych projektowych. Do rozwiązania zadania niezawodnościowej optymalizacji dyskretniej zaproponowano dwa algorytmy. Pierwszy z nich wykorzystuje transformację zadania do ciągłej przestrzeni parametrów. Dopuszczalne wartości dyskretnych parametrów projektowych reprezentowane są za pomocą zmiennych z przedziału $[0, 1]$. Wprowadzenie do zadania dodatkowych ograniczeń pozwala na jednoznaczne określenie optymalnych wartości parametrów dyskretnych. Drugą omawianą metodą jest algorytm kontrolowanego przeglądu. W tym przypadku wprowadzono modyfikację, która polega na wstępnej selekcji kombinacji parametrów projektowych w oparciu o liniową aproksymację wskaźników niezawodności. Dokładnej analizie niezawodności poddawane są tylko te kombinacje, których przybliżone wartości ograniczeń są najbliższe zeru. Taka modyfikacja pozwala na uzyskanie zadowalającego rozwiązania, przy znacznej oszczędności czasu obliczeń. Efektywność obydwu metod porównano na przykładach optymalizacji konstrukcji kratownicowych. Analizie poddano płaską kratownicę złożoną z 10 prętów oraz kopułę kratownicową podatną na utratę stateczności. W przykładach tych, jako zmienne dyskretnie przyjęto pola przekrojów prętów.

Rozdział 4 dotyczy wykorzystania w analizie niezawodności pól losowych. Omówiono podstawowe charakterystyki pól losowych, takie jak: funkcja wartości oczekiwanej, funkcja kowariancji, długość korelacji oraz stacjonarność. Oszacowanie prawdopodobieństwa awarii wymaga przedstawienia pola losowego za pomocą skończonej liczby zmiennych losowych - konieczne jest wykonanie tzw. dyskretyzacji. W pracy przedstawiono następujące algorytmy wykorzystywane w tym celu: metody średnich przestrzennych, metody rozwinięcia w szereg oraz aproksymację liniową. Głównym celem rozdziału 4 jest modelowanie za pomocą pól losowych imperfekcji geometrycznych z uwzględnieniem warunków brzegowych utwierdzenia. Zaproponowana metoda wykorzystuje warunkowanie jednorodnego pola losowego na wartościach i pochodnych jego realizacji w punktach brzegowych. Przedstawiono więc sposób, w jaki można określić warunkowe pole losowe, a w załączonym przykładzie wyprowadzono warunkowe funkcje wartości oczekiwanej i kowariancji gaussowskich pól losowych. Uwzględnienie warunków brzegowych w postaci utwierdzenia wymaga wprowadzenia pojęcia gradientu pola losowego. Ponadto konieczne jest wyznaczenie wzajemnych kowariancji między składowymi gradientu oraz kowariancji wartości pola losowego i składowych jego gradientu. W pracy podano również warunki zapewniające ciągłość i różniczkowalność pól losowych. Przykłady zamieszczone w rozdziale 4 ilustrują zastosowanie zaproponowanej metody do modelowania imperfekcji geometrycznych prętów i płyt z uwzględnieniem warunków ich podparcia.

Rozdział 5 poświęcono omówieniu niezawodności zależnej od czasu. Na wstępie zostały przedstawione podstawowe sformułowania pojęcia prawdopodobieństwa awarii w przedziale czasu: transformacja do zadania niezależnego od czasu oraz klasyczna teoria niezawodności. Przedstawiono również wykorzystywane wspólnie w analizie niezawodności konstrukcji pojęcie prawdopodobieństwa pierwszego przekroczenia powierzchni awarii przez proces losowy modelujący losowe i zależne od czasu parametry projektowe. W dalszej części rozdziału 5 omówiono, kluczowe w przypadku prawdopodobieństwa pierwszego przekroczenia, zagadnienie częstości wyjść procesu losowego do obszaru awarii. Przedstawiono formułę Rice'a dotyczącą procesów ciągłych oraz omówiono metody wyznaczania częstości przekroczeń procesów odnowy o prostokątnych impulsach. Rozdział kończy elementarne zadanie analizy niezawodności zależnej od czasu.

Rozdział 6 dotyczy analizy niezawodności systemów. Przedstawiono definicje pojęć takich, jak element oraz system awarii, a w przykładzie wykorzystano je do określenia globalnej awarii kratownicy statycznie niewyznaczalnej. Omówiono również metody wyznaczania przybliżeń i ograniczeń prawdopodobieństwa awarii systemów.

Rozdział 7 poświęcono zastosowaniu metody wzajemnej entropii do wyznaczenia prawdopodobieństwa awarii konstrukcji. Omówione zostały podstawy teoretyczne tej metody, jak również szczegółowo przedstawiono algorytm numeryczny. Wykorzystano przy tym notację stosowaną w analizie niezawodności konstrukcji. Efektywność prezentowanej metody zbadano na podstawie przykładów testowych odzwierciedlających problemy charakterystyczne w przypadku analizy niezawodności: dużą liczbę zmiennych losowych, małe prawdopodobieństwo awarii, silną nieliniowość powierzchni granicznej, zaszumienie funkcji granicznej, nieróżniczkowalność powierzchni granicznej, złożony kształt systemowych obszarów awarii. Podane wyniki zawierają średnią liczbę symulacji koniecznych do oszacowania prawdopodobieństwa awarii ze współczynnikiem zmienności na poziomie 10%. Na ich podstawie można, więc zidentyfikować zagadnienia, w których metoda wzajemnej entropii jest efektywna. W przykładach omówiono również zagadnienie doboru parametrów algorytmu, zapewniających jego efektywne działanie.

Rozdział 8 zawiera podstawy metod Markov chain Monte Carlo oraz prezentuje ich zastosowanie w analizie niezawodności konstrukcji. Rozdział rozpoczyna się od omówienia wybranych własności łańcuchów Markowa. Dalej opisywany jest algorytm Metropolis-Hastings będący podstawowym algorytmem typu Markov chain Monte Carlo. Poruszane są następujące zagadnienia związane z jego implementacją: klasyfikacja odmian algorytmu, efektywność oszacowania oraz dostosowanie do rozwiązywanego zadania. Część rozdziału poświęcono metodom wyznaczania

stałych normalizujących gęstość rozkładu prawdopodobieństwa. Zagadnienie to jest kluczowe w przypadku zaproponowanego sposobu oszacowania prawdopodobieństwa awarii na podstawie próby losowej z optymalnego rozkładu importance sampling otrzymanej za pomocą Markov chain Monte Carlo. W rozdziale 8 omówiono również inne zagadnienia, które są istotne w przypadku proponowanego podejścia: rozszerzenie poza obszar awarii rozkładu prawdopodobieństwa, z jakiego generowana jest próba losowa, ocena błędu oszacowania za pomocą metody bootstrap oraz dobór parametrów tzw. rozkładu pomocniczego. Zamieszczone przykłady numeryczne prezentują efektywność opracowanej metody w przypadku testowych zadań analizy niezawodności.

Rozdział 9 stanowi przykład analizy niezawodności ściskanej półki blachownicy z imperfekcjami geometrycznymi. Zadanie to przedstawia praktyczne zastosowanie omówionych we wcześniejszej części pracy metod modelowania imperfekcji geometrycznych oraz algorytmów analizy niezawodności.

Rozdział 10 zawiera wnioski, spostrzeżenia i omówienie możliwości kontynuacji badań.

1.3. Przegląd literatury

Dziedzina nauki, której w największym stopniu dotyczy ta praca jest analiza niezawodności konstrukcji. Ze względu na wykorzystanie do oceny bezpieczeństwa stochastycznego modelu konstrukcji, w analizie niezawodności szeroko wykorzystuje się probabilistykę, w tym również komputerowe metody statystyczne. W dalszej części pracy można więc znaleźć odwołania to takich zagadnień probabilistycznych jak teoria pól losowych, symulacje *Monte Carlo*, symulacje *Markov chain Monte Carlo* czy metoda *bootstrap*. W przypadku każdej z wymienionych dziedzin dostępna jest bogata bibliografia. W poniższym przeglądzie, wymieniono głównie pozycje istotne dla tej rozprawy, a także wspomniano prace kluczowe dla rozwoju każdej z wymienionych dziedzin nauki.

Analiza niezawodności konstrukcji jest obecnie dziedziną nauki o dobrze opracowanych podstawach teoretycznych i znajduje zastosowanie w projektowaniu oraz ocenie stanu, zwłaszcza nietypowych, konstrukcji inżynierskich. Dostępnych jest szereg monografii i podręczników dotyczących tej tematyki, można wymienić tutaj prace Ditlevsena i Madsena [28] Fabera [36] oraz Melchersa [73]. Ciekawe podsumowanie oraz omówienie perspektyw rozwoju analizy niezawodności zaprezentował R. Rackwitz w pracy [93]. Warto również wspomnieć o pozycjach polskojęzycznych: monografii J.M. Murzewskiego [82] i książce autorstwa A. Biegusa poświęconej analizie niezawodności konstrukcji stalowych [12]. Tematyce analizy niezawodności konstrukcji całkowicie

poświęcone jest czasopismo "Structural Safety" wydawane przez Elsevier.

Na rynku dostępnych jest kilka pakietów oprogramowania służących do analizy niezawodności, np.: STRUREL [47], COSSAN [108], NESSUS [119], Proban [123]. Przeglądowi komputerowych systemów analizy niezawodności poświęcony został numer 1-2 z 2006 roku czasopisma Structural Safety.

Autorem, powstałej w 1937 roku, jednej z pierwszych prac dotyczących oceny bezpieczeństwa konstrukcji [129] jest polski uczony W. Wierzbicki. Powszechnie za prekursora analizy niezawodności uważa się A. Freudenthala, którego autorstwa jest praca o bezpieczeństwie konstrukcji poddanych losowym obciążeniom [39]. W przypadku współczesnej analizy niezawodności jedną z pierwszych kluczowych publikacji jest praca [51] z 1974 roku, gdzie zaproponowano, nazwany od nazwisk autorów, tzw. wskaźnik niezawodności Hasofera-Linda. Definicja tej miary niezawodności wykorzystuje linearyzację powierzchni granicznej w tzw. punkcie projektowym (punkcie obszaru awarii o największej gęstości prawdopodobieństwa). Najistotniejszą cechą wskaźnika Hasofera-Linda jest niezmienniczość względem definicji stanu granicznego. Brak tej cechy był zasadniczą słabością używanego wcześniej tzw. wskaźnika Cornella [19].

Sformułowanie wprowadzone przez Hasofera i Linda zostało rozwinięte przez Rackwitz i Fiesslera w pracy [94]. Wykorzystali oni transformację niezależnych zmiennych losowych o dowolnym rozkładzie prawdopodobieństwa do standardowego rozkładu normalnego. W pracy tej zaproponowano również algorytm poszukiwania punktu projektowego.

Idea przedstawiona w [94] była rozwijana w pracy Hohenbichlera i Rackwitz [54], gdzie wykorzystano transformację Rosenblatta [100] oraz w pracy Der Kiureghiana i Liu [25], w której zastosowano transformację Natafa [83] do zamiany dowolnych zmiennych losowych na standardowy rozkład normalny. Te dwie metody transformacji są obecnie powszechnie wykorzystywane w analizie niezawodności; pozwalają one na ocenę bezpieczeństwa konstrukcji w przypadkach, kiedy znany jest łączny rozkład prawdopodobieństwa oraz, gdy znane są tylko rozkłady brzegowe i macierz korelacji zmiennych losowych.

Metoda oszacowania prawdopodobieństwa awarii, opierająca się na linearyzacji powierzchni granicznej w przestrzeni standardowych zmiennych normalnych, jest określana skrótem FORM, od *ang. First Order Reliability Method* [94]. Ze względu na dużą efektywność obliczeniową, algorytm ten jest najczęściej stosowany w analizie niezawodności konstrukcji. Dokładniejsze oszacowanie prawdopodobieństwa awarii, można uzyskać na podstawie aproksymacji powierzchni granicznej przez powierzchnię drugiego rzędu. Tak zwana metoda SORM (*ang. Second Order Reliability Method*) została zaproponowana w pracy [38]. Dokładność oraz efektywność numeryczna tej metody była tematem licznych publikacji, najistotniejsze z nich to [14, 24].

Inną metodą stosowaną do oszacowania prawdopodobieństwa awarii jest całkowanie

Monte Carlo. Ponieważ prawdopodobieństwa awarii konstrukcji są zazwyczaj bardzo małe, to efektywne wykorzystanie metody Monte Carlo wymaga użycia metod redukcji wariancji, z których najpopularniejszą jest importance sampling. Powstało wiele prac traktujących o specyficznych problemach związanych z zastosowaniem metod importance sampling do oszacowania prawdopodobieństwa awarii, ciekawsze z nich to [58, 74, 107] oraz praca przeglądowa [34]. Importance sampling jest również wykorzystywana do poprawy oszacowania uzyskanego metodami FORM i SORM [29, 40, 55].

Ze względu na tematykę tej rozprawy, warto wymienić jeszcze publikacje dotyczące oszacowania prawdopodobieństwa awarii za pomocą adaptacyjnych metod symulacyjnych. Metody te wykorzystują informacje, jakie są zdobywane w czasie procesu symulacji do uaktualniania rozkładu prawdopodobieństwa, z którego generowana jest próba losowa. Chyba najczęściej cytowaną pracą w tym zakresie jest artykuł Buchera [17]. Inne publikacje to praca Melchera [75] oraz Wu [130].

Celem analizy niezawodności zależnej od czasu jest uwzględnienie zmienności własności konstrukcji w okresie jej użytkowania. W tym sformułowaniu, jako miarę niezawodności najczęściej przyjmuje się prawdopodobieństwo, że w określonym przedziale czasu, trajektoria funkcji granicznej wyjdzie do obszaru awarii. Oszacowanie tak sformułowanego prawdopodobieństwa awarii, uzyskuje się na podstawie przybliżenia asymptotycznego, które zostało zaproponowane przez Cramera i Leadbettera [20], bądź górnego ograniczenia przypisywanego Bołotinowi (najczęściej cytuje się tutaj pracę [13]). Należy tutaj również wymienić pracę [105], gdzie podano powyższe oszacowania w przypadku modeli, które obok procesów losowych składają się także z niezależnych od czasu zmiennych losowych. Zagadnienie analizy niezawodności zależnej od czasu zostało także poruszone w polskojęzycznej publikacji autorstwa P. Śniadego [87].

Kluczowym zagadnieniem związanym z oszacowaniem prawdopodobieństwa awarii w przedziale czasu jest wyznaczenie częstości, z jaką proces funkcji granicznej wychodzi do obszaru awarii. Podstawowe prace dotyczące tego problemu to, w przypadku wektorowych procesów gaussowskich artykuł Veneziano, Grigoriu i Cornella [126], a w przypadku procesów odnowy o prostokątnych impulsach publikacja Breitunga i Rackwitz [16]. Przegląd metod oszacowania częstości wyjść został zamieszczony w pracy Rackwitz [92]. Problemem jest wciąż oszacowanie częstości wyjść kombinacji procesów losowych. Jak dotąd zadanie to rozwiązano tylko w kilku przypadkach, co ogranicza możliwości praktycznego zastosowania analizy niezawodności zależnej od czasu. Ograniczone zastosowanie praktyczne ma również analiza niezawodności systemów. W ramach tego sformułowania można ocenić globalne bezpieczeństwo konstrukcji, na podstawie lokalnie określonych stanów granicznych. Metody oszacowania prawdopodobieństwa awarii systemów są dobrze opracowane; omówiono je w pracy [27]. Jednak w tym przypadku, zasadniczym problemem jest określenie systemu awarii. Przedstawienie zależności między funkcjami granicznymi jest trudne, a w przypadku złożonych

konstrukcji może okazać się praktycznie niewykonalne. Jedną z publikacji, gdzie porusza się problem definiowania systemów awarii, jest książka Thoft-Christensena i Murotsu [120].

Podstawy teorii pól losowych zostały przedstawione w książce VanMarcke'a [124]. Zastosowanie pól losowych w analizie niezawodności wymaga przeprowadzenia ich dyskretyzacji - aproksymacji przy użyciu skończonej liczby zmiennych losowych w celu oszacowania prawdopodobieństwa awarii. Przegląd metod dyskretyzacji pól losowych zamieszczono w raporcie [116]. Przedmiotem wielu publikacji jest modelowanie parametrów materiałowych za pomocą jednorodnych pól losowych. Analiza imperfekcji geometrycznych jest trudniejsza, gdyż wymaga uwzględnienia warunków brzegowych podparcia konstrukcji. W pracy Mosta, Buchera i Schorlinga [80] zaproponowano wykorzystanie warunkowych pól losowych do reprezentacji imperfekcji geometrycznych konstrukcji swobodnie podpартych.

Jednym z celów pracy jest analiza zastosowania metody wzajemnej entropii (ang. *Cross-Entropy Method*) w analizie niezawodności konstrukcji. Metoda ta została zaproponowana przez R.Y. Rubinsteina w roku 1997, jako adaptacyjny algorytm symulacji rzadkich zdarzeń. Od tego czasu, omawiany algorytm był intensywnie rozwijany i obecnie również znajduje zastosowanie w optymalizacji kombinatorycznej. Metodzie wzajemnej entropii została poświęcona monografia autorstwa R.Y. Rubinsteina i D.P. Kroese [101]. Ze względu na wykorzystanie tego algorytmu w analizie niezawodności, jako metody symulacji rzadkich zdarzeń, warto również wymienić pracę autorstwa T. Homem de Mello i R.Y. Rubinsteina [56]. Wiele publikacji jest dostępnych na stronie internetowej poświęconej metodzie wzajemnej entropii: ew3.technion.ac.il/CE, prowadzonej przez *Technion - Israel Institute of Technology*.

Początki metod Markov chain Monte Carlo sięgają lat pięćdziesiątych XX wieku i badań prowadzonych w laboratorium w Los Alamos. W pracy z 1953 roku N. Metropolis, wraz ze współautorami, przedstawił zastosowanie algorytmu generującego łańcuch Markowa do obliczeń z zakresu fizyki statystycznej. Kolejnym ważnym krokiem w rozwoju tych metod był artykuł Hastingsa z 1970 roku [52], gdzie zaproponowano uogólnienie algorytmu Metropolis. Tak zwany algorytm Metropolis-Hastings jest podstawowym przykładem metod Markov chain Monte Carlo.

Intensywny rozwój metod symulacyjnych generujących łańcuchy Markowa miał miejsce pod koniec XX wieku, wraz z upowszechnieniem się komputerów. W latach dziewięćdziesiątych opublikowano liczne prace dotyczące szczegółów implementacji, optymalizacji oraz analizy własności algorytmów tego rodzaju. Powszechnie cytowany jest artykuł L. Tierney'a [121], który w zwięzły sposób przedstawia podstawowe zagadnienia związane z tą dziedziną. Problematyka Markov chain Monte Carlo jest również przedmiotem monografii [46, 68, 96].

Proponowany sposób oszacowania prawdopodobieństwa awarii, z wykorzystaniem

Markov chain Monte Carlo, wymaga wyznaczenia tzw. stałej normalizującej gęstość prawdopodobieństwa próby losowej. Wśród algorytmów służących do rozwiązania tego zagadnienia warto wskazać metodę *bridge-sampling* [77] oraz jej rozwinięcie tzw. *path-sampling* [42]. W pracy [42] przedstawiono również ciekawe porównanie metod wyznaczania stałych normalizujących. W przykładach, zamieszczonych w dalszej części tej rozprawy, zastosowano algorytmy autorstwa R.M. Neala [84, 85], będące rozwinięciem *bridge-sampling* i wykorzystujące elementy metod Markov chain Monte Carlo.

Próby zastosowania metod Markov chain Monte Carlo w analizie niezawodności były już wcześniej podejmowane, jako przykład można podać artykuł S.K. Au i J.L. Becka [8]. Zaproponowane tam podejście różni się jednak od przedstawionego w tej rozprawie. W pracy [8] nie wykorzystano metod oszacowania stałych normalizujących. Algorytm Metropolisa również zastosowano do symulacji z rozkładu prawdopodobieństwa parametrów projektowych obciążonego do obszaru awarii, ale na podstawie tak uzyskanej próby losowej określono rozkład *importance sampling* za pomocą metody *kernel density estimation* [6].

Analiza niezawodności konstrukcji

2.1. Wstęp

Reakcja konstrukcji inżynierskiej na działanie obciążeń jest uznawana za dopuszczalną, jeżeli spełnia przyjęte kryteria bezpieczeństwa i użytkowania. Wymagania stawiane konstrukcjom inżynierskim dotyczą: wytrzymałości na zniszczenie przy założonym obciążeniu, powstania uszkodzeń, wystąpienia odkształceń uniemożliwiających użytkowanie itp.. Powszechnie wymagania te nazywa się *stanami granicznymi*. Jeżeli konstrukcja jest w stanie, w którym nie spełnia założonych wymagań to mówi się, że stan graniczny jest przekroczony. W analizie niezawodności jako *awarię* określa się przekroczenie stanu granicznego. Tak zdefiniowane pojęcie awarii różni się nieco od potocznego rozumienia tego słowa. Powszechnie, następstwa awarii konstrukcji inżynierskich są utożsamiane z utratą życia, bądź znacznymi stratami materialnymi. W przypadku analizy niezawodności za awarię uznaje się także wystąpienie nadmiernych odkształceń, które zazwyczaj ustępują po zmniejszeniu obciążenia.

W tradycyjnym projektowaniu konstrukcji wykorzystuje się deterministyczne wartości parametrów projektowych. Bezpieczeństwo konstrukcji związane ze zmiennością parametrów konstrukcyjnych zapewnia się poprzez konserwatywny dobór ich wartości i poprzez uwzględnienie w równaniach stanów granicznych współczynników bezpieczeństwa.

Probabilistyczna analiza konstrukcji może być traktowana jako rozszerzenie analizy deterministycznej, poprzez wykorzystanie zmiennych losowych do reprezentacji wartości parametrów konstrukcyjnych. Takie sformułowanie pozwala na jawne uwzględnienie zmienności parametrów projektowych. W rezultacie, możliwa jest budowa modelu matematycznego, który pozwala oszacować, jakie jest prawdopodobieństwo określonego zachowania konstrukcji. Przedmiotem *analizy niezawodności* jest wyznaczenie, na podstawie probabilistycznego modelu konstrukcji, prawdopodobieństwa awarii, która jest rozumiana jako przekroczenie stanu granicznego. W dalszej części tego rozdziału omówione zostanie modelowanie konstrukcji za pomocą zmiennych losowych, metody

obliczania prawdopodobieństwa awarii oraz zostanie zdefiniowana miara bezpieczeństwa, będąca funkcją prawdopodobieństwa awarii.

2.2. Pojęcia podstawowe

Jak wspomniano, probabilistyczna analiza konstrukcji wykorzystuje modele matematyczne, w których przynajmniej część parametrów ma charakter zmiennych losowych. Poniżej zostaną przypomniane niektóre definicje z zakresu probabilistyki, pomocne do zdefiniowania modelu stosowanego w analizie niezawodności konstrukcji. Pełne, przejrzyste wprowadzenie do probabilistyki można znaleźć, między innymi, w pracy [89].

Ogólne sformułowanie problemu analizy niezawodności niezależnej od czasu zakłada, że parametry konstrukcyjne (np. wymiary, parametry materiałowe, mnożniki obciążeń), których wartości charakteryzuje niepewność, są reprezentowane przez n -wymiarową zmienną losową - wektor losowy

$$\mathbf{X} = \{X_1, X_2, \dots, X_n\}, \quad (2.1)$$

gdzie X_1, X_2, \dots, X_n są jednowymiarowymi zmiennymi losowymi (w analizie niezawodności określanymi jako *podstawowe zmienne losowe*). Wektor losowy przyjmuje wartości w n -wymiarowej przestrzeni liczb rzeczywistych \mathbf{R}^n . Funkcja P - rozkładu prawdopodobieństwa wektora losowego wyznacza jego *dystrybuantę*

$$F(\mathbf{x}) = P(\{y_1, y_2, \dots, y_n\} : y_1 < x_1, y_2 < x_2, \dots, y_n < x_n), \quad (2.2)$$

gdzie $\mathbf{x} = (x_1, x_2, \dots, x_n)$ oznacza realizację zmiennej losowej. Zmienna losowa \mathbf{X} jest typu ciągłego, jeśli istnieje nieujemna funkcja gęstości prawdopodobieństwa f taka, że

$$\forall_{x_1, x_2, \dots, x_n} F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(y_1, y_2, \dots, y_n) dy_1, dy_2, \dots, dy_n. \quad (2.3)$$

Stany graniczne w analizie niezawodności wyraża się poprzez, określoną na parametrach losowych, tzw. *uogólnioną funkcję graniczną* $g(\mathbf{X})$. Przyjmuje się, że konstrukcja spełnia założone wymagania, nie występuje przekroczenie stanu granicznego, jeśli $g(\mathbf{X}) > 0$; mówi się również, że konstrukcja jest w stanie bezpiecznym. Jeżeli $g(\mathbf{X}) = 0$ przyjmuje się, że zachodzi stan graniczny. Takie sformułowanie jest związane z bardziej precyzyjną definicją stanu granicznego niż podana we wstępie do tego rozdziału. W pracy [28] stan graniczny odpowiadający danemu warunkowi, jaki powinna spełniać

konstrukcja, określa się jako stan konstrukcji, włączając obciążenia, w którym przestaje ona spełniać ten warunek. W przypadku gdy stan graniczny jest przekroczony zachodzi $g(\mathbf{X}) < 0$. Ponadto przyjmuje się, że konstrukcja jest w stanie awarii, jeżeli $g(\mathbf{X}) \leq 0$. Tak więc w przestrzeni zmiennych losowych można zdefiniować dwa obszary: tzw. *obszar bezpieczny*

$$\Omega_s = \{\mathbf{x} : g(\mathbf{x}) > 0\} \quad (2.4)$$

oraz *obszar awarii*

$$\Omega_f = \{\mathbf{x} : g(\mathbf{x}) \leq 0\}. \quad (2.5)$$

Granicę obszaru awarii $g(\mathbf{x}) = 0$ nazywa się *powierzchnią graniczną*. *Prawdopodobieństwo awarii* jest prawdopodobieństwem zdarzenia losowego określonego przez obszar awarii:

$$P_f = P(\Omega_f) = P(g(\mathbf{X}) \leq 0). \quad (2.6)$$

Jeśli dla rozkładu prawdopodobieństwa istnieje funkcja gęstości, to prawdopodobieństwo awarii można wyrazić za pomocą całki:

$$P_f = \int_{\Omega_f} f(\mathbf{x}) d\mathbf{x}. \quad (2.7)$$

Przynależność punktu do obszaru awarii określa tzw. *funkcja charakterystyczna obszaru awarii*, którą definiuje się jako:

$$I_{\Omega_f}(\mathbf{x}) = \begin{cases} 1, & \text{jeśli } \mathbf{x} \in \Omega_f; \\ 0, & \text{w przeciwnym przypadku.} \end{cases} \quad (2.8)$$

Korzystając z (2.8), prawdopodobieństwo awarii określa się jako *wartość oczekiwana* funkcji charakterystycznej obszaru awarii

$$P_f = \mathbb{E}(I_{\Omega_f}(\mathbf{X})) = \int_{R^n} I_{\Omega_f}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (2.9)$$

W literaturze czasami przyjmuje się założenie, że funkcja graniczna g jest określona tak, że dla powierzchni awarii $\Omega_0 = \{\mathbf{x} : g(\mathbf{x}) = 0\}$ zachodzi $P(\mathbf{X} \in \Omega_0) = 0$. Przyjęcie tego założenia rozwiązuje dylemat czy wystąpienie stanu granicznego uznać za awarię, czy za stan bezpieczny. Wtedy bowiem prawdopodobieństwo obszaru awarii określonego wzorem (2.5) oraz prawdopodobieństwo $\Omega_f = \{\mathbf{x} : g(\mathbf{x}) < 0\}$ są sobie równe. Ważną cechą prawdopodobieństwa awarii zdefiniowanego przez (2.6) jest niezmienniczość

względem postaci funkcji granicznej. Zauważmy, że stan graniczny można wyrazić za pomocą różnych funkcji g , które przyjmują wartości dodatnie, jeżeli konstrukcja jest w stanie bezpiecznym, a ujemne w przypadku awarii.

Konstrukcje inżynierskie projektowane są tak, aby zapewnić jak najwyższy poziom bezpieczeństwa, w praktyce prawdopodobieństwo awarii przyjmuje więc bardzo małe wartości; na przykład z przedziału między $9.96 \cdot 10^{-8}$ a $1.59 \cdot 10^{-1}$. Ponieważ raczej trudno posługiwać się wielkościami z tego zakresu jako miarę bezpieczeństwa wprowadzono tzw. *wskaźnik niezawodności*

$$\beta = -\Phi^{-1}(P_f), \quad (2.10)$$

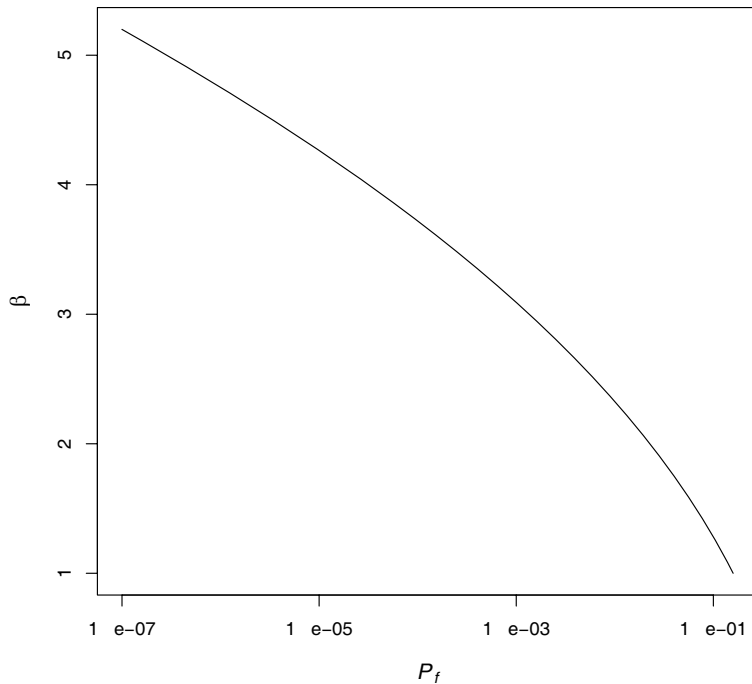
gdzie Φ^{-1} jest funkcją odwrotną dystrybuanty *standardowego rozkładu normalnego* (rozkładu normalnego o zerowej wartości średniej i jednostkowej wariancji). W przypadku podanych wyżej wartości prawdopodobieństwa awarii, β wynosi odpowiednio 5.2 oraz 1. Wykres przedstawiający zależność β od P_f przedstawia rysunek 2.1.

Przykład 2.1: Elementarne zadanie analizy niezawodności. Pojęcia związane z analizą niezawodności często przedstawia się na przykładzie prostych konstrukcji takich, jak rozciągany pręt lub proste ustroje belkowe. Rozpatrzmy belkę wspornikową o długości l i symetrycznym przekroju ze wskaźnikiem wytrzymałości na zginanie W (rys. 2.2). Przyjmijmy, że siła skupiona P działająca na swobodnym końcu belki, oraz granica plastyczności materiału R są modelowane niezależnymi zmiennymi losowymi X_P oraz X_R o rozkładzie logarytmiczno-normalnym. Załóżmy, że wymaga się, aby naprężenia sprężyste powstałe w wyniku zginania nie przekroczyły granicy plastyczności. Ten stan graniczny można wyrazić za pomocą funkcji $g(X_P, X_R) = WX_R - lX_P$, lecz również $g^*(X_P, X_R) = WX_R/lX_P - 1$. Funkcje te mają różne własności (rys. 2.3), ale w przestrzeni zmiennych losowych, gdzie łączna funkcja gęstości prawdopodobieństwa jest niezerowa, wyznaczają taki sam obszar awarii:

$$\begin{aligned} \Omega_f &= \{(x_P, x_R) : g(x_P, x_R) \leq 0 \wedge f(x_P, x_R) > 0\}, \\ &= \{(x_P, x_R) : g^*(x_P, x_R) \leq 0 \wedge f(x_P, x_R) > 0\}. \end{aligned} \quad (2.11)$$

Funkcja $f(x_P, x_R)$ w (2.11) jest łączną gęstością prawdopodobieństwa zmiennych losowych X_P i X_R . Prawdopodobieństwo awarii (2.7) zależy od obszaru awarii, więc dla obu funkcji granicznych jest równe.

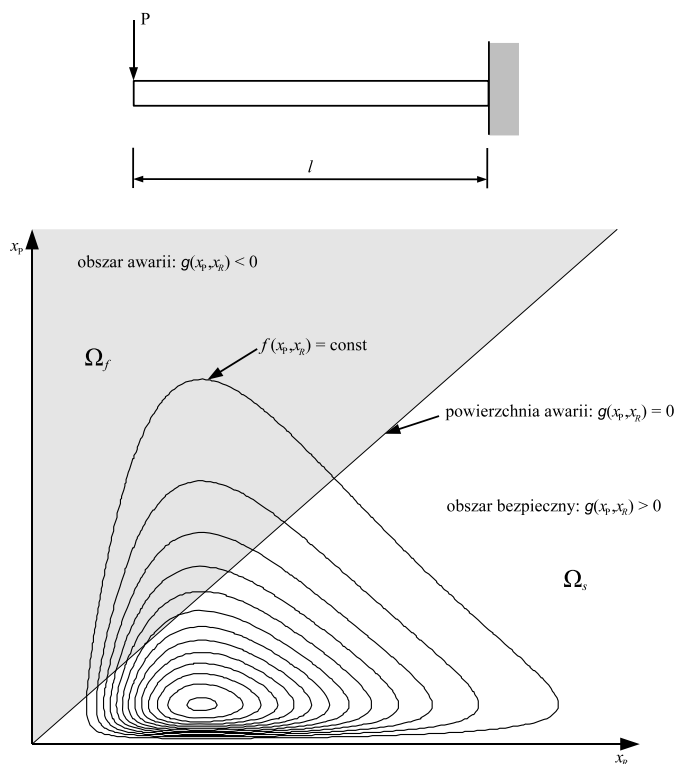
W powyższym modelu pręta wspornikowego przyjęto, że dwa parametry konstrukcyjne mają charakter zmiennych losowych. Przede wszystkim zrobiono tak, ze względu na łatwość graficznej prezentacji dwuwymiarowej przestrzeni zmiennych losowych. Jednak nic nie stoi na przeszkodzie, aby przyjąć, że pozostałe parametry konstrukcyjne mają również charakter zmiennych losowych. Zakładając dodatkowo, że długość wspornika l i wskaźnik wytrzymałości jego przekroju na zginanie W są zmiennymi losowymi, otrzymuje się czterowymiarowe, uogólnione zadanie niezawodności. Rozważany problem analizy niezawodności



RYSUNEK 2.1. Zależność między wskaźnikiem niezawodności β a prawdopodobieństwem awarii P_f .

można również sformułować w nieco inny sposób. Zdefiniujmy następujące funkcje podstawowych zmiennych losowych: $Y_R(X_J, X_R) = X_W X_R$ - wytrzymałości na zginanie, oraz $Y_P(X_L, X_P) = X_L X_P$ - momentu zginającego. Zmienne losowe X_W i X_L modelują odpowiednio wskaźnik wytrzymałości przekroju na zginanie oraz długość pręta. Oczywiście funkcje Y_R i Y_P również są zmiennymi losowymi.

Analizowany stan graniczny przekroczenia granicy plastyczności przez naprężenia sprężyste można wyrazić za pomocą funkcji $\tilde{g}(Y_R, Y_P) = Y_R - Y_P$. Przy założeniu, że Y_R i Y_P



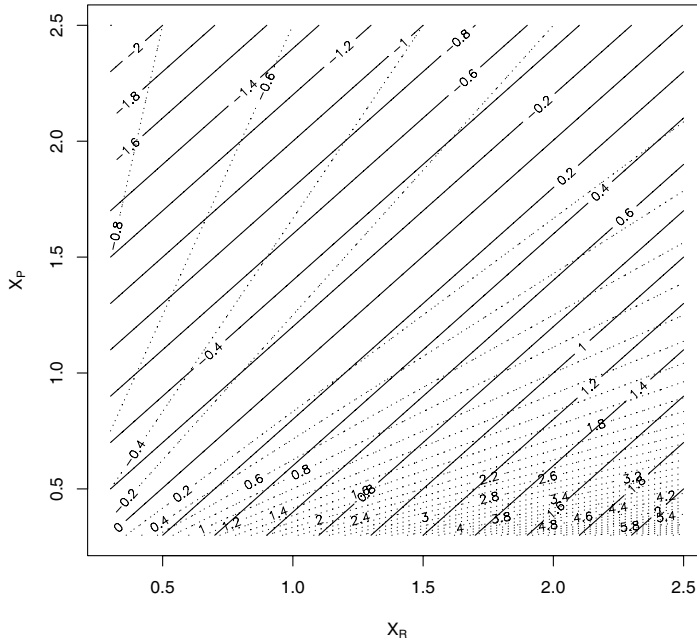
RYSUNEK 2.2. Ilustracja zadania analizy niezawodności pręta wspornikowego.

są niezależnymi zmiennymi losowymi, prawdopodobieństwo awarii (2.7) ma postać:

$$P_f = \int_{-\infty}^{\infty} \int_{-\infty}^{y_P \leq y_R} f_P(y_P) f_R(y_R) dy_P dy_R, \quad (2.12)$$

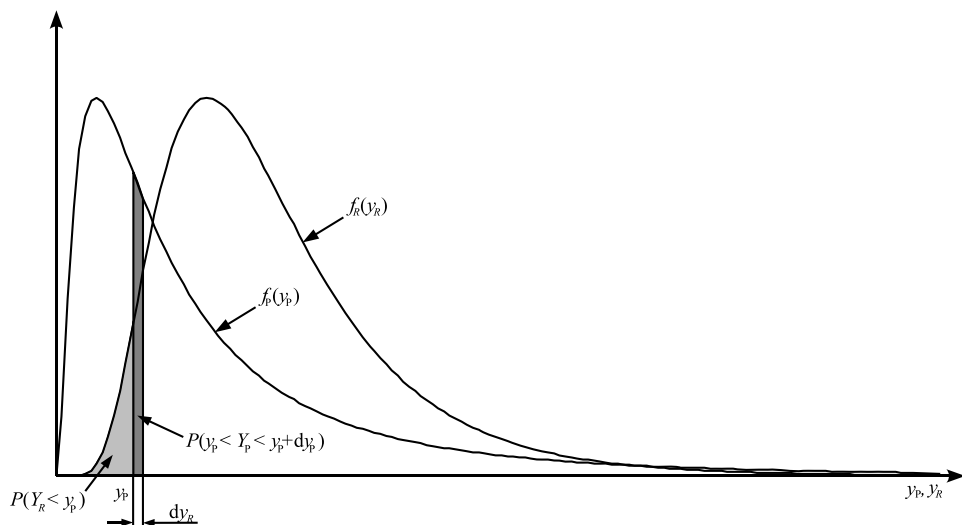
gdzie f_P , f_R są gęstościami rozkładów prawdopodobieństwa, odpowiednio, zmiennych Y_P i Y_R . Korzystając z F_R - dystrybuanty (2.2) zmiennej Y_R , wzór (2.12) można przekształcić do postaci:

$$P_f = \int_{-\infty}^{\infty} F_R(y_P) f_P(y_P) dy_P. \quad (2.13)$$



RYSUNEK 2.3. Funkcje stanów granicznych: $g(X_P, X_R) = WX_R - lX_P$ - linie ciągłe, $g^*(X_P, X_R) = WX_R/lX_P - 1$ - linie przerywane.

Graficzną reprezentację tej całki przedstawia rysunek 2.4, jak widać całkowany jest iloczyn prawdopodobieństw $P(Y_R < y_P)$ i $P(y_P < Y_P < y_P + dy_P)$. Sformułowanie problemu niezawodności z wykorzystaniem zmiennych losowych modelujących charakterystyki wytrzymałościowe oraz odpowiadające im siły wewnętrzne nosi nazwę *podstawowego problemu niezawodności*. Ciekawą cechą tego sformułowania jest zastąpienie wielowymiarowej całki (2.9) całką (2.13), która jest określona w jednym wymiarze. Jak zobaczymy w dalszej części pracy obliczanie całek wielowymiarowych jest trudnym problemem numerycznym. Wykorzystanie sformułowania (2.13) może się więc wydawać atrakcyjną alternatywą wobec (2.9). Zwróćmy jednak uwagę, że zastosowanie (2.13) wymaga określenia rozkładów prawdopodobieństwa zmiennych Y_R i Y_P , co w większości przypadków jest problemem równie złożonym, jak obliczenie całki (2.9).



RYSUNEK 2.4. Graficzna interpretacja wzoru (2.13).

2.3. Metody numeryczne analizy niezawodności

Obliczenie całki we wzorze (2.9) w praktyce okazuje się być trudnym zadaniem. Zastosowanie metod analitycznych jest możliwe tylko w szczególnych przypadkach. Zaskakującym jest fakt, że ograniczone zastosowanie mają również standardowe metody całkowania numerycznego, jak metoda trapezów czy metody oparte na wielomianach. Koszty numeryczne wykorzystania kwadratur rosną bowiem wykładniczo wraz ze wzrostem liczby wymiarów przestrzeni, w której określona jest całka. Uważa się, że kwadratury można stosować, jeśli przestrzeń zmiennych losowych ma co najwyżej pięć wymiarów. Zbudowanie realistycznego modelu konstrukcji wymaga zazwyczaj wykorzystania znacznie większej od pięciu liczby parametrów losowych. Popularną metodą całkowania funkcji określonych w dużej liczbie wymiarów jest *symulacja Monte Carlo*. Znajduje ona szerokie zastosowanie w analizie niezawodności. Ze względu na małe wartości prawdopodobieństwa awarii zazwyczaj wykorzystuje się metodę *importance sampling*, będącą przykładem metod Monte Carlo o zmniejszonej wariancji. Do oszacowania prawdopodobieństwa awarii stosuje się również metody aproksymacyjne, opierające się na szczególnych własnościach *gaussowskiej przestrzeni standardowej*, do której transformuje się problem. Transformacja przestrzeni zmiennych losowych do

gaussowskiej przestrzeni standardowej bywa również stosowana w połączeniu z metodami symulacyjnymi, od jej omówienia rozpoczyna się więc podrozdział przedstawiający metody obliczania prawdopodobieństwa awarii.

2.3.1. Gaussowska przestrzeń standardowa

Wielowymiarową zmienną losową, którą tworzą niezależne zmienne normalne U_i , $i = 1, \dots, n$, o zerowych wartościach średnich i jednostkowych wariancjach, nazywa się *gaussowską przestrzenią standardową*; będziemy ją oznaczali $\mathbf{U} = \{U_1, U_2, \dots, U_n\}$. Zmienna losowa \mathbf{U} jest wielowymiarową zmienną normalną o zerowym wektorze wartości średnich i jednostkowej macierzy kowariancji. Funkcja łącznej gęstości prawdopodobieństwa gaussowskiej przestrzeni standardowej ma następującą postać:

$$\varphi_u(\mathbf{u}) = \varphi(\mathbf{u}, \mathbf{0}, \mathbf{I}) = \prod_{i=1}^n \varphi(u_i, 0, 1) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u_i^2\right), \quad (2.14)$$

gdzie $\varphi(\mathbf{u}, \mathbf{0}, \mathbf{I})$ oznacza funkcję łącznej gęstości prawdopodobieństwa rozkładu normalnego w przypadku realizacji zmiennej losowej \mathbf{u} , zerowego wektora wartości oczekiwanych $\mathbf{0}$ i jednostkowej macierzy kowariancji \mathbf{I} .

Funkcja φ_u jest obrotowo symetryczna względem początku układu współrzędnych, a jej wartość maleje wykładniczo wraz z odległością od tego punktu. Z powyższych własności wynika, że w zbiorze $\mathcal{F} \in \mathbf{U}$ maksimum gęstości prawdopodobieństwa znajduje się w punkcie najbliższym początkowi układu współrzędnych:

$$\max_{\mathbf{u} \in \mathcal{F}} \varphi_u(\mathbf{u}) = \varphi_u(\min_{\mathbf{u} \in \mathcal{F}} |\mathbf{u}|), \quad (2.15)$$

gdzie $|\mathbf{u}| = (\sum_{i=1}^n u_i^2)^{1/2}$ jest normą euklidesową wektora \mathbf{u} .

Niech dana będzie funkcja liniowa

$$g_\beta(\mathbf{u}) = \beta - \sum_{i=1}^n \alpha_i u_i, \quad (2.16)$$

gdzie $\sum_{i=1}^n \alpha_i^2 = 1$ oraz $\beta > 0$. W przypadku (2.16) nierówność określającą obszar awarii $g_\beta(\mathbf{u}) \leq 0$ można zapisać jako

$$\beta \leq \sum_{i=1}^n \alpha_i u_i. \quad (2.17)$$

Ze względu na obrotową symetrię funkcji gęstości prawdopodobieństwa gaussowskiej przestrzeni standardowej, prawdopodobieństwo obszaru awarii określonego przez (2.17)

spełnia następującą tożsamość

$$P\left(\beta \leq \sum_{i=1}^n \alpha_i U_i\right) = P(\beta \leq U), \quad (2.18)$$

gdzie U jest zmienną normalną o zerowej wartości oczekiwanej i jednostkowej wariancji. Prawdopodobieństwo w (2.18) jest określone następującym wyrażeniem:

$$P(\beta \leq U) = \Phi(-\beta), \quad (2.19)$$

gdzie, jak wspomniano wcześniej, Φ jest dystrybuantą rozkładu normalnego. Tak więc prawdopodobieństwa obszaru awarii ograniczonego hiperpłaszczyzną w gaussowskiej przestrzeni standardowej, można oszacować wyznaczając wartość dystrybuanty rozkładu normalnego, zamiast całkować funkcję wielu zmiennych.

Wykorzystanie omówionych własności gaussowskiej przestrzeni standardowej jest podstawą metod aproksymacyjnych stosowanych w analizie niezawodności. Zmienne losowe występujące w modelach analizy niezawodności mogą mieć oczywiście różne rozkłady, jak również mogą być skorelowane. Przeprowadzenie obliczeń w gaussowskiej przestrzeni standardowej wymaga transformacji przestrzeni zmiennych losowych występującej w analizowanym modelu.

2.3.2. Transformacja wektora losowego do gaussowskiej przestrzeni standardowej

Wymaga się, aby stosowana w analizie niezawodności transformacja $T : \mathbf{X} \rightarrow \mathbf{U}$ podstawowych zmiennych losowych \mathbf{X} do gaussowskiej przestrzeni standardowej \mathbf{U} , zachowywała prawdopodobieństwo obszaru awarii. Warunek ten zapisuje się jako

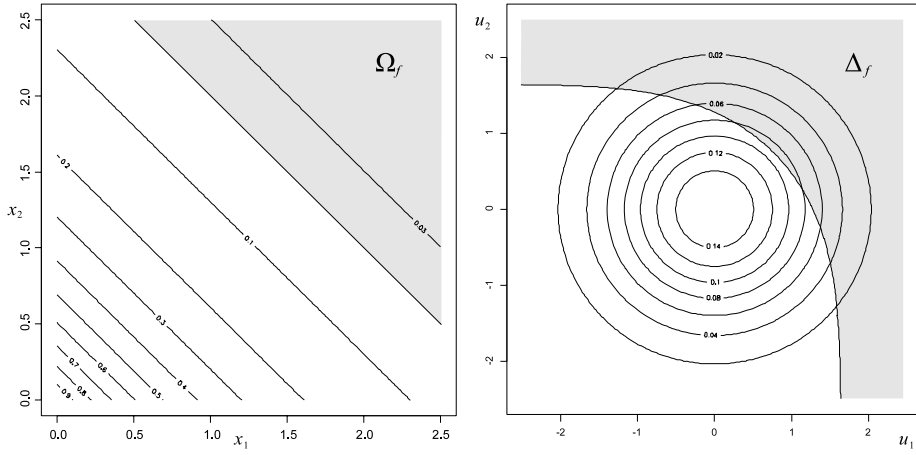
$$P_f = \int_{\Omega_f} f(\mathbf{x}) d\mathbf{x} = \int_{\Delta_f} \varphi_u(\mathbf{u}) d\mathbf{u}, \quad (2.20)$$

gdzie Δ_f jest obszarem awarii w gaussowskiej przestrzeni standardowej. Obszar Δ_f jest określony przez następującą transformację

$$\Omega_f = \{\mathbf{x} : g(\mathbf{x}) \leq 0\} \longrightarrow \Delta_f = \{\mathbf{u} : G(\mathbf{u}) \leq 0\}, \quad (2.21)$$

gdzie $G(\mathbf{U})$ jest funkcją graniczną w gaussowskiej przestrzeni standardowej, określoną przez odwrotną transformację zmiennych losowych $\mathbf{X} = T^{-1}(\mathbf{U})$

$$G(\mathbf{u}) = g(T^{-1}(\mathbf{u})). \quad (2.22)$$



RYSUNEK 2.5. Transformacja obszaru awarii. Obszar awarii Ω_f wyznaczony przez liniową funkcję awarii w przestrzeni \mathbf{X} z niezależnym rozkładem wykładniczym jest transformowany na obszar Δ_f o silnie nieliniowej powierzchni awarii w gaussowskiej przestrzeni standardowej \mathbf{U} . Przykład zaczerpnięty ze [103].

Nieliniowe zależności między rozkładami prawdopodobieństwa powodują zmianę własności geometrycznych obszaru awarii w wyniku transformacji (rys. 2.5).

Jeżeli składowe wektora losowego \mathbf{X} są nieskorelowane, można je przetransformować niezależnie korzystając z tożsamości $\Phi(u_i) = F_{X_i}(x_i)$, gdzie F_{X_i} jest dystrybuantą rozkładu prawdopodobieństwa zmiennej losowej X_i . Tak więc transformacja zmiennej losowej X_i na zmienną losową U_i jest dana wzorem $u_i = \Phi^{-1}(F_{X_i}(x_i))$, a transformacja odwrotna do niej $x_i = F_{X_i}^{-1}(\Phi(u_i))$.

W przypadku, kiedy składowe wektora losowego są skorelowane można zastosować tzw. transformację Rosenblatta [100], którą jako pierwsi w analizie niezawodności wykorzystali Hohenbichler i Rackwitz [54]. Transformacja Rosenblatta przekształca kolejne składowe zmiennej losowej zgodnie z następującą sekwencją

$$\begin{aligned}
 \Phi(u_1) &= F_{X_1}(x_1) , \\
 \Phi(u_2) &= F_{X_2}(x_2|x_1) , \\
 \Phi(u_3) &= F_{X_3}(x_3|x_1, x_2) , \\
 &\vdots \\
 \Phi(u_n) &= F_{X_n}(x_n|x_1, x_2, \dots, x_{n-1}) ,
 \end{aligned} \tag{2.23}$$

gdzie $F_{X_i}(x_i|x_1, x_2, \dots, x_{n-1})$ oznacza dystrybuantę warunkowego rozkładu prawdopodobieństwa X_i pod warunkiem $\{X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}\}$. Jeżeli znana jest funkcja łącznej gęstości prawdopodobieństwa f zmiennej \mathbf{X} , to wtedy F_{X_i} można wyrazić w następujący sposób

$$F_{X_i}(x_i|x_1, x_2, \dots, x_{i-1}) = \frac{\int_{-\infty}^{x_i} f_{X_i}(x_1, x_2, \dots, x_{i-1}, t) dt}{f_{X_{i-1}}(x_1, x_2, \dots, x_{i-1})}, \quad (2.24)$$

gdzie $f_{\mathbf{X}_j}(x_1, x_2, \dots, x_j)$ jest funkcją gęstości prawdopodobieństwa rozkładu brzegowego

$$f_{\mathbf{X}_j}(x_1, x_2, \dots, x_j) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_{j+1}, \dots, dx_n. \quad (2.25)$$

Tak zdefiniowana transformacja nie jest jednoznaczna. Jak zauważył Rosenblatt [100] kolejność transformacji składowych wektora losowego można wybrać na $n!$ sposobów. Kolejność, w jakiej wykonywana jest transformacja, ma wpływ na kształt obszaru awarii w gaussowskiej przestrzeni standardowej, od którego zależy efektywność algorytmów analizy niezawodności [54]. Jeżeli funkcja F_{X_i} nie ma prostej postaci, to transformacja wymaga zastosowania całkowania numerycznego.

2.3.3. Metoda analizy niezawodności pierwszego rzędu (FORM)

Metoda *FORM* (ang. *First Order Reliability Method*) [50] opiera się na wykorzystaniu własności gaussowskiej przestrzeni standardowej. Zagadnienie analizy niezawodności w gaussowskiej przestrzeni standardowej \mathbf{U} sformułujemy posługując się funkcją graniczną $G(\mathbf{U})$. Obszar awarii jest więc określony jako $\Delta_f = \{\mathbf{u} : G(\mathbf{u}) \leq 0\}$, a prawdopodobieństwo awarii wyraża całka

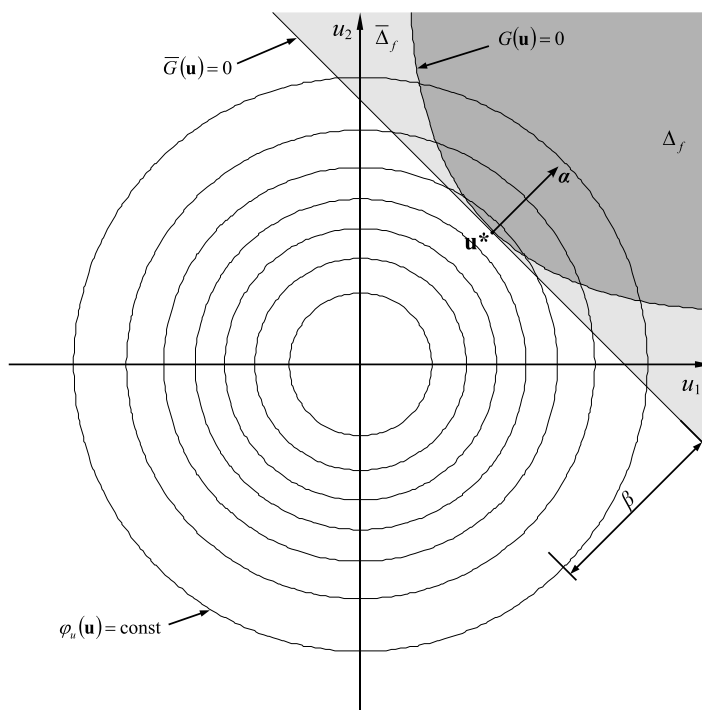
$$P_f = \int_{\Delta_f} \varphi_{\mathbf{u}}(\mathbf{u}) d\mathbf{u}. \quad (2.26)$$

Oszacowanie prawdopodobieństwa awarii metodą FORM wymaga określenia punktu, który jest najbliższy początkowi układu współrzędnych - tzw. *punktu projektowego*:

$$\mathbf{u}^* = \arg \min_{\mathbf{u} \in \Delta_f} \|\mathbf{u}\|. \quad (2.27)$$

Przybliżenie prawdopodobieństwa awarii otrzymuje się na podstawie wyrażenia

$$P_f \approx \Phi(-\beta), \quad (2.28)$$



RYSUNEK 2.6. Przybliżenie obszaru awarii metodą FORM.

gdzie β jest odległością punktu projektowego od środka układu współrzędnych

$$\beta = \|\mathbf{u}^*\|. \quad (2.29)$$

Metoda FORM zamienia więc problem całkowania (2.26) na problem optymalizacji (2.27), który można rozwiązać za pomocą dowolnego algorytmu optymalizacji z ograniczeniami nierównościami.

Zauważmy, że równanie (2.29) sugeruje interpretację indeksu niezawodności (2.10) jako najmniejszej odległości obszaru awarii od początku układu współrzędnych w gausowskiej przestrzeni standardowej.

Przybliżenie prawdopodobieństwa awarii (2.28) opiera się na linearyzacji funkcji

granicznej G w punkcie projektowym:

$$\bar{G}(\mathbf{u}) = \beta - \sum_{i=1}^n \alpha_i u_i, \quad (2.30)$$

gdzie α jest wektorem jednostkowym o kierunku przeciwnym do kierunku gradientu funkcji G w punkcie \mathbf{u}^* :

$$\alpha = - \frac{\nabla G(\mathbf{u})}{\|\nabla G(\mathbf{u})\|} \Big|_{\mathbf{u}=\mathbf{u}^*}. \quad (2.31)$$

Na podstawie własności gaussowskiej przestrzeni standardowej (2.18) prawdopodobieństwo zdarzenia, zdefiniowanego przez półprzestrzeń aproksymującą obszar awarii $\bar{\Delta}_f = \{\mathbf{u} : \bar{G}(\mathbf{u}) \leq 0\}$, określa wzór (2.19), który jest równoważny (2.28).

2.3.4. Metoda analizy niezawodności drugiego rzędu (SORM)

Algorytmy analizy niezawodności, wykorzystujące rozwinięcie funkcji granicznej w szereg Taylora drugiego rzędu, określane są jako metody *SORM* (ang. *Second Order Reliability Method*). Poniżej zostanie przedstawione najpopularniejsze sformułowanie SORM, bazujące na asymptotycznej aproksymacji prawdopodobieństwa obszaru awarii określonego przez przybliżenie drugiego rzędu funkcji granicznej [15].

W metodzie SORM funkcję graniczną aproksymuje się wokół punktu projektowego \mathbf{u}^* zdefiniowanego, tak jak w przypadku metody FORM (por. (2.27)). Zakłada się, że funkcja graniczna jest ciągła i co najmniej dwukrotnie różniczkowalna w otoczeniu punktu projektowego. Znając punkt projektowy wykonuje się transformację ortogonalną przestrzeni \mathbf{U} do przestrzeni \mathbf{V} takiej, że punkt projektowy ma współrzędne $\mathbf{v}^* = [0, 0, \dots, 0, \beta]$, a mieszane drugie pochodne transformowanej funkcji granicznej $G_v(\mathbf{v})$ w tym punkcie są zerowe. Założmy istnienie funkcji f takiej, że

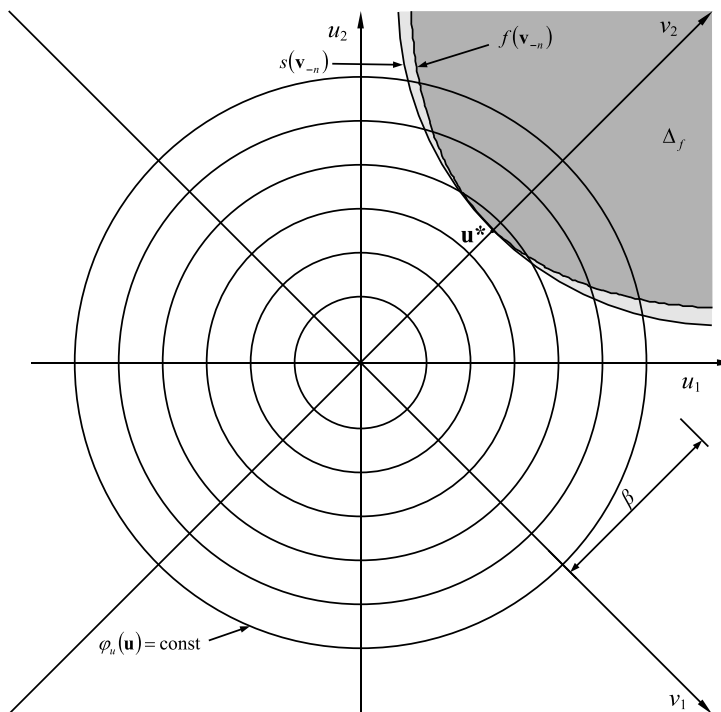
$$G_v(\mathbf{v}_{-n}, f(\mathbf{v}_{-n})) = 0, \quad (2.32)$$

gdzie $\mathbf{v}_{-n} = (v_1, v_2, \dots, v_{n-1})$. Przyjmijmy ponadto, że jeżeli \mathbf{v} jest zawarte w obszarze awarii

$$\Delta_f = \{\mathbf{v} : G_v(\mathbf{v}) < 0\}, \quad (2.33)$$

to zachodzi $v_n > f(\mathbf{v}_{-n})$. Rozwinięcie drugiego rzędu funkcji f w punkcie \mathbf{v}^* ma następującą postać

$$f(\mathbf{v}_{-n}) \approx s(\mathbf{v}_{-n}) = \beta + \frac{1}{2} \sum_{i=1}^{n-1} \kappa_i v_i^2, \quad (2.34)$$



RYSUNEK 2.7. Przybliżenie obszaru awarii metodą SORM.

gdzie κ_i , $i = 1, \dots, n-1$ są krzywiznami głównymi powierzchni awarii $G_v(\mathbf{v})$ w punkcie projektowym. Krzywizny κ_i mają następującą postać

$$\kappa_i = \left. \frac{\partial^2 f(\mathbf{v}_{-n})}{\partial v_i^2} \right|_{\mathbf{v}_{-n}=\mathbf{0}} \quad i = 1, 2, \dots, n-1. \quad (2.35)$$

Korzystając z przybliżenia powierzchni granicznej (2.34) prawdopodobieństwo awarii

wyraża się jako

$$\begin{aligned}
P_f &= \int_{\mathbf{R}^n} I_{\Delta_f}(\mathbf{v}) \varphi_u(\mathbf{v}_{-n}) \varphi_u(v_n) d\mathbf{v}_{-n} dv_n \\
&= \int_{\mathbf{R}^{n-1}} \int_{f(\mathbf{v}_{-n})}^{\infty} \varphi_u v_n dv_n \varphi_u(\mathbf{v}_{-n}) d\mathbf{v}_{-n} \\
&\approx \int_{\mathbf{R}^{n-1}} \int_{s(\mathbf{v}_{-n})}^{\infty} \varphi_u v_n dv_n \varphi_u(\mathbf{v}_{-n}) d\mathbf{v}_{-n} \\
&= \int_{\mathbf{R}^{n-1}} \Phi(-s(\mathbf{v}_{-n})) \varphi_u(\mathbf{v}_{-n}) d\mathbf{v}_{-n} \\
&= \int_{\mathbf{R}^{n-1}} \Phi\left(-\beta - \frac{1}{2} \sum_{i=1}^{n-1} \kappa_i v_i^2\right) \varphi_u(\mathbf{v}_{-n}) d\mathbf{v}_{-n}. \tag{2.36}
\end{aligned}$$

Następnie, za pomocą rozwinięcia w szereg Taylora wyrażenia:

$$\ln \Phi(-\beta - x) \approx \ln \Phi(-\beta) - x\psi(-\beta), \tag{2.37}$$

gdzie $\psi(-\beta) = \varphi_u(-\beta)/\Phi(-\beta)$, oszacowanie (2.36) przekształca się do postaci

$$\begin{aligned}
P_f &\approx \Phi(-\beta) \int_{\mathbf{R}^{n-1}} \exp\left(-\frac{1}{2}\psi(-\beta) \sum_{i=1}^{n-1} \kappa_i v_i^2\right) \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v_i^2\right) d\mathbf{v}_{-n} \\
&= \Phi(-\beta) \prod_{i=1}^{n-1} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v_i^2(1 + \kappa_i\psi(-\beta))\right) dv_i \\
&= \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \kappa_i\psi(-\beta))^{-\frac{1}{2}}. \tag{2.38}
\end{aligned}$$

Powyższy wzór daje dobre oszacowanie prawdopodobieństwa awarii, jeżeli $|\beta| \geq 1$ oraz $1 - \psi(-\beta)\kappa_i > 0$. Wzór (2.38) został zaproponowany przez Hohenbichlera i Rackwitz [55] a jest on modyfikacją oszacowania podanego przez Breitunga [14], asymptotycznie zbieżnego w przypadku $\beta \rightarrow \infty$:

$$P_f \approx \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \kappa_i\beta)^{-\frac{1}{2}}. \tag{2.39}$$

Porównując graficzne interpretacje metod FORM i SORM (rys. 2.6, 2.7) można zauważyć, że oszacowanie drugiego rzędu różni się od oszacowania pierwszego rzędu uwzględnieniem krzywizn powierzchni awarii.

2.3.5. Metoda Monte Carlo

Terminem *metody Monte Carlo* określane są algorytmy, w których do rozwiązania problemu wykorzystuje się analizę statystyczną próby losowej otrzymanej w wyniku symulacji. Metody te są powszechnie stosowane do numerycznego całkowania funkcji wielowymiarowych zmiennych losowych; do tej klasy problemów należy również oszacowanie prawdopodobieństwa awarii (2.6). Metoda Monte Carlo w przypadku analizy niezawodności polega na generowaniu, zgodnie z przyjętym rozkładem prawdopodobieństwa, realizacji parametrów konstrukcyjnych o charakterze losowym. Następnie sprawdza się, czy w przypadku otrzymanych realizacji spełniony jest stan graniczny. Przybliżenie prawdopodobieństwa awarii otrzymuje się na podstawie oszacowania wartości oczekiwanej funkcji charakterystycznej obszaru awarii, otrzymanego za pomocą estymatora właściwego w przypadku przyjętej metody symulacji. Efektywne wykorzystanie Monte Carlo w analizie niezawodności wymaga rozwiązania następujących problemów:

- opracowania efektywnej metody symulacji losowych parametrów konstrukcyjnych (wiarygodne oszacowanie prawdopodobieństwa awarii powinno być możliwe na podstawie jak najmniejszej liczby symulacji);
- oceny wpływu, jaki na przyjętą metodę symulacji ma złożoność obliczeniowa funkcji granicznej,
- przyjęcia metody szacowania prawdopodobieństwa awarii na podstawie uzyskanych symulacji;
- oszacowania błędu zastosowanego oszacowania.

Poniżej zostaną przedstawione, często stosowane w analizie niezawodności, odmiany metody Monte Carlo.

2.3.6. Klasyczne całkowanie Monte Carlo

Klasyczna metoda Monte Carlo wykorzystuje niezależną próbę losową $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}\}$, otrzymaną poprzez generowanie liczb losowych bezpośrednio z rozkładu prawdopodobieństwa zmiennej losowej modelującej parametry konstrukcji - podstawowych zmiennych losowych. Oszacowanie prawdopodobieństwa awarii, traktowanego jako wartość przeciętna funkcji charakterystycznej obszaru awarii (2.9), uzyskuje się za pomocą estymatora

$$\bar{P}_{fn} = \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}^{(i)}), \quad (2.40)$$

który jest nazywany *empiryczną wartością przeciętną*. Na podstawie *Mocnego Prawa Wielkich Liczb* ciąg \bar{P}_{f_n} jest zbieżny z prawdopodobieństwem 1 do $E(I_{\Omega_f}(\mathbf{X}))$. Błąd estymatora (2.40) wyraża się za pomocą jego wariancji, którą w przypadku funkcji charakterystycznej obszaru awarii określa wyrażenie:

$$V_{P_{f_n}} = \text{Var}(\bar{P}_{f_n}) = \frac{1}{n} \int_{\mathbf{R}^n} (I_{\Omega_f}(\mathbf{x}) - E[I_{\Omega_f}(\mathbf{X})])^2 f(\mathbf{x}) dx \quad (2.41)$$

$$= \frac{1}{n} \left(E[I_{\Omega_f}(\mathbf{X})^2] - E[I_{\Omega_f}(\mathbf{X})]^2 \right). \quad (2.42)$$

Oszacowanie $V_{P_{f_n}}$ na podstawie próby losowej $\{\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}\}$ można otrzymać za pomocą estymatora

$$\bar{V}_{P_{f_n}} = \frac{1}{n^2} \sum_{i=1}^n \left(I_{\Omega_f}(\mathbf{X}^{(i)}) - \bar{P}_{f_n} \right)^2. \quad (2.43)$$

Dokładność oszacowania (2.40) często jest określana przy wykorzystaniu tzw. *współczynnika zmienności estymatora*, zdefiniowanego jako:

$$\hat{e} = \frac{(\bar{V}_{P_{f_n}})^{\frac{1}{2}}}{\bar{P}_{f_n}}. \quad (2.44)$$

Alternatywą dla wyznaczenia oszacowania punktowego i jego błędu jest określenie tzw. *przedziału ufności*, w którym z przyjętym poziomem ufności zawiera się prawdopodobieństwo awarii P_f . Przedział ufności estymatora (2.40) można wyznaczyć korzystając z faktu, że zmienna losowa

$$Y_n = \frac{\frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}^{(i)}) - E[I_{\Omega_f}(\mathbf{X})]}{\sqrt{\bar{V}_{P_{f_n}}}}, \quad (2.45)$$

na mocy *Centralnego Twierdzenia Granicznego*, ma asymptotycznie standardowy rozkład normalny $\mathcal{N}(0, 1)$. Niech C będzie przyjętym poziomem ufności, a k - kwantylem rozkładu $\mathcal{N}(0, 1)$ rzędu $\frac{1}{2}(1 - C)$, wtedy dla $n \rightarrow \infty$ zachodzi

$$P(k \leq Y_n \leq -k) = C. \quad (2.46)$$

Mając oszacowanie prawdopodobieństwa awarii (2.40) oraz oszacowanie jego wariancji (2.43), na podstawie wzorów (2.43) i (2.46), otrzymujemy przybliżenie przedziału ufności prawdopodobieństwa awarii:

$$P\left(\bar{P}_{f_n} + k\sqrt{\bar{V}_{P_{f_n}}} \leq P_f \leq \bar{P}_{f_n} - k\sqrt{\bar{V}_{P_{f_n}}}\right) = C. \quad (2.47)$$

Liczbę symulacji jaką należy wykonać, aby otrzymać estymator o żądanym współczynniku zmienności, można określić na podstawie wzoru

$$\hat{e} = \sqrt{\frac{1 - \bar{P}_{f_n}}{n\bar{P}_{f_n}}}, \quad (2.48)$$

który otrzymuje się podstawiając do (2.44) wyrażenie (2.42). Jak widać, liczba koniecznych symulacji nie zależy od liczby wymiarów zmiennej losowej, co jest jedną z zalet klasycznej metody Monte Carlo. Błąd tej metody jest jednak odwrotnie proporcjonalny do wielkości prawdopodobieństwa, które ma być oszacowane. Aby osiągnąć poziom zmienności estymatora $\hat{e} = 0.1$ dla $P_f = 3.16 \cdot 10^{-5}$, co odpowiada wskaźnikowi niezawodności $\beta = 4$, należy wykonać ponad $3 \cdot 10^6$ symulacji. Szczególnie w przypadku, gdy obliczenie funkcji awarii wiąże się z wykonaniem analizy metodą elementów skończonych, koszty numeryczne klasycznej metody Monte Carlo są bardzo duże. Pomimo tego, klasyczne sformułowanie Monte Carlo znajduje zastosowanie praktyczne ze względu na prostotę implementacji.

2.3.7. Metoda importance sampling

Klasyczna metoda Monte Carlo w analizie niezawodności wykorzystuje symulacje otrzymane z rozkładu prawdopodobieństwa losowych parametrów konstrukcji. Prawdopodobieństwo awarii można również oszacować na podstawie próbek z innych rozkładów prawdopodobieństwa. Okazuje się, że odpowiedni dobór rozkładu prawdopodobieństwa do rozwiązywanego problemu pozwala znacznie zmniejszyć liczbę symulacji, potrzebnych do otrzymania oszacowania o wymaganej dokładności. Metoda wykorzystująca to podejście jest nazywana *importance sampling*, jej sformułowanie przedstawione poniżej dotyczy analizy niezawodności, ale znajduje ona powszechne zastosowanie w problemach, gdzie konieczne jest oszacowanie wartości oczekiwanej funkcji zmiennych losowych [53].

Oszacowanie prawdopodobieństwa awarii (2.9) za pomocą metody importance sampling otrzymuje się na podstawie niezależnej próby losowej $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}\}$, uzyskanej z rozkładu prawdopodobieństwa o gęstości q , korzystając z estymatora:

$$E(I_{\Omega_f}(\mathbf{X})) \approx \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}^{(i)}) \frac{f(\mathbf{X}^{(i)})}{q(\mathbf{X}^{(i)})}. \quad (2.49)$$

Powyższa aproksymacja opiera się na następującym przekształceniu wyrażenia (2.9):

$$P_f = \int_{R^n} I_{\Omega_f}(\mathbf{x}) \frac{f(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x}) d\mathbf{x}. \quad (2.50)$$

Jeżeli $\text{supp}(q) \supset \text{supp}(f)$, to estymator (2.49) jest zbieżny do (2.9) tak, jak klasyczny estymator Monte Carlo (2.40). Zbieżność estymatora (2.49) jest więc zapewniona dla prawie dowolnego rozkładu prawdopodobieństwa q , jednak tylko niektóre rozkłady pozwalają uzyskać oszacowanie efektywniejsze niż w przypadku klasycznej metody Monte Carlo. Wariancję prawdopodobieństwa (2.50) określa wzór

$$\text{Var} \left(\frac{I_{\Omega_f}(\mathbf{X})f(\mathbf{X})}{q(\mathbf{X})} \right) = \mathbb{E}_q \left(\frac{I_{\Omega_f}(\mathbf{X})f^2(\mathbf{X})}{q^2(\mathbf{X})} \right) - \left(\mathbb{E}_q \left(\frac{I_{\Omega_f}(\mathbf{X})f(\mathbf{X})}{q(\mathbf{X})} \right) \right)^2, \quad (2.51)$$

wyrażenie to przyjmuje skończoną wartość tylko wtedy, gdy

$$\mathbb{E}_q \left(\frac{I_{\Omega_f}(\mathbf{X})f^2(\mathbf{X})}{q^2(\mathbf{X})} \right) = \int_{R^n} \frac{I_{\Omega_f}(\mathbf{x})f^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} < \infty. \quad (2.52)$$

Z powyższego wzoru wynika, że rozkłady q , które w obszarze awarii mają "lżejsze ogony" niż f , a więc f/q jest nieograniczone, nie mogą być stosowane w importance sampling, ponieważ wtedy wariancja (2.51) jest nieograniczona. Jeżeli symulacja została przeprowadzona z wykorzystaniem q , w przypadku którego f/q jest nieograniczone, to wagi $f(\mathbf{x}_i)/q(\mathbf{x}_i)$ wykazują dużą zmienność i nieproporcjonalnie wyróżniają małą grupę elementów próby losowej. W szczególności, wariancja estymatora (2.51) jest ograniczona, jeżeli spełniony jest warunek

$$\frac{f(\mathbf{x})}{q(\mathbf{x})} < M \quad \forall \mathbf{x} \in \Omega_f, \quad (2.53)$$

gdzie M jest liczbą całkowitą. Pośród rozkładów q zapewniających skończoną wariancję estymatora (2.49) można wskazać optymalny rozkład (minimalizujący wariancję) w przypadku danego obszaru awarii i rozkładu prawdopodobieństwa f . Jest on dany wzorem:

$$q^*(\mathbf{x}) = \frac{I_{\Omega_f}(\mathbf{x})f(\mathbf{x})}{\int_{\mathcal{D}} I_{\Omega_f}(\mathbf{x})f(\mathbf{x})d\mathbf{x}}, \quad (2.54)$$

jak widać jest to rozkład, którego gęstość prawdopodobieństwa jest proporcjonalna do rozkładu f , obciętego do obszaru awarii. Wyznaczenie stałej normalizującej rozkład q^* (mianownik wzoru (2.54)) wymaga jednak znajomości rozwiązania, którego się poszukuje. Wzór (2.54) nie ma więc praktycznego znaczenia jako definicja gęstości rozkładu prawdopodobieństwa, z którego generuje się niezależną próbę losową. Postać optymalnego rozkładu prawdopodobieństwa odzwierciedla fakt, że redukcja wariancji w metodzie importance sampling wymaga uwzględnienia informacji o rozwiązaniu

problemu. W praktyce dąży się do znalezienia q mającego podobne własności jak q^* , dotyczy to zwłaszcza położenia maksimumów gęstości prawdopodobieństwa.

Jednym z najbardziej efektywnych sposobów generowania prób losowych w importance sampling jest metoda *exponential tilting*, nazywana również *exponential biasing*. Wykorzystuje ona rozkład prawdopodobieństwa postaci

$$q(\mathbf{x}) = c \exp(\tau S(\mathbf{x})) f(\mathbf{x}), \quad (2.55)$$

gdzie $S(\mathbf{x})$ jest kombinacją liniową rzeczywistych funkcji składowych realizacji wektora losowego \mathbf{X} :

$$S(\mathbf{x}) = \sum_{j=1}^d s_j(x_j). \quad (2.56)$$

Współczynnik τ określa intensywność "nachylenia" rozkładu, z którego generowana jest próbka losowa, w stosunku do rozkładu oryginalnego, a c jest stałą normalizującą rozkład. Proponowane podejście jest efektywne, jeżeli monotoniczna funkcja S jest dobrym przybliżeniem funkcji, której wartość średnia podlega oszacowaniu. W przypadku analizy niezawodności należy wykorzystać przybliżenie funkcji charakterystycznej obszaru awarii (2.8). Rozważmy zadanie analizy niezawodności transformowane do gaussowskiej przestrzeni standardowej. Funkcję charakterystyczną obszaru awarii, w gaussowskiej przestrzeni standardowej (2.21), można aproksymować za pomocą wyrażenia

$$I_{\Delta_f}(\mathbf{u}) \approx I_{\bar{G}(\mathbf{u}) \leq 0}(\mathbf{u}), \quad (2.57)$$

gdzie \bar{G} jest przybliżeniem metody FORM (2.30). Korzystając z funkcji \bar{G} można sformułować

$$S(\mathbf{u}) = \sum_{i=1}^m \alpha_i u_i, \quad (2.58)$$

gdzie α są dane jak w (2.30). Pozostałe parametry (2.55) niech będą postaci: $\tau = \beta$, $c = \exp(-\beta^2/2)$, gdzie β jest również określone jak w (2.30). Funkcja g , zdefiniowana w taki sposób, jest gęstością prawdopodobieństwa rozkładu normalnego o jednostkowej macierzy korelacji i wartości oczekiwanej w punkcie projektowym (2.27) $\mathbf{u}^* = \beta \alpha_i$

$$q(\mathbf{u}) = \exp\left(-\frac{\beta^2}{2}\right) \exp\left(\beta \sum_{i=1}^m \alpha_i u_i\right) \varphi_{\mathbf{u}}(\mathbf{u}) = \prod_{i=1}^m \varphi(u_i, \beta \alpha_i, 1). \quad (2.59)$$

Przesunięcie do punktu projektowego wartości średniej rozkładu prawdopodobieństwa, z którego generuje się próbę losową, jest przypuszczalnie najczęściej stosowaną w analizie niezawodności strategią importance sampling.

Jeżeli symulacje prowadzone są w przestrzeni oryginalnej, podejście to realizuje się przesuując wartość średnią do punktu w obszarze awarii, o największej gęstości prawdopodobieństwa. Jak widać, zastosowanie omówionej metody wymaga wcześniejszego znalezienia punktu projektowego. Można w tym celu wykorzystać algorytmy optymalizacyjne tak, jak w przypadku metod FORM i SORM.

2.3.8. Adaptacyjne metody Monte Carlo

Jak wspomniano, zmniejszenie wariancji w metodzie importance sampling można osiągnąć, definiując gęstość prawdopodobieństwa q na podstawie informacji o rozkładzie prawdopodobieństwa parametrów konstrukcyjnych i własnościach obszaru awarii. Sformułowanie przedstawione w podrozdziale 2.3.7 wykorzystuje w tym celu położenie punktu projektowego. Jeżeli postać funkcji granicznej uniemożliwia znalezienie punktu projektowego z pomocą optymalizacji, to można zastosować *adaptacyjne metody Monte Carlo* [73] (ang. *adaptive sampling, adaptive monte carlo*). Metody te opierają się na pomysłe systematycznego uaktualniania gęstości prawdopodobieństwa q , na podstawie symulacji wykonanych w trakcie działania algorytmu.

Przybliżeniem optymalnego rozkładu prawdopodobieństwa, z którego generowane są symulacje w metodzie importance sampling q^* (2.54) może być rozkład q (np. normalny) mający wartość średnią i macierz kowariancji taką, jak rozkład optymalny. Przybliżenie takie może nie być idealne, ale w celu uproszczenia implementacji, rozkład ten powinien zapewniać łatwe generowanie próby losowej. Wariancję i wartość średnią q^* można oszacować stosując następującą procedurę iteracyjną [17]. Dla danych w i -tym kroku iteracji algorytmu wektora wartości średnich $\boldsymbol{\mu}^{(i)}$ oraz macierzy kowariancji $\mathbf{C}^{(i)}$ wykonuje się n symulacji. Następnie, na podstawie otrzymanej próby losowej, znajduje się kolejne oszacowania wartości oczekiwanej i składowych macierzy kowariancji:

$$\mu_k^{(i+1)} = \frac{1}{n} \sum_{j=1}^n X_k^{(j)} I_{\Omega_f}(\mathbf{X}^{(j)}) \frac{f(\mathbf{X}^{(j)})}{q(\mathbf{X}^{(j)}, \boldsymbol{\mu}^{(i)}, \mathbf{C}^{(i)})}, \quad (2.60)$$

$$C_{kl}^{(i+1)} = \frac{1}{n} \sum_{j=1}^n (X_k^{(j)} - \mu_k^{(i+1)}) (X_l^{(j)} - \mu_l^{(i+1)}) I_{\Omega_f}(\mathbf{X}^{(j)}) \frac{f(\mathbf{X}^{(j)})}{q(\mathbf{X}^{(j)}, \boldsymbol{\mu}^{(i)}, \mathbf{C}^{(i)})}. \quad (2.61)$$

Możemy przypuszczać, że otrzymany w ten sposób ciąg parametrów $\boldsymbol{\mu}^{(i)}$ oraz $\mathbf{C}^{(i)}$ będzie zbieżny do właściwych wartości, zapewniających dobrą aproksymację rozkładu

q^* . W literaturze nie można jednak znaleźć dowodu zbieżności tak sformułowanego algorytmu.

Optymalizacja niezawodnościowa

3.1. Podstawowe sformułowanie optymalizacji niezawodnościowej

Dzięki dostępności komputerów o dostatecznie dużej mocy obliczeniowej oraz wprowadzeniu efektywnych algorytmów do pakietów wspomagających projektowanie, optymalizacja konstrukcji stała się elementem rutynowej praktyki inżynierskiej. Celem optymalizacji jest uzyskanie projektu konstrukcji o minimalnym koszcie, spełniającej jednak wymagania użytkowe, a także zapewniającej wymagany poziom bezpieczeństwa. Klasyczne sformułowanie optymalizacji opiera się na modelu deterministycznym i wykorzystuje tzw. częściowe współczynniki bezpieczeństwa. Praktyka pokazuje jednak, że konstrukcje projektowane w ten sposób są podatne na występowanie awarii. Dotyczy to zwłaszcza niestandardowych rozwiązań, które nie są brane pod uwagę przy formułowaniu zaleceń normowych. Konstrukcje optymalne są, bowiem szczególnie wrażliwe na zmiany parametrów projektowych, które trudno w pełni uwzględnić za pomocą współczynników bezpieczeństwa.

Jak pokazano w rozdziale 2, wpływ zmienności parametrów na zachowanie konstrukcji może być efektywnie uwzględniony za pomocą analizy niezawodności. W tak zwanej *optymalizacji niezawodnościowej*, do zadania optymalizacji, w sposób jawny, wprowadza się ograniczenia nałożone na wartości wskaźników niezawodności. Uważa się, że takie podejście pozwala lepiej zapewnić bezpieczeństwo konstrukcji, zwłaszcza w przypadku projektów niestandardowych.

Zagadnienie optymalizacji niezawodnościowej, sformułowane jako minimalizacja kosztu początkowego konstrukcji, przy ograniczeniach nałożonych na poziom nieza-

wodności, można sformułować w następujący sposób:

$$\text{minimalizuj } C_I(\mathbf{p}), \quad (3.1)$$

przy ograniczeniach :

$$\beta_i(\mathbf{p}, \mathbf{X}) \geq \beta_i^{\min} \quad i = 1, \dots, n_r, \quad (3.2)$$

$$c_j(\mathbf{p}) \geq 0 \quad j = 1, \dots, n_d, \quad (3.3)$$

$$p_k^{\min} \leq p_k \leq p_k^{\max} \quad k = 1, \dots, n_k, \quad (3.4)$$

gdzie \mathbf{p} jest wektorem parametrów projektowych, a \mathbf{X} wielowymiarową zmienną losową, reprezentującą niepewne parametry konstrukcyjne. Przyjmuje się, że zmienne projektowe określają wielkości uznane za deterministyczne, bądź też parametry rozkładu prawdopodobieństwa \mathbf{X} . Za pomocą β_i , $i = 1, \dots, n_r$, oznaczono wskaźniki niezawodności stanów granicznych określonych w przypadku analizowanej konstrukcji, których β_i^{\min} są minimalnymi wartościami zapewniającymi żądany poziom bezpieczeństwa. Funkcje c_j , $j = 1, \dots, n_d$, definiują ograniczenia deterministyczne, pozwalające uwzględnić wymagania projektowe. Tak zwane ograniczenia bezpośrednie (3.4) definiują zakres zmienności parametrów projektowych. Funkcja celu C_I jest początkowym kosztem konstrukcji. W najprostszych sformułowaniach kosztem jest objętość konstrukcji, jednak sformułowania uwzględniające np. podniesienie standardu wykonania [30], bądź koszty utrzymania i następstw możliwej awarii także są rozważane [104].

Można również spotkać alternatywne postaci sformułowania zagadnienia optymalizacji niezawodnościowej. Na przykład, istotna z praktycznego punktu widzenia jest minimalizacja prawdopodobieństwa awarii przy zadanym koszcie konstrukcji.

Przykłady optymalizacji niezawodnościowej wymiarów oraz kształtu konstrukcji prętowych, żelbetowych, stalowych itp. znajdziemy w literaturze [104, 115]. W większości przypadków wykorzystywany jest tzw. algorytm optymalizacji zagnieżdżonej [114]. Zewnętrzny poziom optymalizacji odpowiada za minimalizację kosztu, natomiast poziom wewnętrzny służy do wyznaczenia ograniczeń niezawodnościowych metodami FORM/SORM (por. 2.3.3, 2.3.4). Ze względu na złożoność obliczeniową algorytmu optymalizacji zagnieżdżonej do jego rozwiązania wykorzystuje się metody gradientowe. Oczywiście zakłada się wtedy różniczkowalność funkcji celu i ograniczeń. Próby zastosowania algorytmów bezgradientowych były również podejmowane [122]. Metody te mogą być stosowane w przypadku funkcji nieróżniczkowalnych, jednak są bardzo wolno zbieżne.

Drugi ze sposobów rozwiązania zadania optymalizacji niezawodnościowej został zaproponowany w pracy [71], a następnie rozszerzony w [62]. W tej metodzie wewnętrzny poziom optymalizacji jest wyeliminowany, poprzez zastąpienie ograniczeń niezawodno-

ściowych warunkami koniecznymi optymalności pierwszego rzędu dla poszczególnych punktów projektowych. Sformułowanie to umożliwi rozwiązanie zadania w rozszerzonej przestrzeni parametrów, przez algorytmy programowania kwadratowego, w ten sposób można uzyskać duże oszczędności nakładów obliczeniowych.

3.2. Optymalizacja niezawodnościowa mieszana i dyskretna

W niektórych zagadnieniach optymalizacji konstrukcji konieczne, bądź pożądane, jest przyjęcie założenia, że część parametrów projektowych przyjmuje wartości ze zbiorów dyskretnych. Przykładem może być tutaj projektowanie prętowych konstrukcji stalowych, których elementy mogą być dobierane praktycznie tylko ze standardowego katalogu gotowych wyrobów hutniczych. Okazuje się, że wprowadzenie ograniczeń na wartości zmiennych projektowych powoduje dodatkowe problemy numeryczne. Dyskretyzacja wymusza, bowiem zastosowanie do rozwiązania zadania mało efektywnej metody przeglądu, bądź metody transformacji, która pozwala wykorzystać algorytmy gradientowe, ale w znacznie poszerzonej przestrzeni parametrów.

Wprowadzając podział na parametry projektowe ciągłe $\mathbf{p}^{(c)}$ i dyskretnie $\mathbf{p}^{(d)}$, na podstawie zadania (3.1) - (3.4), można sformułować zagadnienie *mieszanej optymalizacji niezawodnościowej* (określone dla mieszanych parametrów projektowych: ciągłych i dyskretnych):

$$\text{minimalizuj } C_I(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}), \quad (3.5)$$

przy ograniczeniach :

$$\beta_i(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}, \mathbf{X}) \geq \beta_i^{\min} \quad i = 1, \dots, n_r, \quad (3.6)$$

$$c_j(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}) \geq 0 \quad j = 1, \dots, n_d, \quad (3.7)$$

$$p_k^{(c)\min} \leq p_k^{(c)} \leq p_k^{(c)\max} \quad k = 1, \dots, n_k, \quad (3.8)$$

$$p_l^{(d)} \in \mathcal{P}_l^{(d)} = \{p_{l,1}^{(d)}, p_{l,2}^{(d)}, \dots, p_{l,m_l}^{(d)}\} \quad l = 1, \dots, n_l. \quad (3.9)$$

Ograniczenia (3.9), podobnie jak (3.8) w przypadku parametrów ciągłych, określają dopuszczalne wartości parametrów dyskretnych w zbiorach $\mathcal{P}_l^{(d)} = \{p_{l,1}^{(d)}, p_{l,2}^{(d)}, \dots, p_{l,m_l}^{(d)}\}$, $l = 1, \dots, n_l$. Jeżeli w zadaniu (3.5)-(3.9) nie występują ciągle zmienne projektowe, to mamy do czynienia z "czystym" zadaniem *optymalizacji dyskretniej*.

3.2.1. Transformacja do przestrzeni parametrów ciągłych

Jedną z metod rozwiązania zadania optymalizacji mieszanej (3.5)-(3.9) polega na transformacji do rozszerzonej przestrzeni ciągłych parametrów projektowych [127]. Omówienie tego sformułowania rozpoczniemy od zdefiniowania wektorów zmiennych dyskretnych: $\bar{\mathbf{v}}_l = \{\bar{v}_{l,1}, \bar{v}_{l,2}, \dots, \bar{v}_{l,m_l}\}$, $l = 1, \dots, n_l$, gdzie

$$\bar{v}_{l,j} \in \{0, 1\}, \quad j = 1, \dots, m_l, \quad l = 1, \dots, n_l. \quad (3.10)$$

Ponadto przyjmijmy założenie, że spełnione jest ograniczenie:

$$\sum_{j=1}^{m_l} \bar{v}_{l,j} = 1, \quad l = 1, \dots, n_l. \quad (3.11)$$

Teraz dyskretne parametry projektowe $p_l^{(d)}$, $l = 1, \dots, n_l$ można przedstawić za pomocą \bar{v}_l na podstawie kombinacji liniowej:

$$p_l^{(d)} = \sum_{j=1}^{m_l} p_{l,j}^{(d)} \bar{v}_{l,j}, \quad l = 1, \dots, n_l. \quad (3.12)$$

Tak więc, za pomocą transformacji określonej przez (3.11) i (3.12), parametry $\mathbf{p}^{(d)}$ w zadaniu (3.5)-(3.9) zastąpiono parametrami $\bar{\mathbf{v}}_l$, $l = 1, \dots, n_l$.

W pracy [127] pokazano, że dyskretne parametry $\bar{v}_{l,j}$, można zastąpić parametrami ciągłymi $v_{l,j} \in [0, 1]$, $j = 1, \dots, m_l$, $l = 1, \dots, n_l$, poprzez wprowadzenie ograniczenia kwadratowego

$$Q(\mathbf{v}_l) = \sum_{j=1}^{m_l} j^2 v_{l,j} - \left(\sum_{j=1}^{m_l} j v_{l,j} \right)^2 = 0; \quad (3.13)$$

jeżeli ponadto prawdziwa jest równość

$$\sum_{j=1}^{m_l} v_{l,j} = 1, \quad l = 1, \dots, n_l. \quad (3.14)$$

Zastąpienie zmiennych dyskretnych zmiennymi ciągłymi, jest również możliwe na podstawie innych sformułowań ograniczenia (3.13). Jednak zaletą tej konkretnej postaci (3.13), zaproponowanej w [127], jest to, że nie generuje ona dodatkowych minimów lokalnych.

Teraz, korzystając z transformacji (3.13)-(3.14), można określić zadanie optymalizacji w ciągłej przestrzeni, zawierającej parametry $\mathbf{v} = \left\{ \{v_{1,1}, \dots, v_{1,m_1}\}, \dots, \{v_{n_l,1}, \dots, v_{n_l,m_{n_l}}\} \right\}$, które jest równoważne zadaniu (3.5)-(3.9):

$$\text{minimalizuj } C_I(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}(\mathbf{v})) = \tilde{C}_I(\mathbf{p}^{(c)}, \mathbf{v}), \quad (3.15)$$

przy ograniczeniach :

$$\beta_i(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}(\mathbf{v}), \mathbf{X}) = \tilde{\beta}_i(\mathbf{p}^{(c)}, \mathbf{v}, \mathbf{X}) \geq \beta_i^{\min} \quad i = 1, \dots, n_r, \quad (3.16)$$

$$c_j(\mathbf{p}^{(c)}, \mathbf{p}^{(d)}(\mathbf{v})) = \tilde{c}_j(\mathbf{p}^{(c)}, \mathbf{v}) \geq 0 \quad j = 1, \dots, n_d, \quad (3.17)$$

$$p_k^{(c)\min} \leq p_k^{(c)} \leq p_k^{(c)\max} \quad k = 1, \dots, n_k, \quad (3.18)$$

$$\sum_{j=1}^{m_l} v_{l,j} - 1 = 0 \quad l = 1, \dots, n_l, \quad (3.19)$$

$$Q(\mathbf{v}_l) = 0 \quad l = 1, \dots, n_l, \quad (3.20)$$

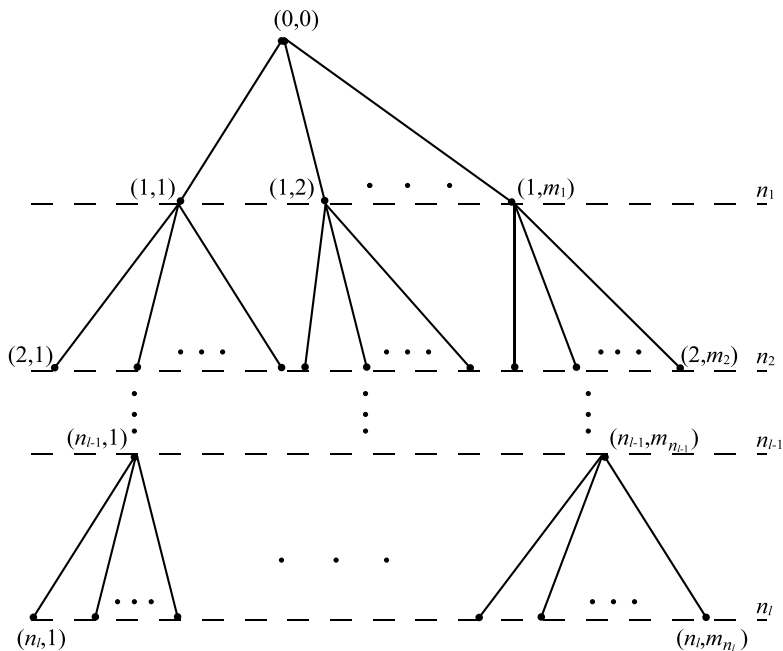
$$0 \leq v_{l,j} \leq 1 \quad j = 1, \dots, m_l, \quad l = 1, \dots, n_l. \quad (3.21)$$

Powyższe zadanie jest równoważne (3.5)-(3.9), jeżeli dla optymalnych wartości \mathbf{v}^* parametrów \mathbf{v} zachodzi:

$$p_l^{(d)*} = \sum_{j=1}^{m_l} p_{l,j} v_{l,j}^* \in \mathcal{P}_l^{(d)} = \left\{ p_{l,1}^{(d)}, p_{l,2}^{(d)}, \dots, p_{l,m_l}^{(d)} \right\} \quad l = 1, \dots, n_l. \quad (3.22)$$

Jeżeli spełnione są warunki (3.19)-(3.21), to dla każdego l tylko jedno $v_{l,j}$ jest równe 1 a pozostałe są równe 0; tak więc wyrażenie (3.22) jest prawdziwe [127].

Zadanie optymalizacji (3.15)-(3.21) można rozwiązać za pomocą dowolnego algorytmu programowania nieliniowego. Wadą tego sformułowania jest wzrost liczby parametrów projektowych, wraz ze wzrostem liczby elementów w zbiorach $\mathcal{P}_l^{(d)}$. Na przykład, mały problem optymalizacji z pięcioma dyskretnymi zmiennymi projektowymi, które przyjmują wartości w zbiorach dwudziestoelementowych (a więc niedużych, mając na



RYSUNEK 3.1. Drzewo metody kontrolowanego przeglądu.

uwadze różnorodność typowych elementów konstrukcyjnych), po transformacji generuje zadanie ze stoma parametrami projektowymi. Jednym ze sposobów zmniejszenia liczby zmiennych typu v jest określenie zbiorów \mathcal{P}_l^d , za pomocą niewielkiej liczby elementów, które są bliskie rozwiązaniu ciągłego zadania optymalizacji (3.1)-(3.4) [115]. Takie podejście zakłada, że dyskretne rozwiązanie optymalne jest bliskie rozwiązaniu w ciągłej przestrzeni parametrów, co w przypadku konstrukcji inżynierskich zazwyczaj jest prawdą. Omówiona metoda została wykorzystana w prezentowanych dalej przykładach.

3.2.2. Metoda kontrolowanego przeglądu

Jeżeli w zadaniu optymalizacji występują jedynie dyskretne parametry projektowe, których zbiory wartości mają ograniczoną liczbę elementów, to liczba kombinacji tych elementów jest oczywiście skończona. Teoretycznie istnieje więc możliwość rozwiąza-

nia zadania, poprzez sprawdzenie ograniczeń optymalizacji dla wszystkich kombinacji wartości parametrów projektowych, a następnie znalezienie kombinacji, która je spełnia i ma najmniejszą wartość funkcji celu. Ponieważ jednak liczba kombinacji jest równa iloczynowi liczby elementów zbiorów wartości, to w przypadku rzeczywistych zagadnień jest ona tak duża, że koszty obliczeniowe uzyskania rozwiązania tą metodą nie mogą być zaakceptowane. Proces sprawdzania własności kombinacji projektowych można jednak usprawnić. Tak zwane *metody kontrolowanego przeglądu* [48] pozwalają uzyskać rozwiązanie, bądź przybliżenie rozwiązania, zadania optymalizacji dyskretnej przy wykorzystaniu rozsądnej mocy obliczeniowej. Poniżej przedstawiony zostanie algorytm optymalizacji dyskretnej konstrukcji kratownicowych, który wykorzystuje to podejście.

Rozpatrzmy zadanie minimalizacji objętości kratownic, w którym dyskretnymi parametrami projektowymi są pola przekrojów prętów $\mathbf{A} = \{A_1, A_2, \dots, A_{n_l}\}$. Funkcja kosztu początkowego jest więc postaci:

$$C_I(\mathbf{A}) = \sum_{l=1}^{n_l} L_l A_l, \quad (3.23)$$

gdzie L_l $l = 1, \dots, n_l$ są całkowitymi długościami elementów, których pola przekrojów są określone przez A_j .

Metodę kontrolowanego przeglądu omówimy posługując się uporządkowanym drzewem z rysunku 3.1, którego węzły zewnętrzne reprezentują wszystkie wartości funkcji celu (3.23). Budowę drzewa rozpoczyna się od korzenia, który jest oznaczony przez $(0, 0)$. Kolejne węzły drzewa generuje się w następujący sposób: węzeł ze wskaźnikami (l, j) jest poprzednikiem węzłów oznaczonych przez $(l + 1, j)$, gdzie $j = 1, \dots, m_{l+1}$. Węzeł, którego pierwszy wskaźnik ma wartość n_l , nie ma następników. Jeżeli wskaźnikiem następnika jest (l, j) , to objętość grupy elementów reprezentowana przez krawędź drzewa, na której końcu się znajduje, wynosi $L_l A_{l,j}$.

Zwróćmy uwagę, że węzły na poziomie 1 odpowiadają możliwym objętościom pierwszej grupy elementów konstrukcyjnych. Węzły na poziomie 2 reprezentują kombinacje objętości elementów grup 1 i 2. Ostatecznie, węzły na poziomie n_l reprezentują możliwe objętości całej konstrukcji.

Przyjmijmy, że elementy zbiorów wartości parametrów projektowych $\mathcal{P}_l^A = \{A_{l,1}, A_{l,2}, \dots, A_{l,m_l}\}$, $l = 1, \dots, n_l$, są uporządkowane rosnąco. Konsekwencją tego założenia jest to, że ścieżka w drzewie o korzeniu $(0, 0)$ i ścieżki w poddrzewach o korzeniach (l, j) przechodzące przez węzły (l, m_l) określają węzeł zewnętrzny, który reprezentuje maksymalną objętość dla ścieżek w tych poddrzewach. Własność ta zostanie wykorzystana w omówionym poniżej algorytmie.

Prezentowany algorytm kontrolowanego przeglądu opiera się na założeniu, że

rozważane zadanie optymalizacji może być rozwiązane w przestrzeni, w której \mathbf{A} są parametrami ciągłymi. Objętość maksymalna w rozwiązaniu ciągłym zadania $C_I(\mathbf{A}^{(c)*})$ jest dolnym ograniczeniem wartości funkcji celu kombinacji zmiennych dyskretnych, spełniających ograniczenia optymalizacji. Tak więc, celem poniższego algorytmu jest znalezienie kombinacji wartości parametrów projektowych $\mathbf{A}^{(d)*} = \{A_{1,j^*}, A_{2,j^*}, \dots, A_{n_l, j^*}\}$, która określa konstrukcję o objętości najbliższej $C_I(\mathbf{A}^{(c)*})$ i spełnia ograniczenia zadania optymalizacji. Wykonanie algorytmu obejmuje następujące kroki:

1. Wyznaczyć $\mathbf{A}^{(c)*}$, które jest rozwiązaniem zadania optymalizacji, przy założeniu ciągłości zmiennych projektowych. Wyznaczyć zbiory wartości dyskretnych parametrów projektowych \mathcal{P}_l^A , $l = 1, \dots, n_l$, wybierając z katalogu produkcyjnego elementy, których parametry są bliskie odpowiednim składowym $\mathbf{A}^{(c)*}$. Wielkość tych zbiorów powinna być uzależniona od dostępnej mocy obliczeniowej [115].
2. Na drzewie z rysunku 3.1 wygenerować ścieżkę, reprezentującą rozwiązanie dyskretne, zgodnie z następującą procedurą:
 - (a) Rozważmy poddrzewa, których korzeniami są węzły oznaczone przez wskaźniki $(1, j)$, gdzie $j = 1, \dots, m_1$, a więc położone na poziomie 1. Maksymalne wartości funkcji celu, reprezentowane przez węzły skrajne (położone na poziomie n_l) tych poddrzew, są określone wyrażeniem:

$$C_{1,j}^{\max} = L_1 A_{1,j} + \sum_{l=2}^{n_l} L_l A_{l,m_l}, \quad j = 1, \dots, m_1. \quad (3.24)$$

Ze zbioru wartości pól przekroju pierwszego elementu wybrać tę, która spełnia warunek:

$$A_{(1,j^\#)} = \min \left\{ A_{1,j} | C_I(\mathbf{A}^{(c)*}) \leq C_{1,j}^{\max} \right\}, \quad j = 1, \dots, m_1. \quad (3.25)$$

Jeżeli zbiór $\{A_{1,j} | C_I(\mathbf{A}^{(c)*}) \leq C_{1,j}^{\max}\}$ jest pusty, to konieczne jest ponowne zdefiniowanie zbioru wartości parametru \mathcal{P}_1^A i powtórne uruchomienie algorytmu.

- (b) Rozważmy poddrzewa o korzeniach na poziomie 2, oznaczonych wskaźnikami $(2, j)$, $j = 1, \dots, m_2$, które należą do poddrzewa o korzeniu $(1, j^\#)$. Maksymalne objętości, reprezentowane przez ścieżki w takich poddrzewach, są określone wzorem

$$C_{2,j}^{\max} = L_1 A_{(1,j^\#)} + L_2 A_{(2,j)} + \sum_{l=3}^{n_l} L_l A_{l,m_l}, \quad j = 1, \dots, m_2. \quad (3.26)$$

Ze zbioru pól przekroju drugiego elementu wybrać wartość, która spełnia warunek:

$$A_{(2,j^\#)} = \min \left\{ A_{2,j} | C_I(\mathbf{A}^{(c)*}) \leq C_{2,j}^{\max} \right\}, \quad j = 1, \dots, m_2, \quad (3.27)$$

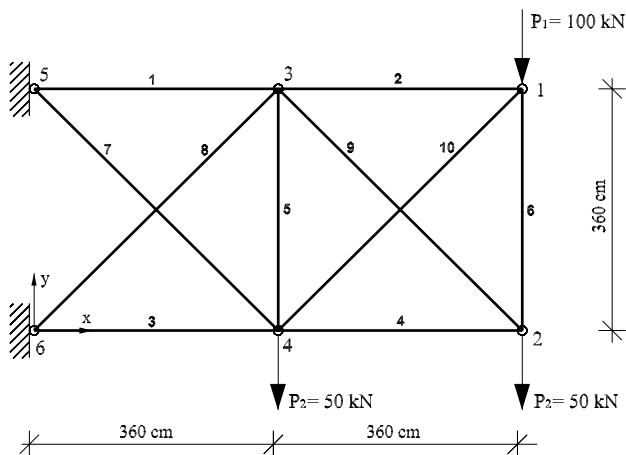
- (c) Kontynuując powyższą procedurę, wybrać kolejne elementy ze zbiorów wartości parametrów $A_{(l,j^\#)}$, $l = 3, \dots, n_l$.

Otrzymana, w powyższych krokach algorytmu, kombinacja parametrów $A_{(l,j^\#)}$, $l = 1, \dots, n_l$ określa ścieżkę $S^\#$, która przechodzi przez węzły $(0,0), (1,j^\#), \dots, (n_l,j^\#)$. Wszystkie ścieżki położone, w drzewie na rysunku 3.1, na lewo od $S^\#$ określają konstrukcje o mniejszej objętości niż $C_I(\mathbf{A}^{(c)*})$. Wśród tych kombinacji nie może być rozwiązania $\mathbf{A}^{(d)*}$, gdyż $C_I(\mathbf{A}^{(c)*})$ jest ograniczeniem dolnym objętości kombinacji projektowych, spełniających ograniczenia optymalizacji. Tak więc kombinacje parametrów projektowych, określone przez ścieżki na lewo od $S^\#$, nie wymagają sprawdzania ograniczeń niezawodnościowych.

3. Obliczyć objętości konstrukcji, reprezentowanych przez ścieżki na prawo od $S^\#$. Te kombinacje, których objętość jest większa od $C_I(\mathbf{A}^{(c)*})$ należy uporządkować rosnąco. Sprawdzać kolejno ograniczenia zadania optymalizacji. Pierwsza kombinacja parametrów projektowych, spełniająca ograniczenia, jest poszukiwanym rozwiązaniem $\mathbf{A}^{(d)*}$ zadania optymalizacji dyskretniej.

W przypadku optymalizacji niezawodnościowej, największe koszty numeryczne związane są z oszacowaniami prawdopodobieństwa awarii. Efektywność metody przeglądu można więc podnieść, poprzez ograniczenie liczby kombinacji projektowych, w przypadku których wykonuje się analizę niezawodności. W pracy [115] zaproponowano, aby wykorzystać w tym celu aproksymację wskaźników niezawodności. Pomysł ten opiera się na intuicyjnym przekonaniu, że wartości wskaźników niezawodności rozwiązania optymalnego w przestrzeni dyskretniej są bliskie wartościom wskaźników rozwiązania w przestrzeni ciągłej. Optymalna kombinacja dyskretnych parametrów projektowych nie może, bowiem określać konstrukcji niebezpiecznej, ale także nie powinna definiować konstrukcji zbyt bezpiecznej, gdyż wtedy można przypuszczać, że elementy są przewymiarowane. Jeżeli w analizie niezawodności wykorzystuje się metody FORM/SORM (roz. 2.3.4, 2.3.3), to wrażliwości wskaźników niezawodności $\partial\beta_i(\mathbf{A})/\partial A_l$, $i = 1, \dots, n_r$, $l = 1, \dots, n_l$, są znane. Powyższy pomysł można więc zrealizować, rezygnując ze sprawdzania ograniczeń niezawodnościowych w przypadku kombinacji projektowych, które nie spełniają następującego warunku:

$$\left| \frac{\partial\beta_i(\mathbf{A})}{\partial A_l} \right|_{\mathbf{A}=\mathbf{A}^{c*}} \Delta A_l \leq \varepsilon, \quad i = 1, \dots, n_r, \quad l = 1, \dots, n_l, \quad (3.28)$$



RYСУNEK 3.2. Kratownica dziesięcioelementowa: początkowa geometria i obciążenie.

gdzie $\Delta A_l = A_l - A_l^{c*}$. Metoda ta, pozwala na ograniczenie nakładów obliczeniowych, ale stwarza możliwość wykluczenia z procesu dokładnego sprawdzania ograniczeń "prawdziwego" rozwiązania zadania optymalizacji dyskretniej. Rozwiązanie, otrzymane z wykorzystaniem takiej modyfikacji algorytmu, należy więc traktować jako przybliżone.

Przykład 3.1: W tym przykładzie zostanie przedstawiona optymalizacja dziesięcioelementowej kratownicy (rys. 3.2) za pomocą metody transformacji do przestrzeni parametrów ciągłych (roz. 3.2.1).

Obciążenia zewnętrzne konstrukcji stanowią: siła $P_1 = 100\text{kN}$ działająca w węźle 1 oraz siły $P_2 = 50\text{kN}$ przyłożone w dolnych węzłach 2 i 4. Model stochastyczny składa się z 21 zmiennych losowych przedstawionych w tabelicy 3.1.

Celem zadania optymalizacji jest wyznaczenie położenia węzłów nieutwierdzonych oraz dobranie przekrojów prętów z katalogu gotowych kątowników równoramiennych, które minimalizują objętość konstrukcji, zapewniając spełnienie przedstawionych poniżej ograniczeń. Tak więc, zadanie zostało określone z wykorzystaniem 14 parametrów projektowych. Pierwszych 10 to wartości średnie zmiennych losowych, reprezentujących pola przekrojów elementów. Założono, że przyjmują one wartości zgodnie z katalogiem kątowników równoramiennych, składającym się z 37 elementów. Pozostałe cztery parametry projektowe to pionowe współrzędne węzłów 1-4.

Par los.	Opis	P	Mom. rozkl.	Wart. pocz.	Par pr.	Opt. ciągła	Opt. dyskr.
1	2	3	4	5	6	7	8
X_1	Pole przekroju elem. Nr. 1 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_1)$ $\sigma(X_1)$	20.0 cm ² 1.0 cm ²	p_1	28.85 cm ²	29.90 cm ² L130 × 130 × 12
X_2	Pole przekroju elem. Nr. 2 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_2)$ $\sigma(X_2)$	20.0 cm ² 1.0 cm ²	p_2	18.57 cm ²	18.70 cm ² L90 × 90 × 11
X_3	Pole przekroju elem. Nr. 3 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_3)$ $\sigma(X_3)$	20.0 cm ² 1.0 cm ²	p_3	40.21 cm ²	43.00 cm ² L150 × 150 × 15
X_4	Pole przekroju elem. Nr. 4 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_4)$ $\sigma(X_4)$	20.0 cm ² 1.0 cm ²	p_4	16.40 cm ²	18.70 cm ² L90 × 90 × 11
X_5	Pole przekroju elem. Nr. 5 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_5)$ $\sigma(X_5)$	20.0 cm ² 1.0 cm ²	p_5	5.09 cm ²	5.69 cm ² L50 × 50 × 6
X_6	Pole przekroju elem. Nr. 6 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_6)$ $\sigma(X_6)$	20.0 cm ² 1.0 cm ²	p_6	5.09 cm ²	5.69 cm ² L50 × 50 × 6
X_7	Pole przekroju elem. Nr. 7 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_7)$ $\sigma(X_7)$	20.0 cm ² 1.0 cm ²	p_7	6.03 cm ²	6.91 cm ² L60 × 60 × 6
X_8	Pole przekroju elem. Nr. 8 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_8)$ $\sigma(X_8)$	20.0 cm ² 1.0 cm ²	p_8	24.60 cm ²	27.50 cm ² L120 × 120 × 12
X_9	Pole przekroju elem. Nr. 9 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_9)$ $\sigma(X_9)$	20.0 cm ² 1.0 cm ²	p_9	5.75 cm ²	5.82 cm ² L60 × 60 × 5
X_{10}	Pole przekroju elem. Nr. 10 <i>staty wsp. zmienności: 5%</i>	LN	$E(X_{10})$ $\sigma(X_{10})$	20.0 cm ² 1.0 cm ²	p_{10}	35.53 cm ²	37.2 cm ² L160 × 160 × 12
X_{11}	Moduł sprężystości wszystkich elementów	LN	$E(X_{11})$ $\sigma(X_{11})$	21000 $\frac{\text{KN}}{\text{cm}^2}$ 1050 $\frac{\text{KN}}{\text{cm}^2}$			
X_{12}	Granica plastyczności wszystkich elementów	LN	$E(X_{12})$ $\sigma(X_{12})$	21 $\frac{\text{KN}}{\text{cm}^2}$ 1 $\frac{\text{KN}}{\text{cm}^2}$			
X_{13}	Współrzędna x węzła Nr. 1	N	$E(X_{13})$ $\sigma(X_{13})$	720 cm 2 cm			
X_{14}	Współrzędna y węzła Nr. 1	N	$E(X_{14})$ $\sigma(X_{14})$	360 cm 2 cm	p_{11}	150.0 cm	150.0 cm
X_{15}	Współrzędna x węzła Nr. 2	N	$E(X_{15})$ $\sigma(X_{15})$	720 cm 2 cm			
X_{16}	Współrzędna y węzła Nr. 2	N	$E(X_{16})$ $\sigma(X_{16})$	0 cm 2 cm	p_{12}	10.1 cm	100.0 cm
X_{17}	Współrzędna x węzła Nr. 3	N	$E(X_{17})$ $\sigma(X_{17})$	360 cm 2 cm			
X_{18}	Współrzędna y węzła Nr. 3	N	$E(X_{18})$ $\sigma(X_{18})$	360 cm 2 cm	p_{13}	259.9 cm	276.4 cm
X_{19}	Współrzędna x węzła Nr. 4	N	$E(X_{19})$ $\sigma(X_{19})$	360 cm 2 cm			
X_{20}	Współrzędna y węzła Nr. 4	N	$E(X_{20})$ $\sigma(X_{20})$	0 cm 2 cm	p_{14}	35.3 cm	38.46 cm
X_{21}	Mnożnik obciążenia 1	LN	$E(X_{21})$ $\sigma(X_{21})$	1.0 0.05			
X_{22}	Mnożnik obciążenia 2	G	$E(X_{22})$ $\sigma(X_{22})$	1.0 0.2			

TABLICA 3.1. Kratownica dziesięcioelementowa: zmienne losowe i projektowe. Kolumna 3 zawiera informację o typach rozkładów prawdopodobieństwa: LN - log-normalny, N - normalny, G - Gumbela.

Obliczenia statyczne wykonano z wykorzystaniem liniowej teorii sprężystości, a odpowiedź konstrukcji była analizowana na podstawie wybranych przemieszczeń i naprężeń. W przykładzie rozważane są następujące stany graniczne:

- Dopuszczalne przemieszczenie węzła Nr. 2:

$$g_1(\mathbf{q}(\mathbf{X}, \mathbf{p})) = 1 - \frac{|q_2(\mathbf{X}, \mathbf{p})|}{q^a}, \quad (3.29)$$

gdzie \mathbf{X} jest wektorem zmiennych losowych, \mathbf{q} jest wektorem przemieszczeń węzłowych, a q_2 oznacza pionowe przemieszczenie węzła Nr. 2, które powinno być mniejsze od wartości dopuszczalnej $q^a = 3.5$ cm. Niejawna zależność funkcji granicznej od zmiennych losowych i parametrów projektowych została zaznaczona w jej definicji.

- Stan graniczny dopuszczalnej wartości naprężenia, bądź utraty stateczności elementów prętowych:

$$g_i(q(\mathbf{X}, \mathbf{p}), \mathbf{X}, \mathbf{p}) = 1 - \frac{|\sigma_{i-1}(q(\mathbf{X}, \mathbf{p}), \mathbf{X}, \mathbf{p})|}{\sigma_{i-1}^a(\mathbf{X}, \mathbf{p})}, \quad i = 2, \dots, 11, \quad (3.30)$$

gdzie σ_j jest naprężeniem w j -tym elemencie, a σ_j^a jego wartością dopuszczalną. W przypadku prętów rozciąganych, w trakcie poszukiwania punktu projektowego wartość σ_j^a jest stała i równa granicy plastyczności. Natomiast w przypadku prętów ściskanych, ze względu na chęć uwzględnienia utraty stateczności, σ_j^a jest zmienne w zależności od parametrów geometrycznych elementu.

Przyjęto, że w przypadku konstrukcji optymalnej wartości β_i , $i = 1, \dots, n_r$, które odpowiadają powyższym stanom granicznym, nie powinny być mniejsze niż 3.7.

Zastosowanie metody transformacji, przy założeniu, że zbiory wartości parametrów dyskretnych $\mathcal{P}_l^{(d)}$, $l = 1, \dots, n_l$, zawierają wszystkie elementy katalogu, wymagałoby określenia zadania optymalizacji z $10 \times 37 + 4 = 374$ zmiennymi projektowymi. Rozwiązanie, za pomocą metod numerycznych, zadania z tak dużą liczbą parametrów projektowych może okazać się trudne. Dlatego w tym przykładzie liczbę elementów zbiorów $\mathcal{P}_l^{(d)}$ ograniczono do kilku, o wartościach bliskich rozwiązaniu zadania optymalizacji z ciągłymi parametrami \mathbf{A} . Zadanie to jest określone w następujący sposób:

$$\text{minimalizuj } C_I(\mathbf{p}) = \sum_{l=1}^{10} p_l L_l(p_{11}, \dots, p_{14}), \quad (3.31)$$

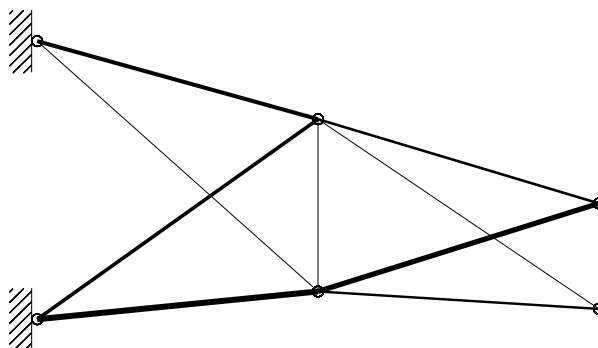
przy ograniczeniach :

$$\beta_i(\mathbf{p}, \mathbf{X}) \geq 3.7 \quad i = 1, \dots, 11, \quad (3.32)$$

$$5.09 \text{ cm}^2 \leq p_l \leq 76.4 \text{ cm}^2 \quad l = 1, \dots, 10, \quad (3.33)$$

$$150 \text{ cm} \leq p_l \leq 450 \text{ cm} \quad l = 11, 13, \quad (3.34)$$

$$-100 \text{ cm} \leq p_l \leq 100 \text{ cm} \quad l = 12, 14, \quad (3.35)$$



RYSUNEK 3.3. Optymalizacja ciągła: projekt optymalny.

gdzie L_l , $l = 1, \dots, 10$, są długościami elementów. Ograniczenia proste (3.33) odpowiadają ekstremom wartości parametrowych elementów katalogu.

Zadanie (3.31)-(3.35) rozwiązano za pomocą programu optymalizacji niezawodnościowej POLSAP-RBO [114]. Początkowa objętość konstrukcji 83929.4 cm^3 została zmniejszona o blisko 17% do 69878.4 cm^3 . Optymalny kształt konstrukcji przedstawia rysunek 3.3. Grubość linii jest proporcjonalna do optymalnych wartości zmiennych projektowych, przedstawionych w siódmej kolumnie tablicy 3.1.

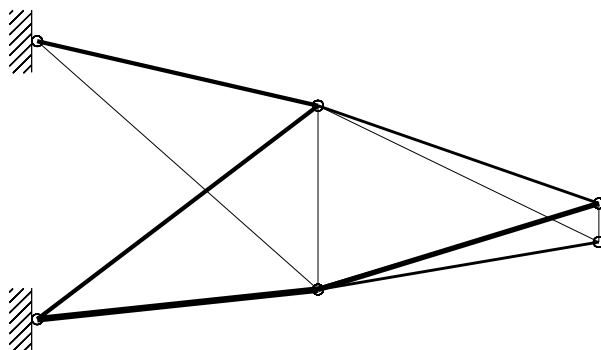
Zbiory wartości parametrów dyskretnych $\mathcal{P}_l^{(d)}$, $l = 1, \dots, 10$, określono wybierając z katalogu kątowników po 3 elementy, o polu przekroju najbliższym rozwiązaniu zadania (3.31)-(3.35) (1 element mniejszy bądź równy, 2 elementy większe). Listy elementów zbiorów $\mathcal{P}_l^{(d)}$ zostały przedstawione w tablicy 3.2.

$\mathcal{P}_{1,j}$		$\mathcal{P}_{2,j}$		$\mathcal{P}_{3,j}$		$\mathcal{P}_{4,j}$	
Symbol	Pole	Symbol	Pole	Symbol	Pole	Symbol	Pole
L120×120×12	27.5	L90×90×9	15.5	L160×160×12	37.2	L90×90×9	15.5
L130×130×12	29.9	L90×90×11	18.7	L150×150×15	43.0	L90×90×11	18.7
L150×150×12	34.8	L100×100×10	19.2	L160×160×15	46.0	L100×100×10	19.2

$\mathcal{P}_{5,j}$		$\mathcal{P}_{6,j}$		$\mathcal{P}_{7,j}$		$\mathcal{P}_{8,j}$	
Symbol	Pole	Symbol	Pole	Symbol	Pole	Symbol	Pole
L45×45×6	5.09	L45×45×6	5.09	L60×60×5	5.82	L120×120×10	23.2
L50×50×6	5.69	L50×50×6	5.69	L60×60×6	6.91	L120×120×12	27.5
L60×60×5	5.82	L60×60×5	5.82	L75×75×5	7.34	L130×130×12	29.9

$\mathcal{P}_{9,j}$		$\mathcal{P}_{10,j}$	
Symbol	Pole	Symbol	Pole
L50×50×6	5.69	L150×150×12	34.8
L60×60×5	5.82	L160×160×12	37.2
L60×60×6	6.91	L150×150×15	43.0

TABLICA 3.2. Katalogi kątowników równoramiennych. Symbole zawierają informacje o wymiarach w milimetrach. Pola przekrojów podano w centymetrach kwadratowych.



RYSUNEK 3.4. Optymalizacja mieszana: projekt optymalny.

Metodę transformacji zastosowano do zredukowanych zbiorów wartości parametrów dyskretnych, w wyniku czego otrzymano następujące zadanie optymalizacji:

$$\text{minimalizuj } \tilde{C}_I(\mathbf{v}, p_{11}, p_{12}, p_{13}, p_{14}) = \sum_{l=1}^{10} \left(\sum_{j=1}^3 p_{l,j} v_{l,j} \right) L_l(p_{11}, \dots, p_{14}), \quad (3.36)$$

przy ograniczeniach :

$$\tilde{\beta}_i(\mathbf{v}, p_{11}, p_{12}, p_{13}, p_{14}, \mathbf{X}) \geq 3.7 \quad i = 1, \dots, 11, \quad (3.37)$$

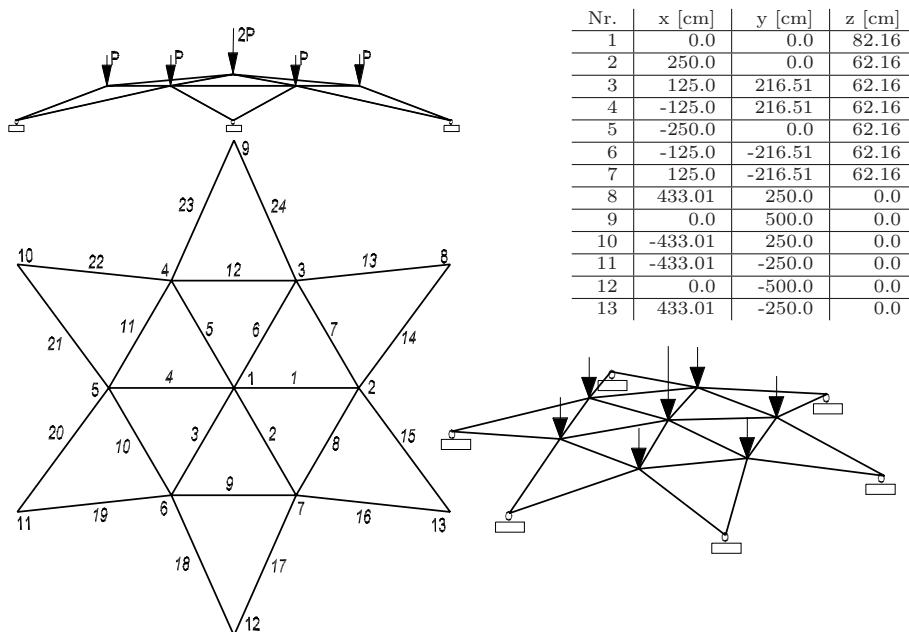
$$\sum_{j=1}^3 v_{l,j} - 1 = 0 \quad l = 1, \dots, 10, \quad (3.38)$$

$$Q(\mathbf{v}_l) = 0 \quad l = 1, \dots, 10, \quad (3.39)$$

$$0 \leq v_{l,j} \leq 1 \quad j = 1, \dots, 3, \quad l = 1, \dots, 10, \quad (3.40)$$

$$150 \text{ cm} \leq p_l \leq 450 \text{ cm} \quad l = 11, 13, \quad (3.41)$$

$$-100 \text{ cm} \leq p_l \leq 100 \text{ cm} \quad l = 12, 14, \quad (3.42)$$

RYSUNEK 3.5. Kopuła kratownicowa: geometria i obciążenie ($P=20\text{ kN}$).

gdzie $p_{l,j} \in \mathcal{P}_l^{(d)}$ $j = 1, \dots, 3$, $l = 1, \dots, 10$, a $v_{l,j}$, $j = 1, \dots, 3$, $l = 1, \dots, 10$ są dodatkowymi zmiennymi ciągłymi metody transformacji (por. roz. 3.2.1).

Ponieważ $n_l = 3$, $l = 1, \dots, 10$, to w przypadku wszystkich zmiennych $v_{l,j}$ przyjęto początkowe wartości $1/3$. Rozwiązanie optymalne o objętości 74391.5 cm^3 zostało uzyskane za pomocą algorytmu sekwencyjnego programowania kwadratowego NLPQL [106]. Projekt kratownicy otrzymany na podstawie zadania (3.36)-(3.42) przedstawia rysunek 3.4, a parametry przekrojów i współrzędne położenia węzłów ujęto w ostatniej kolumnie tablicy 3.1.

Porównując wyniki optymalizacji w przypadku parametrów ciągłych (rys. 3.3) i dyskretnych (rys. 3.4), widzimy, że uwzględniając dostępność materiałów konstrukcyjnych otrzymano wyraźnie inny projekt optymalny.

Przykład 3.2: Porównanie metody transformacji (roz. 3.2.1) i metody kontrolowanego przeglądu (roz. 3.2.2) przedstawimy na przykładzie optymalizacji dyskretniej 24 elementowej kratownicy, o zachowaniu silnie geometrycznie nieliniowym (rys. 3.5). W tym

zadaniu celem jest minimalizacja objętości konstrukcji, przy zachowaniu 26 ograniczeń niezawodnościowych, określonych przez następujące stany graniczne:

- Dopuszczalne przemieszczenie węzła środkowego:

$$g_1(\mathbf{q}(\mathbf{X}, \mathbf{A})) = 1 - \frac{|q_1(\mathbf{X}, \mathbf{A})|}{q^a}, \quad (3.43)$$

gdzie \mathbf{X} jest wektorem zmiennych losowych, przedstawionych w tabelicy 3.4, \mathbf{A} jest wektorem parametrów projektowych, odpowiadających polom przekrojów trzech grup elementów (por. tab. 3.4), q_1 jest przemieszczeniem pionowym węzła środkowego, a $q^a = 3.5$ cm jego wartością dopuszczalną.

- Dopuszczalne naprężenie/lokalna utrata stateczności elementów:

$$g_i(\mathbf{q}(\mathbf{X}, \mathbf{A}), \mathbf{X}, \mathbf{A}) = 1 - \frac{|\sigma_{i-1}(\mathbf{q}(\mathbf{X}, \mathbf{A}), \mathbf{X}, \mathbf{A})|}{\sigma_{i-1}^a(\mathbf{X}, \mathbf{A})}, \quad i = 2, \dots, 25. \quad (3.44)$$

- Globalna utrata stateczności:

$$g_{26}(\lambda_{cr}(\mathbf{X}, \mathbf{A})) = \lambda_{cr}(\mathbf{X}, \mathbf{A}) - 1, \quad (3.45)$$

gdzie λ_{cr} jest krytycznym mnożnikiem obciążenia, który powoduje awarię konstrukcji na skutek przeskoku.

Przyjęto założenie, że elementy kratownicy mają przekroje rurowe o stosunku średnicy zewnętrznej do wewnętrznej między 1.2 a 1.3. W konsekwencji, katalog możliwych do wykorzystania przekrojów ma 75 elementów. Efektywne zastosowanie analizowanych metod optymalizacji dyskretnej wymaga jednak, podobnie jak w przykładzie 3.1, określenia zredukowanych zbiorów wartości parametrów projektowych. Tak więc, optymalizację dyskretną poprzedzono rozwiązaniem zadania optymalizacji ciągłej, które w przypadku rozpatrywanego zagadnienia ma następującą postać:

$$\text{minimalizuj } C_I(\mathbf{A}) = \sum_{l=1}^3 L_l A_l, \quad (3.46)$$

przy ograniczeniach :

$$\beta_i(\mathbf{A}, \mathbf{X}) \geq 3.7, \quad i = 1, \dots, 25, \quad (3.47)$$

$$\beta_{26}(\mathbf{A}, \mathbf{X}) \geq 4.7, \quad (3.48)$$

$$3.13\text{cm}^2 \leq A_l \leq 30.76\text{cm}^2, \quad l = 1, 2, 3. \quad (3.49)$$

Minimalna, dopuszczalna wartość wskaźnika niezawodności w przypadku funkcji g_{26} jest większa niż dla pozostałych stanów granicznych, ze względu na bardziej istotne konsekwencje awarii w postaci globalnej utraty stateczności.

Rozwiązanie zadania (3.46)-(3.49) uzyskano za pomocą algorytmu NLPQL [106]. Optymalne wartości parametrów, w przypadku których objętość konstrukcji wynosi $1.0371 \cdot 10^5 \text{ cm}^3$, zamieszczono w kolumnie 7 tablicy 3.4.

Ograniczone zbiory dyskretnych zmiennych projektowych $\mathcal{P}_l^A = \{A_{l,1}, \dots, A_{l,6}\}$, $l = 1, 2, 3$, określono wybierając z wyjściowego katalogu po 6 elementów, o polach przekrojów najbliższych rozwiązaniu zadania ciągłego (3.46)-(3.49). Zbiory te zostały przedstawione w tablicy 3.3, gdzie liczby w oznaczeniach przekrojów odpowiadają średnicy zewnętrznej i grubości ścianki rur w milimetrach.

$A_{1,j}$		$A_{2,j}$		$A_{3,j}$	
Symbol	Pole [cm ²]	Symbol	Pole [cm ²]	Symbol	Pole [cm ²]
O76×8.5	18.02	O63.5×7	12.43	O70×7	13.85
O83×8	18.85	O70×6.5	12.97	O76×6.5	14.19
O83×8.5	19.89	O70×7	13.85	O76×7	15.17
O89×8	20.36	O76×6.5	14.19	O70×8	15.58
O83×9	20.92	O76×7	15.17	O83×7	16.71
O89×8.5	21.50	O70×8	15.58	O76×8	17.09

TABLICA 3.3. Katalogi przekrojów rurowych.

W przypadku metody transformacji, dla przyjętych zbiorów wartości parametrów dyskretnych, zadanie optymalizacji jest określone z wykorzystaniem 18 zmiennych typu v oraz 6 dodatkowych ograniczeń (w tym trzech określonych przez (3.13) i trzech przez (3.14)), które mają zapewnić równoważność rozwiązania. Objętość konstrukcji optymalnej otrzymanej tą metodą wyniosła $1.0826 \cdot 10^5 \text{ cm}^3$ (por. kol. 8 tab. 3.4). Zadanie rozwiązano algorytmem NLPQL, który wykonał 8 iteracji, zaczynając od wartości zmiennych v równych $1/6$.

W omawianym zadaniu, ograniczenia niezawodnościowe mogą być sprawdzone w przypadku wszystkich kombinacji wartości parametrów projektowych. Ich liczba jest, bowiem względnie mała - zaledwie $6^3 = 216$. Za pomocą algorytmu kontrolowanego przeglądu, omówionego w podrozdziale 3.2.2, otrzymano listę 184 kombinacji. Rozwiązanie zadania znajdowało się na jej 51 miejscu (por. kol. 9 tab. 3.4). Uzyskany projekt ma objętość $1.0762 \cdot 10^5 \text{ cm}^3$, a więc mniejszą niż ten otrzymany za pomocą metody transformacji. Tak więc w przypadku rozwiązania, otrzymanego metodą transformacji, algorytm optymalizacji zakończył działanie w minimum lokalnym.

Jak wspomniano, efektywność metody kontrolowanego przeglądu może być zwiększona poprzez wykorzystanie aproksymacji ograniczeń niezawodnościowych, zgodnie ze wzorem (3.28). Po zastosowaniu tej modyfikacji, lista kombinacji, w przypadku których miały być sprawdzone ograniczenia niezawodnościowe, zawierała projekt optymalny na pierwszym miejscu. Należy jednak podkreślić, że zastosowanie aproksymacji (3.28) w przypadku silnie nieliniowych ograniczeń, może prowadzić do wykluczenia z procesu dokładnej kontroli ograniczeń prawdziwego rozwiązania problemu.

Par. los.	Opis	P	Mom. rozkł.	Wart. pocz.	Par. proj.	Opt. ciągła	Opt. dyskretna	
							transf.	przeł.
1	2	3	4	5	6	7	8	9
X ₁	Przekr. elem. Nr.1-6 wsp. zmienn.: 5%	LN	E(X ₁)	15.0 cm ²	A ₁	19.37 cm ²	O83×8.5	O89×8
			σ(X ₁)	0.75 cm ²				
X ₂	Przekr. elem. Nr.7-12 wsp. zmienn.: 5%	LN	E(X ₂)	15.0 cm ²	A ₂	13.71 cm ²	O70×7	O70×6.5
			σ(X ₂)	0.75 cm ²				
X ₃	Przekr. elem. Nr.13-24 wsp. zmienn.: 5 %	LN	E(X ₃)	15.0 cm ²	A ₃	14.24 cm ²	1O76×7	O76×7
			σ(X ₃)	0.75 cm ²				
X ₄	Moduł sprężystości wszystkich elementów	LN	E(X ₄)	21000 $\frac{\text{kN}}{\text{cm}^2}$				
			σ(X ₄)	1050 $\frac{\text{kN}}{\text{cm}^2}$				
X ₅	Granica plastyczności wszystkich elementów	LN	E(X ₅)	21 $\frac{\text{kN}}{\text{cm}^2}$				
			σ(X ₅)	1 $\frac{\text{kN}}{\text{cm}^2}$				
X ₆	Wsp. x węzła Nr. 1	N	E(X ₆)	0.0 cm				
			σ(X ₆)	1.0 cm				
X ₇	Wsp. y węzła Nr. 1	N	E(X ₇)	0.0 cm				
			σ(X ₇)	1.0 cm				
X ₈	Wsp. z węzła Nr. 1	N	E(X ₈)	82.16 cm				
			σ(X ₈)	1.0 cm				
X ₉	Wsp. x węzła Nr. 2	N	E(X ₉)	250.0 cm				
			σ(X ₉)	1.0 cm				
X ₁₀	Wsp. y węzła Nr. 2	N	E(X ₁₀)	0.0 cm				
			σ(X ₁₀)	1.0 cm				
X ₁₁	Wsp. z węzła Nr. 2	N	E(X ₁₁)	62.16 cm				
			σ(X ₁₁)	1.0 cm				
X ₁₂	Wsp. x węzła Nr. 3	N	E(X ₁₂)	125.0 cm				
			σ(X ₁₂)	1.0 cm				
X ₁₃	Wsp. y węzła Nr. 3	N	E(X ₁₃)	216.51 cm				
			σ(X ₁₃)	1.0 cm				
X ₁₄	Wsp. z węzła Nr. 3	N	E(X ₁₄)	62.16 cm				
			σ(X ₁₄)	1.0 cm				
X ₁₅	Wsp. x węzła Nr. 4	N	E(X ₁₅)	-125.0 cm				
			σ(X ₁₅)	1.0 cm				
X ₁₆	Wsp. y węzła Nr. 4	N	E(X ₁₆)	216.51 cm				
			σ(X ₁₆)	1.0 cm				
X ₁₇	Wsp. z węzła Nr. 4	N	E(X ₁₇)	62.16 cm				
			σ(X ₁₇)	1.0 cm				
X ₁₈	Wsp. x węzła Nr. 5	N	E(X ₁₈)	-250.0 cm				
			σ(X ₁₈)	1.0 cm				
X ₁₉	Wsp. y węzła Nr. 5	N	E(X ₁₉)	0.0 cm				
			σ(X ₁₉)	1.0 cm				
X ₂₀	Wsp. z węzła Nr. 5	N	E(X ₂₀)	62.16 cm				
			σ(X ₂₀)	1.0 cm				
X ₂₁	Wsp. x węzła Nr. 6	N	E(X ₂₁)	-125.0 cm				
			σ(X ₂₁)	1.0 cm				
X ₂₂	Wsp. y węzła Nr. 6	N	E(X ₂₂)	-216.51 cm				
			σ(X ₂₂)	1.0 cm				
X ₂₃	Wsp. z węzła Nr. 6	N	E(X ₂₃)	62.16 cm				
			σ(X ₂₃)	1.0 cm				
X ₂₄	Wsp. x węzła Nr. 7	N	E(X ₂₄)	125.0 cm				
			σ(X ₂₄)	1.0 cm				
X ₂₅	Wsp. y węzła Nr. 7	N	E(X ₂₅)	-216.51 cm				
			σ(X ₂₅)	1.0 cm				
X ₂₆	Wsp. z węzła Nr. 7	N	E(X ₂₆)	62.16 cm				
			σ(X ₂₆)	1.0 cm				
X ₂₇	Współ. obciążenia	LN	E(X ₂₇)	1.0				
			σ(X ₂₇)	0.15				

TABLICA 3.4. Kopuła kratownicowa: zmienne losowe i projektowe. Kolumna 3 zawiera informacje o typach rozkładów prawdopodobieństwa: LN - log-normalny, N - normalny, G - Gumbela

Pola losowe w analizie niezawodności

4.1. Wstęp

W omówionym wcześniej sformułowaniu problemu analizy niezawodności, do reprezentacji niepewnych parametrów konstrukcyjnych wykorzystano zmienne losowe. W ramach takiego podejścia przyjmuje się założenie, że wartość parametru jest określona dla całej konstrukcji, bądź jej wyodrębnionego elementu. Zakłada się również, że wartości parametrów konstrukcyjnych są niezależne od czasu. W przypadku niektórych problemów, przyjęcie takich założeń jest znacznym uproszczeniem, mogącym prowadzić do błędnej oceny poziomu bezpieczeństwa. Jako przykład można podać analizę konstrukcji wykonanych z materiałów o dużej niejednorodności (konstrukcje drewniane, podłoże gruntowe) lub takich, które zmieniają swoje właściwości w wyniku starzenia się. Zróżnicowanie w obrębie konstrukcji wykazują także imperfekcje geometrii i obciążenia. Zazwyczaj zmienne w czasie są wartości obciążeń; nawet jeżeli zmiany te są na tyle powolne, że nie powodują efektów dynamicznych, to mogą one mieć istotny wpływ na ocenę niezawodności. Uwzględnienie w analizie niezawodności przestrzennej zmienności losowych parametrów konstrukcyjnych, wymaga zupełnie innych metod niż modelowanie zmienności tych parametrów w czasie. Jednak modele zależności parametrów losowych od położenia jak i czasu opierają się na wykorzystaniu *funkcji losowych*.

Niech Ω będzie przestrzenią metryczną, a (Γ, \mathcal{E}, P) przestrzenią probabilistyczną, gdzie Γ jest przestrzenią zdarzeń elementarnych, \mathcal{E} jest σ -algebrą podzbiorów Γ , a P jest miarą prawdopodobieństwa. Przekształcenie $H : \Omega \times \Gamma \rightarrow \mathbf{R}^n$ nazywamy *funkcją losową*, jeżeli w przypadku każdego ustalonego $z \in \Omega$, funkcja $H(z, \gamma)$ traktowana jako funkcja argumentu γ , jest zmienną losową.

Funkcja losowa jest więc odwzorowaniem przekształcającym Ω w przestrzeń zmiennych losowych, określonych na przestrzeni probabilistycznej (Γ, \mathcal{E}, P) . Czasami można się również spotkać z określeniem, że funkcja losowa jest indeksowaną przez parametr $z \in \Omega$ rodziną zmiennych losowych $H_\Omega = \{H(z), z \in \Omega\}$ określonych na przestrzeni (Γ, \mathcal{E}, P) .

Funkcję losową można postrzegać w jeszcze w inny sposób. Zauważmy, że w przypadku ustalonych zdarzeń elementarnych γ , funkcja losowa określa nie mającą charakteru losowego funkcję $h(z)$, $z \in \Omega$, którą nazywamy *realizacją funkcji losowej*.

Realizacje funkcji losowej można traktować jako elementy przestrzeni \mathcal{X} wszystkich (spełniających pewne warunki) funkcji zmiennej $z \in \Omega$. Przestrzeń \mathcal{X} jest nazywana *przestrzenią realizacji funkcji losowej*. Zauważmy, że funkcję losową można postrzegać jako uogólnienie zmiennej losowej. Zmienna losowa jest funkcją przekształcającą przestrzeń zdarzeń elementarnych w zbiór liczb rzeczywistych, natomiast funkcja losowa przyporządkowuje zdarzeniom elementarnym funkcje zmiennej $z \in \Omega$, będące elementami przestrzeni \mathcal{X} .

Funkcje losowe mogą być klasyfikowane na podstawie własności ich dziedziny i przeciwdziedziny. Poniżej, podane zostaną jedynie określenia typów stosowanych w analizie niezawodności. Więcej informacji na temat klasyfikacji funkcji losowych można znaleźć, na przykład, w monografii [124].

Funkcje losowe $H(t, \gamma)$, $\gamma \in \Gamma$, $t \in T$, ze skalarnym parametrem t , $T \in \mathbf{R}$, utożsamianym z czasem, nazywa się *procesami losowymi*, bądź *procesami stochastycznymi*. W przypadku modeli zmiennych w czasie parametrów konstrukcyjnych lub parametrów obciążeń, zazwyczaj przyjmuje się, że zbiór T jest przedziałem w zbiorze liczb rzeczywistych. Kiedy brak dostatecznej ilości danych, można jednak budować modele, wykorzystujące przeliczalny zbiór T . Wtedy wartości procesu losowego utożsamia się z obserwacjami wybranego parametru w ustalonych chwilach czasu. Procesy stochastyczne określone na przeliczalnym zbiorze T również wykorzystuje jedna z adaptacyjnych metod symulacyjnych, które przedstawiono w dalszej części pracy (por. 8.1).

Funkcje losowe $H(\mathbf{z}, \gamma)$, $\gamma \in \Gamma$, $\mathbf{z} \in \Omega$, gdzie Ω jest zwartym zbiorem w przestrzeni euklidesowej punktowej \mathcal{E}_n i w przypadku, których z utożsamiane jest ze współrzędną przestrzenną, nazywa się *polami losowymi*. W analizie niezawodności pola losowe zazwyczaj stosuje się do reprezentacji zmienności parametrów projektowych w obszarze konstrukcji. W przypadku pól losowych \mathcal{E}_n zazwyczaj jest wielowymiarowa. Jednak konstrukcje prętowe wygodnie jest modelować za pomocą pól losowych, określonych na parametrze skalarnym $z \in \mathbf{R}$.

Ponieważ w analizie niezawodności pola losowe i procesy losowe są wykorzystywane w różny sposób, to ich omówienie zostanie podzielone na dwa rozdziały. W tym rozdziale przedstawimy zastosowanie skalarnych pól losowych do modelowania przestrzennej zmienności parametrów konstrukcyjnych. Natomiast przedmiotem rozdziału 5 jest analiza niezawodności zależnej od czasu, wykorzystująca procesy losowe.

4.2. Wybrane własności skalarnych pól losowych

Zgodnie z definicją podaną we wstępie, pole losowe o wartościach skalarnych możemy postrzegać jako rodzinę jednowymiarowych zmiennych losowych $H(\mathbf{z})$ określoną przez współrzędną \mathbf{z} . Przyjmijmy założenie, że \mathbf{z} należy do zwartego zbioru $\Omega \subset \mathcal{E}^n$, gdzie przestrzeń euklidesowa punktowa \mathcal{E}^n jest związana z opisywanym przedmiotem materialnym:

$$H(\mathbf{z}) \in \mathbf{R}, \quad \mathbf{z} = \{z_1, \dots, z_n\} \in \Omega \subset \mathcal{E}^n. \quad (4.1)$$

Zmienna losowa $H(\mathbf{z})$ jest nazywana *wartością pola losowego* w punkcie \mathbf{z} .

Omawiając poniżej własności pól losowych przyjmijmy założenie, że są to tzw. *pola losowe drugiego rzędu*. Pole losowe $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, nazywamy polem losowym drugiego rzędu, bądź *połem losowym Hilberta*, jeżeli spełniony jest następujący warunek:

$$\mathbb{E} [|H(\mathbf{z})|^2] < \infty. \quad (4.2)$$

Jeżeli spełniony jest powyższy warunek, to istnieją momenty pierwszego i drugiego rzędu wartości zmiennych losowych $H(\mathbf{z})$, $\mathbf{z} \in \Omega$. W przypadku pola drugiego rzędu można więc określić tzw. *funkcję wartości oczekiwanej*:

$$\mu_H(\mathbf{z}) = \mathbb{E}[H(\mathbf{z})], \quad (4.3)$$

oraz *funkcję wariancji*:

$$\text{Var} [H(\mathbf{z})] = \sigma_H^2(\mathbf{z}) = \mathbb{E} \left[(H(\mathbf{z}) - \mu_H(\mathbf{z}))^2 \right], \quad (4.4)$$

gdzie $\sigma_H(\mathbf{z})$ jest funkcją odchylenia standardowego.

Wartości skalarnego pola losowego $H(\mathbf{z}^{(1)})$ oraz $H(\mathbf{z}^{(2)})$, $\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \in \Omega$, można traktować jako składowe dwuwymiarowej zmiennej losowej. Jeżeli H jest polem drugiego rzędu, to istnieje kowariancja $H(\mathbf{z}^{(1)})$ i $H(\mathbf{z}^{(2)})$. Podobnie, jak w przypadku wartości oczekiwanej i wariancji, możemy więc wprowadzić *funkcję kowariancji* pola losowego (ang. *auto-covariance function*) $K_H : \Omega \times \Omega \rightarrow \mathbf{R}$ określoną jako:

$$\begin{aligned} K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \text{Cov} \left[H(\mathbf{z}^{(1)}), H(\mathbf{z}^{(2)}) \right] \\ &= \mathbb{E} \left[\left(H(\mathbf{z}^{(1)}) - \mu_H(\mathbf{z}^{(1)}) \right) \left(H(\mathbf{z}^{(2)}) - \mu_H(\mathbf{z}^{(2)}) \right) \right] \end{aligned} \quad (4.5)$$

Na podstawie funkcji wariancji i kowariancji, definiuje się *funkcję korelacji* pola losowego (ang. *auto-correlation function*):

$$\rho_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = \frac{K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\sigma_H(\mathbf{z}^{(1)}) \sigma_H(\mathbf{z}^{(2)})}. \quad (4.6)$$

Ważną klasę stanowią pola losowe, których charakterystyki probabilistyczne są niezależne od parametru położenia. Pole losowe $H(\mathbf{z})$ w przypadku, którego istnieją funkcje wartości oczekiwanej i kowariancji określa się jako *ślabo stacjonarne* (ang. *weakly homogenous*) jeżeli

$$\mu_H(\mathbf{z}) = \text{const} \quad \text{oraz} \quad K_H(\mathbf{z}, \mathbf{z} + \boldsymbol{\eta}) = K_H(\boldsymbol{\eta}), \quad (4.7)$$

gdzie $\mathbf{z}, \mathbf{z} + \boldsymbol{\eta} \in \Omega$. Stacjonarne pole losowe $H(\mathbf{z})$ jest *izotropowe*, gdy jego funkcja kowariancji zależy jedynie od odległości między punktami:

$$K_H(\mathbf{z}, \mathbf{z} + \boldsymbol{\eta}) = K_H(\|\boldsymbol{\eta}\|), \quad (4.8)$$

gdzie $\|\boldsymbol{\eta}\|$ jest długością wektora $\boldsymbol{\eta}$.

W teorii, jak również w zastosowaniach, bardzo ważną rolę odgrywają pola losowe, które można zdefiniować za pomocą rozkładu normalnego. Pole losowe $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, określa się jako *pole Gaussa* (gaussowskie, normalne) wtedy i tylko wtedy, gdy każda skończona kombinacja liniowa postaci

$$G = \sum_{i=1}^n \alpha_i H(\mathbf{z}^{(i)}) \quad (4.9)$$

jest zmienną losową o rozkładzie normalnym (por. [111]). Z powyższej definicji wynika, że wartościami gaussowskiego pola losowego $H(\mathbf{z})$, dla każdego $\mathbf{z} \in \Omega$, są zmienne losowe o rozkładzie normalnym. Istotną cechą gaussowskich pól losowych jest możliwość ich jednoznacznego określenia za pomocą funkcji wartości oczekiwanej (4.3) i funkcji kowariancji (4.5). Oczywiście, zamiast funkcji kowariancji, możemy wykorzystać funkcję korelacji (4.6) i funkcję wariancji (4.4).

Stacjonarne, gaussowskie pola losowe często są definiowane za pomocą funkcji korelacji. Na ogół w praktyce spotyka się modele, w których korelacja między wartościami pola losowego $H(\mathbf{z}^{(1)})$ i $H(\mathbf{z}^{(2)})$ maleje wraz z odległością między punktami - $\delta = \|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|$. W takich przypadkach zazwyczaj wykorzystuje się następujące funkcje korelacji:

- trójkątna funkcja korelacji

$$\rho(\delta) = \begin{cases} 1 - \frac{\delta}{a}, & \delta \leq a, \\ 0, & \delta > a, \end{cases} \quad (4.10)$$

gdzie a jest tak zwaną *długością korelacji*;

- wykładnicza funkcja korelacji

$$\rho(\delta) = \exp\left(-\frac{\delta}{a}\right); \quad (4.11)$$

- gaussowska (wykładniczo kwadratowa) funkcja korelacji

$$\rho(\delta) = \exp\left(-\frac{\delta^2}{a^2}\right). \quad (4.12)$$

Występującą w powyższych wzorach długość korelacji a , formalnie określa się jako

$$a = \frac{\int_0^\infty \delta |K_H(\delta)| d\delta}{\int_0^\infty |K_H(\delta)| d\delta}. \quad (4.13)$$

Istnieją problemy, w przypadku których możliwa jest obserwacja realizacji niektórych wartości pól losowych. Przykładem mogą tu być zagadnienia mechaniki gruntów. Modele podłoża wykorzystujące pola losowe, można bowiem uzupełnić wynikami badań parametrów geotechnicznych, przeprowadzonych w określonych miejscach. W modelowaniu imperfekcji geometrycznych istotne jest na ogół, uwzględnienie warunków podparcia konstrukcji. W takim przypadku można przyjąć, że przemieszczenia podpór są deterministyczne, a więc realizacje wartości pola losowego w punktach podparcia są określone.

Jedną z metod definiowania pól losowych, których realizacje przyjmują żądane wartości w pewnych punktach, jest "warunkowanie" stacjonarnych pól losowych [80, 26]. Przyjmijmy, że $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, jest stacjonarnym polem losowym. Wartości tego pola $H(\mathbf{z}), H(\mathbf{z}^{(1)}), \dots, H(\mathbf{z}^{(k)})$ mają pewien łączny rozkład prawdopodobieństwa. Niech zmienne losowe $H(\mathbf{z}^{(1)}), \dots, H(\mathbf{z}^{(k)})$ przyjmują odpowiednio następujące realizacje $h(\mathbf{z}^{(1)}), \dots, h(\mathbf{z}^{(k)})$. "Warunkowanie" pola losowego $H(\mathbf{z})$, $\mathbf{z} \in \Omega$ polega na określeniu pola losowego $G(\mathbf{z})$, $\mathbf{z} \in \Omega$, którego wartości mają warunkowy rozkład prawdopodobieństwa $H(\mathbf{z})$ pod warunkiem $H(\mathbf{z}^{(1)}) = h(\mathbf{z}^{(1)}), \dots, H(\mathbf{z}^{(k)}) = h(\mathbf{z}^{(k)})$. Warto zauważyć, że pole losowe G nie jest stacjonarne, a jego funkcja wariancji ma zerowe wartości w punktach $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(k)}$.

Przykład 4.1: Niech $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, będzie stacjonarnym, gaussowskim polem losowym. Zdefiniujmy pole losowe $G(\mathbf{z})$, $\mathbf{z} \in \Omega$, którego wartości mają warunkowe rozkłady prawdopodobieństwa wartości pola H pod warunkiem $H(\mathbf{z}^{(1)}) = h(\mathbf{z}^{(1)}), H(\mathbf{z}^{(2)}) = h(\mathbf{z}^{(2)}), \dots, H(\mathbf{z}^{(m)}) = h(\mathbf{z}^{(m)})$. Ponieważ pole losowe $G(\mathbf{z})$, $\mathbf{z} \in \Omega$, jest polem gaussowskim, to można je określić poprzez odpowiednie, warunkowe, funkcje wartości oczekiwanej i kowariancji.

Aby wyznaczyć funkcje wartości oczekiwanej i kowariancji pola G , wyprowadzimy wzory, określające warunkową wartość oczekiwaną i warunkową kowariancję normalnych zmiennych losowych. Przyjmijmy, że wektor losowy \mathbf{X} jest związany z wektorem losowym \mathbf{Y} za pomocą transformacji liniowej określonej przez macierz \mathbf{F} o wymiarach $2 \times (n - 2)$:

$$X_1 = Y_1 + \sum_{i=3}^n F_{1i} Y_i \quad (4.14)$$

$$X_2 = Y_2 + \sum_{i=3}^n F_{2i} Y_i \quad (4.15)$$

$$X_i = Y_i, \quad i = 3, \dots, n$$

Założmy ponadto, że spełnione są następujące warunki

$$\text{Cov}[X_j, X_i] = 0, \quad j = 1, 2, \quad i = 3, \dots, n. \quad (4.16)$$

Warunki (4.16) implikują następujące układy równań liniowych (dla $j = 1, 2$) dzięki którym możemy wyznaczyć składowe macierzy transformacji \mathbf{F} :

$$\begin{bmatrix} \text{Cov}[Y_3, Y_3] & \text{Cov}[Y_3, Y_4] & \cdots & \text{Cov}[Y_3, Y_n] \\ \text{Cov}[Y_4, Y_3] & \text{Cov}[Y_4, Y_4] & \cdots & \text{Cov}[Y_4, Y_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[Y_n, Y_3] & \text{Cov}[Y_n, Y_4] & \cdots & \text{Cov}[Y_n, Y_n] \end{bmatrix} \begin{bmatrix} F_{j3} \\ F_{j4} \\ \vdots \\ F_{jn} \end{bmatrix} = - \begin{bmatrix} \text{Cov}[Y_3, Y_j] \\ \text{Cov}[Y_4, Y_j] \\ \vdots \\ \text{Cov}[Y_n, Y_j] \end{bmatrix}, \quad (4.17)$$

gdzie $j = 1, 2$. Wprowadzając oznaczenie \mathbf{C}_{jY} , $j = 1, 2$, dla wektora prawych stron, a \mathbf{C}_{YY} dla macierzy układu równań (4.17), wiersze macierzy transformacji \mathbf{F}_j , $j = 1, 2$, (4.14) możemy wyznaczyć za pomocą

$$\mathbf{F}_j^T = -\mathbf{C}_{jY}^{-1} \mathbf{C}_{YY}, \quad j = 1, 2. \quad (4.18)$$

Korzystając z (4.18), wartość oczekiwaną zmiennych X_j , $j = 1, 2$, wyrażamy przez

$$\mathbb{E}[X_j] = \mathbb{E}[Y_j] + \mathbf{F}_j \boldsymbol{\mu}_Y = \mathbb{E}[Y_j] - \mathbf{C}_{jY}^T \mathbf{C}_{YY}^{-1} \boldsymbol{\mu}_Y, \quad j = 1, 2, \quad (4.19)$$

gdzie $\boldsymbol{\mu}_Y = \{\mathbb{E}[Y_3], \dots, \mathbb{E}[Y_n]\}$ jest wektorem wartości oczekiwanych zmiennych Y_i , $i =$

$3, \dots, n$. Wariancja X_j , $j = 1, 2$, jest następującej postaci:

$$\begin{aligned}
 \text{Var}[X_j] &= \text{Var}\left[Y_j + \sum_{i=3}^n F_{ji}Y_i\right] & (4.20) \\
 &= \text{Var}[Y_j] + 2\sum_{i=3}^n F_{ji}\text{Cov}[Y_j, Y_i] + \sum_{i=3}^n \sum_{k=3}^n F_{ji}\text{Cov}[Y_i, Y_k]F_{jk} \\
 &= \text{Var}[Y_1] + 2\mathbf{F}_j\mathbf{C}_{jY} + \mathbf{F}_j\mathbf{C}_{YY}\mathbf{F}_j^T \\
 &= \text{Var}[Y_j] - 2\mathbf{C}_{jY}^T\mathbf{C}_{YY}^{-1}\mathbf{C}_{jY} + \mathbf{C}_{jY}^T\mathbf{C}_{YY}^{-1}\mathbf{C}_{YY}\mathbf{C}_{YY}^{-1}\mathbf{C}_{jY} \\
 &= \text{Var}[Y_j] - \mathbf{C}_{jY}^T\mathbf{C}_{YY}^{-1}\mathbf{C}_{jY},
 \end{aligned}$$

a kowariancja X_1 i X_2 jest określona jako:

$$\begin{aligned}
 \text{Cov}[X_1, X_2] &= \text{Cov}\left[Y_1 + \sum_{i=3}^n F_{1i}Y_i, Y_2 + \sum_{i=3}^n F_{2i}Y_i\right] & (4.21) \\
 &= \text{Cov}[Y_1, Y_2] + \sum_{i=3}^n F_{1i}\text{Cov}[Y_2, Y_i] + \sum_{i=3}^n F_{2i}\text{Cov}[Y_1, Y_i] \\
 &\quad + \sum_{i=3}^n \sum_{k=3}^n F_{1i}\text{Cov}[Y_i, Y_k]F_{2k} \\
 &= \text{Cov}[Y_1, Y_2] + \mathbf{F}_1\mathbf{C}_{2Y} + \mathbf{F}_2\mathbf{C}_{1Y} + \mathbf{F}_1\mathbf{C}_{YY}\mathbf{F}_2^T \\
 &= \text{Cov}[Y_1, Y_2] - \mathbf{C}_{1Y}\mathbf{C}_{YY}^{-1}\mathbf{C}_{2Y}.
 \end{aligned}$$

(4.22)

Ponieważ \mathbf{X} jest liniową transformacją \mathbf{Y} , to jego rozkład prawdopodobieństwa jest rozkładem normalnym. Korzystając z założenia (4.16) oraz (4.19), (4.20) rozkład prawdopodobieństwa \mathbf{X} możemy przedstawić w następującej postaci:

$$\varphi(\mathbf{x}, \boldsymbol{\mu}_X, \mathbf{C}_X) = \varphi(\mathbf{x}_{12}, \boldsymbol{\mu}_{12}, \mathbf{C}_{12})\varphi(\mathbf{y}, \boldsymbol{\mu}_Y, \mathbf{C}_{YY}), \quad (4.23)$$

gdzie $\boldsymbol{\mu}_X$, \mathbf{C}_X są, odpowiednio, wektorem wartości oczekiwanych i macierzą kowariancji \mathbf{X} , $\boldsymbol{\mu}_{12}$ jest wektorem wartości oczekiwanych wektora losowego $\{X_1, X_2\}$, a \mathbf{C}_{12} jest jego macierzą kowariancji, o składowych określonych przez (4.20) i (4.21).

Korzystając z (4.14), wektor losowy $\mathbf{X}_{12} = \{X_1, X_2\}$ możemy przedstawić za pomocą $\mathbf{Y}_{12} = \{Y_1, Y_2\}$ i $\mathbf{Y}_{3n} = \{Y_3, \dots, Y_n\}$ jako:

$$\mathbf{X}_{12} = \mathbf{Y}_{12} + \mathbf{F}\mathbf{Y}_{3n} = \mathbf{Y}_{12} - \mathbf{C}_{12Y}^T\mathbf{C}_{YY}^{-1}\mathbf{Y}_{3n}. \quad (4.24)$$

Podstawienie powyższej zależności do pierwszego członu prawej strony wyrażenia (4.23)

pozwała uzyskać rozkład warunkowy \mathbf{Y}_{12} pod warunkiem $\mathbf{Y}_{3n} = \mathbf{y}_{3n}$:

$$\begin{aligned} \varphi(\mathbf{x}_{12}, \boldsymbol{\mu}_{12}, \mathbf{C}_{12}) &= \\ &= \varphi(\mathbf{Y}_{12} - \mathbf{C}_{12Y}^T \mathbf{C}_{YY}^{-1} \mathbf{y}_{3n}, \boldsymbol{\mu}_{12Y} - \mathbf{C}_{12Y}^T \mathbf{C}_{YY}^{-1} \boldsymbol{\mu}_{3nY}, \mathbf{C}_{12}) \\ &= \varphi(\mathbf{Y}_{12}, \boldsymbol{\mu}_{12Y} + \mathbf{C}_{12Y}^T \mathbf{C}_{YY}^{-1} (\mathbf{y}_{3n} - \boldsymbol{\mu}_{3nY}), \mathbf{C}_{12}). \end{aligned} \quad (4.25)$$

Z wyrażenia (4.25) wynika wzór na funkcję wartości oczekiwanej $G(\mathbf{z})$, $\mathbf{z} \in \Omega$, która jest warunkową funkcją wartości oczekiwanej $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, pod warunkiem $H(\mathbf{z}^{(1)}) = h(\mathbf{z}^{(1)}), H(\mathbf{z}^{(2)}) = h(\mathbf{z}^{(2)}), \dots, H(\mathbf{z}^{(m)}) = h(\mathbf{z}^{(m)})$:

$$\begin{aligned} \mu_G(\mathbf{z}) &= \mathbb{E} \left[H(\mathbf{z}) | H(\mathbf{z}^{(1)}), H(\mathbf{z}^{(2)}), \dots, H(\mathbf{z}^{(m)}) \right] \\ &= \mu_H(\mathbf{z}) + \\ &\quad + \sum_{i=1}^m \sum_{j=1}^m \text{Cov} \left[H(\mathbf{z}), H(\mathbf{z}^{(i)}) \right] \text{Cov}^{-1} \left[H(\mathbf{z}^{(i)}), H(\mathbf{z}^{(j)}) \right] \left(h(\mathbf{z}^{(j)}) - \mu_H(\mathbf{z}^{(j)}) \right). \end{aligned} \quad (4.26)$$

Z (4.25) można również otrzymać wyrażenie określające funkcję kowariancji pola losowego G :

$$\begin{aligned} K_G(\mathbf{z}^{(01)}, \mathbf{z}^{(02)}) &= K_H(\mathbf{z}^{(01)}, \mathbf{z}^{(02)} | H(\mathbf{z}^{(1)}), \dots, H(\mathbf{z}^{(m)})) \\ &= K_H(\mathbf{z}^{(01)}, \mathbf{z}^{(02)}) + \\ &\quad - \sum_{i=1}^m \sum_{j=1}^m \text{Cov} \left[H(\mathbf{z}^{(01)}), H(\mathbf{z}^{(i)}) \right] \\ &\quad \text{Cov}^{-1} \left[H(\mathbf{z}^{(i)}), H(\mathbf{z}^{(j)}) \right] \text{Cov} \left[H(\mathbf{z}^{(02)}), H(\mathbf{z}^{(j)}) \right], \end{aligned} \quad (4.27)$$

gdzie $K_H(\mathbf{z}^{(01)}, \mathbf{z}^{(02)} | \cdot)$, $\mathbf{z}^{(01)}, \mathbf{z}^{(02)} \in \Omega$, jest warunkową funkcją kowariancji pola losowego H .

4.3. Wybrane analityczne własności realizacji pól losowych

W przypadku funkcji losowych również możemy określić takie pojęcia, jak ciągłość czy pochodna. Jednak w odróżnieniu od klasycznej analizy matematycznej, w analizie funkcji losowych pojęcia te mogą mieć różne interpretacje. Fakt ten jest konsekwencją istnienia alternatywnych definicji funkcji losowej: jako indeksowanej rodziny zmiennych losowych i jako zbioru realizacji. W przypadku posługiwania się pierwszą z definicji,

pojęcie zbieżności funkcji losowej i pojęcie zbieżności ilorazu różnicowego wykorzystuje pojęcia zbieżności probabilistycznej. Tak więc mówimy o ciągłości pochodnej w sensie średniokwadratowym, prawie na pewno i według prawdopodobieństwa [110]. Natomiast traktując funkcję losową jako przekształcenie przestrzeni zdarzeń elementarnych w przestrzeń realizacji, mówimy o ciągłości i pochodnych realizacji funkcji losowych. W tym przypadku, dla każdej realizacji, będziemy posługiwać się pojęciami znanymi z klasycznej analizy matematycznej.

W tej części pracy skupimy się na omówieniu analitycznych własności realizacji pól losowych. Naszym celem będzie podanie warunków, jakie powinny spełniać pola losowe, aby ich realizacje były ciągłe bądź różniczkowalne. Zagadnienie to jest szczególnie istotne w przypadku modelowania imperfekcji geometrycznych. W niektórych problemach istotne jest bowiem, aby realizacje pola losowego modelujące zaburzenia geometrii były ciągłe bądź różniczkowalne.

Podane poniżej twierdzenia są prawdziwe w przypadku tzw. *ośrodkowych pól losowych*. Mówiąc w skrócie, wszystkie własności probabilistyczne ośrodkowego pola losowego są określone przez jego wartości w przeliczalnym, gęstym podziorze punktów obszaru Ω . Na przykład, jeżeli pole losowe nie jest ośrodkowe, to jego supremum może nie być dobrze zdefiniowaną zmienną losową. Bardziej precyzyjną definicję ośrodkowości pól losowych można znaleźć w [3].

Pole losowe $H(\mathbf{z}, \gamma)$, $\gamma \in \Gamma$, $\mathbf{z} \in \Omega$, jest ciągłe na Ω , jeżeli jego realizacje są ciągłe na tym zbiorze prawie na pewno [110]. Tak więc w przypadku każdej sekwencji \mathbf{z}_n takiej, że $\|\mathbf{z}_n - \mathbf{z}\| \rightarrow 0$, przy $n \rightarrow \infty$, zachodzi

$$P\{\gamma : |H(\mathbf{z}_n, \gamma) - H(\mathbf{z}, \gamma)| \rightarrow 0 \text{ przy } n \rightarrow \infty \forall \mathbf{z} \in \Omega\} = 1. \quad (4.28)$$

Dość restrykcyjny dostateczny warunek ciągłości realizacji pól losowych, jako wniosek z bardziej ogólnego twierdzenia, został podany w pracy [10], ale jego wyprowadzenie można również znaleźć w [4]. Pole losowe $H(\mathbf{z}, \gamma)$, $\gamma \in \Gamma$, $\mathbf{z} \in \Omega$, ma ciągłe realizacje na Ω (zwartym podziorze \mathcal{E}_n) z prawdopodobieństwem 1, jeżeli dla $c > 0$, $\alpha > 0$ i $\eta > \alpha$ spełniona jest nierówność:

$$\mathbb{E}[|H(\mathbf{z}) - H(\mathbf{z} + \zeta)|^\alpha] \leq \frac{c\|\zeta\|^{2n}}{|\log \|\zeta\||^{1+\eta}}. \quad (4.29)$$

W przypadku gaussowskich pól losowych o zerowej funkcji wartości oczekiwanej warunek (4.29) można sformułować nieco precyzyjniej [4]. Niech $\tilde{H}(\mathbf{z}, \gamma)$, $\gamma \in \Gamma$, $\mathbf{z} \in \Omega$, będzie gaussowskim polem losowym o zerowej funkcji wartości oczekiwanej i ciągłej funkcji kowariancji. Jeżeli dla pewnych $0 < c < \infty$ oraz $\alpha > 0$,

$$\mathbb{E}\left[\left|\tilde{H}(\mathbf{z}) - \tilde{H}(\mathbf{z} + \zeta)\right|^2\right] \leq \frac{c}{|\log \|\zeta\||^{1+\alpha}} \quad (4.30)$$

dla wszystkich $\mathbf{z}, \mathbf{z} + \boldsymbol{\zeta} \in \Omega$ takich, że $\|\boldsymbol{\zeta}\| < 1$, to pole \tilde{H} ma ciągle realizacje na Ω z prawdopodobieństwem 1. Zauważmy, że gaussowskie pole losowe można przedstawić za pomocą pola o zerowej wartości oczekiwanej na podstawie wyrażenia

$$\tilde{H}(\mathbf{z}) = H(\mathbf{z}) - \mu_H(\mathbf{z}) \quad (4.31)$$

gdzie μ_H jest funkcją wartości oczekiwanej H . Tak więc, jeżeli pole H ma ciągłą funkcję kowariancji K_H , to podstawiając (4.31) do (4.30) uzyskujemy dostateczny warunek ciągłości jego realizacji

$$\sigma_H^2(\mathbf{z}) + \sigma_H^2(\mathbf{z} + \boldsymbol{\zeta}) - 2K_H(\mathbf{z}, \mathbf{z} + \boldsymbol{\zeta}) \leq \frac{c}{|\log \|\boldsymbol{\zeta}\||^{1+\alpha}}, \quad (4.32)$$

gdzie $c, \alpha, \boldsymbol{\zeta}$ spełniają założenia takie, jak w przypadku (4.30).

Przedstawimy teraz wnioski z analizy własności warunku (4.32), jaką można znaleźć w pracy [2]. Na rysunku 4.1 zamieszczono wykresy funkcji będącej prawą stroną wyrażenia (4.32) ($c = 1$ i różne wartości α). Zauważmy, że wykresy tej funkcji zbiegają do zera prawie pionowo. Można więc przypuszczać, że jej pochodna w zerze jest nieskończona. Istotnie, pochodna prawej strony (4.32) - $\partial |\log \|\boldsymbol{\zeta}\||^{-(1+\alpha)} / \partial \|\boldsymbol{\zeta}\|$ dąży do ∞ , kiedy $\|\boldsymbol{\zeta}\| \rightarrow 0$. Tak więc, jeżeli funkcja kowariancji ma skończoną pochodną przy $\|\boldsymbol{\zeta}\| \rightarrow 0$, to istnieje skończone c , zapewniające spełnienie warunku (4.32).

Przyjmując założenie, że realizacje pola losowego $H(\mathbf{z}, \gamma)$, $\mathbf{z} \in \Omega \subset \mathcal{E}_n$, $\gamma \in \Gamma$, są różniczkowalne, możemy określić jego gradient. Jest to wektorowe pole losowe

$$\nabla H(\mathbf{z}, \gamma) \in \mathfrak{E}^n, \quad \mathbf{z} \in \Omega, \gamma \in \Gamma, \quad (4.33)$$

gdzie \mathfrak{E}^n jest n -wymiarową wektorową przestrzenią euklidesową, którego realizacje, w przypadku ustalonych zdarzeń elementarnych γ , są gradientami realizacji pola H . W bazie kartezjańskiej $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ realizacje gradientu można więc przedstawić w następujący sposób:

$$\nabla h(\mathbf{z}, \gamma) = \sum_{i=1}^n \frac{\partial h(\mathbf{z}, \gamma)}{\partial z_i} \mathbf{e}_i. \quad (4.34)$$

Warto zauważyć, że ponieważ operacja różniczkowania jest przekształceniem liniowym, to gradient gaussowskiego pola losowego jest wektorowym polem gaussowskim.

Wartości oczekiwane składowych gradientu $\nabla H_i(\mathbf{z})$, $i = 1, \dots, n$, są pochodnymi cząstkowymi funkcji $\mu_H(\mathbf{z})$. Jeżeli prawie wszystkie realizacje pola losowego są różnicz-

gdzie

$$\nabla \tilde{H}_i(\mathbf{z}) = \nabla H_i(\mathbf{z}) - \frac{\partial \mu_H(\mathbf{z})}{\partial z_i}. \quad (4.37)$$

Przenosząc granice przed operację wartości oczekiwanej, otrzymujemy następujący związek między (4.36) i funkcją kowariancji pola H :

$$\begin{aligned} K_{H\nabla H_i}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)}) \nabla \tilde{H}_i(\mathbf{z}^{(2)}) \right] \\ &= \lim_{\Delta z_i \rightarrow 0} \mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)}) \frac{\tilde{H}(\mathbf{z}^{(2)} + \Delta z_i \mathbf{e}_i) - \tilde{H}(\mathbf{z}^{(2)})}{\Delta z_i} \right] \\ &= \lim_{\Delta z_i \rightarrow 0} \frac{\mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)}) \tilde{H}(\mathbf{z}^{(2)} + \Delta z_i \mathbf{e}_i) \right] - \mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)}) \tilde{H}(\mathbf{z}^{(2)}) \right]}{\Delta z_i} \\ &= \lim_{\Delta z_i \rightarrow 0} \frac{K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)} + \Delta z_i \mathbf{e}_i) - K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\Delta z_i} \\ &= \frac{\partial K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\partial z_i^{(2)}}, \quad i = 1, \dots, n. \end{aligned} \quad (4.38)$$

Analogicznie można pokazać, że:

$$K_{\nabla H_i H}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = \frac{\partial K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\partial z_i^{(1)}}, \quad i = 1, \dots, n. \quad (4.39)$$

W przypadku wektorowych pól losowych możemy rozpatrywać tensory korelacyjne. Składowe tzw. *podstawowego tensora korelacyjnego* rozpatrywanego procesu gradientowego - $\mathbf{K}_{\nabla H}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})$ są określone przez wyrażenia:

$$\begin{aligned} K_{\nabla H_{ij}}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \text{Cov} \left[\nabla H_i(\mathbf{z}^{(1)}), \nabla H_j(\mathbf{z}^{(2)}) \right] = \mathbb{E} \left[\nabla \tilde{H}_i(\mathbf{z}^{(1)}) \nabla \tilde{H}_j(\mathbf{z}^{(2)}) \right] \\ & \quad i = 1, \dots, n, \quad j = 1, \dots, n. \end{aligned} \quad (4.40)$$

Stosując podobne przekształcenia jak powyżej, możemy uzyskać następujące zależności

między składowymi tensora korelacji ∇H a funkcją kowariancji pola H :

$$\begin{aligned}
K_{\nabla H_{ij}}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \mathbb{E} \left[\nabla \tilde{H}_i(\mathbf{z}^{(1)}) \nabla \tilde{H}_j(\mathbf{z}^{(2)}) \right] \\
&= \lim_{\Delta z_i \rightarrow 0} \mathbb{E} \left[\frac{\tilde{H}(\mathbf{z}^{(1)} + \Delta z_i \mathbf{e}_i) - \tilde{H}(\mathbf{z}^{(1)})}{\Delta z_i} \nabla \tilde{H}_j(\mathbf{z}^{(2)}) \right] \\
&= \lim_{\Delta z_i \rightarrow 0} \frac{\mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)} + \Delta z_i \mathbf{e}_i) \nabla \tilde{H}_j(\mathbf{z}^{(2)}) \right] - \mathbb{E} \left[\tilde{H}(\mathbf{z}^{(1)}) \nabla \tilde{H}_j(\mathbf{z}^{(2)}) \right]}{\Delta z_i} \\
&= \lim_{\Delta z_i \rightarrow 0} \frac{K_{H \nabla H_j}(\mathbf{z}^{(1)} + \Delta z_i \mathbf{e}_i, \mathbf{z}^{(2)}) - K_{H \nabla H_j}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\Delta z_i} \\
&= \frac{\partial^2 K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\partial z_i^{(1)} \partial z_j^{(2)}}, \quad i = 1, \dots, n, \quad j = 1, \dots, n. \quad (4.41)
\end{aligned}$$

Pole losowe $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, jest określane jako różniczkowalne na Ω , jeżeli jego realizacje są różniczkowalne na tym zbiorze z prawdopodobieństwem 1 [110]. Funkcje wielu zmiennych klasy C_1 , mające ciągle pochodne w obszarze Ω , są różniczkowalne. Tak więc różniczkowalność pola losowego możemy wyrazić, żądając ciągłości składowych jego gradientu. A zatem, realizacje pola losowego H są różniczkowalne z prawdopodobieństwem 1, jeżeli w przypadku każdej sekwencji \mathbf{z}_k takiej, że $\|\mathbf{z}_k - \mathbf{z}\| \rightarrow 0$, przy $k \rightarrow \infty$, zachodzi

$$P\{\gamma : |\nabla H_i(\mathbf{z}_k, \gamma) - \nabla H_i(\mathbf{z}, \gamma)| \rightarrow 0 \text{ przy } k \rightarrow \infty \forall \mathbf{z} \in \Omega\} = 1 \quad i = 1, \dots, n. \quad (4.42)$$

Konieczne warunki różniczkowalności realizacji pól losowych możemy uzyskać, żądając, aby pochodne cząstkowe miały ciągle realizacje. Pole losowe jest więc różniczkowalne, jeżeli składowe jego gradientu spełniają warunki ciągłości (4.29) lub (4.30). Warunek (4.32), określony za pomocą funkcji kowariancji składowych gradientu, przyjmuje postać

$$K_{\nabla H_{ii}}(\mathbf{z}, \mathbf{z}) + K_{\nabla H_{ii}}(\mathbf{z} + \boldsymbol{\zeta}', \mathbf{z} + \boldsymbol{\zeta}') - 2K_{\nabla H_{ii}}(\mathbf{z}, \mathbf{z} + \boldsymbol{\zeta}') \leq \frac{c'}{|\log \|\boldsymbol{\zeta}'\||^{1+\alpha'}}, \quad i = 1, \dots, n, \quad (4.43)$$

gdzie c' , α' , $\boldsymbol{\zeta}'$ spełniają założenia takie, jak w przypadku (4.32).

4.4. Dyskretyzacja pól losowych

Jak widzieliśmy w poprzednich podrozdziałach, pole losowe w ogólności jest rodziną zmiennych losowych o nieprzeliczalnej liczbie elementów. Taki model nie może być

bezpośrednio wykorzystany w analizie niezawodności, gdyż metody przedstawione w rozdziale 2 działają jedynie dla skończonej liczby zmiennych losowych. Tak więc oszacowanie prawdopodobieństwa awarii, w przypadku problemu określonego na podstawie pola losowego, wymaga jego przybliżenia za pomocą funkcji wielowymiarowej zmiennej losowej.

Dyskretyzacja \hat{H} skalarne pola losowego H jest aproksymacją skalarne pola losowego za pomocą funkcji $\mathcal{H} : \Omega \times \mathbf{X} \rightarrow \mathbf{R}$, gdzie $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ jest wektorem losowym, określonym na tej samej co aproksymowane pole losowe przestrzeni probabilistycznej (Γ, \mathcal{E}, P) :

$$H(\mathbf{z}) \approx \hat{H}(\mathbf{z}) = \mathcal{H}(\mathbf{z}, \mathbf{X}). \quad (4.44)$$

Efektywna dyskretyzacja pola losowego powinna, przy wykorzystaniu możliwie małej liczby zmiennych losowych, zapewniać minimalizację przyjętego kryterium błędu.

Błędem dyskretyzacji jest różnica $H(\mathbf{z}, \gamma) - \hat{H}(\mathbf{z}, \gamma)$, $\gamma \in \Gamma$. Oczywiście błąd ten powinien być jak najbliższy zeru. Tak więc uzasadnione jest dążenie do uzyskania dyskretyzacji, która jest nieobciążona, tzn. spełnia warunek:

$$\mathbb{E} [H(\mathbf{z}) - \hat{H}(\mathbf{z})] = 0. \quad (4.45)$$

Jeżeli spełniony jest powyższy warunek, to za miarę błędu można przyjąć jego wariancję. Często jako punktowe oszacowanie błędu dyskretyzacji stosuje się tzw. *względną wariancję błędu dyskretyzacji*:

$$err(\mathbf{z}) = \frac{\text{Var} [H(\mathbf{z}) - \hat{H}(\mathbf{z})]}{\text{Var} [H(\mathbf{z})]}. \quad (4.46)$$

Globalną miarą błędu jest natomiast *średni błąd kwadratowy*, będący całką wariancji błędu dyskretyzacji po Ω :

$$\overline{err} = \int_{\Omega} \text{Var} [H(\mathbf{z}) - \hat{H}(\mathbf{z})] d\Omega. \quad (4.47)$$

W ogólności dyskretyzację pola losowego można przedstawić w następujący sposób:

$$\mathcal{H}(\mathbf{z}, \mathbf{X}) = g_0(\mathbf{z}) + \sum_{i=1}^n g_i(\mathbf{z}) X_i, \quad (4.48)$$

gdzie g_i , $i = 0, 1, \dots, n$, są funkcjami $z \in \Omega$, a X_i , $i = 1, 2, \dots, n$, są zmiennymi losowymi, na których opiera się dyskretyzacja. Metody dyskretyzacji różnią się sposobem, w jaki określa się funkcje g i zmienne X - to zróżnicowanie jest podstawą ich klasyfikacji. W tej pracy proponujemy następujący podział:

- *Metody lokalnego uśredniania* (ang. *local averaging*) - wykorzystujące podział Ω na rozłączne podzbiory (elementy), w których pole losowe jest reprezentowane przez pojedynczą zmienną losową. Rozkład prawdopodobieństwa tych zmiennych losowych jest określany poprzez procedurę uśredniania pola losowego wewnątrz elementów.
- *Rozwinięcia w szereg* - wykorzystujące reprezentację pola losowego za pomocą obciętego rozwinięcia w szereg funkcji deterministycznych, ze współczynnikami losowymi.
- *Aproksymacja funkcjami kształtu* - wykorzystująca interpolację za pomocą funkcji kształtu i skończonego podzbioru wartości pola losowego.

Najpopularniejsze metody zostaną omówione poniżej. Przedstawione zostaną również przykłady dyskretyzacji pól losowych. W prezentowanych przykładach ograniczymy się jednak tylko do wykorzystania metody optymalnej aproksymacji liniowej, którą uznaliśmy za najbardziej efektywną - uwzględniając łatwość zastosowania i dokładności uzyskanych wyników.

4.4.1. Metody lokalnego uśredniania

Dzięki swojej prostocie metody lokalnego uśredniania są często wykorzystywane do dyskretyzacji pól losowych. Należy jednak zaznaczyć, że podejście będące ich podstawą, jest raczej heurystyczne.

Zastosowanie metod lokalnego uśredniania wymaga podziału obszaru Ω , na którym jest określone pole losowe, na rozłączne, przylegające podobszary (nazywane również elementami pola losowego) Ω_i^e , $i = 1, 2, \dots, n_e$. W przypadku, kiedy analiza stochastyczna jest prowadzona w połączeniu z analizą metodą elementów skończonych, warto wykorzystać podział Ω na elementy skończone. W ramach metody lokalnego uśredniania, pole losowe wewnątrz elementu przybliża się jedną zmienną losową, związaną z pewnym uśrednieniem pola wewnątrz elementu. Najczęściej wykorzystuje się dwa podejścia, tzw. *dyskretyzację punktową* i *uśrednianie po elemencie*.

Dyskretyzacja punktowa. Metoda ta, opiera się na reprezentacji pola losowego w każdym z elementów Ω_i^e , za pomocą wartości pola losowego w wybranym punkcie elementu. Jednym z wariantów tego podejścia jest *metoda punktów środkowych* (ang. *mid point method*) [23], wykorzystująca wartości pola losowego w środkach ciężkości elementów. Niech dany będzie wektor losowy

$$\mathbf{X}_{MP} = \left\{ H \left(\mathbf{z}^{*(1)} \right), H \left(\mathbf{z}^{*(2)} \right), \dots, H \left(\mathbf{z}^{*(n_e)} \right) \right\}, \quad (4.49)$$

gdzie $H(\mathbf{z}^{*(i)})$, $i = 1, 2, \dots, n_e$, są wartościami $H(\mathbf{z})$ w środkach ciężkości elementów. Przyjmuje się, że własności \mathbf{X}_{MP} wynikają z własności pola losowego $H(\mathbf{z})$, na przykład:

$$\text{Cov}[X_m, X_n] = K_H(\mathbf{z}^{*(m)}, \mathbf{z}^{*(n)}), \quad m, n \in \{1, 2, \dots, n_e\}. \quad (4.50)$$

Aproksymację metodą punktów środkowych pola losowego $H(\mathbf{z})$ możemy więc określić w następujący sposób:

$$\hat{H}_{MP}(\mathbf{z}) = \sum_{i=1}^{n_e} I_{\Omega_i^e}(\mathbf{z}) H(\mathbf{z}^{*(i)}), \quad (4.51)$$

gdzie $I_{\Omega_i^e}$ $i = 1, \dots, n_e$, są funkcjami charakterystycznymi elementów pola losowego. Warto zwrócić uwagę, że realizacje $\hat{H}_{MP}(\mathbf{z})$ są stałe wewnątrz elementów i mają nieciągłości na ich brzegach. Można również pokazać [23], że dyskretyzacja za pomocą punktów środkowych ma większą nieregularność niż dyskretyzowane pole losowe.

Metoda średnich przestrzennych. Przykładem metod lokalnego uśredniania jest *metoda średnich przestrzennych* (ang. *spatial average method*) [61, 125]. Składowe wektora zmiennych losowych $\mathbf{X}^{SA} = \{X_1^{SA}, \dots, X_{n_e}^{SA}\}$, który jest wykorzystywany w dyskretyzacji, są średnimi pola losowego w elementach:

$$X_i^{SA} = \frac{\int_{\Omega_i^e} H(\mathbf{z}) d\Omega_i^e}{|\Omega_i^e|}, \quad i = 1, 2, \dots, n_e. \quad (4.52)$$

Dyskretyzacja metodą średnich przestrzennych jest określona tak, jak w metodzie punktów środkowych:

$$\hat{H}_{SA}(\mathbf{z}) = \sum_{i=1}^{n_e} I_{\Omega_i^e}(\mathbf{z}) X_i^{SA}. \quad (4.53)$$

Podjęcie to ma jednak ograniczone zastosowanie, gdyż poza przypadkiem gaussowskich pól losowych, prawie niemożliwe jest wyznaczenie gęstości prawdopodobieństwa wektora \mathbf{X}^{SA} . Wartość średnia i kowariancje składowych \mathbf{X}^{SA} są określone przez całki funkcji wartości średnich i kowariancji pola losowego w elementach dyskretyzacji.

Istotną cechą dyskretyzacji metodą średnich przestrzennych jest jej większa regularność, w porównaniu z aproksymowanym polem losowym. Tak, więc metoda punktów środkowych i metoda średnich przestrzennych stanowią górne i dolne ograniczenie losowej zmienności dyskretyzowanego pola.

4.4.2. Rozwinięcia w szereg

Niech $\{\psi_i(\mathbf{z})\}_{i=1}^{\infty}$ będzie zbiorem funkcji ortogonalnych, tworzących kompletną bazę w $\mathcal{L}^2(\Omega)$ - przestrzeni Hilberta funkcji całkowalnych z kwadratem na Ω . Dla uproszczenia można przyjąć założenie, że baza jest ortonormalna, tzn.:

$$\int_{\Omega} \psi_i(\mathbf{z})\psi_j(\mathbf{z})d\Omega = \delta_{ij}, \quad i = 1, \dots, \infty, j = 1, \dots, \infty. \quad (4.54)$$

Na przykład w pracy [131], do dyskretyzacji gaussowskich pól losowych, jako składowe bazy, wykorzystano wielomiany Legendre'a.

Niech $H(\mathbf{z}, \gamma)$, $\mathbf{z} \in \Omega$, $\gamma \in \Gamma$, będzie polem losowym w przypadku, którego istnieją funkcje wartości oczekiwanej $\mu_H(\mathbf{z})$, $\mathbf{z} \in \Omega$ i kowariancji $K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})$, $\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \in \Omega$. Każda realizacja rozważanego pola losowego $h(\mathbf{z})$, $\mathbf{z} \in \Omega$, jest funkcją $\mathcal{L}^2(\Omega)$, która może być rozwinięta w szereg oparty na funkcjach $\{\psi_i(\mathbf{z})\}_{i=1}^{\infty}$. Tak więc pole losowe H możemy przedstawić w postaci szeregu z losowymi współczynnikami:

$$H(\mathbf{z}, \gamma) = \mu(\mathbf{z}) + \sum_{i=1}^{\infty} X_i(\gamma)\psi_i(\mathbf{z}), \quad (4.55)$$

gdzie $X_i(\gamma)$, $i = 1, \dots, \infty$, $\gamma \in \Gamma$, są zmiennymi losowymi o zerowej wartości oczekiwanej.

Ponieważ funkcje ψ_i są ortogonalne, to zmienne losowe X_i mają postać

$$X_i(\gamma) = \int_{\Omega} (H(\mathbf{z}, \gamma) - \mu_h(\mathbf{z})) \psi_i(\mathbf{z})d\Omega, \quad i = 1, \dots, \infty, \quad (4.56)$$

a składowe ich macierzy kowariancji są określone wyrażeniem:

$$\text{Cov}[X_i, X_j] = \int_{\Omega} \int_{\Omega} K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) \psi_i(\mathbf{z}^{(1)})\psi_j(\mathbf{z}^{(2)}) d\Omega d\Omega, \quad i, j = 1, \dots, \infty. \quad (4.57)$$

Jeżeli H jest gaussowskim polem losowym to z równania (4.56) wynika, że zmienne $\{X_i\}_{i=1}^{\infty}$ są zmiennymi normalnymi o zerowej wartości oczekiwanej.

Szczególnym przypadkiem dyskretyzacji za pomocą szeregu losowego jest, tak zwane rozwinięcie Karhunen-Loeve'a [45]. Metoda ta wykorzystuje funkcje ortogonalne postaci

$$\psi_i(\mathbf{z}) = \sqrt{\lambda_i}\varphi_i(\mathbf{z}) \quad i = 1, \dots, \infty, \quad (4.58)$$

gdzie λ_i i φ_i są odpowiednio wartościami własnymi i funkcjami własnymi następującego jednorodnego równania całkowego Fredholma drugiego rodzaju:

$$\int_{\Omega} K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) \varphi(\mathbf{z}^{(2)}) d\mathbf{z}^{(2)} = \lambda\varphi(\mathbf{z}^{(1)}) . \quad (4.59)$$

Ponieważ jądro $K_H(\cdot, \cdot)$ powyższego równania (będące funkcją kowariancji pola H) jest ograniczone, symetryczne i dodatnio określone, to zbiór funkcji $\{\varphi_i\}$ tworzy kompletną bazę ortogonalną w $\mathcal{L}^2(\Omega)$. Co więcej, wartości własne λ_i są rzeczywiste, dodatnie, a ich jedynym punktem akumulacji jest zero. Rozwinięcie pola losowego w szereg Karhunena-Loeve'a ma więc następującą postać:

$$H(\mathbf{z}, \gamma) = \mu_H(\mathbf{z}) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} Y_i(\gamma) \varphi_i(\mathbf{z}), \quad (4.60)$$

gdzie losowe współczynniki rozwinięcia Y_i , $i = 1, \dots, \infty$, $\gamma \in \Gamma$ są niezależnymi zmiennymi losowymi [45].

Rozwinięcie Karhunena-Loeve'a posiada kilka istotnych własności:

- Wartości własne λ_i można uporządkować w szereg, który jest malejący i zbieżny do zera. Obcinając tak uporządkowany szereg po m -tym wyrazie otrzymujemy dyskretyzację Karhunena-Loeve'a pola losowego $H(\mathbf{z}, \gamma)$:

$$\hat{H}_{KL}(\mathbf{z}) = \mu_H(\mathbf{z}) + \sum_{i=1}^m \sqrt{\lambda_i} Y_i \varphi_i(\mathbf{z}). \quad (4.61)$$

- Rozwinięcie oparte na bazie złożonej z funkcji własnych $\varphi_i(\mathbf{z})$ jest optymalne w sensie minimalizacji średniego błędu kwadratowego, powstałego w wyniku obcięcia szeregu po m -tym wyrazie (w porównaniu do obciętego szeregu opartego na każdej innej bazie):

$$\{\varphi(\mathbf{z}), \lambda\} = \arg \min_{\{\xi, \psi\}} \int_{\Omega} \left(\sum_{i=r+1}^{\infty} \xi_i Y_i \psi_i(\mathbf{z}) \right)^2 d\mathbf{z}. \quad (4.62)$$

- Dzięki ortonormalności funkcji własnych można otrzymać proste wzory, które określają zmienne losowe będące parametrami rozwinięcia:

$$Y_i(\gamma) = \frac{1}{\sqrt{\lambda_i}} \int_{\Omega} (H(\mathbf{z}, \gamma) - \mu_H(\mathbf{z})) \varphi_i(\mathbf{z}) d\Omega. \quad (4.63)$$

- Wariancja błędu, wynikającego z obcięcia szeregu po m wyrazach, jest określona następującym wzorem [116]:

$$\text{Var} [H(\mathbf{z}) - \hat{H}(\mathbf{z})] = \sigma_H^2(\mathbf{z}) - \sum_{i=1}^m \lambda_i \varphi_i^2(\mathbf{z}) = \text{Var} [H(\mathbf{z})] - \text{Var} [\hat{H}(\mathbf{z})]. \quad (4.64)$$

Powyższe wyrażenie jest zawsze dodatnie, ponieważ jest wariancją jakiejś wielkości. Tak więc, pole losowe ma większą nieregularność niż jego dyskretyzacja metodą Karhunena-Loeve'a.

Dyskretyzacja Karhunena-Loeve'a, dzięki swojej optymalności, spotkała się z dużym zainteresowaniem w literaturze. Praktyczne zastosowanie tej metody jest jednak ograniczone do problemów, w przypadku, których można efektywnie rozwiązać równanie (4.59). Rozwiązania analityczne tego problemu istnieją dla kilku typów funkcji korelacji, określonych na obszarach Ω o prostej geometrii. Analityczne rozwiązania problemu (4.59), dla wykładniczych i trójkątnych funkcji korelacji zdefiniowanych w ograniczonych przedziałach, symetrycznych wokół zera, można znaleźć w pracy [45]. W ogólnym przypadku, do rozwiązania równania (4.59) można zastosować metody numeryczne. Również w [45] zaproponowano służącą temu celowi procedurę, opartą na metodzie Galerkina.

4.4.3. Metoda funkcji kształtu

W metodzie funkcji kształtu (ang. *shape function method*) (SF) stosowana jest podobna aproksymacja jak w metodzie elementów skończonych [69, 70]. Obszar, w którym jest określone pole losowe podlega podziałowi na elementy tak jak w przypadku metody lokalnego uśredniania. Aproksymacja w każdym z elementów jest określona za pomocą następującego wyrażenia:

$$\hat{H}_{SF}(\mathbf{z}) = \sum_{i=1}^q N_i(\mathbf{z})H(\mathbf{z}^{(i)}), \quad \mathbf{z} \in \Omega_k^e, k = 1, 2, \dots, n_e, \quad (4.65)$$

gdzie q jest liczbą węzłów w elemencie Ω_k^e , a N_i , $i = 1, \dots, q$, są wielomianowymi funkcjami kształtu, związanymi z węzłami elementów w punktach $\mathbf{z}^{(i)}$. Funkcje wartości oczekiwanej i kowariancji aproksymacji metodą SF są postaci:

$$\mu_{\hat{H}_{SF}}(\mathbf{z}) = \sum_{i=1}^q N_i(\mathbf{z})\mu_H(\mathbf{z}^{(i)}), \quad (4.66)$$

$$K_{\hat{H}_{SF}}(\mathbf{z}^{(01)}, \mathbf{z}^{(02)}) = \sum_{i=1}^q \sum_{j=1}^p N_i(\mathbf{z}^{(01)}) N_j(\mathbf{z}^{(02)}) K_H(\mathbf{z}^{(01)}, \mathbf{z}^{(02)}), \quad (4.67)$$

gdzie q jest liczbą węzłów elementu, zawierającego punkt $\mathbf{z}^{(01)}$ a p jest liczbą węzłów elementu zawierającego punkt $\mathbf{z}^{(02)}$. Warto zauważyć, że realizacje pola losowego \hat{H}_{SF} są funkcjami ciągłymi na obszarze Ω , co niewątpliwie jest zaletą tej metody w porównaniu do metod dyskretyzacji punktowej.

4.4.4. Metoda optymalnej aproksymacji liniowej

Metoda optymalnej aproksymacji liniowej (ang. *Optimal Linear Estimation Method*) (OLE) została zaproponowana przez Li i Der Kiureghiana w pracy [66]. W tym algorytmie wykorzystuje się wektor wartości pola losowego H w punktach węzłowych $\mathbf{z}^{(i)}$, $i = 1, \dots, q$:

$$\mathbf{X} = \left\{ H(\mathbf{z}^{(1)}), H(\mathbf{z}^{(2)}), \dots, H(\mathbf{z}^{(q)}) \right\}. \quad (4.68)$$

Dyskretyzacja metodą OLE jest zdefiniowana następującym wzorem:

$$\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) = a(\mathbf{z}) + \sum_{i=1}^q b_i(\mathbf{z}) X_i, \quad (4.69)$$

gdzie funkcje $a(\mathbf{z})$ i $b_i(\mathbf{z})$ są określone w każdym punkcie $\mathbf{z} \in \Omega$, poprzez następujący problem optymalizacji:

$$\text{minimalizuj} \quad \text{Var}[H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}, \mathbf{X})], \quad (4.70)$$

$$\text{przy ograniczeniu} \quad \mathbb{E}[H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}, \mathbf{X})] = 0. \quad (4.71)$$

Zauważmy, że formułując powyższe zadanie optymalizacji posłużono się warunkami omówionymi na początku tego podrozdziału, które powinny być spełnione przez efektywną dyskretyzację pola losowego. Minimalizacja funkcji celu jest, bowiem równoważna minimalizacji punktowego błędu dyskretyzacji (4.46) a spełnienie ograniczenia zapewnia, że dyskretyzacja jest nieobciążona (por. (4.45)).

Funkcję $a(\mathbf{z})$ możemy określić podstawiając (4.69) do warunku (4.71):

$$a(\mathbf{z}) = \mu_H(\mathbf{z}) - \sum_{i=1}^q b_i(\mathbf{z}) \mathbb{E}[X_i]. \quad (4.72)$$

Poprzez podstawienie powyższego wyrażenia do (4.69), otrzymujemy

$$\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) = \mu_H(\mathbf{z}) + \sum_{i=1}^q b_i(\mathbf{z}) (X_i - \mathbb{E}[X_i]). \quad (4.73)$$

Za pomocą (4.73), funkcję celu (4.70) można przekształcić w następujący sposób:

$$\begin{aligned} \text{Var} \left[H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}) \right] &= \mathbb{E} \left[\left(H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}) \right)^2 \right] = \\ &= \sigma_H^2(\mathbf{z}) - 2 \sum_{i=1}^q b_i(\mathbf{z}) \text{Cov} [H(\mathbf{z}), X_i] + \sum_{i=1}^q \sum_{j=1}^q b_i(\mathbf{z}) b_j(\mathbf{z}) \text{Cov} [X_i, X_j]. \end{aligned} \quad (4.74)$$

Funkcje $b_i(\mathbf{z})$ powinny zapewniać minimalizację funkcji celu (4.74) w każdym punkcie obszaru Ω . Warunek konieczny optymalności, w postaci zerowej pochodnej (4.74) względem $b_i(\mathbf{z})$, jest następującej postaci:

$$\forall i = 1, 2, \dots, q \quad -\text{Cov}[H(\mathbf{z}), X_i] + \sum_{j=1}^q b_j(\mathbf{z})\text{Cov}[X_i, X_j] = 0. \quad (4.75)$$

Ostatecznie wartości b_j w punkcie \mathbf{z} można wyznaczyć rozwiązując następujący układ równań liniowych otrzymany z (4.75):

$$\mathbf{b}(\mathbf{z}) = \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z}), \quad (4.76)$$

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 wektora \mathbf{X} , określonego wzorem (4.68), a $\mathbf{C}_{HX}(\mathbf{z}) = \{\text{Cov}[H(\mathbf{z}), X_1], \text{Cov}[H(\mathbf{z}), X_2], \dots, \text{Cov}[H(\mathbf{z}), X_q]\}$. Podstawiając powyższe wyniki do wyrażenia (4.73), otrzymujemy macierzową postać dyskretyzacji metodą OLE:

$$\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) = \mu_H(\mathbf{z}) + [\mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z})]^T (\mathbf{X} - \mu_X), \quad (4.77)$$

gdzie $\mu_X = \{\mathbb{E}[X_1], \mathbb{E}[X_2], \dots, \mathbb{E}[X_q]\}$. Ponieważ (4.77) spełnia warunek (4.71), to wariancję dyskretyzacji metodą OLE można wyrazić w następujący sposób:

$$\begin{aligned} \text{Var} \left[\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) \right] &= \mathbb{E} \left[\left(\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) - \mathbb{E} \left[\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) \right] \right)^2 \right] \\ &= \mathbb{E} \left[\left(\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) - \mu_H(\mathbf{z}) \right)^2 \right] \\ &= \mathbb{E} \left[\mathbf{C}_{HX}^T(\mathbf{z}) \mathbf{C}_{\mathbf{X}}^{-1} (\mathbf{X} - \mu_X) (\mathbf{X} - \mu_X)^T \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z}) \right] \\ &= \mathbf{C}_{HX}^T(\mathbf{z}) \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z}). \end{aligned} \quad (4.78)$$

Wykorzystując (4.76) wariancję błędu, określoną przez (4.74), przedstawia się w notacji macierzowej, jako

$$\begin{aligned} \text{Var} \left[H(\mathbf{z}) - \hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) \right] &= \sigma_H^2(\mathbf{z}) - 2\mathbf{C}_{HX}^T(\mathbf{z}) \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX} + \mathbf{C}_{HX}^T(\mathbf{z}) \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}(\mathbf{z}) \\ &= \sigma_H^2(\mathbf{z}) - \mathbf{C}_{HX}^T(\mathbf{z}) \mathbf{C}_{\mathbf{X}}^{-1} \mathbf{C}_{HX}. \end{aligned} \quad (4.79)$$

Druga część wyrażenia (4.79) jest równa wariancji dyskretyzacji metodą OLE określonej przez (4.78). Ponieważ wariancja jest zawsze nieujemna to na podstawie (4.79) otrzymujemy:

$$\text{Var} \left[\hat{H}_{OLE}(\mathbf{z}, \mathbf{X}) \right] \leq \sigma_H^2(\mathbf{z}); \quad (4.80)$$

a więc dyskretyzacja metodą OLE jest bardziej regularna niż aproksymowane pole losowe.

W przypadku gaussowskich pól losowych, metoda OLE może być zmodyfikowana tak, aby dyskretyzacja była określona za pomocą standardowych zmiennych normalnych; podobnie jak ma to miejsce w rozwinięciu Karhunen-Loeve'a. Metoda ta oznaczana skrótem EOLE (ang. *expansion optimal linear estimation*) [66], wykorzystuje łatwość transformacji zależnych zmiennych gaussowskich na standardowe zmienne normalne. Wektor gaussowskich zmiennych losowych \mathbf{X} jest związany z wektorem standardowych zmiennych normalnych \mathbf{U} poprzez następującą transformację:

$$\mathbf{X} = \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Phi}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{U}, \quad (4.81)$$

gdzie $\boldsymbol{\Phi}$ jest macierzą, której kolumny są wektorami własnymi a $\boldsymbol{\Lambda}$ jest macierzą diagonalną wartości własnych macierzy kowariancji $\mathbf{C}_{\mathbf{X}}$ wektora \mathbf{X} . Tak, więc spełniona jest poniższa równość

$$\mathbf{C}_{\mathbf{X}}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Lambda}. \quad (4.82)$$

Jeżeli dyskretyzowane pole losowe jest gaussowskie to wektor (4.68) jest wielowymiarową zmienną normalną. Korzystając z transformacji (4.81), wektor (4.68) może być przedstawiony za pomocą standardowego wektora normalnego w następujący sposób:

$$\mathbf{X} = \boldsymbol{\mu}_{\mathbf{X}} + \sum_{i=1}^n \sqrt{\lambda_i} \boldsymbol{\phi}_i U_i, \quad (4.83)$$

gdzie $\boldsymbol{\phi}_i$, λ_i są odpowiednimi wektorami i wartościami własnymi. Rozwiązując problem definiujący metodę OLE (4.70)-(4.71) dla wektora (4.83), otrzymujemy wyrażenie określające dyskretyzację metodą EOLE:

$$\hat{H}_{EOLE}(\mathbf{z}) = \mu_H(\mathbf{z}) + \sum_{i=1}^n \frac{U_i}{\sqrt{\lambda_i}} \boldsymbol{\phi}_i^T \mathbf{C}_{HX}(\mathbf{z}). \quad (4.84)$$

Jeżeli wektory własne są uporządkowane zgodnie z malejącymi wartościami własnymi to sumowanie w wyrażeniu (4.84) może być skrócone do $r < n$ wyrazów.

Wariancja błędu metody EOLE jest następującej postaci:

$$\text{Var} \left[H(\mathbf{z}) - \hat{H}(\mathbf{z}) \right] = \sigma_H^2(\mathbf{z}) - \sum_{i=1}^r \frac{1}{\lambda_i} \left(\boldsymbol{\phi}_i^T \mathbf{C}_{HX} \right)^2. \quad (4.85)$$

Ponieważ drugie wyrażenie w powyższym wzorze jest równe wariancji rozwinięcia EOLE to widzimy, że dyskretyzacja otrzymana za pomocą tej metody ma mniejszą nieregularność niż odpowiadające jej pole losowe.

4.4.5. Dyskretyzacja niegaussowskich pól losowych

Wartości funkcji gęstości prawdopodobieństwa w obszarach awarii są zazwyczaj bardzo małe. Mówimy również, że obszary awarii są położone w tzw. ogonach rozkładu prawdopodobieństwa. Tak, więc metody dyskretyzacji wykorzystywane w analizie niezawodności, powinny zapewniać dokładne przybliżenie nie tylko pierwszych bądź drugich momentów, ale również ogonów rozkładów prawdopodobieństwa wartości aproksymowanych pól losowych. Zaprezentowane powyżej metody spełniają ten warunek, jedynie w przypadku gaussowskich pól losowych. Omawiane metody dyskretyzacji mogą, więc być stosowane dla niegaussowskich pól losowych H_{NG} , jeżeli istnieje następująca nieliniowa transformacja:

$$H_{NG}(\mathbf{z}) = \mathbf{T}[H(\mathbf{z})], \quad (4.86)$$

gdzie H jest polem normalnym. W takim przypadku H_{NG} można przybliżyć poprzez transformację dyskretyzacji pola gaussowskiego

$$\hat{H}_{NG}(\mathbf{z}) = \mathbf{T}[\hat{H}(\mathbf{z})]. \quad (4.87)$$

Stosowanie powyższej metody wiąże się jednak z pewnymi trudnościami. Na przykład w przypadku transformacji dyskretyzacji metodą OLE otrzymane przybliżenie może utracić własność minimalizacji błędu dyskretyzacji.

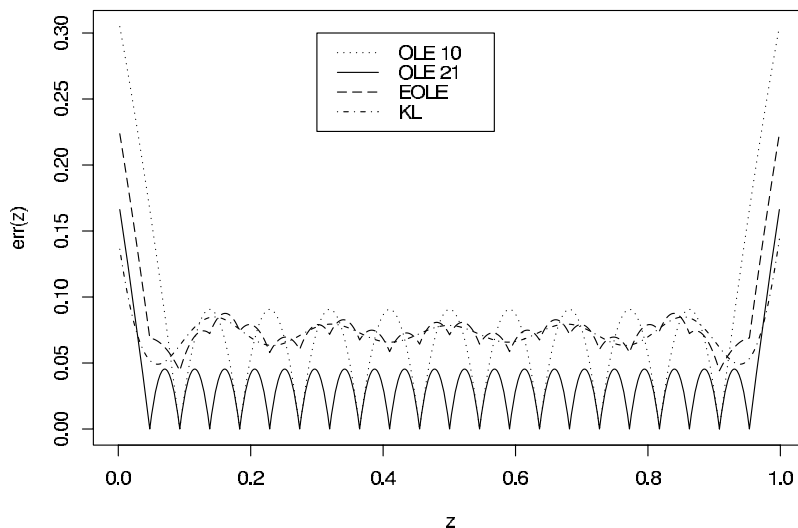
Omawiana metoda znajduje istotne zastosowanie praktyczne gdyż może być efektywnie wykorzystywana do dyskretyzacji logarytmiczno-normalnych pól losowych. Ponieważ realizacje takich pól losowych przyjmują jedynie wartości dodatnie to są one często stosowane do reprezentacji własności materiałowych.

4.4.6. Przykłady dyskretyzacji pól losowych

Przykład 4.2: Omówienie problemów, związanych z modelowaniem losowo i przestrzennie zmiennych parametrów konstrukcyjnych, zaczniemy od elementarnego przykładu, wykorzystującego jednowymiarowe stacjonarne pole losowe.

Stacjonarne pola losowe są zazwyczaj stosowane do modelowania wartości parametrów materiałowych, charakterystyk geometrycznych przekrojów czy parametrów geotechnicznych gruntu. W przypadku, kiedy dokonano pomiarów tych parametrów w wybranych punktach konstrukcji można uwzględnić zaobserwowane wartości poprzez wykorzystanie warunkowego pola losowego [26].

Rozważmy, więc stacjonarne gaussowskie pole losowe ze współrzędną $z \in [0, l]$. Może to być, na przykład, model modułu sprężystości uwzględniający jego zmiany wzdłuż osi belki. Jako pierwszy rozpatrzmy przypadek, kiedy funkcja korelacji jest wykładnicza (4.11) o długości korelacji równej połowie długości przedziału, na którym jest określone

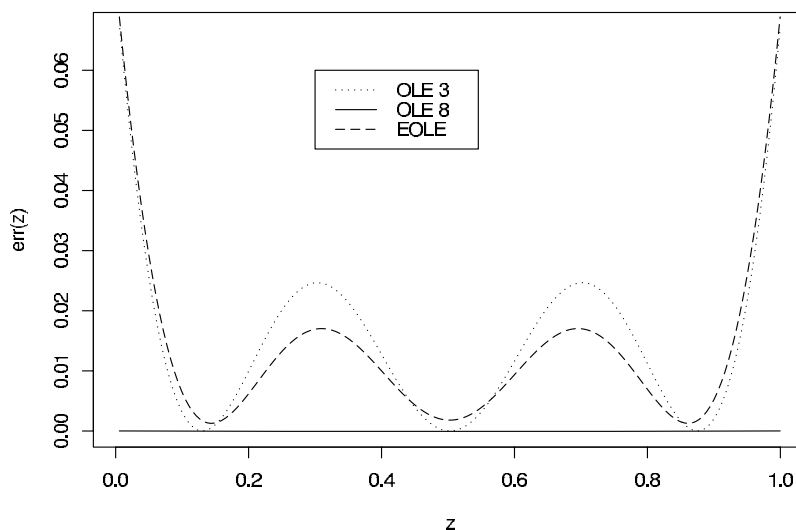


RYСУNEK 4.2. Porównanie punktowej względnej wariancji błędu dyskretyzacji dla metod: OLE 10 - OLE z 10 zmiennymi losowymi, OLE 21 - OLE z 21 zmiennymi losowymi, EOLE - EOLE wykorzystująca 6 z 21 zmiennych losowych, KL - rozwinięcie Karhunenena-Loeve'a rzędu 6. Wykładnicza funkcja korelacji.

pole losowe. W przypadku takiej funkcji korelacji, określonej na przedziale w zbiorze liczb rzeczywistych, istnieje analityczne rozwiązanie problemu (4.59) a więc możliwe jest łatwe zastosowanie dyskretyzacji Karhunenena-Loeve'a (4.60). Dyskretyzację metodą OLE przeprowadzono wykorzystując $q=10$ (OLE 10) i $q=21$ (OLE 21) wartości pola losowego w punktach $z_i = i/(q+1)$, $i = 1, \dots, q$. Na podstawie aproksymacji OLE z 21 zmiennymi losowymi, zgodnie z (4.84), uzyskano dyskretyzację EOLE, w której zastosowano $r=6$ zmiennych losowych. W metodzie Karhunenena-Loeve'a (KL) wykorzystano 6 wyrazów szeregu (4.60). W przypadku rozpatrywanego problemu wzory, które określają funkcje i wartości własne rozwinięcia Karhunenena-Loeve'a można znaleźć w [45]

Rysunek 4.2 przedstawia wykresy błędności dyskretyzacji (4.46) omawianych metod. Jak można zauważyć, porównywalne wartości błędności $err(z)$ wewnątrz obszaru dyskretyzacji zostały uzyskane w przypadku rozwinięcia Karhunenena-Loeve'a, metody OLE dla 10 zmiennych losowych oraz EOLE. Wartości $\int_{\Omega} err(z) dz$ dla rozważanych dyskretyzacji wynoszą: KL - 0.0725, OLE 10 - 0.0789, EOLE - 0.0787.

Podsumowując otrzymane wyniki możemy stwierdzić, że w przypadku rozpatrywanej funkcji korelacji bardzo efektywna jest dyskretyzacja metodą EOLE. Przy wykorzystaniu takiej samej liczby zmiennych losowych daje wyniki zbliżone do, uważanej za optymalną,

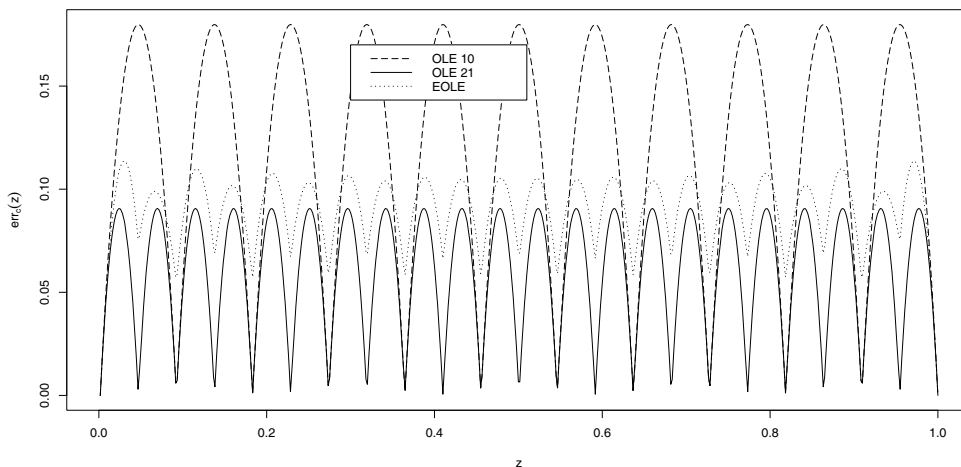


RYSUNEK 4.3. Porównanie punktowej względnej wariancji błędu dyskretyzacji dla metod: OLE 3 - OLE z 3 zmiennymi losowymi, OLE 8 - OLE z 8 zmiennymi losowymi, EOLE - EOLE wykorzystująca 3 z 8 zmiennych losowych. Gaussowska funkcja korelacji.

dyskretyzacji Karhunen-Loeve'a. Zauważmy również, że metoda EOLE wydaje się być bardziej efektywna, jeżeli wykorzystuje się dokładną dyskretyzację pola losowego i relatywnie niski rząd aproksymacji (mała liczba wykorzystanych składników z sumy we wzorze (4.84)), niż stosując zgrubną dyskretyzację i wysoki rząd aproksymacji (duża liczba wykorzystanych składników z sumy we wzorze (4.84)) [66].

Teraz rozważmy ten sam problem dyskretyzacji stacjonarnego normalnego pola losowego, ale z gaussowską funkcją korelacji (4.12). Na rysunku 4.3 przedstawiono wykresy błędów dyskretyzacji w przypadku metody OLE z trzema zmiennymi losowymi (OLE 3), OLE z ośmioma zmiennymi losowymi (OLE 8) oraz EOLE z rozwinięciem skróconym z ośmiu do trzech zmiennych losowych. W tym przypadku nie przedstawiono dyskretyzacji metodą Karhunen-Loeve'a, gdyż dla gaussowskiej funkcji korelacji nie istnieje rozwiązanie analityczne równania (4.59). Jak widzimy dla tej samej liczby zmiennych losowych $err(z)$ jest mniejszy w przypadku metody EOLE niż OLE. Natomiast błąd dyskretyzacji metody OLE z ośmioma zmiennymi losowymi jest pomijalnie mały. Warto zauważyć, że dla tego samego rzędu dyskretyzacji błąd aproksymacji w przypadku gaussowskiej funkcji korelacji jest znacznie mniejszy w porównaniu do funkcji wykładniczej. Fakt ten jest konsekwencją nieróżniczkowalności wykładniczej funkcji korelacji [116].

W raporcie [116] można znaleźć wskazówki, pomagające uzyskać efektywną dyskre-



RYSUNEK 4.4. Porównanie punktowej względnej wariancji błędu dyskretyzacji dla metod: OLE 10 - OLE wykorzystującej 10 zmiennych losowych, OLE 21 - OLE wykorzystującej 21 zmiennych losowych, EOLE - EOLE wykorzystującej 10 z 21 zmiennych losowych. Stacjonarne pole losowe z wykładniczą funkcją korelacji pod warunkiem zerowych wartości realizacji na końcach przedziału.

tyzację pól losowych. Zgodnie z nimi długość elementu pola losowego - odległość między zmiennymi dyskretyzacji - powinna być dobrana w zależności od długości korelacji. W przypadku wykładniczej funkcji korelacji powinna wynosić między $a/10$ i $a/5$, a w przypadku innych funkcji korelacji między $a/4$ i $a/2$.

Warto zaznaczyć, że zmienne losowe wykorzystywane w dyskretyzacji nie mogą być mocno skorelowane. Zbyt duża korelacja powoduje, bowiem niestabilności numeryczne podczas rozwiązywania zagadnienia własnego bądź odwracania macierzy korelacji zmiennych losowych, na których opiera się dyskretyzacja [23]. Z problemem tym spotykamy się w przypadku pól losowych z gaussowską funkcją korelacji gdy do dyskretyzacji chcemy użyć zbyt dużej liczby zmiennych losowych.

Przykład 4.3: Pola losowe znajdują również zastosowanie w modelowaniu imperfekcji geometrycznych. W przypadku małych imperfekcji geometrycznych przyjmuje się założenie, że decydujący wpływ na zachowanie układu zaburzonego ma przemieszczenie $w(\mathbf{z})$, $\mathbf{z} \in \Omega$ - normalne do osi pręta lub powierzchni środkowej konstrukcji powierzchniowej. Wstępne pole ugięć może być uwzględnione przez modyfikację geometrii modelu

bądź uwzględnienie wstępnych przemieszczeń w równaniach kinematycznych i równowagi wewnętrznej ustroju [128].

Modelem wstępnego ugięcia konstrukcji jest jedno lub dwuwymiarowe, odpowiednio dla konstrukcji prętowych i powierzchniowych, skalarne pole losowe. W ogólnym przypadku należy przyjąć założenie, że imperfekcje geometryczne mają charakter niestacjonarnego pola losowego [7]. Taki model można zdefiniować wykorzystując stacjonarne pole losowe wstępnych ugięć, które poddaje się warunkowaniu zgodnie z przyjętymi warunkami brzegowymi [80] (podobnie jak w przykładzie 4.1). To podejście jest uzasadnione, jeżeli przyjmiemy, że imperfekcje są wprowadzane np. podczas produkcji, a ostateczny kształt konstrukcji jest określany przez warunki podparcia. Przemieszczenia w podporach można uznać za deterministyczne, bądź bardziej realistycznie przyjmując, że mają charakter pól lub zmiennych losowych. Warto również zwrócić uwagę na to, że wstępne ugięcie powinno być funkcją ciągłą lub - jeżeli przyjęto takie założenie - różniczkowalną. Aby realizacje pola losowego imperfekcji spełniały to założenie jego funkcja kowariancji musi spełniać warunki ciągłości bądź różniczkowalności, które przedstawiono w podrozdziale 4.3.

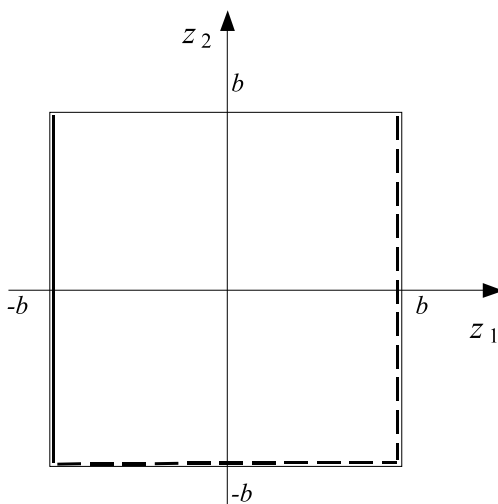
Rozważmy belkę swobodnie podpartą w punktach $z = 0$ i $z = 1$. Budowę modelu wstępnych ugięć zaczniemy od wprowadzenia stacjonarnego, gaussowskiego pola losowego $H(z)$, $z \in [0, 1]$, o zerowej wartości oczekiwanej i długości autokorelacji $a = 1/4$.

W celu uwzględnienia warunków podparcia pręta określimy pole losowe $G(z)$, $z \in [0, 1]$, którego wartości mają warunkowy rozkład prawdopodobieństwa wartości pola H pod warunkiem $H(0) = 0$ i $H(z = 1) = 0$. Określenie pola $G(z)$ wymaga wyznaczenia warunkowej funkcji wartości oczekiwanej $\mu_G(z) = \mathbb{E}[H(z)|H(0) = 0, H(1) = 0]$ i warunkowej funkcji kowariancji $K_H(z^{(1)}, z^{(2)}) = \text{Cov}[H(z^{(1)}), H(z^{(2)})|H(0) = 0, H(1) = 0]$. Ponieważ $h(0) = \mu_H(0)$ i $h(1) = \mu_H(1)$ to na podstawie (4.26) $\mu_G(z) = 0$. Korzystając z (4.27) otrzymujemy warunkową funkcję kowariancji postaci:

$$\begin{aligned} K_G(z^{(1)}, z^{(2)}) &= \sigma_H^2 \rho_H(z^{(1)}, z^{(2)}) \\ &\quad - \frac{\sigma_H^2}{1 - \rho_H(0, 1)^2} \left(\rho_H(z^{(1)}, 0) \rho_H(z^{(2)}, 0) \right. \\ &\quad \quad + \rho_H(z^{(1)}, 1) \rho_H(z^{(2)}, 1) \\ &\quad \quad - \rho_H(0, 1) \left(\rho_H(z^{(1)}, 1) \rho_H(z^{(2)}, 0) \right. \\ &\quad \quad \quad \left. \left. + \rho_H(z^{(1)}, 0) \rho_H(z^{(2)}, 1) \right) \right), \end{aligned} \quad (4.88)$$

gdzie σ_H^2 jest funkcją wariancji, a ρ_H jest funkcją korelacji pola losowego H .

W przypadku wykładniczej funkcji korelacji (4.11), dla przyjętej długości korelacji, wy-



RYSUNEK 4.5. Schemat podparcia płyty.

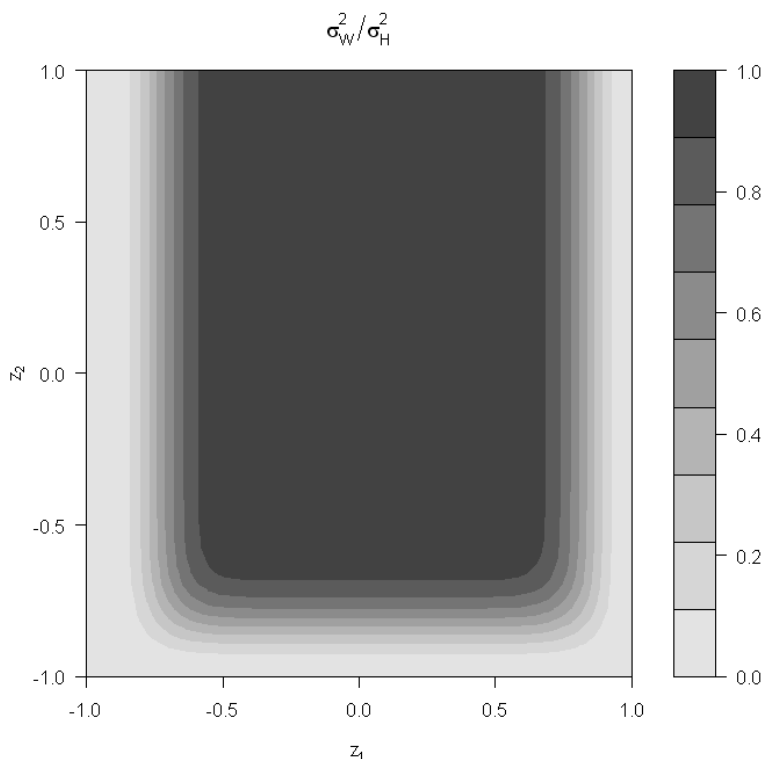
rażenie $(\sigma_G^2(z) + \sigma_G^2(z + \eta) - 2K_G(z, z + \eta)) |\log \|\eta\||^2$ jest zbieżne dla $\eta \rightarrow 0$ i dla $z \in [0, 1]$. Funkcja $K_G(z, z)$ spełnia więc warunek (4.32) zapewniający ciągłość realizacji pola losowego.

Na rysunku 4.4 przedstawiono wykresy błędu dyskretyzacji w przypadku gdy ρ_H jest wykładniczą funkcją korelacji. Rozpatrywana miara błędu dyskretyzacji jest określona wyrażeniem:

$$err_c(z) = \frac{\text{Var} [G(z) - \hat{G}(z)]}{\text{Var} [H(z)]}, \quad (4.89)$$

gdzie $\hat{G}(z)$ jest dyskretyzacją pola losowego $G(z)$. Wykresy przedstawiają wyniki uzyskane za pomocą metody OLE wykorzystującej 10 i 21 zmiennych losowych oraz za pomocą metody EOLE wykorzystującej 10 z 21 zmiennych losowych. Podobnie jak w przypadku jednorodnych pól losowych możemy zaobserwować wysoką efektywność metody EOLE.

Przykład 4.4: W tym przykładzie przedstawimy zagadnienie modelowania imperfekcji geometrycznych konstrukcji dwuwymiarowych. Rozważmy płytę określoną współrzędną $\mathbf{z} \in \Omega = [-b, b] \times [-b, b]$. Niech wzdłuż brzegu $\mathcal{B}_1 = \{\mathbf{z} : z_1 = -b, z_2 \in [-b, b]\}$ płyta będzie utwierdzona. Tak, więc wstępne ugięcie powierzchni środkowej spełnia następujące



RYSUNEK 4.6. Wariancja warunkowego pola losowego.

warunki brzegowe: $w(\mathbf{z}) = 0$ oraz $\partial w(\mathbf{z}) / \partial z_1 = 0$ dla $\mathbf{z} \in \mathcal{B}_1$. Ponadto przyjęto, że wzdłuż brzegów $\mathcal{B}_2 = \{\mathbf{z} : z_1 \in [-b, b], z_2 = -b\}$ oraz $\mathcal{B}_3 = \{\mathbf{z} : z_1 = b, z_2 \in [-b, b]\}$ płyta jest swobodnie podparta, a więc $w(\mathbf{z}) = 0$, dla $\mathbf{z} \in \mathcal{B}_2 \cup \mathcal{B}_3$ (rys. 4.5).

Budowę modelu imperfekcji geometrycznych zaczniemy od przyjęcia stacjonarnego, normalnego pola losowego $H(\mathbf{z})$, $\mathbf{z} \in \Omega$, z gaussovską funkcją korelacji, zerową funkcją wartości oczekiwanej, odchyleniem standardowym $\sigma_H = 0.05b$, oraz długością korelacji $a = 0.3b$. Pole imperfekcji $W(\mathbf{z})$, $\mathbf{z} \in \Omega$, którego realizacje spełniają warunki podparcia, określimy warunkując pole H zgodnie z wartościami ugięcia i jego pochodnych na brzegach. W tym celu, na podstawie wzorów (4.26) i (4.27), zdefiniujemy warunkową funkcję wartości oczekiwanej i warunkową funkcję kowariancji. Warunki podparcia postaci $w(z) = 0$

uwzględniono w 20 punktach na każdym z brzegów $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$. Podobnie w 20 równomiernie rozmieszczonych punktach na brzegu \mathcal{B}_1 uwzględniono warunek $\partial w(z)/\partial z_1 = 0$. Ze względu na przyjęte wartości przemieszczeń na podporach $\mu_W(\mathbf{z}) = \mu_H(\mathbf{z}) = 0$.

Oczywiście, powyższe podejście pozwala otrzymać jedynie przybliżenie funkcji wartości oczekiwanej i kowariancji pod warunkiem $w(\mathbf{z}) = 0$, $\mathbf{z} \in \mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3$ i $\partial w(\mathbf{z})/\partial z_1 = 0$, $\mathbf{z} \in \mathcal{B}_1$. Wzory (4.26) i (4.27) pozwalają, bowiem określić funkcje wartości oczekiwanej i kowariancji pod warunkiem przyjęcia przez realizację wartości w skończonej liczbie punktów, natomiast zbiory $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ są nieprzeliczalne.

Warto zauważyć, że określenie dokładnej funkcji warunkowej kowariancji wymagałoby uogólnienia pojęcia kowariancji. Imperfekcje na brzegach są, bowiem realizacjami jednowymiarowego pola losowego. Jak wspomniano, pole losowe można postrzegać, jako uogólnienie pojęcia zmiennej losowej będące funkcją przekształcającą przestrzeń zdarzeń elementarnych w przestrzeń funkcyjną. Tak, więc aby określić kowariancję wartości pola losowego wewnątrz płyty i jego realizacji na brzegach należałoby zdefiniować kowariancję pola losowego, jako uogólnionej zmiennej losowej i klasycznej zmiennej losowej.

W obliczeniach wykonanych na potrzeby tego przykładu wartości funkcji kowariancji pola W wyznaczano numerycznie na podstawie (4.27). Poniżej przedstawimy wyrażenia, które w rozpatrywanym zadaniu występują we wzorze (4.27). Funkcja kowariancji pochodnych pola losowego H i jego wartości, dla rozważanej gaussowskiej funkcji kowariancji, ma postać:

$$\begin{aligned} K_{\nabla H_1 H}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \frac{\partial \sigma_H^2 \exp\left(-\left(\frac{\|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|}{a}\right)^2\right)}{\partial z_1^{(1)}} \\ &= -\frac{2\sigma_H^2 (z_1^{(1)} - z_1^{(2)})}{a^2} \exp\left(-\left(\frac{\|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|}{a}\right)^2\right), \end{aligned} \quad (4.90)$$

$$\begin{aligned} K_{H \nabla H_1}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \frac{\partial \sigma_H^2 \exp\left(-\left(\frac{\|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|}{a}\right)^2\right)}{\partial z_1^{(2)}} \\ &= \frac{2\sigma_H^2 (z_1^{(1)} - z_1^{(2)})}{a^2} \exp\left(-\left(\frac{\|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|}{a}\right)^2\right). \end{aligned} \quad (4.91)$$

Funkcję kowariancji między pierwszymi składowymi ∇H możemy przedstawić jako

$$\begin{aligned} K_{\nabla H_1 \nabla H_1}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) &= \frac{\partial^2 K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})}{\partial z_1^{(1)} \partial z_1^{(2)}} \\ &= \frac{2\sigma_H^2}{a^2} \left(1 - \frac{2(z_1^{(1)} - z_2^{(2)})^2}{a^2} \right) \exp \left(- \left(\frac{\|\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\|}{a} \right)^2 \right). \end{aligned} \quad (4.92)$$

Zwróćmy uwagę, że dzięki wykorzystaniu wielokrotnie różniczkowalnej funkcji kowariancji K_H funkcja kowariancji warunkowego pola losowego K_W spełniała warunki różniczkowalności realizacji omówione w punkcie 4.3.

Rysunek 4.6 przedstawia wykres funkcji $\sigma_W^2(\mathbf{z})/\sigma_H^2(\mathbf{z})$, gdzie σ_W^2 jest wariancją warunkowego pola losowego W . Jak widzimy obie wariancje są równe w środku płyty. W podporach σ_W^2 jest zerowa, a w porównaniu z brzegiem utwierdzonym, rośnie szybciej wzdłuż brzegów swobodnie podpartych.

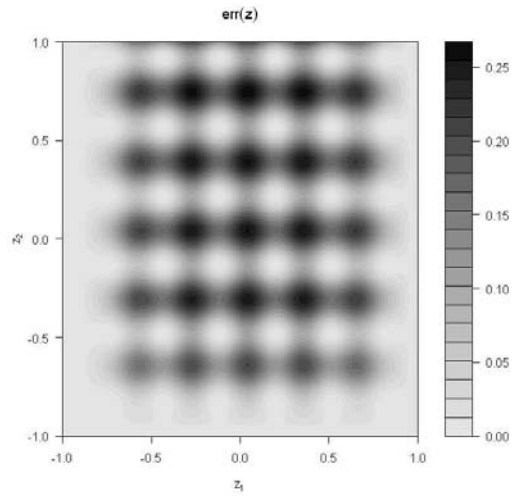
Wykres błędu dyskretyzacji metodą OLE z 36 zmiennymi losowymi zamieszczono na rysunku 4.7. Rysunek 4.8 przedstawia błąd dyskretyzacji metodą EOLE wykorzystującą 36 ze 100 zmiennych losowych. Jak widzimy, również w zadaniu dwuwymiarowym zastosowanie metody EOLE pozwala zmniejszyć maksymalny błąd dyskretyzacji przy wykorzystaniu tej samej liczby zmiennych losowych.

Z wyrażenia określającego dyskretyzację EOLE (4.84) możemy wyodrębnić *funkcje kształtu*, które odpowiadają kolejnym zmiennym losowym:

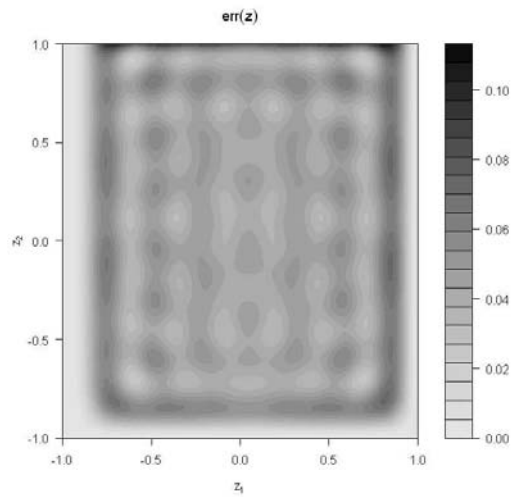
$$\omega_i(\mathbf{z}) = \frac{1}{\sqrt{\lambda_i}} \phi_i^T \mathbf{C}_{HX}(\mathbf{z}). \quad (4.93)$$

Wykresy pierwszych sześciu takich funkcji dla pola losowego W przedstawiono na rysunku 4.10. Warto zwrócić uwagę, że wzdłuż brzegu \mathcal{B}_1 funkcje kształtu spełniają warunek utwierdzenia.

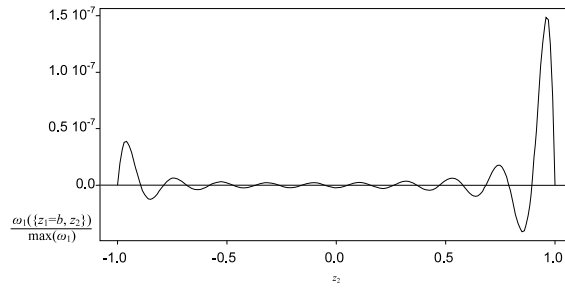
Jak wspomniano wcześniej, dla zagadnień określonych w dwóch wymiarach omawiane pole losowe nie spełnia dokładnie warunków brzegowych. Wykres funkcji kształtu ω_1 wzdłuż brzegu \mathcal{B}_1 został pokazany na rysunku 4.9. Na tym brzegu błąd przybliżenia był największy. Jednak nawet tutaj niepożądane przemieszczenia są kilka rzędów mniejsze w porównaniu do maksimum ω_1 .



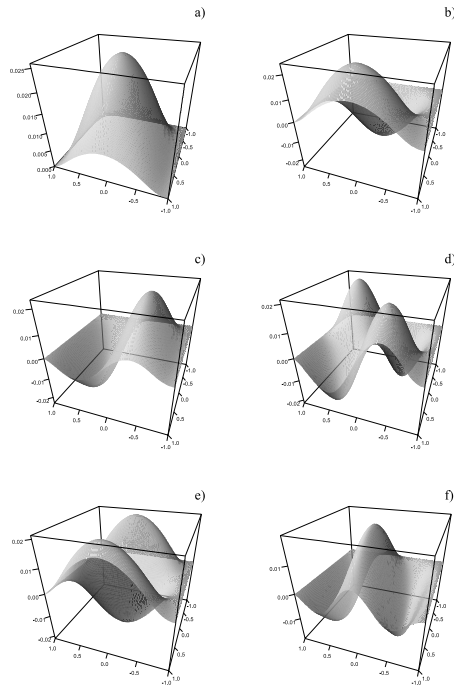
RYSUNEK 4.7. Błąd dyskretyzacji pola losowego metodą OLE.



RYSUNEK 4.8. Błąd dyskretyzacji pola losowego metodą EOLE.



RYSUNEK 4.9. Wykres funkcji kształtu $\omega_1(\mathbf{z}) / \max(\omega_1)$ wzdłuż brzegu \mathcal{B}_1 .



RYSUNEK 4.10. (a - f) Odpowiednio wykresy pierwszych sześciu funkcji kształtu dyskretyzacji EOLE.

Niezawodność zależna od czasu

5.1. Wstęp

W tym rozdziale zajmiemy się analizą niezawodności konstrukcji, których pewne parametry o charakterze losowym są zależne od czasu. Jak już wspomniano we wstępie do rozdziału 4, takie parametry konstrukcyjne są modelowane za pomocą procesów losowych. Obok wspomnianych już obciążeń, istotną zależność od czasu mogą wykazywać niektóre parametry, mające wpływ na wytrzymałość konstrukcji. Przyczyną tego jest starzenie się materiałów, korozja, bądź zmęczenie.

Wprowadzenie do modelu konstrukcji parametrów zależnych od czasu pociąga za sobą konieczność rozszerzenia pojęcia niezawodności. Celem *niezawodności zależnej od czasu* (ang. *time variant reliability*) jest oszacowanie prawdopodobieństwa awarii w określonym przedziale czasu. Najczęściej przedział czasu, w którym jest analizowana niezawodność, odpowiada przewidywanemu okresowi użytkowania konstrukcji. Poniżej przedstawimy kilka sposobów sformułowania tego zagadnienia, jakie można znaleźć w literaturze. Dalej zostało zamieszczone krótkie wprowadzenie do zagadnienia szacowania *częstości wyjść do obszaru awarii*, która jest podstawową charakterystyką procesów losowych wykorzystywaną w analizie niezawodności. Jak zobaczymy, analiza niezawodności zależna od czasu jest pojęciowo i obliczeniowo bardziej skomplikowana niż zagadnienie sformułowane w rozdziale 2.

5.2. Sformułowanie zagadnienia niezawodności zależnej od czasu

Elementarnym pojęciem analizy niezawodności zależnej od czasu jest *chwilowe prawdopodobieństwo awarii* (ang. *instantaneous probability of failure*), określane ja-

ko:

$$P_f(\tau) = P(g(\mathbf{X}(\tau), \tau) \leq 0) = \int_{g(\mathbf{X}(\tau), \tau) \leq 0} f_{X(\tau)}(\mathbf{x}(\tau)) d\mathbf{x}(\tau), \quad (5.1)$$

gdzie $g(\mathbf{X}(\tau), \tau)$ jest zależną od czasu funkcją graniczną a $f_{X(\tau)}(\mathbf{x}(\tau))$ jest funkcją gęstości prawdopodobieństwa wartości procesu losowego $\mathbf{X}(\tau)$. Proces ten jest modelem zmiennych w czasie parametrów konstrukcyjnych. Jak można zauważyć, jest to prawdopodobieństwo awarii znane ze sformułowania niezależnego od czasu (2.6), określone dla wartości procesu losowego $\mathbf{X}(\tau)$ w chwili τ . Posługiwanie się taką miarą niezawodności ma jednak sens jedynie wtedy, gdy mamy pewność, że awaria nie wystąpiła przed chwilą τ bądź też konstrukcja jest w stanie powrócić do stanu bezpiecznego po wystąpieniu awarii. Tak, więc w praktyce definicja (5.1) może być wykorzystana w przypadku stanów granicznych użytkownika.

Jak wspomniano, celem analizy niezawodności zależnej od czasu jest oszacowanie prawdopodobieństwa awarii w przedziale czasu, np. $[0, t]$, które jest określane jako:

$$P_f(0, t) = P(\exists \tau \in [0, t] : g(\mathbf{X}(\tau), \tau) \leq 0), \quad (5.2)$$

bądź alternatywnie, za pomocą wyrażenia:

$$P_f(0, t) = 1 - P(\forall \tau \in [0, t] g(\mathbf{X}(\tau), \tau) > 0). \quad (5.3)$$

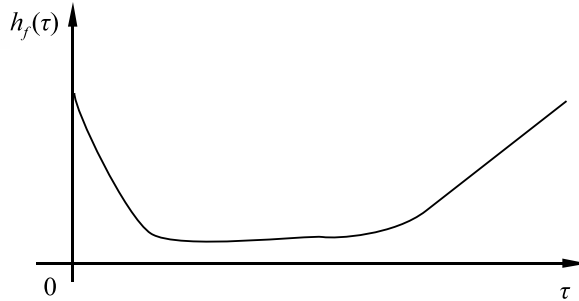
Powyższe prawdopodobieństwa awarii mogą być wyznaczone przez całkowanie (5.1) w przedziale $[0, t]$, jednak wymaga to uwzględnienia korelacji między wartościami procesu losowego $\mathbf{X}(\tau)$ w różnych chwilach czasu. W praktyce stosuje się więc przybliżenia prawdopodobieństwa zdefiniowanego wzorem (5.2).

5.2.1. Transformacja do problemu niezależnego od czasu

Klasycznym sposobem oszacowania prawdopodobieństwa awarii w przedziale czasu, jest transformacja do równoważnego problemu, który od czasu zależny nie jest. Metoda ta wymaga przyjęcia założenia, że parametry wpływające na wytrzymałość konstrukcji są niezależne od czasu. Ponadto zakłada się, że możliwe jest określenie skalarne procesu losowego $Y(\tau)$ reprezentującego kombinacje obciążeń, którym poddana jest konstrukcja (jest to tzw. problem kombinacji obciążeń, omówiony szerzej w [73]). Jeżeli dla procesu $Y(\tau)$ można wyznaczyć zmienną losową $Y_{MAX}(0, t)$, o rozkładzie prawdopodobieństwa maksimum $Y(\tau)$ w przedziale $[0, t]$, to

$$P_f(0, t) = P[g(\mathbf{R}, Y_{MAX}(0, t)) \leq 0], \quad (5.4)$$

gdzie \mathbf{R} jest wektorem losowym reprezentującym parametry wytrzymałościowe.



RYSUNEK 5.1. Typowy przebieg funkcji ryzyka.

5.2.2. Klasyczna teoria niezawodności

W klasycznej torii niezawodności elementów systemów technicznych ocena prawdopodobieństwa awarii uzyskiwana jest na podstawie *funkcji ryzyka* (ang. *hazard function*). Aby określić funkcję ryzyka przedstawimy prawdopodobieństwo awarii w przedziale czasu $[0, t]$ jako:

$$P_f(0, t) = F_f(t), \quad (5.5)$$

gdzie $F_f(\tau)$ jest dystrybuantą czasu wystąpienia awarii. Prawdopodobieństwo awarii w przedziale czasu $\tau + d\tau$, $d\tau \rightarrow 0$, określane również jako *bezw warunkowa częstość awarii* (ang. *unconditional failure rate*), na podstawie (5.5) można przedstawić za pomocą wyrażenia:

$$f_f(\tau) = \frac{dF_f(\tau)}{d\tau}. \quad (5.6)$$

Funkcja ryzyka, także nazywana *warunkową częstością awarii* (ang. *conditional failure rate, age specific failure rate*), definiuje prawdopodobieństwo wystąpienia awarii w przedziale czasu $[\tau, \tau + d\tau]$, $d\tau \rightarrow 0$ przy założeniu, że awaria nie wystąpiła wcześniej:

$$\begin{aligned} h_f(\tau) &= P(\text{awaria w } [\tau, \tau + d\tau] \mid \text{brak awarii w } [0, \tau]) \\ &= \frac{P_f(\tau, \tau + d\tau)}{1 - P_f(0, \tau)} \\ &= \frac{f_f(\tau)}{F_f(0, \tau)}. \end{aligned} \quad (5.7)$$

Dysponując funkcją ryzyka, prawdopodobieństwo awarii w przedziale czasu $[0, t]$ można określić następującym wyrażeniem [73]:

$$P_f(0, t) = F_f(t) = 1 - \exp\left(-\int_0^t h_f(\tau)d\tau\right) \approx \int_0^t h_f(\tau)d\tau. \quad (5.8)$$

Praktyczne zastosowanie (5.8) jest oczywiście uzależnione od możliwości wyznaczenia funkcji ryzyka. Klasyczna teoria niezawodności narodziła się w czasie drugiej wojny światowej wraz z badaniami nad zmniejszeniem awaryjności systemów elektrycznych, które wykorzystywano w urządzeniach wojskowych. Później znalazła zastosowanie w analizie systemów technicznych składających się z wielu elementów tego samego typu, poddanych identycznemu obciążeniu i w praktyce zachowujących się statystycznie niezależnie. W przypadku takich systemów funkcje ryzyka mogą być wyznaczone na podstawie danych eksploatacyjnych.

Zwróćmy uwagę, że w przypadku analizy niezawodności konstrukcji w ujęciu będącym przedmiotem tej pracy nie jest możliwe wyznaczenie funkcji ryzyka na podstawie danych doświadczalnych. Zdarzenia awarii są bowiem bardzo rzadkie i zazwyczaj są konsekwencją sytuacji ekstremalnych. Ponadto elementy wykorzystywane w systemach konstrukcyjnych są często unikatowe.

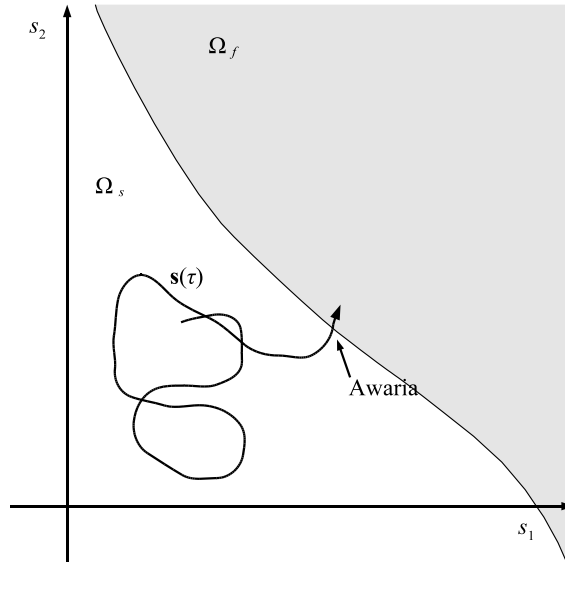
Na rysunku 5.1 przedstawiono, przypominający wannę, wykres funkcji ryzyka, który jest charakterystyczny dla większości systemów technicznych. Zwiększone ryzyko awarii w początkowym okresie użytkowania jest charakterystyczne dla urządzeń w przypadku, których błędy produkcji stanowią istotne zagrożenie wystąpienia awarii. Po czasie, kiedy ujawniają się wady produkcyjne prawdopodobieństwo awarii maleje. W końcowym okresie użytkowania prawdopodobieństwo awarii ponownie wzrasta w wyniku starzenia się komponentów. Warto zauważyć, że znajomość krzywej ryzyka może być bardzo pomocna przy wyznaczaniu harmonogramu inspekcji.

5.2.3. Prawdopodobieństwo pierwszego przekroczenia

Jednym z problemów badanych w ramach teorii procesów stochastycznych jest *zagadnienie przekroczeń* (ang. *barrier crossing, level crossing*). W przypadku skalarnego procesu losowego $X(\tau)$ zagadnienie to wiąże się z oszacowaniem prawdopodobieństwa zdarzenia

$$S = \{X(\tau) > a(\tau), \tau \in [0, t]\}, \quad (5.9)$$

gdzie $[0, t]$ jest badanym przedziałem czasu. Wystąpienie zdarzenia S oznacza, że realizacja procesu $X(\tau)$ przekroczyła deterministyczną barierę $a(\tau)$. Zauważmy, że zdarzenie awarii można postrzegać jako przekroczenie przez skalarny proces losowy $-g(\mathbf{Z}(\tau))$



RYSUNEK 5.2. Graficzna interpretacja problemu pierwszego przekroczenia.

poziomu 0, gdzie \mathbf{Z} jest wektorem losowym reprezentującym parametry konstrukcji. Posługiwanie się takim sformułowaniem awarii nie jest wygodne, gdyż wymaga określenia własności probabilistycznych procesu $g(\mathbf{Z}(\tau))$. Dlatego w praktyce analizuje się uogólnienie problemu przekroczeń: tak zwany *problem wyjść* wektorowego procesu losowego $\mathbf{Z}(\tau)$ do obszaru awarii Ω_f . Graficzną interpretację tego zagadnienia przedstawiono na rysunku 5.2.

Korzystając z pojęcia przekroczeń problem analizy niezawodności w przedziale czasu $[0, t]$ formułuje się jako wyznaczenie $P_f(0, t)$ - prawdopodobieństwa wyjścia do obszaru awarii $\Omega_f(\tau) = \{\mathbf{z}(\tau) : g(\mathbf{z}(\tau), \tau) \leq 0, \tau \in [0, t]\}$ realizacji $\mathbf{z}(\tau)$ wektora losowych parametrów konstrukcji pod warunkiem, że $g(\mathbf{z}(\tau = 0)) > 0$.

Prawdopodobieństwo awarii w przedziale czasu $[0, t]$ można określić za pomocą prawdopodobieństwa pozostania przez $\mathbf{Z}(\tau)$ w obszarze bezpiecznym:

$$P_f(0, t) = 1 - P(N_{\mathbf{Z}}^+(t) = 0 | g(\mathbf{Z}(\tau = 0)) > 0) P(g(\mathbf{Z}(\tau = 0)) > 0), \quad (5.10)$$

gdzie $N_Z^+(t)$ jest liczbą wyjść $\mathbf{Z}(\tau)$ do obszaru awarii w przedziale czasu $[0, t]$.

W dalszej części tego rozdziału będziemy wykorzystywać, zaproponowany w pracy [105], podział parametrów losowych konstrukcji na trzy grupy:

$$\mathbf{Z}(\tau) = \{\mathbf{R}, \mathbf{Q}(\tau), \mathbf{S}(\tau)\}, \quad (5.11)$$

gdzie \mathbf{R} jest wektorem niezależnych od czasu zmiennych losowych, $\mathbf{Q}(\tau)$ jest stacjonarnym, ergodycznym wektorowym procesem losowym modelującym parametry zmieniające się w długich okresach czasu, natomiast $\mathbf{S}(\tau)$ jest procesem losowym, nie koniecznie stacjonarnym, ale wykazującym częste fluktuacje w porównaniu do $\mathbf{Q}(\tau)$.

W celu oszacowania $P_f(0, t)$ na wstępie rozważymy przypadek, gdy w modelu występuje jedynie proces $\mathbf{S}(\tau)$. W ogólnym przypadku wyznaczenie prawdopodobieństwa $P_f(0, t)$ jest trudnym zagadnieniem, gdyż wymaga uwzględnienia historii procesu $\mathbf{S}(\tau)$ w całym przedziale czasu $[0, t]$. Rozwiązanie tego problemu wykorzystywane w analizie niezawodności opiera się na przyjęciu pewnych uproszczeń. Ponieważ systemy inżynierskie projektowane są w sposób zapewniający możliwie najwyższy poziom bezpieczeństwa to przyjmuje się, że awarie (wyjścia realizacji procesu \mathbf{Z} do obszaru awarii) zdarzają się rzadko i w praktyce są to zdarzenia niezależne. Zakłada się, więc że każde wyjście do obszaru awarii jest niezależne od jakiegokolwiek wcześniejszego wyjścia, a ich liczba w przedziale czasu $[0, t]$ ma rozkład Poissona.

Prawdopodobieństwo, że nie wystąpi wyjście do obszaru awarii w przedziale czasu $[0, t]$ przybliża się, więc za pomocą prawdopodobieństwa zerowej liczby zdarzeń rozkładu Poissona [20]:

$$\begin{aligned} P(N_S^+(0, t) = 0) &= \frac{\mathbb{E}[N_S^+(0, t)]^0}{0!} \exp(-\mathbb{E}[N_S^+(0, t)]) \\ &= \exp(-\mathbb{E}[N_S^+(0, t)]), \end{aligned} \quad (5.12)$$

gdzie $N_S^+(0, t)$ jest liczbą wyjść procesu $\mathbf{S}(\tau)$ do obszaru awarii w przedziale czasu $[0, t]$. Wartość oczekiwaną liczby wyjść do obszaru awarii definiuje się jako

$$\mathbb{E}[N_S^+(0, t)] = \int_0^t \nu_S^+(\tau) d\tau, \quad (5.13)$$

gdzie ν_S^+ jest *średnią częstością wyjść* procesu $\mathbf{S}(t)$ do obszaru awarii. W przypadku procesów stacjonarnych, średnia częstość wyjść jest funkcją stałą, a więc $\mathbb{E}[N_S^+(0, t)] = t\nu_S^+$.

Przedstawmy prawdopodobieństwo, że proces $\mathbf{S}(\tau)$ w chwili $\tau = 0$ znajduje się w obszarze bezpiecznym jako

$$P(g(\mathbf{S}(\tau = 0)) > 0) = 1 - P_f(\tau = 0), \quad (5.14)$$

gdzie $P_f(\tau = 0)$ jest prawdopodobieństwem awarii w chwili $\tau = 0$. Wykorzystując wzór (5.14) oraz wyrażenie (5.12), po uwzględnieniu (5.13), prawdopodobieństwo awarii w przedziale czasu $[0, t]$ można wyrazić za pomocą oczekiwanej liczby wyjść $\mathbf{S}(\tau)$ do obszaru awarii [126]:

$$\begin{aligned} P_f(0, t) &= 1 - (1 - P_f(0)) \exp(-\mathbb{E}[N_S^+(0, t)]) \\ &= P_f(0) + (1 - P_f(0)) (1 - \exp(-\mathbb{E}[N_S^+(0, t)])) . \end{aligned} \quad (5.15)$$

Na podstawie zależności $\mathbb{E}[N_S^+(0, t)] > 1 - \exp(-\mathbb{E}[N_S^+(0, t)])$ i po przyjęciu założenia, że $P_f(\tau = 0)$ jest małe, otrzymuje się górne ograniczenie $P_f(0, t)$, które najczęściej podawane jest w postaci [13]:

$$P_f(0, t) \leq P_f(\tau = 0) + \mathbb{E}[N_S^+(0, t)] . \quad (5.16)$$

Ograniczenie to jest szczególnie przydatne w analizie niezawodności, ponieważ uwzględnia możliwość wystąpienia awarii w początkowej chwili czasu $P_f(\tau = 0)$ (na przykład podczas pierwszego obciążenia) oraz prawdopodobieństwo wystąpienia awarii w trakcie eksploatacji $\mathbb{E}[N_S^+(0, t)]$. Przybliżenie (5.16) jest powszechnie stosowane w praktyce i to nie tylko ze względu na swoją prostotę, ale również dlatego, że poza kilkoma specjalnymi przypadkami, nie ma ono praktycznej alternatywy [35].

Proste ograniczenie dolne (5.15) zostało zaproponowane w pracy [109]:

$$P_f(0, t) \geq \max_{0 \leq \tau \leq t} P_f(\tau) , \quad (5.17)$$

gdzie $P_f(\tau)$ jest prawdopodobieństwem awarii w chwili τ , określonym przez (5.1). Należy zaznaczyć, że ograniczenie to jest raczej zgrubne.

Jeżeli wartości $P_f(0, t)$ są małe to dość dokładne jest asymptotyczne przybliżenie prawdopodobieństwa awarii, które ponadto można ograniczyć z góry: [20, 105]:

$$P_f(0, t) \approx 1 - \exp(-\mathbb{E}[N_S^+(0, t)]) \leq \mathbb{E}[N_S^+(0, t)] . \quad (5.18)$$

Jeżeli w modelu obok procesu $\mathbf{S}(\tau)$ występują niezależne od czasu zmienne losowe \mathbf{R} to nieuprawnione jest przyjęcie założenia, że liczba wyjść do obszaru awarii ma rozkład Poissona. Jednak korzystając ze wzoru (5.15) można otrzymać warunkowe prawdopodobieństwo awarii w przedziale $[0, t]$ pod warunkiem $\mathbf{R} = \mathbf{r}$. Oszacowanie całkowitego prawdopodobieństwa awarii wymaga więc całkowania prawdopodobieństwa warunkowego po realizacjach \mathbf{R} . Analogicznie do wzoru (5.18), asymptotyczne przybliżenie całkowitego prawdopodobieństwa awarii oraz jego ograniczenie przedstawia się jako

$$P_f(0, t) \approx 1 - \mathbb{E}_R [\exp(-\mathbb{E}[N_S^+(0, t | \mathbf{R} = \mathbf{r})])] \leq \mathbb{E}_R [\mathbb{E}[N_S^+(0, t | \mathbf{R} = \mathbf{r})]] . \quad (5.19)$$

W ogólnym przypadku, kiedy model wykorzystuje wszystkie wyszczególnione powyżej rodzaje zmiennych losowych \mathbf{R} , $\mathbf{Q}(\tau)$ oraz $\mathbf{S}(\tau)$ to wyznaczenie $P_f(0, t)$ wymaga całkowania po zmiennych \mathbf{R} oraz wyznaczenia wartości oczekiwanej względem wolno zmieniających się w czasie zmiennych $\mathbf{Q}(\tau)$. W pracy [105] zaproponowano następujące przybliżenie:

$$P_f(0, t) \approx 1 - \mathbb{E}_R [\exp(-\mathbb{E}_Q [\mathbb{E} [N_S^+(0, t | \mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q})]])] \quad (5.20)$$

$$\leq \mathbb{E}_R [\mathbb{E}_Q [\mathbb{E} [N_S^+(0, t | \mathbf{R} = \mathbf{r}, \mathbf{Q} = \mathbf{q})]]] . \quad (5.21)$$

Wyrażenie (5.20) jest raczej dobrym przybliżeniem w przypadku stacjonarnym, jeżeli jednak $\mathbf{S}(\tau)$ jest niestacjonarny lub funkcja graniczna wykazuje silną zależność od czasu wyrażenie to powinno być traktowane jako pierwsze przybliżenie.

5.3. Częstość przekroczeń ciągłych procesów losowych

Jak pokazano w podrozdziale 5.2.3, średnia częstość przekroczeń jest charakterystyką procesu losowego pozwalającą wyznaczyć oczekiwaną liczbę wyjść do obszaru awarii (wzór (5.13)), a co za tym idzie oszacować prawdopodobieństwo awarii w przedziale czasu. W przypadku procesu losowego $\mathbf{S}(\tau)$ średnia częstość wyjść do obszaru awarii jest określana w następujący sposób:

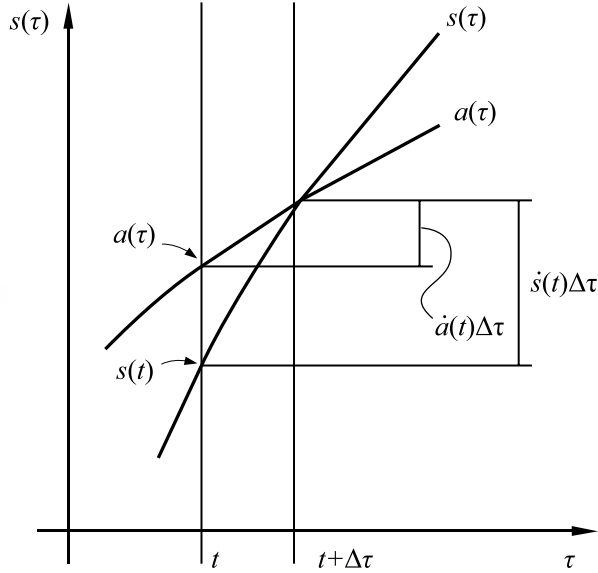
$$\nu_S^+(\tau) = \lim_{\Delta\tau \rightarrow 0} \frac{1}{\Delta\tau} P \left(g(\mathbf{S}(\tau), \tau) > 0 \cap g(\mathbf{S}(\tau + \Delta\tau), \tau + \Delta\tau) \leq 0 \right) . \quad (5.22)$$

Warunkiem istnienia ν_S^+ jest możliwość wykonania przejścia granicznego we wzorze (5.22). Tak, więc średnia częstość wyjść istnieje jedynie dla procesów, które nie wykazują zbyt dużej zmienności w czasie. Ponadto, aby istniało ν_S^+ to prawdopodobieństwo wystąpienia w krótkiej chwili czasu większej liczby przejść niż jedno powinno być pomijalnie małe.

Wyprowadzenie wzoru określającego średnią częstość wyjść do obszaru awarii zaczniemy od przypadku jednowymiarowego. Następnie otrzymane wyniki zostaną uogólnione na przypadek wielowymiarowy.

Rozważmy funkcję graniczną postaci: $g(S(\tau), \tau) = a(\tau) - S(\tau)$, gdzie $S(\tau)$ jest skalarnym procesem stochastycznym, natomiast $a(\tau)$ jest funkcją deterministyczną. W tym przypadku, wyjście procesu $S(\tau)$ do obszaru awarii jest równoważne przekroczeniu z dołu do góry bariery $a(\tau)$. Częstość przekroczeń różniczkowalnego procesu skalarnego przez barierę, określoną funkcją różniczkowalną, wyznacza się na podstawie tzw. *formuły Rice'a* [95].

W celu wyprowadzenia formuły Rice'a, rozpatrzmy pokazaną na rysunku 5.3 realizację $S(\tau)$ w przedziale czasu $[t, t + \Delta\tau]$. Bez straty ogólności można założyć, że w czasie $\Delta\tau$ występuje przekroczenie bariery.



RYSUNEK 5.3. Przekroczenie bariery przez proces skalarny.

Aby miało ono miejsce to w chwili t proces powinien znajdować się poniżej bariery $s(t) < a(t)$, a wartość pochodnej $\dot{s}(t)$ powinna zapewniać przekroczenie bariery podczas rozpatrywanego przyrostu czasu:

$$a(t) + \dot{a}(t)\Delta\tau < s(t) + \dot{s}(t)\Delta\tau, \quad \Delta\tau \rightarrow 0. \quad (5.23)$$

Uwzględniając powyższe warunki prawdopodobieństwo przekroczenia bariery w przedziale czasu $\Delta\tau$ można wyrazić jako:

$$\begin{aligned} P(S(t) < a(t) \cap S(t + \Delta\tau) \geq a(t + \Delta\tau)) &= \\ &= P(a(t) - (\dot{S}(t) - \dot{a}(t))\Delta\tau < X(t) < a(t)) \\ &= \int_{\dot{a}(t)}^{\infty} \int_{a(t) - (\dot{s}(t) - \dot{a}(t))\Delta\tau}^{a(t)} f_{S\dot{S}}(s, \dot{s}) ds d\dot{s} \\ &\stackrel{\Delta\tau \rightarrow 0}{\approx} \Delta\tau \int_{\dot{a}(t)}^{\infty} (\dot{s}(t) - \dot{a}(t)) f_{S\dot{S}}(a(t), \dot{s}) d\dot{s}, \end{aligned} \quad (5.24)$$

gdzie $f_{S\dot{S}}$ jest łączną gęstością prawdopodobieństwa procesu $S(\tau)$ i procesu jego pochodnej po czasie $\dot{S}(\tau)$. Korzystając z (5.24), na podstawie (5.22), średnią częstość przekroczeń bariery przedstawia się w następujący sposób:

$$\nu_S^+ = \int_{\dot{a}(t)}^{\infty} (\dot{s}(t) - \dot{a}(t)) f_{S\dot{S}}(a(t), \dot{s}) d\dot{s}. \quad (5.25)$$

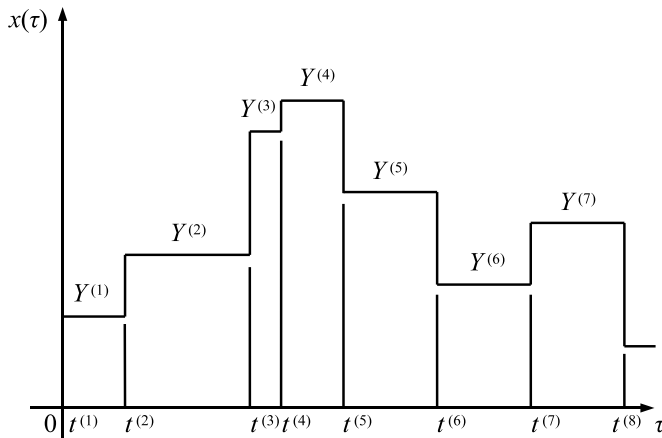
Zauważmy jeszcze, że jeżeli funkcja a jest niezależna od czasu to oczywiście w dolnej granicy powyższej całki $\dot{a}(t) = 0$.

Formułę Rice'a można łatwo uogólnić w celu określenia częstości wyjść wektorowego procesu stochastycznego do obszaru awarii. Niech $\mathbf{S}(\tau)$ będzie wektorowym procesem stochastycznym o m składowych. Przyjmijmy, że $\mathbf{s}(t)$ znajduje się na powierzchni awarii $\partial\Omega_f = \{\mathbf{s} : g(\mathbf{s}(t)) = 0\}$ w punkcie A . Wyjście do obszaru awarii wystąpi w chwili t , jeżeli $\mathbf{s}(t + \Delta\tau)$, $\Delta\tau \rightarrow 0$, znajduje się poza obszarem bezpiecznym. Przyrost realizacji $\mathbf{s}(\tau)$ w przedziale czasu $\Delta\tau \rightarrow 0$ możemy przedstawić jako $\dot{\mathbf{s}}(t)\Delta\tau$, gdzie $\dot{\mathbf{S}}(\tau) = \{\dot{S}_1(\tau), \dots, \dot{S}_m(\tau)\}$ jest wektorem pochodnych po czasie składowych $\mathbf{S}(\tau)$. Aby wystąpiło wyjście do obszaru awarii $\mathbf{n}(A, t)$ - wektor normalny do powierzchni awarii w punkcie A , skierowany na zewnątrz obszaru bezpiecznego - powinien spełniać warunek: $\mathbf{n}(A, t) \cdot \dot{\mathbf{s}}(t) > 0$. Dla składowych w bazie kartezjańskiej iloczyn skalarny wektorów $\dot{\mathbf{s}}$ i \mathbf{n} jest postaci: $\dot{s}_n(t) = \mathbf{n}(A, t) \cdot \dot{\mathbf{s}} = \sum_{i=1}^N n_i(A, t) s_i(t)$. Porównując rozważany problem z zagadnieniem jednowymiarowym widzimy, że \dot{s}_n odpowiada \dot{s} . Tak, więc (5.25) można uogólnić dla procesu wektorowego w następujący sposób:

$$\nu_S^+ = \int_{\partial\Omega_f} d\mathbf{s} \int_0^{\infty} \dot{x}_n f_{S\dot{S}_n}(\mathbf{s}, \dot{s}_n) d\dot{x}_n, \quad (5.26)$$

gdzie $f_{S\dot{S}_n}$ jest gęstością łącznego rozkładu prawdopodobieństwa procesów losowych \mathbf{S} i \dot{S}_n , którego realizacjami są $\dot{\mathbf{s}}_n$. Pierwsza z całek we wzorze (5.26) zapewnia, że uwzględniona zostanie możliwość wyjścia w każdym z punktów powierzchni awarii. Formalne wyprowadzenie (5.26) można znaleźć w pracy [9].

Należy zaznaczyć, że rozwiązanie problemu (5.26) nie jest łatwe. Istnieją rozwiązania analityczne dla stałych w czasie obszarów awarii przy założeniu, że $\mathbf{S}(\tau)$ i $\dot{\mathbf{S}}(\tau)$ są wzajemnie niezależne. Niestety modele zjawisk występujących w rzeczywistości nie spełniają tych założeń. W przypadku pewnych klas procesów opracowano metody umożliwiające przybliżenie, bądź wyznaczenie ograniczeń częstości wyjść do obszaru awarii. Przegląd tych metod zamieszczono w pracy [92].



RYSUNEK 5.4. Realizacja skalarnej realizacji procesu odnowy o prostokątnych impulsach.

5.4. Częstość wyjść procesu odnowy o prostokątnych impulsach

Zjawiska, których intensywność jest wartością losową zmieniającą się w losowych chwilach czasu mogą być dobrze modelowane przez tzw. *procesy odnowy o prostokątnych impulsach* (ang. *rectangular wave renewal processes*). Procesy tego rodzaju znajdują zastosowanie, jako modele obciążenia stropów w obiektach takich jak biura, szpitale czy parkingi. Wykres realizacji skalarnej realizacji procesu odnowy o prostokątnych impulsach przedstawia rysunek 5.4.

Ogólnie proces odnowy o prostokątnych impulsach $X(\tau)$, $\tau \in [0, t]$ można przedstawić w następujący sposób:

$$X(\tau) = \sum_{k=1}^N I_{T^{(k)}}(\tau) Y^{(k)}, \quad (5.27)$$

gdzie $Y^{(k)}$ są niezależnymi zmiennymi losowymi o identycznych rozkładach prawdopodobieństwa, $t^{(k)}$ są chwilami, w których realizacja procesu zmienia wartość - występują *odnowienia procesu*, N jest liczbą odnowień, a $I_{T^{(k)}}$ są funkcjami charakterystycznymi przedziałów czasu $T^{(k)} = [t^{(k)}, t^{(k+1)})$, $k = 1, \dots, N$. Istotną w analizie niezawodności charakterystyką procesu odnowy o prostokątnych impulsach jest *intensywność*

odnowień λ , którą możemy określić jako:

$$\lambda = \frac{1}{\mathbb{E}[t^{(k+1)} - t^{(k)}]}, \quad (5.28)$$

gdzie $\mathbb{E}[t^{(k+1)} - t^{(k)}]$ jest oczekiwaną długością przedziału czasu pomiędzy odnowieniami.

Elementarnym przykładem procesu odnowy o prostokątnych impulsach jest *proces Borgesa* [37]. W tym przypadku, $t^{(k)}$ są ustalonymi, równoodległymi chwilami czasu. Ponieważ T_b , będące długościami przedziałów $[t^{(k)}, t^{(k+1)})$ są stałe to $N = t/T_b$, a $\lambda = 1/T_b$. Wykres realizacji procesu Borgesa jest sekwencją prostokątnych fal o równej długości.

Innym przykładem procesu należącego do omawianej klasy jest tzw. *proces Poissona o prostokątnych impulsach* (ang. *Poisson square wave process*). W tym przypadku chwile $t^{(k)}$ są generowane przez proces Poissona $N(t)$. Jeżeli przyjmiemy założenie, że $P(N(\tau = 0) = 0) = 1$, to $N = N(t)$, gdzie $N(t)$ jest zmienną losową o rozkładzie Poissona. Prawdopodobieństwo wystąpienia k odnowień jest, więc określone wzorem:

$$P(N(t) = k) = \frac{\mu(t)^k}{k!} \exp(-\mu(t)), \quad (5.29)$$

gdzie $\mu(t) = \mathbb{E}[N(t)]$. Jeżeli $\mu(t) = t\lambda$, to proces $N(t)$ jest jednorodnym procesem Poissona. W przypadku procesu niejednorodnego można wprowadzić *funkcję intensywności procesu* $\lambda(\tau)$, dla której zachodzi związek: $\lambda(t) = \int_0^t \nu(\tau) d\tau$.

Zastosowanie w analizie niezawodności procesów odnowy o prostokątnych impulsach wymaga oszacowania oczekiwanej liczby wyjść do obszaru awarii. Załóżmy, że $\mathbf{S}(\tau)$ jest wektorowym procesem losowym, którego składowe są niezależnymi stacjonarnymi procesami odnowy o prostokątnych impulsach. W pracy [16] pokazano, że oczekiwaną liczbę wyjść $\mathbf{S}(\tau)$ do obszaru awarii w przedziale czasu $[0, t]$ można wyznaczyć w następujący sposób:

$$\begin{aligned} \mathbb{E}[N_S^+(0, t)] &= t \sum_{i=1}^m \lambda_i (P(\{\mathbf{S}_i^- \in \Omega_s\} \cap \{\mathbf{S}_i^+ \in \Omega_f\})) \\ &= t \sum_{i=1}^m \lambda_i (P(\{\mathbf{S}_i^+ \in \Omega_f\}) - P(\{\mathbf{S}_i^+ \in \Omega_f\} \cap \{\mathbf{S}_i^- \in \Omega_f\})) \\ &\leq t \sum_{i=1}^m \lambda_i P(\{\mathbf{S} \in \Omega_f\}), \end{aligned} \quad (5.31)$$

gdzie m jest liczbą niezależnych składowych procesu, λ_i , $i = 1, \dots, m$, są intensywnościami odpowiednich składowych, \mathbf{S}_i^- jest wektorem losowym tuż przed odnowieniem

i -tej składowej, a \mathbf{S}_i^+ jest wektorem losowym tuż po odnowieniu i -tej składowej. Zazwyczaj (a w szczególności w przypadku małych prawdopodobieństw awarii) prawdopodobieństwo $P(\{\mathbf{S}_i^+ \in \Omega_f\} \cap \{\mathbf{S}_i^- \in \Omega_f\})$ jest na tyle małe, że wystarczającą dokładność zapewnia posługiwanie się, wygodnym obliczeniowo, oszacowaniem górnym (5.31). Powyższe wzory są prawdziwe przy założeniu, że składowa S_i zmienia jedną wartość losową na drugą (nie występuje powrót do zera lub wartości średniej).

W przypadku, kiedy powierzchnia awarii jest hiperpłaszczyzną określoną w gausowskiej przestrzeni standardowej $\partial\Omega_f = \boldsymbol{\alpha}^T \mathbf{u} + \beta$ (wartości składowych procesu są standardowymi zmiennymi normalnymi) częstość wyjść procesu odnowy o prostokątnych impulsach do obszaru awarii może być określona za pomocą następującego wzoru [16]:

$$\nu_S^+ = \sum_{i=1}^m \lambda_i (\Phi(-\beta) - \Phi_2(-\beta, -\beta, \rho_i)) , \quad (5.32)$$

gdzie Φ_2 jest dwuwymiarową dystrybuantą rozkładu normalnego, a $\rho_i = 1 - \alpha_i^2$ jest współczynnikiem korelacji tego rozkładu.

Jeżeli w rozważanym modelu występują, wprowadzone w podrozdziale 5.2.3, zmienne losowe \mathbf{Q} oraz \mathbf{R} to oczekiwaną liczbę wyjść, określoną wzorem (5.30), należy traktować jako wartość warunkową pod warunkiem $\mathbf{Q} = \mathbf{q}$, $\mathbf{R} = \mathbf{r}$. Warunkowa postać górnego ograniczenia oczekiwanej liczby wyjść do obszaru awarii (5.31) jest więc dana wzorem:

$$\mathbb{E} [N_S^+(0, t | \mathbf{q}, \mathbf{r})] \leq t \sum_{i=1}^m \lambda_i P(\{\mathbf{S} \in \Omega_f | \mathbf{q}, \mathbf{r}\}) . \quad (5.33)$$

Wzór, który określa oczekiwaną liczbę przekroczeń niestacjonarnego procesu odnowy o prostokątnych impulsach w zasadzie nie jest bardziej skomplikowany niż dla procesu stacjonarnego. W tym przypadku zakłada się, że $\lambda_i(\tau)$, $i = 1, \dots, m$, - intensywności niestacjonarnych składowych $\mathbf{S}(\tau)$ wolno zmieniają się w czasie. Ponadto przyjmuje się, że $\mathbf{r}(\tau)$ - parametry rozkładu prawdopodobieństwa wartości $\mathbf{S}(\tau)$ mogą zależeć od czasu oraz obszar awarii może być funkcją czasu, np.: $\Omega_f = \{g(\mathbf{s}, \mathbf{q}, \mathbf{r}, \tau)\}$. W przypadku tak sformułowanego problemu warunkową wartość oczekiwaną wyjść do obszaru awarii (5.33) należy zmodyfikować do następującej postaci:

$$\begin{aligned} \mathbb{E} [N_S^+(0, t | \mathbf{r})] &\approx \mathbb{E}_Q \left[\int_0^t \sum_{i=1}^m \lambda_i(\tau) P(\{\mathbf{S} \in \Omega_f | \mathbf{Q}, \mathbf{r}, \tau\}) d\tau \right] \\ &= \int_0^t \int_{\Omega_f} \sum_{i=1}^m \lambda_i(\tau) f_{S,Q}(\mathbf{s}, \mathbf{q}, \tau | \mathbf{r}) d\mathbf{s} d\mathbf{q} d\tau , \end{aligned} \quad (5.34)$$

gdzie $f_{S,Q}$ jest zależną od czasu gęstością łącznego rozkładu prawdopodobieństwa wartości procesów \mathbf{S} i \mathbf{Q} .

Jeden ze sposobów oszacowania (5.34) opiera się na założeniu, że w przypadku małych prawdopodobieństw awarii istotną rolę przy wyznaczeniu zawartych w tym wzorze całek odgrywa jedynie masa prawdopodobieństwa skupiona wokół punktu w obszarze awarii $\mathbf{z}^* = \{\mathbf{s}^*, \mathbf{q}^*, t^*\}$, który ma największą gęstość prawdopodobieństwa. Takie założenie pozwala zastosować metody znane z niezależnej od czasu analizy niezawodności jak FORM/SORM czy importance sampling. Jednak w tym celu należy nieznacznie zmodyfikować (5.34). Ponieważ przyjęto założenie, że $\lambda_i(\tau)$ wolno zmieniają się w czasie to wielkości te można przenieść przed całkę po czasie w (5.34) z wartościami $\lambda_i(t^*)$. Ponadto dla czasu τ należy wprowadzić dodatkowy rozkład prawdopodobieństwa o gęstości $f_t(\tau) = t^{-1}$. Teraz (5.34) można przedstawić w następującej postaci

$$\mathbb{E} [N_S^+(0, t|\mathbf{r})] = t \sum_{i=1}^m \lambda_i(t^*) \int_{\mathbf{R}_1} \int_{\Omega_f} f_{S,Q}(\mathbf{s}, \mathbf{q}, \tau|\mathbf{r}) f_t(\tau) ds d\mathbf{q} d\tau . \quad (5.35)$$

Oszacowanie prawdopodobieństwa awarii (5.20) wymaga wyznaczenia wartości oczekiwanej funkcji $\exp(-\mathbb{E} [N_S^+(0, t|\mathbf{r})])$ względem wektora losowego \mathbf{R} . W ogólnym przypadku można się w tym celu posłużyć metodą importance sampling ([91]), wykorzystującą następujące wyrażenie całkowite

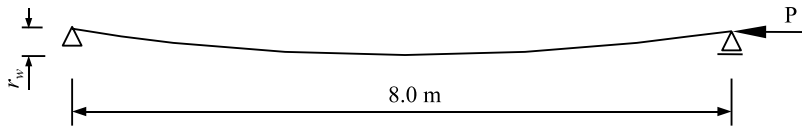
$$\begin{aligned} P_f(0, t) &\approx 1 - \mathbb{E}_R [\exp(-\mathbb{E}_Q [\mathbb{E} [N_S^+(0, t|\mathbf{R}, \mathbf{Q})]])] \\ &= 1 - \int_{\mathbf{R}_{M_r}} (\exp(-\mathbb{E} [N_S^+(0, t|\mathbf{r})])) \frac{f_R(\mathbf{r})}{h_R(\mathbf{r})} h_R(\mathbf{r}) d\mathbf{r} , \end{aligned} \quad (5.36)$$

gdzie M_r oznacza wymiar wektora losowego \mathbf{R} , f_R jest gęstością prawdopodobieństwa \mathbf{R} , a h_R jest gęstością rozkładu, z którego generowana jest próba losowa.

Zauważmy, że całka we wzorze (5.36) nie może być wyznaczona metodami FORM/SORM, ze względu na występującą w niej funkcję exp. Metody te można jednak zastosować do obliczenia ograniczenia (5.21), gdyż występujące w tym wzorze wartości oczekiwane można przedstawić za pomocą całek gęstości prawdopodobieństwa po obszarze awarii [22]:

$$\begin{aligned} &\mathbb{E}_R [\mathbb{E}_Q [\mathbb{E} [N_S^+(0, t|\mathbf{R}, \mathbf{Q})]]) \\ &\leq t \sum_{i=1}^M \lambda_i(t^*) \mathbb{E}_R [\mathbb{E}_Q [P(\{\mathbf{S} \in \Omega_f | \mathbf{Q}, \mathbf{R}, t^*\})]] \\ &= t \sum_{i=1}^M \lambda_i(t^*) \int_{\mathbf{R}_1} \int_{\Omega_f} f_{S,Q,R}(\mathbf{s}, \mathbf{q}, \mathbf{r}|\tau) f_t(\tau) ds d\mathbf{q} d\mathbf{r} d\tau , \end{aligned} \quad (5.37)$$

gdzie $f_{S,Q,R}$ jest gęstością łącznego rozkładu prawdopodobieństwa w zmiennych \mathbf{S} , \mathbf{Q} , \mathbf{R} .



RYSUNEK 5.5. Schemat obciążenia i podparcia ściskanego pręta ze wstępnym ugięciem.

Przykład 5.1: Analiza niezawodności zależnej od czasu ściskanego pręta ze wstępnym ugięciem.

Celem przykładu jest oszacowanie prawdopodobieństwa awarii pręta przedstawionego na rysunku 5.1. Przyjęto, że awaria ma miejsce w wyniku przekroczenia wartości dopuszczalnej przez obciążenie P , które ma postać procesu odnowy o prostokątnych impulsach.

Założono, że istotne w dalszej części przykładu, deterministyczne parametry projektowe mają wartości: długość $L = 8.0\text{m}$, pole przekroju $A = 5.250 \cdot 10^{-3}\text{m}^2$, moment bezwładności przekroju $J = 6.901 \cdot 10^{-5}\text{m}^4$, odległość skrajnego punktu przekroju od osi obojętnej $z = 0.125\text{m}$, moduł Younga $E = 2.1 \cdot 10^5\text{MPa}$. Ponadto przyjęto, że następujące parametry mają charakter zmiennych losowych (parametry typu R, zgodnie z klasyfikacją w roz. 5.2.3): maksymalne wstępne ugięcie pręta $R_w \sim \mathcal{N}(\mu_w = 0, \sigma_w = L/1000)$ oraz granica plastyczności materiału $R_S \sim \mathcal{N}(\mu_f = 500\text{MPa}, \sigma_f = 25\text{MPa})$.

Siła P - będąca obciążeniem pręta - jest modelowana przez stacjonarny proces odnowy o prostokątnych impulsach S_P . Jego intensywność wynosi $\lambda = 0.2$ a amplituda P ma rozkład normalny o wartości oczekiwanej $\mu_P = 0.5P_E$ i odchyleniu standardowym $\sigma_P = 0.1\mu_P$ ($P_E = \pi^2 EJ/L^2$ jest siłą Eulera w pręcie bez wstępnego ugięcia).

Prawdopodobieństwo awarii zostanie wyznaczone w przedziale czasu o długości $\Delta t = 10$, którego jednostki mogą być traktowane jako lata. Ponieważ przyjęliśmy $\lambda = 0.2$ możemy oczekiwać, że obciążenie zmienia swoją intensywność średnio co pięć lat.

W przypadku analizowanego pręta, wartość dopuszczalnego obciążenia wyznaczmy za pomocą formuły Perry-Robertsona:

$$P_d = \frac{P_S + (1 + \eta)P_E}{2} - \sqrt{\left(\frac{P_S + (1 + \eta)P_E}{2}\right)^2 - P_S P_E}, \quad (5.38)$$

gdzie $P_S = Ar_S$ oraz $\eta = zr_w A/J$.

Ponieważ w omawianym zadaniu nie występują wielkości losowe typu \mathbf{Q} (por. roz. 5.2.3) to przybliżenie prawdopodobieństwa awarii można uzyskać za pomocą wzorów (5.19).

Wykorzystując oszacowanie oczekiwanej liczby wyjść do obszaru awarii (5.33), górne

ograniczenie prawdopodobieństwa awarii wyznaczono za pomocą następującego wyrażenia:

$$P_f(\Delta t) \leq \lambda \Delta t P(P_d(R_w, R_S) - P < 0) = 9.398 \cdot 10^{-6}. \quad (5.39)$$

Zauważmy, że w omawianym przypadku wyznaczenie górnego ograniczenia prawdopodobieństwa awarii w przedziale czasu wymaga obliczenia prawdopodobieństwa obszaru awarii, a więc rozwiązania takiego samego problemu jak w przypadku analizy niezawodności niezależnej od czasu.

Przybliżenie asymptotyczne prawdopodobieństwa awarii uzyskano na podstawie wzoru:

$$P_f(\Delta t) \approx 1 - \mathbb{E}_{R_w, R_S}(\exp(-\lambda \Delta t P(P_d(r_w, r_S) - P < 0 | r_w, r_S))) = 9.346 \cdot 10^{-6}. \quad (5.40)$$

Jak widzimy, w przypadku przyjętej intensywności procesu $\lambda = 0.2$ uzyskane górne ograniczenie prawdopodobieństwa awarii jest dość bliskie przybliżeniu asymptotycznemu. Większą różnicę między tymi wartościami otrzymamy przyjmując wyższą wartość λ . W przypadku $\lambda = 12$ (obciążenie zmienia się średnio co miesiąc) ograniczenie górne wynosi $5.640 \cdot 10^{-4}$, a przybliżenie asymptotyczne $4.622 \cdot 10^{-4}$.

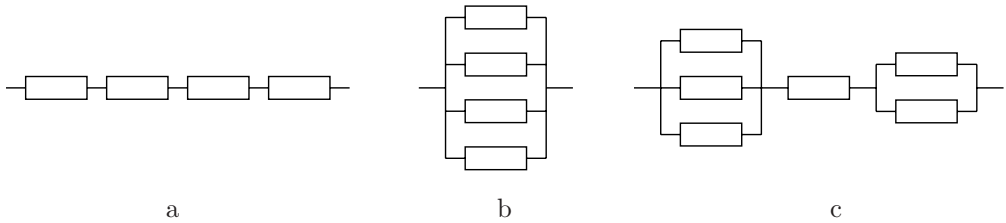
Niezawodność systemów konstrukcyjnych

6.1. Wstęp

W poprzednich rozdziałach omówiono sformułowanie problemu analizy niezawodności z wykorzystaniem pojedynczej funkcji granicznej. Uwzględnienie jednej funkcji granicznej oznacza, że jesteśmy w stanie oszacować niezawodność konstrukcji tylko ze względu na przekroczenie jednego stanu granicznego. Jednak nawet w przypadku prostych konstrukcji bezpieczne projektowanie wymaga przeanalizowania co najmniej kilku stanów granicznych. Zazwyczaj konieczne jest sprawdzenie stanów granicznych związanych z wytrzymałością, statecznością oraz stanów granicznych użytkowania (przekroczenia dopuszczalnych wartości przemieszczeń). W przypadku systemów konstrukcyjnych lub złożonych konstrukcji uwzględnia się stany graniczne określone dla elementów jak i dla całej konstrukcji. Tak, więc na podstawie sformułowań problemu analizy niezawodności przedstawionych w poprzednich rozdziałach możemy otrzymać zbiór prawdopodobieństw awarii. Taki zbiór trudno jest wykorzystać jako użyteczną miarę niezawodności całej konstrukcji. Zobaczymy dalej, że wybranie z tego zbioru maksimum bądź sumowanie jego elementów, pozwala jedynie na bardzo zgrubną ocenę prawdopodobieństwa awarii konstrukcji jako całości. Dokładniejsze oszacowanie prawdopodobieństwa awarii systemu konstrukcyjnego wymaga, bowiem uwzględnienia korelacji pomiędzy stanami granicznymi. W dalszej części tego rozdziału zostaną przedstawione zasady reprezentacji zależności pomiędzy stanami granicznymi za pomocą systemów logicznych. Modele te są podstawą metod, które służą do szacowania niezawodności systemów konstrukcyjnych na podstawie lokalnych stanów granicznych.

6.2. Modelowanie konstrukcji za pomocą systemów

Przedstawione w tym rozdziale metody analizy niezawodności wykorzystują reprezentację konstrukcji za pomocą systemu logicznego, składającego się ze skończonej liczby *elementów awarii*. Zakłada się, że każdemu elementowi awarii odpowiada jedna



RYСУNEK 6.1. System szeregowy (a); równoległy (b); połączenie systemów równoległych w system szeregowy (system mieszany) (c).

funkcja graniczna. Element awarii może znajdować się w stanie bezpiecznym lub w stanie awarii. Stan ten jest oczywiście określony przez wartość związanej z elementem funkcji granicznej.

Warto zwrócić uwagę, aby pomimo podobnych nazw nie utożsamiać elementów awarii z elementami konstrukcyjnymi. Na przykład, w przypadku pręta kratownicy można zdefiniować co najmniej dwie funkcje graniczne: stanu granicznego dopuszczalnych naprężeń i stanu granicznego stateczności pręta. A więc w tym przypadku jednemu elementowi konstrukcyjnemu odpowiadają dwa elementy awarii.

Poprzez łączenie elementów awarii definiuje się *system awarii*. Celem budowy takiego systemu jest stworzenie modelu, który reprezentuje zależność pomiędzy awarią całej konstrukcji a przekroczeniami stanów granicznych związanych z elementami awarii. Wyróżnia się dwa podstawowe rodzaje systemów awarii: *równoległe* i *szeregowy*. W przypadku skomplikowanych konstrukcji zazwyczaj istnieje potrzeba określenia *systemów mieszanych*, które są połączeniem systemów szeregowych i równoległych. Wygodną formą reprezentacji systemów awarii są schematy przedstawione na rysunku 6.1.

Diagram reprezentujący system szeregowy został przedstawiony na rysunku 6.1(a). Przyjmuje się, że system szeregowy jest w stanie awarii, jeżeli jeden z jego elementów jest w stanie awarii. Obszar awarii systemu szeregowego określa się jako

$$\Omega_f^S = \bigcup_{i=1}^m \Omega_{f_i}, \quad (6.1)$$

gdzie Ω_{f_i} są obszarami awarii m elementów, z których zbudowany jest system. Przykładem konstrukcji, której awarię można przedstawić za pomocą systemu szeregowego, jest kratownica statycznie wyznaczalna. Taka konstrukcja traci nośność po awarii jednego z elementów prętowych. System szeregowy, którego elementy odpowiadają stanom granicznym nośności prętów kratownicy będzie więc jej dobrym modelem.

Konstrukcje, które jako całość ulegają awarii dopiero po przekroczeniu kilku stanów granicznych, są modelowane za pomocą systemów równoległych. Diagram takiego

systemu przedstawia rysunek 6.1(b). Zakłada się, że system równoległy jest w stanie awarii, jeżeli wszystkie jego elementy znajdują się w stanie awarii. Obszar awarii systemu równoległego określa się, więc jako

$$\Omega_f^P = \bigcap_{i=1}^m \Omega_{f_i} . \quad (6.2)$$

Przykładem awarii, której analiza wymaga zastosowania systemu równoległego, jest utrata nośności kratownic statycznie niewyznaczalnych. Awaria konstrukcji tego typu jest konsekwencją zniszczenia co najmniej kilku elementów prętowych. W przypadku kratownic statycznie niewyznaczalnych zazwyczaj można zdefiniować kilka grup prętów, których awaria prowadzi do różnych form awarii konstrukcji. Modelem awarii grupy prętów jest system równoległy łączący odpowiadające im elementy awarii. Uwzględnienie postaci utraty nośności kratownicy, określonych przez wszystkie grupy prętów, wymaga połączenia systemów równoległych w system szeregowy.

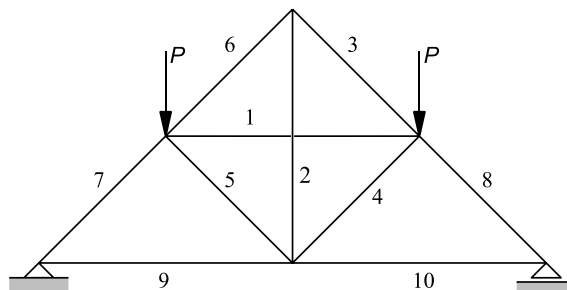
Połączenie systemów równoległych w system szeregowy tworzy system mieszany. Każdy system awarii można przedstawić w postaci systemu szeregowego, którego elementami są systemy równoległe. Przykładowy diagram systemu mieszanego przedstawia rysunek 6.1(c). Obszar awarii konstrukcji ze względu na przekroczenia wielu stanów granicznych można, więc przedstawić jako obszar awarii systemu mieszanego:

$$\Omega_f^M = \bigcup_{j=1}^{n_p} \Omega_{f_j}^P = \bigcup_{j=1}^{n_p} \bigcap_{i=1}^{m_j} \Omega_{f_{ij}} , \quad (6.3)$$

gdzie n_p jest liczbą systemów równoległych połączonych w system szeregowy, $\Omega_{f_j}^P$, $j = 1, \dots, n_p$, są obszarami awarii systemów równoległych, m_j są liczbami elementów w systemach równoległych, a $\Omega_{f_{ij}}$ oznacza obszar awarii i -tego elementu awarii w j -otym systemie równoległym.

System szeregowy ulega awarii, jeżeli awarii ulega, będący jego elementem, system równoległy. Zbiory elementów awarii, które tworzą takie systemy równoległe nazywa się *cięciami* (awaria systemu równoległego skutkuje "przecięciem" systemu szeregowego). Ciecie nazywamy *cięciem minimalnym*, jeżeli z niepełnej liczby jego elementów nie można utworzyć innego cięcia. Tak, więc system awarii można efektywnie przedstawić jako system szeregowy cięć minimalnych.

Przykład 6.1: Budowę systemu awarii przedstawimy na przykładzie kratownicy statycznie niewyznaczalnej, który został zaczerpnięty z pracy [112]. Rysunek 6.2 przedstawia kratownicę jednokrotnie statycznie niewyznaczalną, której pręty o numerach 1 i 2 przecinają się. Rozważana kratownica stanie się statycznie wyznaczalna po usunięciu (wystąpieniu



RYSUNEK 6.2. Kratownica statycznie niewyznaczalna.

awarii) jednego z prętów 1,2,3,4,5 lub 6. Jeżeli więc awarii ulegnie jeszcze jeden element z wymienionej grupy to nastąpi awaria całej konstrukcji. Utrata nośności całej konstrukcji ma również miejsce, jeżeli awarii ulega jeden z elementów o numerach 7,8,9 lub 10. System awarii w przypadku rozważanej konstrukcji można więc określić jako złożony z dziewiętnastu elementów system szeregowy, którego piętnaście elementów jest dwuelementowymi systemami równoległymi. System szeregowy składa się z następujących zbiorów elementów awarii: $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{1, 5\}$, $\{1, 6\}$, $\{2, 3\}$, $\{2, 4\}$, $\{2, 5\}$, $\{2, 6\}$, $\{3, 4\}$, $\{3, 5\}$, $\{3, 6\}$, $\{4, 5\}$, $\{4, 6\}$, $\{5, 6\}$, $\{7\}$, $\{8\}$, $\{9\}$ oraz $\{10\}$. Łatwo zauważyć, że wszystkie te systemy równoległe są cięciami minimalnymi.

System awarii w przykładzie 6.1 udało się określić dość łatwo. Jednak uwzględnia on tylko powstanie ustroju kinematycznie zmiennego, natomiast pominięto wpływ na postać awarii własności wytrzymałościowych prętów czy sposobu obciążenia. Na przykład, utrata nośności pręta 1 może doprowadzić do takiej redystrybucji sił wewnętrznych, że zostaną przekroczone dopuszczalne naprężenia w prętach 2 i 6. Tak, więc cała konstrukcja straci nośność w wyniku awarii jednego pręta. Aby uwzględnić taką postać zniszczenia to w systemie awarii, zdefiniowanym w przykładzie 6.1, należałoby określić minimalne cięcia składające się z elementu 1.

W ogólnym przypadku określenie systemu awarii wymaga uwzględnienia modeli wytrzymałościowych elementów, które definiują ich zachowanie po wystąpieniu awarii: całkowita lub częściowa utrata nośności. Konieczne jest również prześledzenie możliwych scenariuszy redystrybucji obciążeń po awarii kolejnych elementów. W przypadku złożonych systemów konstrukcyjnych brakuje metod pozwalających na wykonanie takich analiz. Tak, więc zastosowanie systemów awarii do oceny bezpieczeństwa rzeczywistych konstrukcji, przy obecnym stanie wiedzy, jest ograniczone. Pomimo to metody

oszacowania prawdopodobieństwa awarii systemów są dobrze opracowane.

6.3. Analiza niezawodności systemów

Prawdopodobieństwo awarii systemu jest prawdopodobieństwem zdarzenia losowego określonego przez obszar awarii systemu, zdefiniowany przez jeden ze wzorów: (6.1), (6.2) lub (6.3). Jest to oczywiście uogólnienie pojęcia prawdopodobieństwa awarii (2.6) na przypadek gdy obszar awarii jest określony przez więcej niż jedną funkcję graniczną.

Przyjmując, że $g_i(\mathbf{x})$, $i = 1, \dots, m$ są funkcjami granicznymi to *prawdopodobieństwo awarii systemu szeregowego* określa się za pomocą wyrażenia:

$$P_f^S = P[\Omega_f^S] = P\left[\bigcup_{i=1}^m \Omega_{f_i}\right] = P\left[\bigcup_{i=1}^m \{g_i(\mathbf{X}) \leq 0\}\right], \quad (6.4)$$

a *prawdopodobieństwo awarii systemu równoległego* jako:

$$P_f^P = P[\Omega_f^P] = P\left[\bigcap_{i=1}^m \Omega_{f_i}\right] = P\left[\bigcap_{i=1}^m \{g_i(\mathbf{X}) \leq 0\}\right]. \quad (6.5)$$

Prawdopodobieństwo awarii systemu mieszanego przedstawia się w następujący sposób:

$$P_f^M = P[\Omega_f^M] = P\left[\bigcup_{j=1}^{n_p} \bigcap_{i=1}^{m_j} \Omega_{f_{ij}}\right] = P\left[\bigcup_{j=1}^{n_p} \bigcap_{i=1}^{m_j} \{g_{ij}(\mathbf{X}) \leq 0\}\right], \quad (6.6)$$

gdzie $g_{ij}(\mathbf{X})$ jest funkcją graniczną j -tego elementu awarii w i -tym systemie równoległym.

Prawdopodobieństwo awarii systemu można, analogicznie do wzoru (2.7), wyrazić jako całkę z gęstości prawdopodobieństwa parametrów losowych konstrukcji $f_X(\mathbf{x})$ po obszarze awarii systemu Ω_f^{sys} (indeks *sys* zastępuje używane powyżej oznaczenia typów systemów):

$$P_f^{sys} = \int_{\Omega_f^{sys}} f_X(\mathbf{x}) d\mathbf{x} \quad (6.7)$$

oraz, konsekwentnie, jako wartość oczekiwaną funkcji charakterystycznej systemowego obszaru awarii $I_{\Omega_f^{sys}}(\mathbf{x})$:

$$P_f^{sys} = \mathbb{E}\left[I_{\Omega_f^{sys}}(\mathbf{X})\right] = \int_{\mathbf{R}^n} I_{\Omega_f^{sys}}(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x}. \quad (6.8)$$

Korzystając z prawdopodobieństwo awarii systemu P_f^{sys} , definiuje się *uogólniony wskaźnik niezawodności systemu*:

$$\beta^{sys} = -\Phi^{-1} \left(P_f^{sys} \right). \quad (6.9)$$

W dalszej części tego rozdziału przedstawione zostaną metody dzięki, którym możliwe jest oszacowanie prawdopodobieństwa awarii systemów.

6.4. Zastosowanie FORM w analizie niezawodności systemów

Przedmiotem rozważań w tym podrozdziale są systemy zbudowane z m elementów awarii. Każdemu elementowi awarii jest przypisana funkcja graniczna $g_i(\mathbf{x})$, $i = 1, 2, \dots, m$. Niech $\mathbf{U} = T(\mathbf{X})$ oznacza transformację podstawowych zmiennych losowych \mathbf{X} do gaussowskiej przestrzeni standardowej (por. 2.3.2). Funkcje graniczne w gaussowskiej przestrzeni standardowej są określone jako $h_i(\mathbf{u}) = g_i(T^{-1}(\mathbf{x}))$. Poniżej zostanie przedstawione zastosowanie metod pierwszego rzędu (por. 2.3.3) do oszacowania prawdopodobieństwa awarii systemów.

6.4.1. Systemy szeregowe

Przybliżenie metodą FORM prawdopodobieństwa awarii i -tego elementu awarii jest postaci:

$$\begin{aligned} P_{f_i} &= P [g_i(\mathbf{X}) \leq 0] = P [g_i(T^{-1}(\mathbf{U})) \leq 0] = P [h_i(\mathbf{U}) \leq 0] \\ &\approx P [\beta_i - \boldsymbol{\alpha}_i^T \mathbf{U} \leq 0] = \Phi(-\beta_i), \end{aligned} \quad (6.10)$$

gdzie β_i jest wskaźnikiem niezawodności wyznaczonym dla funkcji granicznej, opowiadającej i -temu elementowi awarii (tzw. *elementowy wskaźnik niezawodności*). W przypadku i -tej funkcji granicznej wektor $\boldsymbol{\alpha}_i$ jest określony wzorem (2.31). Na podstawie (6.1) prawdopodobieństwo awarii systemu szeregowego można wyrazić jako:

$$P_f^S = P \left[\bigcup_{i=1}^m \Omega_{f_i} \right] = P \left[\bigcup_{i=1}^m \{g_i(\mathbf{X}) \leq 0\} \right] = P \left[\bigcup_{i=1}^m \{h_i(\mathbf{U}) \leq 0\} \right]. \quad (6.11)$$

Przybliżenie FORM prawdopodobieństwa awarii systemu szeregowego uzyskuje się poprzez uwzględnienie linearyzacji funkcji granicznych w ich punktach projektowych:

$$P_f^S \approx P \left[\bigcup_{i=1}^m \{\beta_i - \boldsymbol{\alpha}_i^T \mathbf{U} \leq 0\} \right]. \quad (6.12)$$

Stosując prawo De Morgana wzór (6.12) przekształca się do następującej postaci

$$P_f^S \approx 1 - P \left[\bigcap_{i=1}^m \{ \boldsymbol{\alpha}_i^T \mathbf{U} < \beta_i \} \right]. \quad (6.13)$$

Zauważmy, że wyrażenie $P[\bigcap_{i=1}^m \{ \boldsymbol{\alpha}_i^T \mathbf{U} < \beta_i \}]$ jest wartością dystrybuanty łącznego rozkładu prawdopodobieństwa zmiennych losowych $\boldsymbol{\alpha}_i^T \mathbf{U}$, $i = 1, \dots, m$, w punkcie $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_m\}$. Ponieważ wektory $\boldsymbol{\alpha}_i$ mają jednostkową długość, a zmienna \mathbf{U} ma standardowy rozkład normalny to $\boldsymbol{\alpha}_i^T \mathbf{U}$, $i = 1, \dots, m$, są normalnymi zmiennymi losowymi o zerowej wartości oczekiwanej i jednostkowej wariancji. Ponadto współczynnik korelacji zmiennych $\boldsymbol{\alpha}_i^T \mathbf{U}$ i $\boldsymbol{\alpha}_j^T \mathbf{U}$ wynosi

$$\rho_{ij} = \boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_j. \quad (6.14)$$

Tak, więc wyrażenie (6.13) można przedstawić jako:

$$P_f^S \approx 1 - \Phi_m(\boldsymbol{\beta}, \mathbf{0}, \mathbf{R}), \quad (6.15)$$

gdzie Φ_m jest dystrybuantą m -wymiarowego rozkładu normalnego o wartości oczekiwanej $\mathbf{0}$ i macierzy kowariancji \mathbf{R} , której elementy określa (6.14).

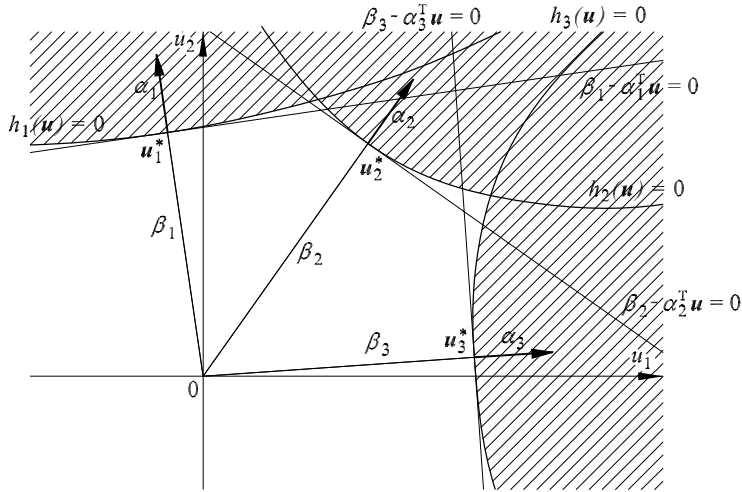
Uogólniony wskaźnik niezawodności systemu szeregowego określa się na podstawie oszacowania prawdopodobieństwa awarii (6.15), za pomocą wyrażenia

$$\beta^S = -\Phi^{-1}(1 - \Phi_m(\boldsymbol{\beta}, \mathbf{0}, \mathbf{R})). \quad (6.16)$$

Ideę zastosowania metody FORM do oszacowania prawdopodobieństwa awarii systemu szeregowego ilustruje rysunek 6.3. Analizowany system awarii składa się z trzech elementów, którym przypisano funkcje graniczne $g_i(T^{-1}(\mathbf{U})) = h_i(\mathbf{U})$, $i = 1, 2, 3$. Obszar awarii systemu, będący sumą obszarów awarii elementów, został zakreskowany. Na rysunku przedstawiono elementowe wskaźniki niezawodności β_i , $i = 1, 2, 3$, oraz powierzchnie graniczne, zlinearyzowane we właściwych punktach projektowych \mathbf{u}_i^* , $i = 1, 2, 3$. Błąd oszacowania prawdopodobieństwa awarii systemu szeregowego metodą FORM (6.11) zależy oczywiście od dokładności przybliżenia powierzchni granicznych $h_i(\mathbf{u} = 0)$ przez hiperpłaszczyzny $\beta_i - \boldsymbol{\alpha}_i^T \mathbf{u} = 0$.

Na podstawie wzoru (6.14) oraz własności wektorów $\boldsymbol{\alpha}_i$ można zaproponować interpretację współczynnika ρ_{ij} jako kosinusa kąta θ zawartego między zlinearyzowanymi powierzchniami granicznymi $l_i(\mathbf{u}) = 0$ i $l_j(\mathbf{u}) = 0$:

$$\cos \theta = \boldsymbol{\alpha}_i^T \boldsymbol{\alpha}_j = \rho_{ij}. \quad (6.17)$$



RYСУNEK 6.3. Zastosowanie metody FORM w analizie niezawodności systemów szeregowych.

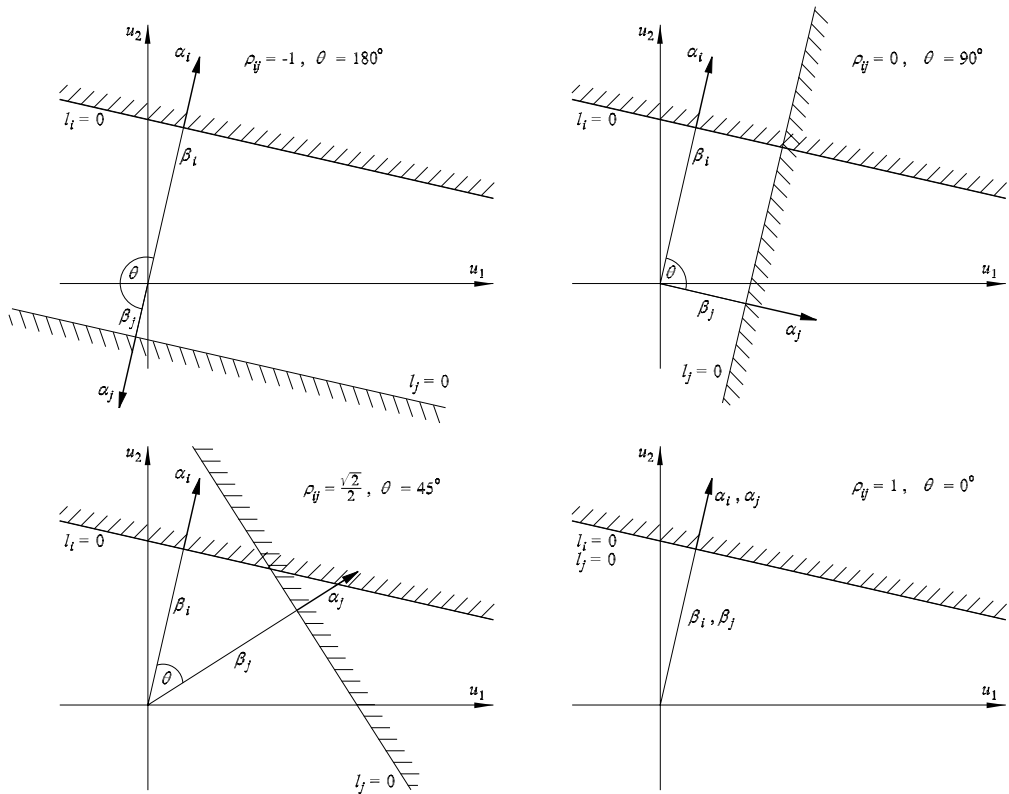
Tak, więc na podstawie współczynnika korelacji możliwa jest ocena kształtu obszaru awarii systemu szeregowego. Na rysunku 6.4 przedstawiono wzajemne położenie dwóch liniowych powierzchni granicznych i odpowiadającym im obszarów awarii dla ρ_{12} o wartościach -1.0 , 0.0 , $\sqrt{0.5}$ oraz 1.0 . Przyjmując $\beta_i = \beta_j = 3.0$ obliczono wartości systemowego wskaźnika niezawodności dla przyjętych wartości współczynnika korelacji. Wynoszą one odpowiednio 2.7821, 2.7824, 2.7921 oraz 3.0.

Na rysunku 6.5 przedstawiono wykres uogólnionego wskaźnika niezawodności dwuelementowego systemu szeregowego w funkcji współczynnika korelacji. Zauważmy, że założenie pełnej korelacji między elementami szeregowego systemu awarii, jeżeli tak nie jest w danym przypadku, prowadzi do zaniżenia oszacowania prawdopodobieństwa awarii systemu.

6.4.2. Systemy równoległe

Na podstawie definicji obszaru awarii systemu równoległego (6.2) jego prawdopodobieństwo określa się wzorem:

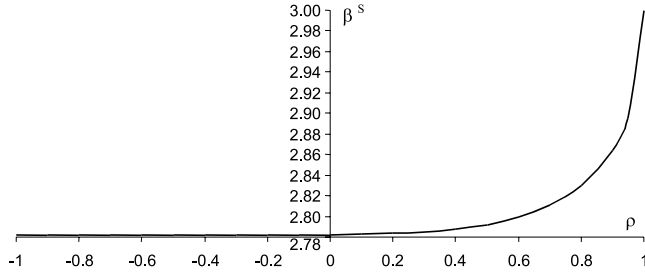
$$P_f^P = P \left[\bigcap_{i=1}^m \Omega_{f_i} \right] = P \left[\bigcap_{i=1}^m \{g_i(\mathbf{X}) \leq 0\} \right] = P \left[\bigcap_{i=1}^m \{h_i(\mathbf{U}) \leq 0\} \right]. \quad (6.18)$$

RYSUNEK 6.4. Powierzchnie awarii dla różnych wartości ρ_{ij} .

Schemat zastosowania metody FORM do oszacowania prawdopodobieństwa awarii systemu równoległego przedstawiono na rysunku 6.6. Po transformacji do gaussowskiej przestrzeni standardowej, wewnątrz obszaru awarii systemu równoległego Δ_f^P można wyznaczyć punkt położony najbliżej środka układu współrzędnych. Punkt ten to tzw. *wspólny punkt projektowy*, który precyzyjniej definiuje się jako:

$$\bar{\mathbf{u}}^* = \arg \min_{\mathbf{u} \in \Delta_f^P} \|\mathbf{u}\| \quad (6.19)$$

(por. rys. 6.6). Znalezienie wspólnego punktu projektowego wymaga rozwiązania odpowiedniego zadania optymalizacji z ograniczeniami. W tym celu można zastosować ogólny algorytm optymalizacyjny jak NLPQL [106] lub specjalnie opracowany dla tego



RYSUNEK 6.5. $\beta^S(\rho) = -\Phi^{-1}(1 - \Phi_2(\{3.0, 3.0\}, \mathbf{0}, \rho))$.

zagadnienia algorytm JOINT3 [32].

Funkcje graniczne, których wartość w punkcie $\bar{\mathbf{u}}^*$ jest równa zero są określane jako *aktywne funkcje graniczne*. Funkcje te mogą być zlinearyzowane w punkcie $\bar{\mathbf{u}}^*$ w następujący sposób:

$$l_i(\mathbf{u}) = -\bar{\alpha}_i^T \mathbf{u} + \beta_i^J, \quad i = 1, 2, \dots, m_A, \quad (6.20)$$

gdzie m_A jest liczbą aktywnych funkcji granicznych,

$$\bar{\alpha}_i = -\frac{\nabla h_i(\mathbf{u})}{\|\nabla h_i(\mathbf{u})\|} \Big|_{\mathbf{u}=\bar{\mathbf{u}}^*} \quad (6.21)$$

oraz

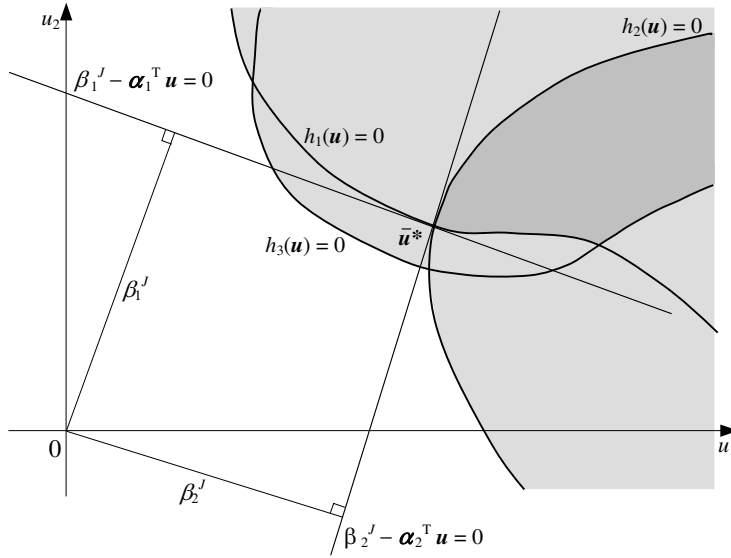
$$\beta_i^J = \bar{\alpha}_i^T \bar{\mathbf{u}}^*. \quad (6.22)$$

Zwróćmy uwagę, że we wzorze (6.20) funkcje graniczne nie są linearyzowane w ich punktach projektowych, tak jak w przypadku systemów szeregowych, ale we wspólnym punkcie projektowym. Na podstawie aproksymacji funkcji granicznych, określonych przez (6.20) oraz z wyrażenia (6.18), oszacowanie prawdopodobieństwa awarii systemu równoległego przybiera postać:

$$P_f^P \approx P \left[\bigcap_{i=1}^{m_A} \{-\bar{\alpha}_i^T \mathbf{U} + \beta_i^J \leq 0\} \right] = P \left[\bigcap_{i=1}^{m_A} \{\bar{\alpha}_i^T \mathbf{U} \geq \beta_i^J\} \right] = \Phi_{m_A}(-\boldsymbol{\beta}^J, \mathbf{0}, \bar{\mathbf{R}}), \quad (6.23)$$

gdzie $\boldsymbol{\beta}^J = \{\beta_1^J, \beta_2^J, \dots, \beta_{m_A}^J\}$, a $\bar{\mathbf{R}}$ jest macierzą kowariancji o elementach

$$\bar{\rho}_{ij} = \bar{\alpha}_i^T \bar{\alpha}_j. \quad (6.24)$$

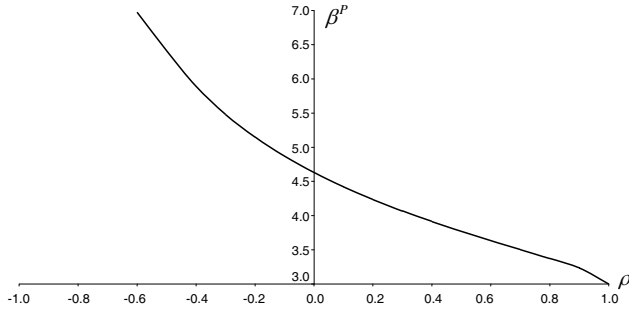


RYSUNEK 6.6. Zastosowanie metody FORM do oszacowania prawdopodobieństwa awarii systemu równoległego.

Oczywiście mając zdefiniowane prawdopodobieństwo awarii systemu równoległego można wprowadzić *uogólniony wskaźnik niezawodności systemu równoległego* w następującej postaci:

$$\beta^P = -\Phi^{-1} \left(\Phi_{m_A} \left(-\beta^J, \mathbf{0}, \mathbf{R} \right) \right). \quad (6.25)$$

Podobnie jak w przypadku systemów szeregowych współczynniki $\bar{\rho}_{ij}$ określają kształt zlinearyzowanego obszaru awarii systemu równoległego. Ponadto wartość uogólnionego wskaźnika niezawodności systemu równoległego, który został otrzymany za pomocą FORM, silnie zależy od współczynników korelacji. Poniżej przedstawiamy analizę tego zagadnienia podobną do tej, jaką wykonano na stronie 118 w przypadku systemów szeregowych. Rysunek 6.7 przedstawia wykres β^P dwuelementowego systemu równoległego w funkcji współczynnika korelacji $\bar{\rho}_{ij}$ (przyjęto $\beta_1^j = \beta_2^j = 3.0$). Jak można zauważyć, β^P przyjmuje wartości z przedziału od 3.0 do ∞ dla - odpowiednio - w pełni dodatnio i w pełni ujemnie skorelowanych funkcji granicznych.



RYSUNEK 6.7. $\beta^P(\rho) = -\Phi^{-1}(1 - \Phi_2(\{-3.0, -3.0\}, \mathbf{0}, \bar{\rho}))$.

6.4.3. Systemy mieszane

Oszacowanie metodą FORM prawdopodobieństwa awarii systemu mieszanego (6.6) otrzymuje się na podstawie jego przedstawienia jako systemu szeregowego złożonego z systemów równoległych. Oszacowanie to można wyrazić - podobnie do (6.15) - w następujący sposób:

$$P_f^M \approx 1 - \Phi_{n_P}(\beta^P, \mathbf{0}, \mathbf{R}^P), \quad (6.26)$$

gdzie $\beta^P = \{\beta_1^P, \beta_2^P, \dots, \beta_{n_P}^P\}$ jest wektorem uogólnionych wskaźników niezawodności n_P połączonych szeregowo systemów równoległych (6.25) a \mathbf{R}^P jest macierzą przybliżonych współczynników korelacji pomiędzy tymi systemami równoległymi.

Wyznaczenie współczynników ρ_{ij}^P , $i, j = 1, 2, \dots, n_P$, będących elementami \mathbf{R}^P , wymaga przybliżenia systemów równoległych za pomocą liniowej funkcji granicznej

$$l_i^P(\mathbf{U}) = -\alpha_i^{P^T} \mathbf{U} + \beta_i^P, \quad i = 1, 2, \dots, n_P, \quad (6.27)$$

gdzie α_i^P jest wektorem znormalizowanych wrażliwości wskaźników β_i^P ze względu na zmiany wspólnego punktu projektowego systemu równoległego. W pracy [33] zaproponowano następujące oszacowanie współczynników α_i^P :

$$\alpha_i^P = \frac{\mathbf{a}_i^P}{\|\mathbf{a}_i^P\|}, \quad (6.28)$$

gdzie składowe wektora \mathbf{a}_i^P mają następującą postać:

$$(a_i^P)_j = \frac{1}{\varphi(-\beta_i^P)} \sum_{k=1}^{m_{A_i}} \left\{ \left(\alpha_{kj}^i + \frac{d\alpha_k^{i,T}}{d\bar{u}_j^{*i}} \bar{\mathbf{u}}^{*i} \right) \frac{\partial \Phi_{m_{A_i}}(-\beta^{J^i}, \mathbf{R}^i)}{\partial \beta_k^{J^i}} + 2 \sum_{l=1}^{k-1} \frac{\partial \Phi_{m_{A_i}}(-\beta^{J^i}, \mathbf{R}^i)}{\partial \rho_{kl}^i} \frac{d\rho_{kl}^i}{d\bar{u}_j^{*i}} \right\}. \quad (6.29)$$

We wzorze (6.29) gęstość standardowego rozkładu normalnego oznaczono przez φ , a m_A jest liczbą aktywnych funkcji granicznych i -tego systemu równoległego. Wyrażenia $d\rho_{kl}/d\bar{u}_j^*$ w (6.29) (dla uproszczenia pominięto indeksy i) mogą być wyznaczone na podstawie wzoru:

$$\frac{d\rho_{kl}}{d\bar{u}_j^*} = \frac{d\alpha_k^T}{d\bar{u}_j^*} \alpha_l + \alpha_k^T \frac{d\alpha_l}{d\bar{u}_j^*} \quad (6.30)$$

gdzie

$$\frac{d\alpha_k}{d\bar{u}_j^*} = \left(\frac{-\mathbf{I}}{\|\nabla h_k\|} + \frac{\nabla h_k \nabla h_k^T}{\|\nabla h_k\|^3} \right) \frac{\partial \nabla h_k}{\partial \bar{u}_j^*}, \quad (6.31)$$

w którym \mathbf{I} jest macierzą jednostkową. Ostatecznie otrzymujemy następujące wyrażenie, które pozwala oszacować elementy macierzy współczynników korelacji między systemami równoległymi \mathbf{R}^P :

$$\rho_{ij}^P = \alpha_i^{P,T} \alpha_j^P. \quad (6.32)$$

Tak, więc wszystkie wielkości potrzebne do wyznaczenia oszacowania (6.26) zostały określone.

6.5. Przedziały prawdopodobieństwa awarii systemu

Przedmiotem tego podrozdziału jest omówienie algorytmów pozwalających na oszacowanie ograniczeń prawdopodobieństwa awarii systemu. Posługiwanie się tymi metodami nie wymaga obliczania dystrybuanty rozkładu normalnego określonego w dużej liczbie wymiarów, które występują w przybliżeniach (6.15) i (6.23).

Zauważmy, że prawdopodobieństwo awarii systemu szeregowego można przedstawić

za pomocą następującej sumy:

$$P \left[\bigcup_{i=1}^m \Omega_{f_i} \right] = P [\Omega_{f_1}] + P [\Omega_{f_2} \cap \overline{\Omega_{f_1}}] + P [\Omega_{f_3} \cap \overline{\Omega_{f_2} \cup \Omega_{f_1}}] + \dots + P \left[\Omega_{f_m} \cap \overline{\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}} \right], \quad (6.33)$$

gdzie $\overline{\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}}$ jest zdarzeniem przeciwnym do zdarzenia $\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}$. Korzystając z równości $A = (A \cap B) \cup (A \cap \bar{B})$, gdzie A i B są dowolnymi zdarzeniami losowymi, uzyskuje się poniższe przekształcenie prawdopodobieństw w sumie (6.33):

$$P \left[\Omega_{f_m} \cap \overline{\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}} \right] = P [\Omega_{f_m}] - P \left[\Omega_{f_m} \cap (\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}) \right]. \quad (6.34)$$

Podstawiając (6.34) do (6.33) otrzymujemy prawdopodobieństwo awarii systemu szeregowego w następującej postaci:

$$\begin{aligned} P \left[\bigcup_{i=1}^m \Omega_{f_i} \right] &= P [\Omega_{f_1}] \\ &+ P [\Omega_{f_2}] - P [\Omega_{f_2} \cap \Omega_{f_1}] \\ &+ P [\Omega_{f_3}] - P [\Omega_{f_3} \cap (\Omega_{f_2} \cup \Omega_{f_1})] \\ &\vdots \\ &+ P [\Omega_{f_m}] - P \left[\Omega_{f_m} \cap (\Omega_{f_{m-1}} \cup \dots \cup \Omega_{f_1}) \right]. \end{aligned} \quad (6.35)$$

Uwzględnienie we wzorze (6.35) jedynie *wyrażeń pierwszego rzędu* $P [\Omega_{f_i}]$ prowadzi do górnego oszacowania prawdopodobieństwa awarii. Oszacowaniem dolnym, które wykorzystuje jedynie wyrażenia pierwszego rzędu, jest oczywiście maksymalne prawdopodobieństwo awarii elementu. *Oszacowanie proste* (lub pierwszego rzędu) przedziału zawierającego prawdopodobieństwo awarii systemu szeregowego (ang. *simple bounds*) określa się, więc jako:

$$\max \{ P [\Omega_{f_1}], P [\Omega_{f_2}], \dots, P [\Omega_{f_m}] \} \leq P \left[\bigcup_{i=1}^m \Omega_{f_i} \right] \leq \sum_{i=1}^m P [\Omega_{f_i}]. \quad (6.36)$$

Warto zauważyć, że dolne ograniczenie jest dokładną wartością P_f^S , jeżeli wszystkie elementy systemu szeregowego są całkowicie skorelowane.

Analogicznie ograniczenia proste można wyznaczyć w przypadku wskaźnika niezawodności systemu szeregowego (6.16). Korzystając ze wskaźników niezawodności elementów systemu β_i , $i = 1, 2, \dots, m$, przedstawimy je w następujący sposób

$$-\Phi^{-1}\left(\sum_{i=1}^m \Phi(-\beta_i)\right) \leq \beta^S \leq \min\{\beta_1, \beta_2, \dots, \beta_m\}. \quad (6.37)$$

Niestety, jeżeli prawdopodobieństwo awarii jednego z elementów nie jest znacząco większe od pozostałych to przedział określony ograniczeniami prostymi jest zazwyczaj bardzo szeroki. W takich przypadkach konieczne jest zastosowanie ograniczeń, które są dokładniejsze, ale również bardziej skomplikowane obliczeniowo.

Dokładniejsze ograniczenia prawdopodobieństwa awarii systemu szeregowego możemy uzyskać uwzględniając człony wyrażenia (6.35), które pominięto w ograniczeniach prostych. Zauważmy, że

$$\Omega_{f_k} \cap \left(\Omega_{f_{k-1}} \cup \dots \cup \Omega_{f_1}\right) = \bigcup_{i=1}^{k-1} (\Omega_{f_k} \cap \Omega_{f_i}). \quad (6.38)$$

Podstawiając wyrażenia $(\Omega_{f_k} \cap \Omega_{f_i})$ do wzoru (6.35) w miejsce Ω_{f_i} otrzymujemy prawdopodobieństwo zdarzeń określonych przez (6.38):

$$\begin{aligned} P\left[\bigcup_{i=1}^{k-1} (\Omega_{f_k} \cap \Omega_{f_i})\right] &= P[\Omega_{f_k} \cap \Omega_{f_1}] \\ &+ P[\Omega_{f_k} \cap \Omega_{f_2}] - P[(\Omega_{f_k} \cap \Omega_{f_2}) \cap (\Omega_{f_k} \cap \Omega_{f_1})] \\ &+ P[\Omega_{f_k} \cap \Omega_{f_3}] \\ &- P[(\Omega_{f_k} \cap \Omega_{f_3}) \cap ((\Omega_{f_k} \cap \Omega_{f_2}) \cup (\Omega_{f_k} \cap \Omega_{f_1}))] \\ &\vdots \\ &+ P[\Omega_{f_k} \cap \Omega_{f_{k-1}}] \\ &- P\left[(\Omega_{f_k} \cap \Omega_{f_{k-1}}) \cap \left((\Omega_{f_k} \cap \Omega_{f_{k-2}}) \cup \dots \cup (\Omega_{f_k} \cap \Omega_{f_1})\right)\right]. \end{aligned} \quad (6.39)$$

Uwzględniając jedynie prawdopodobieństwa *wyrażeń drugiego rzędu* $\Omega_{f_k} \cap \Omega_{f_i}$, na podstawie wzoru (6.39), otrzymuje się następujące ograniczenie górne:

$$P\left[\Omega_{f_k} \cap \left(\Omega_{f_{k-1}} \cup \dots \cup \Omega_{f_1}\right)\right] \leq \sum_{i=1}^{k-1} P[\Omega_{f_k} \cap \Omega_{f_i}]. \quad (6.40)$$

Dolne ograniczenie drugiego rzędu prawdopodobieństwa awarii systemu szeregowego można uzyskać podstawiając (6.40) do prawej strony (6.35). W praktyce często używa się nieco zmodyfikowanej postaci ograniczenia dolnego drugiego rzędu. Jest to tak zwane *dolne ograniczenie Ditlevsena*:

$$P\left[\bigcup_{i=1}^m \Omega_{f_i}\right] \geq P[\Omega_{f_1}] + \sum_{i=2}^m \max\left\{0, P[\Omega_{f_i}] - \sum_{j=1}^{i-1} P[\Omega_{f_i} \cap \Omega_{f_j}]\right\}. \quad (6.41)$$

Górne ograniczenie drugiego rzędu prawdopodobieństwa awarii systemu szeregowego wyznacza się podstawiając do wzoru (6.35) ograniczenia dolne członów, które są odejmowane. Ograniczenie dolne prawdopodobieństw zbiorów w (6.38) można określić tak jak ograniczenie dolne pierwszego rzędu - w postaci maksymalnej wartości prawdopodobieństwa części wspólnych zbiorów:

$$P\left[\Omega_{f_k} \cap \left(\Omega_{f_{k-1}} \cup \dots \cup \Omega_{f_1}\right)\right] \geq \max\left\{P\left[\Omega_{f_k} \cap \Omega_{f_{k-1}}\right], \dots, P\left[\Omega_{f_k} \cap \Omega_{f_1}\right]\right\}. \quad (6.42)$$

Podstawiając (6.42) do (6.35) otrzymujemy górne ograniczenie drugiego rzędu prawdopodobieństwa awarii systemu szeregowego (górne ograniczenie Ditlevsena):

$$P\left[\bigcup_{i=1}^m \Omega_{f_i}\right] \leq \sum_{i=1}^m P[\Omega_{f_i}] - \sum_{i=2}^m \max\left\{P\left[\Omega_{f_i} \cap \Omega_{f_{i-1}}\right], \dots, P\left[\Omega_{f_i} \cap \Omega_{f_1}\right]\right\}. \quad (6.43)$$

Podobnie uzyskuje się ograniczenia drugiego rzędu wskaźnika niezawodności systemu szeregowego. Korzystając z przybliżeń FORM prawdopodobieństwa awarii systemów równoległych (por. 6.4.2) można je przedstawić w następującej postaci:

$$\beta^S \geq -\Phi^{-1}\left(\sum_{i=1}^m \Phi(-\beta_i) - \sum_{i=2}^m \max\left\{\Phi_2\left(\{-\beta_i^J, -\beta_{i-1}^J\}, \mathbf{0}, \bar{\rho}_{i(i-1)}\right), \dots, \Phi_2\left(\{-\beta_i^J, -\beta_1^J\}, \mathbf{0}, \bar{\rho}_{i1}\right)\right\}\right), \quad (6.44)$$

$$\beta^S \leq -\Phi^{-1}\left(\Phi(-\beta_1) + \sum_{i=2}^m \max\left\{0, \Phi(-\beta_i) - \sum_{j=1}^{i-1} \Phi_2\left(\{-\beta_i^J, -\beta_j^J\}, \mathbf{0}, \bar{\rho}_{ij}\right)\right\}\right). \quad (6.45)$$

Ograniczenia Ditlevsena zazwyczaj są dokładniejsze niż ograniczenia proste, jednak wymagają bardziej złożonych obliczeń. Do ich wyznaczenia konieczne jest, bowiem oszacowanie prawdopodobieństw części wspólnych obszarów awarii elementów.

Wartość przedstawionych ograniczeń drugiego rzędu jest zależna od numeracji elementów awarii. Zaleca się, aby elementy awarii były uporządkowane zgodnie z malejącym prawdopodobieństwem awarii. Tak, więc uzyskanie dobrego ograniczenia wymaga spełnienia warunku: $P[\Omega_{f_1}] \geq [\Omega_{f_1}] \geq \dots \geq [\Omega_{f_m}]$.

Ograniczenia prawdopodobieństwa awarii pierwszego i drugiego rzędu można również określić w przypadku systemów równoległych. Zauważając, że obszar awarii każdego z elementów systemu równoległego zawiera obszar awarii systemu definiuje się poniższe ograniczenia proste (pierwszego rzędu):

$$0 \leq P \left[\bigcap_{i=1}^m \Omega_{f_i} \right] \leq \max \{ P[\Omega_{f_1}], P[\Omega_{f_2}], \dots, P[\Omega_{f_m}] \}. \quad (6.46)$$

Jeżeli prawdopodobieństwa awarii elementów systemu równoległego spełniają zależność

$$P \left[\bigcap_{i=1}^j \Omega_{f_i} \right] \geq P \left[\bigcap_{i=1}^{j-1} \Omega_{f_i} \right] P[\Omega_{f_j}], \quad (6.47)$$

to prawdziwe jest dokładniejsze ograniczenie dolne:

$$\begin{aligned} P \left[\bigcap_{i=1}^m \Omega_{f_i} \right] &= P[\Omega_{f_1}] \frac{P[\Omega_{f_1} \cap \Omega_{f_2}]}{P[\Omega_{f_1}]} \dots \frac{P[\Omega_{f_1} \cap \Omega_{f_2} \cap \dots \cap \Omega_{f_m}]}{P[\Omega_{f_1} \cap \Omega_{f_2} \cap \dots \cap \Omega_{f_{m-1}}]} \\ &\geq \prod_{i=1}^m P[\Omega_{f_i}]. \end{aligned} \quad (6.48)$$

Zwróćmy jeszcze uwagę, że ograniczenie (6.48) jest równe dokładnej wartości prawdopodobieństwa awarii systemu równoległego, jeżeli elementy awarii nie są skorelowane.

Ograniczeniom prawdopodobieństwa awarii (6.46) i (6.48) odpowiadają proste ograniczenia wskaźnika niezawodności systemu równoległego [112]:

$$\max \{ \beta_1^J, \beta_2^J, \dots, \beta_{m_A}^J \} \leq \beta^P \leq -\Phi^{-1} \left(\prod_{i=1}^{m_A} \Phi(-\beta_i^J) \right), \quad (6.49)$$

gdzie wskaźniki niezawodności β_i^J są określone przez (6.22).

W większości przypadków ograniczenia proste prawdopodobieństwa awarii systemów równoległych, tak jak w przypadku systemów szeregowych, są mało dokładne. Dokładniejsze ograniczenie górne drugiego rzędu zaproponowano w pracy [81]. Zauważmy, że obszar awarii systemu równoległego zawiera się w części wspólnej obszarów

awarii dowolnych dwóch elementów systemu, a więc prawdziwa jest nierówność:

$$P \left[\bigcap_{i=1}^m \Omega_{f_i} \right] \leq \min_{i,j} \left\{ P \left[\Omega_{f_i} \cap \Omega_{f_j} \right] \right\}. \quad (6.50)$$

Na podstawie aproksymacji FORM definiuje się - odpowiadające (6.50) - ograniczenie wskaźnika niezawodności:

$$\beta^P \geq -\Phi^{-1} \left(\max_{i,j} \left\{ \Phi_2(-\beta_i^J, -\beta_j^J, \mathbf{0}, \bar{\rho}_{ij}) \right\} \right), \quad (6.51)$$

gdzie β_i^J i $\bar{\rho}_{ij}$ wyznaczone są we wspólnym punkcie projektowym obszaru $\Omega_{f_i} \cap \Omega_{f_j}$.

6.6. Metody Monte Carlo w analizie niezawodności systemów

Oszacowanie prawdopodobieństwa awarii systemu za pomocą klasycznej metody Monte Carlo (por. 2.3.6) można otrzymać stosując estymator (2.40) do wyznaczenia całki (6.8). Zastosowanie metody importance sampling (por. 2.3.7) również opiera się na wyznaczeniu przybliżenia całki (6.8). Jednak w porównaniu do problemów z jedną funkcją graniczną, oszacowanie prawdopodobieństwa awarii systemu zazwyczaj wymaga określenia bardziej skomplikowanego rozkładu, z którego generowana jest próba losowa importance sampling.

Przedstawione w rozdziale 2 metody analizy niezawodności wykorzystują założenie, że w obszarze awarii można określić punkt projektowy, w którym gęstość rozkładu prawdopodobieństwa ma maksimum globalne o wartości znacznie większej od ewentualnych maksimów lokalnych. W przypadku analizy systemów przyjęcie takiego założenia może prowadzić do dużych błędów oszacowania prawdopodobieństwa awarii. Skomplikowany kształt systemowej powierzchni granicznej powoduje, że funkcja gęstości prawdopodobieństwa może mieć w systemowym obszarze awarii kilka istotnych maksimów lokalnych. W przypadku systemów szeregowych, określonych w gaussowskiej przestrzeni standardowej, lokalne maksima gęstości prawdopodobieństwa pokrywają się z punktami projektowymi elementowych funkcji granicznych. Warto zwrócić uwagę, że na wykorzystaniu tego faktu opiera się aproksymacja FORM prawdopodobieństwa awarii systemu szeregowego (por. 6.4.1). W przypadku problemu analizy niezawodności określonego jedną funkcją graniczną, w podrozdziale 2.3.7, zaproponowano generowanie próby losowej z rozkładu prawdopodobieństwa o gęstości (2.59). Funkcja gęstości prawdopodobieństwa (2.59) ma jedno maksimum, które pokrywa się z punktem projektowym. Zastosowanie takiego *jednomodowego* rozkładu prawdopodobieństwa (ang. *unimodal distribution*) zapewnia uzyskanie próby losowej skupionej w

obszarze, który jest najbardziej istotny dla oszacowania prawdopodobieństwa awarii. Uzyskanie efektywnego oszacowania prawdopodobieństwa awarii systemu szeregowe- go najczęściej wymaga wygenerowania próby losowej, której elementy są rozłożone w pobliżu lokalnych maksimów gęstości prawdopodobieństwa. Uzyskanie takiej próby z rozkładu jednomodowego jest niemożliwe. W tym przypadku konieczne jest zastosowa- nie *rozkładu wielomodowego* (ang. *multimodal distribution*), którego gęstość ma kilka istotnych maksimów lokalnych.

W analizie niezawodności systemów za pomocą importance sampling stosuje się *rozkłady mieszane* (ang. *mixed distribution*) [53]. Gęstość prawdopodobieństwa rozkła- du mieszanego określa się jako kombinację liniową gęstości rozkładów składowych q_i , $i = 1, \dots, k$:

$$q(\mathbf{x}) = \sum_{i=1}^k \lambda_i q_i(\mathbf{x}), \quad (6.52)$$

gdzie $\sum_{i=1}^k \lambda_i = 1$ oraz $\lambda_i \geq 0$. W przypadku problemu analizy niezawodności systemu szeregowego, sformułowanego w gaussowskiej przestrzeni standardowej, dobre rezultaty daje zastosowanie kombinacji liniowej rozkładów (2.59) określonych dla elementowych funkcji awarii. W tym podejściu przyjmuje się, że współczynniki λ są funkcjami indek- sów niezawodności elementów awarii [76].

Metoda wzajemnej entropii w analizie niezawodności konstrukcji

7.1. Adaptacyjne metody symulacyjne

Jak wspomniano w podrozdziale 2.3.7 efektywność metody importance sampling zależy od doboru rozkładu prawdopodobieństwa, z którego generowana jest próba losowa. Typowa implementacja importance sampling (2.59) wymaga określenia punktu projektowego. Najbardziej efektywnie punkt projektowy można wyznaczyć za pomocą gradientowych metod optymalizacji. Jednak zastosowanie tego podejścia wymaga różniczkowalności funkcji granicznej. Jeżeli warunek ten nie jest spełniony należy zastosować inne metody; jedną z alternatyw jest wykorzystanie adaptacyjnych metod symulacyjnych. Przykład takiego algorytmu przedstawiono w podrozdziale 2.3.8.

Adaptacyjne metody symulacyjne cieszą się zainteresowaniem ze względu na możliwość zastosowania w przypadku nieregularnych funkcji granicznych, przy niższych nakładach obliczeniowych w porównaniu do klasycznej metody Monte Carlo. Jak wspomniano we wstępie do tej rozprawy, w literaturze można znaleźć propozycje takich metod, które mimo iż sprawdzają się w pewnych przypadkach nie są opracowane teoretycznie, a zwłaszcza brakuje dowodów ich zbieżności. Przedmiotem tego rozdziału jest omówienie wykorzystania w analizie niezawodności konstrukcji *metody wzajemnej entropii* (ang. *cross-entropy method*) [21]. W kolejnym rozdziale zostaną przedstawione *metody Markov chain Monte Carlo* [121]. Te dwie adaptacyjne metody Monte Carlo znalazły szerokie zastosowanie w różnych dziedzinach wykorzystujących techniki symulacyjne. Obie mają dobrze opracowane podstawy teoretyczne, w tym również zbadano warunki ich zbieżności.

Metoda wzajemnej entropii jest dość nowym podejściem pozwalającym rozwiązywać zagadnienia, które można sformułować w ramach optymalizacji kombinatorycznej bądź symulacji tzw. rzadkich zdarzeń (ang. *rare events*). Oszacowanie prawdopodobieństwa awarii można zaliczyć do drugiej z tych grup.

Omawiany algorytm jest typowym przykładem adaptacyjnych metod symulacyjnych. Metody te wykorzystują do generowania próby losowej rozkład prawdopodobieństwa, który jest dostosowywany do danego zadania w trakcie wykonywania algorytmu. Zazwyczaj dobór rozkładu prawdopodobieństwa wykorzystywanego w celu symulacji jest ograniczony do pewnej rodziny parametrycznej. W takim przypadku dostosowanie rozkładu polega na określeniu odpowiednich wartości parametrów. Kryterium wyboru rodziny rozkładów prawdopodobieństwa może być łatwość generowania próby losowej. Częstość podejściem jest wykorzystanie tej samej rodziny, do której należą opisujące zadanie zmienne losowe. Na przykład, w przypadku problemu analizy niezawodności sformułowanego w gaussowskiej przestrzeni standardowej jest to wielowymiarowy rozkład normalny.

Istotą metody wzajemnej entropii jest dobór parametrów rozkładu, z którego generowana jest próba losowa, poprzez minimalizację wzajemnej entropii jego i rozkładu minimalizującego wariancję estymatora importance sampling (2.54). Jak to zostanie pokazane w dalszej części rozdziału wykorzystanie takiego kryterium, w pewnych przypadkach, pozwala szczególnie łatwo określić parametry rozkładu próby losowej.

Metody Markov chain Monte Carlo można zastosować do otrzymania próby losowej z danego rozkładu \mathcal{F} , znając jedynie funkcję proporcjonalną do jego gęstości prawdopodobieństwa. Metody te generują ergodyczny łańcuch Markowa z rozkładem stacjonarnym \mathcal{F} .

Zastosowanie metod Markov chain Monte Carlo w analizie niezawodności opiera się na pomysłе wykorzystania ich do otrzymania próby z rozkładu optymalizującego metodę importance sampling. Jednak wbrew potocznym opiniom, oszacowanie prawdopodobieństwa awarii w ten sposób wymaga więcej niż wygenerowania jednego elementu próby losowej. Istotą proponowanej metody jest wyznaczenie stałej normalizującej funkcji $I_{\Omega_f}(\mathbf{x})f(\mathbf{x})$ (por. (2.54)), która jest równa prawdopodobieństwu awarii. Podejście to wymaga zastosowania algorytmu pozwalającego oszacować stałą normalizującą gęstość prawdopodobieństwa próby losowej, otrzymanej za pomocą Markov chain Monte Carlo.

7.2. Metoda wzajemnej entropii

W niniejszym podrozdziale przedstawiono teoretyczne podstawy metody wzajemnej entropii. Posłużono się w tym celu zagadnieniem oszacowania prawdopodobieństwa awarii określonej wzorem (2.5) [59]. Metoda wzajemnej entropii może być również stosowana do wyznaczenia prawdopodobieństwa rzadkich zdarzeń, które zdefiniowano w inny sposób. Na przykład, w pracy [21] wykorzystano to podejście do oszacowania prawdopodobieństwa zdarzenia danego nierównością $g(\mathbf{Y}) > \gamma$, gdzie g jest funkcją

wielowymiarowej zmiennej losowej \mathbf{Y} , a γ jest liczbą rzeczywistą.

Powiedzieliśmy już, że w praktyce wybór rozkładu, z którego generowana jest próba losowa w metodzie importance sampling jest ograniczony do pewnej rodziny parametrycznej o gęstościach prawdopodobieństwa $f = \{f(\mathbf{x}, \mathbf{v})\}$, $\mathbf{v} \in V$, gdzie V jest zbiorem parametrów rozkładu.

Parametry \mathbf{v} można określić rozwiązując dla elementów rodziny f zadanie minimalizacji wariancji estymatora (2.49) określonej wzorem (2.51). Zadanie to definiuje się jako

$$\min_{\mathbf{v} \in V} \text{Var}_{f_{\mathbf{v}}} \left[I_{\Omega_f}(\mathbf{x}) \frac{f(\mathbf{x})}{f_{\mathbf{v}}(\mathbf{x})} \right], \quad (7.1)$$

gdzie $f_{\mathbf{v}}(\mathbf{x}) = f(\mathbf{x}, \mathbf{v})$ jest elementem rodziny f . Równoważne sformułowanie (7.1) jest następującej postaci (por. (2.51) oraz [56]):

$$\min_{\mathbf{v} \in V} \left\{ \mathbb{E}_{f_{\mathbf{v}}} \left[I_{\Omega_f}(\mathbf{x}) \frac{f^2(\mathbf{x})}{f_{\mathbf{v}}^2(\mathbf{x})} \right] \right\}. \quad (7.2)$$

To zadanie optymalizacji można rozwiązać wykorzystując poniższe oszacowanie wartości oczekiwanej;

$$\min_{\mathbf{v} \in V} \left\{ \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f_{\mathbf{v}}(\mathbf{X}_i)} \frac{f(\mathbf{X}_i)}{f_{v_s}(\mathbf{X}_i)} \right\}, \quad (7.3)$$

gdzie f_{v_s} jest gęstością rozkładu prawdopodobieństwa zmiennych losowych \mathbf{X}_i .

Sformułowanie (7.3) jest podstawą prostego algorytmu iteracyjnego (por. algorytm 8.4.1 w [102]):

1. Przyjąć $f_{v_s} = f$.
2. Wygenerować próbę losową $\mathbf{X}_1, \dots, \mathbf{X}_n$ z rozkładu o gęstości f_{v_s} i rozwiązać zadanie optymalizacji (7.3). Oznaczmy rozwiązanie jako $\hat{\mathbf{v}}^*$. Niech $\hat{\mathbf{v}}^*$ będzie oszacowaniem optymalnej wartości parametru \mathbf{v}^* .
3. Na podstawie próby losowej z gęstości prawdopodobieństwa $q(\mathbf{x}) = f(\mathbf{x}, \hat{\mathbf{v}}^*)$ oszacować prawdopodobieństwo awarii, korzystając przy tym z estymatora (2.49).

W celu uzyskania dokładniejszego oszacowania parametru \mathbf{v}^* można powtórzyć drugi krok algorytmu kilkakrotnie, przyjmując $\mathbf{v}_s = \hat{\mathbf{v}}^*$.

Oczywiście stosowanie adaptacyjnych metod Monte Carlo ma sens, jeżeli uzyskanie oszacowania $\hat{\mathbf{v}}^*$, które zapewni zmniejszenie wariancji (2.49), wymaga mniej obliczeń niż oszacowanie prawdopodobieństwa awarii metodą Monte Carlo.

Innym podejściem pozwalającym wyznaczyć parametry \mathbf{v} jest wybór rozkładu prawdopodobieństwa, który jest w pewnym sensie podobny do optymalnego rozkładu metody importance sampling q^* (2.54). Okazuje się, że w celu porównania rozkładów prawdopodobieństwa wygodnie jest posłużyć się ich wzajemną entropią (ang. *cross-entropy*), która jest również określana jako odległość Kullbacka-Leiblera (ang. *Kullback-Leibler distance*). Wzajemną entropię dwóch rozkładów prawdopodobieństwa o gęstościach $f(\mathbf{x})$ i $q(\mathbf{x})$ definiuje się jako:

$$\mathcal{D}(f, q) = \int f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}. \quad (7.4)$$

Warto zauważyć, że wzajemna entropia nie jest odległością w sensie formalnym, ponieważ w ogólności nie spełnia warunku symetrii: $\mathcal{D}(f, g) \neq \mathcal{D}(g, f)$.

Wzajemna entropia optymalnego rozkładu importance sampling q^* oraz elementów rodziny f jest określona wyrażeniem:

$$\begin{aligned} \mathcal{D}(q^*(\mathbf{x}), f(\mathbf{x}, \mathbf{v})) &= \int P_f^{-1} I_{\Omega_f}(\mathbf{x}) f(\mathbf{x}) \ln \frac{P_f^{-1} I_{\Omega_f}(\mathbf{x}) f(\mathbf{x})}{f(\mathbf{x}, \mathbf{v})} d\mathbf{x} \\ &= \mathbb{E}_f \left[P_f^{-1} I_{\Omega_f}(\mathbf{x}) \ln \frac{P_f^{-1} I_{\Omega_f}(\mathbf{x}) f(\mathbf{x})}{f(\mathbf{x}, \mathbf{v})} \right] \\ &= \mathbb{E}_f \left[P_f^{-1} I_{\Omega_f}(\mathbf{x}) \ln P_f^{-1} I_{\Omega_f}(\mathbf{x}) f(\mathbf{x}) \right] \\ &\quad - \mathbb{E}_f \left[P_f^{-1} I_{\Omega_f}(\mathbf{x}) \ln f(\mathbf{x}, \mathbf{v}) \right]. \end{aligned} \quad (7.5)$$

Jako optymalne parametry \mathbf{v}^* rozkładu, z którego generowana jest próba losowa, można więc przyjąć te, które minimalizują wzajemną entropię (7.5):

$$\min_{\mathbf{v} \in V} \{ \mathcal{D}(q^*(\mathbf{x}), f(\mathbf{x}, \mathbf{v})) \}. \quad (7.6)$$

Zauważmy, że rozwiązanie powyższego zadania jest równoważne określeniu parametru maksymalizującego drugi człon wyrażenia (7.5):

$$\max_{\mathbf{v} \in V} \{ D(\mathbf{v}) = \mathbb{E}_f [I_{\Omega_f}(\mathbf{x}) \ln f(\mathbf{x}, \mathbf{v})] \}. \quad (7.7)$$

Korzystając z estymatora importance sampling wartości oczekiwanej, zadanie (7.7) przekształca się do postaci

$$\max_{\mathbf{v} \in V} \left\{ \hat{D}_n(\mathbf{v}) = \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \mathbf{v}_s)} \ln f(\mathbf{X}_i, \mathbf{v}) \right\}, \quad (7.8)$$

gdzie $\mathbf{X}_1, \dots, \mathbf{X}_N$ jest próbą losową z rozkładu prawdopodobieństwa o gęstości $f(\mathbf{x}, \mathbf{v}_s)$.

Rozwiązanie problemów (7.3) i (7.8) ma ten sam cel, którym jest określenie parametrów w pewnym sensie optymalnego rozkładu wykorzystywanego w metodzie importance sampling. Ponieważ w przypadku zadania (7.3) otrzymujemy rozkład minimalizujący wariancję poszukiwanego oszacowania to wydaje się, że stosowanie rozkładu określonego za pomocą metody wzajemnej entropii nie ma sensu. Jednak metoda wzajemnej entropii wykorzystuje znacznie łatwiejszy do rozwiązania problem optymalizacji, który w pewnych przypadkach można nawet rozwiązać analitycznie [21]. Tak, więc stosowanie metody wzajemnej entropii jest uzasadnione, jeżeli oszczędność nakładów obliczeniowych koniecznych do rozwiązania zadania optymalizacji rekompensuje wykorzystanie rozkładu, który nie zapewnia minimalizacji wariancji estymatora. W pracy [56] przedstawiono analizę pokazującą równoważność rozwiązania problemu minimalizacji wariancji i minimalizacji wzajemnej entropii w przypadku, kiedy prawdopodobieństwo awarii dąży do zera (dowód dotyczy ogólniejszego zagadnienia symulacji rzadkich zdarzeń).

7.3. Algorytm metody wzajemnej entropii

W tym podrozdziale omówimy implementację algorytmów, które pozwalają na efektywne oszacowanie prawdopodobieństwa awarii za pomocą metody wzajemnej entropii.

Zauważmy, że jeżeli funkcja \hat{D}_n w (7.8) jest wypukła i różniczkowalna względem \mathbf{v} , to gradient tej funkcji jest zerowy dla rozwiązania zadania (7.8) (por. [21]): $\nabla \hat{D}_n(\hat{\mathbf{v}}^*) = \mathbf{0}$. Tak więc oszacowanie parametru $\hat{\mathbf{v}}^*$, minimalizującego wzajemną entropię q^* i rozkładu z rodziny $f = \{f(\mathbf{x}, \mathbf{v})\}$, można otrzymać rozwiązując następujący układ równań:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})} \frac{\partial \ln f(\mathbf{X}_i, \mathbf{v})}{\partial v_1} &= 0 \\ \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})} \frac{\partial \ln f(\mathbf{X}_i, \mathbf{v})}{\partial v_2} &= 0 \\ &\vdots \\ \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})} \frac{\partial \ln f(\mathbf{X}_i, \mathbf{v})}{\partial v_k} &= 0, \end{aligned}$$

(7.9)

gdzie $\mathbf{v} = \{v_1, v_2, \dots, v_k\}$.

W przypadku niektórych rodzin rozkładów prawdopodobieństwa równania (7.9) przybierają szczególnie łatwą do rozwiązania postać. Dla zastosowań w analizie niezawodności szczególnie interesujący jest przypadek wielowymiarowej zmiennej normalnej o niezależnych składowych. Gęstość prawdopodobieństwa takiej zmiennej losowej jest określona wzorem:

$$\varphi_I(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\sigma}) = \prod_{j=1}^k \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right), \quad (7.10)$$

gdzie $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_k\}$ i $\boldsymbol{\sigma} = \{\sigma_1, \dots, \sigma_k\}$, to odpowiednio wektory wartości oczekiwanych i odchyłeń standardowych. Zauważmy, że pochodne logarytmu gęstości (7.10) względem wartości oczekiwanych są postaci:

$$\frac{\partial \ln \varphi_I(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\sigma})}{\partial \mu_j} = \frac{x_j - \mu_j}{\sigma_j^2} \quad j = 1, \dots, k. \quad (7.11)$$

Po prostych przekształceniach, korzystając z (7.11) oraz (7.9), otrzymujemy następujące oszacowania wartości oczekiwanych

$$\hat{\mu}_j^* = \frac{\frac{1}{n} \sum_{i=1}^n X_{ji} I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})}}{\frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})}} \quad j = 1, \dots, k. \quad (7.12)$$

W celu wyznaczenia optymalnych odchyłeń standardowych, które występują w (7.10), wykorzystamy następujące pochodne

$$\frac{\partial \ln \varphi_I(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\sigma})}{\partial \sigma_j} = \frac{(x_j - \mu_j)^2 - \sigma_j^2}{\sigma_j^3} \quad j = 1, \dots, k. \quad (7.13)$$

Podobnie jak w przypadku wartości oczekiwanych również odchylenia standardowe mogą być oszacowane za pomocą prostych wyrażeń:

$$\hat{\sigma}_j^{*2} = \frac{\frac{1}{n} \sum_{i=1}^n (X_{ji} - \hat{\mu}_j^*)^2 I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})}}{\frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i) \frac{f(\mathbf{X}_i)}{f(\mathbf{X}_i, \tilde{\mathbf{v}})}} \quad j = 1, \dots, k. \quad (7.14)$$

Oczywiście w mianownikach (7.12) i (7.14) znajduje się oszacowanie prawdopodobieństwa awarii. Nasuwa się więc pytanie czy wykorzystywanie takich oszacowań

parametrów rozkładu ma sens, jeżeli wymagają one przybliżenia poszukiwanego rozwiązania zadania. W prezentowanych dalej testach numerycznych wariancja estymatora (7.12) była znacząco niższa od wariancji oszacowania prawdopodobieństwa awarii, które występuje w mianowniku tego wzoru. Wyrażenia (7.12) i (7.14) są równoważne alternatywnemu dla (2.49) estymatorowi metody importance sampling (por. *ratio estimate* w [53]) określanemu jako:

$$\mathbb{E}[\chi(\mathbf{X})] \approx \frac{\sum_{i=1}^n \chi(\mathbf{X}_i) \frac{\phi(\mathbf{X}_i)}{\psi(\mathbf{X}_i)}}{\sum_{i=1}^n \frac{\phi(\mathbf{X}_i)}{\psi(\mathbf{X}_i)}}, \quad (7.15)$$

gdzie \mathbf{X}_i , $i = 1, \dots, n$ jest próbą losową o gęstości ψ , ϕ jest gęstością prawdopodobieństwa zmiennej \mathbf{X} , a χ jest funkcją, której oszacowanie wartości oczekiwanej jest poszukiwane. Estymator (7.15) w wielu przypadkach ma mniejszą wariancję niż estymator (2.49).

Efektywność oszacowania optymalnych parametrów metody wzajemnej entropii za pomocą (7.9) zależy od odpowiedniego doboru \mathbf{v}_s - parametrów rozkładu prawdopodobieństwa, z którego generowana jest próbka losowa. Oczywiście w przypadkach, gdy prawdopodobieństwo awarii jest małe określenie właściwych parametrów może być trudne. W pracy [56] (algorytm 5.2) do rozwiązania tego problemu zaproponowano algorytm iteracyjny. Poniżej jest on przedstawiony z wykorzystaniem notacji stosowanej w analizie niezawodności.

Omawiany algorytm wykorzystuje prostą modyfikację funkcji granicznej:

$$\tilde{g}(\mathbf{x}) = g(\mathbf{x}) + \gamma, \quad (7.16)$$

gdzie $\gamma \geq 0$. Obszar awarii określony przez funkcję \tilde{g} ma oczywiście większe prawdopodobieństwo niż obszar zdefiniowany za pomocą g . Dzięki temu określenie parametrów rozkładu prawdopodobieństwa przy wykorzystaniu estymatora (7.9) jest łatwiejsze.

Celem algorytmu jest wyznaczenie na podstawie kolejnych symulacji sekwencji parametrów: $\gamma_0, \gamma_1, \dots, \gamma_k$ określających modyfikacje funkcji granicznej (7.16) oraz $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_k$ definiujących rozkłady prawdopodobieństwa, z których generowane są próby losowe.

Algorytm wymaga przyjęcia następujących parametrów: stałej ρ , $0 < \rho \leq 1$ (w [56] są sugerowane wartości $0.01 \leq \rho \leq 0.1$), stałych $\delta > 0$ i $\alpha > 1$ oraz określenia liczby symulacji n , które są wykonywane w jednym kroku iteracji.

Wykonanie algorytmu obejmuje poniższe kroki:

1. Inicjalizacja algorytmu. Niech $\rho_0 = \rho$. Na podstawie próby losowej $\mathbf{X}_1^{(0)}, \dots, \mathbf{X}_n^{(0)}$, otrzymanej z rozkładu prawdopodobieństwa o gęstości $f(\mathbf{x}, \mathbf{v}^{(0)})$,

oszacować $\hat{\gamma}_0$, które jest kwantylem empirycznym rzędu ρ_0 funkcji próby losowej $g(\mathbf{X}_1^{(0)}), \dots, g(\mathbf{X}_n^{(0)})$ (por. 2.2):

$$P[g(\mathbf{X}_i^{(0)}) \leq \hat{\gamma}_0] = \rho_0, \quad i = 1, \dots, n. \quad (7.17)$$

Następnie niech $t = 1$.

- Wykorzystując próbę losową $\mathbf{X}_1^{(t-1)}, \dots, \mathbf{X}_n^{(t-1)}$ wyznaczyć składowe wektora parametrów $\mathbf{v}^{(t)}$ poprzez rozwiązanie zadania optymalizacji:

$$\hat{\mathbf{v}}^{(t)} = \arg \max_{\mathbf{v} \in V} \left\{ \frac{1}{n} \sum_{i=1}^n I_{\{g(\mathbf{x}) \leq \hat{\gamma}_{t-1}\}}(\mathbf{X}_i^{(t-1)}) \frac{f(\mathbf{X}_i^{(t-1)})}{f(\mathbf{X}_i^{(t-1)}, \hat{\mathbf{v}}^{(t-1)})} \ln f(\mathbf{X}_i^{(t-1)}, \mathbf{v}) \right\}, \quad (7.18)$$

gdzie $I_{\{g(\mathbf{x}) \leq \hat{\gamma}_{t-1}\}}$ jest funkcją charakterystyczną obszaru, w którym $g(\mathbf{x}) \leq \hat{\gamma}_{t-1}$ (por. 2.2).

- Wygenerować próbę losową $\mathbf{X}_1^{(t)}, \dots, \mathbf{X}_n^{(t)}$ z rozkładu prawdopodobieństwa o gęstości $f(\mathbf{x}, \hat{\mathbf{v}}^{(t)})$ oraz przyjąć $\rho_t = \rho$.
- Na podstawie $\mathbf{X}_1^{(t)}, \dots, \mathbf{X}_n^{(t)}$ wyznaczyć kwantyl empiryczny $\hat{\gamma}_t$ rzędu ρ_t funkcji próby losowej $g(\mathbf{X}_1^{(t)}), \dots, g(\mathbf{X}_n^{(t)})$.
- Jeżeli $\hat{\gamma}_t \leq 0$, przyjąć $\hat{\gamma}_t = 0$ oraz wyznaczyć oszacowanie $\hat{\mathbf{v}}_*^{(t)}$ parametru \mathbf{v}^* , które jest rozwiązaniem zadania optymalizacji:

$$\hat{\mathbf{v}}_*^{(t)} = \arg \max_{\mathbf{v} \in V} \left\{ \frac{1}{n} \sum_{i=1}^n I_{\Omega_f}(\mathbf{X}_i^{(t)}) \frac{f(\mathbf{X}_i^{(t)})}{f(\mathbf{X}_i^{(t)}, \hat{\mathbf{v}}^{(t-1)})} \ln f(\mathbf{X}_i^{(t)}, \mathbf{v}) \right\}. \quad (7.19)$$

Następnie przejść do punktu 7.

- Jeżeli $\hat{\gamma}_t > 0$, sprawdzić czy istnieje $\bar{\rho}$ takie, że $\bar{\gamma}$ będące kwantylem empirycznym rzędu $\bar{\rho}$ funkcji próby losowej $g(\mathbf{X}_1^{(t-1)}), \dots, g(\mathbf{X}_n^{(t-1)})$ spełnia $\bar{\gamma} \leq \max\{0, \hat{\gamma}_{t-1} - \delta\}$:
 - Jeżeli takie $\bar{\rho}$ istnieje i $\bar{\rho} = \rho_t$ to przyjąć $t = t + 1$ i powtórzyć iteracje od kroku 2;
 - Jeżeli takie $\bar{\rho}$ istnieje oraz $\bar{\rho} < \rho_t$, to przyjąć $\rho_t = \bar{\rho}$ i wrócić do kroku 4;
 - W przeciwnym przypadku (na przykład jeżeli takie $\bar{\rho}$ nie istnieje) powiększyć liczbę symulacji w jednej iteracji $n = \alpha n$ i powrócić do kroku 3.

7. Wyznaczyć oszacowanie prawdopodobieństwa awarii na podstawie estymatora importance sampling (2.49) wykorzystując próbę losową otrzymaną z rozkładu o gęstości $f(\mathbf{x}, \hat{\mathbf{v}}_*^{(t)})$.

Jeżeli dla każdego rozkładu prawdopodobieństwa z rodziny $f = \{f(\mathbf{x}, \mathbf{v})\}$ obszar awarii ma niezerowe prawdopodobieństwo:

$$\forall \mathbf{v} \in V \int I_{\Omega_f}(\mathbf{x}) f(\mathbf{x}, \mathbf{v}) d\mathbf{x} > 0 \quad (7.20)$$

to powyższy algorytm jest zbieżny w skończonej liczbie iteracji do rozwiązania problemu (7.8) prawie na pewno. Dowód tego twierdzenia podano w pracy [56].

Zauważmy, że w omawianym algorytmie oszacowania kolejnych wartości parametrów \mathbf{v}_t , γ_t są kontrolowane przez ciąg parametrów ρ_t . Dobierając odpowiednio dużą wartość parametru ρ zapewniamy, że liczba realizacji próby losowej w zmodyfikowanym obszarze awarii $\{\mathbf{x} : g(\mathbf{x}) \leq \hat{\gamma}_{t-1}\}$ będzie wystarczająca do dobrego oszacowania parametrów \mathbf{v} . W praktyce algorytm jest inicjowany dla dość dużego ρ i wartość ta nie musi być zmniejszana w trakcie jego wykonywania. Dzieje się tak, ponieważ dzięki konstrukcji problemu wzajemnej entropii rozkład prawdopodobieństwa $g(\mathbf{X})$, gdy $\mathbf{X} \sim f(\mathbf{x}, \mathbf{v}_{t+1})$, zazwyczaj ma cięższy dolny ogon, niż kiedy $\mathbf{X} \sim f(\mathbf{x}, \mathbf{v}_t)$, a więc $\gamma(\mathbf{v}_{t+1}, \rho) < \gamma(\mathbf{v}_t, \rho)$ dla wszystkich ρ . W przypadkach, kiedy ta własność nie jest spełniona zbieżność algorytmu zapewniają modyfikacje parametrów wykonywane w kroku 6 [56].

7.4. Ocena efektywności metody wzajemnej entropii

Przedmiotem tego rozdziału jest ocena efektywności numerycznej metody wzajemnej entropii na podstawie przykładów testowych. W tym celu posłużymy się zestawem zadań, które w pracy [34] zostały wykorzystane do porównania różnych algorytmów importance sampling. Zadania te wykorzystują jawnie określone funkcje graniczne mające cechy, które utrudniają zastosowanie importance sampling. Są to np.: duża liczba zmiennych losowych, małe prawdopodobieństwo awarii, silna nieliniowość powierzchni granicznej, wielokrotne punkty projektowe, zaszumiona funkcja graniczna. Jako kryterium oceny algorytmu posłużymy się liczbą symulacji, jaką należy wykonać, aby otrzymać estymator o współczynniku zmienności $\hat{e} = 0.1$ (2.44).

Prezentowane poniżej zadania zostały sformułowane w gaussowskiej przestrzeni standardowej tak, więc jako rodzinę rozkładów prawdopodobieństwa, z których generowane są próby losowe przyjęto wielowymiarową zmienną normalną o niezależnych składowych i jednostkowych wariancjach. W tym przypadku wyznaczenie optymalnych

parametrów rozkładu prawdopodobieństwa ogranicza się do oszacowania wartości oczekiwanych za pomocą (7.12).

W podanych przykładach zastosowano nieco zmodyfikowany algorytm adaptacyjny niż ten przedstawiony w podrozdziale 7.3. W wykorzystanym algorytmie symulacje w danej iteracji prowadzi się do momentu, gdy oszacowania wartości oczekiwanych kolejnego rozkładu mają mniejsze współczynniki zmienności niż przyjęta wartość $\hat{\epsilon}_v$. Drugim z parametrów, jaki należy określić dla algorytmu jest omawiane już ρ . Ponieważ dzięki tej modyfikacji nie zachodzi potrzeba wykonania kroku 6 to nie ma potrzeby określenia parametrów δ i α . Wartości parametrów ρ i $\hat{\epsilon}_v$ podawane w wynikach eksperymentów numerycznych zapewniały najlepszą efektywność algorytmu spośród kilkudziesięciu analizowanych kombinacji. Jednak wartości te nie mogą być traktowane jako optymalne.

Przedstawione liczby symulacji \bar{n} , jakie były potrzebne do otrzymania estymatora o wymaganym współczynniku zmienności są wartościami średnimi, które uzyskano dla kilkuset rozwiązań tego samego zadania ze stałymi parametrami algorytmu.

Przykład 7.1: Ocena efektywności w z zależności od liczby zmiennych losowych i wielkości prawdopodobieństwa awarii. Funkcja graniczna wykorzystywana w tym przykładzie określa powierzchnię graniczną w postaci hiperpłaszczyzny:

$$g_1 = \sqrt{k}\beta - \sum_{i=1}^k U_i, \quad (7.21)$$

gdzie U_i , $i = 1, 2, \dots, k$, są standardowymi zmiennymi losowymi, a β jest wskaźnikiem niezawodności. Zmieniając wartości k i β można łatwo różnicować liczbę zmiennych losowych i prawdopodobieństwo awarii. Testy numeryczne przeprowadzono dla następujących parametrów funkcji granicznej: $\beta = 1.0$, $\beta = 5.0$, $\beta = 10.0$ oraz $k = 2$, $k = 10$, $k = 50$. Wyniki eksperymentów przedstawiono w tablicy 7.1. Jak można zaobserwować konieczna do uzyskania wymaganego poziomu zmienności estymatora liczba symulacji wzrasta wraz z wymiarem przestrzeni zmiennych losowych i wartością indeksu niezawodności. Przekroczenie limitu zaznaczone dla $\beta = 10$ i $k = 10$ oraz $k = 50$ oznacza, że wymagana dokładność oszacowania nie została osiągnięta po dopuszczalnej (bardzo dużej) liczbie symulacji.

Jak widzimy w prezentowanym przykładzie koszty numeryczne konieczne do uzyskania oszacowania o wymaganej dokładności nie są małe. Jednak biorąc pod uwagę, że jest to całkowity koszt analizy, a algorytm wymaga jedynie doboru dwóch parametrów to w opinii autora efektywność prezentowanej metody jest zadowalająca.

Przykład 7.2: Zastosowanie metody wzajemnej entropii w zagadnieniu z silnie nieliniową funkcją graniczną. W tym przykładzie funkcja graniczna jest określona

		$\beta = 1.0$	$\beta = 5.0$	$\beta = 10.0$
$k = 2$	ρ	0.3	0.3	0.3
	\hat{e}_v	0.0389	0.0473	0.5527
	\bar{n}	230	884	1912
$k = 10$	ρ	0.5	0.5	0.45
	\hat{e}_v	0.5	0.5	0.05
	\bar{n}	360	1375	3181 16% ponad limit
$k = 50$	ρ	0.5	0.5	0.45
	\hat{e}_v	0.1	0.1	0.05
	\bar{n}	882	2968	8569 38% ponad limit

TABLICA 7.1. Ocena efektywności metody wzajemnej entropii dla funkcji granicznej w postaci hiperpłaszczyzny.

wzorem:

$$g_2 = \pm \sum_{i=1}^k X_i \mp C, \quad (7.22)$$

gdzie X_i , $i = 1, 2, \dots, k$, są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym określonym przez parametr λ . Po transformacji do gaussowskiej przestrzeni standardowej funkcja graniczna (7.22) staje się silnie nieliniowa:

$$G_2 = \pm \sum_{i=1}^k \ln [\Phi(-U_i)] / \lambda \pm C, \quad (7.23)$$

gdzie U_i , $i = 1, \dots, k$, są zmiennymi losowymi o standardowym rozkładzie normalnym. Obliczenia w przykładzie wykonano dla $\lambda = 1$ oraz $k = 20$. Wartości parametru C przedstawiono w tabelcy 7.2.

Kiedy we wzorze (7.23) występuje odejmowanie - powierzchnia graniczna ma silnie ujemną krzywiznę. W takim przypadku, jak pokazują wyniki w tabelcy 7.2, ilość symulacji koniecznych do otrzymania estymatora o żądanej dokładności zwiększa się wraz z wartością parametru C . Dzieje się tak z dwóch powodów. Po pierwsze zwiększa się odległość obszaru awarii od środka układu współrzędnych, który jest punktem startowym algorytmu. Drugą z przyczyn jest zmniejszanie się obszaru awarii wraz z rosnącą krzywizną w wyniku czego coraz mniejsza część wykonywanych symulacji "wpada" do istotnego dla oszacowania obszaru. Podane w tabelcy 7.2 liczby symulacji, jakie wykonano dla uzyskania oszacowania

C	16.175	11.077	8.951	7.435	6.277	5.343
P_f	0.20	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
β	0.841	2.328	3.093	3.722	4.268	4.765
ρ	0.475	0.3	0.3	0.3	0.2	0.2
\hat{e}_v	0.313	0.024	0.015	0.01	0.005	0.001
\bar{n}	493	668	3014	5512	8505 1.0% ponad limit	15022 1.5% ponad limit
P_f	0.20	$1.0 \cdot 10^{-2}$	$0.97 \cdot 10^{-3}$	$0.98 \cdot 10^{-4}$	$0.97 \cdot 10^{-5}$	$0.93 \cdot 10^{-6}$

TABLICA 7.2. Ocena efektywności metody wzajemnej entropii w przypadku silnie nieliniowej funkcji granicznej z ujemną krzywizną.

o wymaganej dokładności są duże. Jednak biorąc pod uwagę, że omawiana metoda nie wymaga dodatkowego algorytmu poszukiwania punktu projektowego można rozważyć jej zastosowanie dla problemów tego typu.

Jeżeli we wzorze (7.23) występuje suma to obszar awarii ma krzywiznę dodatnią. W takim przypadku istotna część gęstości prawdopodobieństwa w obszarze awarii znajduje się z dala od punktu projektowego. W zadaniach tego typu metoda importance sampling nie jest efektywna, jeżeli próba losowa jest generowana z jednomodowego rozkładu prawdopodobieństwa. Wyniki symulacji przedstawione w 7.3 pokazują, że tak jest również w tym przypadku. Uzyskane oszacowania wymagały bardzo dużej liczby symulacji, a dla dużych C praktycznie nie można było uzyskać oszacowania o żądanej dokładności. Ponadto uzyskane oszacowania prawdopodobieństwa awarii wraz z rosnącym C stawały się coraz bardziej obciążone. Można stwierdzić, że w przypadku powierzchni granicznej z dodatnią krzywizną testowana metoda nie powinna być stosowana.

Przykład 7.3: Funkcja graniczna z "szumem". W tym przykładzie wykorzystano funkcję graniczną następującej postaci:

$$g_3 = X_1 + 2X_2 + 2X_3 + X_4 - 5X_5 - 5X_6 + 0.001 \sum_{i=1}^6 \sin(100X_i), \quad (7.24)$$

gdzie wszystkie zmienne losowe są niezależne i mają rozkład logarytmiczno-normalny. Zmienne X_1, X_2, X_3, X_4 mają wartości oczekiwane równe 120.0 i odchylenia standardowe 12.0. Zmienna X_5 ma wartość oczekiwaną 50.0 i odchylenie standardowe 15.0 a zmienna X_6 wartość oczekiwaną 40.0 i odchylenie standardowe 12.0. Prawdopodobieństwo obszaru awarii określonego funkcją graniczną g_3 wynosi $P_f = 1.23 \cdot 10^{-2}$.

Ostatni człon wyrażenia (7.24) wprowadza do funkcji granicznej małe zaburzenie, które

C	25.90	31.856	36.720	41.050	45.067
P_f	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
β	1.282	2.328	3.093	3.722	4.268
ρ	0.5	0.4	0.3	0.3	0.2
\hat{e}_v	0.398	0.0316	0.0562	0.0562	0.0421
\bar{n}	1106.27	1407.58	8020.14 4.0% ponad limit	13921.73 13.3% ponad limit	28240.91 63.33% ponad limit
P_f	$0.971 \cdot 10^{-1}$	$0.963 \cdot 10^{-2}$	$0.907 \cdot 10^{-3}$	$0.840 \cdot 10^{-4}$	$0.736 \cdot 10^{-5}$

TABLICA 7.3. Ocena efektywności metody wzajemnej entropii dla silnie nieliniowej funkcji granicznej z dodatnią krzywizną.

określa się jako szum. Powierzchnia awarii określona przez funkcję graniczną tego typu jest bardzo nieregularna, a metody gradientowe FORM/SORM mają w tym przypadku problemy z poprawnym oszacowaniem prawdopodobieństwa awarii. Tak, więc adaptacyjne metody symulacyjne stanowią podstawową klasę algorytmów służących do rozwiązywania zadań z zaszumioną funkcją graniczną. W praktyce szum występuje bardzo często, jego przyczyną może być nie tylko natura rozwiązywanego zagadnienia, ale również błędy generowane przez procedury numeryczne, które są wykorzystywane do wyznaczania wartości funkcji granicznej.

W tym przykładzie obliczenia wykonano przy następujących wartościach parametrów algorytmu: $\rho = 0.4$ i $\hat{e}_v = 0.398$. Średnia liczba symulacji koniecznych do uzyskania oszacowania o 10% współczynnika zmienności wyniosła 710, z których przeciętnie 564 było wykorzystywanych do końcowego oszacowania prawdopodobieństwa awarii. Średnia wartość oszacowanego prawdopodobieństwa awarii to $P_f = 1.21 \cdot 10^{-2}$.

Ilość symulacji wymaganych do końcowego oszacowania prawdopodobieństwa awarii jest porównywalna z liczbą 500, która jest podawana w pracy [34]. Natomiast konieczność wykonania ok. 150 symulacji w celu oszacowania wartości oczekiwanych jest akceptowalna.

Przykład 7.4: Efektywność w przypadku oceny niezawodności systemów awarii. Oszacowanie prawdopodobieństwa awarii rzeczywistych konstrukcji często wymaga zastosowania analizy niezawodności systemów. Dlatego też istotne jest, aby adaptacyjne metody symulacyjne pozwalały na efektywne oszacowanie prawdopodobieństwa sum i części wspólnych elementowych obszarów awarii (por. 6.3).

W omawianym przykładzie obszar awarii systemu szeregowego (6.4) jest sumą obszarów

awarii określonych przez następujące funkcje graniczne:

$$\begin{aligned} g_{41} &= X_1 + 2X_3 + 2X_4 + X_5 - 5X_6 - 5X_7, \\ g_{42} &= X_1 + 2X_2 + X_4 + X_5 - 5X_6, \\ g_{43} &= X_2 + 2X_3 + X_4 - 5X_7, \end{aligned} \quad (7.25)$$

gdzie wszystkie zmienne losowe są niezależne. Zmienne X_1, \dots, X_5 mają rozkład logarytmiczno-normalny o wartości oczekiwanej 120.0 i odchyleniu standardowym 12.0, X_6 ma rozkład Gumbela z wartością oczekiwaną 20.0 i odchyleniem standardowym 6.0 a X_7 ma rozkład Gumbela z wartością oczekiwaną 25.0 i odchyleniem standardowym 7.5. Oszacowane za pomocą klasycznej metody Monte Carlo prawdopodobieństwo awarii omawianego systemu wynosi $P_f = 2 \cdot 10^{-2}$. Do rozwiązania tego problemu wykorzystano algorytm wzajemnej entropii z parametrami $\rho = 0.4$ i $\hat{\epsilon}_v = 0.398$. Do obliczeń przyjęto funkcję awarii systemu postaci:

$$g_4 = \min(g_{41}, g_{42}, g_{43}). \quad (7.26)$$

Uzyskanie oszacowania o żądanej dokładności wymagało średnio 1050 symulacji, z których przeciętnie 886 było koniecznych do oszacowania prawdopodobieństwa awarii. Jego średnia wartość wyniosła $\bar{P}_f = 0.019$. Należy jednak podkreślić, że omawiany algorytm wykorzystuje do oszacowania prawdopodobieństwa awarii próbę losową z rozkładu jednomodowego. Obszary awarii systemów szeregowych mają zazwyczaj wielokrotne punkty projektowe; mamy, więc do czynienia z sytuacją podobną jak dla nieliniowej funkcji z dodatnią krzywizną (przykład 7.2). W takim przypadku bardziej efektywne jest wykorzystanie próby losowej z mieszanego rozkładu prawdopodobieństwa.

System równoległy analizowany w tym przykładzie jest określony za pomocą następujących funkcji granicznych:

$$\begin{aligned} g_{51} &= 2.677 - U_1 - U_2, \\ g_{52} &= 2.500 - U_2 - U_3, \\ g_{53} &= 2.323 - U_3 - U_4, \\ g_{54} &= 2.250 - U_4 - U_5, \end{aligned} \quad (7.27)$$

gdzie U_i , $i = 1, \dots, 5$, są standardowymi zmiennymi normalnymi. Podana w pracy [34] dokładna wartość prawdopodobieństwa awarii systemu (7.27) wynosi $P_f = 2.087 \cdot 10^{-4}$.

Parametry algorytmu wykorzystanego w tym przypadku miały wartości $\rho = 0.2$ i $\hat{\epsilon}_v = 0.2$. Średnia oszacowana wartość prawdopodobieństwa awarii wyniosła $2.11 \cdot 10^{-4}$, a osiągnięcie żądanej zmienności estymatora wymagało średnio 2818 symulacji. Biorąc pod uwagę postać obszaru awarii, liczbę zmiennych losowych oraz wielkość prawdopodobieństwa awarii należy stwierdzić, że jest to duża liczba. Przyczyny braku efektywności w tym przykładzie należy upatrywać w postaci systemowej funkcji granicznej, jaka została wyko-

rzystana do obliczeń:

$$g_5 = \max(g_{51}, g_{52}, g_{53}, g_{54}). \quad (7.28)$$

W przypadku takiej nieregularnej funkcji granicznej trudno jest uzyskać wystarczająco dokładne oszacowanie parametrów rozkładu, z którego generuje się kolejne próby losowe.

Podsumowując wyniki uzyskane w powyższych przykładach można stwierdzić, że istnieją problemy, w których prezentowany algorytm oparty na metodzie wzajemnej entropii może być efektywnie stosowany. Są to zagadnienia, gdzie prawdopodobieństwo w obszarze awarii jest skupione w pobliżu jednego punktu. Istotne jest, że omawiany algorytm może być stosowany w przypadku funkcji granicznych z szumem. Niestety omawiana metoda nie powinna być stosowana, gdy obszar awarii ma wielokrotne punkty projektowe. Niezadowalającą efektywność uzyskano również w przypadku systemu równoległego.

Analiza niezawodności za pomocą metod wykorzystujących łańcuchy Markowa

8.1. Markov chain Monte Carlo (MCMC)

W tym rozdziale omówione zostaną algorytmy, które umożliwiają otrzymanie próby losowej o rozkładzie prawdopodobieństwa bliskim danemu rozkładowi \mathcal{F} bez konieczności bezpośredniej symulacji z tego rozkładu (np. poprzez zastosowanie transformacji (2.23) do próby losowej z rozkładu jednorodnego na przedziale $[0, 1]$). Metody te, nazywane w literaturze anglojęzycznej Markov chain Monte Carlo (MCMC), wykorzystują ergodyczny łańcuch Markowa z rozkładem stacjonarnym \mathcal{F} .

Zaczynając od przyjętego stanu $\mathbf{x}^{(0)}$ generuje się łańcuch Markowa $\mathbf{X}^{(t)}$, $t = 0, 1, \dots$, z jądrem przejścia, którego rozkładem stacjonarnym jest \mathcal{F} - nazywany również *rozkładem celu*. W ten sposób zapewniona jest zbieżność według prawdopodobieństwa ciągu $\mathbf{X}^{(t)}$ do zmiennej losowej o rozkładzie \mathcal{F} . Jeżeli łańcuch Markowa jest ergodyczny to wpływ rozkładu stanu początkowego na elementy próby losowej maleje wraz z wartością indeksu t . Można, więc przyjąć, że dla dostatecznie dużego t_0 próba losowa $\mathbf{X}^{(t_0)}, \mathbf{X}^{(t_0+1)}, \dots$ ma rozkład prawdopodobieństwa bliski rozkładowi celu. Tak otrzymana próba losowa może być wykorzystana do oszacowania wartości oczekiwanych względem rozkładu \mathcal{F} .

Dodatnia korelacja między elementami próby losowej otrzymanej za pomocą MCMC jest chyba największą wadą tej metody. Wewnętrzna korelacja silnie zmniejsza efektywność estymatorów. Otrzymanie oszacowania o podobnym współczynniku zmienności wymaga znacznie większej liczby symulacji w przypadku metod MCMC niż w przypadku metod wykorzystujących próbę o niezależnych elementach jak np. importance sampling. Stosowanie metod MCMC jest jednak uzasadnione możliwością ich dość łatwego zastosowania w szerokiej klasie problemów. Dotyczy to zwłaszcza zagadnień, które są trudne do rozwiązania za pomocą importance sampling.

8.2. Wybrane informacje o łańcuchach Markowa

Poniżej zostaną przedstawione wybrane informacje o łańcuchach Markowa, które dalej będą przydatne do omówienia metod MCMC.

Łańcuchy Markowa są przykładem procesów losowych (por. roz. 4) określonych na parametrze czasu, który jest zbiorem przeliczalnym. Łańcuch Markowa jest więc sekwencją zmiennych losowych $\mathbf{X}^{(t)}$, $t = 0, 1, 2, \dots$, spełniających warunek:

$$P\left(\mathbf{X}^{(t+1)} \in A \mid \mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(t)}\right) = P\left(\mathbf{X}^{(t+1)} \in A \mid \mathbf{x}^{(t)}\right), \quad (8.1)$$

gdzie A jest zdarzeniem losowym, a $\mathbf{x}^{(t)}$, $t = 0, 1, 2, \dots$, są realizacjami $\mathbf{X}^{(t)}$, $t = 0, 1, 2, \dots$. Własność (8.1) jest często wyrażana poglądowo następującymi słowami: *przyszłość nie zależy od przeszłości, gdy znana jest teraźniejszość*.

Jednym ze sposobów definiowania łańcuchów Markowa jest określenie tzw. *prawdopodobieństwa przejścia* między jego stanami. Jeżeli zmienne losowe tworzące łańcuch Markowa są ciągle to wygodnie jest wykorzystać w tym celu tzw. *jądro przejścia* K (ang. *transition kernel*), będące gęstością warunkowego rozkładu prawdopodobieństwa kolejnego stanu procesu pod warunkiem aktualnego stanu procesu:

$$P\left(\mathbf{X}^{(t+1)} \in A \mid \mathbf{x}^{(t)}\right) = \int_A K\left(\mathbf{x}^{(t)}, \mathbf{x}^{(t+1)}\right) d\mathbf{x}^{(t+1)}. \quad (8.2)$$

Mówimy, że rozkład prawdopodobieństwa \mathcal{F} , o gęstości f , jest *niezmienniczy* względem jądra przejścia K i związanego z nim łańcucha Markowa, jeżeli spełniony jest następujący warunek

$$\int_A f(\mathbf{y}) d\mathbf{y} = \int_{\mathbb{R}^n} \int_A K(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) d\mathbf{x} d\mathbf{y}. \quad (8.3)$$

Rozkład niezmienniczy może być również postrzegany jako rozkład stacjonarny. Jeżeli bowiem \mathcal{F} jest rozkładem niezmienniczym oraz $\mathbf{X}^{(t)} \sim \mathcal{F}$ to $\mathbf{X}^{(t+n)} \sim \mathcal{F}$ dla dowolnego n . Rozkład prawdopodobieństwa stanów łańcucha Markowa jest, więc niezależny od parametru t , co oznacza, że łańcuch jest *ściśle stacjonarny*. Zwróćmy uwagę, że tak zdefiniowane pojęcie stacjonarności ma nieco węższy sens niż pojęcie słabej stacjonarności (4.7) określonej dla pól losowych (por. [110]).

Przyjmując oznaczenie $K^1(x, A)$ dla jądra jednego przejścia łańcucha Markowa - jądro dla n przejść można zdefiniować za pomocą wzoru:

$$K^n(\mathbf{x}, A) = \int_{\mathbb{R}^n} K^{n-1}(\mathbf{x}, \mathbf{y}) K(\mathbf{y}, A) d\mathbf{y}. \quad (8.4)$$

Rozkład prawdopodobieństwa $\mathbf{X}^{(0)}$ określa się jako *rozkład początkowy* łańcucha Markowa.

Za pomocą jądra n -przejsć (8.4) można zdefiniować warunkowy rozkład prawdopodobieństwa $\mathbf{X}^{(n)}$ pod warunkiem $\mathbf{X}^{(0)} = \mathbf{x}^{(0)}$:

$$P\left(\mathbf{X}^{(n)} \in A | \mathbf{x}^{(0)}\right) = K^n\left(\mathbf{x}^{(0)} | A\right). \quad (8.5)$$

Rozkład niezmienniczy \mathcal{F} jest *rozkładem równowagi* (ang. *equilibrium distribution* lub *limiting distribution*) łańcucha Markowa, jeżeli dla prawie każdego $\mathbf{x}^{(0)}$ zachodzi zbieżność:

$$\lim_{n \rightarrow \infty} K^n\left(\mathbf{x}^{(0)}, A\right) = P_{\mathcal{F}}(A), \quad (8.6)$$

gdzie $P_{\mathcal{F}}(A)$ jest prawdopodobieństwem zdarzenia A określonym przez rozkład \mathcal{F} .

Mówimy, że Łańcuch Markowa z rozkładem niezmienniczym \mathcal{F} jest *nieograniczony* (ang. *irreducible*), jeżeli dla dowolnego stanu początkowego istnieje większe od zera prawdopodobieństwo wejścia procesu do dowolnego zbioru, którego prawdopodobieństwo określone przez rozkład \mathcal{F} jest niezerowe.

Łańcuch Markowa określa się jako *okresowy* (ang. *periodic*), jeżeli istnieją obszary w przestrzeni stanów, do których proces ten może wchodzić jedynie w regularnych odstępach czasu. W przeciwnym przypadku łańcuch jest *nieokresowy* (ang. *aperiodic*).

Jeżeli \mathcal{F} jest rozkładem niezmienniczym nieokresowego i nieograniczonego łańcucha Markowa to \mathcal{F} jest jednoznacznie określonym rozkładem niezmienniczym tego procesu oraz jest jego rozkładem równowagi (dowód tego twierdzenia zamieszczono w [121]).

Najbardziej istotną konsekwencją zbieżności (8.6) dla zastosowania w metodach symulacyjnych jest zbieżność średniej $\frac{1}{N} \sum_{n=1}^N g(\mathbf{X}^{(n)})$ do wartości oczekiwanej $\mathbb{E}_{\mathcal{F}}[g(\mathbf{X})]$, gdzie g jest funkcją rzeczywistą, dla której istnieje powyższa wartość oczekiwana. Warunki, jakie są konieczne, aby łańcuch Markowa był ergodyczny, a więc aby możliwe było oszacowanie wartości oczekiwanej na podstawie jego realizacji, zostaną omówione szczegółowo dla algorytmu Metropolis-Hastings (por. 8.3)

Uzupełnieniem pojęcia ergodyczności, które w przypadku łańcuchów Markowa jest traktowane jako uogólnienie prawa wielkich liczb, jest wprowadzenie centralnych twierdzeń granicznych określających rozkład wartości średnich funkcji realizacji. Aby podać jedno z twierdzeń granicznych konieczne jest jeszcze zdefiniowanie pewnej ważnej własności łańcuchów Markowa.

Łańcuch Markowa $\mathbf{X}^{(t)}$ jest określany jako *odwracalny*, jeżeli rozkład prawdopodobieństwa $\mathbf{X}^{(t+1)}$ pod warunkiem $\mathbf{X}^{(t+2)} = \mathbf{x}^{(t+2)}$ jest identyczny jak rozkład $\mathbf{X}^{(t+1)}$ pod warunkiem $\mathbf{X}^{(t)} = \mathbf{x}^{(t)}$. Zobaczmy, że własność ta jest naturalną cechą procesów generowanych przez algorytm Metropolis-Hastings (por. 8.3).

Jeżeli łańcuch Markowa jest nieograniczony, nieokresowy, odwracalny oraz spełniony jest następujący warunek:

$$0 < \gamma_g^2 = \text{Var} \left(g \left(\mathbf{X}^{(0)} \right) \right) + 2 \sum_{k=1}^{\infty} \text{Cov} \left[g \left(\mathbf{X}^{(0)} \right), g \left(\mathbf{X}^{(k)} \right) \right] < +\infty \quad (8.7)$$

to

$$\frac{1}{\sqrt{N}} \sum_{n=1}^N \left(g \left(\mathbf{X}^{(n)} \right) - \mathbb{E}_{\mathcal{F}} [g(\mathbf{X})] \right) \rightsquigarrow \mathcal{N} \left(0, \gamma_g^2 \right), \quad (8.8)$$

gdzie \rightsquigarrow oznacza zbieżność według dystrybuant. Dowód tego twierdzenia można znaleźć w pracach [96, 121].

Metody MCMC opierają się na wykorzystaniu jądra przejścia o rozkładzie granicznym, z którego jest pożądana próba losowa. Przy czym dobierając odpowiednie jądro przejścia korzysta się z faktu, że rozkład stacjonarny jest również rozkładem granicznym.

8.3. Algorytm Metropolis-Hastings (M-H)

Podstawowym przykładem metod MCMC jest *algorytm Metropolis-Hastings* (M-H). Jak wspomniano wynikiem działania algorytmów tego typu jest realizacja łańcucha Markowa, którego rozkładem stacjonarnym jest rozkład celu. Realizacja algorytmu M-H polega na generowaniu elementów z tzw. *rozkładu pomocniczego*, a następnie dołączaniu ich do realizacji łańcucha na podstawie kryterium związanego z rozkładem celu. Wykorzystanie rozkładu pomocniczego jest jedną z cech decydujących o atrakcyjności algorytmu M-H. Jako rozkład pomocniczy można bowiem przyjąć rozkład, z którego łatwo otrzymać realizacje, a więc uprościć rozwiązanie zadania - przynajmniej w zakresie generowania liczb losowych. Omawiana metoda również pozwala na otrzymanie próby losowej z rozkładów spełniających bardzo słabe wymagania dotyczące regularności funkcji gęstości prawdopodobieństwa. Algorytm ten można więc stosować w szerokiej klasie zagadnień. Ponadto dla zastosowania w analizie niezawodności istotne jest to, że gęstość prawdopodobieństwa rozkładu celu może być znana jedynie z dokładnością do stałej normalizującej.

Jednym z istotnych zagadnień teorii procesów Markowa jest badanie istnienia rozkładów niezmienniczych łańcuchów określonych przez dane jądro przejścia. W teorii algorytmów MCMC rozwiązuje się problem odwrotny: poszukuje się jądra przejścia, którego rozkładem niezmienniczym jest rozkład celu.

Algorytm M-H wykorzystuje jądro przejścia następującej postaci:

$$K(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}, \mathbf{y}) + r(\mathbf{x})\delta_x(\mathbf{y}), \quad (8.9)$$

gdzie $p(\mathbf{x}, \mathbf{x}) = 0$, δ_x jest masą Diraca w \mathbf{x} oraz

$$r(\mathbf{x}) = 1 - \int_{\mathbf{R}^n} p(\mathbf{x}, \mathbf{y}) d\mathbf{y} \quad (8.10)$$

jest prawdopodobieństwem pozostania łańcucha punkcie \mathbf{x} . Ponieważ możliwe jest, że $r(\mathbf{x}) \neq 0$ to całka $p(\mathbf{x}, \mathbf{y})$ po \mathbf{y} nie musi być równa 1. Jeżeli funkcja $p(\mathbf{x}, \mathbf{y})$ spełnia *warunek odwracalności* (ang. *reversibility condition*):

$$f(\mathbf{x})p(\mathbf{x}, \mathbf{y}) = f(\mathbf{y})p(\mathbf{y}, \mathbf{x}) \quad (8.11)$$

to f jest niezmienniczą gęstością prawdopodobieństwa jądra K zdefiniowanego wzorem 8.9 [121]. Twierdzenie to można wykazać za pomocą poniższych przekształceń:

$$\begin{aligned} \int_{\mathbf{R}^n} \int_A K(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) d\mathbf{x} d\mathbf{y} &= \int_{\mathbf{R}^n} \int_A p(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) d\mathbf{x} d\mathbf{y} + \int_{\mathbf{R}^n} \int_A r(\mathbf{x}) \delta_x(\mathbf{y}) f(\mathbf{x}) d\mathbf{x} d\mathbf{y} \\ &= \int_{\mathbf{R}^n} \int_A p(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) d\mathbf{y} d\mathbf{x} + \int_A r(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{R}^n} \int_A p(\mathbf{y}, \mathbf{x}) f(\mathbf{y}) d\mathbf{y} d\mathbf{x} + \int_A r(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_A (1 - r(\mathbf{y})) f(\mathbf{y}) d\mathbf{y} + \int_A r(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ &= \int_A f(\mathbf{y}) d\mathbf{y}. \end{aligned} \quad (8.12)$$

Warunek odwracalności (8.11) interpretuje się jako wymaganie, aby gęstość prawdopodobieństwa przejścia z \mathbf{x} , otrzymanego z rozkładu f , do \mathbf{y} (lewa strona równania (8.11)) była równa gęstości prawdopodobieństwa przejścia z \mathbf{y} , otrzymanego z rozkładu prawdopodobieństwa f , do \mathbf{x} .

Podstawą działania algorytmu M-H jest określenie funkcji p , która spełnia warunek odwracalności dla danego rozkładu celu. Rozwiązanie tego zadania rozpoczyna się od przyjęcia rozkładu pomocniczego, z którego generowane są propozycje kolejnych stanów łańcucha. Ponieważ próba losowa otrzymana poprzez algorytm MCMC jest łańcuchem Markowa to rozkład pomocniczy w ogólności może zależeć od bieżącego stanu procesu. Gęstość tego rozkładu oznaczmy przez $q(\mathbf{y}|\mathbf{x})$, zachodzi oczywiście

$\int_{\mathbf{R}^n} q(\mathbf{y}|\mathbf{x})d\mathbf{x} = 1$. Jeżeli gęstość q spełnia warunek odwracalności (8.11) dla wszystkich \mathbf{x} i \mathbf{y} to za funkcję p można przyjąć gęstość q . Zazwyczaj jednak tak nie jest. Na przykład, dla pewnych \mathbf{x} , \mathbf{y} może zachodzić nierówność:

$$q(\mathbf{y}|\mathbf{x})f(\mathbf{x}) > q(\mathbf{x}|\mathbf{y})f(\mathbf{y}), \quad (8.13)$$

co oznacza, że łańcuch przechodzi ze stanu \mathbf{x} do stanu \mathbf{y} zbyt często, a ze stanu \mathbf{y} do stanu \mathbf{x} zbyt rzadko. Zauważmy, że "równowagę" procesu można przywrócić zmniejszając liczbę przejść z \mathbf{x} do \mathbf{y} poprzez wprowadzenie *prawdopodobieństwa wykonania przejścia* $\rho(\mathbf{x}, \mathbf{y}) \leq 1$ [52]. Można więc zaproponować funkcję p następującej postaci:

$$p_{MH}(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}, \mathbf{y}). \quad (8.14)$$

Ponieważ żąda się, aby funkcja p_{MH} spełniała szczegółowy warunek równowagi to powinna zachodzić równość:

$$q(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}, \mathbf{y})f(\mathbf{x}) = q(\mathbf{x}|\mathbf{y})\rho(\mathbf{y}, \mathbf{x})f(\mathbf{y}). \quad (8.15)$$

Załóżmy, że q spełnia (8.13), więc liczba przejść z \mathbf{y} do \mathbf{x} jest zbyt mała. Sensownym jest zatem przyjęcie $\rho(\mathbf{y}, \mathbf{x})$ równego 1, gdyż nie zmniejsza ono liczby przejść ze stanu \mathbf{y} do \mathbf{x} . Po uwzględnieniu takiego założenia, równanie (8.15) przyjmuje postać $q(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}, \mathbf{y})f(\mathbf{x}) = q(\mathbf{x}|\mathbf{y})f(\mathbf{y})$, która pozwala określić prawdopodobieństwa wykonania przejścia z \mathbf{x} do \mathbf{y} :

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{q(\mathbf{x}|\mathbf{y})f(\mathbf{y})}{q(\mathbf{y}|\mathbf{x})f(\mathbf{x})}. \quad (8.16)$$

Oczywiście, jeżeli ma miejsce sytuacja odwrotna do (8.13) to przyjmujemy $\rho(\mathbf{x}, \mathbf{y})$ równe 1 i odpowiednie $\rho(\mathbf{y}, \mathbf{x})$. A więc wprowadzając prawdopodobieństwa $\rho(\mathbf{x}, \mathbf{y})$ i $\rho(\mathbf{y}, \mathbf{x})$, które "równoważą" wyrażenie (8.15) zapewniamy się, że funkcja (8.14) spełnia szczegółowy warunek równowagi. Podsumowując: funkcja p_{MH} spełnia szczegółowy warunek równowagi, jeżeli dla danej gęstości pomocniczej prawdopodobieństwo przejścia jest następującej postaci:

$$\rho(\mathbf{x}, \mathbf{y}) = \begin{cases} \min \left\{ \frac{f(\mathbf{y})q(\mathbf{x}|\mathbf{y})}{f(\mathbf{x})q(\mathbf{y}|\mathbf{x})}, 1 \right\}, & \text{jeżeli } f(\mathbf{x})q(\mathbf{y}|\mathbf{x}) \neq 0 ; \\ 1, & \text{w p.p. .} \end{cases} \quad (8.17)$$

Prawdopodobieństwo, że proces zdefiniowany przez algorytm M-H pozostanie w stanie \mathbf{x} można teraz wyrazić jako:

$$r(\mathbf{x}) = 1 - \int_{\mathbf{R}^n} q(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}, \mathbf{y})d\mathbf{y}. \quad (8.18)$$

Ostatecznie jądro przejścia algorytmu M-H jest postaci:

$$K_{MH}(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{x}, \mathbf{y})q(\mathbf{y}|\mathbf{x}) + \left(1 - \int_{\mathbf{R}^n} q(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}, \mathbf{y})d\mathbf{y}\right) \delta_x(\mathbf{y}), \quad (8.19)$$

która jest szczególnym przypadkiem (8.9). Gęstość prawdopodobieństwa w przypadku, której zdefiniowano p_{MH} tak, aby spełniała dla niej szczegółowy warunek równowagi, jest niezmienniczą gęstością prawdopodobieństwa jądra K_{MH} .

Jako podsumowanie powyższych rozważań przedstawimy schemat implementacji algorytmu M-H. Realizacja algorytmu składa się z następujących kroków:

1. Przyjąć punkt startowy $\mathbf{x}^{(0)}$.
2. Powtarzać dla sekwencji $t = 1, 2, \dots, n$:
 - Wygenerować $\mathbf{Y}^{(t)} \sim q(\mathbf{y}|\mathbf{x}^{(t)})$ oraz $u \sim \mathcal{U}(0, 1)$.
 - Przyjąć

$$\mathbf{X}^{(t+1)} = \begin{cases} \mathbf{Y}^{(t)}, & \text{jeżeli } u \leq \rho(\mathbf{x}^{(t)}, \mathbf{Y}^{(t)}); \\ \mathbf{x}^{(t)}, & \text{w p.p.} \end{cases} \quad (8.20)$$

3. Wynikiem działania algorytmu jest następująca realizacja łańcucha Markowa $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}$.

Zauważmy, że ostateczna postać algorytmu zależy od zastosowanego pomocniczego rozkładu prawdopodobieństwa. Sposoby doboru tego rozkładu zostaną przedstawione w kolejnym podrozdziale.

Istotną cechą realizacji łańcucha Markowa, będącej wynikiem działania algorytmu M-H, jest powtarzanie się stanów procesu, a więc występowanie zdarzeń $\{\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)}\}$. Co ciekawe własność ta jest warunkiem dostatecznym, aby łańcuchy Markowa generowane przez algorytm M-H były nieokresowe [96]. Prawdopodobieństwo powtórzenia stanu jest niezerowe, kiedy

$$P\left[f\left(\mathbf{X}^{(t)}\right)q\left(\mathbf{Y}^{(t)}|\mathbf{X}^{(t)}\right) \leq f\left(\mathbf{Y}^{(t)}\right)q\left(\mathbf{X}^{(t)}|\mathbf{Y}^{(t)}\right)\right] < 1. \quad (8.21)$$

Tak, więc jeżeli rozkład pomocniczy spełnia warunek (8.21) to łańcuch Markowa powstały w wyniku działania algorytmu M-H jest nieokresowy.

Warunkiem dostatecznym, aby łańcuch generowany przez algorytm M-H był nieograniczony jest możliwość osiągnięcia przez proces dowolnego zbioru zawartego w σ -algebrze \mathcal{E} , w jednym kroku:

$$\forall_{(\mathbf{x}, \mathbf{y}) \in \mathcal{E} \times \mathcal{E}} \quad q(\mathbf{y}|\mathbf{x}) > 0. \quad (8.22)$$

Warunek (8.22) wydaje się być dość restrykcyjny, jednak okazuje się, że w praktyce może być dość łatwo spełniony; warunek (8.22) jest spełniony, jeżeli jako rozkład pomocniczy zostanie wykorzystany rozkład normalny - co jest dość częstą praktyką.

Nieco mniej restrykcyjny warunek, którego spełnienie zapewnia nieograniczoność i nieokresowość, został zaproponowany w pracy [97]. Łańcuch generowany przez algorytm M-H jest nieograniczony i nieokresowy, jeżeli pomocniczy rozkład prawdopodobieństwa przypisuje punktom położonym w sąsiedztwie aktualnego stanu procesu dodatnie prawdopodobieństwo to znaczy istnieją dodatnie liczby ϵ i δ takie, że:

$$q(\mathbf{y}|\mathbf{x}) > \epsilon, \quad \text{jeżeli} \quad |\mathbf{x} - \mathbf{y}| < \delta. \quad (8.23)$$

Jak widzimy rozkład q w (8.23) zapewnia możliwość przejścia łańcucha do punktu położonego w otoczeniu $\mathbf{x}^{(t)}$ o promieniu ograniczonym z dołu. Jeżeli ponadto rozkład \mathcal{F} jest taki, że $\rho(\mathbf{x}^{(t)}, y)$ jest dodatnie w tym otoczeniu to każdy podzbiór σ -algebry \mathcal{E} może być osiągnięty przez łańcuch Markowa w odpowiednio dużej liczbie kroków. Omawiane twierdzenie daje podstawy do zastosowania np. rozkładu jednorodnego określonego wewnątrz wielowymiarowej sfery, jako rozkładu pomocniczego.

Ponieważ nieograniczoność i nieokresowość są warunkami dostatecznymi, aby rozkład niezmienniczy był rozkładem równowagi łańcucha Markowa (por. 8.2) to (8.21) i (8.22) lub (8.22) są warunkami dostatecznymi dla zbieżności jądra n przejść algorytmu M-H do rozkładu niezmienniczego \mathcal{F} . Mówiąc precyzyjniej zachodzi tzw. zbieżność względem normy całkowitej zmienności (ang. *total variation norm*):

$$\lim_{n \rightarrow \infty} \left\| \int K_{MH}^n(\mathbf{x}, A) \mu(\mathbf{x}) d\mathbf{x} - P_{\mathcal{F}}(A) \right\|_{TV} = 0 \quad (8.24)$$

gdzie μ jest gęstością dowolnego rozkładu początkowego, a K_{MH}^n jest jądrem algorytmu M-H dla n przejść. Normę występującą we wzorze (8.24) definiuje się jako: $|P_{\mathcal{G}_1}(A) - P_{\mathcal{G}_2}(A)|_{TV} = \sup_A |P_{\mathcal{G}_1}(A) - P_{\mathcal{G}_2}(A)|$, gdzie $P_{\mathcal{G}_1}(A)$, $P_{\mathcal{G}_2}(A)$ są prawdopodobieństwami zdarzenia A określonymi przez rozkłady \mathcal{G}_1 oraz \mathcal{G}_2 .

Podstawowym estymatorem wartości oczekiwanej wykorzystującym realizację łańcucha generowanego przez algorytm M-H jest średnia wartości funkcji realizacji:

$$\bar{g}_n = \frac{1}{n - t_0} \sum_{t=t_0}^n g(\mathbf{X}^{(t)}). \quad (8.25)$$

Jeżeli spełnione są warunki dostateczne: generowany łańcuch jest nieokresowy i nieograniczony, a dla funkcji g zachodzi (8.7) to rozkład estymatora (8.25) dąży do rozkładu normalnego określonego przez (8.8). Ostatni dostateczny warunek istnienia granicy (8.8) - odwracalność łańcucha - jest własnością wszystkich procesów generowanych przez algorytmu M-H.

Początkowe elementy realizacji łańcucha (indeksy $t = 0, \dots, t_0 - 1$, tzw. ang. *burn in period*) zostały pominięte w oszacowaniu (8.25), gdyż uważa się, że ich rozkład prawdopodobieństwa jest obciążony wpływem przyjęcia dowolnego punktu startowego $\mathbf{x}^{(0)}$. Jest to typowa praktyka stosowana w przypadku metod MCMC.

8.4. Rodzaje algorytmu Metropolis-Hastings

Jedną z cech decydujących o atrakcyjności algorytmu M-H jest wykorzystanie rozkładu pomocniczego do generowania próby losowej z rozkładu celu. Jak wspomniano zbieżność algorytmu jest zapewniona dla dość szerokiej klasy rozkładów pomocniczych. Jednak z praktycznego punktu widzenia sens ma jedynie stosowanie rozkładów zapewniających, że odpowiednio duża część realizacji łańcucha Markowa znajdzie się w obszarze, gdzie jest skupiona masa prawdopodobieństwa rozkładu celu. Dobór właściwego dla danego problemu rozkładu pomocniczego ma decydujący wpływ na efektywność otrzymanego oszacowania.

Poniżej zostaną przedstawione podstawowe sposoby doboru rozkładu pomocniczego. Ze względu na łatwość tworzenia podejść hybrydowych bardzo trudne jest stworzenie klasyfikacji metod doboru rozkładu pomocniczego. Różnorodność tych metod chyba najpełniej, jak dotąd, została przedstawiona w [68].

8.4.1. Przypadkowe błędzenie

Naturalne podejście do realizacji algorytmu M-H opiera się na zastosowaniu rozkładu pomocniczego, który jest zależny od aktualnego stanu procesu. Bardzo popularna, a zarazem prosta realizacja tej metody polega na generowaniu propozycji kolejnego stanu łańcucha $\mathbf{Y}^{(t)}$ z rozkładu następującej postaci:

$$\mathbf{Y}^{(t)} = \mathbf{X}^{(t)} + \boldsymbol{\varepsilon}^{(t)}, \quad (8.26)$$

gdzie $\boldsymbol{\varepsilon}^{(t)}$ jest zaburzeniem losowym z rozkładem niezależnym od \mathbf{X}_t . W tym przypadku gęstość rozkładu pomocniczego można przedstawić jako

$$q(\mathbf{y}|\mathbf{x}) = \varrho_r(\mathbf{y} - \mathbf{x}), \quad (8.27)$$

gdzie ϱ_r jest gęstością prawdopodobieństwa zaburzenia $\boldsymbol{\varepsilon}$. Łańcuch Markowa związany z tak zdefiniowanym rozkładem pomocniczym określa się jako *przypadkowe błędzenie* (ang. *random walk*).

Najczęściej stosowane w praktyce rozkłady zaburzenia $\boldsymbol{\varepsilon}$ to rozkład normalny oraz rozkład jednorodny, określony w kuli ze środkiem w początku układu współrzędnych.

Wykorzystanie takich rozkładów prawdopodobieństwa zapewnia spełnienie warunku (8.23), który wystarcza aby istniał rozkład równowagi a proces był ergodyczny.

Zastosowanie symetrycznego rozkładu ϱ_r (tzn. $\varrho_r(\mathbf{x}) = \varrho_r(-\mathbf{x})$) prowadzi do tzw. *algorytmu Metropolisa* [79], który, zaproponowany w latach 50-tych XX wieku, był historycznie pierwszym algorytmem MCMC. W tym przypadku prawdopodobieństwo przejścia (8.17) upraszcza się do postaci

$$\rho(\mathbf{x}, \mathbf{y}) = \min \left\{ \frac{f(\mathbf{y})}{f(\mathbf{x})}, 1 \right\}. \quad (8.28)$$

8.4.2. Łańcuch niezależny

Omawiana implementacja algorytmu M-H jest przeciwieństwem przypadkowego błędzenia w tym sensie, że propozycja kolejnego stanu procesu $\mathbf{Y}^{(t)}$ jest generowana z rozkładu, który jest niezależny od stanu procesu:

$$q(\mathbf{y}|\mathbf{x}) = \varrho_i(\mathbf{y}). \quad (8.29)$$

Podejście jest określane jako *łańcuch niezależny* (ang. *independence chain* [121]).

W przypadku rozważanej implementacji prawdopodobieństwo przejścia (8.17) jest następującej postaci:

$$\rho(\mathbf{x}, \mathbf{y}) = \min \left\{ \frac{f(\mathbf{y})\varrho_i(\mathbf{x})}{f(\mathbf{x})\varrho_i(\mathbf{y})}, 1 \right\}. \quad (8.30)$$

Warto zauważyć, że chociaż propozycje kolejnych stanów procesu $\mathbf{Y}^{(t)}$ są niezależnymi zmiennymi losowymi to zależność prawdopodobieństwa przejścia od stanów łańcucha $\mathbf{X}^{(t)}$ wprowadza korelacje między kolejnymi elementami generowanej próby losowej.

Zbieżność tak zdefiniowanego algorytmu M-H zależy od własności rozkładu pomocniczego. Jeżeli gęstość ϱ_i spełnia warunek (8.22) to algorytm ten jest zbieżny.

Efektywność omawianej odmiany algorytmu M-H, podobnie jak w metodzie importance sampling, zależy od doboru odpowiedniej dla danego zagadnienia gęstości prawdopodobieństwa ϱ_i . W pracy [78] wykazano, że jeżeli istnieje stała M taka, że

$$f(\mathbf{x}) \leq M\varrho_i(\mathbf{x}), \quad \mathbf{x} \in \text{supp} f \quad (8.31)$$

to łańcuch generowany przez niezależny algorytm M-H jest *jednostajnie ergodyczny* (ang. *uniformly ergodic*), a więc zachodzi

$$\|K_{MH}^n(\mathbf{x}, A) - P_{\mathcal{F}}(A)\|_{TV} \leq 2 \left(1 - \frac{1}{M}\right)^n. \quad (8.32)$$

Wyrażenie $\text{supp}f$ we wzorze 8.31 oznacza dziedzinę gęstości prawdopodobieństwa rozkładu celu \mathcal{F} . Jak widzimy spełnienie warunku jednostajnej ergodyczności zapewnia szybką zbieżność rozkładu $\mathbf{X}^{(t)}$ do rozkładu równowagi, a tym samym zapewnia efektywne uzyskanie oszacowania wartości oczekiwanej.

8.5. Efektywna implementacja algorytmu Metropolis-Hastings

Algorytm M-H jest zbieżny, jeżeli rozkład pomocniczy spełnia, przedstawione w podrozdziale 8.3, niezbyt restrykcyjne warunki. Niestety o ile dość łatwo jest zapewnić samą zbieżność, to dobór rozkładu pomocniczego pozwalającego uzyskać zadowalającą szybkość zbieżności (np. 8.32) może okazać się trudnym problemem.

Praktyczne zastosowanie algorytmu M-H wymusza zastosowanie rozkładu pomocniczego, z którego jest łatwo generować liczby losowe. Dlatego też najczęściej stosowane są rozkład normalny i rozkład jednorodny. Efektywne wykorzystanie tych rozkładów wymaga ostrożnego doboru ich parametrów. Poniżej przedstawione zostaną zagadnienia związane z doбором parametrów rozkładu pomocniczego w przypadku łańcucha niezależnego i przypadkowego błędzenia.

8.5.1. Ocena efektywności algorytmu Metropolis-Hastings

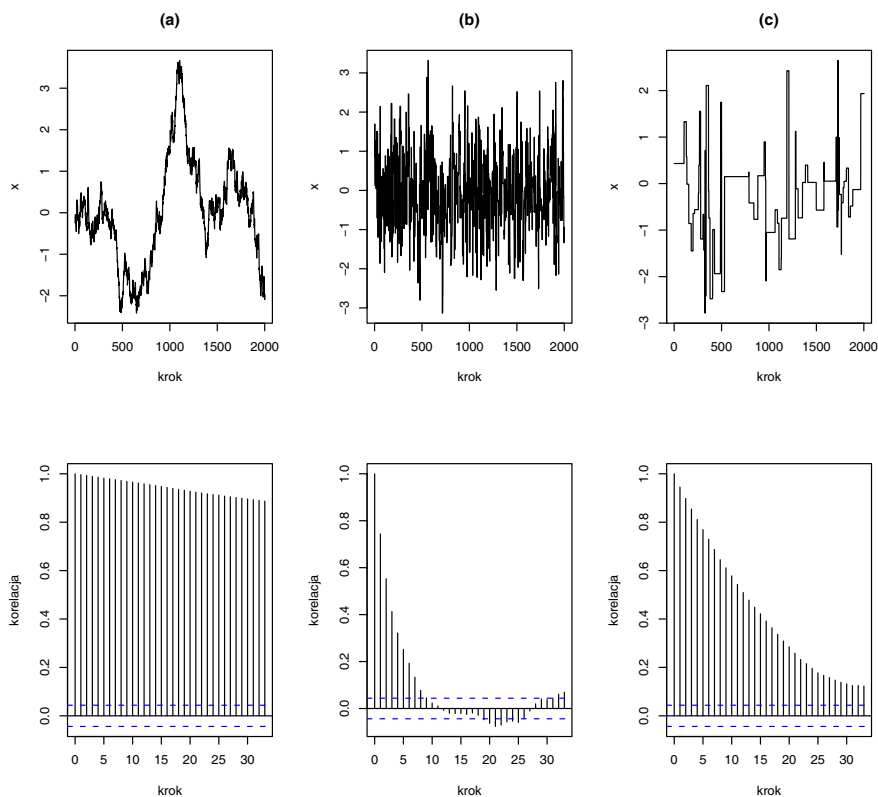
Porównanie efektywności algorytmów typu M-H wymaga określenia odpowiedniego kryterium. W tym celu wprowadza się tzw. *całkowity czas autokorelacji* (ang. *integrated autocorrelation time*), który dla całkowalnej z kwadratem funkcji g definiuje się jako

$$\tau_g = 1 + 2 \sum_{i=1}^{\infty} \text{Corr} \left(g \left(\mathbf{X}^{(0)} \right), g \left(\mathbf{X}^{(i)} \right) \right), \quad (8.33)$$

gdzie zakłada się, że $\mathbf{X}^{(0)}$ ma rozkład równowagi generowanego łańcucha. Jeżeli dla estymatora (8.25) wartości oczekiwanej g zachodzi centralne twierdzenie graniczne (8.8) to jego wariancję można oszacować za pomocą całkowitego czasu autokorelacji:

$$\text{Var}(\bar{g}_N) = \text{Var} \left(\frac{1}{N} \sum_{n=1}^N g \left(\mathbf{X}^{(n)} \right) \right) \approx \frac{\tau_g}{N} \text{Var} (g(\mathbf{X})). \quad (8.34)$$

Na podstawie tej zależności efektywność oszacowania \bar{g} określa się jako iloraz wariancji estymatora klasycznej metody Monte Carlo (2.41) uzyskanego z próby losowej o



RYSUNEK 8.1. Wykres realizacji przypadkowego błędzenia w zależności od doboru wariancji rozkładu pomocniczego: (a)-zbyt małej, (b)-optymalnej, (c)-zbyt dużej.

niezależnych elementach i estymatora korzystającego z algorytmu M-H:

$$\text{eff}_g = \frac{\frac{1}{N} \text{Var}(g(\mathbf{X}))}{\text{Var}(\bar{g}_n)} = \tau_g^{-1}, \quad (8.35)$$

Tak zdefiniowana miara efektywności jest oczywiście zależna od funkcji g . Jednak w pracach [43], [98], [99] wykazano, że przynajmniej w przypadku pewnej klasy problemów dla liczby wymiarów dążącej do nieskończoności wszystkie miary efektywności eff_g są sobie równoważne. Ponadto pokazano, że wielkością związaną z efektywnością

algorytmu jest tzw. *oczekiwany wskaźnik przejść* (ang. *expected acceptance ratio*) [43]. Zakładając, że łańcuch znajduje się w stanie stacjonarnym, parametr ten określa się w następujący sposób:

$$a = \int_{\mathbf{R}^d} \int_{\mathbf{R}^d} \rho(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) q(\mathbf{y}|\mathbf{x}) d\mathbf{x} d\mathbf{y}. \quad (8.36)$$

Alternatywnie, korzystając z ergodyczności, oczekiwany wskaźnik przejść można wyrazić jako długookresowy stosunek liczby kroków, w których łańcuch przeszedł do nowego stanu do całkowitej liczby kroków:

$$a = \lim_{n \rightarrow \infty} \frac{\#\{t : \mathbf{X}^{(t+1)} = \mathbf{Y}^{(t)}\}}{n}, \quad (8.37)$$

gdzie $\#$ oznacza liczbę elementów w zbiorze. Wyniki te, uzyskane w pracach [43, 98, 99], stanowią teoretyczne uzasadnienie stosowanej wcześniej, praktycznej zasady zgodnie, z którą parametry rozkładu pomocniczego powinny być dobrane tak, aby otrzymany łańcuch charakteryzował się odpowiednią częstością przejść.

Okazuje się, że uważane za efektywne algorytm niezależny i przypadkowe błędzenie charakteryzują się znacząco różnymi wartościami wskaźnika przejść. W przypadku algorytmu niezależnego wskaźnik przejść powinien być jak najwyższy, natomiast dla przypadkowego błędzenia parametr ten powinien mieć odpowiednią, na ogół niezbyt dużą wartość.

Zapewnienie maksymalnej wartości a w przypadku algorytmu niezależnego wymaga, aby gęstość pomocnicza ϱ_i była możliwie bliska gęstości rozkładu celu f [96]. Na przykład, jeżeli gęstość ϱ_i jest za bardzo lub zbyt mało skupiona w porównaniu do f , to można się spodziewać dużej zmienności prawdopodobieństwa przejścia (8.30), a tym samym małej wartości wskaźnika przejść.

Zupełnie inne wymagania odnośnie wartości wskaźnika przejść dotyczą efektywnej implementacji przypadkowego błędzenia. Jest tak, ponieważ gęstość pomocnicza tego algorytmu zależy od aktualnego stanu procesu. Wysoka liczba przejść nie koniecznie oznacza, że proces przemieszcza się w pożądanym sposób. Jeżeli $\mathbf{x}^{(t)}$ oraz $\mathbf{y}^{(t)}$ są blisko siebie to $f(\mathbf{x}^{(t)})$ i $f(\mathbf{y}^{(t)})$ są zazwyczaj prawie równe, a więc prawdopodobieństwo przejścia (8.28) jest bliskie 1. Jednak w tym przypadku zbieżność algorytmu jest bardzo wolna, gdyż ruchy procesu w przestrzeni stanów są za krótkie. Przypadek ten jest pokazany na rysunku 8.1(a). Jak widać korelacje pierwszego stanu łańcuchu zmniejszają się bardzo powoli skutkiem czego jest wysoka wartość całkowitego czasu autokorelacji.

Z drugiej strony, jeżeli wartość oczekiwana wskaźnika przejść jest mała oznacza to, że wartości $f(\mathbf{y}^{(t)})$ są znacznie mniejsze niż $f(\mathbf{x}^{(t)})$. W takim przypadku proces

wykonuje skoki w domenie f , propozycje nowych stanów często wpadają w ogony rozkładu, a łańcuch "blokuje się" w punktach o dużej gęstości prawdopodobieństwa (por. rys. 8.1(c)).

Tak, więc efektywne zastosowanie przypadkowego błędzenia wymaga dość precyzyjnego doboru gęstości pomocniczej do rozwiązywanego zagadnienia (rys. 8.1(b)). Przyjęcie zbyt skupionego rozkładu spowoduje zbieżność, a rozkład zbyt rozproszony spowoduje, że często będą proponowane punkty w mało istotnych obszarach, gdzie gęstość f jest mała.

8.5.2. Zalecana wartość wskaźnika przejść przypadkowego błędzenia

Przedmiotem prac [43, 98, 99] jest optymalizacja procesu generowania próby losowej z rozkładu prawdopodobieństwa o gęstości

$$f(\mathbf{x}) = \prod_{i=1}^d \psi(x_i) \quad (8.38)$$

za pomocą przypadkowego błędzenia z normalnym rozkładem pomocniczym

$$q_r(\mathbf{y}) = \varphi(\mathbf{y}, \mathbf{x}, \sigma^2 \mathbf{I}). \quad (8.39)$$

Zakłada się, że funkcje gęstości prawdopodobieństwa ψ we wzorze (8.38) są ciągłe. Jak wspomniano w poprzednim podrozdziale, w przypadku tak sformułowanego zadania dla dużej liczby wymiarów wszystkie miary efektywności (8.35) są sobie równoważne, a ponadto efektywność jest związana z wartością oczekiwanego wskaźnika przejść.

W [43] pokazano, że dla dużej liczby wymiarów d optymalna wartość odchylenia standardowego σ w (8.39) jest rzędu $\mathcal{O}(d^{-1/2})$, a optymalna wartość wskaźnika przejść a wynosi 0.234. Dla tak dobranych parametrów efektywność algorytmu (8.35) jest rzędu $\mathcal{O}(d^{-1})$ tak, więc liczba iteracji potrzebnych do otrzymania oszacowania o żądanej dokładności wzrasta liniowo wraz z liczbą wymiarów.

W pracy [43], na podstawie testów numerycznych, określono optymalne parametry omawianego sformułowania dla wymiarów $d = 1 \dots 10$. Wykazano, że przedstawione wcześniej wyniki asymptotyczne można w przybliżeniu przyjąć za prawdziwe już dla problemu określonego w sześciu wymiarach. W przypadku mniejszej liczby wymiarów bardziej korzystne jest zastosowanie rozkładu pomocniczego o większym odchyleniu standardowym. Dla zadania określonego w przestrzeni jednowymiarowej optymalna wartość σ wynosi 2.40, a towarzyszy jej wskaźnik przejść 0.441. W pracy [43] pokazano również, że efektywność algorytmu w punkcie optymalnym nie jest zbyt wrażliwa na wartość wskaźnika przejść. Dla wartości a między 0.15 a 0.40 algorytm osiąga 80%

maksymalnej efektywności. Tak, więc dobór parametrów rozkładu dający zbliżoną do optymalnej proporcję przejść powinien zapewnić akceptowalny poziom efektywności.

Zgodnie z przedstawionymi wynikami, w zastosowaniach praktycznych często dąży się do otrzymania przypadkowego błędzenia o wskaźniku przejść zbliżonym do $1/4$. Jednak ta praktyczna reguła powinna być stosowana z ostrożnością. W pracy [99] podano wstępne wyniki badań dotyczących zastosowania przypadkowego błędzenia do generowania próby losowej z rozkładu o nieciągłej gęstości prawdopodobieństwa. W tym przypadku nie obowiązuje jedno z istotnych założeń, na których opiera się wprowadzenie omawianej reguły. Podawana w [99] optymalna wartość wskaźnika przejść dla nieciągłej gęstości rozkładu celu wynosi 0.13, a optymalne odchylenie standardowe jest rzędu $\mathcal{O}(d^{-1})$.

8.6. Adaptacyjne algorytmy Metropolis-Hastings

Jak pokazano w podrozdziale 8.5 dobór rozkładu pomocniczego ma kluczowe znaczenie dla efektywnej implementacji algorytmu M-H. Najbardziej bezpośrednim podejściem do zagadnienia właściwej parametryzacji rozkładu pomocniczego jest metoda prób i błędów. Można wykonać serię wstępnych symulacji i na ich podstawie tak dobrać parametry rozkładu pomocniczego, aby generowany łańcuch miał odpowiednią wartość wskaźnika przejść. Ta metoda jest proponowana w niektórych publikacjach. Jednak takie podejście ze względu na wymagane nakłady obliczeniowe i brak możliwości algorytmizacji jest mało zadowalające. Pożądane jest, więc opracowanie metod pozwalających w systematyczny sposób optymalizować parametry rozkładu pomocniczego na potrzeby rozwiązania konkretnego zadania.

"Automatyczne" metody wykorzystujące informacje pozyskane w czasie działania algorytmu MCMC nazywane są *adaptive Markov chain Monte Carlo*. W ostatnich latach zaprezentowano kilka sposobów rozwiązania tego zagadnienia [5, 88, 117]; niektóre z nich zostaną omówione w dalszej części tego podrozdziału.

Budowa adaptacyjnych algorytmów MCMC nie jest prostym zadaniem. Wymaga, bowiem rozwiązania przynajmniej dwóch trudnych problemów. Wszystkie dowody dotyczące zbieżności algorytmu M-H są przeprowadzone przy założeniu, że parametry rozkładu pomocniczego są stałe. Oczywiście w przypadku algorytmu adaptacyjnego pożądana jest możliwość uaktualniania tych parametrów w trakcie jego działania. Drugim problemem, jaki wiąże się z adaptacją rozkładu pomocniczego jest to, że uaktualnianie parametrów na podstawie poprzednich iteracji wprowadza zależność między dalekimi stanami procesu, a w rezultacie otrzymany proces stochastyczny nie jest już łańcuchem Markowa. W literaturze spotyka się dwa sposoby budowy algorytmów adaptacyjnych: przeprowadzenie adaptacji we wstępnej fazie działania algorytmu i prowadzenie symu-

lacji dla ustalonych parametrów lub wykorzystanie zmiennego rozkładu pomocniczego, co wiąże się jednak z koniecznością analizy własności takiego algorytmu.

8.6.1. Kryteria adaptacji

W celu sformułowania zagadnienia optymalizacji algorytmu M-H wygodnie jest posłużyć się funkcją celu określoną na parametrach rozkładu pomocniczego. Funkcja ta powinna być związana ze stochastycznymi miarami efektywności algorytmu, których przykłady zostały omówione w podrozdziale 8.5.1. Punkt optymalny takiej funkcji celu musi oczywiście odpowiadać parametrom maksymalizującym efektywność algorytmu.

W wielu przypadkach funkcję pozwalającą optymalizować algorytm M-H można przedstawić w ogólnej postaci zaproponowanej w pracy [5]:

$$\eta(\mathbf{v}) = \mathbb{E} [\tilde{\eta}(\mathbf{v}, (\mathbf{x}, \mathbf{y}))] = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \tilde{\eta}(\gamma, (\mathbf{x}, \mathbf{y})) f(\mathbf{x}) q_{\mathbf{v}}(\mathbf{y}|\mathbf{x}) d\mathbf{x} d\mathbf{y}, \quad (8.40)$$

gdzie \mathbf{v} jest parametrem lub grupą parametrów rozkładu pomocniczego $q_{\mathbf{v}}$, a $\tilde{\eta}$ jest pewną funkcją związaną z wybranymi własnościami statystycznymi generowanego łańcucha Markowa.

Jednym z przykładów funkcji celu wykorzystywanej w optymalizacji algorytmów M-H jest *funkcja wymuszająca wartość oczekiwanego wskaźnika przejść* (ang. *coerced acceptance ratio*). W podrozdziale 8.5.1 omówiono związek oczekiwanego wskaźnika przejść z efektywnością algorytmu. Przypomnijmy, że w przypadku optymalnego błędzenia istnieje pewna optymalna wartość tego parametru. Żądaną wartość wskaźnika przejść można wymusić przyjmując parametry algorytmu, które minimalizują następującą funkcję [88]:

$$\eta_{CAR}(\gamma) = \left(\int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \rho_{\mathbf{v}}(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) q_{\mathbf{v}}(\mathbf{y}|\mathbf{x}) d\mathbf{x} d\mathbf{y} - a^* \right)^2, \quad (8.41)$$

gdzie $\rho_{\mathbf{v}}$ jest prawdopodobieństwem przejścia - w ogólności zależnym od parametru \mathbf{v} - natomiast a^* jest wymaganą wartością wskaźnika przejść.

Wartości parametrów optymalizujące algorytm M-H można również wyznaczyć maksymalizując kwadrat różnicy kolejnych wartości funkcji g (ang. *Expected Squared Function Variation*), której wartość oczekiwana podlega oszacowaniu:

$$\eta_{ESFV}(\gamma) = \mathbb{E} \left[\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)}) \right)^2 \right]. \quad (8.42)$$

Jak zostanie to wykazane, maksymalizacja h_{ESFV} jest równoważna minimalizacji $\text{Corr} \left(g(\mathbf{X}^{(t)}), g(\mathbf{X}^{(t+1)}) \right)$, co prowadzi do minimalizacji wariancji (8.34) estymatora (8.25) wartości oczekiwanej funkcji g .

Zauważmy, że w przypadku łańcucha będącego wynikiem algorytmu M-H wariancja funkcji $g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})$ ma następujące własności:

$$\begin{aligned} \text{Var}\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right) &= \mathbb{E}\left[\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right) \right. \\ &\quad \left. - \mathbb{E}\left[g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right]\right)^2] \\ &= \mathbb{E}\left[\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right)^2\right], \end{aligned} \quad (8.43)$$

gdzie przekształcenie (8.43) jest prawdziwe po przyjęciu założenia, że w stanie stacjonarnym łańcucha $\mathbf{X}^{(t)}$ i $\mathbf{X}^{(t+1)}$ mają takie same rozkłady prawdopodobieństwa, a więc $\mathbb{E}[g(\mathbf{X}^{(t+1)})] - \mathbb{E}[g(\mathbf{X}^{(t)})] = 0$. Następnie przyrównując prawą stronę równości

$$\begin{aligned} \text{Var}\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right) &= \text{Var}\left(g(\mathbf{X}_t)\right) + \text{Var}\left(g(\mathbf{X}^{(t+1)})\right) \\ &\quad - 2\text{Cov}\left[g(\mathbf{X}^{(t)}), g(\mathbf{X}^{(t+1)})\right] \\ &= 2\left(1 - \text{Corr}\left(g(\mathbf{X}^{(t)}), g(\mathbf{X}^{(t+1)})\right)\right) \\ &\quad \times \text{Var}\left(g(\mathbf{X}^{(t)})\right), \end{aligned} \quad (8.44)$$

która jest prawdziwa dzięki założeniu stacjonarności łańcucha, do prawej strony równania (8.43) otrzymujemy wyrażenie:

$$\begin{aligned} \eta_{ESFV}(\gamma) &= \mathbb{E}\left[\left(g(\mathbf{X}^{(t+1)}) - g(\mathbf{X}^{(t)})\right)^2\right] \\ &= 2\left(1 - \text{Corr}\left(g(\mathbf{X}^{(t)}), g(\mathbf{X}^{(t+1)})\right)\right) \text{Var}\left(g(\mathbf{X}^{(t)})\right). \end{aligned} \quad (8.45)$$

Wzór (8.45) pokazuje, że maksymalizacja η_{ESFV} jest równoważna minimalizacji korelacji pierwszego rzędu funkcji kolejnych stanów łańcucha $\text{Corr}(h(\mathbf{X}^{(t)}), h(\mathbf{X}^{(t+1)}))$. Minimalizację korelacji pierwszego rzędu można traktować jako zgrubne ujęcie minimalizacji całkowitego czasu autokorelacji (8.33). Jeżeli autokorelacje wyższego rzędu maleją monotonicznie to zmniejszenie autokorelacji pierwszego rzędu prowadzi do zwiększenia efektywności algorytmu określonej wzorem (8.35).

Funkcja η_{ESFV} jest uogólnieniem, zaproponowanego w pracy [88], *oczekiwane kwadratu długości przejścia* (ang. *expected squared jumping distance* (ESJD)). W przypadku jednowymiarowym wielkość ta jest określana jako $ESJD(\mathbf{v}) = \mathbb{E}\left[|X^{(t+1)} - X^{(t)}|^2\right]$, a w przypadku wielowymiarowym wykorzystuje się kwadrat od-

ległości Mahalanobisa:

$$ESJD(\mathbf{v}) = \mathbb{E} \left[\left(\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)} \right)^T \mathbf{C}^{-1} \left(\mathbf{X}^{(t+1)} - \mathbf{X}^{(t)} \right) \right], \quad (8.46)$$

gdzie \mathbf{C} jest macierzą kowariancji rozkładu równowagi. Taki dobór optymalizowanej funkcji jest wynikiem przyjęcia jako "naturalnej" miary efektywności algorytmu (8.35) wariancji oszacowania wartości oczekiwanej zmiennej losowej \mathbf{X} . Warto zwrócić uwagę, że w przypadku rozkładu celu określonego w wielu wymiarach, którego macierz kowariancji \mathbf{C} nie jest znana oszacowanie $ESJD$ może stwarzać problemy numeryczne. Z tego względu wydaje się, że wykorzystanie funkcji danej wzorem (8.42) jest wygodniejsze.

Funkcja η_{ESFV} posiada cechy, które ułatwiają jej zastosowanie do optymalizacji algorytmu M-H. Wartość oczekiwana we wzorze (8.42) można przedstawić w następujący sposób:

$$\begin{aligned} & \mathbb{E} \left[\left(g \left(\mathbf{x}^{(t+1)} \right) - g \left(\mathbf{x}^{(t)} \right) \right)^2 \right] = \\ & = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \left(g \left(\mathbf{x}^{(t+1)} \right) - g \left(\mathbf{x}^{(t)} \right) \right)^2 K \left(\mathbf{x}^{(t)}, \mathbf{x}^{(t+1)} \right) f \left(\mathbf{x}^{(t)} \right) d\mathbf{x}^{(t+1)} d\mathbf{x}^{(t)} \\ & = \int_{\mathbf{R}^n} \int_{\mathbf{R}^n} \left(g \left(\mathbf{x}^{(t+1)} \right) - g \left(\mathbf{x}^{(t)} \right) \right)^2 \rho \left(\mathbf{x}^{(t)}, \mathbf{x}^{(t+1)} \right) \\ & \quad q \left(\mathbf{x}^{(t+1)} | \mathbf{x}^{(t)} \right) f \left(\mathbf{x}^{(t)} \right) d\mathbf{x}^{(t+1)} d\mathbf{x}^{(t)}. \end{aligned} \quad (8.47)$$

Powyższe przekształcenia są możliwe, gdyż

$$\left(g \left(\mathbf{x}^{(t+1)} \right) - g \left(\mathbf{x}^{(t)} \right) \right)^2 \left(1 - r \left(\mathbf{x}^{(t+1)} \right) \right) \delta_{\mathbf{x}^{(t)}} \left(\mathbf{x}^{(t+1)} \right) = 0. \quad (8.48)$$

Eliminacja wyrazu związanego z pozostawianiem łańcucha w tym samym stanie pozwala na pominięcie w obliczeniach wyrażenia $r(\mathbf{x}^{(t+1)})$, którego wyznaczenie wymagałoby obliczenia dodatkowej całki (por. (8.18)). Funkcja $\tilde{\eta}$ we wzorze (8.40) związana z η_{ESFV} ma więc postać

$$\tilde{\eta}_{ESFV} \left(\mathbf{v}, \left(\mathbf{x}^{(t)}, \mathbf{x}^{(t+1)} \right) \right) = \left(g \left(\mathbf{x}^{(t+1)} \right) - g \left(\mathbf{x}^{(t)} \right) \right)^2 \rho_{\mathbf{v}} \left(\mathbf{x}^{(t)}, \mathbf{x}^{(t+1)} \right). \quad (8.49)$$

Inne przykłady funkcji η zostały przedstawione w pracach [5, 88].

8.6.2. Dostosowanie rozkładu pomocniczego

Wykorzystanie początkowych symulacji w celu optymalizacji parametrów pomocniczego rozkładu prawdopodobieństwa jest naturalnym sposobem implementacji adaptacyjnych algorytmów MCMC (przykłady można znaleźć między innymi w pracach [41, 49]). Poniżej przedstawiono algorytm zaproponowany w niepublikowanej pracy [88].

Koncepcja omawianego algorytmu jest bardzo prosta. Początkowa faza symulacji, która zazwyczaj i tak jest pomijana przy oszacowaniu wartości oczekiwanej, wykorzystywana jest do optymalizacji parametrów rozkładu pomocniczego. Generowanie łańcucha Markowa rozpoczyna się od wykorzystania rozkładu pomocniczego z przyjętymi początkowymi wartościami parametrów \mathbf{v} . Po każdym n krokach uaktualnia się wartości parametrów optymalizując oszacowanie funkcji (8.40). Obserwując zbieżność sekwencji $\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(j)}$ uzyskuje się oszacowanie optymalnych parametrów rozkładu pomocniczego $\hat{\mathbf{v}}^*$. W drugiej fazie algorytmu za pomocą zoptymalizowanego rozkładu pomocniczego generowane są stany łańcucha, które później wykorzystuje się do oszacowania wartości oczekiwanej. Implementacja algorytmu obejmuje, więc wykonanie następujących kroków:

1. Przyjąć początkowe wartości parametrów rozkładu pomocniczego $\mathbf{v}^{(0)}$ - na przykład wykorzystując wskazówki przedstawione w podrozdziale 8.5. Korzystając z rozkładu pomocniczego o parametrach \mathbf{v} wygenerować, składającą się z n_0 elementów, realizację łańcucha Markowa $\{\mathbf{X}_0^{(1)}, \mathbf{X}_0^{(2)}, \dots, \mathbf{X}_0^{(n_0)}\}$ oraz związaną z nią sekwencję proponowanych punktów przejścia $\{\mathbf{Y}_0^{(1)}, \mathbf{Y}_0^{(2)}, \dots, \mathbf{Y}_0^{(n_0)}\}$.
2. Na podstawie przyjętej funkcji η wyznaczyć $\mathbf{v}^{(1)}$ - przybliżenie parametrów \mathbf{v}^* , korzystając z oszacowania wartości oczekiwanej (8.40) za pomocą wyrażenia:

$$\bar{\eta}(\mathbf{v}|\mathbf{v}^{(0)}) = \frac{\sum_{i=1}^{n_0} \tilde{\eta}(\mathbf{v}, \mathbf{X}_0^{(i)}, \mathbf{Y}_0^{(i)}) w_{\mathbf{v}|\mathbf{v}^{(0)}}(\mathbf{X}_0^{(i)}, \mathbf{Y}_0^{(i)})}{\sum_{i=1}^{n_0} w_{\mathbf{v}|\mathbf{v}^{(0)}}(\mathbf{X}_0^{(i)}, \mathbf{Y}_0^{(i)})}, \quad (8.50)$$

gdzie

$$w_{\mathbf{v}|\mathbf{v}^{(0)}}(\mathbf{x}, \mathbf{y}) = \frac{q_{\mathbf{v}}(\mathbf{x}, \mathbf{y})}{q_{\mathbf{v}^{(0)}}(\mathbf{x}, \mathbf{y})}, \quad (8.51)$$

a $q_{\mathbf{v}^{(\cdot)}}$ jest gęstością pomocniczego rozkładu prawdopodobieństwa dla parametrów $\mathbf{v}^{(\cdot)}$. Zwiększyć wartość licznika iteracji: $j = 1$.

3. Wygenerować n_j elementowy łańcuch Markowa $\{\mathbf{X}_j^{(1)}, \mathbf{X}_j^{(2)}, \dots, \mathbf{X}_j^{(n_j)}\}$ oraz towarzyszący mu zbiór proponowanych punktów przejścia $\{\mathbf{Y}_j^{(1)}, \mathbf{Y}_j^{(2)}, \dots, \mathbf{Y}_j^{(n_j)}\}$.

Wyznaczyć przybliżenie $\mathbf{v}^{(j+1)}$, korzystając z przybliżenia funkcji (8.40) postaci:

$$\bar{\eta}(\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)}) = \frac{\sum_{k=0}^j \sum_{i=1}^{n_k} \tilde{\eta}(\mathbf{v}, \mathbf{X}_k^{(i)}, \mathbf{Y}_k^{(i)}) w_{\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)}}(\mathbf{X}_k^{(i)}, \mathbf{Y}_k^{(i)})}{\sum_{k=0}^j \sum_{i=1}^{n_k} w_{\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)}}(\mathbf{X}_k^{(i)}, \mathbf{Y}_k^{(i)})}, \quad (8.52)$$

gdzie

$$w_{\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)}}(\mathbf{x}, \mathbf{y}) = \frac{q_{\mathbf{v}}(\mathbf{x}, \mathbf{y})}{\sum_{k=0}^j \frac{n_k}{n} q_{\mathbf{v}^{(k)}}(\mathbf{x}, \mathbf{y})}, \quad (8.53)$$

w którym $n = \sum_{k=0}^j n_k$. Zwiększyć wartość licznika iteracji: $j = j + 1$.

4. Sprawdzić przyjęte kryterium zbieżności ciągu oszacowań parametrów rozkładu $\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(j+1)}$. Jeżeli kryterium to jest spełnione, przyjąć $\hat{\mathbf{v}}^* = \mathbf{v}^{(j+1)}$, w przeciwnym przypadku powtórzyć krok 3.
5. Oszacować żadaną wartość oczekiwaną korzystając z realizacji łańcucha Markowa, którą uzyskano za pomocą algorytmu M-H z rozkładem pomocniczym o gęstości $q_{\mathbf{v}^*}$.

Warunki zapewniające zbieżność przedstawionego algorytmu zostały przedstawione w pracy [88]. Estymator $\bar{\eta}(\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)})$ jest zbieżny, jeżeli łańcuch Markowa wykorzystywany w oszacowaniu jest nieograniczony i nieokresowy (spełnione są warunki zbieżności oszacowania wartości oczekiwanej omówione w podrozdziale 8.3). Ciąg oszacowań $\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(j)}$ jest zbieżny do wartości \mathbf{v}^* - optymalizującej funkcję $\eta(\mathbf{v})$, jeżeli funkcje $\bar{\eta}(\mathbf{v}|\mathbf{v}^{(j)}, \dots, \mathbf{v}^{(0)})$ i $\eta(\mathbf{v})$ są wypukłe i dwukrotnie różniczkowalne względem \mathbf{v} . W ogólnym przypadku badanie wypukłości estymatora (8.52) jest trudnym zadaniem. W pracy [88] wykazano zbieżność powyższej metody dla algorytmu M-H z normalnym rozkładem pomocniczym i funkcją η w postaci oczekiwanego kwadratu długości przejścia (8.46).

W wyrażeniach (8.50) i (8.52) wartości oczekiwane względem rozkładu $q_{\mathbf{v}}$ przybliżane są za pomocą wspomnianego już estymatora reciprocal importance sampling (7.15).

Estymator (8.52) wykorzystuje próbę losową otrzymaną z kilku rozkładów prawdopodobieństwa. Przyjmuje się założenie, że gęstość rozkładu prawdopodobieństwa tej próby losowej jest gęstością rozkładu mieszanego [53]:

$$q(\mathbf{x}) = \sum_{k=1}^j \lambda_k q_k(\mathbf{z}), \quad (8.54)$$

gdzie q_k są gęstościami rozkładów wykorzystanych do otrzymania próby, a λ_k spełniają warunek $\sum_{k=1}^J \lambda_k = 1$. Współczynniki λ_k są proporcjonalne do liczby elementów, które zostały wygenerowane z rozkładu o gęstości q_k .

8.7. Wyznaczanie stałych normalizujących

Przypomnijmy, że gęstość prawdopodobieństwa (2.54) optymalizującą zastosowanie importance sampling w analizie niezawodności można przedstawić w ogólnej postaci

$$\psi(\mathbf{x}) = \frac{\phi(\mathbf{x})}{c}, \quad (8.55)$$

gdzie

$$c = \int_{R^n} \phi(\mathbf{x}) d\mathbf{x} \quad (8.56)$$

jest stałą normalizującą funkcję ϕ . W przypadku analizy niezawodności ϕ jest gęstością losowych parametrów konstrukcji, którą obcięto do obszaru awarii, a c jest prawdopodobieństwem awarii.

Oszacowanie prawdopodobieństwa awarii za pomocą algorytmów MCMC opiera się na pomysłach wygenerowania próby losowej z rozkładu prawdopodobieństwa (2.54). Taka próba losowa może być otrzymana za pomocą algorytmu M-H, gdyż w tym celu wymagana jest jedynie znajomość funkcji ϕ (w analizie niezawodności $\phi = I_{\Omega_f} f(\mathbf{x})$). Okazuje się jednak, że dysponowanie tylko taką próbą losową nie wystarcza do oszacowania prawdopodobieństwa awarii. W tym celu konieczne jest wyznaczenie stałej normalizującej c , a więc w przypadku analizy niezawodności oszacowanie prawdopodobieństwa awarii. Problem oszacowania stałej normalizującej gęstość rozkładu prawdopodobieństwa jest spotykany w niektórych zagadnieniach probabilistyki oraz w fizyce statystycznej [42]. Poniżej przedstawimy kilka ogólnych metod służących do oszacowania stałych normalizujących, które można znaleźć w literaturze, a w dalszej części podejmiemy próbę zastosowania przynajmniej części z nich do oszacowania prawdopodobieństwa awarii.

8.7.1. Estymator wykorzystujący przybliżenie Laplace'a

Zaproponowany w pracy [65] *estymator Laplace'a-Metropolis'a* (ang. Laplace-Metropolis estimator) do oszacowania stałej normalizującej wykorzystuje przybliżenie Laplace'a całki gęstości prawdopodobieństwa. Parametry tego przybliżenia są wyznaczone na podstawie realizacji łańcucha Markowa, otrzymanej za pomocą algorytmu

M-H. W rozważanym przypadku, ten rodzaj aproksymacji może być również postrzegany jako przybliżenie rozkładu równowagi łańcucha za pomocą rozkładu normalnego.

Teoretycznie, aproksymacja Laplace'a jest prawdziwa dla ciągłych, ograniczonych funkcji z wyraźnym maksimum globalnym. Jednak w praktyce metoda ta może być stosowana w przypadku funkcji gęstości prawdopodobieństwa, które nie spełniają tych wszystkich założeń. Istotne jest, aby rozkład równowagi był *jednomodowy*, a więc jego gęstość powinna być skupiona w pobliżu jednego punktu.

Stałą normalizującą gęstość f przybliża się za pomocą następującego wyrażenia:

$$\bar{c}_L = \frac{f(\bar{\mathbf{x}}^*)}{\varphi(\bar{\mathbf{x}}^*, \bar{\mathbf{x}}^*, \bar{\mathbf{C}})} = (2\pi)^{\frac{d}{2}} |\bar{\mathbf{C}}|^{\frac{1}{2}} f(\bar{\mathbf{x}}^*), \quad (8.57)$$

gdzie f jest nieznormalizowaną gęstością rozkładu równowagi łańcucha, $\bar{\mathbf{x}}^*$ jest oszacowaniem $\arg \max f(\mathbf{x})$, a $|\bar{\mathbf{C}}|$ jest wyznacznikiem empirycznej macierzy kowariancji próby losowej uzyskanej za pomocą algorytmu M-H.

Oszacowanie $\bar{\mathbf{x}}^*$, w którym występuje maksimum funkcji f może być wyznaczone poprzez wybranie punktu z realizacji łańcucha, dla którego wartość f jest największa, bądź też poprzez wyznaczenie wartości oczekiwanej lub mediany próby losowej. Zwrócimy jeszcze uwagę, że występująca w wzorze (8.57) macierz kowariancji jest aproksymacją hesjanu gęstości f w punkcie $\arg \max f(\mathbf{x})$.

8.7.2. Odwrotna metoda importance sampling

Ponieważ stała normalizująca jest określona w postaci całki (8.56) to można ją wyznaczyć korzystając z metody importance sampling (por. 2.3.7). W ogólnym przypadku estymator jest postaci

$$\bar{c}_{IS} = \frac{1}{N} \sum_{n=1}^N \frac{\phi(\mathbf{X}^{(n)})}{q(\mathbf{X}^{(n)})}, \quad (8.58)$$

gdzie q jest gęstością rozkładu prawdopodobieństwa próby losowej $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$. Oczywiście oszacowanie prawdopodobieństwa awarii za pomocą importance sampling (2.49) można postrzegać jako oszacowanie stałej normalizującej funkcji $\phi(\mathbf{x}) = I_{\Omega_f}(\mathbf{x})f(\mathbf{x})$. Efektywność oszacowania (8.58) zależy od doboru gęstości $q(\mathbf{x})$, która powinna spełniać wymagania analogiczne do opisanych w podrozdziale 2.3.7.

Podobny do (8.58) estymator stałej normalizującej można określić dla próby losowej otrzymanej za pomocą algorytmu M-H. Niech $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}\}$ będzie łańcuchem Markowa z rozkładem równowagi o gęstości proporcjonalnej do funkcji ϕ . Oszacowanie stałej normalizującej za pomocą *odwrotnej metody importance sampling* (ang. *recipro-*

cal importance sampling) określa się jako:

$$\bar{c}_{RIS} = \left(\frac{1}{N} \sum_{n=1}^N \frac{s(\mathbf{X}^{(n)})}{\phi(\mathbf{X}^{(n)})} \right)^{-1}, \quad (8.59)$$

gdzie $s(\cdot)$ jest funkcją gęstości prawdopodobieństwa. Jeżeli s ma odpowiednio lekkie ogony, a zwłaszcza jeżeli zachodzi $\int_{\mathbf{R}^n} s^2/\phi < \infty$, wtedy estymator (8.59) zachowuje się poprawnie (por. [86]). Zauważmy, że wymaganie to jest odwrotne niż w przypadku *importance sampling*, gdzie gęstość prawdopodobieństwa, z której generowana jest próba losowa powinna mieć cięższe ogony niż gęstość ϕ . Natomiast podobnie jak w *importance sampling* wariancja \bar{c}_{RIS} jest mała, jeżeli s jest dobrym przybliżeniem ψ .

Zaletą odwrotnej metody *importance sampling* jest możliwość wykorzystania otrzymanej realizacji łańcucha Markowa do zdefiniowania rozkładu s . Oczywiście najprostszym podejściem jest określenie rozkładu normalnego o wartości oczekiwanej i macierzy kowariancji wyznaczonych na podstawie analizowanej próby losowej. Należy jednak pamiętać, że wykorzystanie tych samych elementów próby do określenia gęstości s i wyznaczenia średniej (8.59) może prowadzić do uzyskania obciążonego oszacowania stałej normalizującej.

8.7.3. Bridge sampling

Niech dane będą dwie gęstości prawdopodobieństwa $p_i(\mathbf{x})$, $i = 1, 2$, określone na mających część wspólną dziedzinach \mathcal{D}_i . Gęstości te znane są z dokładnością do stałej normalizującej: $p_i(\mathbf{x}) = q_i(\mathbf{x})/c_i$. Celem metody *bridge sampling*, którą zaproponowano w pracy [77], jest oszacowanie ilorazu stałych normalizujących c_1/c_2 na podstawie prób losowych z rozkładów prawdopodobieństwa o gęstościach p_i . Próby te można oczywiście otrzymać za pomocą algorytmu M-H.

Niech dana będzie funkcja $\alpha(\mathbf{x})$, określona na $\mathcal{D}_1 \cap \mathcal{D}_2$ - wspólnej części dziedzin q_1 i q_2 , która spełnia następujący warunek

$$0 < \left| \int_{\mathcal{D}_1 \cap \mathcal{D}_2} \alpha(\mathbf{x}) q_1(\mathbf{x}) q_2(\mathbf{x}) d\mathbf{x} \right| < \infty. \quad (8.60)$$

Jeżeli α spełnia warunek (8.60) to zbiór $\mathcal{D}_1 \cap \mathcal{D}_2$ jest niezdegenerowany, ponieważ

$$\int_{\mathcal{D}_1 \cap \mathcal{D}_2} \alpha(\mathbf{x}) p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} > 0. \quad (8.61)$$

Zauważmy, że wartość całki we wzorze (8.61) jest miarą "nakładania" się rozkładów prawdopodobieństwa o gęstościach p_1 i p_2 . Jeżeli funkcja α spełnia warunek (8.61) to

prawdziwa jest równość

$$\frac{\int_{\mathcal{D}_2} q_1(\mathbf{x})\alpha(\mathbf{x})p_2(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{D}_1} q_2(\mathbf{x})\alpha(\mathbf{x})p_1(\mathbf{x})d\mathbf{x}} = \frac{c_1}{c_2} \cdot \frac{\int_{\mathcal{D}_1 \cap \mathcal{D}_2} \alpha(\mathbf{x})p_1(\mathbf{x})p_2(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{D}_1 \cap \mathcal{D}_2} \alpha(\mathbf{x})p_1(\mathbf{x})p_2(\mathbf{x})d\mathbf{x}}, \quad (8.62)$$

z której wynika tożsamość będąca podstawą metody bridge sampling:

$$r = \frac{\mathbb{E}_{p_2} [q_1(\mathbf{x})\alpha(\mathbf{x})]}{\mathbb{E}_{p_1} [q_2(\mathbf{x})\alpha(\mathbf{x})]}, \quad (8.63)$$

gdzie $r = c_1/c_2$. Odpowiednio dobierając funkcję α możemy otrzymać różne metody szacowania stałych normalizujących. Jeżeli przyjmiemy

$$\alpha = q_2^{-1} \quad (8.64)$$

to otrzymamy wyrażenie będące podstawą oszacowania stałej normalizującej za pomocą metody importance sampling

$$r_{IS} = \mathbb{E}_{p_2} \left[\frac{q_1(\mathbf{x})}{q_2(\mathbf{x})} \right]. \quad (8.65)$$

Zakładając, że $c_2 = 1$ a $q_2 = p_2$ otrzymuje się estymator (8.58). W przypadku, kiedy w pełni znana jest gęstość p_1 ($c_1 = 1$) to wyrażenie (8.65) jest odwrotnością estymatora reciprocal importance sampling (8.59). W pracy [77] oprócz (8.64) zaproponowano następujące postaci funkcji α :

- *Funkcja geometryczna* $\alpha = (q_1 q_2)^{-1/2}$, która implikuje formułę

$$r_G = \frac{\mathbb{E}_{p_2} [(q_1/q_2)^{1/2}]}{\mathbb{E}_{p_1} [(q_2/q_1)^{1/2}]}. \quad (8.66)$$

Zauważmy, że w porównaniu do metody importance sampling, omawiane oszacowanie jest bardziej stabilne numerycznie. Wyrażenia postaci $(q_1/q_2)^{1/2}$ mają, bowiem bardziej zrównoważone wartości niż odpowiednie wagi metody importance sampling. Zastosowanie pierwiastka kwadratowego w wyrażeniu (8.66) zapewnia ponadto, że $(q_1/q_2)^{1/2}$ oraz $(q_2/q_1)^{1/2}$ są całkowalne z kwadratem względem odpowiednio q_2 i q_1 .

- *Rodzina wykładnicza*, $\alpha(k, A) = [q_1^{1/k} + (Aq_2)^{1/k}]^{-k}$, gdzie $A > 0$ i $k > 0$, wraz z wyrażeniem określającym wartość ilorazu stałych normalizujących

$$r_P(k, A) = \frac{\mathbb{E}_{p_2} [1 + (Aq_2/q_1)^{1/k}]^{-k}}{\mathbb{E}_{p_1} [(q_1/q_2)^{1/k} + A^{1/k}]^{-k}}. \quad (8.67)$$

Zastosowanie tego sformułowania wymaga ostrożnego doboru parametrów A oraz k w zależności od stosowanej metody symulacji [77].

- *Funkcja stała*, $\alpha = 1$ implikująca wyrażenie

$$r_C = \frac{\mathbb{E}_{p_2}[q_1]}{\mathbb{E}_{p_1}[q_2]}. \quad (8.68)$$

Rozważmy jeszcze funkcję α w następującej postaci:

$$\alpha(\mathbf{x}) = \frac{q_{12}(\mathbf{x})}{q_1(\mathbf{x})q_2(\mathbf{x})}, \quad \mathbf{x} \in \mathcal{D}_1 \cap \mathcal{D}_2, \quad (8.69)$$

gdzie $q_{12}(\mathbf{x})$ jest pewną nieznormalizowaną gęstością prawdopodobieństwa, określoną na zbiorze $\mathcal{D}_1 \cap \mathcal{D}_2$. Oznaczenie q_{12} zostało wprowadzone w celu podkreślenia, że pożądane jest, aby gęstość q_{12} była "pomiędzy" gęstościami q_1 oraz q_2 . Podstawiając (8.69) do wzoru (8.63) otrzymujemy

$$r_{12} = \frac{c_1}{c_2} = \frac{c_{12}/c_2}{c_{12}/c_1} = \frac{\mathbb{E}_{p_2}[q_{12}(\mathbf{x})/q_2(\mathbf{x})]}{\mathbb{E}_{p_1}[q_{12}(\mathbf{x})/q_1(\mathbf{x})]}. \quad (8.70)$$

W oszacowaniu uzyskanym na podstawie (8.70), zamiast wykorzystać (8.68) bezpośrednio do wyznaczenia c_1/c_2 , uzyskuje się c_{12}/c_2 oraz c_{12}/c_1 , a następnie przez dzielenie eliminuje się c_{12} . Takie postępowanie ma na celu zwiększenie efektywności oszacowania. Wiemy, że efektywność metody importance sampling zależy od tego jak "podobne" do siebie są gęstości q_1 i q_2 . Wprowadzając odpowiednio dobraną funkcję q_{12} , będącą swego rodzaju "mostem" między q_1 i q_2 (stąd nazwa metody bridge sampling), możemy uzyskać lepsze oszacowanie c_1/c_2 za pomocą (8.70) niż dzięki (8.68). Wyrażenie (8.63) należy więc postrzegać jako uogólnienie (8.70), gdzie wprowadzono funkcję α , której celem jest zwiększenie efektywności oszacowania.

8.7.4. Annealed Importance Sampling (AIS) oraz Linked Importance Sampling (LIS)

Przedstawiona w poprzednim podrozdziale idea poprawy oszacowania ilorazu stałych normalizujących poprzez wprowadzenie rozkładu pośredniego może być rozwijana. W niektórych przypadkach gęstości prawdopodobieństwa q_1 i q_2 mogą się różnić tak bardzo, że nawet wykorzystanie optymalnego rozkładu pośredniego q_{12} (rozkład ten jest zdefiniowany w pracy [77]) nie zapewni zadowalającej efektywności estymatora uzyskanego na podstawie (8.70). W takim przypadku można wielokrotnie zastosować metodę bridge sampling wprowadzając sekwencję pośrednich rozkładów prawdopodobieństwa o gęstościach $q_k(\mathbf{x})$, $k = 1, \dots, K$, gdzie q_1 i q_K odpowiadają odpowiednio

q_1 i q_2 . Dla rozkładów ϱ_k , z których generowana jest próba losowa, ponadto definiuje się pośrednie gęstości prawdopodobieństwa $\varrho_{k+1/2}(\mathbf{x})$, $k = 1, \dots, K-1$, służące do wyznaczenia oszacowania. Iloraz $r = c_1/c_K$ stałych normalizujących gęstości ϱ_1 i ϱ_K można przedstawić za pomocą ilorazów stałych normalizujących rozkłady pośrednie:

$$\frac{c_1}{c_K} = \prod_{k=1}^K \frac{c_{k+1/2}/c_k}{c_{k+1/2}/c_{k+1}} = \prod_{k=1}^K \frac{\mathbb{E}_{\varrho_k}[\varrho_{k+1/2}/\varrho_k]}{\mathbb{E}_{\varrho_{k+1}}[\varrho_{k+1/2}/\varrho_{k+1}]} \quad (8.71)$$

Jako rozwinięcie pomysłu, który zastosowano w (8.70) powyższe wyrażenie określa "most o wielu przęsłach". Estymator korzystający z (8.70) jest nieobciążony; tej własności nie posiada jednak estymator otrzymany na podstawie (8.71) [85]. Problem ten został jednak rozwiązany w algorytmach *annealed importance sampling* (AIS) [84] oraz *linked importance sampling* (LIS) [85].

Metody AIS i LIS do oszacowania ilorazu stałych normalizujących wykorzystują sekwencję rozkładów pośrednich. Motywacją do opracowania tych algorytmów było wykorzystanie pomysłu, który jest podstawą metod *simulated annealing*. Metody te wykorzystują rozkłady pośrednie do generowania prób losowych z rozkładów wielomodowych [44] lub do optymalizacji funkcji z wieloma minimami lokalnymi [60]. A więc AIS i LIS, przy wykorzystaniu odpowiednich rozkładów pośrednich, powinny wykazywać dobrą efektywność w zadaniach, gdzie konieczne jest oszacowanie stałych normalizujących rozkładów wielomodowych.

W celu efektywnej implementacji metody AIS wygodnie jest przyjąć rozkład o gęstości ϱ_1 , z którego możliwe jest otrzymanie niezależnej próby losowej. W przypadku rozkładów o gęstościach $\varrho_2, \dots, \varrho_{K-1}$ zakłada się, że można dla nich uzyskać próbę losową za pomocą algorytmu M-H o jądrach przejścia odpowiednio K_2, \dots, K_{K-1} .

Oszacowanie ilorazu stałych normalizujących za pomocą algorytmu AIS wymaga wygenerowania sekwencji punktów $\mathbf{X}_1^{(n)}, \dots, \mathbf{X}_{K-1}^{(n)}$, $n = 1, \dots, N$, z kolejnych rozkładów prawdopodobieństwa, zgodnie z następującym algorytmem:

- Wygenerować $\mathbf{X}_1^{(n)} \sim \varrho_1(\mathbf{x})$
- Wygenerować $\mathbf{X}_2^{(n)} \sim K_2(\mathbf{X}_1^{(n)}, \mathbf{x})$
- \vdots
- Wygenerować $\mathbf{X}_{K-1}^{(n)} \sim K_{K-1}(\mathbf{X}_{K-2}^{(n)}, \mathbf{x})$

Iloraz stałych normalizujących gęstości ϱ_1 i ϱ_K przybliża się za pomocą estymatora:

$$\bar{r}_{AIS} = \frac{1}{N} \sum_{n=1}^N \frac{\varrho_2(\mathbf{X}_1^{(n)})}{\varrho_1(\mathbf{X}_1^{(n)})} \frac{\varrho_3(\mathbf{X}_2^{(n)})}{\varrho_2(\mathbf{X}_2^{(n)})} \cdots \frac{\varrho_K(\mathbf{X}_{K-1}^{(n)})}{\varrho_{K-1}(\mathbf{X}_{K-1}^{(n)})} \quad (8.72)$$

Pomimo, że pojedyncze przejścia M-H zazwyczaj nie wystarczają do osiągnięcia rozkładu równowagi to oszacowanie \bar{r}_{AIS} jest nieobciążone i zbieżne z rosnącym N , jeżeli nie istnieją obszary, w których ϱ_k jest równe zero, a ϱ_{k+1} , jest niezerowe, $k = 1, \dots, K - 1$ [84].

W przypadku metody LIS wykorzystuje się sekwencję rozkładów o gęstościach $\varrho_1, \dots, \varrho_K$, z których generuje się próby losowe za pomocą algorytmu M-H. Zakłada się, że generowanie próby losowej z rozkładu ϱ_1 można rozpocząć od pojedynczego punktu otrzymanego niezależnie od ϱ_1 . Generowanie prób losowych o gęstości ϱ_k , $k > 1$, rozpoczyna się wybierając tzw. *stan łączący* (ang. *link state*), \mathbf{x}_{k-1*k} z próby o gęstości ϱ_{k-1} . Dla każdego ϱ_k zaczynając od stanu łączącego i wykorzystując jądro przejścia K_k generuje się realizację łańcucha Markowa złożoną z n_k elementów. Stany łączące wybiera się na podstawie odpowiednio dobranych rozkładów pośrednich ϱ_{k*k+1} , które są określone za pomocą gęstości ϱ_k i ϱ_{k+1} . W pracy [85] proponuje się, aby wyznaczyć M oszacowań ilorazów stałych normalizujących $\bar{r}_{LIS}^{(1)}, \dots, \bar{r}_{LIS}^{(M)}$, a następnie je uśrednić. Takie podejście zapewnia uzyskanie stabilniejszego przybliżenia w przypadku rozkładów wielomodowych. Każde z oszacowań $\bar{r}_{LIS}^{(m)}$ jest otrzymywane poprzez wykonanie następującego algorytmu:

1. Wylosować wskaźnik ν_1 z rozkładu jednorodnego określonego na zbiorze liczb całkowitych $0, \dots, n_1$. Zainicjować algorytm przypisując stanowi $\mathbf{x}_1^{(\nu_1)}$ próby z rozkładu o gęstości ϱ_1 pewną niezależnie wylosowaną wartość.
2. Dla $k = 1, \dots, K$ wygenerować $n_k + 1$ elementowe próby losowe o gęstościach ϱ_k poprzez wykonanie następujących kroków:
 - Jeżeli $k > 1$ to wylosować wskaźnik ν_k z rozkładu jednorodnego, określonego na $0, \dots, n_k$ i przyporządkować $\mathbf{x}_k^{(\nu_k)}$ wartość stanu przejścia \mathbf{x}_{k-1*k} .
 - Rozpoczynając od stanu $\mathbf{x}_k^{(\nu_k)}$, za pomocą jądra przejścia $K_k \left(\mathbf{x}_k^{(t)}, \mathbf{x}_k^{(t+1)} \right)$ algorytmu M-H, wygenerować "do przodu" stany realizacji łańcucha Markowa $\mathbf{x}_k^{(t)}$ o indeksach $t = \nu_k + 1, \dots, n_k$. Następnie korzystając z $K_k \left(\mathbf{x}_k^{(t)}, \mathbf{x}_k^{(t-1)} \right)$, zgodnie z malejącymi indeksami $t = \nu_k - 1, \dots, 0$, wygenerować stany realizacji łańcucha Markowa "wstecz".
 - Jeżeli $k < K$, to wylosować wskaźnik μ_k ze zbioru $\{0, \dots, n_k\}$ zgodnie

rozkładem prawdopodobieństwa

$$P\left(\mu_i | \mathbf{x}_k^{(\mu_i)}\right) = \frac{\frac{\varrho_{k^{**}k+1}\left(\mathbf{x}_k^{(\mu_i)}\right)}{\varrho_k\left(\mathbf{x}_k^{(\mu_i)}\right)}}{\sum_{i=0}^{n_k} \frac{\varrho_{k^{**}k+1}\left(\mathbf{x}_k^{(\mu_i)}\right)}{\varrho_k\left(\mathbf{x}_k^{(\mu_i)}\right)}}, \quad (8.73)$$

a następnie przyporządkować stanowi przejścia $\mathbf{x}_{k^{**}k+1}$ stan realizacji łańcucha $\mathbf{x}_k^{(\mu_k)}$.

3. Zgodnie z rozkładem jednorodnym przyporządkować μ_K wartość ze zbioru $\{0, \dots, n_K\}$. (Ten krok algorytmu nie ma wpływu na oszacowanie, ale jest konieczny do dowodu poprawności algorytmu [85].)
4. Oszacować wartość ilorazu stałych normalizujących dla tego przebiegu algorytmu:

$$\bar{r}_{LIS}^{(m)} = \prod_{k=1}^{K-1} \frac{\frac{1}{n_k + 1} \sum_{i=0}^{n_k} \frac{\varrho_{k^{**}k+1}\left(\mathbf{x}_k^{(i)}\right)}{\varrho_k\left(\mathbf{x}_k^{(i)}\right)}}{\frac{1}{n_{k+1} + 1} \sum_{i=0}^{n_{k+1}} \frac{\varrho_{k^{**}k+1}\left(\mathbf{x}_{k+1}^{(i)}\right)}{\varrho_{k+1}\left(\mathbf{x}_{k+1}^{(i)}\right)}}. \quad (8.74)$$

Po M -krotnym wykonaniu powyższego algorytmu otrzymuje się końcowe oszacowanie ilorazu stałych normalizujących:

$$\bar{r}_{LIS} = \frac{1}{M} \sum_{m=1}^M \bar{r}_{LIS}^{(m)}. \quad (8.75)$$

Istotną cechą metody LIS jest to, że przechodząc między rozkładami ϱ_k i ϱ_{k+1} stan łączący $\mathbf{x}_{k^{**}k+1}$ jest wybierany losowo z realizacji $\{\mathbf{x}_k^{(0)}, \dots, \mathbf{x}_k^{(n_k)}\}$. Stan łączący może być więc postrzegany jako element próby losowej związanej zarówno z gęstością ϱ_k jak i ϱ_{k+1} . W pracy [85] wykazano, że estymator (8.74) jest nieobciążony.

8.8. Oszacowanie prawdopodobieństwa awarii za pomocą metod MCMC

We wstępie do rozdziału 7 zaproponowano oszacowanie prawdopodobieństwa awarii poprzez wyznaczenie stałej normalizującej funkcji $q_f(\mathbf{x}) = I_{\Omega_f}(\mathbf{x})f(\mathbf{x})$ (por. (2.54)) na podstawie próby losowej otrzymanej za pomocą MCMC. Po omówieniu w poprzednich podrozdziałach koniecznych w tym celu metod, przedstawimy dokładniej realizację proponowanego podejścia.

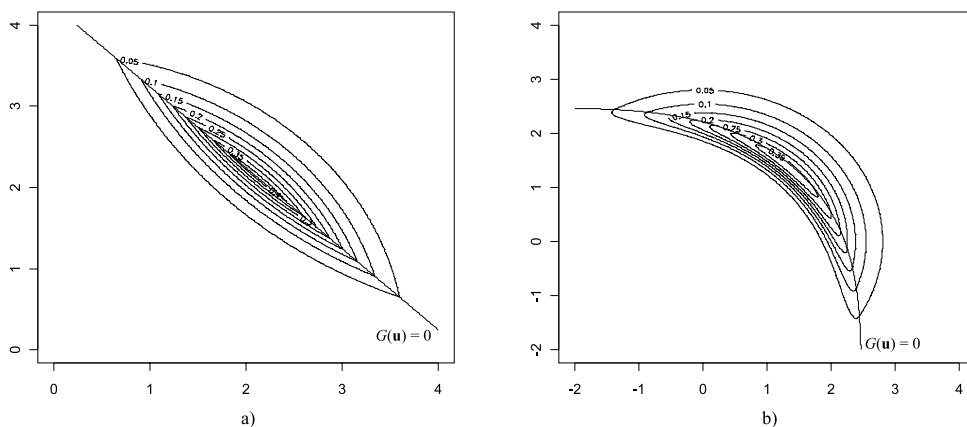
Jednym z problemów związanych z zastosowaniem algorytmu M-H do otrzymania próby losowej z rozkładu o gęstości proporcjonalnej do q_f jest określenie punktu startowego o niezerowej gęstości prawdopodobieństwa. Prawdopodobieństwo przejścia (8.17) jest bowiem określone w punktach, w których gęstość prawdopodobieństwa rozkładu równowagi jest większa od zera. Algorytm M-H można wystartować z punktu o zerowej gęstości prawdopodobieństwa, jednak wtedy proces pozostanie w nim do chwili, kiedy z rozkładu pomocniczego zostanie wygenerowany punkt o niezerowej gęstości prawdopodobieństwa. Jeżeli punkt startowy zostanie przyjęty daleko od obszaru awarii to liczba kroków podczas, których łańcuch będzie pozostawał w tym samym miejscu może być duża. Wydaje się, więc uzasadnione zastosowanie dodatkowego algorytmu poszukującego punktów w obszarze awarii. Oczywiście można się spodziewać, że efektywność takiego algorytmu będzie uzależniona od wartości prawdopodobieństwa awarii.

Rozwiązanie problemu określenia punktu startowego jest jednym z powodów, dla których warto poszukiwać modyfikacji funkcji $q_f = I_{\Omega_f}(\mathbf{x})f(\mathbf{x})$ w celu przypisania dodatniej gęstości prawdopodobieństwa realizacjom zmiennych losowych na zewnątrz obszaru awarii. Prostą modyfikacją pozwalającą określić dodatnią gęstość prawdopodobieństwa w obszarze, gdzie zdefiniowana jest funkcja graniczna, można otrzymać wykorzystując w metodzie exponential tilting (2.55) funkcję kary, która jest stosowana w algorytmach poszukiwania punktu projektowego metod FORM/SORM [1]. Podejście to pozwala zachować pewne własności rozkładu $q_f = I_{\Omega_f}(\mathbf{x})f(\mathbf{x})$, a zwłaszcza rozmieszczenie maksimum gęstości prawdopodobieństwa.

Rozpatrzmy funkcję gęstości prawdopodobieństwa o następującej postaci:

$$q_t(\mathbf{x}) = \frac{1}{c_t} \exp(-\mu\rho(g(\mathbf{x}))) f(\mathbf{x}), \quad (8.76)$$

gdzie c_t jest stałą normalizującą, μ jest współczynnikiem funkcji kary, $g(\mathbf{x})$ jest funkcją graniczną, a $\rho(\cdot)$ jest funkcją o własnościach podobnych do tych, jakie ma funkcja kary wykorzystywana w optymalizacji z ograniczeniami [113]. Zakłada się więc spełnienie



RYSUNEK 8.2. Modyfikacja standardowego rozkładu normalnego za pomocą wyrażenia (8.76), (a) - $q_t(\mathbf{u}) = (6.89e - 3)^{-1} \exp(-8.0(\sqrt{2} * 3 - u_1 + u_2))\varphi_u(\mathbf{u})$, (b) - $q_t(\mathbf{u}) = (6.27e - 2)^{-1} \exp(-3.0(\log(\Phi(-u_1)) + \log(\Phi(-u_2)) + 5.0)^2)\varphi_u(\mathbf{u})$.

warunku:

$$\rho(y) = \begin{cases} 0 & \text{dla } y \leq 0, \\ > 0 & \text{dla } y > 0. \end{cases} \quad (8.77)$$

Funkcja $\rho(\cdot)$ powinna być dobrana tak, aby wartość gęstości prawdopodobieństwa szybko malała wraz ze wzrostem wartości funkcji granicznej g . Własność ta pozwala na zachowanie położenia modów gęstości q_f . Ponadto dobór $\rho(\cdot)$, jak i współczynnika μ powinny ułatwiać implementację algorytmu M-H. Oczywiście podstawowym wymaganiem jest, aby dla punktu startowego $\mathbf{x}^{(0)}$, $q_t(\mathbf{x}^{(0)})$ było dodatnie.

Przykład 8.1: Rysunek 8.2 przedstawia funkcje gęstości rozkładu prawdopodobieństwa otrzymane w wyniku modyfikacji gęstości standardowego rozkładu normalnego za pomocą wzoru (8.76). Wykres na rysunku 8.2 (a) przedstawia gęstość uzyskaną dla liniowej funkcji granicznej (por. (7.21)) $G(\mathbf{u}) = \sqrt{2} \cdot 3 - u_1 + u_2$ z parametrami modyfikującymi gęstość rozkładu prawdopodobieństwa o wartościach: $c_t = 6.89e - 3$, $\mu = 8.0$. W punktach, gdzie $G(\mathbf{u}) > 0$ funkcja kary ma postać $\rho(y) = y$. Na rysunku 8.2 (b) zamieszczono wykres zmodyfikowanej gęstości prawdopodobieństwa (8.76) dla nieliniowej funkcji granicznej (por. (7.22)) $G(\mathbf{u}) = \log(\Phi(-u_1)) + \log(\Phi(-u_2)) + 5.0$ z parametrami: $c_t = 6.27e - 2$, $\mu = 3.0$. W tym przypadku funkcja kary dla dodatnich wartości funkcji granicznej ma postać $\rho(y) = y^2$.

Oszacowanie prawdopodobieństwa awarii, na podstawie realizacji łańcucha Markowa $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}\}$ z rozkładem równowagi o gęstości $q(\mathbf{x})/c$, można uzyskać za pomocą estymatora importance sampling (2.49), który w tym przypadku przyjmuje postać

$$\bar{P}_{f_{M-H}} = \frac{1}{N} \sum_{n=1}^N I_{\Omega_f}(\mathbf{X}^{(n)}) \frac{f(\mathbf{X}^{(n)})}{q(\mathbf{X}^{(n)})/c}. \quad (8.78)$$

Jeżeli q ma postać (8.76), a funkcja kary spełnia warunek (8.77) to wyrażenie (8.78) można przedstawić jako:

$$\bar{P}_{f_{M-H}} = c_t \frac{1}{N} \sum_{n=1}^N I_{\Omega_f}(\mathbf{X}^{(n)}). \quad (8.79)$$

Tak, więc oszacowanie prawdopodobieństwa awarii za pomocą proponowanego podejścia wymaga rozwiązania dwóch zadań: oszacowania wartości oczekiwanej funkcji charakterystycznej obszaru awarii względem rozkładu o gęstości q_t oraz oszacowania stałej normalizującej gęstość q_t - na podstawie algorytmów przedstawionych w podrozdziale 8.7.

Efektywną realizację podejścia, które określa estymator (8.79) można uzyskać wykorzystując metody AIS i LIS. Przypomnijmy, że metody te pozwalają na oszacowanie ilorazu stałych normalizujących $r = c_K/c_1$. Niech c_K będzie stałą normalizującą gęstość prawdopodobieństwa q_K określoną przez (8.76) ze współczynnikiem kary μ_K . Ponadto przyjmijmy, że c_1 jest stałą normalizującą gęstość standardowego rozkładu normalnego, którą również można określić za pomocą (8.76) ze współczynnikiem kary $\mu_1 = 0$; oczywiście $c_1 = 1$. Sekwencję pośrednich rozkładów prawdopodobieństwa q_2, \dots, q_{K-1} , która jest podstawą oszacowania prawdopodobieństwa awarii za pomocą AIS i LIS, można zdefiniować na podstawie (8.76) poprzez sekwencję parametrów kary: $0 < \mu_2 < \dots < \mu_{K-1}$. Korzystając z tej sekwencji możemy oszacować stałą normalizującą w (8.79).

Oszacowanie prawdopodobieństwa awarii za pomocą metody AIS uzyskuje się na podstawie estymatora:

$$\bar{P}_{f_{AIS}} = \frac{1}{N} \sum_{n=1}^N I_{\Omega_f}(\mathbf{X}_K^{(n)}) \frac{q_1(\mathbf{X}_1^{(n)})}{\varphi_u(\mathbf{X}_1^{(n)})} \frac{q_3(\mathbf{X}_2^{(n)})}{q_2(\mathbf{X}_2^{(n)})} \cdots \frac{q_K(\mathbf{X}_{K-1}^{(n)})}{q_{K-1}(\mathbf{X}_{K-1}^{(n)})}, \quad (8.80)$$

gdzie $\mathbf{X}_K^{(n)}$ $n = 1, \dots, N$ jest próbą losową z rozkładu o gęstości q_K , którą otrzymano wydłużając o jeden krok algorytm AIS ze strony 172.

W przypadku metody LIS próbę losową o gęstości q_K można otrzymać generując łańcuch Markowa, który rozpoczyna się od punktu przejścia z próby o gęstości q_{K-1} . Tak, więc estymator prawdopodobieństwa awarii, który jest średnią oszacowań otrzymanych za pomocą M przebiegów metody LIS określa się jako:

$$\bar{P}_{fLIS} = \frac{1}{M} \sum_{m=1}^M \frac{\bar{r}_{LIS}^{(m)}}{n_K} \sum_{i=0}^{n_K} I_{\Omega_f} \left(\mathbf{X}_K^{(i)(m)} \right), \quad (8.81)$$

gdzie $\bar{r}_{LIS}^{(m)}$ są oszacowaniami ilorazu c_K/c_1 na podstawie (8.74), a $\mathbf{X}_K^{(i)(m)}$, $i = 0, \dots, n_K$, $m = 1, \dots, M$, są stanami realizacji łańcucha Markowa, m -tego przebiegu algorytmu, wygenerowanej z rozkładu o gęstości q_K .

8.9. Ocena błędu oszacowania za pomocą metody bootstrap

Oszacowanie wariancji estymatora prawdopodobieństwa awarii, uzyskanego klasycznymi metodami Monte Carlo, opiera się na prawdziwości centralnego twierdzenia granicznego dla średniej empirycznej funkcji charakterystycznej obszaru awarii (por. 2.3.6). W przypadku estymatorów stałych normalizujących, omówionych w podrozdziale 8.7, przyjmowanie założenia, że mają one asymptotycznie rozkład normalny jest nieuzasadnione. Co prawda pokazano, że w pewnych szczególnych przypadkach rozkłady prawdopodobieństwa tych estymatorów dążą do rozkładu normalnego [77, 85] to jednak analiza ich zachowania asymptotycznego jest konieczna dla nieróżniczkowalnych gęstości prawdopodobieństwa spotykanych w sformułowaniach analizy niezawodności, które przedstawiono w podrozdziale 8.8.

Alternatywą dla oszacowania błędu wykorzystującego asymptotyczne własności estymatora są metody polegające na generowaniu prób losowych, których elementy są powtórnie losowane z analizowanej próby losowej. Dla każdej z otrzymanych prób można wyznaczyć poszukiwane oszacowanie. W ten sposób otrzymuje się próbę losową wartości estymatora, która pozwala wyznaczyć przybliżenie jego rozkładu prawdopodobieństwa, a co za tym idzie jego obciążenie, wariancję oraz przedziały ufności. W tym podrozdziale przedstawiony zostanie jeden z podstawowych algorytmów wykorzystujących omawiane podejście tzw. metoda *bootstrap* [31].

Przyjmijmy, że w celu wyznaczenia poszukiwanego oszacowania posługujemy się statystyką g , która jest funkcją niezależnej N -elementowej próby losowej: $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(N)}\}$ z rozkładu \mathcal{F} .

Badanie własności oszacowania $g(\mathbf{X})$ za pomocą metody bootstrap składa się z następujących kroków:

- Wygenerowanie tzw. *prób losowych bootstrap* $\mathbf{Z}_m = \{\mathbf{Z}_m^{(1)}, \mathbf{Z}_m^{(2)}, \dots, \mathbf{Z}_m^{(N)}\}$, $m = 1, \dots, M$, które są otrzymywane poprzez losowanie ze zwracaniem elementów próby \mathbf{X} . Elementy próby \mathbf{Z}_m , $m = 1, \dots, M$, można więc określić jako: $\mathbf{Z}_m^{(n)} = \mathbf{X}^{(k)}$, $n = 1, \dots, N$, $k \sim \mathcal{U}(1, \dots, N)$, gdzie $\mathcal{U}(1, \dots, N)$ jest rozkładem jednorodnym określonym na zbiorze indeksów $\{1, \dots, N\}$.
- Wyznaczenie próby losowej składającej się z wartości statystyki dla prób bootstrap: $\mathbf{G} = \{g(\mathbf{Z}_1), \dots, g(\mathbf{Z}_M)\}$.
- Analiza rozkładu prawdopodobieństwa próby losowej \mathbf{G} .

Wartości dystrybuanty empirycznej statystyki g otrzymuje się za pomocą wyrażenia

$$\bar{F}_g(x) = \frac{1}{M} \sum_{m=1}^M I(g(\mathbf{Z}_m) \leq x). \quad (8.82)$$

Oszacowanie wariancji statystyki, które jest równoważne oszacowaniu wariancji estymatora, określa się wzorem:

$$\overline{\text{Var}}(g) = \frac{1}{M} \sum_{m=1}^M g(\mathbf{Z}_m)^2 - \left(\frac{1}{M} \sum_{m=1}^M g(\mathbf{Z}_m) \right)^2. \quad (8.83)$$

Na podstawie próby bootstrap można również oszacować przedział ufności estymatora. Wyznaczając kwantyle empiryczne rozkładu statystyki $\bar{q}(p)$, rzędu $p_1 = \frac{1}{2}(1 - C)$ i $p_2 = \frac{1}{2}(1 + C)$, otrzymujemy przedział $[\bar{q}(p_1), \bar{q}(p_2)]$ o poziomie ufności C .

Metoda bootstrap pozwala również uzyskać oszacowanie obciążenia estymatora $T(\mathbf{X})$, które określa się jako:

$$\overline{\text{Bias}}(g) = \frac{1}{M} \sum_{m=1}^M g(\mathbf{Z}_m) - g(\mathbf{X}). \quad (8.84)$$

Zauważmy, że powyższe wyrażenie może być również wykorzystane do oszacowania obciążenia estymatora, którego własności asymptotyczne są znane, wynikającego np. z małej liczby elementów próby losowej.

Jak wspomniano wcześniej metoda bootstrap może być stosowana, jeżeli próba losowa używana w oszacowaniu jest niezależna. Bootstrap w przedstawionej formie nie może być, więc wykorzystana w przypadku próby losowej otrzymanej za pomocą MCMC. W literaturze można znaleźć propozycje modyfikacji metody bootstrap, które pozwalają zastosować ją do analizy realizacji procesów stochastycznych [11], [57]. Zasadniczo modyfikacje te sprowadzają się do wprowadzenia procedury powtórnego

losowania, która pozwala zachować strukturę autokorelacji procesu. Należy jednak zaznaczyć, że zastosowanie bootstrap do analizy procesów stochastycznych jest bardziej skomplikowane i niestety mniej efektywne niż w przypadku prób niezależnych.

Jedno z podejść do implementacji metody bootstrap w analizie procesów stochastycznych wykorzystuje losowanie ze zwracaniem bloków kolejnych stanów zaobserwowanej realizacji $\mathbf{Y} = \{\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(N)}\}$. Łączenie danych w bloki zachowuje korelacje między elementami próby bootstrap. Tzw. *blokowa metoda bootstrap* (ang. *block bootstrap*) ma wiele odmian - w zależności od sposobu, w jaki losowane są bloki. W metodzie MBB (ang. *Moving Block Bootstrap*) losuje się ze zwracaniem nachodzące na siebie bloki $\mathcal{B}(i, l) = \{\mathbf{Y}^{(i)}, \mathbf{Y}^{(i+1)}, \dots, \mathbf{Y}^{(i+l-1)}\}$, $i = 1, \dots, N - l + 1$, $1 \leq l \leq N$. Realizacje metody bootstrap

$$\begin{aligned} \mathbf{S}_m(l) &= \{\mathcal{B}(I_1, l), \mathcal{B}(I_2, l), \dots, \mathcal{B}(I_k, l)\} \\ &= \left\{ \mathbf{Y}^{(I_1)}, \dots, \mathbf{Y}^{(I_1+l-1)}, \mathbf{Y}^{(I_2)}, \dots, \mathbf{Y}^{(I_2+l-1)}, \dots, \mathbf{Y}^{(I_k)}, \dots, \mathbf{Y}^{(I_k+l-1)} \right\}, \end{aligned} \quad (8.85)$$

gdzie $m = 1, \dots, M$ otrzymuje się łącząc bloki zgodnie z indeksami $I_i = 1, \dots, N - l + 1$, które wylosowano ze zwracaniem z rozkładu jednorodnego. Dla uproszczenia założymy, że $N = k \cdot l$. Jeżeli statystyka g jest ciągłą funkcją wartości średniej realizacji procesu $\bar{Y} = \frac{1}{N} \sum_{n=1}^N \mathbf{Y}^{(n)}$ to estymatory obciążenia i wariancji metody MBB są odpowiednio postaci [63]:

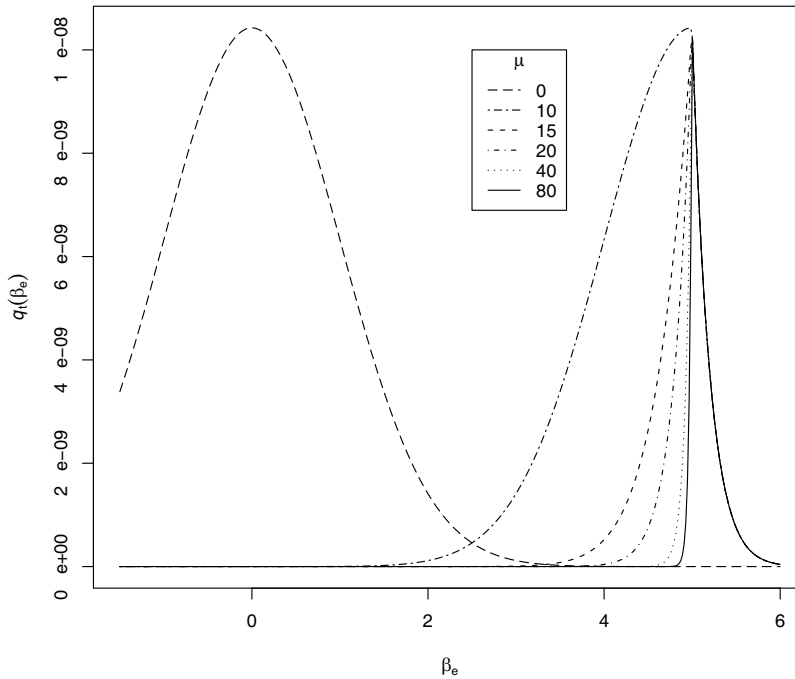
$$\overline{\text{Bias}}(g, l) = \frac{1}{M} \sum_{m=1}^M g(\mathbf{S}_m(l)) - g(\mathbf{Y}), \quad (8.86)$$

$$\overline{\text{Var}}(g, l) = \frac{1}{M} \sum_{m=1}^M g(\mathbf{S}_m(l))^2 - \left(\sum_{m=1}^M g(\mathbf{S}_m(l)) \right)^2. \quad (8.87)$$

Efektywność oszacowań (8.86), (8.87) zależy od doboru długości bloków l . Zbyt mała długość bloków powoduje obciążenie oszacowania, natomiast zbyt duża zwiększa wariancję. Zagadnienie doboru długości bloków w zależności od korelacji procesu i własności estymatora zostało omówione np. w pracy [63].

8.10. Optymalizacja algorytmu Metropolis-Hastings do zastosowań w analizie niezawodności

Zagadnienie efektywnej implementacji algorytmu MCMC było już omawiane w podrozdziale 8.5. Jednak przytoczone tam wyniki zostały uzyskane przy założeniu, że funkcje gęstości rozkładu prawdopodobieństwa spełniają pewne warunki gładkości. Problem efektywnej implementacji algorytmu M-H dla nieciągłych rozkładów praw-



RYSUNEK 8.3. Wykresy funkcji $\exp(-5.0 \cdot \mu \sqrt{k} \sum_{i=1}^k u_i) \cdot \varphi_u(\mathbf{u})$, wzdłuż prostej $\mathbf{u} = \beta_e \cdot 1/\sqrt{k}\mathbf{i}$. Dla $\mu = 0$ zmieniono skalę tak aby ułatwić porównanie wykresów.

dopodobieństwa, poza będącą ciągle w przygotowaniu pracą [85], nie był dotąd poruszany w literaturze. Zauważmy, że problem ten jest szczególnie istotny w przypadku zastosowania MCMC w analizie niezawodności. Jest tak ponieważ gęstości rozkładów, których wykorzystanie jest proponowane (8.76), nie są różniczkowalne, a w zasadzie są przybliżeniami rozkładów nieciągłych. W tym podrozdziale zostanie przedstawiona numeryczna optymalizacja algorytmu przypadkowego błądzenia (roz. 8.4.1) zastosowanego do generowania próby losowej z rozkładów (8.76). Otrzymane wyniki pozwoliły wyciągnąć pewne wnioski przydatne w implementacji metod szacowania stałych normalizujących (roz. 8.7).

Rozważmy algorytm przypadkowego błądzenia, którego zaburzenie losowe (8.26) ma rozkład normalny z macierzą kowariancji $\sigma^2 \mathbf{I}$, a więc taki jak w podrozdziale 8.5.2, gdzie omawiano efektywność M-H dla gładkich gęstości prawdopodobieństwa.

Przyjmijmy, że rozważany problem analizy niezawodności jest określony w gausowskiej przestrzeni standardowej, a funkcja graniczna ma postać hiperpłaszczyzny (7.21) ze współczynnikiem $\beta = 5.0$. Poniższe obliczenia wykonano dla przypadków określonych w $k = 5, 10, 20, 30, 40, 50$ wymiarach.

Zmodyfikowaną gęstość prawdopodobieństwa zdefiniowano zgodnie ze wzorem (8.76) wykorzystując funkcję kary $\rho(y) = y$. Dobierając współczynnik funkcji kary o wartościach $\mu = 0, 10, 15, 20, 40, 80$ otrzymano funkcje gęstości o różnym stopniu nieciągłości. Nieznormalizowane wykresy tych funkcji wzdłuż prostej $\mathbf{u} = \beta_e \cdot 1/\sqrt{k}\mathbf{i}$, zostały przedstawione na rysunku (8.3).

Własności gęstości prawdopodobieństwa (8.76) zmieniają się w zależności od współczynnika funkcji kary. W rozważanym zadaniu $\mu = 0$ określa gładką gęstość standardowego rozkładu normalnego, a $\mu \rightarrow \infty$ definiuje gęstość standardowego rozkładu normalnego obciążonego do obszaru awarii wyznaczonego przez funkcję graniczną $-5.0 \cdot \sqrt{k} \sum_{i=1}^k u_i$.

Optymalne wartości odchylenia standardowego rozkładu pomocniczego σ zostały wyznaczone za pomocą algorytmu przedstawionego w podrozdziale 8.6.2 poprzez maksymalizację oczekiwanego kwadratu długości przejścia (8.46). Prezentowane wyniki są średnimi uzyskanymi po przeprowadzeniu 20 eksperymentów dla każdej z kombinacji liczby wymiarów k i parametru μ .

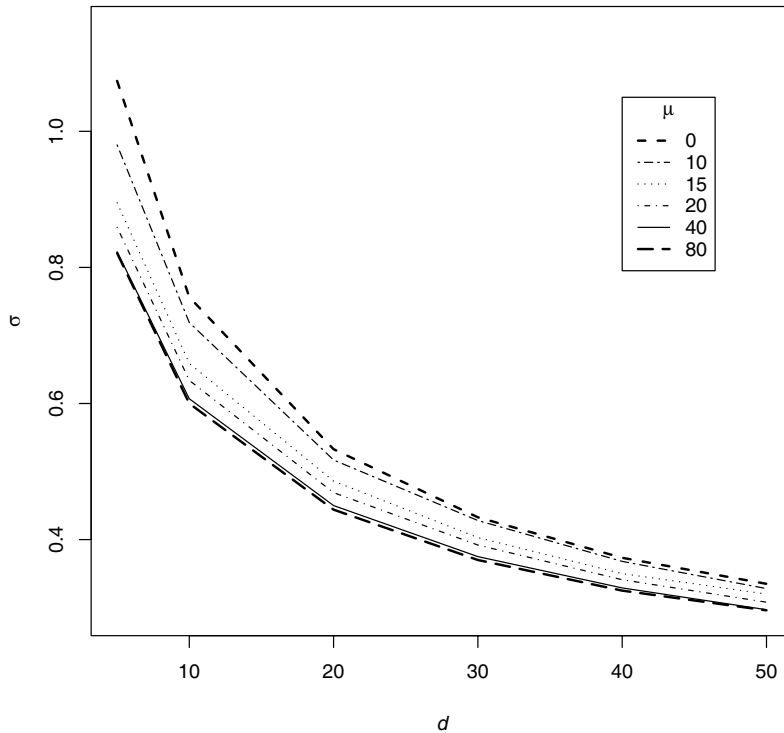
Wykresy optymalnych wartości σ , dla rozkładów celu o gęstości prawdopodobieństwa określonych przez parametr $\mu = 0, 10, 15, 20, 40, 80$, przedstawia rysunek 8.3.

Wykres dla $\mu = 0$, odpowiada krzywej $2.38/\sqrt{k}$, a więc jest zgodny z wynikami cytowanymi w podrozdziale 8.5.2 na podstawie pracy [43].

Jak widzimy na rysunku (8.3), w przypadku rozważanego zadania $\mu = 80$ zapewnia już dość dobre przybliżenie optymalnego rozkładu importance sampling. Tak więc w zależności od gładkości gęstości prawdopodobieństwa rozkładu celu optymalne wartości σ przyjmują wartości między krzywą dla $\mu = 0$ i $\mu = 80$ (w przybliżeniu).

Zakres zmienności oczekiwanego wskaźnika przejść a (8.37), dla optymalnych wartości σ określają dwie pogrubione linie na rysunku 8.5. Ograniczenie górne stanowi wykres a dla $\mu = 0$, który wraz z rosnącą liczbą wymiarów zbiega do wartości ≈ 0.240 . Otrzymana wartość graniczna jest nieco większa niż 0.234, która jest podawana w podrozdziale 8.5.2 jako optymalna w przypadku gładkich gęstości prawdopodobieństwa. Przybliżenie dolnego ograniczenia zmienności a stanowi wykres uzyskany dla $\mu = 80$, który wydaje się być zbieżny z dołu do wartości ok. 1.85. Tak, więc w przypadku rozważanych nieciągłych gęstości prawdopodobieństwa optymalne wartości a są mniejsze niż dla rozkładu normalnego i zbliżone do wartości 0.13, która została podana w podrozdziale 8.5.2.

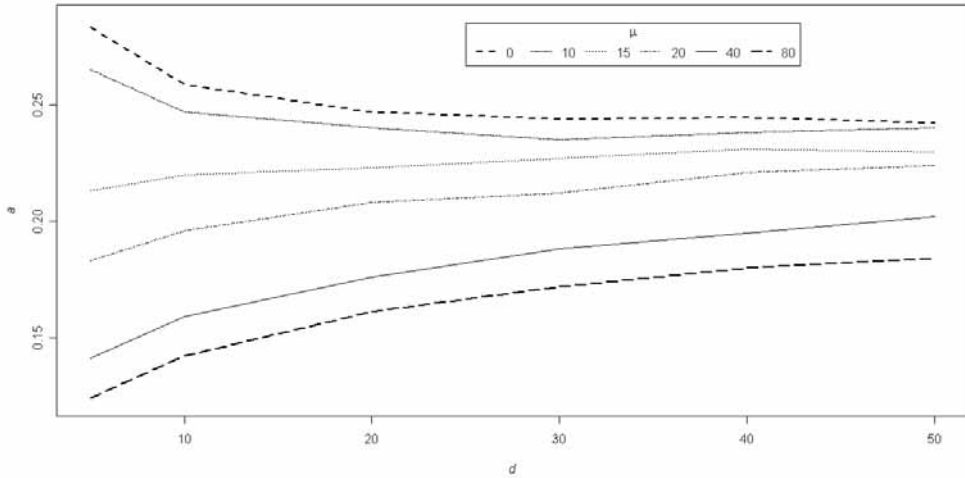
Zauważmy, że oszacowanie stałych normalizujących na podstawie wyrażenia (8.71) wymaga prób losowych z rozkładów o gęstościach określonych różnymi wartościami



RYSUNEK 8.4. Optymalne wartości odchylenia standardowego rozkładu pomocniczego σ , w przypadku różnych wartości współczynnika funkcji kary μ .

parametru μ . Oczywiście może się okazać, że nie jest możliwe wyznaczenie optymalnych parametrów rozkładu pomocniczego w przypadku każdej wartości μ . Podjęto więc próbę oceny efektywności omawianej implementacji przypadkowego błędzenia z nieoptymalnym parametrem σ zastosowanego do wyznaczenia wielkości występujących w (8.71).

Przeprowadzono następujący eksperyment. W celu otrzymania próby losowej z rozkładu określonego przez $\mu = 80$ wykorzystano σ^* , które jest optymalne dla gładkiej funkcji gęstości prawdopodobieństwa z parametrem $\mu = 0$. Natomiast w przypadku rozkładu zdefiniowanego przez $\mu = 10$ zastosowano σ_{80}^* - optymalne dla $\mu = 80$. Do porównania rozważanych implementacji przypadkowego błędzenia wykorzystano efektywność oszacowania (8.35) wartości oczekiwanej odległości punktu od początku



RYSUNEK 8.5. Wartości oczekiwanego wskaźnika przejść dla optymalnych implementacji przypadkowego błędzenia.

układu współrzędnych: $g_d(\mathbf{u}) = \mathbf{u}^T \mathbf{I} \mathbf{u}$. Parametr ten wybrano, ponieważ jest on argumentem funkcji gęstości prawdopodobieństwa występujących we wzorze (8.71).

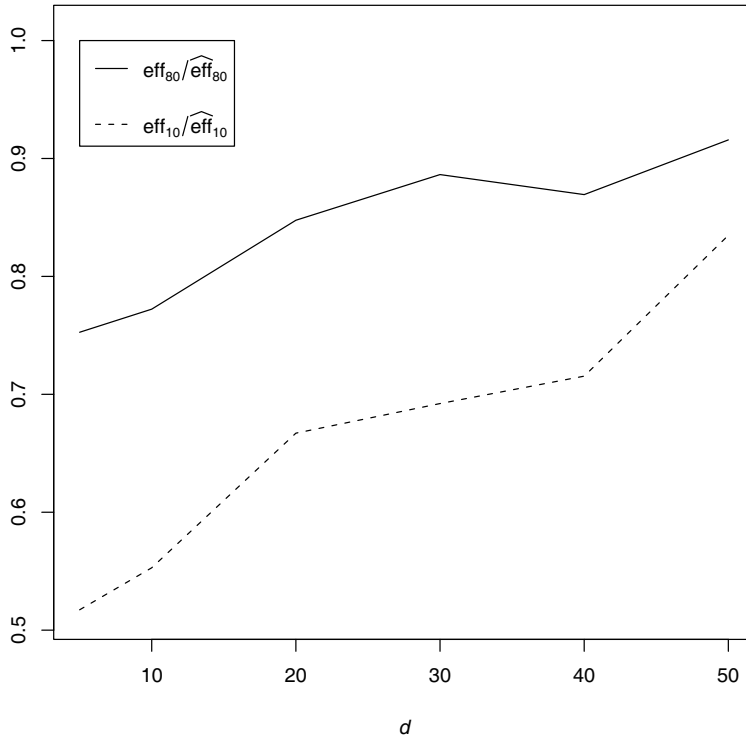
W przypadku rozkładu prawdopodobieństwa określonego przez $\mu = 80$ wyznaczono eff_{80} dla rozkładu pomocniczego z parametrami σ^* , oraz $\widehat{\text{eff}}_{80}$ dla rozkładu pomocniczego z parametrami σ_{80}^* .

W przypadku rozkładu określonego przez $\mu = 10$ wyznaczono efektywność eff_{10} dla rozkładu pomocniczego z σ_{80}^* i $\widehat{\text{eff}}_{10}$ z σ_{10}^* .

Rysunek 8.6 przedstawia wykresy $\text{eff}_{80}/\widehat{\text{eff}}_{80}$ oraz $\text{eff}_{10}/\widehat{\text{eff}}_{10}$ w zależności od liczby wymiarów. Jak widzimy zastosowanie σ optymalnego dla funkcji gładkich do generowania próby losowej z obciętej gęstości prawdopodobieństwa powoduje mniejszy spadek efektywności niż zastosowanie σ optymalnego dla $\mu = 80$ do gęstości określonej przez $\mu = 10$, która definiuje gęstość o cechach pośrednich. Tak, więc w przypadku konieczności generowania prób losowych z rozkładów określonych przez różne wartości μ rozsądne wydaje się zastosowanie wartości σ optymalnych dla funkcji gładkich. Wniosek ten jest bardzo wygodny, gdyż uzasadnia wykorzystanie wartości parametrów, których wartości są znane.

Przykład 8.2: Ocena efektywności metody AIS i LIS.

W przykładzie 7.2 pokazano, że implementacja metody wzajemnej entropii, która wy-



RYSUNEK 8.6. Porównanie efektywności nieoptymalnych implementacji przypadkowego błędzenia.

korzysta rozkład normalny z jednostkową macierzą kowariancji, jest nieefektywna w przypadku silnie nieliniowej powierzchni granicznej z dodatnią krzywizną. W tym przykładzie przedstawimy zastosowanie metod AIS i LIS do oszacowania prawdopodobieństwa obszaru awarii określonego taką powierzchnią graniczną.

Przyjmując liniową funkcję kary we wzorze (8.76), poza obszarem awarii pośrednie gęstości prawdopodobieństwa w przypadku funkcji granicznej (7.23) są określone wyrażeniem:

$$q_k(\mathbf{U}) = \exp\left(-\mu_k \left(\sum_{i=1}^d \ln[\Phi(-U_i)]/\lambda + C\right)\right) \varphi_u(\mathbf{U}), \quad k = 1, \dots, K. \quad (8.88)$$

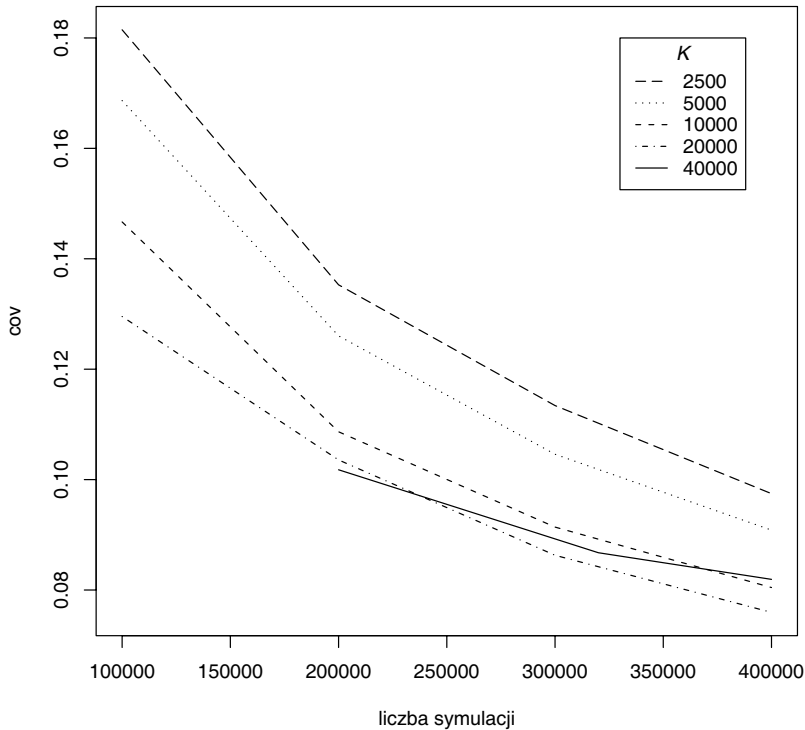
W celu oszacowania prawdopodobieństwa awarii za pomocą algorytmów AIS i LIS, odpowiednio na podstawie estymatorów (8.80) i (8.81), należy rozwiązać dwa zagadnienia:

μ		$d = 5$	$d = 10$	$d = 20$	$d = 30$	$d = 40$	$d = 50$
0.0	σ	1.074	0.756	0.532	0.432	0.373	0.335
	a	0.283	0.258	0.246	0.243	0.244	0.242
10.0	σ	0.980	0.719	0.531	0.428	0.368	0.328
	a	0.265	0.247	0.248	0.235	0.238	0.240
15.0	σ	0.895	0.659	0.486	0.403	0.350	0.320
	a	0.213	0.220	0.223	0.227	0.232	0.227
20.0	σ	0.859	0.634	0.469	0.392	0.341	0.308
	a	0.183	0.196	0.208	0.212	0.221	0.224
40.0	σ	0.823	0.607	0.450	0.375	0.329	0.297
	a	0.141	0.159	0.176	0.188	0.195	0.202
80.0	σ	0.821	0.600	0.444	0.370	0.325	0.296
	a	0.124	0.142	0.161	0.172	0.180	0.184

TABLICA 8.1. Efektywność algorytmu M-H dla zmodyfikowanego liniowego obszaru awarii.

pierwsze z nich dotyczy doboru pośrednich gęstości prawdopodobieństwa, a drugie przyjęcia parametrów rozkładu pomocniczego. W omawianym zadaniu pośrednie gęstości prawdopodobieństwa są określone przez sekwencję parametrów μ_k , $k = 1, \dots, K$. Na podstawie przeprowadzonych testów numerycznych stwierdzono, że w tak sformułowanym zadaniu liniowa sekwencja μ_k nie zapewnia zadowalającej efektywności oszacowania. Dobre wyniki otrzymano w przypadku sekwencji $\mu_k = \mu_0 2^{s_k}$, $k = 1, \dots, K$, gdzie μ_0 jest stałą zależną od parametru funkcji granicznej C , a s_k są równomiernie rozmieszczone w przedziale $[-12, 0]$. Korzystając z wniosków przedstawionych w podrozdziale 8.10 w prezentowanych testach numerycznych jako rozkład pomocniczy przyjęto rozkład normalny o niezależnych składowych z odchyleniami standardowymi $2.38/\sqrt{20}$.

Dylematem, jaki należy rozstrzygnąć implementując metody AIS i LIS jest przyjęcie liczby pośrednich rozkładów prawdopodobieństwa K . Można przypuszczać, że dłuższe sekwencje pozwolą na otrzymanie oszacowania o żądanej dokładności w mniejszej liczbie przebiegów algorytmów. Jednak z drugiej strony większa liczba przebiegów powinna pozwolić lepiej uwzględnić własności obszarów awarii o skomplikowanym kształcie. Rysunek 8.7 przedstawia wykresy efektywności AIS wykorzystującej sekwencje o różnej długości do oszacowania prawdopodobieństwa obszaru awarii, który jest określony w 20 wymiarach przez funkcję graniczną (7.22) z parametrem $C = 41.05$. W tym eksperymencie najbardziej efektywne okazały się przebiegi algorytmu wykorzystujące sekwencje rozkładów pośrednich o długości 20000. Pozwalają one na oszacowanie prawdopodobieństwa awarii ze współczynnikiem zmienności 0.1 po wykonaniu ok. 225000 symulacji. Ponieważ oszacowanie za pomocą przebiegu o długości 40000 jest mniej efektywne to wydają się, że istnieje pewna



RYSUNEK 8.7. Dokładność oszacowania w przypadku oszacowania za pomocą sekwencji o różnych długościach.

optymalna długość sekwencji gęstości rozkładów prawdopodobieństwa.

Wyniki zastosowania metod AIS i LIS do oszacowania prawdopodobieństwa awarii, określonego przez funkcję graniczną (7.22) w dwudziestowymiarowej standardowej przestrzeni normalnej, przedstawiono w tabelicy 8.2. Metody te pozwalają na dobre oszacowanie nawet małych prawdopodobieństw awarii, jednak koszty numeryczne ich zastosowania są bardzo duże. W ostatnim wierszu tabelicy 8.2 zamieszczono przewidywane koszty numeryczne uzyskania odpowiednich oszacowań o współczynniku zmienności 0.1 za pomocą klasycznej metody Monte Carlo. Jak widzimy w przypadku prawdopodobieństwa awarii rzędu 10^{-3} klasyczny algorytm jest bardziej efektywny niż AIS i LIS. Jednak wraz ze zmniejszaniem się prawdopodobieństwa awarii stosowanie proponowanych metod staje się coraz bardziej atrakcyjne.

Ze względu na małą liczbę elementów w próbach losowych współczynniki zmienności w

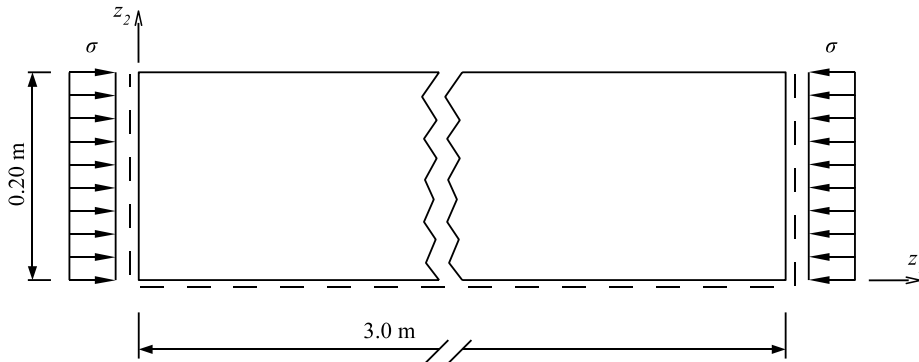
C	36.720	41.050	45.067	48.932
P_f	10^{-3}	10^{-4}	10^{-5}	10^{-6}
β	3.093	3.722	4.268	4.765
Metoda AIS				
K	20000	20000	30000	80000
μ_0	60	60	60	80
M	10	12	15	8
\hat{e}	0.0991	0.0981	0.0937	0.0932
\hat{P}_f	$0.994 \cdot 10^{-3}$	$0.984 \cdot 10^{-4}$	$0.995 \cdot 10^{-5}$	$0.963 \cdot 10^{-6}$
$K \cdot M$	$0.2 \cdot 10^6$	$0.24 \cdot 10^6$	$0.45 \cdot 10^6$	$0.64 \cdot 10^6$
Metoda LIS				
K	75	100	200	200
μ_0	60	60	60	80
n_k	200	200	200	200
M	16	16	10	12
\hat{e}	0.1022	0.105	0.0994	0.0923
\hat{P}_f	$0.957 \cdot 10^{-3}$	$1.061 \cdot 10^{-4}$	$1.007 \cdot 10^{-5}$	$1.006 \cdot 10^{-6}$
$K \cdot M$	$0.24 \cdot 10^6$	$0.32 \cdot 10^6$	$0.4 \cdot 10^6$	$0.48 \cdot 10^6$
MC	$0.1 \cdot 10^6$	$1.0 \cdot 10^6$	$10.0 \cdot 10^6$	$10.0 \cdot 10^7$

TABLICA 8.2. Ocena efektywności metod LIS i AIS w przypadku nieliniowej funkcji granicznej z dodatnią krzywizną.

tablicy 8.2 wyznaczono metodą *bootstrap* (por. 8.9).

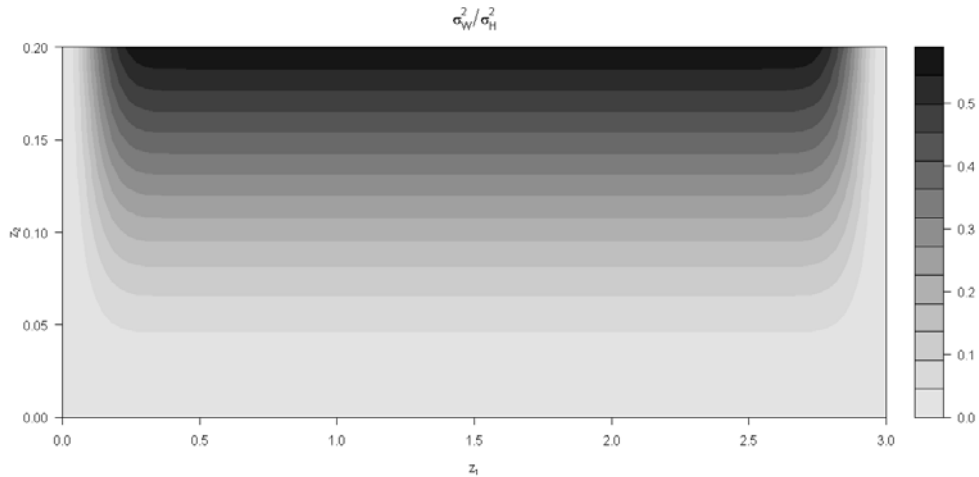
W omawianym przykładzie czasu obliczeń za pomocą metod AIS i LIS zostały skrócone dzięki równoległemu wykonaniu przebiegów algorytmów. W przypadku statystycznych obliczeń równoległych, konieczne jest posłużenie się generatorem liczb losowych zapewniającym brak korelacji między sekwencjami liczb losowych, które są generowane w różnych procesach. Wykorzystano w tym celu interfejs *rsprng* [67] środowiska *R* [90] do równoległego skalowalnego generatora liczb losowych SPRNG [72].

Analiza niezawodności ściskanej półki blachownicy



RYSUNEK 9.1. Schemat ściskanej półki dwuteownika.

W tym rozdziale przedstawimy wykorzystanie zaproponowanych metod analizy niezawodności do oszacowania prawdopodobieństwa awarii ściskanej półki dwuteownika z imperfekcjami geometrycznymi. W praktyce jako model takiego elementu konstrukcyjnego przyjmuje się płytę swobodnie podpartą wzdłuż trzech brzegów. Imperfekcje geometryczne w postaci odchyłek powierzchni środkowej nie mają tak istotnego wpływu na nośność graniczną płyt jak w przypadku prętów czy powłok. Jednak ścisnienie siłą bliską krytycznej jednostronnie podparych pasm płytowych z imperfekcjami powoduje powstanie przemieszczeń ich swobodnego brzegu. Tak, więc w tym zadaniu określimy prawdopodobieństwo przekroczenia dopuszczalnej wartości przemieszczeń w kierunku normalnym do powierzchni środkowej, które jest wynikiem dużego naprężenia



RYСУNEK 9.2. Wariancja warunkowego pola losowego, będącego modelem imperfekcji półki dwuteownika.

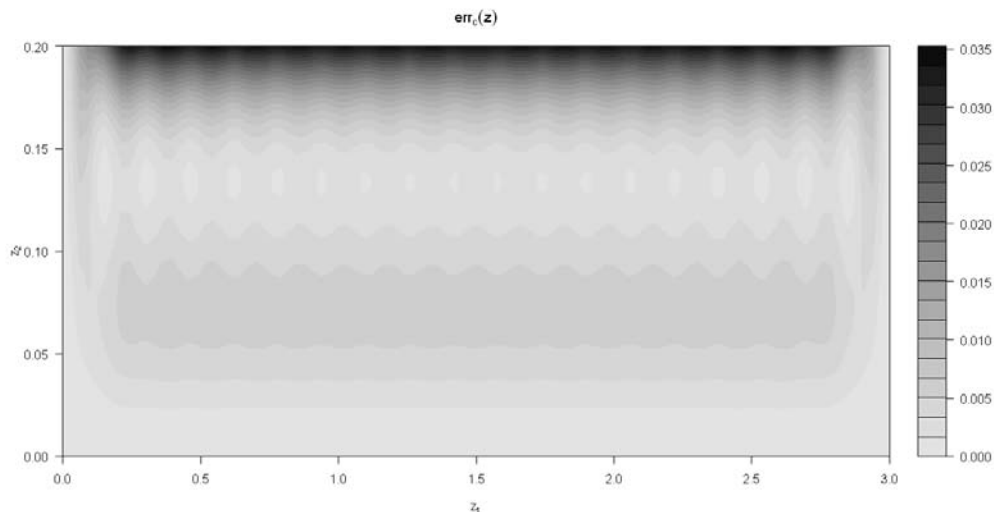
ściskającego.

Wymiary oraz schematy obciążenia i podparcia modelu analizowanego elementu konstrukcyjnego zostały pokazane na rysunku 9.1. Ze względu na przyjęte wymiary: szerokość $b = 0.2\text{m}$ oraz grubość $t = 0.01\text{m}$ (zgodnie z PN-90/B-03200) przekrój z taką ścianką należy do klasy 4, a więc traci nośność przy naprężeniach ściskających mniejszych od granicy plastyczności. Przyjmując, że element jest wykonany ze stali o wytrzymałości obliczeniowej 235MPa - smukłość względna ścianki wynosi (zgodnie z PN-90/B-03200 wzór (7)):

$$\bar{\lambda}_p = \frac{0.20}{0.01} \cdot \frac{2.2 + 0.8}{56} \sqrt{\frac{235}{215}} = 1.12$$

Zgodnie z tabelicą 9 w PN-90/B-03200 w przypadku ścianki o takiej smukłości względnej wartość współczynnika niestateczności miejscowej wynosi: $\varphi_p = 0.67$. Na podstawie wzoru (9) w PN-90/B-03200 można określić wartość największego naprężenia normalnego, która spełnia warunek stateczności lokalnej: $\sigma_c \leq 0.67 \cdot 235 = 157.45\text{MPa}$.

Jako model imperfekcji geometrycznych przyjęto gaussowskie pole losowe $W(\mathbf{z})$, $\mathbf{z} \in \{[0.0, 3.0] \times [0.0, 0.2]\}$, którego realizacje spełniają warunki brzegowe podparcia. Pole to zostało zdefiniowane przy pomocy metody przedstawionej w rozdziale 4. Za podstawę modelu przyjęto stacjonarne gaussowskie pole losowe $H(\mathbf{z})$, $\mathbf{z} \in \{[0.0, 3.0] \times$



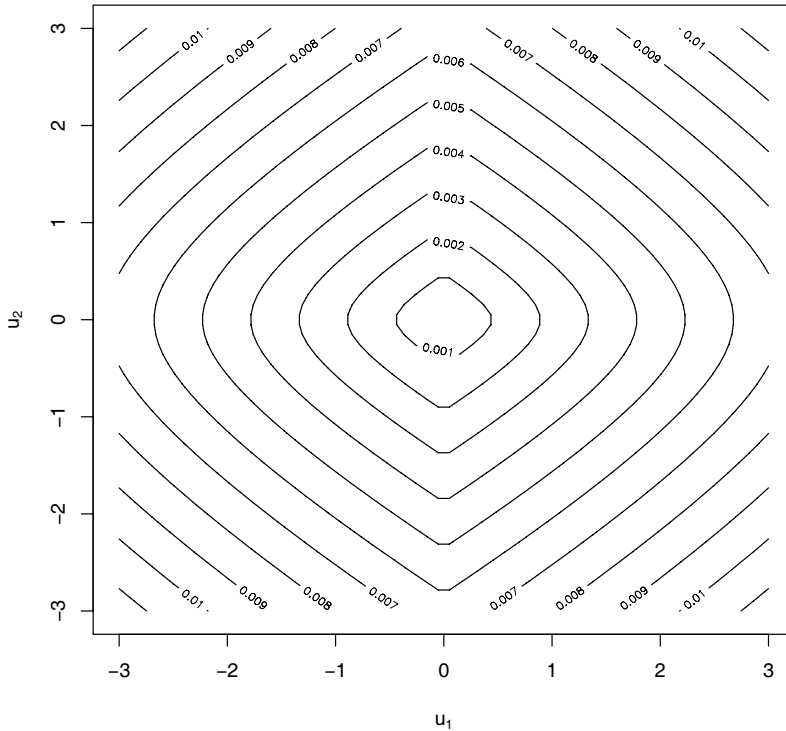
RYSUNEK 9.3. Błąd dyskretyzacji pola losowego imperfekcji geometrycznych.

$[0.0, 0.2]$ }, o zerowej wartości oczekiwanej $\mu_H = 0$ i anizotropowej funkcji kowariancji:

$$K_H(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = \sigma_H^2 \exp\left(-\left(\frac{(z_1^{(1)} - z_1^{(2)})^2}{0.2^2} + \frac{(z_2^{(1)} - z_2^{(2)})^2}{0.3^2}\right)\right). \quad (9.1)$$

Pole W zdefiniowano warunkując pole H na jego wartościach równych zero w sześciu, równomiernie rozmieszczonych, punktach na każdym z krótszych brzegów i 40, również równomiernie rozmieszczonych, punktach na brzegu dłuższym. Wykres względnej wariancji tak określonego warunkowego pola losowego W przedstawia rysunek 9.2.

Dyskretyzację pola losowego \hat{W} uzyskano za pomocą metody EOLE (roz. 4.4). Wykorzystano w tym celu 240 wartości pola losowego rozmieszczonych w węzłach siatki 40×6 . Prezentowane w dalszej części przykładu wyniki obliczeń numerycznych otrzymano za pomocą pierwszych 18 wyrazów rozwinięcia EOLE. Jak można zobaczyć na rysunku 9.3, pomimo tak małej liczby zmiennych losowych błąd dyskretyzacji (4.89) nie przekracza 3.5%. Wykresy wybranych funkcji kształtu metody EOLE zamieszczono na rysunku 9.5. Zauważmy, że składowe generowanych realizacji pola losowego spełniają warunki brzegowe podparcia. W omawianym modelu pierwszych 18 funkcji własnych ma postać coraz gęstszych fal wzdłuż dłuższego boku płyty. Najbardziej



RYSUNEK 9.4. Maksymalne przemieszczenie normalne w funkcji dwóch pierwszych zmiennych losowych dyskretyzacji EOLE.

złożony przypadek tego typu odpowiada osiemnastej funkcji rozwinięcia (Rys. 9.5(d)). W przypadku kolejnych funkcji własnych pojawiają się bardziej złożone formy odkształcenia wzdłuż krótszego boku płyty. Funkcję kształtu tego rodzaju przedstawia rysunek 9.5(e). Zwróćmy uwagę, że jest to odpowiednik trzeciego wektora własnego (Rys. 9.5(c)), jednak jego wartości są kilkakrotnie mniejsze.

W omawianym przykładzie, jako funkcję graniczną przyjęto przekroczenie dopuszczalnej wartości $v_{dop} = 0.02[\text{m}]$ przez przemieszczenie v w kierunku normalnym do

powierzchni środkowej płyty:

$$g(\mathbf{U}) = 1 - \frac{\max |v(\hat{W}(\mathbf{U}, \mathbf{z}), \mathbf{z})|}{v_{\text{dop}}}. \quad (9.2)$$

Za pomocą \mathbf{U} we wzorze (9.2) oznaczono standardowy wektor normalny rozwinięcia EOLE. Jak wspomniano w rozważanym zadaniu jest on określony w osiemnastu wymiarach. Zauważmy, że w przypadku realizacji imperfekcji, które są swoimi odbiciami względem osi symetrii płyty, funkcja graniczna (9.2) przyjmuje takie same wartości. Z własności tej wynikają symetrie obszaru awarii względem początku układu współrzędnych przestrzeni zmiennych losowych. W przypadku rozważanego zadania można przypuszczać, że powierzchnia graniczna ma kształt symetrycznej nieregularnej bryły. Jakościowo problem ten ilustruje rysunek 9.4, gdzie przedstawiono wykres maksymalnego przemieszczenia normalnego w funkcji dwóch pierwszych zmiennych losowych dyskretyzacji EOLE. Oszacowanie prawdopodobieństwa awarii w analizowanym przypadku jest trudnym problemem, gdyż wymaga uwzględnienia co najmniej kilku lokalnych maksimów gęstości prawdopodobieństwa w obszarze awarii.

Obliczenia statyczne w tym przykładzie wykonano za pomocą programu metody elementów skończonych *FEAPpv* [118]. Wykorzystano w tym celu 960 elementów powłokowych na siatce 60×16 .

W przykładzie 7.2 pokazano, że metoda wzajemnej entropii wykorzystująca do generowania próby losowej rozkład normalny o jednostkowej macierzy kowariancji jest nieefektywna w przypadku silnie nieliniowych funkcji granicznych o dodatniej krzywiznie. Na podstawie rysunku 9.4 wystąpienia podobnego problemu można spodziewać się również w przypadku funkcji (9.2). Aby umożliwić zastosowanie w omawianym zadaniu metody wzajemnej entropii, stan graniczny określony przez (9.2) zastąpimy szeregowym systemem awarii. W tym celu zdefiniujemy następujące funkcje graniczne

$$g^{\max}(\mathbf{U}, \mathcal{Z}) = 1 - \frac{\max(v(\hat{W}(\mathbf{U}, \mathbf{z}), \mathbf{z}) | \mathbf{z} \in \mathcal{Z})}{v_{\text{dop}}} \quad (9.3)$$

oraz

$$g^{\min}(\mathbf{U}, \mathcal{Z}) = 1 + \frac{\min(v(\hat{W}(\mathbf{U}, \mathbf{z}), \mathbf{z}) | \mathbf{z} \in \mathcal{Z})}{v_{\text{dop}}}, \quad (9.4)$$

gdzie \mathcal{Z} są podzbiorami w obszarze definiującym płytę $\{[0.0, 3.0] \times [0.0, 0.2]\}$. Płytę podzielono wzdłuż jej dłuższego brzegu na pięć obszarów: $\mathcal{Z}_1 = \{\mathbf{z} : 0.0 \leq z_1 < 0.6\}$, $\mathcal{Z}_2 = \{\mathbf{z} : 0.6 \leq z_1 < 1.2\}$, $\mathcal{Z}_3 = \{\mathbf{z} : 1.2 \leq z_1 < 1.8\}$, $\mathcal{Z}_4 = \{\mathbf{z} : 1.8 \leq z_1 < 2.4\}$, $\mathcal{Z}_5 = \{\mathbf{z} : 2.4 \leq z_1 < 3.0\}$. Teraz korzystając z (9.3) oraz (9.4) można określić system

szeregowy, którego prawdopodobieństwo awarii jest równe prawdopodobieństwu awarii określonej przez funkcję graniczną (9.2):

$$P_f(\{g(\mathbf{U}) \leq 0\}) = P_f\left(\bigcup_{i=1}^5\{g^{\max}(\mathbf{U}, \mathcal{Z}_i) \leq 0\} \bigcup \bigcup_{i=1}^5\{g^{\min}(\mathbf{U}, \mathcal{Z}_i) \leq 0\}\right). \quad (9.5)$$

	$g^{\max}(\mathcal{Z}_1)$	$g^{\max}(\mathcal{Z}_2)$	$g^{\max}(\mathcal{Z}_3)$	$g^{\max}(\mathcal{Z}_4)$	$g^{\max}(\mathcal{Z}_5)$
i	1	2	3	4	5
n_i	4480	6784	5248	5952	3840
β_i	5.340	4.914	4.750	4.818	5.277
	$g^{\min}(\mathcal{Z}_1)$	$g^{\min}(\mathcal{Z}_2)$	$g^{\min}(\mathcal{Z}_3)$	$g^{\min}(\mathcal{Z}_4)$	$g^{\min}(\mathcal{Z}_5)$
i	6	7	8	9	10
n_i	3968	4160	4928	3648	5568
β_i	5.232	5.422	5.097	4.970	5.370

TABLICA 9.1. Wyniki oszacowania parametrów metody wzajemnej entropii.

Prawdopodobieństwo awarii w wyrażeniu (9.5) oszacowano za pomocą metody importance sampling na podstawie próby losowej z rozkładu o gęstości:

$$q(\mathbf{u}) = \sum_{i=1}^{10} \lambda_i \varphi(\mathbf{u}, \hat{\mathbf{u}}_i^*, \mathbf{I}), \quad (9.6)$$

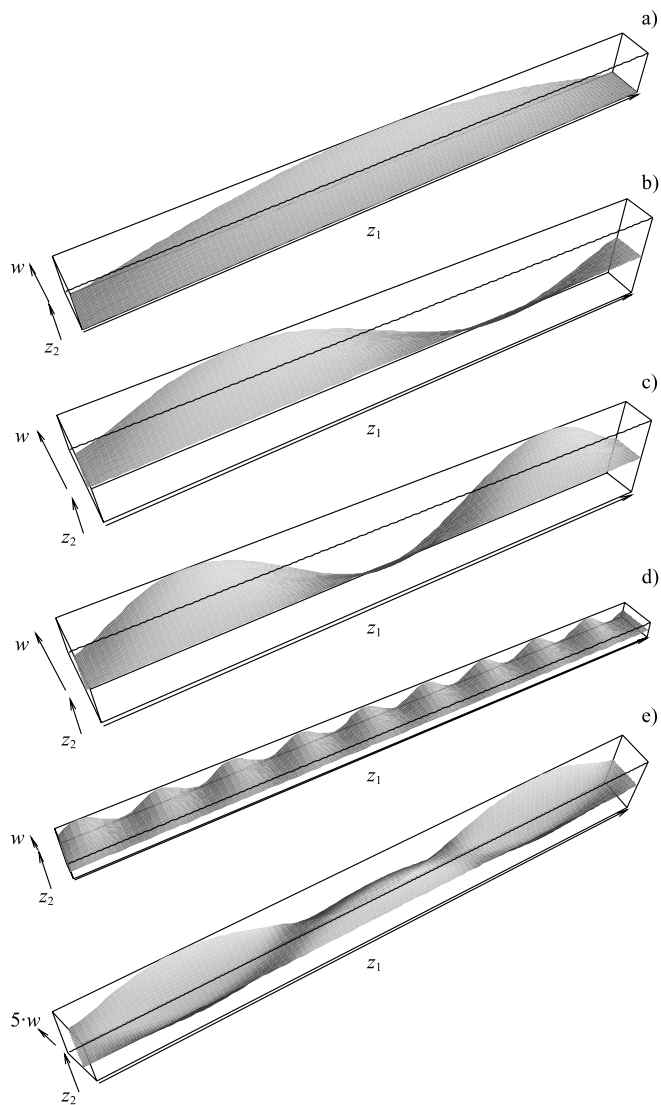
gdzie $\lambda_i > 0$, $\sum_{i=1}^{10} \lambda_i = 1$, a $\hat{\mathbf{u}}_i^*$, $i = 1, \dots, 10$, są wektorami wartości oczekiwanych składowych zmiennej \mathbf{U} w obszarach awarii, które określono przez kolejne funkcje graniczne $g^{\max}(\mathcal{Z}_1), \dots, g^{\max}(\mathcal{Z}_5), g^{\min}(\mathcal{Z}_1), \dots, g^{\min}(\mathcal{Z}_5)$. Wektory $\hat{\mathbf{u}}_i^*$ zostały oszacowane za pomocą metody wzajemnej entropii. W przypadku wszystkich funkcji granicznych przyjęto następujące parametry algorytmu: $\rho = 0.1$ oraz $\hat{e}_v = 0.005$ (por. roz. 7.4). Liczby symulacji, n_i , $i = 1, \dots, 10$, jakich wymagało oszacowanie z żadaną dokładnością odpowiednich $\hat{\mathbf{u}}_i^*$, $i = 1, \dots, 10$, zostały przedstawione w tabelicy 9.1. Ponieważ odległości od środka układu współrzędnych (wskaźniki niezawodności) β_i , punktów $\hat{\mathbf{u}}_i^*$ $i = 1, \dots, 10$, (tab. 9.1) mają zbliżone wartości to zdecydowano się przyjąć równe współczynniki składowych gęstości prawdopodobieństwa w (9.6): $\lambda_i = 0.1$, $i = 1, \dots, 10$. Z każdego rozkładu o gęstości $\varphi(\mathbf{u}, \hat{\mathbf{u}}_i^*, \mathbf{I})$ $i = 1, \dots, 10$ wygenerowano próby losowe zawierające 1600 elementów. Oszacowanie prawdopodobieństwa awarii (9.5) otrzymano za pomocą estymatora

$$P_{f_{CE}}(\{g(\mathbf{U}) \leq 0\}) \approx \frac{1}{10 \cdot 1600} \sum_{i=1}^{10} \sum_{j=1}^{1600} I_{\{g(\mathbf{x}) \leq 0\}}(\mathbf{x}_{ij}) \frac{\varphi(\mathbf{x}_{ij}, \mathbf{0}, \mathbf{I})}{\varphi(\mathbf{x}_{ij}, \hat{\mathbf{u}}_i^*, \mathbf{I})}, \quad (9.7)$$

gdzie \mathbf{x}_{ij} , $i = 1, \dots, 10$, $j = 1, \dots, 1600$ są elementami prób losowych. Obliczone w ten sposób prawdopodobieństwo awarii wyniosło $P_{f_{CE}}(\{g(\mathbf{U}) \leq 0\}) \approx 3.217 \cdot 10^{-6}$, ze współczynnikiem zmienności $\hat{e}_{CE} = 0.0953$. Tak, więc otrzymane oszacowanie, które można postrzegać jako połączenie prób losowych metody wzajemnej entropii wymagało wykonania 64576 symulacji.

Metody AIS i LIS powinny dawać dobre oszacowania prawdopodobieństwa awarii nawet w przypadku skomplikowanych powierzchni granicznych (por. 8.7.4). Korzystając z estymatorów (8.80) i (8.81), podjęto próbę rozwiązania omawianego zadania bezpośrednio na podstawie funkcji (9.2). W przypadku metody AIS gęstości prawdopodobieństwa określono za pomocą sekwencji parametrów $\varsigma_k = 12000 \cdot 2^{\nu_k}$, $k = 1, \dots, 19999$, gdzie ν_k są równomiernie rozłożone w przedziale $[-12, 0]$. Po wykonaniu 32 przebiegów algorytmu na podstawie (8.80) otrzymano oszacowanie prawdopodobieństwa awarii: $P_{f_{AIS}}(\{g(\mathbf{U}) \leq 0\}) \approx 3.108 \cdot 10^{-6}$, którego współczynnik zmienności oszacowany metodą bootstrap wyniósł $\hat{e}_{AIS} = 0.1095$. Pośrednie gęstości prawdopodobieństwa metody LIS określono tak jak w przypadku AIS jednak było ich jedynie 99. Na każdym z poziomów wygenerowano po 200 przejść łańcucha Markowa. Ponieważ wykonano 32 przebiegi algorytmu to liczba wykonanych symulacji była taka sama jak w oszacowaniu metodą AIS. Za pomocą wzoru (8.81) uzyskano następujące oszacowanie prawdopodobieństwa awarii: $P_{f_{LIS}}(\{g(\mathbf{U}) \leq 0\}) \approx 3.317 \cdot 10^{-6}$ o współczynniku zmienności $\hat{e}_{LIS} = 0.1088$.

W omawianym przykładzie oszacowanie prawdopodobieństwa awarii, o takim samym współczynniku zmienności, wymagało dziesięciokrotnie większej liczby symulacji w przypadku metod AIS i LIS niż za pomocą metody wzajemnej entropii. Jednak metody generujące łańcuchy Markowa nie wymagały definiowania systemu awarii, a wymagane nakłady obliczeniowe były kilkudziesięciokrotnie mniejsze niż w przypadku klasycznej metody Monte Carlo.



RYСУNEK 9.5. Wykresy "funkcji kształtu" dyskretyzacji EOLE imperfekcji pasa dwuteownika, odpowiadające kolejnym wartościom własnym: (a) - 1 (wartość własna), (b) - 2, (c) - 3, (d) - 18, (e) - 21.

Wnioski, spostrzeżenia i kierunki dalszych badań

Jednym z zagadnień przedstawionych w tej pracy jest zastosowanie metody wzajemnej entropii i metod Markov chain Monte Carlo do oszacowania prawdopodobieństwa awarii. Metoda wzajemnej entropii jest algorytmem symulacji rzadkich zdarzeń, jej zastosowanie w analizie niezawodności wymaga jedynie uwzględnienia charakterystycznej definicji stanu granicznego. Pomimo zalet tego algorytmu, jakimi są: eleganckie sformułowanie i dobre opracowanie podstaw teoretycznych, wydaje się, że autor jako pierwszy podjął próbę zastosowania metody wzajemnej entropii do wyznaczenia prawdopodobieństwa awarii konstrukcji.

W celu zastosowania do analizy niezawodności metod generujących łańcuchy Markowa, oszacowanie prawdopodobieństwa awarii sformułowano jako zadanie wyznaczenia stałej normalizującej. Korzystając z zaproponowanego podejścia, przedstawiono przykłady oszacowania prawdopodobieństwa awarii, które określono w postaci ilorazu stałych normalizujących funkcję gęstości prawdopodobieństwa parametrów losowych konstrukcji obciążoną do obszaru awarii i gęstość standardowego rozkładu normalnego. W przykładach wykorzystano algorytmy Annealed Importance Sampling i Linked Importance Sampling. Do implementacji tych metod posłużono się sekwencjami rozkładów prawdopodobieństwa, które zdefiniowano na podstawie funkcji granicznej.

Innym zagadnieniem omawianym w tej rozprawie jest modelowanie imperfekcji geometrycznych za pomocą pól losowych. Szczególną uwagę poświęcono opracowaniu metody definiowania pól losowych, których realizacje spełniają warunki brzegowe utwierdzenia. Wykorzystano w tym celu warunkowanie jednorodnych pól losowych na wartościach ich gradientu w punktach brzegowych.

Wymienione powyżej metody znalazły zastosowanie w analizie niezawodności elementu konstrukcyjnego z imperfekcjami geometrycznymi. Opisując rozwiązanie tego przykładu, zwrócono uwagę na konieczność uwzględnienia symetrii powierzchni gra-

nicznej związanej z symetriami konstrukcji. Pominięcie tego problemu prowadzi do grubych błędów oszacowania prawdopodobieństwa awarii.

Jeden z rozdziałów rozprawy poświęcono optymalizacji niezawodnościowej konstrukcji z uwzględnieniem dyskretnego charakteru zmiennych projektowych. Przedstawiono zastosowanie metody transformacji do ciągłej przestrzeni parametrów i metody kontrolowanego przeglądu, w celu rozwiązania zagadnień tego rodzaju. Efektywność obu algorytmów została oceniona na przykładach optymalizacji konstrukcji kratownicowych. Zaproponowano również sposoby poprawy efektywności tych metod wykorzystujące charakterystyczne własności zadania optymalizacji niezawodnościowej.

Rozważania i eksperymenty numeryczne zawarte w tej pracy skłaniają do sformułowania następujących wniosków.

1. Metoda wzajemnej entropii może być efektywnie zastosowana w analizie niezawodności konstrukcji. Algorytm ten pozwala na oszacowanie prawdopodobieństwa awarii w przypadkach, gdy funkcja graniczna jest nieróżniczkowalna, nieciągła bądź zaszumiona. Wykorzystywana w tej metodzie iteracyjna procedura oszacowania parametrów rozkładu prawdopodobieństwa, z którego generowana jest próba losowa, wymaga dodatkowych obliczeń funkcji granicznej. Jednak w większości przypadków, całkowity koszt obliczeń numerycznych koniecznych do oszacowania prawdopodobieństwa awarii konstrukcji za pomocą tego algorytmu jest możliwy do zaakceptowania.

Prezentowane przykłady pokazały, że analizowane sformułowanie metody wzajemnej entropii z rodziną rozkładów normalnych o jednostkowej macierzy kowariancji i zmiennej wartości oczekiwanej jest mało efektywne, gdy powierzchnia graniczna ma silnie dodatnią krzywiznę. Sposobem rozwiązania tego problemu może być zastosowanie do generowania próby losowej rodziny bardziej złożonych rozkładów prawdopodobieństwa. Efektywne wykorzystanie algorytmu wymaga, aby takie rozkłady spełniały co najmniej dwa warunki: rozwiązanie zadania minimalizacji wzajemnej entropii względem ich parametrów powinno być możliwe do uzyskania analitycznie, a ponadto oszacowanie wartości tych parametrów musi być efektywne numerycznie. W opinii autora, właśnie ze względu na trudności z oszacowaniem numerycznym, trudnym do wykorzystania parametrem jest macierz kowariancji. W przypadku dużej liczby wymiarów jej dokładne wyznaczenie wymaga przeprowadzenia wielu symulacji, natomiast zbyt grube przybliżenia powodują kłopoty ze zbieżnością procedury iteracyjnej.

W zadaniach, gdzie gęstość prawdopodobieństwa nie jest skupiona wokół jednego punktu obszaru awarii, do generowania próby losowej metody importance sampling wykorzystuje się rozkłady mieszane. Jednak ze względu na specyficzne problemy związane z określeniem parametrów rozkładów mieszanych [18] wydaje

się, że ich zastosowanie w metodzie wzajemnej entropii może okazać się trudnym problemem.

2. Prawdopodobieństwo awarii może być oszacowane za pomocą algorytmów symulacyjnych generujących łańcuchy Markowa. Konieczne w tym celu jest zastosowanie metod służących do wyznaczania stałych normalizujących. Prezentowane przykłady pokazały, że zastosowanie tego podejścia wiąże się z koniecznością wykonania dużej liczby symulacji. Jednak metody Annealed Importance Sampling oraz Linked Importance Sampling zostały z powodzeniem wykorzystane do rozwiązania zadania, w którym obszar awarii jest symetryczny względem początku układu współrzędnych. Przykłady testowe pokazały ponadto, że w przypadku małych prawdopodobieństw awarii efektywność tych algorytmów jest znacząco wyższa, niż tradycyjnej metody Monte Carlo. Wydaje się, że wykorzystanie metod Annealed Importance Sampling i Linked Importance Sampling jest warte rozważenia w przypadku złożonych obszarów awarii o wielu punktach projektowych.

Dalsze prace nad zastosowaniem Annealed Importance Sampling i Linked Importance Sampling w analizie niezawodności powinny dotyczyć poprawy efektywności oszacowania prawdopodobieństwa awarii. Zapewne dużą poprawę można osiągnąć poprzez optymalizację sekwencji pośrednich gęstości prawdopodobieństwa. Warto w tym celu dokładniej zbadać własności zaproponowanej metody definiowania gęstości prawdopodobieństwa za pomocą funkcji granicznej. Ponieważ Annealed Importance Sampling i Linked Importance Sampling generują punkty próby losowej tak jak algorytm Metropolis-Hastings to celowe wydaje się racjonalne dostosowanie, przedstawionych w tej pracy, metod optymalizacji rozkładu pomocniczego.

3. W opinii autora jednym z ciekawszych osiągnięć tej pracy jest opracowanie metody definiowania pól losowych, których realizacje spełniają warunki brzegowe utwierdzenia. Celowość wykorzystania proponowanego podejścia wymaga jednak uzasadnienia. Losowy model imperfekcji można przecież uzyskać za pomocą sumy postaci wyboczenia ze współczynnikami losowymi. Ponadto zauważmy, że funkcje kształtu metody EOLE przypominają kolejne postacie wyboczenia. Co więcej, realizacje pola losowego uzyskane za pomocą proponowanej metody spełniają warunki brzegowe w sposób przybliżony, a postacie wyboczenia spełniają je dokładnie. Otóż wydaje się, że metoda warunkowania pozwala łatwiej uwzględnić zależność od kierunków funkcji kowariancji pola losowego. W przykładzie analizy niezawodności pasa blachownicy osiągnięto ten cel wprowadzając nieizotropową funkcję kowariancji jednorodnego pola losowego. W przypadku wykorzystania postaci wyboczenia, konieczne byłoby określenie tensora sztywności o odpowied-

nio dobranych składowych.

Przykłady zamieszczone w tej pracy dotyczą jedynie zagadnień określonych w kartezyjskim układzie współrzędnych. Istotne praktycznie znaczenie ma analiza stateczności powłok z imperfekcjami geometrycznymi. Jednak wykorzystanie w tym celu pól losowych wymaga teoretycznego opracowania metod definiowania funkcji kowariancji w krzywoliniowych układach współrzędnych. Pogłębienia wymagają również podstawy teoretyczne proponowanej metody. Konieczne jest zwłaszcza dokładniejsze zbadanie własności zaproponowanej postaci gradientu pola losowego.

4. Optymalizacja niezawodnościowa jest zagadnieniem wymagającym długotrwałych obliczeń numerycznych. Uwzględnienie dyskretnego charakteru niektórych zmiennych projektowych stanowi dodatkowe utrudnienie, gdyż uniemożliwia zastosowanie gradientowych algorytmów optymalizacji. W pracy tej zaproponowano skrócenie czasu obliczeń poprzez wykorzystanie rozwiązania zadania określonego przy założeniu ciągłego charakteru parametrów projektowych. Wokół tego rozwiązania można określić kilkuelementowe zbiory wartości parametrów dyskretnych. Dzięki temu w przypadku metody transformacji uzyskuje się zmniejszenie liczby wymiarów przestrzeni parametrów, a w przypadku metody przeglądu redukuje się liczbę kombinacji parametrów projektowych wymagających sprawdzenia ograniczeń zadania optymalizacji. Ponadto z procesu dokładnego sprawdzania ograniczeń optymalizacji można wykluczyć te kombinacje projektowe, których przybliżenia liniowe ograniczeń niezawodnościowych są dalekie od zera. Aproksymację liniową ograniczeń niezawodnościowych można łatwo uzyskać na podstawie gradientów wskaźników niezawodności rozwiązania w ciągłej przestrzeni parametrów. Oczywiście otrzymane w ten sposób rozwiązanie może nie być "dokładne", jednak w większości przypadków pozwoli na zadowalające zmniejszenie początkowego kosztu konstrukcji.
5. W zadaniu dotyczącym analizy niezawodności ściskanej półki blachownicy zwrócono uwagę na zależność między symetriami konstrukcji a symetriami obszaru awarii. Zgodnie z wiedzą autora zagadnienie to nie było dotąd poruszane w literaturze pomimo, że jest ono bardzo istotne. Nie uwzględnienie symetrii powierzchni granicznej i zastosowanie standardowych metod analizy niezawodności zazwyczaj prowadzi do grubych błędów w oszacowaniu prawdopodobieństwa awarii. Wydaje się, więc pożądane, aby opracować metody pozwalające identyfikować symetrie obszaru awarii wynikające z symetrii konstrukcji i postaci funkcji granicznych.
6. Obliczenia za pomocą algorytmów prezentowanych w tej pracy zostały wykonane w środowisku R [90]. Jest to darmowa implementacja obiektowego języka $S+$, dedykowanego do analizy danych i statystyki numerycznej. Pakiet R wydaje się

cieszyć dużą popularnością wśród naukowców zajmujących się komputerowymi metodami statystyki. Dzięki temu często można znaleźć implementacje nowatorskich algorytmów zrealizowane w tym środowisku. Tak jest w przypadku metod Annealed Importance Sampling i Linked Importance Sampling, których kody dostępne są pod adresem: <http://www.cs.toronto.edu/~radford/ftp/lis.tar>.

Symulacje Monte Carlo wykonane na potrzeby prezentowanych w pracy zadań, zostały zrealizowane za pomocą obliczeń równoległych. W tym celu wykorzystano pakiet *biopara* [64], który jest dostępny w środowisku R. Pozwala on na równoległe wykonywanie funkcji wywoływanych w interpreterze R. Połączenie w ten sposób kilku komputerów bądź procesorów jest bardzo łatwe, a osiągnięta szybkość komunikacji jest zadowalająca. Dzięki temu pakietowi można również w pełni wykorzystać procesory dwurdzeniowe, które stają się obecnie standardem w komputerach osobistych.

Bibliografia

- [1] T. Abdo, R. Rackwitz. Reliability of uncertain structural systems. *Proc. Finite Elements in Engineering Applications*, strony 161–176, Stuttgart, 1990. INTES GmbH.
- [2] P. Abrahamsen. A review of gaussian random fields and correlation functions - second edition. Raport instytutowy 917, Norsk Regnesentral/Norwegian Computing Center, Oslo, 1997.
- [3] J.R. Adler, J.E. Taylor. Random fields and geometry. W przygotowaniu, 2006.
- [4] R.J. Adler. *The Geometry of Random Fields*. Wiley, London, 1972.
- [5] C. Andrieu, C. P. Robert. Controlled mcmc for optimal sampling. Raport instytutowy, Université Paris-Dauphine, 2001.
- [6] G.L. Ang, A.H-S. Ang, W.H. Tang. Optimal importance sampling density estimator. *Journal of Engineering Mechanics ASCE*, 118(6):1146–63, 1992.
- [7] J. Arbocz, J.H. Starnes Jr. Future directions and challenges in shell stability analysis. *Thin-Walled Structures*, 40:729–754, 2002.
- [8] S.K. Au, J.L. Beck. A new adaptive importance sampling scheme for reliability calculations. *Structural Safety*, 21(2):135–158, 1999.
- [9] Y. K. Belyaev. On the number of exits across the boundary of a region by a vector stochastic processes. *Theory of Probability and Its Applications*, 13(2):320–324, 1968.
- [10] Y.K. Belyaev. Point processes and first passage problems. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, wolumen 2, strony 1–17, Berkeley, 1972. Univ. of California Press.

-
- [11] P. Bühlmann. Bootstraps for time series. *Statistical Science*, 17(1):52–72, 2002.
- [12] A. Biegus. *Probabilistyczna analiza konstrukcji stalowych*. Wydawnictwo Naukowe PWN, Warszawa - Wrocław, 1999.
- [13] V.V. Bolotin. *Wahrscheinlichkeitsmethoden zur Berechnung von Konstruktionen*. VEB Verlag für Bauwesen, Berlin, 1981.
- [14] K. W. Breitung. Asymptotic approximations for multinormal integrals. *Journal of the Engineering Mechanics Division, ASCE*, 110:357–366, 1984.
- [15] K. W. Breitung. *Asymptotic Approximations for Probability Integrals*. Springer-Verlag, 1994.
- [16] K. W. Breitung, R. Rackwitz. Nonlinear combination of load processes. *Journ. Struct. Mech.*, 10:145–166, 1982.
- [17] C. G. Bucher. Adaptive sampling - an iterative fast monte carlo procedure. *Structural Safety*, 5:119–126, 1988.
- [18] G. Celeux, M. Hurn, C.P. Robert. Computational and inferential difficulties with mixture posterior distributions. *J. American Statist. Assoc.*, 95:957–970, 2000.
- [19] C.A. Cornell. A probability-based structural code. *Journal of American Concrete Institute*, 66:974–985, 1969.
- [20] H. Cramer, M. R. Leadbetter. *Stationary and Related Stochastic Processes*. Wiley, 1967.
- [21] P.T. de Boer, D.P. Kroese, S. Mannor, Rubinstein R.Y. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, 2005.
- [22] G. Defaux, R. Rackwitz. Improvement of outcrossing rates by importance sampling. *EURODYN2002*, Munich, Germany, September 2002.
- [23] A. Der Kiureghian, J.-B. Ke. The stochastic finite element method in structural reliability. *Probabilistic Engineering Mechanics*, 3(2):83–91, 1988.
- [24] A. Der Kiureghian, H-Z. Lin, S-J. Hwang. Second-order reliability approximations. *Journal of Engineering Mechanics Division, ASCE*, 113(8):1208–1225, 1987.
- [25] A. Der Kiureghian, Liu P-L. Structural reliability under incomplete probability information. *Journal of Engineering Mechanics, ASCE*, 112(1):85–104, 1986.

-
- [26] O. Ditlevsen. Random field interpolation between point by point measured properties. P.D. Spanos, C.A. Brebbia, redaktorzy, *Proc. 1st Int. Conf. on Computational Stochastic Mechanics*, strony 801–812, Corfu, Greece, September 17-19 1991. Computational Mechanics Publications.
- [27] O. Ditlevsen, P. Bjerager. Methods of structural systems reliability. *Structural Safety*, 3(3-4):195–229, 1986.
- [28] O. Ditlevsen, H. O. Madsen. *Structural Reliability Methods*. Wiley, 1996.
- [29] K. Doliński. Importance sampling techniques in reliability calculations. *Prace IPPT*, 37, 1988.
- [30] K. Doliński, R. Stocki. Reliability-based structural optimisation accounting for manufacturing and material quality. *Engineering Transactions*, 49(4):623–636, 201.
- [31] B. Efron, R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.
- [32] I. Enevoldsen, J.D. Sørensen. Optimization algorithms for calculation of the joint design point in parallel systems. *Structural Optimization*, 4:121–127, 1992.
- [33] I. Enevoldsen, J.D. Sørensen. Reliability-based optimization of series systems of parallel systems. *Journal of Structural Engineering, ASCE*, 119:1069–1084, 1993.
- [34] S. Engelund, R. Rackwitz. A benchmark study on importance sampling in structural reliability. *Structural Safety*, 12:255–276, 1993.
- [35] S. Engelund, R. Rackwitz, C. Lange. Approximations of first-passage times for differentiable processes based on higher-order threshold crossings. *Probabilistic Engineering Mechanics*, 10:53–60, 1995.
- [36] M.H. Faber. Risk and safety in civil, surveying and environmental engineering. Lecture Notes, dostępne na: www.ibk.ethz.ch, 2006.
- [37] J. Ferry Borges, M. Castanheta. *Structural Safety*. Laboratório Nacional de Engenharia Civil, 1971.
- [38] B. Fiessler, H-J. Neumann, R. Rackwitz. Quadratic limit states in structural reliability. *Journal of Engineering Mechanics, ASCE*, 105(4):661–676, 1979.

-
- [39] A.M. Freudenthal. Safety and probability of structural failure. *Transactions, ASCE*, 121:1337–1375, 1956.
- [40] M. Fujita, R. Rackwitz. Updating first- and second-order reliability estimates by importance sampling. *Structural Engineering / Earthquake Engineering*, 5:53–59, 1988.
- [41] A. E. Gelfand, S. K. Sahu. On markov chain monte carlo acceleration. *Journal of Computational and Graphical Statistics*, 3(3):261–276, 1994.
- [42] A. Gelman, X.-L. Meng. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*, 13(2):163–185, 1998.
- [43] A. Gelman, G.O. Roberts, W. R. Gilks. *Bayesian Statistics 5*, rozdzia/1 Efficient Metropolis jumping rules, strony 599–607. Oxford University Press, 1996.
- [44] C.J. Geyer, E.A. Thompson. Annealing markov chain monte carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90:909—920, 1995.
- [45] R. G. Ghanem, P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, 1991.
- [46] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, W.R. Spiegelhalter, redaktorzy. *Markov Chain Monte Carlo in Practice*. CRC Press, 1996.
- [47] S. Gollwitzer, B. Kirchgäßner, R. Fischer, R. Rackwitz. Permas-ra/strurel system of programs for probabilistic reliability analysis. *Structural Safety*, 28(1-2):108–129, 2006.
- [48] H. Greenberg. *Integer programming*. Academic Press, New York, 1971.
- [49] H. Haario, E. Saksman, J. Tamminen. Adaptive proposal distribution for random walk metropolis algorithm. *Computational Statistics*, 14:375–395, 1999.
- [50] A. M. Hasofer, N. C. Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division, ASCE*, 100:111–121, 1974.
- [51] A.M. Hasofer, N.C. Lind. Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division, ASCE*, 100(1):111–121, 1974.
- [52] W. K. Hastings. Montec carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

-
- [53] T. C. Hesterberg. *Advances in Importance Sampling*. Praca doktorska, Stanford University, 1988.
- [54] M. Hohenbichler, R. Rackwitz. Non-normal dependent vectors in structural safety. *Journal of the Engineering Mechanics Division, ASCE*, 107:1227–1238, 1981.
- [55] M. Hohenbichler, R. Rackwitz. Improvement of second-order reliability estimates by importance sampling. *Journal of the Engineering Mechanics Division, ASCE*, 114:2195–2199, 1988.
- [56] T. Homem-de Mello, R.Y. Rubinstein. Rare event estimation for static models via cross-entropy and importance sampling. Submitted for publication, 2002.
- [57] W. Härdle, J. Horowitz, J.P. Kreiss. Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459, 2003.
- [58] Y. Ibrahim. Observations of applications of importance sampling in structural reliability analysis. *Structural Safety*, 9:269–281, 1991.
- [59] S. Jendo, K. Kolanek. Evaluation of adaptive monte carlo methods for reliability analysis. *Journal of Konbin*, wolumen 1, strony 219–228, Kraków, 2006. Air Force Institute of Technology.
- [60] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671—680, 1983.
- [61] W. Knabe, J. Przewłócki, Różyński. Spatial averages for linear elements for two-parameter random fields. *Probabilistic Engineering Mechanics*, 13(3):147–167, 1998.
- [62] N. Kuschel, R. Rackwitz. Optimal design under time-variant reliability constraints. *Structural Safety*, 22:113–128, 2000.
- [63] S.N. Lahiri. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):389–404, 1999.
- [64] P. Lazar, D. Schoenfeld. *biopara: Self-contained parallel system for R*. R package version 1.4.
- [65] S.M. Lewis, A.E. Raftery. Estimating bayes factors via posterior simulation with the laplace-metropolis estimator. *Journal of the American Statistical Association*, 92:648–655, 1997.

-
- [66] C.-C. Li, A. Der Kiureghian. Optimal discretization of random fields. *Journal of Engineering Mechanics ASCE*, 119(6):1136–1154, 1993.
- [67] Na (Michael) Li. *rsprng: R interface to SPRNG (Scalable Parallel Random Number Generators)*, 2006. R package version 0.3-3.
- [68] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag, New York, 2001.
- [69] W.-K. Liu, T. Belytschko, A. Mani. Probabilistic finite elements for non linear structural dynamics. *Comp. Meth. App. Mech. Eng.*, 56:61–86, 1986.
- [70] W.-K. Liu, T. Belytschko, A. Mani. Random field finite elements. *Int. J. Num. Meth. Eng.*, 26(10):1831–1845, 1986.
- [71] H. O. Madsen, F. Friis Hansen. A comparison of some algorithms for reliability based structural optimization and sensitivity analysis. R. Rackwitz, P. Thoft-Christensen, redaktorzy, *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf.*, strony 443—451. Springer-Verlag, 1992.
- [72] Michael Mascagni, Ashok Srinivasan. Algorithm 806: Sprng: a scalable library for pseudorandom number generation. *ACM Trans. Math. Softw.*, 26(3):436–461, 2000.
- [73] R. E. Melchers. *Structural Reliability Analysis and Prediction*. Wiley, 1999.
- [74] R.E. Melchers. Importance sampling in structural systems. *Structural Safety*, 6:3–10, 1989.
- [75] R.E. Melchers. Search-based importance sampling. *Structural Safety*, 9(2):117–128, 1990.
- [76] R.E. Melchers. Simulation in time-invariant and time-variant reliability problems. R. Rackwitz, P. Thoft-Christensen, redaktorzy, *Reliability and Optimization of Structural Systems '91, Proc. 4th IFIP WG 7.5 Conf., Munich, 11-13 September 1991*, strony 39–82. Springer-Verlag, 1992.
- [77] X.L. Meng, W.H. Wong. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, 6:831–860, 1996.
- [78] K. L. Mengersen, R. L. Tweedie. Rates of convergence of the hastings and metropolis algorithms. *Annals of Statistics*, 24(1):101–121, 1996.

-
- [79] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- [80] T. Most, C. Bucher, Y. Schorling. Dynamic stability analysis of non-linear structures with geometrical imperfections under random loading. *Journal of Sound and Vibration*, 276:381–400, 2004.
- [81] Y. Murotsu, M. Okada, M. Yonezawa, K. Taguchi. Reliability assessment of redundant structures. *Structural Safety and Reliability, ICOSSAR 81*, strony 315–321. Elsevier Scientific Publishing Company, 1981.
- [82] J.M. Murzewski. *Niezawodność konstrukcji inżynierskich*. Arkady, 1989.
- [83] A. Nataf. Determination des distributions de probabilités dont les marges sont données. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 255:42–43, 1962.
- [84] R. M. Neal. Annealed importance sampling. *Statistics and Computing*, 11:125–139, 2001.
- [85] R. M. Neal. Estimating ratios of normalizing constants using linked importance sampling. Technical Report No. 0511, Dept. of Statistics, University of Toronto, 2005.
- [86] M.A. Newton, A.E. Raftery. Approximate bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society*, 56:3–48, 1994.
- [87] P. Śniady. *Podstawy stochastycznej dynamiki konstrukcji*. Ofic.Wyd.Polit.Wrocł., 2000.
- [88] C. Pasarica, A. Gelman. Adaptively scaling the metropolis algorithm using expected squared jumped distance. 2003.
- [89] A. Plucińska, E. Pluciński. *Probabilistyka*. WNT, 2000.
- [90] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.
- [91] R. Rackwitz. On the combination of non-stationary rectangular wave renewal processes. *Structural Safety*, 13(1+2):21–28, 1993.

-
- [92] R. Rackwitz. Computational techniques in stationary and non-stationary load combination - a review and some extensions. *Journal of Structural Engineering, SERC*, 25(1):1–20, 1998.
- [93] R. Rackwitz. Reliability analysis - a review and some perspectives. *Structural Safety*, 23:365–395, 2001.
- [94] R. Rackwitz, B. Fiessler. Structural reliability under combined random load sequences. *Computers and Structures*, 9:484–494, 1978.
- [95] S. O. Rice. Mathematical analysis of random noise. *Bell Systems Tech. J.*, 23:282–332, 1944.
- [96] C. P. Robert, G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1999.
- [97] G. O. Roberts, R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, 83:96–110, 1996.
- [98] G.O. Roberts, A. Gelman, W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Probability*, 7:110–120, 1997.
- [99] G.O. Roberts, J.S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical Science*, 16:351–367, 2001.
- [100] M. Rosenblatt. Remarks on multivariate transformation. *The Annals of Mathematical Statistics*, 23:470–472, 1952.
- [101] R.Y. Rubinstein, D.P. Kroese. *The Cross-Entropy Method*. Springer-Verlag, 2004.
- [102] R.Y. Rubinstein, B. Melamed. *Modern Simulation and Modeling*. Wiley, 1998.
- [103] Engelund S., Rackwitz R. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety*, 12:255–276, 1993.
- [104] K.C. Sarma, Adeli H. Cost optimization of concrete structures. *Journal of Structural Engineering*, 124(5):570–578, 1988.
- [105] G. Schall, M. Faber, R. Rackwitz. The ergodicity assumption for aea atates in the reliability assessment of offshore structures. *Journ. Offshore Mech. and Arctic Eng., ASME*, 113:241–246, 1991.

-
- [106] 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.
- [107] G. I. Schueller, Stix R. A critical appraisal of methods to determine failure probabilities. *Structural Safety*, 4:293–309, 1987.
- [108] G.I. Schuëller, H.J. Pradlwarter. Computational stochastic structural analysis (cossan) – a software tool. *Structural Safety*, 28(1-2):68–82, 2006.
- [109] M. Shinozuka. Probability of failure under random loading. *Journ. Eng. Mech. Div., ASCE*, 90:147–170, 1964.
- [110] K. Sobczyk. *Zarys Teorii Prawdopodobieństwa*. Wydawnictwo Politechniki Poznańskiej, 1985.
- [111] K. Sobczyk. *Stochastyczne Równania Różniczkowe*. Wydawnictwa Naukowo-Techniczne, 1996.
- [112] J.D. Sørensen, I. Enevoldsen. Reliability of systems. *Materials of the training course on Reliability-Based Structural Design, Poznań, 20-21, October, 1996*.
- [113] A. Stachurski, A. P. Wierzbicki. *Podstawy Optymalizacji*. Oficyna Wydawnicza Politechniki Warszawskiej, 1999.
- [114] R. Stocki. *Niezawodnościowa optymalizacja konstrukcji prętowych w zakresie dużych przemieszczeń - teoria i program komputerowy*. Praca doktorska, Instytut Podstawowych Problemów Techniki PAN, 1999.
- [115] 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.
- [116] B. Sudret, A. Der Kiureghian. Stochastic finite element methods and reliability a state-of-the-art report. Raport instytutowy UCB/SEMM-2000/08, Department of Civil and Environmental Engineering, University of California, Berkeley, 2000.
- [117] J. Tamminen. Adaptive markov chain monte carlo algorithms with geophysical applications. Finnish Meteorological Institute Contributions No. 47, 2004.
- [118] R.L. Taylor. *FEAPpv - - A Finite Element Analysis Program*. Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, California, 2005.

- [119] B.H. Thacker, D.S. Riha, S.H.K. Fitch, L.J. Huysse, J.B. Pleming. Probabilistic engineering analysis using the nessus software. *Structural Safety*, 28(1-2):83–107, 2006.
- [120] P. Thoft-Christensen, Y. Murotsu. *Application of Structural Systems Reliability Theory*. Springer-Verlag, 1986.
- [121] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(5):1701–1762, 1994.
- [122] Y. Tsompanakis, M. Papadrakakis. Robust and efficient methods for reliability-based structural optimization. M. Papadrakakis, redaktor, *Fourth International Colloquium on Computation of Shell & Spatial Structures*, Chania - Crete, 2000. IASS-IACM.
- [123] L. Tvedt. Proban – probabilistic analysis. *Structural Safety*, 28(1-2):150–163, 2006.
- [124] E. VanMarcke. *Random Fields: Analysis and Synthesis*. The MIT Press, 1983.
- [125] E. VanMarcke, M. Grigoriu. Stochastic finite element analysis of simple beams. *Journal of Engineering Mechanics ASCE*, 109(5):1203–1214, 1983.
- [126] D. Veneziano, M. Grigoriu, C.A. Cornell. Vector-process models for system reliability. *Journal of the Engineering Mechanics Division, ASCE*, 103:441–460, 1977.
- [127] S. Wang, K.L. Teo, H.W.J. Lee. A new approach to nonlinear mixed discrete programming problems. *Engineering Optimization*, 30:249–262, 1998.
- [128] Z. Waszczyszyn, C. Cichoń, M. Radwańska. *Metoda elementów skończonych w stateczności konstrukcji*. ARKADY, 1990.
- [129] W. Wierzbicki. W sprawie bezpieczeństwa pręta wyciąganego osiowo. *Czasopismo techniczne*, 50:273–277, 1937.
- [130] Y-T Wu. An adaptive importance sampling method for structural system reliability analysis. *Reliability technology - 1992; Proceedings of the Symposium, ASME Winter Annual Meeting*, strony 217–231, Anaheim, CA; UNITED STATES, 8-13 Nov. 1992.
- [131] J. Zhang, B. Ellingwood. Orthogonal series expansion of random fields in reliability analysis. *Journal of Engineering Mechanics*, 120(12):2660–2677, 1994.