

Wydział Informatyki, Elektrotechniki i Telekomunikacji  
Uniwersytet Zielonogórski

Rozkłady  $\alpha$ -stabilne w ewolucyjnych algorytmach  
globalnej optymalizacji parametrycznej

Przemysław Prętki

Rozprawa doktorska  
promotor: dr hab. inż. Andrzej Obuchowicz

Zielona Góra, 13 Luty 2008

# Spis treści

Wstęp	1
<b>1 Cel i zakres pracy</b>	<b>10</b>
1.1 Cel rozprawy	10
1.2 Wkład rozprawy do aktualnego stanu wiedzy	10
1.3 Zadania szczegółowe	12
1.4 Układ rozprawy	12
<b>2 Rozkłady stabilne</b>	<b>14</b>
2.1 Stabilna zmienna losowa	14
2.1.1 Wybrane właściwości	15
2.2 Wielowymiarowy rozkład stabilny	21
2.2.1 Izotropowy rozkład $\alpha$ -stabilny	25
2.3 Podsumowanie	30
<b>3 Klasyczne algorytmy ewolucyjne</b>	<b>34</b>
3.1 Podstawy optymalizacji ewolucyjnej	34
3.2 Ewolucyjne algorytmy fenotypowe	37
3.2.1 Strategie ewolucyjne	41
3.2.2 Programowanie ewolucyjne	43
3.2.3 Przeszukiwanie ewolucyjne z miękką selekcją	44
3.3 Podsumowanie	45
<b>4 Mutacja stabilna w fenotypowych algorytmach ewolucyjnych</b>	<b>47</b>
4.1 Efekt symetrii	47
4.1.1 Badania eksperymentalne	50
4.1.2 Podsumowanie	53
4.2 Mutacja o sferycznej symetrii	53
4.2.1 Eksploracja a eksploatacja - kompromis według mutacji stabilnej	53
4.2.2 Efekt otoczenia	64
4.2.3 Analiza zbieżności lokalnej algorytmu ewolucyjnego	71
4.2.4 Odporność mutacji stabilnej	77
4.2.5 Adaptacja izotropowego rozkładu stabilnego	85
4.3 Mutacja bazująca na dyskretnej mierze spektralnej	96
4.3.1 Adaptacja dyskretnej miary spektralnej	102
4.3.2 Podsumowanie	108

<b>5</b>	<b>Mutacja kierunkowa</b>	<b>110</b>
5.1	Wstęp i motywacja . . . . .	110
5.1.1	Kłątwa wymiarowości dla mutacji izotropowych . . . . .	111
5.1.2	Problem symetrii rozkładu . . . . .	113
5.2	Rozkłady kierunkowe . . . . .	115
5.3	Badania eksperymentalne . . . . .	121
5.4	Podsumowanie . . . . .	132
<b>6</b>	<b>Algorytmy ewolucyjne w zadaniach projektowania systemów wykrywania uszkodzeń</b>	<b>134</b>
6.1	Statyczne systemy FDI oparte o neuronowe generatory residuum . . . . .	135
6.1.1	Sztuczne sieci neuronowe . . . . .	135
6.1.2	Niepewność parametryczna modelu neuronowego . . . . .	137
6.1.3	Minimalizacja niepewności parametrycznej modelu neuronowego . . . . .	140
6.1.4	Ocena dokładności linearyzacji modelu neuronowego w zadaniach wnioskowania statystycznego . . . . .	147
6.1.5	Ewolucyjne przedziały ufności . . . . .	149
6.2	System wykrywania uszkodzeń systemu przemysłowego . . . . .	152
6.3	Podsumowanie . . . . .	156
<b>7</b>	<b>Algorytmy ewolucyjne w zadaniach projektowania nieliniowych obserwatorów stanu</b>	<b>158</b>
7.1	Wstęp i podstawowe pojęcia . . . . .	158
7.2	Projektowanie odpornych obserwatorów nieliniowych . . . . .	161
7.3	Metoda Monte-Carlo . . . . .	163
7.4	Aspekty optymalizacyjne . . . . .	163
7.4.1	Dedykowana strategia ewolucyjna . . . . .	164
7.5	Badania symulacyjne . . . . .	166
7.6	Podsumowanie . . . . .	168
<b>8</b>	<b>Wnioski końcowe i kierunki dalszych badań</b>	<b>170</b>
8.1	Podsumowanie wyników pracy . . . . .	170
8.2	Plany dalszych badań . . . . .	171
<b>A</b>	<b>Funkcja gęstości izotropowego rozkładu stabilnego</b>	<b>180</b>
<b>B</b>	<b>Twierdzenie o istnieniu momentów</b>	<b>181</b>
<b>C</b>	<b>Parametr skali w generatorze gęstości</b>	<b>183</b>
<b>D</b>	<b>Rozkład normy izotropowego wektora stabilnego</b>	<b>184</b>
<b>E</b>	<b>Rozkład zmiennej losowej <math>\ Z + z_0\ </math>, <math>Z \sim X_\alpha^s</math></b>	<b>185</b>
<b>F</b>	<b>Rozkład zmiennej losowej <math>V = \ Z + z_0\ ^2 / \ z_0\ ^2</math></b>	<b>187</b>
<b>G</b>	<b>Rozkład zmiennej losowej <math>V^{b/2}</math></b>	<b>188</b>
<b>H</b>	<b>Funkcje testowe</b>	<b>189</b>

## Podziękowania

W pierwszej kolejności chciałbym podziękować promotorowi dr hab. Andrzejowi Obuchowiczowi, za zaufanie jakim obdarzył mnie na początku pracy na Uniwersytecie, wprowadzenie w świat obliczeń ewolucyjnych oraz wiele inspirujących dyskusji, bez których niniejsza rozprawa nie mogłaby powstać.

Szczególne wyrazy wdzięczności należą się również Marcinowi Witczakowi za grad pomysłów oraz znacząca pomoc w realizacji rozdziałów 6, 7. Autor pragnie także wyrazić swoją wdzięczność Ministerstwu Nauki i Szkolnictwa Wyższego, przy którego wsparciu (grant promotorski Nr 3T11A01530) powstała rozprawa. W końcu, niezwykle podziękowania zmierzają w kierunku Mojego Honorulka, której nieustanne wsparcie oraz wiara w to co robię, rozświetlały tunel, na końcu którego ukryta była ostateczna wersja niniejszej rozprawy.

# Wykaz oznaczeń

## Symbole

$\phi(\cdot)$  - funkcja celu,

$\mathbf{x}^*$  - optimum globalne funkcji celu

$\mathbf{x}_\epsilon$  - rozwiązanie  $\epsilon$ -optymalne,

$\mathcal{D}$  - przestrzeń przeszukiwań,

$\mathbb{D}$  -  $\sigma$ -ciało podzbiorów przestrzeni przeszukiwań,

$\mathbf{A}_D$  - deterministyczny algorytm optymalizacji,

$\mathbf{A}_S$  - stochastyczny algorytm optymalizacji,

$\|\cdot\|$  - norma euklidesowa wektora,

$\pi_S$  - rozkład eksploracyjny,

$\varphi(\cdot)$  - funkcja charakterystyczna,

$\alpha$  - indeks stabilności,

$\sigma$  - parametr skali rozkładu stabilnego,

$\beta$  - parametr skośności rozkładu stabilnego,

$\mu$  - parametr położenia rozkładu stabilnego,

$\kappa$  - parametr rozproszenia rozkładu kierunkowego,

$S_\alpha(\sigma, \beta, \mu)$  - stabilna zmienna losowa,

$S_\alpha S(\sigma)$  - symetryczna, stabilna zmienna losowa ( $\beta = 0, \mu = 0$ ),

$Z_{i:\lambda}$  -  $i$ -ta statystyka rankingowa z ciągu  $\lambda$ -elementowego,

$\Gamma$  - miara spektralna wektora stabilnego,

$\xi$  - punkty skupienia dyskretnej miary spektralnej,

$\gamma$  - wartości wag dyskretnej miary spektralnej,

$\mathbf{X}_\xi^\gamma(\alpha)$  - wektor  $\alpha$ -stabilny opisany dyskretną miarą spektralną  $\Gamma = (\xi, \gamma)$ ,

$\mathcal{M}(\mu, \kappa)$  - rozkład kierunkowy z wyróżnionym kierunkiem  $\mu$ ,

**vMF** - rozkład kierunkowy von Mises-Fishera,

$\mathbf{I}_n$  - macierz jednostkowa o rozmiarze  $n \times n$ ,

$\mathbf{u}^{(n)}$  - rozkład równomierny na powierzchni  $n$ -wymiarowej, jednostkowej kuli euklidesowej,

$\mathbf{X}^s$  - izotropowy wektor losowy,  
 $\mathbf{X}_\alpha^s$  - izotropowy stabilny wektor losowy,  
 $\mathbf{X}_\alpha$  - stabilny wektor losowy o niezależnych składowych  $X_i \sim S_\alpha S(\sigma)$ ,  
 $S_r^{(n)}(\mathbf{x})$  -  $n$ -wymiarowa kula euklidesowa o promieniu  $r$  i środku  $\mathbf{x}$ ,  
 $\partial S_r^{(n)}(\mathbf{x})$  - powierzchnia  $S_r^{(n)}(\mathbf{x})$ ,  
 $n_p$  - liczba parametrów modelu neuronowego,  
 $n_u$  - liczba wejść modelu neuronowego,  
 $n_h$  - liczba neuronów warstwy ukrytej,  
 $\Theta_{c,L}$  - obszar ufności dla parametrów modelu liniowego,  
 $Y_{c,L}$  - przedział ufności dla odpowiedzi modelu liniowego,  
 $\Theta_{c,N}$  - obszar ufności dla parametrów modelu nieliniowego,  
 $Y_{c,N}$  - przedział ufności dla odpowiedzi modelu nieliniowego,

## Funkcje

$\mathbf{I}(A)$  - funkcja wskaźnikowa zbioru  $A$ ,  
 $\mathbf{I}_n(x)$  - funkcja Bessela pierwszego rodzaju,  
 $\Gamma(x)$  - funkcja Gamma,  
 $G_\alpha(x)$  - generator gęstości izotropowego wektora stabilnego,  
 $\det(\mathbf{A})$  - wyznacznik macierzy  $\mathbf{A}$ ,  
 $\text{trace}(\mathbf{A})$  - ślad macierzy  $\mathbf{A}$ ,

## Skróty

**DSM** - Dyskretna Miara Spektralna,  
**ESSS** - algorytm przeszukiwań ewolucyjnych z miękką selekcją,  
**ESSS $_\alpha$**  - algorytm przeszukiwań ewolucyjnych z miękką selekcją oraz mutacją stabilną,  
**ARS** - algorytm adaptacyjnych przeszukiwań losowych,  
**ARS $_\alpha$**  - algorytm adaptacyjnych przeszukiwań losowych oraz mutacją stabilną,

# Wstęp

Termin optymalizacja, a w szczególności optymalizacja globalna, w dobie burzliwego rozwoju komputerowych technik obliczeniowych, nabiera szczególnego znaczenia. Obecnie, w niemalże każdej dziedzinie życia, zaobserwować można potrzebę rozwiązywania problemów, które mają na celu poprawienie pewnych aspektów już istniejących rozwiązań. Uwzględniając ograniczenia wynikające z natury obliczeń wykonywanych na maszynach cyfrowych najbardziej adekwatnej definicji problemu optymalizacji globalnej dostarcza definicja:

**Definicja 1** *Problem optymalizacji globalnej polega na odnalezieniu przynajmniej jednego punktu  $\mathbf{x}_\epsilon$  należącego do zbioru*

$$M_\epsilon = \{\mathbf{x} \in \mathcal{D} : \phi(\mathbf{x}) < \phi(\mathbf{x}^*) + \epsilon\}, \quad \epsilon > 0, \quad (1)$$

gdzie  $\mathcal{D}$  jest zbiorem rozwiązań dopuszczalnych,  $\phi : \mathcal{D} \rightarrow \mathbb{R}$  funkcją celu, oraz

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathcal{D}} \phi(\mathbf{x}), \quad (2)$$

W literaturze poświęconej optymalizacji globalnej, optimum  $\mathbf{x}_\epsilon$  określane jest mianem minimum pierwszego rodzaju [121] lub rozwiązaniem  $\epsilon$ -optymalnym [119].

Na przestrzeni ostatnich dziesięcioleci zaproponowano wiele numerycznych metod mogących posłużyć do rozwiązania problemu wyrażonego w definicji 1. Niemalże wszystkie algorytmy optymalizacji numerycznej są procedurami iteracyjnymi. Oznacza to, że dysponując pewnym początkowym przybliżeniem rozwiązania globalnego, na poszczególnych etapach procesu optymalizacji, rozwiązanie to jest modyfikowane według znamienych, dla każdego algorytmu, reguł. Charakterystyczne jest, że każdy algorytm, bez względu na swoją strukturę wewnętrzną, w każdej iteracji typuje pewne przybliżenie optimum globalnego. Tym samym powstaje pewien ciąg rozwiązań  $v_m$ , uporządkowany względem rosnących indeksów  $T = \{1, 2, \dots\}$ , t.j.

$$v_m = \{(\mathbf{x}_1, \phi_1), (\mathbf{x}_2, \phi_2), \dots, (\mathbf{x}_m, \phi_m)\}, \quad (3)$$

gdzie  $\mathbf{x}_k \in \mathcal{D}$  oraz  $\phi_k = \phi(\mathbf{x}_k)$  dla  $k = 1, \dots, m$ . Dyskretny zbiór indeksów  $T$  utożsamiany jest z kolejnymi iteracjami. W celu zdefiniowania algorytmu optymalizacji globalnej oznaczmy zbiór  $\mathcal{V}_m = (\mathcal{D} \times \mathbb{R})^m$  jako przestrzeń wszystkich możliwych ciągów  $v_m$  oraz przyjmijmy  $\mathcal{V} = \cup_{m>0} \mathcal{V}_m$ . W literaturze poświęconej optymalizacji globalnej można odnaleźć wiele kryteriów służących do klasyfikacji poszczególnych technik [119]. Zdecydowaną większość algorytmów można jednak przypisać do jednej z dwóch głównych klas: rozwiązań deterministycznych bądź stochastycznych.

**Definicja 2** *Deterministycznym algorytmem optymalizacji  $\mathbf{A}_D$ , nazywamy odwzorowanie*

$$\mathbf{A}_D : \mathcal{V} \rightarrow \mathcal{D}. \quad (4)$$

W przypadku metod deterministycznych odwzorowanie (4) jest ściśle określoną, niekoniecznie różnowartościową funkcją, taką że

$$\mathbf{A}_D(v_m^{(1)}) \neq \mathbf{A}_D(v_m^{(2)}) \Rightarrow v_m^{(1)} \neq v_m^{(2)}, \quad (5)$$

Oznacza to, że algorytm deterministyczny posiadając listę odwiedzonych wcześniej punktów przestrzeni przeszukiwań, przekształca ciąg  $\mathbf{v}_m$  na jeden i zawsze ten sam element zbioru  $\mathcal{D}$ . Co więcej, ustalone warunki początkowe algorytmu w pełni determinują wszystkie wyrazy ciągu rozwiązań (3). Niezwykle silna wrażliwość na dobór warunków początkowych powoduje, że skuteczność algorytmów deterministycznych zagwarantowana jest tylko w przypadku pewnej, niezwykle wąskiej klasy funkcji celu [81].

Stochastyczne algorytmy optymalizacji, w odróżnieniu od swoich deterministycznych odpowiedników, ten sam ciąg  $v_m$ , mogą przekształcać na różne punkty przestrzeni  $\mathcal{D}$ . Formalna definicja, podkreślająca probabilistyczny charakter algorytmu, może zostać wyrażona w następujący sposób

**Definicja 3** *Niech  $(\mathcal{V}, \mathbb{V}, \nu)$  oznacza przestrzeń probabilistyczną, gdzie  $\mathbb{V}$  jest  $\sigma$ -ciałem zbioru  $\mathcal{V}$ , a  $\nu$  jest pewną miarą probabilistyczną. Wówczas algorytmem stochastycznym nazywamy następującą funkcję mierzalną*

$$\mathbf{A}_S : (\mathcal{V}, \mathbb{V}) \rightarrow (\mathcal{D}, \mathbb{D}), \quad (6)$$

gdzie  $\mathbb{D}$  oznacza  $\sigma$ -ciało zbioru rozwiązań dopuszczalnych  $\mathcal{D}$ .

Kluczową rolę dla analizy działania algorytmu stochastycznego odgrywa pojęcie tzw. rozkładu eksploracyjnego.

**Definicja 4** *Rozkładem eksploracyjnym  $\pi_S$ , nazywamy miarę probabilistyczną indukowaną przez algorytm stochastyczny  $\mathbf{A}_S$ , tj.  $\pi_S = \nu \circ \mathbf{A}_S^{-1}$  taką, że*

$$\pi_S(D \in \mathbb{D}) = \nu(\{\mathbf{v} \in \mathcal{V} : \mathbf{A}_S(\mathbf{v}) \in D\}). \quad (7)$$

Tym samym, rozkład eksploracyjny określa prawdopodobieństwo otrzymania kolejnego przybliżenia rozwiązania z dowolnego podzbioru przestrzeni rozwiązań  $\mathcal{D}$ .

Istnieje przynajmniej kilka powodów, dla których na przestrzeni ostatnich lat, algorytmy stochastyczne zaczęły odgrywać dominującą rolę w optymalizacji globalnej. Bodajże najistotniejszą zaletą tej klasy metod, jest ich odporność na nieregularności funkcji celu [118, 138]. Zdecydowana większość algorytmów stochastycznych nie wymaga różniczkowalności, a nawet ciągłości funkcji celu [25, 53, 137]. Co więcej, techniki te mogą być z powodzeniem stosowane w przypadku wystąpienia niepewności o charakterze losowym związanych z wartościami funkcji celu [12, 13, 14, 80, 92, 114] lub jej niestacjonarnego charakteru [28, 77, 89, 120, 124, 125]. Ponadto, na ogół prosta i intuicyjna struktura algorytmów stochastycznych sprawia, iż w bardzo łatwy sposób można wzbogacić procedurę optymalizacyjną o dodatkowe heurystyki, możliwość adaptacji czy dokonać zrównoleglenia procesu optymalizacji [9, 78, 79, 90].



Niniejsza rozprawa dotyczy wyróżnionej klasy problemów tzw. zadań parametrycznej optymalizacji ciągłej [81, 119] oraz specjalnej grupy metod dedykowanych do ich rozwiązywania - algorytmów ewolucyjnych. Cechą charakterystyczną, decydującą o odrębności poruszanej klasy problemów, jest utożsamienie zbioru rozwiązań dopuszczalnych  $\mathcal{D}$  ze skończeniem wymiarową przestrzenią euklidesową  $\mathbb{R}^n$ . W dalszej części rozprawy, o ile nie zostanie to zaznaczone inaczej, przestrzeń przeszukiwań zawsze będzie rozumiana właśnie w taki sposób. Algorytmy ewolucyjne (AE) stanowią szczególny przypadek stochastycznych, metod optymalizacji globalnej. Ich główna idea polega na cyklicznym przetwarzaniu zbioru punktów z dziedziny rozwiązań dopuszczalnych zwanych próbą losową lub populacją. W odróżnieniu od metod pasywnych [8, 119], dla których rozwiązanie otrzymuje się poprzez losowanie z pewnym, ustalonym rozkładem prawdopodobieństwa, dla technik ewolucyjnych, rozkład ten zmienia się podczas procesu optymalizacji. Adaptacja rozkładu eksploracyjnego następuje dzięki mechanizmom naśladującym proces ewolucji żywych organizmów. Największego znaczenia nabierają tutaj takie zjawiska jak mutacja kodów fenotypowych [10, 109] oraz naturalna selekcja [74, 85].

Algorytmy ewolucyjne od momentu swojego powstania niemalże całkowicie, zdominowane zostały przez zastosowanie mechanizmu eksploracyjnego opartego o losowy rozkład normalny  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Słynne twierdzenie *No Free Lunch Theorem* [134] oraz liczne przykłady zamieszczone w literaturze [85, 135, 136], dobitnie dowodzą, że rozkład normalny, w wielu przypadkach, nie stanowi podstawy do konstrukcji najefektywniejszego operatora mutacji.

Niniejsza praca wychodzi na przeciw takiemu stanowi rzeczy. Jej głównym celem jest zaprojektowanie efektywnych algorytmów ewolucyjnej optymalizacji globalnej w oparciu o stabilne rozkłady eksploracyjne. W tym celu, w centrum zainteresowania rozprawy leżą takie aspekty optymalizacyjne jak: odporność mutacji stabilnej na wybór warunków początkowych, rozstrzygnięcie kompromisu pomiędzy eksploatacją a eksploracją przestrzeni rozwiązań w obrębie jednej procedury optymalizacyjnej, analiza zbieżności lokalnej, zniwelowanie efektu otoczenia oraz efektu symetrii poprzez zaprojektowanie specjalnych, stabilnych operatorów ewolucyjnych.

# Rozdział 1

## Cel i zakres pracy

### 1.1 Cel rozprawy

Niniejsza rozprawa dotyczy wykorzystania stabilnych rozkładów prawdopodobieństwa w algorytmach ewolucyjnej optymalizacji globalnej. W szczególności jej głównym celem jest zaprojektowanie efektywnych oraz odpornych operatorów mutacji dla popularnych algorytmów fenotypowych. Analiza wstępnych rezultatów, uzyskanych w początkowej fazie badań, skłoniły autora do sformułowania następującej tezy pracy:

Operator mutacji oparty o rozkłady stabilne pozwala na ustalenie efektywnego kompromisu pomiędzy eksploracją oraz eksploatacją przestrzeni rozwiązań, co prowadzi do zwiększenia skuteczności algorytmów ewolucyjnych w zadaniach globalnej optymalizacji parametrycznej.

W celu potwierdzenia powyższej tezy, niezbędne stało się przeprowadzenie kompleksowej analizy skuteczności optymalizacyjnej wybranych algorytmów ewolucyjnych, zarówno pod względem zbieżności lokalnej jak i globalnej.

### 1.2 Wkład rozprawy do aktualnego stanu wiedzy

Algorytmy ewolucyjne od momentu swojego powstania niemalże całkowicie, zdominowane zostały przez zastosowanie mechanizmu eksploracyjnego opartego o losowy rozkład normalny  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Istnieje wiele powodów przemawiających na korzyść takiego właśnie wyboru. Nie bez znaczenia dla rozpopularyzowania rozkładu normalnego na polu obliczeń ewolucyjnych, jest fakt, iż model ten, należy bodajże do jednych z najlepiej poznanych ciągłych rozkładów prawdopodobieństwa. Mianowicie, znane są analityczne postacie funkcji: gęstości, generującej momenty, charakterystycznej, rozkładów brzegowych oraz warunkowych. Prosta struktura modelu, wyrażająca się poprzez wektor wartości oczekiwanych oraz dodatkowo określoną macierz kowariancji, wraz z zagwarantowaną symetrią, umożliwiającą intuicyjne konstruowanie heurystyk usprawniających przeszukiwanie przestrzeni rozwiązań. Ponadto spośród wszystkich rozkładów ciągłych o

ustalonej wariancji, rozkład normalny wnosi największą ilość informacji [52] o przestrzeni rozwiązań, co ma kluczowe znaczenie w przypadku braku wiedzy a-priori o rozważanym problemie. Niemniej jednak stosowanie rozkładu normalnego w operatorze mutacji algorytmów ewolucyjnych nie jest wolne od wad. Wspomniana wyżej symetria względem wektora oczekiwanego, narzuca bowiem rygorystyczne ograniczenie na częstość polepszania rozwiązania [95, 96]. Mianowicie dla najprostszych, liniowych funkcji celu prawdopodobieństwo otrzymania rozwiązania o lepszej jakości niż najlepsze dotychczas znane wynosi zaledwie 0.5 [20]. Ponadto, w przypadku optymalizacji wielowymiarowej, wiele komplikacji rodzi, znamienny dla rozkładu normalnego tzw. efekt otoczenia [85], przejawiający się pogorszeniem jakości otrzymywanych rozwiązań. Kolejnym powodem, przemawiającym na niekorzyść normalnego modelu przeszukiwań jest niewystarczająca elastyczność rozkładu w pogodzeniu zadań *eksploracji* oraz *eksploatacji* przestrzeni rozwiązań [42, 46, 85]. Granica pomiędzy pojęciami eksploracji i eksploatacji, w kontekście optymalizacji, ma bardzo często formę umowną. W potocznym rozumieniu, eksploracja odnosi się przeważnie do zdobywania nowej wiedzy, natomiast eksploatacja ma na celu wykorzystanie wiedzy już posiadanej. Zatem przekładając powyższe pojęcia na język algorytmów ewolucyjnych, zdolności eksploatacyjne mogą być postrzegane jako zdolności algorytmu do przeszukiwania bezpośredniego otoczenia rozwiązania bazowego o niewielkiej średnicy. Z drugiej strony, zdolności eksploracyjne, mogą być utożsamiane ze zdolnościami przeszukiwania obszarów wykraczających poza obszar uważany za bezpośrednie sąsiedztwo rozwiązań już znanych. Największej przeszkody na drodze do zaprojektowania uniwersalnego algorytmu optymalizacji globalnej, należy upatrywać w trudności polegającej na połączeniu dwóch wspomnianych, wzajemnie wykluczających się zadań [46]. Rozkład normalny natomiast, można skonfigurować jedynie w sposób zapewniający bądź to efektywne przeszukiwanie bezpośredniego otoczenia aktualnie najlepszego rozwiązania bądź obszarów znacznie od niego oddalonych. Wszystkie przytoczone wyżej fakty, przyczyniły się poszukiwania bardziej niezawodnych rozkładów eksploracyjnych.

Na przestrzeni ostatnich dziesięciu lat wzrosło zainteresowanie, społeczności skupionej wokół obliczeń ewolucyjnych, rozkładami Cauchy’go [85],[104], [135],[136], [30] a także sporadycznie rozkładem Levy’ego [36, 88]. Oba modele, obok rozkładu normalnego, są jedynymi rozkładami stabilnymi, dla których istnieje analityczna postać funkcji gęstości. Klasa  $\alpha$ -stabilnych rozkładów prawdopodobieństwa cieszy się obecnie dużym, i stale rosnącym zainteresowaniem, szerokiej rzeszy specjalistów. Rozkłady te znalazły szerokie zastosowanie w ekonomii [40], finansach [72], przy modelowaniu ryzyka [45], modelowaniu ruchu w sieci internetowej [64], przetwarzania sygnałów radarowych [5], [16], analizy szeregów czasowych [7] i wielu innych dziedzinach [82].

Według najlepszej wiedzy autora, jedynymi pracami traktującymi klasę rozkładów stabilnych, jako narzędzie optymalizacji globalnej, w sposób kompleksowy, są prace Gutowskiego [46], Lee i Yao [63] oraz Obuchowicza i Prętkiego [88]. Niemniej jednak wszystkie cytowane powyżej pozycje, uwzględniały jedynie wektory stabilne złożone z niezależnych zmiennych losowych. Jakkolwiek taki sposób konstrukcji stabilnego wektora losowego nawiązuje do budowy powszechnie stosowanego normalnego rozkładu prawdopodobieństwa, to należy wziąć pod uwagę fakt, iż w klasie wielowymiarowych rozkładów stabilnych, wektory takie stanowią zaledwie szczególny przypadek.

Wszystkie przytoczone wyżej fakty, stanowią bezpośrednią przesłankę dla po-

stawienia, przyjętego w pracy, celu badawczego. W kontekście stale rosnącego zapotrzebowania na skuteczne metody optymalizacji globalnej, kompleksowa analiza skuteczności nowego narzędzia, jakim są rozkłady stabilne, wydaje się jak najbardziej pożądana.

### 1.3 Zadania szczegółowe

Udowodnienie słuszności przyjętej tezy pracy, zamierzamy osiągnąć poprzez realizację następujących zadań szczegółowych:

- zaprojektowanie operatorów mutacji opartych o wielowymiarowe rozkłady stabilne,
- analiza zbieżności lokalnej oraz globalnej proponowanych algorytmów,
- analiza odporności stabilnego operatora mutacji na niepewność związaną z doбором warunków początkowych algorytmu ewolucyjnego,
- zaprojektowanie oraz weryfikacja efektywności procedury adaptacyjnej dla niesymetrycznych mutacji stabilnych
- weryfikacja skuteczności proponowanych rozwiązań na obliczeniowych przykładach natury inżynierskiej.

### 1.4 Układ rozprawy

Organizacja niniejszej rozprawy podporządkowana została realizacji przyjętych zadań szczegółowych i przedstawia się następująco.

Rozdział 2 zawiera wprowadzenie w tematykę związaną z losowymi rozkładami stabilnymi. Przytoczone właściwości rozkładów stabilnych, zostały potraktowane jako punkt wyjścia do rozważań na temat ich potencjalnych zastosowań w algorytmach optymalizacji stochastycznej. Ponadto, przedstawione zostały powody, na podstawie których można domniemać, iż rozkłady te mogą stanowić niezwykle silną konkurencję, dla istniejących już rozwiązań.

Rozdział 3 stanowi przegląd klasycznych algorytmów oraz powszechnie wykorzystywanych operatorów ewolucyjnych, które stanowią punkt wyjścia do rozważań w dalszej części rozprawy.

Rozdziały 4 oraz 5 stanowią właściwy przedmiot rozprawy z oryginalnymi osiągnięciami autora. Między innymi odniesiono się do tzw. efektu symetrii oraz efektu otoczenia. W szczególowy sposób zaprezentowano przyczyny powstawania obu zjawisk oraz zaproponowano operatory genetyczne, mogące załagodzić ich niekorzystny wpływ na skuteczność optymalizacyjną. Dokonano również analizy zbieżności lokalnej strategii ewolucyjnej  $(1 + 1)ES_{\alpha}$  z izotropową mutacją stabilną. Odniesiono się do kwestii odporności algorytmu ewolucyjnego na niepewność związaną z ustaleniem optymalnej konfiguracji operatora mutacji. Ponadto, przedstawiono procedurę optymalizacyjną wykorzystującą probabilistyczny model stabilny do estymacji rozkładu osobników populacji.

Rozdział 5 stanowi próbę przewyżczenia problemów pojawiających się podczas optymalizacji globalnej o znacznej liczbie zmiennych decyzyjnych. Przedstawiono konstrukcję specjalnej klasy rozkładów kierunkowych, które wraz ze znanymi heurystykami wyznaczania kierunku mutacji, posłużyły do zaprojektowania

efektywnych algorytmów ewolucyjnych.

Pracę zamykają rozdziały obrazujące zastosowania proponowanych rozwiązań do złożonych problemów technicznych. Celem rozdziału 6 było zaprezentowanie skuteczności algorytmów ewolucyjnych w zadaniach projektowania odpornych układów diagnostyki przemysłowej. Przede wszystkim przedstawiono sposób określenia niepewności parametrycznej sztucznych sieci neuronowych. Ponadto wykorzystując technikę planowania eksperymentu oraz algorytmy oparte o stabilną klasę rozkładów prawdopodobieństwa pokazano, w jaki sposób można rozwiązać problem minimalizacji niepewności parametrycznej modeli nieliniowych. W rozdziale tym, proponowane algorytmy posłużyły do projektowania odpornych, nieliniowych obserwatorów stanu. Zamieszczony przykład, wykorzystujący trójfazowy silnik indukcyjny, wykazał skuteczność zaproponowanego algorytmu ewolucyjnego w tym niezwykle skomplikowanym zadaniu optymalizacji globalnej.

Treść rozprawy zwieńczona została krótkim podsumowaniem najważniejszych wyników, oraz przedstawieniem kierunków przyszłych badań.

## Rozdział 2

# Rozkłady stabilne

W rozdziale tym, przedstawione zostały definicje oraz podstawowe właściwości stabilnych rozkładów prawdopodobieństwa. W ramach niniejszej pracy, skupiono się jedynie na aspektach, które w bezpośredni sposób rzutują na zagadnienia związane z analizą działania algorytmów ewolucyjnych. W pierwszej kolejności przytoczono definicję stabilnej zmiennej losowej. Między innymi podano twierdzenia dotyczące istnienia momentów, asymptotycznego zachowania ogonów funkcji gęstości oraz zaprezentowano procedurę generowania liczb pseudo-losowych. W drugiej części rozdziału skupiono się na przedstawieniu najważniejszych zagadnień związanych ze stabilnymi wektorami losowymi. W centrum zainteresowania znalazły się izotropowe, symetryczne oraz niesymetryczne wersje wektorów stabilnych. Wprowadzono pojęcie miary spektralnej oraz przedstawiono ilościowe wskaźniki służące do opisu zależności statystycznych pomiędzy poszczególnymi składowymi wektora losowego.

### 2.1 Stabilna zmienna losowa

Rozkłady stabilne można zdefiniować przynajmniej na cztery równoważne sposoby [107]. Podstawowa definicja, uwzględniająca fakt, że rodzina rozkładów stabilnych, tworzy grupę z operacją sumy poszczególnych zmiennych losowych wyraża się następująco [82]

**Definicja 5** *Zmienna losowa  $X$  posiada rozkład stabilny, jeśli dla dowolnych liczb  $A, B > 0$  istnieją liczby  $C > 0$  oraz  $D \in \mathbb{R}$ , takie że:*

$$A X_1 + B X_2 \stackrel{d}{=} C X + D, \quad (2.1)$$

gdzie  $X_1, X_2$  oznaczają niezależne kopie  $X$ , a symbol  $\stackrel{d}{=}$  oznacza, że zmienne losowe po obu stronach znaku mają identyczny rozkład.

Początki rozkładów stabilnych sięgają lat pięćdziesiątych dziewiętnastego wieku, kiedy to Augustin Cauchy (1853) udowodnił, że klasa funkcji  $f_\alpha$  spełniających warunek

$$\int_{-\infty}^{\infty} \exp(i k x) f_\alpha(x) = \exp(-\sigma^\alpha \|k\|^\alpha) \quad (2.2)$$

zachowuje swoją niezmienną formę względem operacji splotu

$$(Af_\alpha(A)) * (Bf_\alpha(B)) = Cf_\alpha(A), \quad (2.3)$$

dla  $A, B > 0$  oraz  $C$  zależnego od  $A, B$ . Późniejsze prace George'a Polya, Paul Levy'ego oraz Aleksandra Khinchine'a pozwoliły na sformułowanie bardziej precyzyjnej definicji rozkładu stabilnego [107]

**Definicja 6** *Zmienna losowa  $X$  posiada rozkład stabilny, jeśli istnieją parametry:  $\alpha \in (0, 2]$ ,  $\sigma \geq 0$ ,  $\beta \in [-1, 1]$  oraz  $\mu \in \mathbb{R}$  takie, że jej funkcja charakterystyczna posiada następującą postać:*

dla  $\alpha \neq 1$

$$\varphi(k) = \exp\left(-\sigma^\alpha |k|^\alpha \left(1 - i\beta \operatorname{sign}(k) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu k\right) \quad (2.4)$$

dla  $\alpha = 1$

$$\varphi(k) = \exp\left(-\sigma |k| \left(1 + i\beta \operatorname{sign}(k) \frac{2}{\pi} \ln(|k|)\right) + i\mu k\right) \quad (2.5)$$

gdzie  $\alpha$  jest indeksem stabilności,  $\beta$  jest parametrem skośności,  $\sigma$  oznacza parametr skali rozkładu,  $\mu$  jest parametrem lokalizacji oraz

$$\operatorname{sign}(x) \begin{cases} 1 & \text{jeśli } x > 0 \\ 0 & \text{jeśli } x = 0 \\ -1 & \text{jeśli } x < 0 \end{cases} \quad (2.6)$$

Funkcja charakterystyczna, pojawiająca się w definicji (6) może być postrzegana jako wartość oczekiwana zespolonej zmiennej losowej  $\exp(jkX)$  lub transformata Fouriera funkcji gęstości  $X$ :

$$\varphi(k) = E[\exp(jkX)] = \int_{-\infty}^{\infty} \exp(jkx) dF(x) = \int_{-\infty}^{\infty} \exp(jkx) f(x) dx, \quad (2.7)$$

gdzie  $F(x)$  i  $f(x)$ , są odpowiednio dystrybuantą i funkcją gęstości zmiennej losowej  $X$ . Aby podkreślić fakt, że rozkłady stabilne określone są przez cztery parametry rzeczywiste, dla ich oznaczenia zarezerwowano notację  $S_\alpha(\sigma, \beta, \mu)$  [82, 107]. W wielu miejscach niniejszej rozprawy, w centrum zainteresowania będą leżały symetryczne rozkłady stabilne, dla których zachodzi  $X \stackrel{d}{=} -X$ . Rozkłady o zadanej własności otrzymuje się poprzez przyjęcie  $\beta = 0$  oraz  $\mu = 0$ . Dla ich oznaczenia, w literaturze stosuje się osobne oznaczenie  $S_\alpha S(\sigma)$ . Warto zauważyć, że funkcja charakterystyczna symetrycznych, stabilnych zmiennych losowych przyjmuje znacznie prostszą postać

$$\varphi(k) = \exp\left(-\sigma^\alpha |k|^\alpha\right) \quad (2.8)$$

### 2.1.1 Wybrane właściwości

Charakterystyczną cechą rozkładów stabilnych jest zachowanie rozkładu podczas operacji sumowania niezależnych zmiennych losowych. Naturalnie, podczas tej operacji zmianie ulegają parametry rozkładu wynikowego, co znajduje swoje odzwierciedlenie w poniższym twierdzeniu [107]

**Twierdzenie 1** Niech  $X_1$  i  $X_2$  będą niezależnymi zmiennymi losowymi o rozkładach odpowiednio  $X_1 \sim S_\alpha(\sigma_1, \beta_1, \mu_1)$  oraz  $X_2 \sim S_\alpha(\sigma_2, \beta_2, \mu_2)$ . Wówczas suma  $X_1 + X_2 \sim S_\alpha(\sigma, \beta, \mu)$ , gdzie poszczególne parametry rozkładu wynikowego związane są następującymi zależnościami:

$$\sigma = (\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}, \quad \beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \mu = \mu_1 + \mu_2 \quad (2.9)$$

Parametr  $\sigma$  najczęściej określany jest mianem parametru skali. Szczególne znaczenie parametru  $\sigma$  dla klasy rozkładów stabilnych, znajduje potwierdzenie w twierdzeniu

**Twierdzenie 2** Niech  $X \sim S_\alpha(\sigma, \beta, \mu)$  oraz  $a$  oznacza różną od zera liczbę rzeczywistą. Wówczas, prawdziwe są poniższe równości:

- dla  $\alpha \neq 1$

$$aX \sim S_\alpha(|a|\sigma, \text{sign}(a)\beta, a\mu) \quad (2.10)$$

- dla  $\alpha = 1$

$$aX \sim S_1(|a|\sigma, \text{sign}(a)\beta, a\mu - \frac{2}{\pi} a \ln(|a|)\sigma\beta) \quad (2.11)$$

**Uwaga 1** Możliwość skalowania rozkładów stabilnych, wyrażona w twierdzeniu 2, nabiera szczególnego znaczenia w kontekście ich potencjalnych zastosowań w algorytmach ewolucyjnych. Mianowicie, klasyczna metoda adaptacji operatora mutacji tzw. reguła 1/5 sukcesu (ang. one-fifth success rule) [10, 21, 20] polega na utrzymywaniu częstości mutacji skutkujących poprawieniem rozwiązania w okolicach wartości 0.2. Heurystyka, która jest za to odpowiedzialna, bazuje na wymnażaniu mutującej zmiennej losowej przez odpowiednio dobrane stałe. Metodę tą, można zatem bezpośrednio zastosować do mutacji opartej o rozkłady stabilne, gdyż podczas takiej operacji wszystkie statystyki pozycyjne ulegną przeskalowaniu, zachowując swoje pierwotne znaczenie. Możliwe jest zatem zaprojektowanie odpowiednich heurystyk adaptacyjnych dla standardowego rozkładu stabilnego ( $\sigma = 1$ ), a następnie ich bardzo proste skalowanie w zależności od aktualnego parametru skali  $\sigma$ .

**Uwaga 2** Niezmienniczość względem operatora sumy zmiennych losowych, może z kolei okazać się bardzo istotną zaletą przy konstruowaniu adaptacyjnych procedur opartych o analizę skumulowanej ścieżki ewolucyjnej [48], [19] (ang. cumulative path-length control). W podejściu tym, odpowiedni wybór skali rozkładu eksploracyjnego  $\sigma$ , zostaje dokonany na podstawie porównania teoretycznej długości ścieżki, jaką powinien pokonać algorytm ewolucyjny przy braku presji selekcyjnej<sup>1</sup>, z długością obserwowaną. Niewątpliwą zaletą zachowania rozkładu podczas kumulowania poszczególnych przemieszczeń w przestrzeni rozwiązań, jest fakt, iż zawsze znamy teoretyczny rozkład długości ścieżki. Mamy zatem możliwość utworzenia heurystyki sterującej zasięgiem mutacji w oparciu o pewne statystyki charakteryzujące rozkład długości ścieżki, co byłoby nie możliwe w przypadku nieznaności teoretycznego rozkładu długości ścieżki ewolucyjnej.

Rozkłady stabilne, co głównie stoi na drodze do ich jeszcze szerszego spopularyzowania, generalnie nie posiadają analitycznych postaci funkcji gęstości. Jednakże, są one znane dla trzech szczególnych przypadków:

<sup>1</sup>zazwyczaj termin ten odnosi się do sytuacji, w której funkcja dopasowania nie faworyzuje żadnych rozwiązań, co sprowadza się do przypadku błędzenia losowego



- $S_2(\sigma, 0, \mu)$  - rozkład normalny

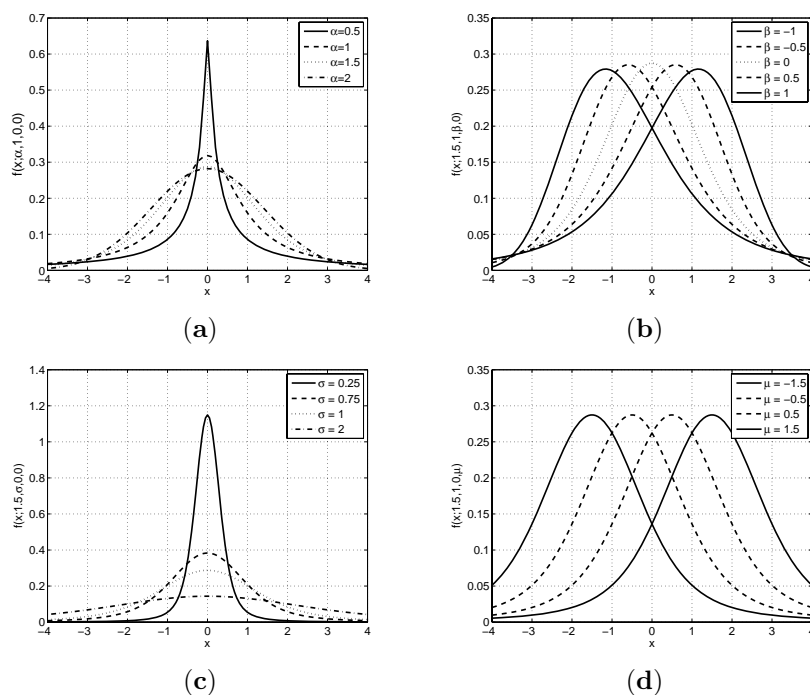
$$f_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty, \quad (2.12)$$

- $S_1(\sigma, 0, \mu)$  - rozkład Cauchy'ego

$$f_C(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x-\mu)^2}, \quad -\infty < x < \infty, \quad (2.13)$$

- $S_{1/2}(\sigma, 1, \mu)$  - rozkład Levy'ego

$$f_L(x) = \sqrt{\frac{\sigma}{2\pi}} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\sigma}{2(x-\mu)}\right), \quad \mu < x < \infty. \quad (2.14)$$



Rysunek 2.1: Wpływ parametrów  $(\alpha, \beta, \sigma, \mu)$  na postać funkcji gęstości rozkładu prawdopodobieństwa

Rozkłady maksymalnie prawostronnie skośne ( $\beta = 1$ ), odgrywają w całej klasie rozkładów stabilnych szczególną rolę. Mianowicie, mogą one być postrzegane jako podstawowe bloki, z których zbudowany jest dowolny rozkład tej klasy, o czym mówi poniższe twierdzenie [107]

**Twierdzenie 3** Niech  $X$  będzie stabilną zmienną losową  $X \sim S_\alpha(\sigma, \beta, \mu)$  dla  $\alpha \in (0, 2)$ . Wówczas, istnieją dwie niezależne zmiennne losowe  $Y_1$  oraz  $Y_2$  o identycznym rozkładzie  $S_\alpha(\sigma, 1, 0)$ , takie że

- $\alpha \neq 1$

$$X \stackrel{d}{=} \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_1 - \left(\frac{1-\beta}{2}\right)^{1/\alpha} Y_2 \quad (2.15)$$

- $\alpha = 1$

$$X \stackrel{d}{=} \left(\frac{1+\beta}{2}\right) Y_1 - \left(\frac{1-\beta}{2}\right) Y_2 + \sigma \left( \frac{1+\beta}{\pi} \ln\left(\frac{1+\beta}{2}\right) - \frac{1-\beta}{\pi} \ln\left(\frac{1-\beta}{2}\right) \right) \quad (2.16)$$

W licznych zastosowaniach praktycznych, gdzie rozkłady prawdopodobieństwa służą do modelowania różnej natury niepewności, podstawowymi wielkościami charakteryzującymi zmienne losowe są ich dwa pierwsze momenty centralne t.j. wartość oczekiwana  $E(X)$  oraz wariancja  $\text{Var}(X) = E(X^2) - (E(X))^2$ . Fakt istnienia tzw. ciężkich ogonów, które są konsekwencją wolnej zbieżności asymptotycznej funkcji gęstości, skutecznie uniemożliwia zastosowanie wspomnianych statystyk, o czym przekonuje następujące twierdzenie [107]

**Twierdzenie 4** Niech  $X \sim S_\alpha(\sigma, \beta, \mu)$  dla  $\alpha \in (0, 2)$ . Wówczas zachodzi:

$$E(|X|^p) = \begin{cases} < \infty & \text{dla każdego } 0 < p < \alpha \\ = \infty & \text{dla każdego } p \geq \alpha \end{cases} \quad (2.17)$$

**Uwaga 3** Z twierdzenia 4 wynika brak wariancji dla wszystkich rozkładów stabilnych z wyjątkiem rozkładu normalnego ( $\alpha = 2$ ). Co więcej, dla  $\alpha \leq 1$  nie istnieje nawet wartość oczekiwana stabilnej zmiennej losowej.

Ilościowe ujęcie zjawiska polegającego na przemieszczaniu masy prawdopodobieństwa z centralnej części rozkładu do jego ogonów, zawarte jest w kolejnym twierdzeniu [107]

**Twierdzenie 5** Niech  $X \sim S_\alpha(\sigma, \beta, \mu)$  dla  $\alpha \in (0, 2)$ . Wówczas

$$\lim_{x \rightarrow \infty} P(X > x) = x^{-\alpha} C_\alpha \frac{1+\beta}{2} \sigma^\alpha, \quad (2.18)$$

gdzie

$$C_\alpha = \left( \int_0^\infty x^{-\alpha} \sin(x) dx \right)^{-1} \quad (2.19)$$

Brak wariancji, a w niektórych przypadkach także wartości oczekiwanej ma swoje daleko idące konsekwencje dla przebiegu procesu przeszukiwania przestrzeni rozwiązań. W celu zobrazowania niektórych z nich posłużmy się następującym scenariuszem. Rozpatrzmy mianowicie jedną z najprostszych strategii ewolucyjnych (1+1)ES, w której zastosujemy stabilną zmienną losową zaburzającą rozwiązania  $\sigma Z \sim S_\alpha S(\sigma)$ . Załóżmy, że inicjując wspomniany algorytm z punktu  $x_0 \in \mathbb{R}$ , w przeciągu  $N$  kolejnych iteracji obserwowano same sukcesy tzn. każda mutacja skutkowałą poprawą rozwiązania. To, z pozoru wyidealizowane, założenie niekoniecznie musi być oderwane od rzeczywistości, gdyż dla wielu funkcji celu oraz niewielkiego  $N$ , sytuacja taka jak najbardziej może mieć miejsce. Wówczas, rozwiązanie po  $N$  iteracjach, w sensie statystycznym, wyrażone jest przez zależność

$$x_N = x_{N-1} + \sigma Z_k = x_0 + \sigma \sum_{i=1}^N Z_k, \quad (2.20)$$

Z zaprezentowanych dotychczas właściwości stabilnych zmiennych losowych wynika, że  $\sum_{k=1}^N Z_k$  posiada rozkład stabilny  $S_\alpha S(N^{1/\alpha})$ . Rozważmy przemieszczenie jakie w trakcie wspomnianych iteracji nastąpiło w przestrzeni rozwiązań:

$$\Delta x = |x_N - x_0| = \sigma N^{1/\alpha} |Z| \quad (2.21)$$

Warto zauważyć, że w przypadku rozkładu normalnego ( $\alpha = 2$ ), oczekiwana wartość przemieszczenia wynosi:

$$E[\Delta x] = \sigma \sqrt{N} \frac{1}{\pi} \quad (2.22)$$

Tym samym, w przypadku mutacji gaussowskiej, otrzymujemy prostą zależność: im więcej generacji  $N$ , tym większe prawdopodobieństwo eksploracji dalej położonych obszarów przestrzeni rozwiązań. Reguła ta nie znajduje potwierdzenia w przypadku stabilnych rozkładów zaburzających dla  $\alpha < 1$ , gdyż mamy

$$E[\Delta x] = \sigma N^{1/\alpha} E[|Z|] = \infty \quad (2.23)$$

Oznacza to, że oczekiwana wartość przemieszczenia, nie zależy od liczb wykonanych kroków, a znacznie oddalone obszary mogą zostać eksplorowane równie dobrze w przeciągu jednej, jak i kilku iteracji. Cecha, ta przejawia się możliwością częstego wystąpienia tzw. makromutacji, czyli przemieszczeń w przestrzeni przeszukiwań, znacznie różniących się od najczęściej obserwowanych wartości. Mechanizm makromutacji, z pewnością umożliwi łatwiejsze omijanie szerokich basenów przyciągania rozwiązań lokalnych, co może mieć bezpośredni wpływ na polepszenie skuteczności optymalizacyjnej algorytmów ewolucyjnych wykorzystujących rozkłady stabilne. W nawiązaniu do powyższych uwag, rodzi się jednocześnie pytanie dotyczące, równie ważnej, optymalizacji lokalnej. Mianowicie, istotnym zagadnieniem jest problem utraty precyzji podczas eksploatacji bezpośredniego otoczenia punktu bazowego  $x_0$ , który jest konsekwencją braku wartości oczekiwanej dla przemieszczenia (2.23). Pierwsza, bardziej intuicyjna, próba odniesienia się do tej kwestii, może polegać na obserwacji wykresów funkcji gęstości (rysunek 2.1 - **a**). W klasie rozkładów stabilnych można bowiem zaobserwować następującą prawidłowość: dla dowolnych indeksów stabilności  $0 < \alpha_1 < \alpha_2 \leq 2$  istnieje liczba  $t > 0$ , taka że spełniona jest nierówność

$$P(|S_{\alpha_1} S(\sigma)| < t) > P(|S_{\alpha_2} S(\sigma)| < t) \quad (2.24)$$

Nierówność (2.24) oznacza istnienie bezpośredniego otoczenia punktu bazowego, w którym wygenerowanie rozwiązania alternatywnego, będzie bardziej prawdopodobne dla rozkładów o mniejszych indeksach stabilności. Oczywiście nie jest sprawą do końca jasną, na ile fakt istnienia takiego otoczenia, w rzeczywistości przekłada się skuteczność optymalizacji lokalnej.

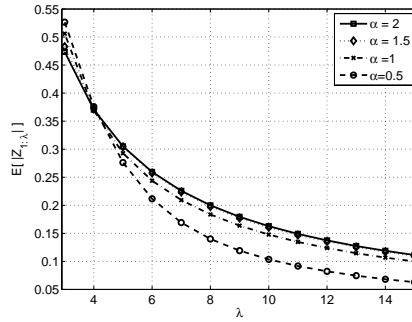
Do problemu eksploatacji przestrzeni rozwiązań za pomocą rozkładów stabilnych, można odnieść się również w bardziej wiarygodny i przekonujący sposób. Mianowicie, warto zauważyć, że większość algorytmów ewolucyjnych, tworząc populację rozwiązań alternatywnych, dokonuje więcej niż jednokrotnego zaburzenia rozwiązań rodzicielskich. Najbardziej oczywistym przykładem takiej sytuacji jest strategia ewolucyjna  $(1, \lambda)$ ES, czy algorytm ESSS, dla którego operator miękkiej selekcji powoduje powielanie najlepszych rozwiązań. Oznaczając

ciąg realizacji zmiennych losowych  $Z_1, Z_2, \dots, Z_\lambda$  gdzie  $Z_i \sim S_\alpha S(\sigma)$ , stosowanych do zaburzenia rozwiązania  $x_0$ , spróbujmy odnieść się do następującego pytania: czy braku momentów dla przemieszczenia dokonywanego na podstawie jednokrotnego zaburzenia oznacza jednocześnie, iż wartość ta nie będzie istnieć także w przypadku zaburzenia wielokrotnego? Otóż odpowiedź na powyższe pytanie jest negatywna, a jej uzasadnienie, bazujące na pojęciu statystyki rankingowej (ang. *order statistics*), podaje następujące twierdzenie (dowód zamieszczony został w dodatku B)

**Twierdzenie 6** Niech  $Z \sim |S_\alpha S(\sigma)|$  będzie zmienną losową. Wówczas  $k$ -ty moment zmiennej losowej  $Z_{i:\lambda}$  istnieje, wtedy i tylko wtedy, gdy zachodzi następująca nierówność:

$$k - \alpha(\lambda - i + 1) < 0 \quad (2.25)$$

Powyższe twierdzenie mówi dokładnie tyle, iż uporządkowanie zaburzających zmiennych losowych według schematu:  $Z_{1:\lambda} < Z_{2:\lambda} < \dots < Z_{\lambda:\lambda}$  sprawia, że nawet dla indeksu stabilności  $\alpha < 1$  występują wartości oczekiwane przemieszczenia (2.23). Co więcej, na wykresie przedstawiono, wyznaczone numerycznie, wartości oczekiwane zmiennej losowej  $Z_{\min} = Z_{1:\lambda}$  dla różnych indeksów stabilności.



Rysunek 2.2: Wartości oczekiwane zmiennej losowej  $Z_{1:\lambda}$  w zależności od liczby zaburzeń  $\lambda$

**Uwaga 4** Z twierdzenia 6 wynika, że wartość oczekiwana zmiennej losowej, która odpowiada najbliższemu zaburzeniu spośród  $\lambda$  prób, w przypadku rozkładów stabilnych, istnieje, jeśli  $\lambda > \frac{1}{\alpha}$ . Znacznie bardziej zadziwiające jednak jest to, że zaczynając od  $\lambda > 3$  wartości oczekiwane dla rozkładów o cięższych ogonach przyjmują mniejsze wartości. Stwarza to sytuację, w której dokonując  $\lambda > 3$  prób zaburzeń punktu bazowego  $x_0$ , z rozkładami o mniejszym indeksie stabilności  $\alpha$ , mamy gwarancję bardziej precyzyjnej eksploatacji otoczenia punktu  $x_0$ . Pamiętając jednocześnie o tym, iż mniejsze  $\alpha$  umożliwia otrzymywanie częstszych makromutacji, odpowiedzialnych za eksplorację przestrzeni rozwiązań, istnieje nadzieja, że klasa rozkładów stabilnych może stanowić panaceum na odwieczny problem połączenia dwóch wykluczających się zadań: eksploracji oraz eksploatacji.

Algorytmy ewolucyjne, z racji bardzo prostej heurystyki ukierunkowującej proces przeszukiwania przestrzeni rozwiązań, zazwyczaj wymagają dużej liczby

epok, aby osiągnąć zadowalający wynik. Często przez to, są postrzegane jako algorytmy o dużej złożoności obliczeniowej, które w dobie obecnych komputerów, nie znajdują zastosowania w optymalizacji odbywającej się w czasie rzeczywistym. W świetle powyższych uwag, szczególnego znaczenia nabiera kwestia złożoności obliczeniowej procedury, służącej do generowania liczb pseudolosowych o rozkładzie stabilnym. Okazuje się, że jeśli rozpatrzmy niezależne zmienne losowe:  $W \sim \text{EXP}(1)$ <sup>2</sup> oraz  $U \sim U(\frac{-\pi}{2}, \frac{\pi}{2})$ <sup>3</sup> wykorzystując twierdzenie 7 można w łatwy sposób zbudować generator odpowiednich liczb pseudolosowych.

**Twierdzenie 7** Niech  $W \sim \text{EXP}(1)$  i  $U \sim U(\frac{-\pi}{2}, \frac{\pi}{2})$  będą dwoma niezależnymi zmiennymi losowymi. Wówczas zmienna losowa  $X$  postaci

$$X \stackrel{d}{=} \frac{\sin(\alpha U)}{(\sin(U))^{1/\alpha}} \left( \frac{\cos((1-\alpha)U)}{W} \right)^{\frac{1-\alpha}{\alpha}} \quad (2.26)$$

posiada standardowy rozkład  $\alpha$ -stabilny, t.j.  $X \sim S_\alpha(1, 0, 0)$ .

Kolejne twierdzenie zostanie przytoczone ze względu na swoją kluczową rolę jaką ogrywa przy konstrukcji rozkładów stabilnych o sferycznej symetrii (rozdział 2.2.1)

**Twierdzenie 8** Niech  $X_1 \sim S_{\alpha'}(\sigma, 0, 0)$  oraz  $X_2 \sim S_{\alpha/\alpha'}\left(\cos(\frac{\pi\alpha'}{2\alpha})^{\frac{\alpha'}{\alpha}}, 1, 0\right)$  dla  $0 < \alpha < \alpha' \leq 2$ . Wówczas zachodzi:

$$X_1 X_2^{1/\alpha'} \sim S_\alpha(\sigma, 0, 0) \quad (2.27)$$

**Uwaga 5** Przyjmując  $\alpha' = 2$  w twierdzeniu 8, można zauważyć, że dowolny rozkład stabilny  $S_\alpha(\sigma, 0, 0)$  może zostać otrzymany z rozkładu normalnego, t.j.  $N(0, A)$ , tzn. gdy odchylenie standardowe jest stabilną zmienną losową  $A \sim S_{\alpha/2}\left(\cos(\frac{\pi\alpha}{4})^{\frac{2}{\alpha}}, 1, 0\right)$ .

## 2.2 Wielowymiarowy rozkład stabilny

W kontekście zastosowań rozkładów stabilnych w stochastycznych algorytmach optymalizacji globalnej największe znaczenie odgrywają rozkłady wielowymiarowe. Ogólnie, rozkłady te można zdefiniować w sposób podobny, jak to miało miejsce w przypadku ich jednowymiarowych odpowiedników.

**Definicja 7** Wektor losowy  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  jest stabilnym wektorem w  $\mathbb{R}^n$  jeśli dla dowolnych liczb  $A, B > 0$  istnieje liczba  $C > 0$  oraz wektor  $\mathbf{D} \in \mathbb{R}^n$ , taki że zachodzi równość

$$A \mathbf{X}^{(1)} + B \mathbf{X}^{(2)} \stackrel{d}{=} C \mathbf{X} + \mathbf{D}, \quad (2.28)$$

gdzie  $\mathbf{X}^{(1)}$  oraz  $\mathbf{X}^{(2)}$  są niezależnymi kopiami wektora losowego  $\mathbf{X}$ .

Ogólną definicję 7, która uwypukla najbardziej charakterystyczną cechę stabilnych wektorów losowych, można sprecyzować narzucając warunki konieczne i wystarczające na postać funkcji charakterystycznej

<sup>2</sup>EXP( $\lambda$ ) oznacza rozkład wykładniczy z wartością oczekiwaną równą 1.

<sup>3</sup>U( $a, b$ ) oznacza rozkład równomierny na przedziale ( $a, b$ )

**Definicja 8** Wektor losowy  $\mathbf{X}$  posiada wielowymiarowy rozkład stabilny, jeśli jego funkcję charakterystyczną  $\varphi(\mathbf{k}) = \mathbb{E}[\exp(-i\mathbf{k}^T \mathbf{X})]$  można przedstawić w następującej postaci

dla  $\alpha \neq 1$

$$\varphi(\mathbf{k}) = \exp\left(-\int_{S^{(d)}} |\mathbf{k}^T \mathbf{s}|^\alpha (1 - i \operatorname{sgn}(\mathbf{k}^T \mathbf{s}) \tan(\frac{\pi\alpha}{2})) \Gamma(ds) + i\mathbf{k}^T \boldsymbol{\mu}_0\right) \quad (2.29)$$

dla  $\alpha = 1$

$$\varphi(\mathbf{k}) = \exp\left(-\int_{S^{(n)}} |\mathbf{k}^T \mathbf{s}| (1 - i \frac{2}{\pi} \operatorname{sgn}(\mathbf{k}^T \mathbf{s}) \ln|\mathbf{k}^T \mathbf{s}|) \Gamma(ds) + i\mathbf{k}^T \boldsymbol{\mu}_0\right) \quad (2.30)$$

gdzie  $\Gamma(\cdot)$  jest miarą spektralną, a  $\boldsymbol{\mu}$  oznacza wektor położenia.

### Wybrane właściwości oraz pojęcia

Jednym z niezbędnych warunków na to, aby wektor losowy  $\mathbf{X}$  posiadał rozkład tej samej klasy co zmienna losowa  $X$ , jest istnienie funkcji  $c(\mathbf{b})$ , takiej, aby dla każdego  $\mathbf{b} \in \mathbb{R}^n$  zachodziła równość:  $\mathbf{b}^T \mathbf{X} = c(\mathbf{b}) X$ . Oznacza to, iż w przypadku wektora stabilnego  $\mathbf{X}$ , każda kombinacja liniowa jego elementów, musi posiadać jednowymiarowy rozkład stabilny. Poniższe twierdzenie pokazuje, że związek ten rzeczywiście zachodzi dla wektorów stabilnych.

**Twierdzenie 9** Każda kombinacja liniowa  $Y = \mathbf{b}^T \mathbf{X} = \sum_{i=1}^n b_i X_i$  poszczególnych składników stabilnego wektora losowego  $\mathbf{X}$  jest stabilną zmienną losową, t.j.  $Y \sim S_\alpha(\sigma_{\mathbf{b}}, \beta_{\mathbf{b}}, \mu_{\mathbf{b}})$ , gdzie poszczególne parametry przyjmują następujące wartości:

$$\sigma_{\mathbf{b}} = \left( \int_{S^{(n)}} |\mathbf{b}^T \mathbf{s}|^\alpha \Gamma(ds) \right)^{1/\alpha} \quad (2.31)$$

$$\beta_{\mathbf{b}} = \sigma_{\mathbf{b}}^{-\alpha} \int_{S^{(n)}} \operatorname{sign}(\mathbf{b}^T \mathbf{s}) |\mathbf{b}^T \mathbf{s}|^\alpha \Gamma(ds), \quad (2.32)$$

$$\mu_{\mathbf{b}} = \begin{cases} \mathbf{b}^T \boldsymbol{\mu} & \alpha \neq 1 \\ \mathbf{b}^T \boldsymbol{\mu} - \frac{2}{\pi} \int_{S^{(n)}} \mathbf{b}^T \mathbf{s} \ln(|\mathbf{b}^T \mathbf{s}|) \Gamma(ds) & \alpha = 1 \end{cases} \quad (2.33)$$

**Uwaga 6** Stosując zależności występujące w twierdzeniu 9, można w bardzo łatwy sposób wyznaczyć dowolne rozkłady brzegowe stabilnego wektora losowego. W odniesieniu do stochastycznych metod optymalizacji, zależności te, mogą posłużyć do zdefiniowania rozkładu prawdopodobieństwa dla poszczególnych podprzestrzeni zbioru rozwiązań dopuszczalnych. Stabilna postać rozkładów brzegowych, zapewnia również, że każda podprzestrzeń jednowymiarowa również będzie próbkowana z rozkładem ciężkoogonowym, co w znacznej mierze może ułatwić interpretację uzyskiwanych wyników.

Bardzo ważnym wynikiem, ułatwiającym zrozumienie związku pomiędzy miarą spektralną stabilnego wektora  $\mathbf{X}$  a jego rozkładem, jest twierdzenie [11]

**Twierdzenie 10**

$$\lim_{r \rightarrow \infty} \frac{P(\mathbf{X} \in \operatorname{Cone}(A) \mid \|\mathbf{X}\| > r)}{P(\|\mathbf{X}\| > r)} = \frac{\Gamma(A)}{\Gamma(S^n)}, \quad (2.34)$$

gdzie  $\operatorname{Cone}(A) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| > 0, \frac{\mathbf{x}}{\|\mathbf{x}\|} \in A \subset S^n\}$

**Uwaga 7** W kontekście zastosowań wielowymiarowego rozkładu stabilnego do zaburzeń rozwiązań w stochastycznych algorytmach optymalizacji globalnej, twierdzenie 10 ma szczególnie ważne znaczenie. Otóż znając rozkład miary spektralnej na sferze jednostkowej, wiemy dokładnie w jakich kierunkach najczęściej będą zachodziły makromutacje.

Najbardziej intuicyjny wgląd w charakter wielowymiarowych rozkładów prawdopodobieństwa, dostarcza postać ich funkcji gęstości. Niestety w przypadku wektorów stabilnych, gęstości te dane są jedynie w postaci całkowej [3]

dla  $\alpha \neq 1$

$$p(\mathbf{x}) = \int_{S_d} g_{\alpha,n} \left( \frac{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{s}}{\sigma(\mathbf{s})}, \beta(\mathbf{s}) \right) \sigma^{-n}(\mathbf{s}) d\mathbf{s} \quad (2.35)$$

gdzie

$$g_{\alpha,n}(v, \beta) = \frac{1}{(2\pi)^n} \int_0^\infty \cos(vu - \beta \tan\left(\frac{\pi\alpha}{2}\right)u^\alpha) u^{n-1} \exp(-u^\alpha) du \quad (2.36)$$

dla  $\alpha = 1$

$$p(\mathbf{x}) = \int_{S_d} g_{1,n} \left( \frac{(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{s}}{\sigma(\mathbf{s})}, \beta(\mathbf{s}) \right) \sigma^{-n}(\mathbf{s}) d\mathbf{s} \quad (2.37)$$

gdzie

$$g_{1,n}(v, \beta) = \frac{1}{(2\pi)^n} \int_0^\infty \cos\left(vu - \frac{2}{\pi} u \beta \ln(u)\right) u^{n-1} \exp(-u) du \quad (2.38)$$

Bodajże najprostszym wektorem stabilnym  $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$  jest wersja zawierająca wzajemnie niezależne, symetryczne elementy stabilne, tj.  $X_1, X_2, \dots, X_n \sim S_\alpha S(\sigma)$ . Wektory takie, wielokrotnie wykorzystywane w obliczeniach ewolucyjnych [10, 23, 85, 88], mogą być postrzegane jako uogólnienie wektora normalnego  $\mathcal{N}(\mathbf{0}, \sigma \mathbf{I})$ . Wykorzystując zależności (2.35)-(2.37), na rysunku 2.3, zostały zaprezentowane przykładowe funkcje gęstości powyżej rozważanych wektorów stabilnych.

**Uwaga 8** W miarę obniżania wartości indeksu stabilności  $\alpha$  obserwujemy następującą prawidłowość. Masa prawdopodobieństwa w coraz większym stopniu zostaje skupiona wokół osi układu współrzędnych. Powoduje to stopniowe odejście od symetrii sferycznej. Stosując rozważane wektory w algorytmach optymalizacji stochastycznej, należy zatem pamiętać, iż decydując się na mniejszy indeks stabilności, sprawiamy, że kierunki wyznaczone przez osie układu współrzędnych, będą próbkowane intensywniej niż inne kierunki przestrzeni rozwiązań.

Efekt braku sferycznej symetrii dla wektorów stabilnych ma swoje poważne konsekwencje dla sposobu przeszukiwania przestrzeni rozwiązań. Na rysunku 2.4 zaznaczone zostały trajektorie dwuwymiarowej cząsteczki zaburzanej przez rozważany stabilny wektor losowy. Zachowanie bardzo zbliżone do tego przedstawionego na rysunku 2.4 można bardzo często zaobserwować w przypadku algorytmów ewolucyjnych przy braku wyraźnej presji selekcyjnej.

**Obserwacja 1** *Błądzenie losowe cząsteczki zaburzanej przez wektor stabilny, należy do klasy samopodobnych procesów stochastycznych. Oznacza to niezmienniczość rozkładu względem operacji skalowania przestrzeni, w którym porusza się cząsteczka. Efekt ten, bardzo dobrze widoczny jest na rysunku 2.4, gdzie powiększenia wybranych fragmentów trajektorii, charakteryzują się podobnymi cechami jak obraz oryginalny. Wspomniany wcześniej efekt gromadzenia masy prawdopodobieństwa na kierunkach wyznaczanych przez osie układu współrzędnych, można zweryfikować obserwując kierunki, w których zachodzą makromutacje dla  $\alpha < 2$ .*

Dla powszechnie stosowanego, w algorytmach ewolucyjnych wielowymiarowego rozkładu normalnego, pojęcie macierzy kowariancji pozwala jednoznacznie zdefiniować wektor losowy. Kowariancja zmiennych losowych o rozkładzie normalnym posiada prostą, intuicyjną interpretację i przez to, znajduje również szerokie zastosowanie przy projektowaniu procedur adaptacyjnych dla gaussowskiego rozkładu eksploracyjnego. Niestety pojęcie to nie jest zdefiniowane dla wektorów stabilnych. Niemniej jednak istnieją miary, które potrafią oceniać wzajemną zależność pomiędzy poszczególnymi elementami stabilnego wektora losowego. Pierwszym, z powszechnie używanych pojęć, jest tzw. kowariacja (ang. *covariation*) zdefiniowana następująco

**Definicja 9** *Niech  $\mathbf{X} = [X_1, X_2]^T$  oznacza symetryczny wektor stabilny dla  $\alpha > 1$ . Wówczas kowariacja zmiennych  $X_1, X_2$  zdefiniowana jest jako liczba rzeczywista:*

$$[X_1, X_2]_\alpha = \int_{S^{(2)}} s_1 s_2 \langle \alpha - 1 \rangle \Gamma(d\mathbf{s}), \quad (2.39)$$

gdzie  $\Gamma(\cdot)$  jest miarą spektralną wektora  $\mathbf{X}$ , oraz  $a \langle p \rangle = |a|^p \text{sign}(a)$ .

Pojęcie kowariacji, w przypadku  $\alpha = 2$ , odpowiada kowariancji zmiennych losowych t.j.  $[X_1, X_2]_\alpha = \frac{1}{2} \text{cov}(X_1, X_2)$ . Dla stabilnych wektorów symetrycznych  $\mathbf{X} = [X_1, X_2]^T$  autokowariacja równa jest  $[X_1, X_1]_\alpha = \sigma_{X_1}^\alpha$ , a jeśli tylko  $X_1, X_2$  są statystycznie niezależne zachodzi  $[X_1, X_1]_\alpha = 0$ . Wspomniane właściwości wyraźnie wskazują na analogię do rozkładów normalnych. Niestety, w odróżnieniu od kowariancji, miara (2.39) nie zawsze jest symetryczna oraz liniowa. Co więcej, kowariacja, określona jest wyłącznie dla symetrycznych wektorów stabilnych z indeksem  $\alpha > 1$ , co przyczynia się do jej ograniczonego stosowania. Innym pojęciem powszechnie stosowanym do oceny zależności pomiędzy stabilnymi zmiennymi losowymi, jest tzw. kodyferencja (ang. *codifference*).

**Definicja 10** *Niech  $\mathbf{X} = [X_1, X_2]^T$  oznacza symetryczny wektor stabilny dla  $0 < \alpha \leq 2$ . Wówczas kodyferencja zmiennych  $X_1, X_2$  zdefiniowana jest jako liczba rzeczywista:*

$$\tau_{X_1, X_2} = \|X_1\|_\alpha^\alpha + \|X_2\|_\alpha^\alpha - \|X_1 - X_2\|_\alpha^\alpha \quad (2.40)$$

gdzie  $\|X\|_\alpha = ([X_1, X_2]_\alpha)^{1/\alpha}$  jest normą indukowaną przez kowariację.

W odróżnieniu od (2.39) kodyferencja jest zawsze symetryczna, określona dla całej dziedziny indeksu stabilności. W przypadku  $\alpha = 2$  pojęcie to sprowadza się do kowariancji t.j.:  $\tau_{X_1, X_2} = \text{cov}(X_1, X_2)$ . Co więcej, można pokazać [107] iż macierz  $\{\tau_{X_i, X_j}, i, j = 1, \dots, n\}$  symetrycznego wektora stabilnego  $\mathbf{X} = [X_1, \dots, X_n]^T$  jest dodatnio określona. Ponadto, dla  $0 < \alpha < 1$  równość



$\tau_{X_1, X_2} = 0$  oznacza niezależność zmiennych losowych. Warty podkreślenia jest również zależność, która wskazuje na podobieństwo do pojęcia kowariancji rozkładu normalnego.

**Właściwość 1** *Rozpatrzmy dwa symetryczne, stabilne wektory losowe:  $\mathbf{X}_1 = [X_{1,1}, X_{1,2}]^T$  oraz  $\mathbf{X}_2 = [X_{2,1}, X_{2,2}]^T$ , dla których rozkłady brzegowe odznaczają się identycznymi parametrami skali, t.j:  $\|X_{1,1}\|_\alpha = \|X_{1,2}\|_\alpha = \|X_{2,1}\|_\alpha = \|X_{2,2}\|_\alpha$ . Wówczas, dla każdego  $c > 0$  prawdziwa jest zależność:*

$$\tau_{X_{1,1}, X_{1,2}} \leq \tau_{X_{2,1}, X_{2,2}} \Rightarrow P(|X_{1,1} - X_{2,1}| > c) \geq P(|X_{1,2} - X_{2,2}| > c) \quad (2.41)$$

Zależność (2.41) mówi o tym, że im większa wartość współczynnika kodyferencji, tym silniejszy związek pomiędzy zmiennymi losowymi, i tym samym mniejsze prawdopodobieństwo, że będą się one od siebie znacząco różnić.

### 2.2.1 Izotropowy rozkład $\alpha$ -stabilny

Rozkłady izotropowe odgrywają szczególną rolę w algorytmach optymalizacji globalnej. W przypadku nieograniczonej przestrzeni przeszukiwań  $\mathbb{R}^n$ , można doszukać się pewnej analogii pomiędzy nimi, a rozkładem równomiernie określonym na ograniczonym zbiorze rozwiązań dopuszczalnych. Ten ostatni nie wyróżnia żadnego spośród podzbiorów przestrzeni przeszukiwań, natomiast rozkłady o sferycznej symetrii, nie faworyzują żadnego z kierunków przeszukiwań. Z tego względu, ich stosowanie, jest szczególnie uzasadnione na początku procesu optymalizacji, gdzie nie posiadamy żadnej wiedzy o optymalizowanym problemie. Pojęcie symetrii pewnej funkcji  $f(\cdot)$ , nieodłącznie łączy się z zagadnieniem jej niezmienniczości względem określonej grupy przekształceń.

**Definicja 11** *Przekształcenie  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , jest niezmiennicze względem grupy przekształceń  $\mathcal{G}$ , jeśli zachodzi:*

$$f(\mathbf{x}) = f(g(\mathbf{x})), \quad (2.42)$$

dla każdego  $\mathbf{x} \in \mathbb{R}^n$  i  $g \in \mathcal{G}$ .

Do najpopularniejszych typów symetrii można zaliczyć

- *symetrię przez odwrotność*, dla której grupa  $\mathcal{G}$  składa się z przekształceń postaci:  $g(\mathbf{x}) = g(-\mathbf{x})$ ,
- *symetrię permutacyjną*, dla której grupa  $\mathcal{G}$  składa się z przekształceń postaci:  $g(\mathbf{x}) = \mathbf{H}\mathbf{x}$ , gdzie  $\mathbf{H} \in \mathbb{R}^{n \times n}$  jest macierzą permutacji, tj. zawiera dokładnie jedną jedynkę w każdym wierszu i w każdej kolumnie.
- *symetrię sferyczną*, dla której grupa  $\mathcal{G}$  składa się z przekształceń postaci:  $g(\mathbf{x}) = \mathbf{H}\mathbf{x}$ , gdzie  $\mathbf{H} \in \mathbb{R}^{n \times n}$  jest macierzą ortonormalną, tj. spełniającą warunek  $\mathbf{H}\mathbf{H}^T = \mathbf{I}_n$ .
- *symetrię eliptyczną*, dla której grupa  $\mathcal{G}$  składa się z przekształceń postaci:  $g(\mathbf{x}) = \mathbf{H}\mathbf{x}$ , gdzie  $\mathbf{H} \in \mathbb{R}^{n \times n}$  jest macierzą spełniającą warunek  $\mathbf{H}^T \mathbf{H} \succ \mathbf{0}$ <sup>4</sup>.

<sup>4</sup> $\mathbf{H} \succ \mathbf{0}$  oznacza, że  $\mathbf{H}$  jest macierzą dodatnio określoną

W rozdziale tym naszą uwagę skupimy na mutacji, która bazuje na rozkładach stabilnych o sferycznej symetrii, tzn. takich których rozkład prawdopodobieństwa nie ulega zmianie pod wpływem obrotów układu odniesienia. Generalnie, istnieje co najmniej kilka sposobów na zdefiniowanie izotropowego rozkładu losowego. Z punktu widzenia problemów poruszanych w niniejszej rozprawie najistotniejsza wydaje się definicja [38]

**Definicja 12** Niech  $\mathbf{X}^s = (X_1, \dots, X_n)^T$  oznacza  $n$ -wymiarowy wektor losowy,  $\mathcal{O}(n)$  będzie klasą ortogonalnych macierzy rozmiaru  $n \times n$ , oraz niech  $\mathbf{u}^{(n)}$  oznacza wektor losowy o rozkładzie równomiernym na powierzchni  $n$ -wymiarowej kuli jednostkowej. Wówczas, wektor  $\mathbf{X}^s$  posiada rozkład o sferycznej symetrii wtedy i tylko wtedy, gdy jeden z poniższych warunków jest spełniony:

- 1)  $\mathbf{X}^s \stackrel{d}{=} \mathbf{O} \mathbf{X}^s$  dla każdej macierzy  $\mathbf{O} \in \mathcal{O}(n)$ ,
- 2) funkcja charakterystyczna  $\mathbf{X}^s$  posiada postać  $\phi(\boldsymbol{\theta}^T \boldsymbol{\theta})$ , gdzie  $\phi(\cdot)$  jest pewną funkcją skalarną, zwaną generatorem charakterystycznym,
- 3)  $\mathbf{X}^s$  może być przedstawiony w postaci stochastycznej dekompozycji  $\mathbf{X}^s \stackrel{d}{=} r \mathbf{u}^{(n)}$ , dla pewnej zmiennej losowej  $r \geq 0$  niezależnej od  $\mathbf{u}^{(n)}$ ,
- 4) dla każdego  $\mathbf{a} \in \mathbb{R}^n$  zachodzi  $\mathbf{a}^T \mathbf{X}^s \stackrel{d}{=} \|\mathbf{a}\| x_1$ .

W gruncie rzeczy uogólnienie dowolnego rozkładu jednowymiarowego do jego wielowymiarowego izotropowego odpowiednika, nie jest zadaniem trywialnym, chociażby z tego powodu, iż definicja nie dostarcza żadnych wskazówek, jak należy tego dokonać. Na szczęście, w przypadku rozkładów stabilnych istnieje specjalna podklasa rozkładów stabilnych tzw. warunkowo-gausowskie rozkłady stabilne [107] (ang. *sub-Gaussian distribution*), które umożliwiają zdefiniowanie wektorów o rozkładzie izotropowym [107]:

**Twierdzenie 11** Niech  $\mathbf{X}_\alpha^s = A^{1/2} \mathbf{G}$ , gdzie  $A \sim S_{\alpha/2}(\cos(\frac{\pi\alpha}{4})^{\frac{2}{\alpha}}, \sigma^2, 0)$ , dla  $\alpha \in (0, 2)$ ,  $\mathbf{G} \sim \mathcal{N}(0, \sigma^2 I_n)$  oraz  $A$  i  $\mathbf{G}$  są stochastycznie niezależne. Wówczas:

a) Funkcja charakterystyczna  $\mathbf{X}_\alpha^s$  jest postaci:

$$\varphi(\boldsymbol{\theta}) = E[\exp(-j\boldsymbol{\theta}^T \mathbf{Z})] = \exp(-2^{-\alpha/2} \sigma^\alpha \|\boldsymbol{\theta}\|^\alpha) \quad (2.43)$$

b) Wektor  $\mathbf{X}_\alpha^s$  posiada równomiernie rozłożoną miarę spektralną na powierzchni kuli  $n$ -wymiarowej.

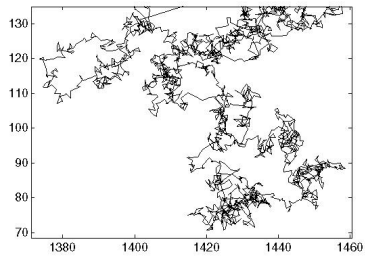
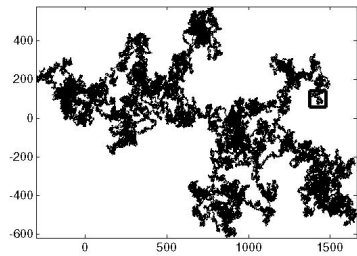
Spełnienie, któregokolwiek z warunków wyrażonych w twierdzeniu 11 przez wektor losowy, oznacza jednocześnie, że wektor ten charakteryzuje się rozkładem izotropowym. Aby się o tym przekonać, wystarczy zauważyć, że funkcja charakterystyczna (2.43) zależy jedynie od wartości  $\|\boldsymbol{\theta}\|$ , co w nawiązaniu do punktu drugiego twierdzenia 12, natychmiast potwierdza istnienie sferycznej symetrii.

**Uwaga 9** Twierdzenie 11 pozwala wykorzystać dekompozycję stochastyczną  $\mathbf{X}_\alpha^s = A^{1/2} \mathbf{G}$  do zaprojektowania generatora wektorów losowych o izotropowym rozkładzie stabilnym. Wykorzystując twierdzenie 7 oraz jedną z powszechnie znanych procedur otrzymywania liczb pseudolosowych o standaryzowanym rozkładzie normalnym, możemy z łatwością otrzymać stabilny wektor izotropowy  $\mathbf{X}_\alpha^s$ . Co ważne, w stosunku do procedur formujących wektory normalne, całkowity koszt otrzymania wektora izotropowego jest powiększony o nakłady poniesione na wygenerowanie tylko jednej pseudolosowej liczby stabilnej.

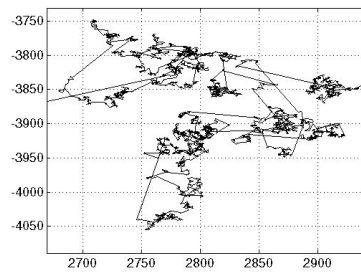
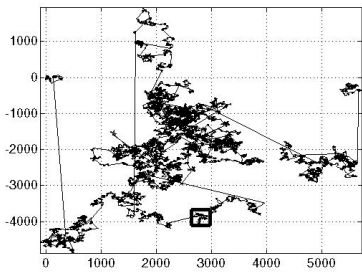
### Wybrane właściwości oraz pojęcia

Wiadomo, że każda kombinacja liniowa wektora stabilnego musi być stabilną zmienną losową (twierdzenie 9). W przypadku wektora izotropowego  $\mathbf{X}^s$  zmienna ta jest symetryczna  $S_\alpha S(\sigma_b)$ , gdzie  $\sigma_b$  wyznaczone jest z zależności (2.31). W przełożeniu na operator mutacji, oparty o rozważaną klasę rozkładów, oznacza to, że każda z jednowymiarowych podprzestrzeni zbioru rozwiązań, będzie przeszukiwana zgodnie z rozkładem  $S_\alpha S(\sigma)$ . Stwierdzenie to oczywiście ma charakter statystyczny, gdyż poszczególne współrzędne muszą wykazywać wzajemną zależność.

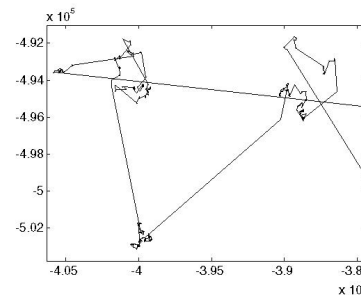
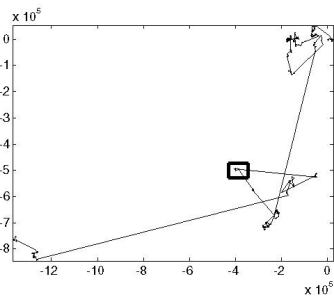
Sferyczna symetria rozkładu eksploracyjnego oznacza, sprawiedliwe traktowanie każdego kierunku przeszukiwań. Równomierny rozkład miary spektralnej, o którym mowa w twierdzeniu 11, w połączeniu z twierdzeniem 10, powoduje również, że prawdopodobieństwo wystąpienia makromutacji nie zależy od kierunku w przestrzeni rozwiązań. Izotropowy wektor stabilny, ustala zupełnie inne reguły przeglądu zbioru rozwiązań, niż to miało miejsce dla wektora z niezależnymi komponentami stabilnymi. Dla zobrazowania różnic pomiędzy tym dwoma przypadkami, na rysunku 2.5 przedstawiono przykładowe trajektorie błędzenia losowego dla zaburzeń realizowanych w postaci stabilnego wektora izotropowego  $\mathbf{X}_\alpha^s$ .



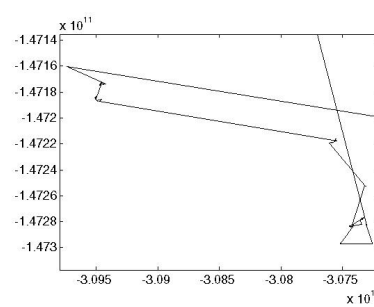
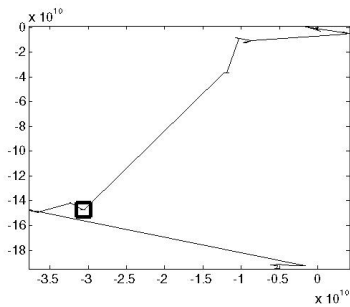
(a)



(b)



(c)



(d)

Rysunek 2.5: Symulacja błędzenia losowego na podstawie 1000000 modyfikacji cząsteczki zgodnie z izotropowymi rozkładami stabilnymi. Zakreślone trajektorie po prawej stronie odpowiadają różnym wykładnikom stabilności: (a) -  $\alpha = 2$ , (b) -  $\alpha = 1.5$ , (c) -  $\alpha = 1$ , (d) -  $\alpha = 0.5$ . Wykresy umieszczone z lewej strony przedstawiają powiększone zaznaczone obszary

**Obserwacja 2** *Przykład błędzenia losowego ujawnia fraktalną naturę stabilnych rozkładów izotropowych. Można zaobserwować, iż powiększone wybrane fragmenty trajektorii, do złudzenia przypominają strukturę trajektorii pierwotnej. Godnym podkreślenia, jest również fakt występowania makromutacji w dowolnych kierunkach przestrzeni przeszukiwań.*

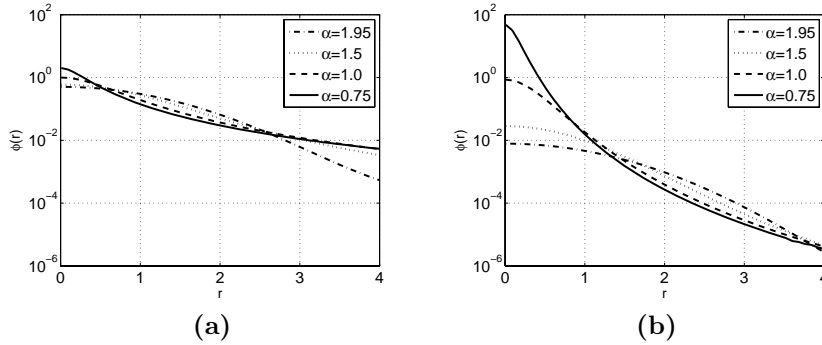
W przypadku algorytmów stochastycznych, które korzystają z różnych rozkładów eksploracyjnych, znajomość funkcji gęstości, daje przybliżone pojęcie o sposobie przeszukiwania przestrzeni rozwiązań. Mając na względzie fakt, iż analityczna postać funkcji gęstości stabilnych zmiennych losowych (rozdział 2.1) nie jest znana, naiwnością byłoby oczekiwanie, że postać taka będzie znana w przypadku ich izotropowych odpowiedników. Jednakże pewną, ogólną wiedzą, na temat postaci funkcji gęstości rozkładów izotropowych, jednak dysponujemy. Mianowicie, wiadomo, że każda funkcja o sferycznej symetrii, w tym funkcja gęstości, musi być funkcją iloczynu skalarnego  $\mathbf{x}^T \mathbf{x}$ , który jest funkcją maksymalnie niezmienniczą [123] względem przekształceń zachowujących symetrię. Ponadto, wykorzystując definicję odwrotnej transformaty Fouriera, można pokazać (dodatek A), że funkcja gęstości  $h(\mathbf{x})$   $n$ -wymiarowego, izotropowego, stabilnego wektora losowego może zostać przedstawiona w następującej postaci całkowej:

$$h(\mathbf{x}) = C(n) \int_0^\infty \int_0^1 r^{n-1} \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \cos(r\|\mathbf{x}\|t) (1-t^2)^{\frac{n-3}{2}} dt dr, \quad (2.44)$$

gdzie

$$C(n) = \frac{2^{2-n}}{\pi^{\frac{n+1}{2}} \Gamma(\frac{n-1}{2})}. \quad (2.45)$$

Aby zobrazować sposób, w jaki masa prawdopodobieństwa, rozważanych wektorów losowych, rozkłada się w przestrzeni przeszukiwań, na rysunku 2.6 przedstawiono przekroje  $\phi(r)$  funkcji gęstości (2.44), tj.  $\phi(h) = h(\mathbf{r}e_1)$ .



Rysunek 2.6: Przekrój poprzeczny funkcji gęstości izotropowego rozkładu stabilnego dla dwóch przykładowych wymiarów wektora losowego: (a) -  $n = 2$  (b) -  $n = 6$

**Uwaga 10** *Na podstawie rysunku 2.6, można zaobserwować analogiczną własność wektorów stabilnych, do tej prezentowanej w rozdziale 2.1, poświęconym*

ich jednowymiarowym odpowiednikom. Mianowicie, istnieje liczba  $r > 0$ , taka że wewnątrz kuli  $S_r^n(\mathbf{0})$  można zaobserwować uporządkowanie funkcji gęstości względem indeksu stabilności  $\alpha$ . Jednym ze sposobów, wyrażenia wspomnianego porządku, jest zależność:

$$\alpha_1 < \alpha_2 \Rightarrow \exists r > 0 P(\mathbf{X}_{\alpha_1}^s(\sigma) \in S_r^n(\mathbf{0})) > P(\mathbf{X}_{\alpha_2}^s(\sigma) \in S_r^n(\mathbf{0})). \quad (2.46)$$

Odnosząc (2.46) do problemu zaburzania rozwiązań w operatorze mutacji, uzasadniona jest hipoteza, mówiąca o tym, iż rozkłady o cięższych ogonach, znacznie intensywniej przeszukują przestrzeń rozwiązań w bezpośrednim sąsiedztwie punktu bazowego.

Jedną z konsekwencji sferycznej symetrii, jest fakt, iż funkcja gęstości ulega zmianie jedynie w przypadku zmiany odległości euklidesowej od środka symetrii rozkładu. W wielu przypadkach, pozwala to na sprowadzenie zadania wnioskowania statystycznego do analizy rozkładów jednowymiarowych. W takich sytuacjach ogromnego znaczenia nabiera pojęcie generatora gęstości rozkładu. Jest to funkcja skalarna  $G : \mathbb{R} \rightarrow \mathbb{R}$ , która umożliwi przedstawienie wielowymiarowej funkcji gęstości rozkładu izotropowego  $g(\cdot)$  w postaci:

$$g(\mathbf{x}) = c G(\mathbf{x}^T \mathbf{x}), \quad (2.47)$$

gdzie  $c$  jest stałą normalizującą<sup>5</sup>, jeśli tylko spełniony jest warunek [123]:

$$\int_0^\infty y^{n/2-1} G(y) dy < \infty. \quad (2.48)$$

Wykorzystując wzór (2.44), generator izotropowych rozkładów stabilnych może zostać przedstawiony w postaci

$$G_\alpha(k) = C(n) \int_0^\infty \int_0^1 r^{n-1} \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \cos(r\sqrt{kt}) (1-t^2)^{\frac{n-3}{2}} dt dr, \quad (2.49)$$

gdzie  $C(n)$  zostało zdefiniowane w (2.45). Dla podkreślenia wagi, generatorów gęstości, wystarczy nadmienić, że pozwalają on w zręczny sposób wyrazić rozkład normy losowego wektora izotropowego, która stanowi podstawę do analizy zdolności eksploracyjnych oraz eksploatacyjnych operatora mutacji (dodatek D):

$$P(\|\mathbf{Z}\| < R) = c \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^R r^{n-1} G(r^2) dr. \quad (2.50)$$

Wzór (2.50) nabiera szczególnego znaczenia przy analizie zjawiska określanego mianem efektu otoczenia [85], któremu dużo miejsca poświęcono w rozdziale 4.2.2.

## 2.3 Podsumowanie

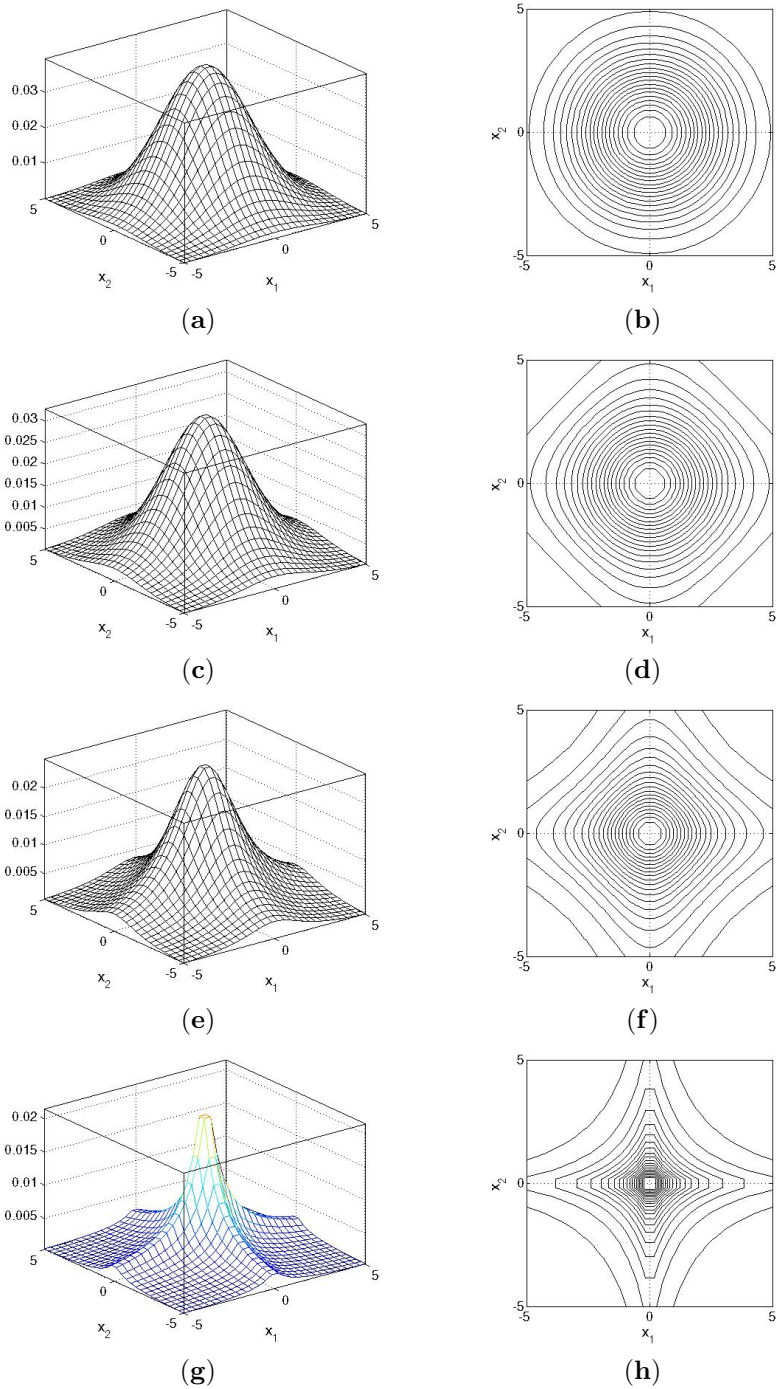
W pierwszej części rozdziału zaprezentowano najważniejsze definicje oraz pojęcia związane z klasą jednowymiarowych rozkładów stabilnych. W szczególności przedstawiono wpływ poszczególnych parametrów  $(\alpha, \beta, \sigma, \mu)$  na postać

<sup>5</sup>niezależną od  $\mathbf{x}$ , taką że  $\int_{\mathbb{R}^n} c G(\mathbf{x}^T \mathbf{x}) = 1$

rozkładu stabilnego, twierdzenia mówiące o braku momentów centralnych oraz asymptotycznym zachowaniu funkcji gęstości. Zwrócono uwagę na bardzo istotną cechę rozkładów stabilnych tzw. makromutacje, które mają kolosalne znaczenie dla skuteczności oraz odporności algorytmów stochastycznych. Ponadto pokazano, iż w przypadku stabilnej zmiennej losowej  $\alpha < 2$ , oczekiwana wartość przemieszczenia w przestrzeni rozwiązań, nie zależy od liczby iteracji. Prezentację najważniejszych właściwości, zamyka procedura generowania liczb pseudolosowych o zadanym rozkładzie stabilnym. Wnioski wyciągnięte z obserwacji przykładowych wykresów funkcji gęstości, pozwalają sądzić, że rozkłady o mniejszym indeksie stabilności mogą posiadać zarówno lepsze właściwości eksploracyjne jak i eksploatacyjne. W celu analitycznej weryfikacji wspomnianej hipotezy, udowodnione zostało twierdzenie mówiące o istnieniu momentów stabilnej statystyki porządkowej.

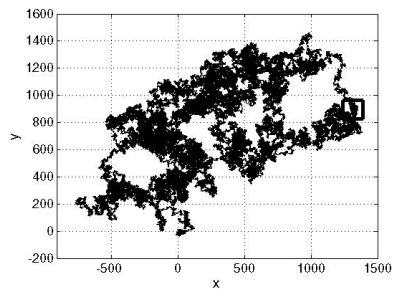
W części drugiej powyższego rozdziału zaprezentowane zostało pojęcie stabilnego wektora losowego. Podobnie jak w przypadku wielowymiarowego rozkładu normalnego, dla którego każda kombinacja liniowa poszczególnych składowych wektora losowego jest zmienną losową o rozkładzie normalnym, tak też w przypadku wektorów stabilnych, każdy wynik kombinacji liniowej jest jednowymiarowym rozkładem stabilnym. Niestety, w dużej mierze, na tym kończą się podobieństwa pomiędzy wspomnianymi rozkładami. Podczas gdy dla jednoznacznego określenia rozkładu normalnego wystarcza znajomość macierzy kowariancji, w przypadku wektorów stabilnych  $\alpha < 2$  pojęcie takie w ogóle nie istnieje. Dla określenia stopnia zależności stochastycznej pomiędzy elementami wektora stabilnego, stosuje się natomiast takie wskaźniki ilościowe jak kowariancja oraz kodyferencja. Pośród nietypowych właściwości stabilnych wektorów losowych, odnajdziemy regresję nieliniową, brak symetrii rozkładu warunkowego wokół wartości oczekiwanej, czy brak stabilności rozkładów warunkowych. Podobnie, jak to miało miejsce w przypadku stabilnych zmiennych losowych, także dla ich wielowymiarowych odpowiedników nieznane są analityczne postaci funkcji gęstości. Zamiast tego, do precyzyjnego zdefiniowania rozkładu stosuje się funkcję charakterystyczną, która w ogólności wymaga określenia tzw. miary spektralnej. Pojęcie miary spektralnej powoduje, że rozkłady stabilne wykazują się dużą elastycznością w modelowaniu skomplikowanych zależności pomiędzy zmiennymi decyzyjnymi. Pamiętając o tym, że jedynym sposobem na zwiększenie skuteczności stochastycznych algorytmów optymalizacji, jest precyzyjne dopasowanie rozkładu eksploracyjnego do rozważanego problemu, zasadność stosowania rozkładów stabilnych nie powinna budzić wątpliwości.

Brak dostatecznej wiedzy o optymalizowanym problemie, wymusza na algorytmach stosowanie rozkładu izotropowego. Rozkład ten zapewnia sprawliwie traktowanie wszystkich kierunków przestrzeni przeszukiwań. W związku z tym, zaprezentowane zostało również pojęcie stabilnego rozkładu o sferycznej symetrii. W szczególności wyprowadzone zostały zależności całkowite, umożliwiające zdefiniowanie tzw. generatora gęstości, który zostanie wykorzystany w dalszej części rozprawy do analizy zbieżności lokalnej strategii ewolucyjnej. Przykład błędzenia losowego dla stabilnych, izotropowych zaburzeń cząsteczki, obrazuje sposób, w jaki następować będzie proces eksploracji przestrzeni rozwiązań. Przytoczona zależność (2.50) posłuży w rozdziale 4.2.2 do wytłumaczenia efektu, polegającego na obserwowanym pogarszaniu skuteczności algorytmów ewolucyjnych w miarę zwiększania wymiaru przestrzeni przeszukiwań.

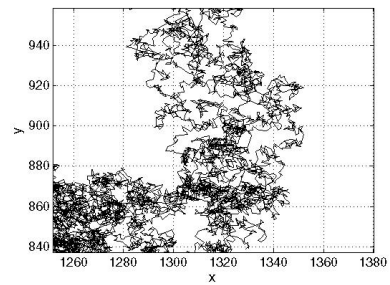


Rysunek 2.3: Funkcje gęstości wektora stabilnego  $\mathbf{X} = [X_1, X_2]^T$  z niezależnymi elementami  $X_1, X_2 \sim S_\alpha S(1)$ . Indeksy stabilności:  $\alpha = 2.0$  (a)-(b),  $\alpha = 1.5$  (c)-(d),  $\alpha = 1.0$  (e)-(f),  $\alpha = 0.5$  (g)-(h)

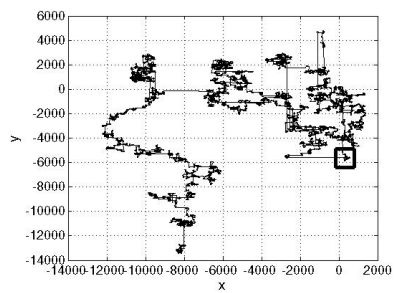




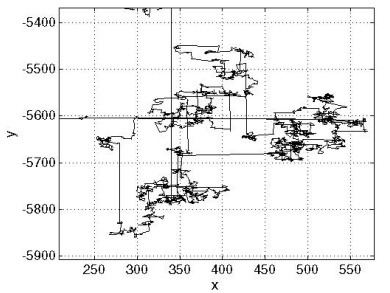
(a)



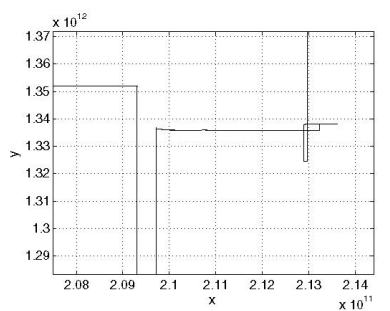
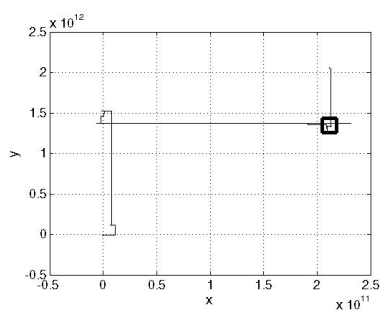
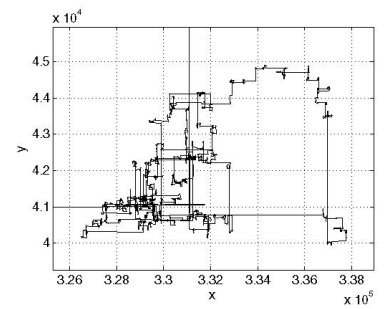
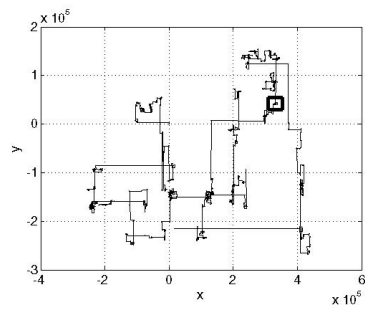
(b)



(c)



(d)



Rysunek 2.4: Symulacja błędzenia losowego na podstawie 1000000 modyfikacji cząsteczki zgodnie z izotropowymi rozkładami stabilnymi. Zakreślone trajektorie po prawej stronie odpowiadają różnym wykładnikom stabilności: (a) -  $\alpha = 2$ , (b) -  $\alpha = 1.5$ , (c) -  $\alpha = 1$ , (d) -  $\alpha = 0.5$ . Wykresy umieszczone z lewej strony przedstawiają powiększone zaznaczone obszary.

## Rozdział 3

# Klasyczne algorytmy ewolucyjne

Niniejszy rozdział dokonuje krótkiego wprowadzenia w ewolucyjne techniki optymalizacji globalnej. W szczególności przedstawione zostały klasyczne algorytmy fenotypowe: strategie ewolucyjne, programowanie ewolucyjne, oraz algorytm przeszukiwania ewolucyjnego z miękką selekcją. Zdefiniowane zostały podstawowe operatory genetyczne, których transparentna struktura posłuży w dalszej części rozprawy do przeprowadzenia wiarygodnej analizy działania algorytmów ewolucyjnych z mutacją stabilną.

### 3.1 Podstawy optymalizacji ewolucyjnej

Metody heurystyczne optymalizacji globalnej, stanowią odpowiedź na problemy związane z niemożnością przeglądu wszystkich rozwiązań. Wiele z tych technik opiera się na prostych prawach rządzących naturą, które w dziewiętnastym wieku doczekały się systematycznego opisu w pracach Charlsa Darwina (1859) oraz Johanna Gregora Mendela (1865). W świetle teorii ewolucji, gatunki doskonalały swoje cechy oraz zdolności w wyniku selekcji naturalnej. Lepiej dopasowane do środowiska osobniki posiadają większą szansę na przeżycie oraz posiadanie potomstwa. Tym samym, istnieje większa szansa na to, iż geny, które decydują o lepszym dopasowaniu, zostaną rozpowszechnione w całej populacji. Ponadto, w trakcie przekazywania informacji w postaci materiału genetycznego dochodzi do szeregu zaburzeń tzw. mutacji, w wyniku których pojawiają się nowe cechy, decydujące o indywidualności każdego osobnika. Powyżej przedstawiony proces, powtarza się cyklicznie z pokolenia na pokolenie, skutkując coraz lepszym dopasowaniem określonych gatunków do warunków w otaczających je świecie. Idea ta stała się podstawą do opracowania nowej klasy metod optymalizacji globalnej [100], [39] tzw. algorytmów ewolucyjnych.

W nazewnictwie właściwym algorytmom ewolucyjnym, bardzo często przejawia się nawiązanie do ich korzeni. W poszczególnych epokach (generacjach lub iteracjach), rozpatruje się populację osobników (zbiór rozwiązań problemu)  $P(k) = \{\mathbf{a}_i(k); i = 1, \dots, \eta\}$ . Poszczególne osobniki  $\mathbf{a}_i(k) \in \mathcal{G}$  reprezentuje element tzw. przestrzeni genotypowej, która stanowi najniższy możliwy poziom opisu osobnika. Zapis w postaci genotypu, definiuje jednocześnie wszelkie cechy określo-

ne, z kolei na przestrzeni fenotypowej  $\mathcal{D}$ . Jednoznaczność zapisu genotypowego oraz dualność jego reprezentacji, podkreślona jest poprzez funkcję dekodującą  $\zeta : \mathcal{G} \rightarrow \mathcal{D}$ . Stopień przydatności cech reprezentowanych przez materiał genetyczny osobnika, zostaje poddany weryfikacji środowiska za sprawą tzw. funkcji dopasowania (funkcji celu w zadaniu optymalizacji)  $\phi : \mathcal{D} \rightarrow \mathbb{R}$ .

Początkowa populacja, tworzona jest na ogół, poprzez przypisanie każdemu osobnikowi losowo wybranego punktu z przestrzeni genotypowej. Następnie, w procesie reprodukcji następuje wyodrębnienie części populacji  $S : G^\eta \rightarrow G^{\eta'}$ , która w dalszej kolejności otrzyma szansę na przekazanie swojego materiału genetycznego przyszłym pokoleniom. W zależności od klasy algorytmu, rozważa się różne techniki reprodukcji, zazwyczaj wymagające ustalenia pewnych parametrów zewnętrznych  $\theta_s$ . Osobniki, które znalazły się w uprzywilejowanej części populacji, wystawione są na działanie operatorów genetycznych: rekombinacji  $R : G^{\eta'} \rightarrow G^{\eta''}$  oraz mutacji  $M : G^{\eta''} \rightarrow G^{\eta'''}$ . Obie operacje przebiegają pod kontrolą pewnych dodatkowych parametrów sterujących:  $\theta_r$  oraz  $\theta_m$ . W wyniku ich działania, dochodzi do powstania nowej populacji  $P'(k) = \{\mathbf{a}'_i(k); i = 1, \dots, \eta'''\}$  o zwykle odmiennym zapisie genotypowym. Następnie, za pomocą funkcji dekodującej, dochodzi do oceny nowo-uformowanej populacji  $\phi'_i(k) = \phi(\zeta(\mathbf{a}'_i(k)))$ . Na jej podstawie operator  $T : P(k)' \times P(k) \rightarrow P(k+1)$  odpowiadający mechanizmowi selekcji naturalnej, tworzy kolejne pokolenie osobników. Cały proces jest cyklicznie kontynuowany do czasu spełnienia pewnego kryterium stopu  $\iota(P(k+1))$ . Powyższy schemat zaprezentowany został w tabeli 3.1.

Tabela 3.1: Ogólny schemat algorytmu ewolucyjnego [85]

---

### I. Inicjacja

a) *Losowe utworzenie populacji początkowej*

$$P(0) = \{\mathbf{a}_i(0); i = 1, \dots, \eta\}$$

b) *Ocena populacji początkowej*

$$P(0) \rightarrow \Phi(P(0)) = \{\phi_i(0) = \phi(\zeta(\mathbf{a}_i(0))); i = 1, \dots, \eta\}$$

c)  $k = 1$

### II. Powtarzaj

a) *Reprodukcja*

$$P'(k) = S(P(k)|\boldsymbol{\theta}_s) = \{\mathbf{a}'_i(k); i = 1, \dots, \eta'\}$$

b) *Rekombinacja*

$$P''(k) = R(P'(k)|\boldsymbol{\theta}_r) = \{\mathbf{a}''_i(k); i = 1, \dots, \eta''\}$$

c) *Mutacja*

$$P'''(k) = M(P''(k)|\boldsymbol{\theta}_m) = \{\mathbf{a}'''_i(k); i = 1, \dots, \eta''\}$$

d) *Ocena*

$$P'''(k) \rightarrow \Phi(P'''(k)) = \{\phi_i(k) = \phi(\zeta(\mathbf{a}'''_i(k))); i = 1, \dots, \eta''\}$$

e) *Sukcesja*

$$P''(k) = T(P'''(k) \cup P(k)|\boldsymbol{\theta}_T) = \{\mathbf{a}_i(k+1); i = 1, \dots, \eta\}$$

f)  $k = k + 1$

**Dopóki**  $\iota(P(k)) = \text{prawda}$

---

Postęp w dziedzinie komputerowych technik obliczeniowych, doprowadził na początku lat osiemdziesiątych, do burzliwego rozwoju metod optymalizacji globalnej wzorowanych na schemacie 3.1. Obecnie, można już mówić o tradycyjnym podziale algorytmów ewolucyjnych na cztery główne podklasy [74, 111]:

- algorytmy genetyczne,
- strategie ewolucyjne,
- programowanie ewolucyjne,
- programowanie genetyczne.

Różnice, które decydują o przynależności poszczególnych rozwiązań do wyżej wymienionych grup, polegają głównie na sposobie reprezentacji osobnika oraz stosowaniu odmiennych operatorów genetycznych. Szczegółowy opis każdej klasy można znaleźć w monografiach [85, 74]. Zakres problematyki, podjętej w niniejszej rozprawie, ogranicza się jedynie do pewnej specyficznej podgrupy algorytmów ewolucyjnych tzw. algorytmów fenotypowych.

## 3.2 Ewolucyjne algorytmy fenotypowe

Cechą charakterystyczną algorytmów fenotypowych jest definicja osobnika, która nie uwzględnia przestrzeni genotypowej. Tym samym, każdy osobnik zdefiniowany jest jedynie na przestrzeni, która bezpośrednio stanowi dziedzinę rozważanego problemu. Jak to zostało ujęte we wstępie, rozprawa prawie wyłącznie skupia się na problemach, dla których zbiór rozwiązań utożsamiany jest z  $\mathbb{R}^n$ . Poszczególne osobniki, zyskują zatem reprezentację  $\mathbf{a} = (\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\theta} \in \Theta)$ , gdzie  $\mathbf{x}$  reprezentuje rozwiązanie problemu,  $\boldsymbol{\theta}$  natomiast oznacza dodatkowy wektor parametrów, wykorzystywany w bardziej zaawansowanych technikach obliczeniowych. Funkcja dekodująca, w tej sytuacji, sprowadza się do bardzo prostej postaci  $\zeta(\mathbf{a}) = \mathbf{x}$ . Określenie wszystkich operatorów genetycznych z wykorzystaniem wiedzy dziedzinowej, bardzo często skutkuje znacznie większą przejrzystością oraz efektywnością algorytmów fenotypowych, w stosunku do ich genetycznych odpowiedników. Fakt, iż każdy osobnik w bezpośredni sposób stanowi rozwiązanie problemu optymalizacji, automatycznie narzuca odmienną postać operatorów genetycznych.

### Rekombinacja

Operator rekombinacji odgrywa specyficzną rolę w fenotypowych algorytmach ewolucyjnych. Osobniki powstałe w wyniku jego działania, dzielą cechy fenotypowe swoich rodziców. Ogólnie operatory rekombinacji można rozdzielić na dwie klasy: dyskretne oraz uśredniające. Przed wykonaniem każdej operacji rekombinacji, w pierwszej kolejności dokonuje się losowego wyboru  $\eta'$  rodziców  $P' = \{\mathbf{a}_{h_1}, \dots, \mathbf{a}_{h_{\eta'}}\}$ , którzy będą uczestniczyli w tworzeniu nowego potomka  $\mathbf{a}_r'' = (\mathbf{x}_r'', \boldsymbol{\theta}_r'')$ . Rekombinacja dyskretna polega na formowaniu wektorów  $\mathbf{x}_r, \boldsymbol{\theta}_r$  nowotworzonego potomka, poprzez kopiowanie elementów z losowo wybieranych rodziców. Tym sposobem, osobnik  $\mathbf{a}_r''$  powstaje w wyniku przetasowania danych przechowywanych w populacji. Inaczej przedstawia się sprawa w przypadku rekombinacji uśredniającej. Nowy osobnik powstaje według reguły:

$$\mathbf{x}'' = \frac{1}{\rho} \sum_{i=1}^{\rho} \mathbf{x}'_i$$
$$\boldsymbol{\theta}'' = \frac{1}{\rho} \sum_{i=1}^{\rho} \boldsymbol{\theta}'_i$$

Cechą charakterystyczną operatorów rekombinacji jest fakt, iż w wyniku ich stosowania, następuje eksploatacja przestrzeni ograniczona do hiperkostki zawierającej aktualną populację rozwiązań. Marginalny wpływ, obu typów operatorów rekombinacji, jaki wywierają na poprawę skuteczności optymalizacyjnej powoduje, że zwykle operatory te są pomijane w numerycznych implementacjach algorytmów fenotypowych.

### Mutacja

W odróżnieniu od operatora rekombinacji, który bazował na informacji zawartej w fenotypowej reprezentacji osobników rodzicielskich, celem mutacji jest wprowadzenie zróżnicowania do populacji rozwiązań. W algorytmach fenotypowych

operator ten, polega na zaburzaniu rozwiązań, za pomocą pewnego wektora losowego:

$$\mathbf{x}''' = \mathbf{x}'' + \mathbf{Z}, \quad (3.1)$$

Z teoretycznego punktu widzenia, najistotniejszym warunkiem, który narzuconym na wektor  $\mathbf{Z}$  jest zapewnienie osiągalności każdego obszaru przestrzeni rozwiązań przy skończonej liczbie mutacji. Jest to zarazem warunek konieczny zbieżności algorytmu ewolucyjnego do rozwiązania globalnego. Do najczęściej stosowanych rozkładów zaburzających spotykanych w literaturze należą rozkłady: normalny oraz Cauchy'ego.

### Selekcja

Rolą mutacji oraz w mniejszym stopniu rekombinacji jest wprowadzenie zmian w strukturę fenotypową populacji. Algorytm ewolucyjny bazujący jedynie na dwóch powyższych operatorach, przyjął by postać puli osobników, których zachowanie do złudzenia przypominałoby błędzenie losowe. Trudno zatem przypuszczać, iż proces taki mógłby posłużyć jako skuteczny algorytm optymalizacji globalnej. W algorytmach ewolucyjnych zadanie ukierunkowania procesu ewolucji w stronę obszarów o lepszym dopasowaniu spoczywa na operatorze selekcji. Wyróżnić można dwa odmienne podejścia do projektowania operatora selekcji. Pierwszą z nich jest tzw. selekcja twarda (ang. *hard selection*), które gwarantuje przetrwanie tylko osobnikowi lub osobnikom o najlepszym dopasowaniu. Najczęściej proces ten ma charakter deterministyczny, i polega na uszeregowaniu populacji względem wartości funkcji celu i uwzględnianiu tylko jej pewnej części w formowaniu następnego pokolenia. Jeśli dodatkowo dopuszcza się możliwość konkurowania rodziców ze swoimi potomkami, otrzymujemy algorytm, dla którego ciąg najlepszych rozwiązań jest ciągiem nierosnącym (przy zadaniu minimalizacji) bądź niemalejącym (zadanie maksymalizacji funkcji celu). Alternatywnym rozwiązaniem jest tzw. selekcja miękka, umożliwiająca przeżycie także najgorzej dopasowanym osobnikom. Wzorując się na podstawach selekcji naturalnej, osobniki reprezentujące korzystniejsze rozwiązania, zyskują większe prawdopodobieństwo przeżycia. Ponadto, zastosowanie selekcji miękkiej umożliwia powstanie tzw. dryftu populacji [42] - zjawiska ułatwiającego eksplorację środowiska.

Dodatkowo, niemalże każdy operator selekcji udostępnia parametr, za pomocą którego możliwe jest płynne kontrolowanie tzw. presji selekcyjnej. Presja selekcyjna może być postrzegana jako pojęcie określające stopień dominacji najlepszych osobników podczas formowania kolejnego pokolenia i nabiera szczególnego znaczenia przy analizie zbieżności algorytmów ewolucyjnych. Do najbardziej popularnych metod selekcji miękkiej, stosowanych również w kolejnych rozdziałach rozprawy, należą:

- selekcja turniejowa,
- selekcja proporcjonalna.

**Selekcja turniejowa** (ang. *tournament selection*) polega na utworzeniu tzw. grupy turniejowej  $P_G$ , złożonej z losowo wybranych  $T_G$  osobników aktualnej populacji. Każdy osobnik posiada identyczne prawdopodobieństwo znalezienia się w  $P_G$ , przy czym losowanie odbywa się bez zwracania. Do następnej populacji przeżywa tylko najlepszy osobnik z grupy turniejowej. Cały proces powtarza

się  $\eta$  razy, aż do chwili utworzenia kompletnej populacji potomnej. Rozmiar turnieju, pozwala kontrolować presję selekcyjną. Z jednej strony, gdy  $T_G = \eta$  otrzymujemy populację potomną w zupełności zdominowaną przez najlepszego osobnika. Jeśli  $T_G = 1$  funkcja dopasowania nie jest brana w ogóle pod uwagę, a populacja potomna tworzona jest poprzez losowy wybór osobników.

Tabela 3.2: Selekcja turniejowa

---

**Dane wejściowe**

*Populacja osobników*  $P(k) = \{\mathbf{a}_i(k); i = 1, \dots, \eta\}$

*Dopasowanie osobników*  $\Phi(P_G) = \{\phi_i(k); i = 1, \dots, \eta\}$

*Rozmiar turnieju*  $T_G \leq \eta$

**Dane wyjściowe**

*Populacja po selekcji*  $P(k+1) = \{\mathbf{a}_i(k+1); i = 1, \dots, \eta\}$

**Algorytm**

$i = 0$

$P(k+1) = \emptyset$

*Powtarzaj*

$j = 0$

$P_g = \emptyset$

*Powtarzaj*

$\mathbf{a}_j^G = \text{random}(P(k) \setminus P_g)$

$P_G \leftarrow \mathbf{a}_j^G$

$j = j + 1$

*Dopóki*  $j < T_G$

$P(k+1) \leftarrow \min(\Phi(P_G))$

$i = i + 1$

*Dopóki*  $i < \eta$

---

**Selekcja proporcjonalna** została początkowo zaproponowaną przez Holandę [54] jako operator wykorzystywany w algorytmach genetycznych. Metoda opiera się na podziale odcina prostej rzeczywistej  $[0, 1]$ . Każdemu osobnikowi odpowiada pewien odcinek  $I_{p_i} \subset [0, 1]$ , którego długość jest proporcjonalna do jakości reprezentowanego przez niego rozwiązania. Następnie, z rozkładem równomiernym na przedziale  $[0, 1]$ , wylosowany zostaje punkt  $r \in [0, 1]$ . Do następnej generacji przeżywa ten osobnik, którego przedział  $I_{p_i}$  zawiera  $r$ . Prawdopodobieństwo wylosowania osobnika jest zatem proporcjonalne do wartości jego funkcji dopasowania.

Tabela 3.3: Selekcja proporcjonalna

---

**Dane wejściowe**

*Populacja osobników*  $P(k) = \{\mathbf{a}_i(k); i = 1, \dots, \eta\}$

*Dopasowanie osobników*  $\Phi(P_G) = \{\phi_i(k); i = 1, \dots, \eta\}$

**Dane wyjściowe**

*Populacja po selekcji*  $P(k+1) = \{\mathbf{a}_i(k+1); i = 1, \dots, \eta\}$

**Algorytm**

$i = 1$

$s_0 = 0$

*Powtarzaj*

$$s_i = s_{i-1} + \frac{\phi_i(k)}{\sum_{j=1}^{\eta} \phi_j(k)}$$

$i = i + 1$

*Dopóki*  $i \leq \eta$

$P(k+1) = \emptyset$

$i = 1$

*Powtarzaj*

$r = \text{random}([0, 1])$

$P(k+1) \leftarrow \mathbf{a}_l$  gdzie  $l$  spełnia  $s_{l-1} < r < s_l$

$i = i + 1$

*Dopóki*  $i \leq \eta$

---

Jest oczywiste, iż schemat selekcji proporcjonalnej zamieszczony w tabeli 3.3 odnosi się jedynie do zadania maksymalizacji i to tylko wówczas, gdy funkcja celu przyjmuje wartości dodatnie. W celu umożliwienia stosowania wspomnianego operatora także w zadaniach nie spełniających powyższych warunków, stosuje się następujące modyfikacje zbioru jakości dopasowania  $\Phi(P_G) = \{\phi_i(k); i = 1, \dots, \eta\}$  [85]:

- dla zadania maksymalizacji

$$\phi_i(k) = \phi(\mathbf{x}_i(k)) - \phi_{\min}(k) + \left(\frac{1}{\eta}\right)^2 \quad (3.2)$$

- dla zadania minimalizacji

$$\phi_i(k) = \phi_{\max}(k) - \phi(\mathbf{x}_i(k)) + \left(\frac{1}{\eta}\right)^2 \quad (3.3)$$

gdzie  $\phi_{\min}(k)$  oraz  $\phi_{\max}(k)$  oznaczają najmniejszą i największą wartość w zbiorze  $\Phi(P_G)$ .



### 3.2.1 Strategie ewolucyjne

Strategie ewolucyjne [20] korzystają ze specyficznej reprezentacji osobnika w postaci:  $\mathbf{a} = \{\mathbf{x}, \boldsymbol{\theta}, \phi(\mathbf{x})\}$ , gdzie  $\mathbf{x}$  oznacza wektor zmiennych decyzyjnych,  $\phi(\mathbf{x})$  jest wartością dopasowania osobnika a  $\boldsymbol{\theta}$  jest wektorem parametrów warunkujących proces mutacji. Ogólny sposób notacji  $(\mu/\rho + \lambda)$ ES odzwierciedla strukturę algorytmu. Liczba osobników biorących udział w powstaniu pojedynczego potomka (ang. mixing number) oznaczana jest jako  $\rho$ . W przypadku gdy  $\rho = 1$ , otrzymujemy algorytmy pozbawione operatora rekombinacji, dla których przyjęło się stosować notację postaci  $(\mu, \lambda)$ ES lub  $(\mu + \lambda)$ ES -  $\mu, \lambda$  oznaczają rozmiar populacji rodzicielskiej oraz potomnej. Symbole „+” oraz „,” obecne w notacji strategii ewolucyjnej, określają sposób selekcji, która dla algorytmów tej klasy jest w zupełności deterministyczna i polega na wyborze  $\mu$  najlepszych osobników z puli selekcyjnej. Symbol „+” oznacza, że nowa populacja formowana jest na podstawie połączonych populacji: rodzicielskiej oraz potomnej. Zastosowanie symbolu „,” oznacza, iż populacja rodzicielska nie bierze udziału w selekcji. Ogólny schemat strategii ewolucyjnej, w przypadku nieskorelowanych mutacji, oraz definicji osobnika wzbogaconej o zestaw odchyłeń standardowych t.j.:  $\boldsymbol{\theta} = \boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_n]^T$  przedstawiony został w tabeli 3.4.

Tabela 3.4: Ogólny schemat strategii ewolucyjnej  $(\mu/\rho^+, \lambda)$ -ES

---

**Dane wejściowe**

- $\rho$  - parametr rekombinacji  
 $\mu$  - liczebność populacji rodzicielskiej  
 $\lambda$  - liczebność populacji potomnej  
 $\{", +", ", "\}$  - sposób selekcji

**I. Inicjacja**

- a)  $k = 1$   
b) Losowe utworzenie populacji początkowej  
 $P(0) = \{\{\mathbf{x}_i(0), \sigma_i(0), \phi_i(0)\}; i = 1, \dots, \mu\}$

**II. Powtarzaj**

- $i = 1$   
 $P_\lambda = \emptyset$   
Powtarzaj
- a) Reprodukacja  
 $P'(k) = S(P(k)|\rho) = \{\mathbf{a}'_{h_1}(k), \dots, \mathbf{a}'_{h_\rho}(k)\}$
- b) Rekombinacja parametrów  
 $\sigma''_i(k) = R_\sigma(P'(k))$
- c) Rekombinacja rozwiązania  
 $\mathbf{x}''_i(k) = R_X(P'(k))$
- d) Mutacja parametrów  
 $\sigma'''_{i,j}(k) = \sigma''_{i,j}(k) \exp(\tau' \mathcal{N}(0, 1) + \tau \mathcal{N}_j(0, 1)), \quad j = 1, \dots, n$
- e) Mutacja rozwiązania  
 $\mathbf{x}'''_{i,j}(k) = \mathbf{x}''_{i,j}(k) + \sigma'''_{i,j}(k) \mathcal{N}(0, 1), \quad j = 1, \dots, n$
- f) Ocena  
 $\phi'''_i(k) = \phi(\mathbf{x}'''_i(k))$
- g) Przypisz  
 $P_\lambda \leftarrow \mathbf{a}'''_i(k) = \{\mathbf{x}'''_i(k), \sigma'''_i(k), \phi'''_i(k)\}$   
 $i = i + 1$
- Dopóki  $i \leq \lambda$
- Selekcja  
 $P(k+1) = T(P_\lambda \cup P(k))$  jeśli  $", +"$   
lub  
 $P(k+1) = T(P_\lambda)$  jeśli  $", "$
- Przypisz  $k = k + 1$

**Dopóki  $\iota(P(k)) = \text{prawda}$**

---

### 3.2.2 Programowanie ewolucyjne

Początków programowania ewolucyjnego upatrywać należy w procesie modelowania sztucznej inteligencji na drodze samoczynnej organizacji [39]. Problemem, który w bezpośredni sposób przyczynił się do powstania tej klasy algorytmów ewolucyjnych, był problem odkrywania gramatyki nieznanego języka. Przy danym zestawie symboli oraz przykładach wyrażeń syntaktycznie poprawnych, gramatyka była modelowana za pomocą automatu skończonego. Zadanie to polegało na odnalezieniu zbiorów stanów, postaci funkcji przejść oraz funkcji wyjść wspomnianego automatu. Funkcja dopasowania osobnika - automatu skończonego, polegała na określeniu liczby prawidłowych klasyfikacji (jako poprawnych syntaktycznie) wyrażeń poznawanego języka. Osobnik mógł podlegać modyfikacją, która w przypadku programowania ewolucyjnego, polegała wyłącznie na dokonywaniu mutacji: *dodanie stanu, usunięcie stanu, zmiana stanu początkowego, zmiana stanu wyjść, zmiana funkcji przejść*. Choć wyniki prac były wprawdzie zachęcające - obserwowano uczenie się automatów, jednakże programowanie ewolucyjne w swojej pierwotnej formie nie zyskało na popularności.

Lata osiemdziesiąte przyniosły ponowne zainteresowanie programowaniem ewolucyjnym, które ulegało znacznym zmianom, upodabniającym je do strategii ewolucyjnych. Modyfikacje dotyczyły przede wszystkim dziedziny zastosowań - schemat programowania ewolucyjnego został rozwinięty przez Dawida Fogela, w kierunku optymalizacji numerycznej. Jego wkład polegał na wprowadzeniu reprezentacji rzeczywistoliczbowej oraz dostosowanego do niej operatora mutacji, wykorzystującego nieskorelowany wielowymiarowy rozkład normalny o zerowej wartości oczekiwanej. Fogel wzbogacił genotyp osobnika o chromosom, zawierający wektor standardowych odchyłeń (po jednej wartości dla każdej zmiennej niezależnej) oraz wprowadził mechanizmy modyfikacji zasięgu mutacji. W początkowych implementacjach programowania ewolucyjnego, nie były uwzględniane operatory rekombinacji i preselekcji. Każdy rodzic generował osobniki na drodze losowej mutacji, która postępowała zgodnie z rozkładem równomiernym. Tylko połowa najlepszych osobników ze zbioru rodziców i potomków mogły przeżyć do następnej epoki. W roku 1991 Fogel zaproponował nieco odmienny algorytm, w którym każdy osobnik reprezentowany był przez uporządkowaną parę  $\{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\sigma} \in \mathbb{R}_+^n\}$ , gdzie  $\mathbf{x}$  reprezentuje fenotyp osobnika,  $\boldsymbol{\sigma}$  jest samoadaptacyjnym wektorem odchylenia standardowego wykorzystywanego w operatorze mutacji.

Inicjacja algorytmu programowania ewolucyjnego następuje poprzez losowe utworzenie populacji początkowej  $P(0) = \{\mathbf{a}_i = \{\mathbf{x}_i(0), \boldsymbol{\sigma}_i(0)\}, i = 1, \dots, \eta\}$ , gdzie elementy  $\mathbf{x}_i(0)$  oraz  $\boldsymbol{\sigma}_i(0)$  każdego osobnika, zostają wybrane zgodnie z rozkładem równomiernym określonym na wcześniej zdefiniowanych hiperkostkach. Cechą charakterystyczną programowania ewolucyjnego jest fakt, iż każdy osobnik populacji poddawany jest mutacji w następujący sposób:

$$x_{i,j}(k) = x_{i,j}(k) + \mathcal{N}(0, \sigma_{i,j}), \quad (3.4)$$

$$\sigma_{i,j}(k) = \sigma_{i,j}(k) \exp(\tau' \mathcal{N}(0, 1) + \tau \mathcal{N}_j(0, 1)) \quad (3.5)$$

dla  $i = 1, \dots, \eta, j = 1, \dots, n$ .

W procesie formowania nowego pokolenia biorą udział wspólnie populacja rodziców oraz potomków. Dla każdego osobnika wybiera się losowo  $q$  elementów z połączonych populacji oraz zapamiętuje liczbę osobników o gorszym dopasowaniu. Tworzony na tej podstawie ranking służy następnie do stworzenia nowej

generacji.

Popularność programowania ewolucyjnego wynika także z faktu istnienia dowodu asymptotycznej zbieżności algorytmu do rozwiązania globalnego [39].

### 3.2.3 Przeszukiwanie ewolucyjne z miękką selekcją

Przeszukiwanie ewolucyjne z miękką selekcją (ang. Evolutionary Search with Soft Selection - ESSS) bazuje na bardzo uproszczonym modelu ewolucji [41]. Osobnik, w odróżnieniu od wyżej przytoczonych algorytmów fenotypowych, reprezentowany jest jedynie przez rozwiązanie zadania optymalizacji  $\mathbf{a} = \mathbf{x} \in \mathbb{R}^n$ . Proces optymalizacji przebiega na podstawie kolejno powtarzanych procedur: mutacji oraz selekcji. Algorytm zakłada przebieg procesu ewolucji w środowisku, zdefiniowanym przez funkcję przyjmującą wartości dodatnie. Początkowa populacja generowana jest przez losowe zaburzenia punktu startowego  $\mathbf{x}_0$  z rozkładem normalnym  $\mathcal{N}(\mathbf{0}, \sigma \mathbf{I}_n)$ . Mutacji podlegają wszystkie osobniki populacji, a powstała w wyniku selekcji proporcjonalnej (tabela 3.3) populacja formuje następne pokolenie. Proces ewolucji kontynuowany jest do czasu wykonania założonej z góry liczby iteracji  $T_{\max}$ .

Okazuje się, że miękka selekcja zastosowana w algorytmie, oraz brak adaptacji skali rozkładu podczas optymalizacji, skutkuje brakiem zbieżności algorytmu do rozwiązania globalnego. Analiza analityczna systemu dynamicznego, symulującego proces ewolucji algorytmu ESSS wykazała, iż nie jest on zbieżny do rozwiązania globalnego. Z tego też względu, w wielu zastosowaniach praktycznych, metodę tą wzbogaca się o technikę optymalizacji lokalnej. Szczegółowy schemat algorytmu zamieszczony został w tabeli 5.2

Tabela 3.5: Algorytm przeszukiwań ewolucyjnych z miękką selekcją - ESSS

**Dane wejściowe**

- $\eta$  – rozmiar populacji;
- $T_{\max}$  – maksymalna liczba generacji,
- $\sigma$  – odchylenie standardowe mutacji,
- $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  – funkcja celu,
- $\mathbf{x}_0^0$  – punkt startowy,
- $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_n)$  – losowy wektor zaburzający.

**I. Inicjacja**

$$P(0) = (\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_\eta^0), \quad \mathbf{x}_k^0 = \mathbf{x}_0^0 + \mathbf{Z}, \quad k = 1, 2, \dots, \eta$$

$$k = 1$$

**II. Powtarzaj**

a) *Ocena*

$$\Phi(P(t)) = (\phi_1^t, \phi_2^t, \dots, \phi_\eta^t), \quad \text{gdzie } \phi_k^t = \phi(\mathbf{x}_k^t), \quad k = 1, 2, \dots, \eta.$$

b) *Selekcja proporcjonalna*

$$P(t) \longrightarrow P(t)' = (\mathbf{x}_{h_1}^t, \mathbf{x}_{h_2}^t, \dots, \mathbf{x}_{h_\eta}^t),$$

c) *Mutacja*

$$P(t)' \longrightarrow P(t+1);$$

$$\mathbf{x}_k^{t+1} = \mathbf{x}_{h_k}^t + \sigma \mathbf{Z}, \quad k = 1, 2, \dots, \eta.$$

$k = k + 1$

**Dopóki**  $k < T_{\max}$ .

### 3.3 Podsumowanie

Algorytmy ewolucyjne stanowią potężne narzędzie optymalizacji globalnej. Wśród licznych zalet, na które zwrócono uwagę w powyższym rozdziale, na pierwszy plan wysuwają się kwestie związane z uniwersalnością omawianych metod. Cecha ta przejawia się przede wszystkim brakiem założeń dotyczących różniczkowalności czy nawet ciągłości funkcji celu - właściwości niezbędnych przy zastosowaniu tradycyjnych metod optymalizacji. Na przestrzeni ostatniej dekad, zaproponowano wiele rozwiązań optymalizacyjnych wzorowanych na procesie ewolucji żywych organizmów. W każdym z nich można jednak doszukać się wspólnych, dla klasy algorytmów, procedur: rekombinacji, mutacji oraz selekcji. Przytoczone operatory mają za zadanie utworzenie pewnej puli rozwiązań alternatywnych oraz późniejsze wyselekcjonowanie stosunkowo najlepszych osobników. Iteracyjny proces optymalizacji sprowadza się w ten sposób do ewolucji rozwiązań alternatywnych w stronę obszarów przestrzeni rozwiązań o lepszym dopasowaniu.

W rozdziale przedstawiono szczegółowo klasyczne algorytmy fenotypowe: stra-

tegie ewolucyjne [20], programowanie ewolucyjne [39] oraz algorytm przeszukiwań ewolucyjnych z miękką selekcją. Ich wspólną cechą jest sposób definiowania osobnika z pominięciem przestrzeni genotypowej. Oznacza to, że proces ewolucji odbywa się jedynie w przestrzeni będącej równocześnie dziedziną problemu optymalizacji. W algorytmach tego typu, do rzadkości należą implementacje wykorzystujące operator rekombinacji. Największe znaczenie dla skuteczności algorytmów fenotypowych, odgrywa operator mutacji, a zwłaszcza rozkład wektora losowego, wykorzystywanego do formowania rozwiązań alternatywnych. Obecnie, po przytłaczającej fali rozwiązań bazujących na rozkładzie normalnym, ostatnie lata przeniosły nieco ciężar zainteresowań w stronę rozkładu Cauchy'ego [107],[104],[85]. Z tego punktu widzenia problematyka podjęta w rozprawie wydaje się niezwykle ważna, gdyż oba rozkłady należą do znaczenie szerszej klasy, jaką jest klasa rozkładów stabilnych.

## Rozdział 4

# Mutacja stabilna w fenotypowych algorytmach ewolucyjnych

W rozdziale tym przedstawiamy rodzinę operatorów mutacji, zdefiniowanych za pomocą stabilnych wektorów losowych. W pierwszej kolejności zwrócono uwagę na wady operatora mutacji polegającego na zaburzaniu każdej współrzędnej wektora rozwiązań za pomocą niezależnych, identycznych, stabilnych zmiennych losowych. Scharakteryzowany został tzw. efekt symetrii, który może mieć negatywny wpływ na skuteczność ewolucyjnych technik optymalizacji. Próby zniwelowania wspomnianego problemu, poszukuje się w dalszej części rozdziału, proponując operator mutacji oparty o izotropowe rozkłady stabilne. W szczególności, za pomocą badań eksperymentalnych, przeprowadzona została analiza pod kątem pogodzenia dwóch przeciwstawnych cech algorytmu optymalizacji: eksploracji oraz eksploatacji. Scharakteryzowany został tzw. efekt otoczenia - zjawisko wyjaśniające pogorszenie jakości otrzymywanych rozwiązań wraz ze wzrostem wymiaru przestrzeni przeszukiwań. Ponadto, dokonana została teoretyczna analiza zbieżności lokalnej strategii ewolucyjnej  $(1 + 1)$ ES z wykorzystaniem sferycznej funkcji celu. Dużo miejsca poświęcono również kwestii odporności mutacji stabilnej na wybór początkowych warunków algorytmu ewolucyjnego. Za pomocą środowiska optymalizacyjnego zawierającego szereg popularnych funkcji testowych, skonfigurowano optymalną strategię adaptacyjną dla izotropowych mutacji stabilnych.

### 4.1 Efekt symetrii

Jednym z najprostszych sposobów uogólnienia stabilnych zmiennych losowych do ich wielowymiarowych odpowiedników, jest złożenie wektora  $\mathbf{X}_\alpha$  z niezależnych komponentów o identycznym rozkładzie, t.j:

$$\mathbf{X}_\alpha = [X_\alpha, \dots, X_\alpha]^T, \quad (4.1)$$

gdzie  $X_\alpha \sim S_\alpha S(\sigma)$ . Wykorzystując jedynie twierdzenie 1 o rozkładzie sumy zmiennych losowych, można w bardzo prosty sposób udowodnić stabilność ca-

tego wektora (4.1). Praktyka ta, w przypadku rozkładu normalnego jest bardzo często stosowana w algorytmach optymalizacji stochastycznej [23],[20],[112],[85],[88].

O ile, w przypadku modelu gaussowskiego, konstrukcja ta powoduje uzyskanie wektora o sferycznej symetrii, o tyle w przypadku rozkładów o indeksie stabilności  $\alpha < 2$ , wektory te, takiej symetrii już nie posiadają. Przekonują o tym wykresy funkcji gęstości - rysunek 2.3 zamieszczony w rozdziale 2.2.

Brak sferycznej symetrii musi wpływać na skuteczność optymalizacyjną algorytmów ewolucyjnych. Pierwszym, który zwrócił na to zjawisko uwagę, był Obuchowicz [86, 85]. Przeprowadzając szereg badań symulacyjnych, z zastosowaniem anizotropowej mutacji opartej o rozkład Cauchy'ego, zaobserwował, iż duże znaczenie ogrywa kwestia zorientowania układu współrzędnych [85].

W [88] oraz [61] autorzy, analizując skuteczność mutacji stabilnej postaci (4.1) zwrócili uwagę na fakt wzrostu średniej liczby epok, potrzebnej do pokonania siodła ewolucyjnego, w zależności od jego zorientowania, w stosunku do kierunków wyznaczonych przez osie układu odniesienia.

Obserwując proces zagęszczenia masy prawdopodobieństwa wokół kierunków wyznaczonych przez osie układu współrzędnych (wykresy funkcji gęstości 2.3 zamieszczone w rozdziale 2.2), można przypuszczać iż zjawisko to, będzie nasilało się dla mniejszych wartości indeksu  $\alpha$ . W celu ilościowego scharakteryzowania wspomnianego efektu, rozważmy zbiór  $n$ -wymiarowych wektorów  $\{\mathbf{b}_k\}_{k=1}^n$   $\mathbf{b}_k \in \mathbb{R}^n$  skonstruowany według następującej zasady:

$$\begin{aligned}\mathbf{b}_1 &= [1, 0, \dots, 0]^T, \\ \mathbf{b}_2 &= \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0\right]^T, \\ \mathbf{b}_3 &= \left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, \dots, 0\right]^T, \\ \mathbf{b}_n &= \left[\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right]^T\end{aligned}$$

Warto zauważyć, że poszczególne współrzędne dobrane zostały w taki sposób, aby spełniać zależność  $\|\mathbf{b}_k\| = 1$ ,  $k = 1, \dots, n$ . W dalszej kolejności, rozważmy zmienną losową  $Y_k$ , powstałą z kombinacji liniowej wektora stabilnego (4.1) oraz poszczególnych wektorów  $\mathbf{b}_k$ , t.j.

$$Y_k = \mathbf{b}_k^T \mathbf{X}_\alpha. \quad (4.2)$$

Z twierdzenia 9, wynika, że każdy rozkład  $Y_k$  będzie rozkładem stabilnym  $S_\alpha(\sigma, \beta, \mu)$ . Biorąc pod uwagę, że poszczególne niezależne składowe wektora  $\mathbf{X}_\alpha$  posiadają rozkład  $S_\alpha(\sigma, 0, 0)$ , ze wzorów (2.9) w rozdziale 2.1 wynika natychmiast, iż parametry zmiennej losowej  $Y_k$  są równe  $\beta = 0$   $\mu = 0$ . Parametr skali zmiennej (4.2) wynosi z kolei:

$$\sigma_{\mathbf{b}_k} = \sigma k^{\frac{1}{\alpha} - \frac{1}{2}}. \quad (4.3)$$

Zmienna losowa (4.2) ma bardzo ciekawą właściwość, którą można wykorzystać do zobrazowania efektu rozkładu masy prawdopodobieństwa wektora stabilnego  $\mathbf{X}$  w zależności od wektorów kierunkowych  $\mathbf{b}_k$ . Mianowicie należy zauważyć, iż funkcja gęstości  $f(y_k)$  kombinacji liniowej (4.2), w punkcie  $y_k = 0$  jest równa całce z funkcji gęstości wektora (4.1), obliczonej po płaszczyźnie przechodzącej przez początek układu współrzędnych oraz prostopadłej do wektora  $\mathbf{b}_k$ :

$$f(y_k = 0) = \int_{\Omega} f(x_1, x_2, \dots, x_n) d\mathbf{x}, \quad (4.4)$$



gdzie  $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T \mathbf{b}_k = 0\}$ . Wartość (4.4) nie mówi nic o kształcie przekroju poprzecznego wielowymiarowej funkcji gęstości wektora (4.1), a jedynie obrazuje *jego rozmiar* w postaci całki. Oczywiście, aby wyznaczyć (4.4), nie istnieje konieczność obliczania wielowymiarowych całek. Możemy wykorzystać fakt, iż znamy rozkład stabilnej zmiennej losowej (4.2). Mimo, iż postać analityczna funkcji gęstości, jednowymiarowych rozkładów stabilnych, nie jest znana, to jej wartość w punkcie  $y_k = 0$  może zostać wyznaczona korzystając bezpośrednio z definicji funkcji charakterystycznej:

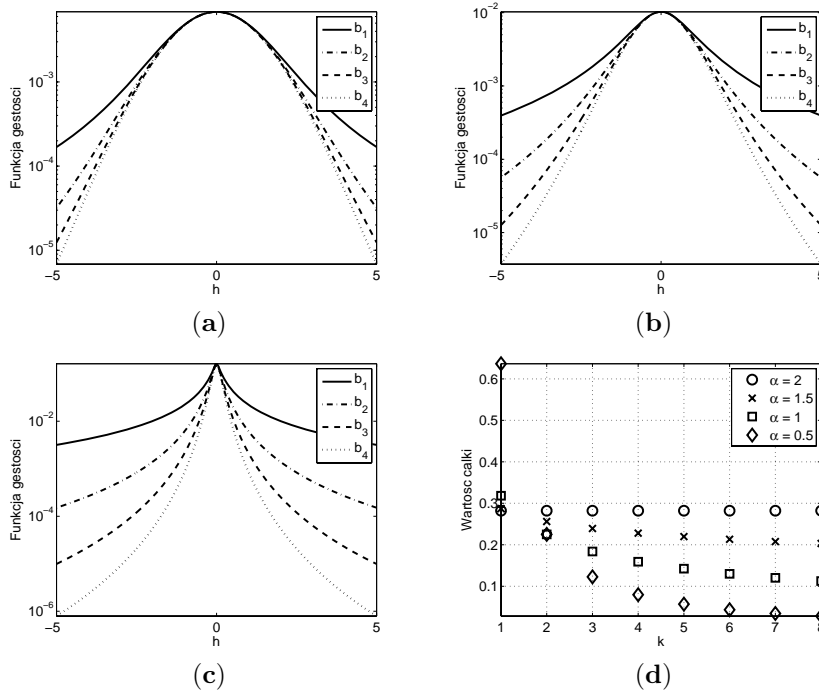
$$f(y; \alpha, \sigma, 0, 0) = \frac{1}{\pi} \int_0^{\infty} \exp\left(-(\sigma k)^\alpha\right) dk = \frac{\Gamma(1/\alpha)}{\sigma \alpha \pi} \quad (4.5)$$

Podstawiając (4.3) do (4.6) otrzymujemy:

$$f(y_k; \alpha, \sigma, 0, 0) = \frac{\Gamma(1/\alpha)}{\sigma k^{\frac{1}{\alpha} - \frac{1}{2}} \alpha \pi} \quad (4.6)$$

Na rysunku 4.1 - (d) przedstawiono wartości (4.6) dla ośmiowymiarowego wektora stabilnego (4.1) z parametrem  $\sigma = 1$ . Ponadto, na tym samym wykresie zamieszczono wykresy wartości czterowymiarowej funkcji gęstości wektora (4.1), wyznaczone na kierunkach  $\mathbf{b}_1 = [1, 0, 0, 0]^T$ ,  $\mathbf{b}_2 = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0]^T$ ,  $\mathbf{b}_3 = [\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0]^T$ ,  $\mathbf{b}_4 = [0.5, 0.5, 0.5, 0.5]^T$ .

**Obserwacja 3** *Poszczególne przekroje poprzeczne przedstawione na wykresie (4.1) wyraźnie obrazują sposób, w jaki poszczególne kierunki przestrzeni przeszukiwać, będą traktowane przez stabilny wektor losowy  $\mathbf{X}_\alpha$ . Mianowicie, należy się spodziewać, że wektor losowy  $\mathbf{X}_\alpha$  znacznie częściej będzie powodował zaburzenie rozwiązania wzdłuż kierunków wskazywanych przez osie układu współrzędnych.*



Rysunek 4.1: Wartości funkcji gęstości  $f(\mathbf{h}\mathbf{b}_k)$  stabilnego wektora losowego  $\mathbf{X}_\alpha \in \mathbb{R}^4$  dla: (a) -  $\alpha = 1.5$ , (b) -  $\alpha = 1$ , (c) -  $\alpha = 0.5$ . Wykres (d) przedstawia wartości  $\int_{-\infty}^{\infty} f(\mathbf{h}\mathbf{b}_k)dh$  dla  $n = 8$

Niesymetryczny sposób rozkładu masy prawdopodobieństwa musi w znacznej mierze wpływać na skuteczność algorytmów ewolucyjnych. Dalsza część rozdziału, poświęcona zostanie badaniom eksperymentalnym, których celem będzie bardziej szczegółowa analiza tego zjawiska.

#### 4.1.1 Badania eksperymentalne

W części eksperymentalnej, posłużymy się dwoma najczęściej spotykanymi środowiskami optymalizacyjnymi: funkcja Ackley'a -  $\phi_{14}$  oraz Rastrigin'a -  $\phi_{13}$  (dodatek H). Obydwa środowiska odznaczają się występowaniem wielu optimum lokalnych zlokalizowanych na siatce prostopadłej. W związku z tym, iż badania dotyczą czterowymiarowej funkcji gęstości  $n = 4$ , zdefiniujemy wektory kierunkowe:

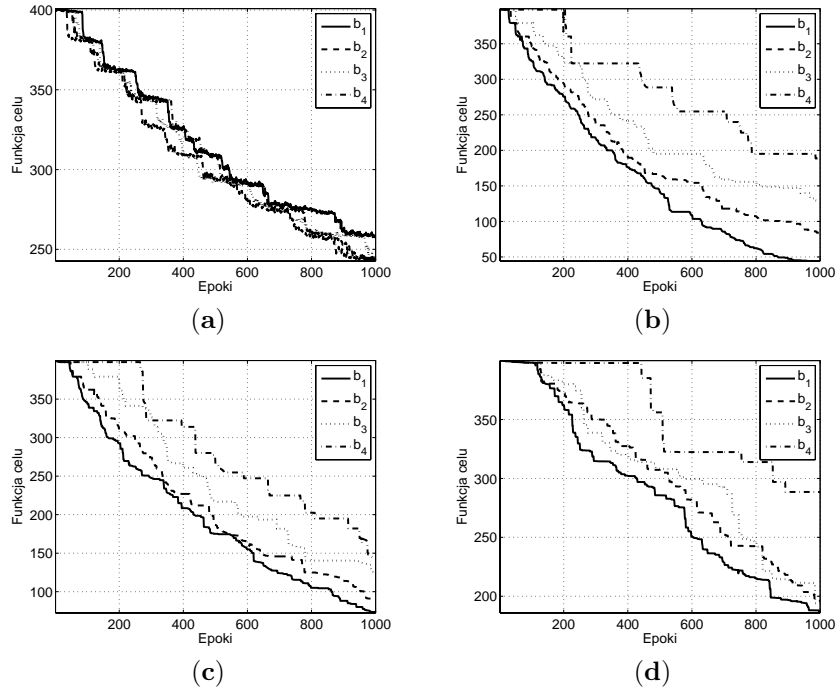
$$\begin{aligned} \mathbf{b}_1 &= [1, 0, 0, 0]^T, \\ \mathbf{b}_2 &= \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right]^T, \\ \mathbf{b}_3 &= \left[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0\right]^T, \\ \mathbf{b}_4 &= [0.5, 0.5, 0.5, 0.5]^T \end{aligned}$$

oraz odpowiadające im macierze obrotu:

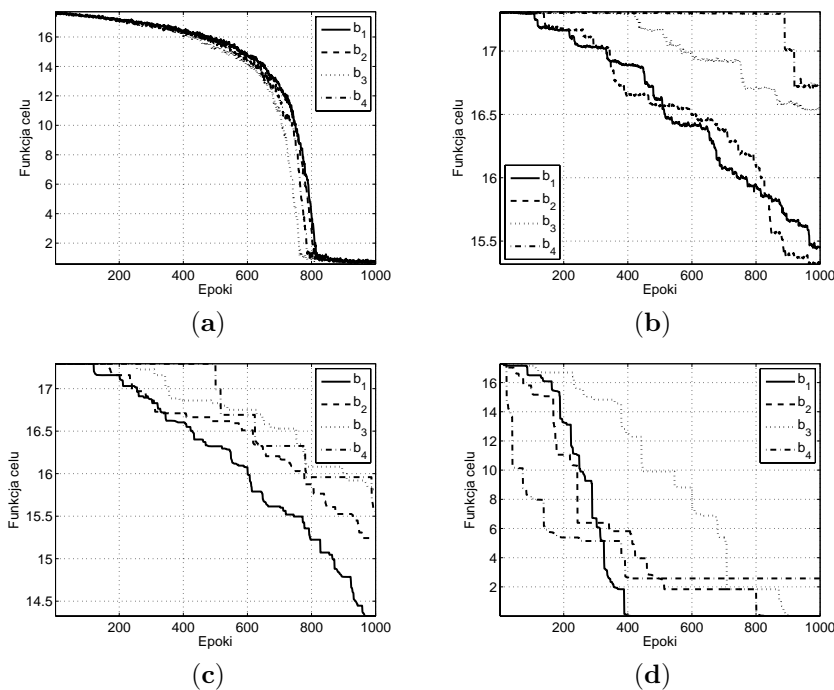
$$M_1 = I_4,$$

$$M_k = [I_4 - 2vv^T], \quad \frac{\mathbf{b}_1 - \mathbf{b}_k}{\|\mathbf{b}_1 - \mathbf{b}_k\|}, \quad k = 2, 3, 4.$$

Wspomniane funkcje testowe podane zostaną przekształceniu według zależności  $\phi(M_k \mathbf{x})$ ,  $k = 1, 2, 3, 4$ . W badaniach wykorzystano algorytm przeszukiwań ewolucyjnych z miękką selekcją [41]. Operator mutacji oparty został o wektor stabilny  $\sigma \mathbf{X}_\alpha$  z parametrem skali  $\sigma$  dobowanym w taki sposób, aby zagwarantować zbieżność algorytmu w trakcie  $T_{\max} = 1000$  pokoleń. Ponadto, pozostałe parametry algorytmu ustalone zostały w następujący sposób: rozmiar populacji  $\eta = 20$ , punkt startowy  $\mathbf{x}_0 = M_k [10, 10, 10, 10]^T$ , rozmiar grupy turniejowej  $P_G = 4$ . Na wykresach 4.2 - 4.3 przedstawiony został, proces ewolucji najlepszego osobnika w populacji, uśredniony po 50 niezależnych uruchomieniach algorytmu.



Rysunek 4.2: Proces ewolucji najlepszego rozwiązania dla funkcji Rastrigin'a ( $n = 4$ ). Algorytm przeszukiwań ewolucyjnych z miękką selekcją oraz stabilnym wektorem zaburzającym rozwiązania: (a) -  $\alpha = 2$ , (b) -  $\alpha = 1.5$ , (c) -  $\alpha = 1$ , (d) -  $\alpha = 0.5$



Rysunek 4.3: Proces ewolucji dla czterowymiarowej funkcji Ackley’*a*. Algorytm przeszukiwań ewolucyjnych z miękką selekcją oraz stabilnym wektorem zaburzającym rozwiązania: (a) -  $\alpha = 2$ , (b) -  $\alpha = 1.5$ , (c) -  $\alpha = 1$ , (d) -  $\alpha = 0.5$ .

**Obserwacja 4** Analizując wyniki eksperymentu należy zauważyć, że ortogonalne macierze obrotu, nie zmieniają postaci funkcji celu, a jedynie zmieniają wzajemne położenie najkorzystniejszych obszarów w przestrzeni rozwiązań oraz kierunków preferowanych przez operator mutacji. Wyraźnie widać, iż niekiedy nawet niewielkie zmiany, utożsamiane przez wektor  $\mathbf{b}_2$ , mogą spowodować znaczne pogorszenie uwarunkowania problemu optymalizacji ewolucyjnej. Znamienne jest, iż w przypadku rozkładu stabilnego  $\alpha = 2$  nie obserwuje się znaczących różnic w skuteczności algorytmu ewolucyjnego. Niewątpliwie świadczy to o sferycznej symetrii rozkładu eksploracyjnego, a co za tym idzie o sprawiedliwym traktowaniu wszystkich kierunków przeszukiwań.

**Obserwacja 5** W zależności od wielkości kąta obrotu oryginalnej przestrzeni przeszukiwań, warto również zwrócić uwagę, na przebieg samego procesu ewolucji. W sytuacji, gdy rozwiązania lokalne funkcji testowych, rozmieszczone są na prostopadłej siatce ( $\mathbf{b}_1$ ), uśredniona trajektoria najlepszego rozwiązania jest nacechowana znacznie mniejszymi skokami, niż ma to miejsce w przypadku ( $\mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4$ ). Oznacza to, iż proces ewolucji w dużej mierze polega na pokonywaniu sąsiadujących ze sobą szczytów ewolucyjnych, leżących na drodze do optimum globalnego. Staje się to możliwe, dzięki temu, iż szczyty ewolucyjne zlokalizowane są dokładnie na preferowanych kierunkach mutacji. Zupełnie inaczej proces ewolucji przebiega w przypadku obrotu przestrzeni zdefiniowanego przez wektor  $\mathbf{b}_4$ . Można zaobserwować, iż średnia trajektoria najlepszego rozwiązania, cechuje

*się stosunkowo długimi okresami w trakcie których rozwiązanie nie ulega poprawie. Po pewnym czasie jednak, następują gwałtowne zmiany wartości funkcji celu, co może świadczyć o wystąpieniu szczęśliwej makromutacji. Zjawiska te, występują jednak na tyle rzadko, iż spadek skuteczność algorytmu ewolucyjnego jest wyraźnie z zauważalny.*

#### 4.1.2 Podsumowanie

Najczęściej nasuwająca się konstrukcja wektora losowego, polegająca na utożsamieniu każdej współrzędnej z niezależną zmienną losową o identycznym rozkładzie stabilnym, może w praktyce przyczynić się do powstania poważnych problemów. Jakkolwiek rozwiązanie takie jest powszechnie stosowane w algorytmach ewolucyjnych z rozkładem normalnym, należy być ostrożnym przy próbie przełożenia tej idei na rozkłady stabilne. Główny problem, tzw. efekt symetrii [85], przejawia się brakiem sferycznej symetrii wektora zaburzającego rozwiązania. Oznacza, to, że wielowymiarowe rozkłady stabilne, wykorzystujące wspomnianą konstrukcję, w sposób uprzywilejowany, traktują kierunki leżące wzdłuż osi układu współrzędnych. Tym samym, można zaobserwować spadek skuteczności algorytmu ewolucyjnego w sytuacji, gdy obszary przestrzeni przeszukiwań o lepszym dopasowaniu nie będą zorientowane zgodnie ze wspomnianymi kierunkami.

W niniejszym rozdziale zamieszczono szczegółową analizę opisywanego zjawiska, wraz z badaniami eksperymentalnymi przeprowadzonymi dla dwóch popularnych funkcji testowych. Wnioski, wyciągnięte na podstawie otrzymanych wyników, jednoznacznie wskazują na konieczność uwzględnienia efektu symetrii, przy okazji projektowania wydajnych i uniwersalnych operatorów mutacji algorytmów ewolucyjnych. Prób złagodzenia skutków jakie wywołuje efektu symetrii, należy upatrywać w dwóch podejściach: zastosowanie izotropowej mutacji stabilnej lub procedury adaptacyjnej, niesymetrycznego operatora mutacji. Zagadnieniom tym, poświęcone zostały kolejne rozdziały rozprawy.

## 4.2 Mutacja o sferycznej symetrii

Rozkłady izotropowe odgrywają szczególną rolę w algorytmach optymalizacji globalnej [10, 20, 88, 104]. W przypadku nieograniczonej przestrzeni przeszukiwań  $\mathbb{R}^n$ , można doszukać się pewnej analogii pomiędzy nimi, a rozkładem równomiernym określonym na ograniczonym zbiorze rozwiązań dopuszczalnych. Ten ostatni nie wyróżnia żadnego spośród podzbiorów przestrzeni przeszukiwań, natomiast rozkłady o sferycznej symetrii, nie faworyzują żadnego z kierunków przeszukiwań. Z tego względu, ich stosowanie, jest szczególnie uzasadnione na początku procesu optymalizacji, gdzie nie posiadamy żadnej wiedzy o optymalizowanym problemie.

### 4.2.1 Eksploracja a eksploatacja - kompromis według mutacji stabilnej

Biorąc pod uwagę przestrzeń wszystkich możliwych funkcji  $\Omega = \{\phi(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}\}$ , problem optymalizacji doczekał się satysfakcjonującego rozwiązania, jedy-

nie dla niezwykle wąskiej podklasy zbioru  $\Omega$ . Warunki konieczne i wystarczające na zaklasyfikowanie punktu jako optimum lokalnego, pozwoliły na skonstruowanie wydajnych algorytmów służących do optymalizacji unimodalnych funkcji gładkich [81]. Pojęcie optimum lokalnego, odnosi się jedynie do bezpośredniego sąsiedztwa rozważanego rozwiązania, które z matematycznego punktu widzenia, może być nieskończenie małe. Oznacza to, że nawet jeśli procedura testowa oparta o wspomniane wyżej warunki, wskazywać będzie na istnienie rozwiązania lokalnego, to w dalszym ciągu nie możemy powiedzieć nic o jakości rozwiązań zlokalizowanych poza pewnym, bezpośrednio przylegającym do niego obszarem. Warto w tym miejscu nadmienić, że rozmiar tego obszaru, w przypadku funkcji różniczkowalnych, jest uwarunkowany przez tzw. stałą Lipschitza [81].

Ogólnie, największej przeszkody na drodze zaprojektowania uniwersalnego algorytmu optymalizacji globalnej, należy upatrywać w trudności polegającej na połączeniu dwóch wzajemnie wykluczających się zadań [46]: eksploracji oraz eksploatacji przestrzeni rozwiązań. Precyzyjna definicja tych dwóch pojęć, jak również ich wzajemny wpływ na skuteczność stochastycznych metod optymalizacji jest sprawą dyskusyjną. Wraz z rozwojem komputerowych technik optymalizacji, zagadnienie to, było szeroko dyskutowane zarówno w literaturze poświęconej metodom optymalizacji [46], [55], [108], [119], [51] jak i w znacznie szerszym kontekście adaptacyjnych systemów uczących [110], [54], [43]. Istota niniejszego problemu polega na tym, że z jednej strony algorytm powinien znajdować rozwiązanie z możliwie dużą dokładnością, co a ogół wiąże się z koniecznością przeszukiwania bardzo wąskiego obszaru wokół rozwiązania. Z drugiej zaś strony, nie ma żadnej pewności, że znalezione optimum jest ekstremum globalnym, co pociąga za sobą konieczność kontynuowania procedury przeszukiwania, w innych obszarach przestrzeni rozwiązań. Niezbędnym wydaje się zatem pogodzenie, w obrębie jednej procedury optymalizacyjnej, dwóch wzajemnie wykluczających się celów: szybkiego zlokalizowania rozwiązań oraz ich skutecznego omijania.

### **Eksploatacja przestrzeni rozwiązań - zbieżność lokalna**

Na pierwszy rzut oka, analiza stochastycznych metod optymalizacji pod kątem ich zdolności do rozwiązywania problemów unimodalnych wydaje się mieć małe znaczenie. Istnieją bowiem, algorytmy deterministyczne, które odznaczają się dużą efektywnością w odnajdywaniu rozwiązań lokalnych [81]. Należy przy tym jednak zaznaczyć, że znakomita większość wspomnianych technik, ogranicza się do funkcji spełniających pewne, niejednokrotnie zbyt wygórowane założenia np. ciągłą różniczkowalność do rzędu 2 [81] lub znajomość stałej Lipschitza [119]. Co więcej, w przypadku niespełnienia przez funkcję celu, któregośkolwiek z niezbędnych założeń, można najczęściej zaobserwować rozbieżność algorytmów oraz utratę ich stabilności numerycznej. Algorytmy stochastyczne, pod tym względem, mogą być postrzegane jako rozwiązania w dużym stopniu odporne tzn. nie wykazują znacznego spadku skuteczności, nawet w przypadku nie spełnienia przez funkcję celu, któregośkolwiek z pierwotnych założeń. Dlatego też, analiza zbieżności lokalnej może w dużym stopniu przyczynić się do lepszego zrozumienia natury działania algorytmów ewolucyjnych, projektowania procedur adaptacyjnych oraz strojenia zewnętrznych parametrów sterujących. Oceniając efektywność algorytmów stochastycznych, w klasie funkcji unimodalnych, należy wziąć pod uwagę przynajmniej dwa czynniki: nakład obliczeniowy potrzebny do zlokalizowania rozwiązania oraz jakość otrzymywanych rozwiązań.

Kwestia oceny skuteczności algorytmu ewolucyjnego może jedynie być dyskusowana przy uwzględnieniu odpowiednich warunków stopu, będących integralną częścią każdej techniki optymalizacji. Dlatego też, nakład obliczeniowy, jest zazwyczaj utożsamiany z liczbą obliczeń funkcji celu, niezbędnych do zatrzymania pracy algorytmu. Mimo iż dla wielu popularnych algorytmów ewolucyjnych istnieją teoretyczne dowody stochastycznej zbieżności do rozwiązania globalnego, fakt ten rzadko kiedy znajduje swoje odzwierciedlenie w ich komputerowej implementacji. Ograniczenia narzucane przez skończoną długość słowa maszynowego, powodują znaczne różnice pomiędzy zachowaniem obserwowanym, a tym wynikającym z teoretycznego opisu. Problem ten staje się szczególnie widoczny w przypadku generatorów sekwencji pseudolosowych. Jak wiadomo, niemożliwe jest na przykład otrzymanie liczb pseudolosowych z teoretycznego rozkładu normalnego, przekraczających wartości  $5\sigma$ , choć oczywiście, teoria pozwala na otrzymanie liczb ze wspomnianego zakresu [10]. Powyżej naszkicowany problem skłania projektantów procedur optymalizacyjnych, do wprowadzenia dodatkowych zabezpieczeń, wymuszających zatrzymanie algorytmu. Najczęściej kryteria stopu oparte są na maksymalnej liczbie obliczeń funkcji celu lub na zbiorze reguł heurystycznych mających na celu wykrycie stagnacji algorytmu [10, 85]<sup>1</sup>. W świetle przytoczonych faktów, dwa wspomniane kryteria oceny algorytmów w problemach unimodalnych, nie powinny budzić kontrowersji.

Ogólnie analityczna analiza zagadnień związanych ze zbieżnością lokalną algorytmów ewolucyjnych nie jest zadaniem trywialnym. Zadanie to, w znacznej mierze utrudnia fakt, iż rezultatem działania algorytmu ewolucyjnego jest pewien stochastyczny rozkład wynikowych wartości funkcji celu. Ponadto, rozkład ten, ze względu na stopień złożoności, nieliniowość oraz niestacjonarność procedur stochastycznych, najczęściej pozostaje nieznany.

Problem jakości otrzymywanych rozwiązań, otrzymywanych za pomocą algorytmu Przeszukiwań Ewolucyjnych z Miękką Selekcją, został poruszony w [56]. Korzystając z modelu ewolucji opartym na dwuelementowej populacji rozwiązań, Iwona Karcz-Dulęba sprowadziła analizę algorytmu do analizy pewnego systemu dynamicznego, opisanego w przestrzeni stanów. W [56] wykazała, że w przypadku ewolucji nadzorowanej przez rozkład normalny ze stałym odchyleniem standardowym, populacja rozwiązań nie jest zbieżna do rozwiązania globalnego, lecz osiąga punkt równowagi selekcyjno-mutacyjnej w pewnej odległości od optimum globalnego. Odległość ta, warunkuje oczywiście jakość rozwiązania, co w bezpośredni sposób nawiązuje do wspomnianych na początku rozdziału kryteriów oceny.

Porównanie szybkości z jaką algorytm zbiega do rozwiązania, można odleźć w [135, 136]. Przeprowadzone badania symulacyjne dotyczyły nieizotropowych wersji rozkładów normalnego oraz Cauchy'ego, wzbogaconych o mechanizm samoadaptacji parametrów skali. Wyniki otrzymane na podstawie siedmiu wybranych funkcji unimodalnych skłoniły autorów do wskazania na rozkład Cauchy'ego, jako na ten który wykazuje znacznie lepszą poprawę rozwiązania na początku procesu optymalizacji. Z mutacją opartą o rozkład Cauchy'ego, nieodzownie wiąże się jednak następująca wada: po osiągnięciu pewnej jakości rozwiązania, następuje stagnacja algorytmu z znacznej odległości od rozwiązania [135]. Z drugiej zaś strony, strategia ewolucyjna z Gaussowskim rozkładem eks-

<sup>1</sup>sytuacji, w której algorytm przez określony czas nie powoduje poprawy najlepszego znalezionego do tej pory rozwiązania

ploracyjny, odznacza się znacznie szybszą zbieżnością w sytuacjach, gdy rozwiązanie problemu, wymaga stosunkowo niewielkich zaburzeń rozwiązań. Generalnie, analiza wyników badań eksperymentalnych [135, 136] pozwoliła na sformułowanie hipotezy, iż w przypadku problemów optymalizacji lokalnej, posługiwanie się rozkładem gaussowskim daje nieco lepsze rezultaty [135].

Problem zbieżności lokalnej algorytmów ewolucyjnych z mutacjami Cauchy'ego oraz Gaussa rozważany był również w [85]. Przeprowadzone badania eksperymentalne skupiał się głównie na analizie przebiegu procesu ewolucji warunkowanej przez izotropowe jak i nieizotropowe wersje obu rozkładów. Prezentowane wyniki [85] pozwoliły na sformułowanie dwóch najważniejszych czynników, mających bezpośredni wpływ na skuteczność optymalizacyjną: tzw. efektu otoczenia (ang. *surrounding effect*) oraz efektu symetrii (ang. *symmetry effect*). Pierwszy z nich może być postrzegany jako ewolucyjna wersja tzw. kłątwy wielowymiarowości [119] i staje się szczególnie widoczny w przypadku problemów optymalizacyjnych o znacznej liczbie zmiennych decyzyjnych. Efekt otoczenia, który polega na zmianie rozkładu normy wektora zaburzającego, posłużył do wytłumaczenia obserwowanego pogorszenia skuteczności procedur stochastycznych w miarę wzrostu rozmiaru problemu. Efekt otoczenia jest szerzej komentowany w rozdziale 4.2.2.

Rozkłady stabilne posiadają wszystkie cechy niezbędne do tego, aby zagwarantować wysoką wydajność algorytmów ewolucyjnych w zadaniach optymalizacji lokalnej. Najbardziej przekonującego dowodu dostarczają funkcje gęstości zaprezentowane na wykresie 2.1-(a) (rozdział 2.1). Mianowicie rozkłady ciężkoogonowe wypiętrzają funkcję gęstości w pobliżu punktu bazowego, co świadczy o znacznie intensywniejszym próbkowaniu bezpośredniego otoczenia rodzica. Pamiętając o tym, iż efektywność algorytmu ewolucyjnego w zadaniach optymalizacji lokalnej, uwarunkowana jest przez częste, skuteczne mutacje, których prawdopodobieństwo wystąpienia jest znacznie większe w pobliżu rodzica, korzyści ze stosowania rozkładów stabilnych wydają się nader oczywiste. W celu weryfikacji powyższych przypuszczeń, w następnym punkcie przeprowadzono serię eksperymentów symulacyjnych, które wyniki mają posłużyć do oceny efektywności klasy izotropowych rozkładów stabilnych.

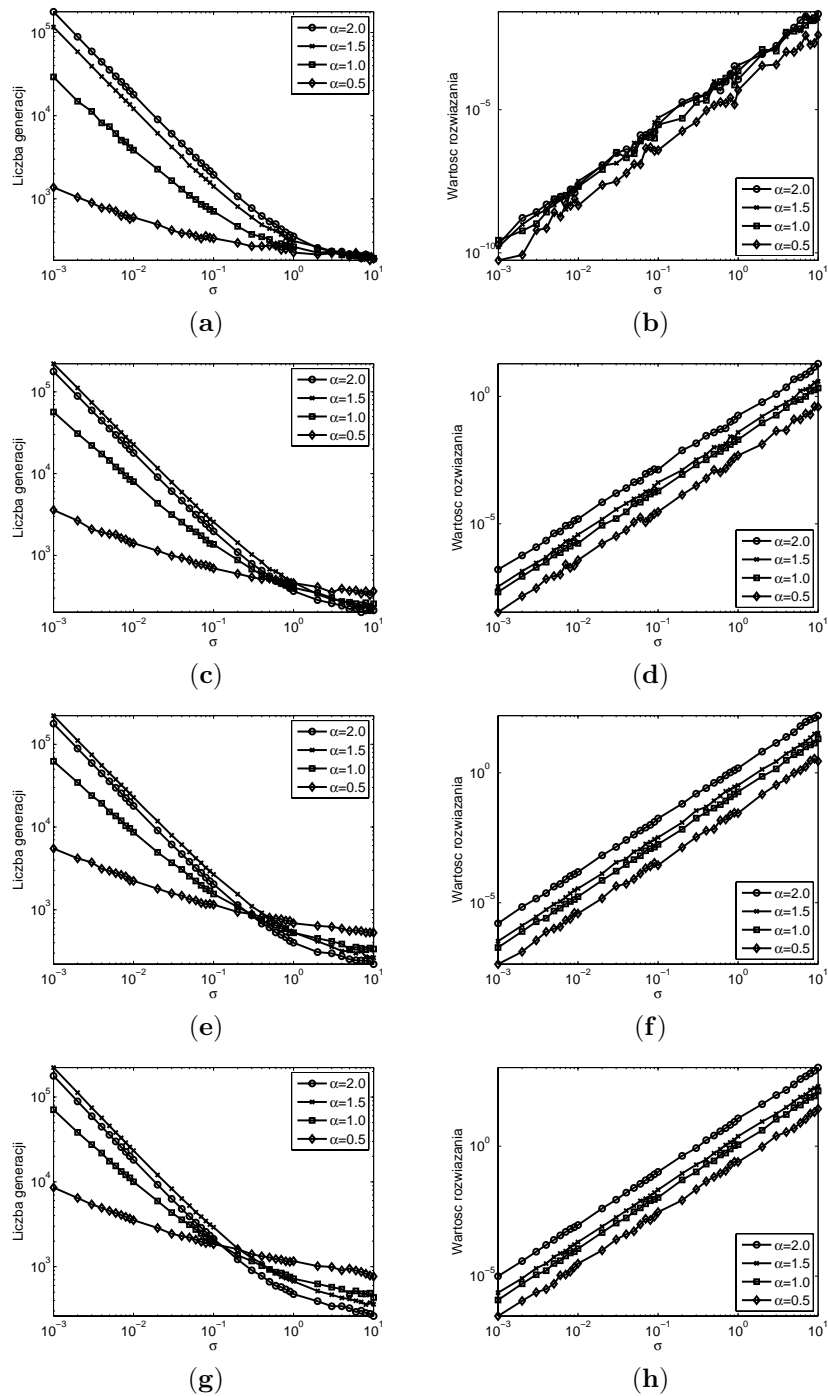
## Badania symulacyjne

W celu bliższego scharakteryzowania wpływu rozkładów  $\alpha$ -stabilnych na zbieżność lokalną algorytmu ewolucyjnego, posłużymy się następującym scenariuszem. Rozważmy zadanie minimalizacji funkcji postaci  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ , gdzie  $\mathbf{x} \in \mathbb{R}^n$  oraz załóżmy, że dane jest rozwiązanie początkowe  $(\mathbf{x}_0)_i = \frac{100}{\sqrt{n}}$ ,  $i = 1, \dots, n$ . Problem ten spróbujemy rozwiązać za pomocą najprostszej strategii ewolucyjnej (1+1)ES wyposażonej w operator mutacji zaburzający rozwiązania z izotropowym rozkładem stabilnym t.j.  $\mathbf{X}_\alpha^s$  (rozdział 2.2.1). Przestrzeń parametrów  $(0, 2] \times \mathbb{R}_+$ , rozważanych rozkładów stabilnych, została zdyskretyzowana, w taki sposób, aby badania zostały przeprowadzone dla indeksu stabilności  $\alpha = 2.0, 1.5, 1.0, 0.5$ , oraz skali  $\sigma = 0.001, 0.002, \dots, 0.009, 0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 0.9, 1, 2, \dots, 10$ . Celem badań była analiza szybkości zbieżności oraz jakości otrzymywanych rozwiązań, dla każdej pary wymienionych wyżej parametrów. W każdym przypadku, zastosowano kryterium stopu, według którego, algorytm był zatrzymywany jeśli w przeciągu stu następujących po sobie generacji, rozwiązanie nie uległo zmianie. Pamiętając o tym, iż wyniki uzyskiwane przez algorytm stochastyczny,



mają w istocie charakter losowy, zdecydowano się na przeprowadzenie  $N_t = 100$  niezależnych uruchomień strategii  $(1 + 1)ES_\alpha$ , dla każdego zestawu parametrów  $(\alpha_i, \sigma_j)$ ,  $i = 1, \dots, 4$ ,  $j = 1, \dots, 37$ , oraz dla różnych wymiarów przestrzeni przeszukiwań  $n = 1, 3, 6, 12$ . Tym samym przeprowadzenie niniejszego eksperymentu wymagało 59200 niezależnych uruchomień algorytmu.

Otrzymane wyniki w postaci średniej liczby iteracji oraz średniej wartości rozwiązania zamieszczone zostały na rysunku 4.4.



Rysunek 4.4: (a), (c), (e), (g) - średnia liczba epok potrzebna do zlokalizowania rozwiązania, odpowiednio dla wymiaru przestrzeni przeszukiwań  $n = 1, 3, 6, 12$ . (b), (d), (f), (h) - średnia wartość funkcji celu otrzymywana przez algorytm ewolucyjny z mutacją  $\alpha$ -stabilną dla wymiarów przestrzeni  $n = 1, 3, 6, 12$

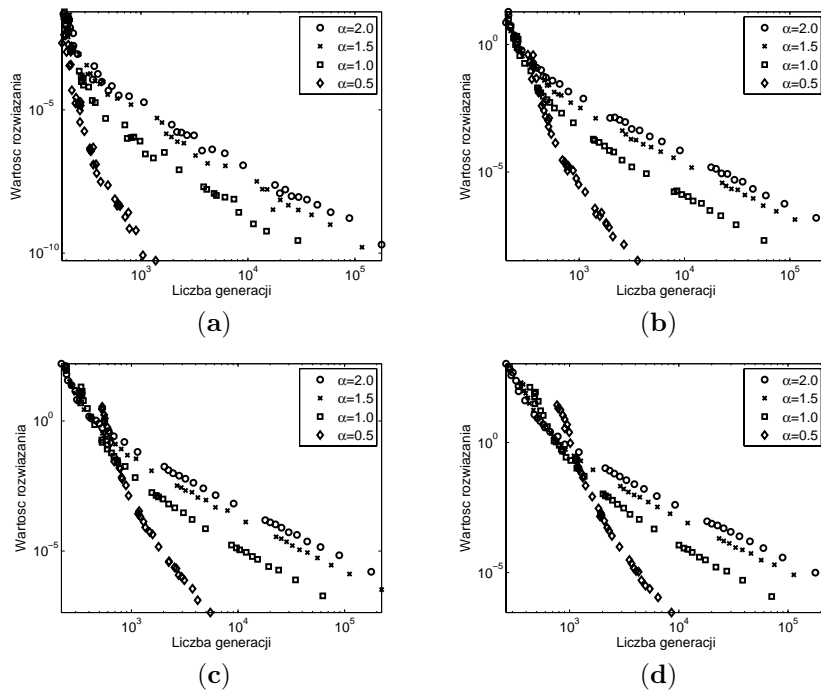
**Obserwacja 6** Bazując na wynikach przedstawionych na rysunku 4.4 znamienym jest, że dla przypadków, w których otrzymano zadowalającą jakość rozwiązania (na ogół  $\sigma < 0.1$ ),<sup>2</sup> można zaobserwować następującą regularność: dla identycznych parametrów skali, średnia precyzja, z jaką algorytm ewolucyjny lokalizuje rozwiązanie, jest większa dla mutacji zachodzących z rozkładami o mniejszym indeksie stabilności. Znacznie więcej dowodów na przewagę rozkładów o cięższych ogonach dostarczają wykresy przedstawiające średnią liczbę epok niezbędną do osiągnięcia wspomnianych rezultatów (wykresy (b), (d), (f), (h) - rysunek 4.4). Mianowicie, analizując rezultaty otrzymywane dla strategii ewolucyjnej z mutacją  $\alpha = 0.5$ , zdumiewające jest, że pozwala ona na osiągnięcie lepszych wyników, w czasie wielokrotnie krótszym od innych modeli probabilistycznych.

**Obserwacja 7** Należy również zwrócić uwagę na regularność polegającą na pogarszaniu się jakości otrzymywanych rozwiązań, w miarę zwiększania wymiaru przestrzeni przeszukiwań. Obserwacja ta, dotyczy wszystkich mutacji izotropowych bazujących na klasie rozkładów stabilnych. Zjawisko to, określone jako efekt otoczenia, po raz pierwszy, w systematyczny sposób, zostało scharakteryzowane dla rozkładu normalnego oraz Cauchy'ego w [85]. Efekt otoczenia, ma swoje źródło, w zmianach rozkładu normy wektora losowego stosowanego do modyfikacji rozwiązań w algorytmie ewolucyjnym i jest szerzej dyskutowany w rozdziale 4.2.2.

**Obserwacja 8** Kolejna, niezwykle ważna obserwacja, jaką można poczynić analizując wyniki przedstawione na rysunku 4.4, dotyczy zagadnienia odporności strategii  $(1+1)ES_\alpha$  na dobór parametru  $\sigma$ . Algorytm, który mimo znacznych odchyżeń od optymalnej konfiguracji, w dalszym ciągu mógłby służyć do otrzymywania akceptowalnych rozwiązań, jest z praktycznego punktu widzenia, szczególnie pożądany. Z doświadczeń zdobytych przy rozwiązywaniu szeregu problemów natury inżynierskiej, wynika, że najczęściej strojenie algorytmu optymalizacji, dokonuje się metodą prób i błędów. Biorąc pod uwagę wspomniany scenariusz oraz analizując wrażliwość parametru skali  $\sigma$  na jakość otrzymywanych rozwiązań (rysunek 4.4), jasnym jest, że zastosowanie rozkładów o mniejszym indeksie stabilności, umożliwiłoby zwiększenie szans na szybsze odnalezienie optymalnej konfiguracji.

Biorąc pod uwagę problem wyboru najskuteczniejszej strategii ewolucyjnej w zadaniach optymalizacji lokalnej, zupełnie naturalnym jest, że za najbardziej pożądany rozkład eksploracyjny należałoby uznać ten, który gwarantuje najlepszą jakość rozwiązania i jednocześnie wymaga niewielkiej liczby nakładów obliczeniowych do osiągnięcia zamierzonych rezultatów. W ten sposób, efektywność izotropowych rozkładów stabilnych może być oceniana na podstawie dwukryterialnego wskaźnika jakości:  $J(\alpha, \sigma) = [J_1(\alpha, \sigma), J_2(\alpha, \sigma)]^T$ , gdzie  $J_1(\alpha, \sigma)$  jest jakością rozwiązania, uzyskaną dla rozkładu stabilnego o parametrach  $\alpha, \sigma$  oraz  $J_2(\alpha, \sigma)$  jest liczbą generacji niezbędną do uzyskania najlepszego rozwiązania dla tej samej konfiguracji. W ten sposób, pytanie o najbardziej skuteczny izotropowy rozkład stabilny może zostać rozstrzygnięte w oparciu o pojęcie Pareto-optymalności [66]. Na rysunku 4.5 możemy zaobserwować układ obu wskaźników jakości zamieszczonych na wspólnym wykresie.

<sup>2</sup>przyjmując, że dla rozważanego zadania optymalizacji, rozwiązanie jest jest wyznaczone z akceptowalną dokładnością jeśli wartość funkcji celu  $\phi(\mathbf{x}^*) < 0.01$



Rysunek 4.5: Wykresy przedstawiające średnią jakość rozwiązania otrzymywanego przez strategie ewolucyjne  $(1+1)ES_\alpha$  w zależności od średniej liczby pokoleń niezbędnej do spełnienia założonego kryterium stopu. Poszczególne punkty na wykresach odpowiadają malejącym parametrom skali  $\sigma$  (od strony lewej do prawej)

**Obserwacja 9** W przypadku wyników, które mogą być uważane za zadowalające (przypis w obserwacji 6), znamienne jest, że rozkłady o cięższych ogonach mają tendencję do dominowania nad rozkładami o większym indeksie stabilności  $\alpha$ . Pojęcie dominacji w sensie Pareto, w odniesieniu do rozważanej strategii ewolucyjnej oznacza, że zapewniają one zarówno lepszą jakość rozwiązania jak i mniejsze nakłady obliczeniowe niezbędne do jego osiągnięcia. Regularność tą potwierdzają wyniki zaprezentowane na rysunku 4.5, gdzie większą część frontu Pareto zajmują rozwiązania odpowiadające  $\alpha = 0.5$ .

### Eksploracja przestrzeni rozwiązań - zbieżność globalna

Przeprowadzona w poprzednim rozdziale ocena rozkładów eksploracyjnych, w kontekście zadań optymalizacji lokalnej, dostarcza bogatej wiedzy na temat sposobu eksploatacji przestrzeni rozwiązań. Pozwala między innymi na, zaobserwowanie wzajemnych zależności występujących w przestrzeni parametrów izotropowych rozkładów stabilnych oraz daje ogólne pojęcie na temat jak zależności te przekładają się na skuteczność algorytmu ewolucyjnego. Niemniej jednak analiza zbieżności lokalnej, nie pozwala na wyciąganie wiarygodnych wniosków dotyczących procesu ewolucji w bardziej złożonych środowiskach optymalizacyjnych. Ponadto, intuicyjnie uzasadnionym wydaje się pogląd mówiący o tym,

że algorytmy o najlepszych właściwościach eksploatacyjnych zwykle nie dają zadowalających rezultatów, w zadaniach optymalizacji globalnej. Zachłanność w eksploatacji najbliższego otoczenia, która cechuje wspomniane algorytmy, przejawia się szybkim ukierunkowaniem procesu ewolucji odsuwając niejako na bok, eksplorację jej niezbadanych dotychczas regionów. Strategia taka, jest jak najbardziej uzasadniona w zadaniach optymalizacji lokalnej, lecz w przypadku środowiska multimodalnego, może powodować przede wszystkim przedwczesną zbieżność algorytmu lub problemy związane z opuszczeniem basenów przyciągania optimum lokalnych.

Pojęcie eksploracji przestrzeni rozwiązań może być postrzegane jako zdolności do generowania rozwiązań, które w dużej mierze różnią się od tych już znanych. Cecha ta, jest niezwykle istotna z punktu widzenia zadań jakie stoją przed efektywnym algorytmem optymalizacji globalnej. Na podstawie badań prowadzonych na przestrzeni ostatnich lat, zaobserwowano, że zdolności eksploracyjne algorytmów ewolucyjnych uwarunkowane są przez przynajmniej dwa mechanizmy. Pierwszym z nich jest możliwość wystąpienia tzw. makromutacji, czyli uzyskiwania na drodze mutacji potomków o strukturze znacznie różniącej się od osobnika rodzicielskiego [85, 135]. Drugi mechanizm, który umożliwia eksplorację przestrzeni rozwiązań, opiera się na pojęciu dryftu populacji [42]. Dryft obserwowalny jest tylko w przypadku nielicznych populacji, i opiera się przypadkowym zaburzeniu naturalnego kierunku ewolucji. Zjawisko to może zajść tylko w przypadku algorytmów ewolucyjnych, które dopuszczają możliwość formowania populacji przyszłych rodziców, z osobników o gorszej jakości dopasowania - tzw. miękka selekcja [41, 85]. Wystąpienie stosunkowo małoprawdopodobnych zdarzenia podczas selekcji, może spowodować zmianę kierunku ewolucji w stosunku do jej naturalnego przebiegu. Oznacza, to że najlepsze rozwiązania, typowane przez algorytm ewolucyjny, mogą ulec pogorszeniu. Ostatecznie, proces ten nie może utrzymywać się przez długi okres czasu, po którym, ewolucji przywracany jest jej naturalny bieg. Niemniej jednak czas w którym obserwowany jest dryft populacji, może sprawić, że algorytm uwolni się z basenu przyciągania optimum lokalnego [43], a tym samym zapoczątkuje eksploatację nie znanych dotychczas obszarów przestrzeni rozwiązań.

Kolejny punkt niniejszego rozdziału poświęcony został, weryfikacji zdolności eksploracyjnych algorytmu ewolucyjnego, bazującego na klasie izotropowych rozkładów stabilnych.

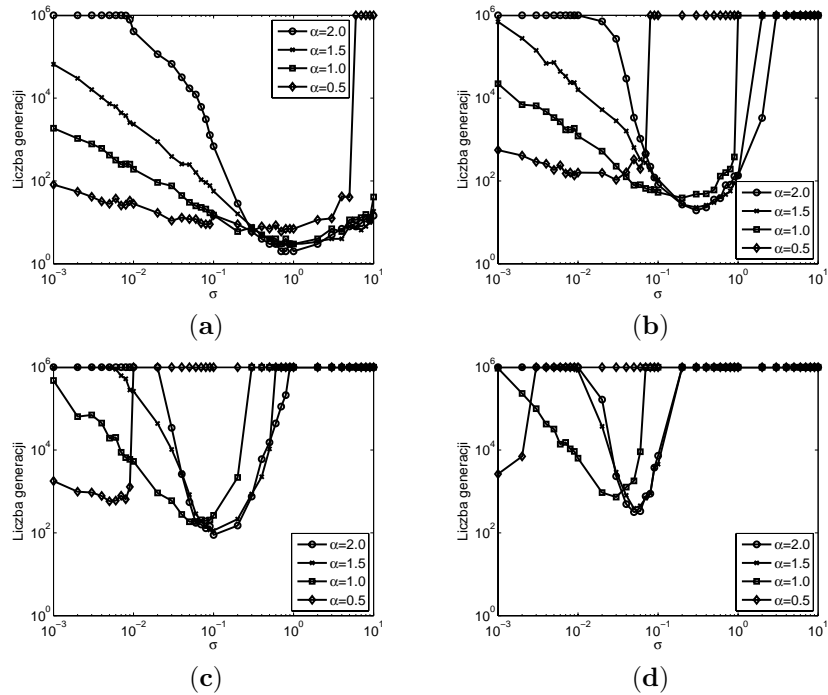
## Badania symulacyjne

W rozdziale tym, posłużymy się bardzo prostym modelem ewolucji w postaci strategii ewolucyjnej  $(1,2)ES_\alpha$  do eksperymentalnej analizy zdolności eksploracyjnych izotropowej mutacji stabilnej. Wybór algorytmu ewolucyjnego pozostającego przedmiotem zainteresowania niniejszych badań, nie jest przypadkowy i poza mechanizmem makromutacji, właściwym dla rozkładów ciężkoogonowych, ma umożliwić również ewentualne wystąpienie wspomnianego wcześniej dryftu populacji. Jako, że zdolności eksploracyjne są zazwyczaj tożsame ze zdolnościami omijania przez algorytm rozwiązań lokalnych, eksperyment opiera się na problemie przekraczania siodła pomiędzy dwoma wzniesieniami gaussowskimi. Mianowicie, cel jaki postawiono przed algorytmem  $(1,2)ES_\alpha$  to maksymalizacja

funkcji celu o postaci:

$$\phi(\mathbf{x}) = \frac{1}{2} \exp(-5\|\mathbf{x}\|^2) + \exp(-5\|\mathbf{x} - \mathbf{e}_1\|^2), \quad (4.7)$$

gdzie  $\mathbf{e}_1 = [1, 0, 0, \dots, 0]$ . Pierwszy osobnik rodzicielski umiejscawiany jest na szczycie wzgórza o gorszym dopasowaniu. Tym samym, celem każdego algorytmu było zlokalizowanie obszaru wewnątrz którego jakość rozwiązania przekracza wartość 0.55, co jest tożsamy z przekroczeniem punktu siodłowego pomiędzy dwoma optimami lokalnymi. Sformułowane na tej podstawie kryterium stopu, posłużyło do obserwacji liczby obliczeń funkcji celu niezbędnej do opuszczenia przez algorytm basenu przyciągania gorszego z rozwiązań lokalnych. Przestrzeń parametrów rozkładów izotropowych została zdyskretyzowana identycznie, jak to miało miejsce w przypadku eksperymentu opisanego w poprzednim punkcie. Wyniki w formie średniej liczby epok niezbędnych do przekroczenia siodła, uzyskane na podstawie stu niezależnych uruchomień algorytmu, dla każdego zestawu parametrów  $\{\alpha_i, \sigma_j\}$ , zaprezentowane zostały na rysunku 4.6.



Rysunek 4.6: Liczba generacji niezbędna do przekroczenia siodła pomiędzy dwoma wzgórzami gaussowskimi, dla różnych wymiarów przestrzeni przeszukiwań: (a) - 1, (b) - 3, (c) - 6, (d) - 12.

**Obserwacja 10** Obserwując średnią liczbę epok potrzebną do przekroczenia punktu siodłowego, dla różnych mutacji stabilnych, rzeczą która najbardziej przykuwa uwagę jest wrażliwość wyników na parametr skali  $\sigma$ . Generalnie można wysnuć wniosek, iż skuteczność algorytmów wykorzystujących rozkład eksploracyjny o

mniejszym indeksie stabilności, jest w mniejszym stopniu zależna od właściwego doboru skali rozkładu. Powyższy wniosek uwidocznia się szczególnie wyraźnie dla niewielkiej liczby wymiarów przestrzeni przeszukiwań. Szerokość okna ewolucyjnego, formowanego przez wartości skali rozkładu  $\sigma$ , dla których strategia  $(1,2)ES_\alpha$  pokonuje punkt siodłowy, zmienia się w zależności od wykładnika stabilności  $\alpha$ . W rezultatach prezentowanych na rysunkach 4.6 (a)-(b) można doszukać się prawidłowości polegającej na, że rozmiar okna ewolucyjnego zmniejsza się wraz ze wzrostem  $\alpha$ . W sytuacji, gdy należy skonfigurować algorytm optymalizacyjny nie dysponując żadną wiedzą na temat ewentualnych szerokości siodła ewolucyjnych, obserwacja ta może przemawiać na korzyść rozkładów z cięższymi ogonami. Obserwując zmiany w skuteczności strategii ewolucyjnej jakie zachodzą pod wpływem zwiększania wymiarowości przestrzeni przeszukiwań, wyraźnie widać, że okno ewolucyjne dla wszystkich rozkładów stabilnych ulega zmniejszeniu. Efekt ten, jest specyficznym przejawem tzw. kłutwy wymiarowości, i musi być widoczny dla wszystkich rozkładów izotropowych. Szczegółowe omówienie obserwowalnej prawidłowości, wraz ze wskazaniem jej źródeł, znajduje się w rozdziale 5.1.1.

Godnym podkreślenia jest fakt, iż wszystkie rozkłady uwzględnione w eksperymencie, mogą zagwarantować bardzo zbliżoną skuteczność, pod warunkiem odpowiedniego skonfigurowania parametru skali.

## Podsumowanie

Problem związany z zaprojektowaniem uniwersalnego algorytmu optymalizacji globalnej, to przede wszystkim problem polegający na ustaleniu kompromisu pomiędzy dwoma wykluczającymi się aspektami: eksploracją oraz eksploatacją przestrzeni rozwiązań. Rozdział ten, zainspirowany został przez szereg publikacji [65],[46],[47],[85], wskazujących na możliwość znalezienia skutecznego kompromisu, za sprawą specjalnych rozkładów prawdopodobieństwa stosowanych w operatorze mutacji algorytmów ewolucyjnych. Stosowany powszechnie w algorytmach ewolucyjnych izotropowy normalny rozkład eksploracyjny, powoduje bardzo szybką zbieżność algorytmu do rozwiązań lokalnych jak i skutecznie radzi sobie z przekraczaniem siodła ewolucyjnych. Niestety, obie sytuacje zachodzą jedynie w przypadku precyzyjnego ustalenia parametru skali rozkładu  $\sigma$ . Co gorsza, obie sytuacje zwykle wymagają zupełnie różnych wartości  $\sigma$ . W ten sposób skuteczność algorytmu w nieznanym środowisku, zależy głównie od tego, w jakim stopniu rozkład normalny potrafi pogodzić te dwa wzajemnie wykluczające się cele optymalizacji. W tym miejscu powstaje pytanie o zasadność stosowania rozkładu normalnego. Być może wszechobecny rozkład normalny nie jest najodpowiedniejszym modelem przeszukiwań przestrzeni rozwiązań? Być może istnieją rozkłady eksploracyjne, które mogą ustanowić znacznie efektywniejszy kompromis pomiędzy eksploracją i eksploatacją? W rozdziale tym, podjęta została próba odpowiedzi na powyższe pytania, z uwzględnieniem klasy stabilnych rozkładów izotropowych.

Na podstawie dwóch prostych eksperymentów przeprowadzonych dla strategii ewolucyjnej: optymalizacji funkcji sferycznej (będącej podstawowym modelem przy badaniu skuteczności lokalnych metod optymalizacji) oraz przekraczania siodła między dwoma wzniesieniami gaussowskimi (powszechnie stosowanym w analizie skuteczności algorytmów optymalizacji globalnej), wykazano przewagę rozkładów ciężkoogonowych. W przypadku optymalizacji lokalnej, rozkłady o

indeksie stabilności  $\alpha = 0.5$  niemalże całkowicie dominują nad innymi rozwiązaniami. Oznacza to, iż powodują znacznie szybszą zbieżność, a zarazem gwarantują otrzymanie rozwiązania o lepszej jakości. Zaprezentowane wyniki, wyraźnie przemawiają za odejściem od powszechnie przyjętej praktyki polegającej na stosowaniu rozkładu normalnego w operatorze mutacji algorytmów ewolucyjnych. Jakkolwiek, słuszność powyższych obserwacji, jest jak najbardziej uzasadniona dla niewielkiej liczby wymiarów przestrzeni rozwiązań, to sytuacja ta ulega zmianie przy problemach o większej liczbie zmiennych decyzyjnych. Efekt ten, przejawia się szczególnie widocznie zawężeniem okna ewolucyjnego dla rozkładów ciężkoogonowych w dwunastowymiarowym problemie przekraczania siodeł. Źródła, oraz szersza dyskusja powodów obserwowanej sytuacji, są przedmiotem opracowania rozdziału 4.2.2.

### 4.2.2 Efekt otoczenia

Obserwowane pogorszenie jakości otrzymywanych rozwiązań lokalnych w miarę zwiększania wymiaru przestrzeni przeszukiwań, było szczegółowo analizowane i tłumaczone tzw. efektem otoczenia w [85]. Generalnie rzecz biorąc, efekt ten jest skutkiem zmian rozkładu normy wektora losowego stosowanego do zaburzeń rozwiązań, jakie zachodzą w wyniku zmian wymiaru rozkładu eksploracyjnego. Sztandarowym przykładem, obrazującym naturę efektu otoczenia, jest rozkład normalny o sferycznej symetrii tj.  $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ . W tym przypadku, norma wektora  $\mathbf{X}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  posiada dobrze znany rozkład  $\|\mathbf{X}^s\| \sim \chi_n$ , dla którego można wyznaczyć zarówno wartość oczekiwaną jak i modę rozkładu:

- wartość oczekiwana:  $\sqrt{2} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$ ,
- moda:  $\sqrt{n-1}$ .

Warto zauważyć, iż wektor losowy  $\mathbf{X}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  można przedstawić w postaci  $\mathbf{X}^s \stackrel{d}{=} \mathbf{u}^{(n)} \chi_n$  (definicja 12). Z powyższej dekompozycji stochastycznej wynika jasno, że stopień podobieństwa pomiędzy rodzicem  $\mathbf{x}$  a jego potomkiem  $\mathbf{x}'$  zależy tylko i wyłącznie od zmiennej losowej  $\|\mathbf{x} - \mathbf{x}'\| \sim \chi_n$ . Z wymienionych wyżej właściwości rozkładu  $\chi_n$  wynika jasno, że zwiększenie wymiaru przestrzeni przeszukiwań, powoduje jednocześnie wzrost różnic pomiędzy rodzicem oraz jego potomkiem. Oczywiście efekt ten nie pozostaje bez wpływu na skuteczności algorytmu ewolucyjnego. Odwołując się do jednej z najwcześniejszych heurystyk adaptacyjnych tzw. reguły 1/5 sukcesu, wiemy, iż wraz ze wzrostem średniej odległości od rodzica, maleje również prawdopodobieństwo sukcesu [23]. Bez prawdopodobieństwa sukcesu utrzymywanego na odpowiednio poziomie, nie jest możliwy dalszy postęp procesu ewolucji. Dlatego też reguła 1/5 sukcesu, w takich sytuacjach, powoduje zawężenie rozkładu eksploracyjnego. Bazując na tych spostrzeżeniach można przypuszczać, iż skuteczność algorytmu ewolucyjnego w optymalizacji lokalnej, w dużej mierze zależy od częstych, skutecznych mutacji o niewielkiej sile zaburzenia. Obserwując wartości statystyk rozkładu  $\chi_n$ , jasnym jest, że wraz ze wzrostem wymiaru przestrzeni  $n$ , otrzymujemy zupełnie odwrotną sytuację, która musi siłą rzeczy przyczynić się do pogorszenia skuteczności algorytmu ewolucyjnego. Pogorszenie to, może ujawnić się w dwojaki sposób. W przypadku algorytmów z selekcją elitarną, można zaobserwować znaczne zwiększenie liczby epok niezbędnych do osiągnięcia rozwiązania o akceptowalnej jakości. Natomiast dla algorytmów ewolucyjnych z selekcją miękką,



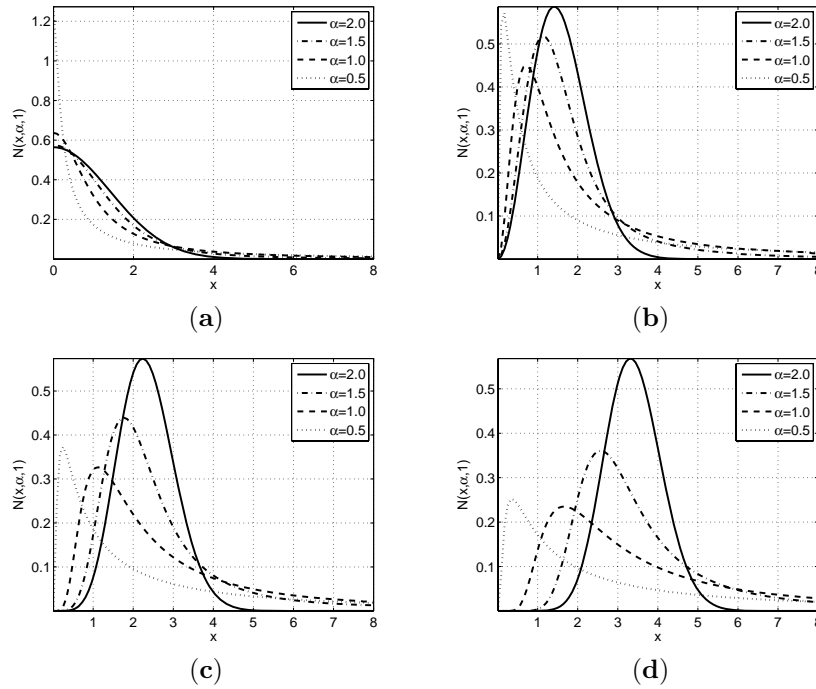
efekt ten najwyraźniej przejawia się zwiększeniem odległości, w której populacja osiąga stan równowagi selekcyjno-mutacyjnej [56].

Według najlepszej wiedzy autora, w przypadku stabilnej klasy rozkładów eksploracyjnych, efekt ten jak dotychczas nie doczekał się bliższego scharakteryzowania.

Rozważmy izotropowy wektor stabilny  $\mathbf{X}_\alpha^s$  (definicja 11), który analogicznie do zastosowanej wyżej dekompozycji rozkładu normalnego, można przedstawić w postaci  $\mathbf{X}_\alpha^s \stackrel{d}{=} \mathbf{u}^{(n)} \|\mathbf{X}_\alpha^s\|$ . Niestety, rozkład  $\|\mathbf{X}_\alpha^s\|$  nie posiada, żadnej znanej analitycznej postaci. Dlatego też, próbę ilościowego opisu efektu otoczenia dla rozkładów stabilnych można jedynie oprzeć na numerycznych metodach obliczeniowych oraz intuicyjnych badaniach eksperymentalnych. W pierwszej kolejności wzór (2.50) może posłużyć do wyznaczenia funkcji gęstości  $N(x; \alpha, \sigma)$  zmiennej losowej  $\|\mathbf{X}_\alpha^s\|$ . Stosując (2.50) otrzymujemy

$$N(x; \alpha, \sigma) = c \frac{2\pi^{n/2}}{\Gamma(n/2)} x^{n-1} G(x^2), \quad (4.8)$$

gdzie  $G(\cdot)$  jest generatorem gęstości stabilnego wektora losowego (2.49). Na wykresie (4.7) zaprezentowane zostały gęstości (4.8) dla najbardziej charakterystycznych indeksów stabilności  $\alpha = 2, 1.5, 1, 0.5$  oraz przykładowych wymiarów przestrzeni przeszukiwań  $n = 1, 3, 6, 12$ .



Rysunek 4.7: Funkcja gęstości zmiennej losowej  $\|\mathbf{X}_\alpha^s\|$ ,  $\mathbf{X}_\alpha^s \in \mathbb{R}^n$ : (a) -  $n = 1$ , (b) -  $n = 3$ , (c) -  $n = 6$ , (d) -  $n = 12$

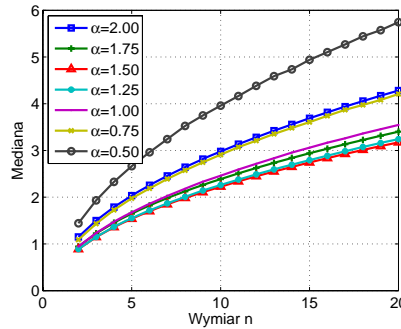
**Obserwacja 11** *Efekt otoczenia powodowany przez zmianę rozkładu normy wektora losowego, jest obecny również dla całej klasy rozkładów stabilnych. Obser-*

wując jednak funkcje gęstości dla różnych wymiarów przestrzeni  $n$ , wyraźnie widać, że proces ten przybiera różną postać w zależności od indeksu stabilności  $\alpha$ . Cechą charakterystyczną jest to, że dla rozkładów o mniejszych indeksach stabilności, więcej masy prawdopodobieństwa skupionej jest w mniejszej odległości od początku układu współrzędnych. Obserwacja ta, pozwala przypuszczać, iż dla rozkładów o cięższych ogonach, efekt otoczenia będzie znacznie mniej uciążliwy, niż ma to miejsce w przypadku rozkładu normalnego.

Jak można zaobserwować na wykresach (4.7) rozkłady norm stabilnych wektorów izotropowych są rozkładami jednomodalnymi. Ponadto, cała klasa wydaje się być uporządkowana względem indeksu stabilności tzn. czym mniejsza wartość  $\alpha$ , tym mniejsza wartość mody rozkładu. W celu bardziej precyzyjnego scharakteryzowania rozważanych rozkładów prawdopodobieństwa, podejmiemy się zadania wyznaczenia median zmiennej zaburzającej  $\|\mathbf{X}_\alpha^s\|$ . Zadanie to, bazując na obliczeniach numerycznych, można sprowadzić do następującego problemu optymalizacji lokalnej:

$$\text{Me}(\alpha, n) = \arg \min_{\text{Me} > 0} \left[ \int_0^{\text{Me}} N(h; \alpha, 1, n) dh - 1/2 \right]^2 \quad (4.9)$$

Wyniki uzyskane przez rozwiązanie zadania (4.9) przedstawione zostały na rysunku 4.8.



Rysunek 4.8: Mediana zmiennej losowej  $\|\mathbf{X}_\alpha^s\|$ ,  $\mathbf{X}_\alpha^s \in \mathbb{R}^n$

Ponadto, okazuje się, że wartości mediany, wyrażonej w funkcji indeksu stabilności oraz wymiaru przestrzeni przeszukiwań, można przedstawić za pomocą następującej zależności (z błędem względnym mniejszym od 0.05):

$$\text{Me}(\alpha, n) = (0.8\sqrt{n} - 0.35)\alpha^2 + (1.02 - 2.25\sqrt{n})\alpha + 2.29\sqrt{n} - 0.88 \quad (4.10)$$

Efekt otoczenia najdobitniej przejawia się w przypadku algorytmów ewolucyjnych z miękką selekcją. Jak to już zostało powiedziane, algorytmy te dążą do osiągnięcia równowagi selekcyjno-mutacyjnej, która skutecznie wyhamowuje proces ewolucji. Populacja rozwiązań zatrzymuje się w pewnej odległości od rozwiązania lokalnego tworząc tym samym tzw. martwy obszar. Jest to otoczenie rozwiązania lokalnego, charakteryzujące się tym, iż prawdopodobieństwo wygenerowania osobnika leżącego w jego wnętrzu jest bardzo małe. Ze względu

na stopień złożoności algorytmów ewolucyjnych z miękką selekcją, precyzyjne określenie rozmiaru martwego obszaru, jest możliwe tylko dla najprostszych przypadków ( np. dwuelementowej populacji ewoluującej w jednowymiarowym środowisku - [56]). W przypadku stabilnych stabilnych wektorów zaburzających, jesteśmy zmuszeni polegać jedynie na analizie symulacji komputerowych.

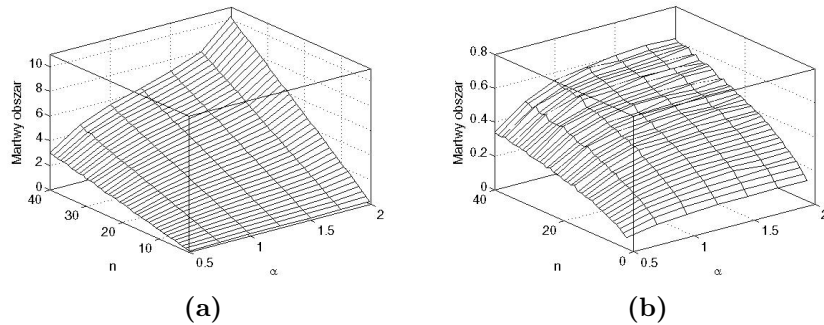
### Badania eksperymentalne

Aby zobrazować efekt otoczenia dla klasy izotropowych rozkładów stabilnych posłużymy się algorytmem przeszukiwań ewolucyjnych z miękką selekcją. Przedmiotem eksperymentu jest problem minimalizacji funkcji celu  $\phi(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{x}}$ . Wybór środowiska optymalizacyjnego pozwala na skojarzenie średniej wartości funkcji celu ze średnią odległością populacji od rozwiązania globalnego. Jako, że celem eksperymentu było ustalenie przeciętnego rozmiaru martwego obszaru, dla różnych indeksów stabilności oraz różnych wymiarów wektorów losowych, przeprowadzono symulacje dla następujących konfiguracji parametrów:

$$\alpha = \{2, 1.75, 1.5, 1.25, 1, 0.75, 0.5\}$$

$$n = \{2, 3, 4, \dots, 40\}$$

Eksperyment zakładał stałą liczebność populacji  $\eta = 20$  oraz zastosowanie selekcji turniejowej (tabela 3.2) z rozmiarem grupy turniejowej równym  $T_G = 4$ . W celu oszacowania rozmiaru martwego obszaru, algorytm ESSS inicjowany był z minimum globalnego  $\mathbf{x}_0 = [0, 0, \dots, 0]^T$ , a po osiągnięciu równowagi selekcyjno-mutacyjnej (przyjęto, że stan ten osiągany był po upływie 50 pokoleń), zapamiętywana była wartość dopasowania najlepszego osobnika. Dla każdej konfiguracji  $\{\alpha, n\}$  proces ten przeprowadzany był 50 razy, a zgromadzone dane poddawane były uśrednieniu. Wyniki zaprezentowane zostały na wykresie 4.9 - (a).



Rysunek 4.9: Wielkość martwego obszaru określona na podstawie eksperymentu z zastosowaniem algorytmu ESSS. (a) - mutacja wektorem  $\mathbf{u}^{(n)} \|\mathbf{X}_\alpha^{(s)}\|$  (b) - mutacja wektorem  $\mathbf{u}^{(n)} |S_\alpha S(1)|$

**Obserwacja 12** Jak można zaobserwować na rysunku 4.9-(a), wraz ze wzrostem wymiaru przestrzeni przeszukiwań oraz indeksu stabilności zwiększa się również rozmiar martwego obszaru. Co ciekawe, dla każdego testowanego  $\alpha$  wzrost ten następuje w sposób zbliżony do liniowego, w zależności od  $n$  oraz

nieliniowy w stosunku do wzrostu  $\alpha$ . Wyniki te, potwierdzają obserwacje dotyczące skuteczności rozkładów ciężkoogonowych, poczynione w rozdziale 4.2.1. Mianowicie, rozkłady o niskich indeksach stabilności powodują, iż populacja rozwiązań osiąga stan równowagi selekcyjno-mutacyjnej w mniejszej odległości od rozwiązania globalnego. Innymi słowy, należy się spodziewać, iż w wyniku zastosowania algorytmu ESSS opartego o rozkład  $\alpha = 0.5$  uzyskamy rozwiązanie o najlepszej jakości.

### Zniwelowanie efektu otocznia

Efekt otoczenia, jak to już zostało wielokrotnie podkreślone w treści tego rozdziału, powodowany jest zmianami rozkładu normy wektorów zaburzających rozwiązania. Naturalnym więc wydaje się pytanie, czy efekt ten można wyeliminować poprzez zastosowanie innych rozkładów zaburzających w operatorze mutacji? Poniżej, postaramy się ustosunkować do powyższego pytania, pozostając dalej w obrębie klasy rozkładów stabilnych.

Biorąc pod uwagę dekompozycję stochastyczną wyrażoną w definicji rozkładu izotropowego 12 postaci

$$\mathbf{X}^s \stackrel{d}{=} \mathbf{u}^{(n)} r, \quad (4.11)$$

naturalnym sposobem na wyeliminowanie efektu otoczenia jest niezależnienie, zmiennej losowej  $r$ , od wymiaru przestrzeni  $n$ . Jedną z możliwości, jest zastosowanie zaburzającego wektora losowego, dla którego dekompozycja (4.11) przybierze postać:

$$\mathbf{X}^s \stackrel{d}{=} \mathbf{u}^{(n)} |S_\alpha S(\sigma)|. \quad (4.12)$$

Warto nadmienić, że  $n$ -wymiarową bazę rozkładu  $\mathbf{u}^{(n)}$  można w prosty sposób otrzymać, przyjmując [38]

$$\mathbf{u}^{(n)} \stackrel{d}{=} \mathbf{X} / \|\mathbf{X}\|, \quad (4.13)$$

gdzie  $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I}_n)$ . Wykorzystując w operatorze mutacji wektor (4.12), można całkowicie wyeliminować zjawisko polegające na zwiększeniu średnich odległości pomiędzy rodzicami a ich potomkami w miarę wzrostu  $n$ . Wystarczy bowiem zauważyć, że zachodzi  $\|\mathbf{X}^s\| = |S_\alpha S(\sigma)|$ , co oznacza, że rozkład normy wektora zaburzającego nie zależy od  $n$ . Na wykresie 4.7-(a) zaprezentowane zostały gęstości zmiennej losowej  $|S_\alpha S(\sigma)|$ . Aby przekonać się, czy rzeczywiście opisywany powyżej zabieg, pozwoli całkowicie wyeliminować efekt otoczenia, przeprowadźmy kolejny eksperyment z wykorzystaniem algorytmu ESSS. Eksperyment przeprowadzony został w identyczny sposób, jak to miało miejsce w poprzednim przypadku (punkt 4.2.2 niniejszego rozdziału), z jedyną różnicą polegającą na zastąpieniu wektora mutującego  $\mathbf{X}_\alpha^s$  wektorem (4.12). Wyniki zaprezentowane zostały na wykresie 4.9-(b).

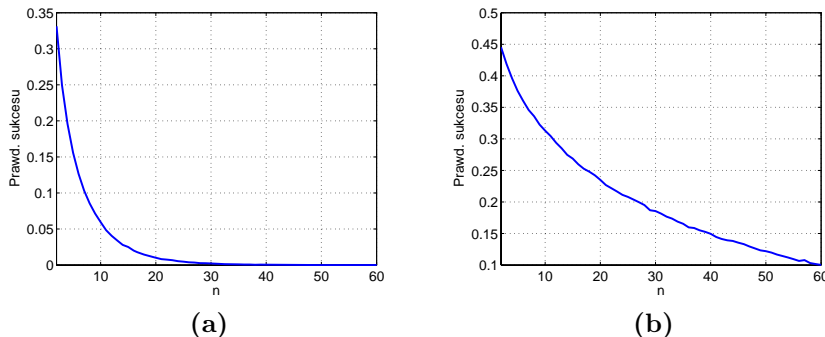
**Obserwacja 13** *Zastosowanie wektora zaburzającego, którego rozkład normy nie zależy od wymiaru przestrzeni przeszukiwań, zdecydowanie zmniejsza rozmiar martwego obszaru. Niemniej jednak, w dalszym ciągu można zaobserwować, niewielką tendencję do pogarszania średniej jakości rozwiązań, w miarę wzrostu  $n$ . Co ciekawe postać tych zmian, ma podobny charakter dla wszystkich indeksów stabilności  $\alpha$ .*

Analiza wyników zaprezentowanych na wykresie 4.9-(b), skłania do konkluzji, iż zjawisko powiększania martwego obszaru, musi mieć, obok efektu otocznia, także

inne źródło. Powodu, dla którego populacja algorytmu ESSS w stanie równowagi selekcyjno-mutacyjnej, fluktuuje w coraz to większej odległości od rozwiązania globalnego, należy upatrywać w bazie rozkładu  $\mathbf{u}^{(n)}$ . Mianowicie, aby zaszła mutacja skutkująca poprawieniem dopasowania rodzica, muszą zostać spełnione dwa warunki:

1. należy wybrać kierunek w przestrzeni rozwiązań, w którym funkcja dopasowania przyjmuje wartości mniejsze od punktu bazowego,
2. dokonać zaburzenia wcześniej obranego kierunku, w taki sposób, aby trafić w obszar o lepszym dopasowaniu.

W przypadku funkcji celu  $\phi(\mathbf{x}) = \sqrt{\mathbf{x}^T \mathbf{x}}$  oraz zastosowania rozkładu równomiernego na sferze jednostkowej  $\mathbf{u}^{(n)}$ , następuje zmniejszenie prawdopodobieństwa wylosowania kierunku, na którym można znaleźć rozwiązania o lepszym dopasowaniu. Aby się o tym przekonać, przeprowadźmy kolejny eksperyment. Punkt  $\mathbf{x} = [1, 0, 0, \dots, 0]^T$  zaburzymy, tym razem wektorem losowym  $\mathbf{u}^{(n)}$  oraz obserwujemy jak zmienia się prawdopodobieństwo sukcesu w miarę wzrostu  $n$ . Oczywiście, wszystkie zaburzenia będą w tym wypadku leżeć na sferze jednostkowej o promieniu 1 i środku  $\mathbf{x} = [1, 0, 0, \dots, 0]^T$ . Zaburzenia, charakteryzujące się lepszą jakością dopasowania, muszą leżeć wewnątrz kuli jednostkowej o środku w początku układu współrzędnych. Na rysunku 4.10, przedstawiono wartość tego prawdopodobieństwa, wyznaczoną na podstawie metody Monte-Carlo. Dla porównania powtórzono eksperyment z nieco dalej położonym punktem bazowym  $\mathbf{x} = [3, 0, 0, \dots, 0]^T$ .



Rysunek 4.10: Prawdopodobieństwo wylosowania kierunku umożliwiającego poprawę rozwiązania za pomocą rozkładu  $\mathbf{u}^{(n)}$  i punktu przestrzeni: (a) -  $\mathbf{x} = [1, 0, 0, \dots, 0]^T$ , (b) -  $\mathbf{x} = [3, 0, 0, \dots, 0]^T$ .

**Obserwacja 14** Dla rozważanej funkcji celu, obszar o lepszym dopasowaniu przyjmuje formę  $n$ -wymiarowej kuli o promieniu  $\|\mathbf{x}\| - S_{\|\mathbf{x}\|}^{(n)}(\mathbf{0})$ , na brzegu której umieszczony jest punkt bazowy  $\mathbf{x}$ . Punkt bazowy zaburzany jest przez wektor  $\mathbf{u}^{(n)}$ , formując równomierny rozkład osobników na powierzchni kuli jednostkowej -  $S_1^{(n)}(\mathbf{x})$ . W ten sposób, prawdopodobieństwo, wylosowania kierunku umożliwiającego poprawę rozwiązania, może być z dużą dokładnością przybliżone, przez stosunek tej części sfery  $\partial S_1^{(n)}(\mathbf{x})$ , która leży wewnątrz  $S_{\|\mathbf{x}\|}^{(n)}(\mathbf{0})$ , do pola powierzchni  $\partial S_1^{(n)}(\mathbf{x})$ . Z wartości przedstawionych na wykresach można wyciągnąć

*dwa ważne wnioski. Po pierwsze dla ustalonych punktów bazowych, wraz ze wzrostem wymiaru przestrzeni coraz rzadziej wylosowywany zostaje kierunek poprawy rozwiązania. Z drugiej strony widać, że jedyną receptą na zwiększenie wartości tego prawdopodobieństwa jest oddalanie punktu bazowego. Uwaga ta staje się oczywista, jeżeli uświadomimy sobie, że zwiększając promień kuli  $S_{\|x\|}^{(n)}(\mathbf{0})$  (czyli odsuwając coraz dalej punkt bazowy), jej brzeg, z punktu widzenia przylegającej do niej sfery jednostkowej, będzie się prostował. Stosując przejście graniczne, można powiedzieć, iż prawdopodobieństwo to, będzie dążyć do wartości 0.5.*

Powyższe uwagi, w zupełności wystarczą do wytłumaczenia efektu zwiększania martwego obszaru, zaobserwowanego na rysunku 4.9-(b). Mianowicie, proces ewolucji w przypadku algorytmu ESSS, zostaje zatrzymany, w miejscu, w którym prawdopodobieństwo sukcesu osiąga pewien ściśle określony poziom. Poziom ten, musi być na tyle duży aby przynajmniej część potomków posiadała dopasowanie zbliżone do swoich rodziców. Należy przypuszczać, że wartość wspomnianego prawdopodobieństwa sukcesu nie zależy od wymiaru przestrzeni, a jedynie od rozmiaru populacji i zastosowanej selekcji. W przypadku, gdy zarówno rozmiar populacji jak i operator mutacji pozostają niezmiennie, jedyną szansą na osiągnięcie progu prawdopodobieństwa gwarantującego równowagę selekcyjno-mutacyjną jest oddalenie od rozwiązania globalnego (patrz obserwacja 14). Skutkuje to natychmiast powiększeniem rozmiaru martwego obszaru zaobserwowanego na rysunku 4.9-(b).

## Podsumowanie

Efekt otoczenia, powoduje szereg komplikacji związanych z ewolucyjną optymalizacją problemów o znacznej liczbie zmiennych decyzyjnych. Zjawisko to przejawia się w dwojaki sposób. W przypadku algorytmów z selekcją elitarną, obserwujemy znaczny spadek współczynnika zbieżności lokalnej (patrz: wyniki eksperymentów przeprowadzone w punkcie 4.2.1). Dla algorytmów z selekcją miękką, znamienne jest pogorszenie jakości otrzymywanych rozwiązań, wraz ze wzrostem wymiaru przestrzeni przeszukiwań. Pogorszenie to, świadczy o powstaniu tzw. martwego obszaru. Ten specyficzny podzbiór uformowany wokół rozwiązania globalnego, cechuje mała wartość prawdopodobieństwa znalezienia w jego wnętrzu osobnika populacji.

W rozdziale tym, efekt otoczenia został szczegółowo scharakteryzowany dla klasy stabilnych rozkładów izotropowych. Wyznaczone charakterystyki rozkładu normy wektora stabilnego oraz ich zależność od wymiaru przestrzeni, pozwalają przypuszczać, że efekt otoczenia będzie obecny dla wszystkich modeli stabilnych. Badania symulacyjne przeprowadzone dla algorytmu przeszukiwań ewolucyjnych, jak najbardziej potwierdziły tą hipotezę. Przy okazji badań eksperymentalnych oszacowano wielkość martwego obszaru w funkcji indeksu stabilności oraz wymiaru przestrzeni rozwiązań. Ponadto zaprezentowany został sposób wyeliminowania efektu otoczenia, poprzez konstrukcję wektora zaburzającego, którego rozkład normy pozostaje niezmienny w stosunku do zmian wymiaru przestrzeni. Adekwatne badania potwierdziły słuszność wcześniej przyjętych założeń. W treści rozdziału zwrócono również uwagę na odmienne źródło spadku efektywności algorytmów ewolucyjnych w problemach wielowymiarowych. Ma on bezpośredni związek ze spadkiem prawdopodobieństwa zlokalizowania kierunku, który umożliwiłby poprawę rozwiązania. Zjawisko to, obecne dla wszyst-

kich rozkładów izotropowych, jest przedmiotem dociekań rozdziału 5.1.1.

### 4.2.3 Analiza zbieżności lokalnej algorytmu ewolucyjnego

Jednym z najważniejszych wskaźników jakości algorytmów optymalizacyjnych jest tzw. współczynnik zbieżności. Generalnie wielkość ta charakteryzuje szybkość, z jaką algorytm iteracyjny zbiega do rozwiązania optymalnego. Przyjmując, że  $\{\mathbf{x}_k\}$  jest ciągiem w przestrzeni  $\mathbb{R}^n$  zbieżnym do  $\mathbf{x}^*$ , można wyróżnić następujące klasy zbieżności [81]:

- zbieżność *Q-liniowa*, jeśli istnieje stała  $r \in (0, 1)$ , taka że, dla dostatecznie dużego  $k$ , zachodzi

$$\frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \leq r, \quad (4.14)$$

- zbieżność *Q-superliniowa*

$$\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 0, \quad (4.15)$$

- zbieżność *Q-kwadratowa*, jeśli dla pewnego  $M > 0$  i dostatecznie dużego  $k$ , zachodzi

$$\frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|^2} \leq M, \quad (4.16)$$

- zbieżność *R-liniowa* - jeśli istnieje ciąg o nieujemnych wartościach  $v_k$  zbieżny *Q-liniowo* do zera, taki że

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq v_k \quad (4.17)$$

Warto zauważyć, że najsłabszą, z wymienionych form zbieżności, jest zbieżność *R-liniowa*, która polega na minimalizacji błędów w całym horyzoncie czasowym, a nie zaś, jak to ma miejsce w przypadku innych form zbieżności, na minimalizacji błędów w poszczególnych iteracjach. Przedstawione definicje znajdują głównie zastosowanie w przypadku algorytmów deterministycznych, i to w dodatku dla specjalnych klasy funkcji np. funkcji wypukłych [53], [27].

#### Zbieżność lokalna strategii ewolucyjnej

W przypadku algorytmów optymalizacji stochastycznej, w tym także algorytmów ewolucyjnych, odchodzi się od wyżej przedstawionych klas zbieżności, definiując tzw. *współczynnik zbieżności lokalnej* (ang. *Progress rate*) [20, 95]. Wielkość ta jest postrzegana jako oczekiwana zmiana odległości od rozwiązania optymalnego, w następujących po sobie iteracjach

$$\varphi = \mathbb{E} \left[ \|\mathbf{x}_k - \mathbf{x}^*\| - \|\mathbf{x}_{k+1} - \mathbf{x}^*\| \right] \quad (4.18)$$

W przypadku algorytmów ewolucyjnych, wyznaczenie wielkości (4.18) jest możliwe jedynie dla najprostszych algorytmów oraz kilku najprostszych funkcji celu [15, 23, 24, 103, 95]. Niemniej jednak analiza współczynnika zbieżności lokalnej

(4.18) może dostarczyć wiele cennej wiedzy na temat praw rządzących dynamiką populacji. W szczególności badania nad współczynnikiem zbieżności lokalnej oraz odpowiadającym mu tzw. prawdopodobieństwem skutecznej mutacji [20], doprowadziło do opracowania najbardziej znanej reguły adaptacji rozkładu prawdopodobieństwa tzw. *reguły 1/5 sukcesu*. Z tego powodu, w dalszej części rozdziału, przedstawiona zostanie analiza zbieżności lokalnej dla izotropowego rozkładu stabilnego.

Rozważmy funkcję celu następującej postaci:

$$\phi(\mathbf{x}) = a(\mathbf{x}^T \mathbf{x})^{b/2}, \quad (4.19)$$

dla  $a, b \geq 0$ . Posługując się dobrze uwarunkowanym zadaniem optymalizacji lokalnej (4.19), oraz prostą strategią ewolucyjną  $(1+1)$ ES, w rozdziale tym spróbujemy odpowiedzieć na pytanie: który przedstawiciel izotropowej klasy rozkładów stabilnych gwarantuje najszybszą zbieżność lokalną.

Biorąc pod uwagę elitarny schemat selekcji stosowany w algorytmie  $(1+1)$ ES, rozwiązanie w iteracji  $k+1$  wyraża się poprzez zależność:

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{x}_k + \mathbf{Z}_k & \text{jeśli } \phi(\mathbf{x}_k + \mathbf{Z}_k) < \phi(\mathbf{x}_k) \\ \mathbf{x}_k & \text{jeśli } \phi(\mathbf{x}_k + \mathbf{Z}_k) \geq \phi(\mathbf{x}_k) \end{cases}, \quad (4.20)$$

gdzie  $\mathbf{Z}_k$  jest realizacją wektora losowego  $\mathbf{X}_\alpha^s$ . Jak już to zostało powiedziane, istnieje wiele kryteriów, mogących posłużyć jako wskaźnik jakości przy ocenianiu zbieżności lokalnej algorytmu optymalizacyjnego. Postać rozważanej funkcji celu (4.19) oraz struktura algorytmu (4.20) skłania do przyjęcia następującej definicji współczynnika zbieżności lokalnej [103]

$$c = \frac{\phi(\mathbf{x}_{k+1})}{\phi(\mathbf{x}_k)} = \left( \frac{\mathbf{x}_{k+1}^T \mathbf{x}_{k+1}}{\mathbf{x}_k^T \mathbf{x}_k} \right)^{b/2}, \quad (4.21)$$

Warto w tym miejscu nadmienić, że przyjęta definicja współczynnika zbieżności (4.21) jest specjalną wersją wyrażenia służącego do oceny zbieżności Q-liniowej (4.14) oraz Q-superliniowej (4.15). Mianowicie pamiętając, że dla funkcji (4.19)  $\mathbf{x}^* = \mathbf{0}$  oraz, że  $\|\mathbf{x}\| = (\mathbf{x}^T \mathbf{x})^{1/2}$ , współczynnik (4.21) można ogólnie przedstawić w postaci, która nawiązuje do wspomnianych typów zbieżności, [?], tj.

$$c^{1/b} = \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \quad (4.22)$$

Co więcej (4.21), dla funkcji sferycznej, odpowiada ściśle definicji współczynnika zbieżności lokalnej (4.18), gdyż

$$\varphi = \|\mathbf{x}_k\|(c-1) \quad (4.23)$$

Uwzględniając stochastyczny charakter rozwiązania w iteracji  $k+1$ , tj.  $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{Z}_k$ , przeformułujmy (4.21), wprowadzając dodatkową zmienną losową  $V$ , oznaczającą iloraz rozkładu wartości funkcji celu po wykonaniu mutacji i jakości aktualnego rozwiązania, tj.:

$$V = \left( \frac{[\mathbf{x}_k + \mathbf{Z}]^T [\mathbf{x}_k + \mathbf{Z}]}{\mathbf{x}_k^T \mathbf{x}_k} \right)^{b/2} \quad (4.24)$$



Biorąc pod uwagę regułę decyzyjną (4.20), współczynnik zbieżności lokalnej (4.21) przyjmuje postać zmiennej losowej  $C$ :

$$C = \min\{V, 1\}, \quad (4.25)$$

Należy zaznaczyć, że powyższa zmienna losowa nie posiada funkcji gęstości, gdyż cała masa prawdopodobieństwa równa  $P(V > 1) > 0$  jest skupiona w punkcie  $v = 1$ . Jako, że rozważany współczynnik (4.25) mający na celu ocenę zbieżności strategii ewolucyjnej (1 + 1)ES jest zmienną losową, dalsza analiza opierać się będzie na wartości oczekiwanej (4.25), tj.

$$\begin{aligned} E[C] &= E\left[\min\{V, 1\}\right] = 1 + \int_0^1 v f_V(v; \alpha, \delta, b) dv - \int_0^1 f_V(v; \alpha, \delta, b) dv \\ &= 1 - \int_0^1 f_V(v; \alpha, \delta, b)[1 - v] dv, \end{aligned} \quad (4.26)$$

gdzie  $f_V(v, \alpha, \delta)$  oznacza funkcję gęstości zmiennej (4.24) o postaci (dodatek F)

$$f_V(v; \alpha, \delta, b) = \frac{2\delta^n v^{\frac{n-b}{b}} \pi^{\frac{n-1}{2}}}{b\Gamma(\frac{n-1}{2})} \int_{-1}^1 G_\alpha\left(\delta^2(v^{2/b} + 2v^{1/b}t + 1)\right)(1-t^2)^{\frac{n-3}{2}} dt, \quad (4.27)$$

gdzie  $\delta = \frac{\|\mathbf{x}_k\|}{\sigma}$ ,  $G_\alpha(\cdot)$  jest generatorem gęstości (2.49). Zakładając, że najbardziej efektywnym izotropowym rozkładem stabilnym, będzie ten, dla którego (4.26) przyjmuje wartość najmniejszą, należy znaleźć optymalną wartość  $\delta^*$  w funkcji indeksu stabilności  $\alpha$ . Uwzględniając (4.26), zadanie to sprowadza się następującego problemu optymalizacji

$$\delta^*(\alpha) = \arg \max_{\delta > 0} \int_0^1 f_V(v, \alpha, \delta)[1 - v] dv \quad (4.28)$$

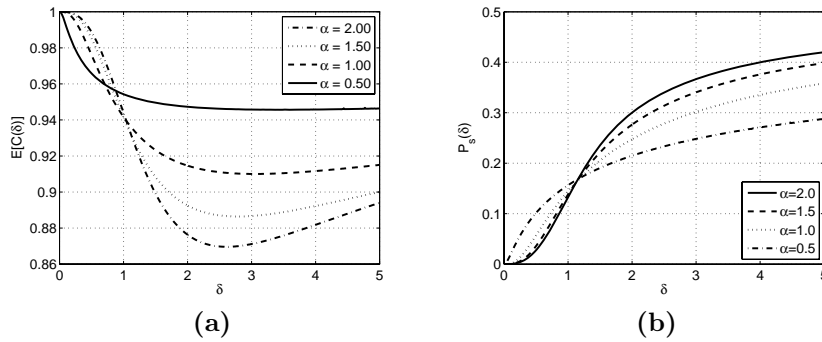
Niestety, ze względu na skomplikowaną postać (4.27), zadania (4.28) nie sposób rozwiązać stosując analityczne metody. Aby przewyciężyć ten problem należy posłużyć się numerycznymi, a zarazem przybliżonymi metodami optymalizacji.

### Analiza zbieżności dla $n = 3$

Podstawiając za  $n = 3$  w (4.27), całka (4.28) zyskuje następującą postać [?]

$$\frac{2}{b} \delta^3 \pi \int_0^1 \int_{-1}^1 [1 - v] v^{\frac{3-b}{b}} G_\alpha\left(\delta^2(v^{2/b} + 2v^{1/b}t + 1)\right) dt dv \quad (4.29)$$

Wykorzystując metody numeryczne do maksymalizacji (4.29) można odnaleźć szukane  $\delta^*(\alpha)$  [?]. Ich wartości oraz odpowiadające im wartości współczynników zbieżności lokalnej zostały zamieszczone odpowiednio w tabelach 4.1 i 4.2. Dodatkowo na rysunku 4.11 zobrazowano wartość współczynnika jakości w funkcji parametru  $\delta$ .



Rysunek 4.11: (a) - współczynnik zbieżności lokalnej dla strategii ewolucyjnej  $(1+1)ES$  dla trójwymiarowej funkcji celu  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ , oraz odpowiadające im prawdopodobieństwa sukcesu - (b)

Tabela 4.1: Pseudo-optymalne wartości współczynnika  $\delta^*$ , gwarantujące naj-  
szybszą zbieżność algorytmu  $(1+1)ES_\alpha$  dla trójwymiarowej funkcji celu  $\phi(\mathbf{x}) =$   
 $(\mathbf{x}^T \mathbf{x})^b$

$\alpha$	$b = 0.1$	$b = 0.25$	$b = 0.5$	$b = 1.0$	$b = 2.0$	$b = 4.0$
2.00	2.28	2.30	2.35	2.43	2.61	2.96
1.75	2.29	2.32	2.38	2.49	2.69	3.08
1.50	2.33	2.37	2.44	2.56	2.78	3.24
1.25	2.40	2.44	2.51	2.65	2.90	3.37
1.00	2.49	2.58	2.62	2.78	3.07	3.54
0.75	2.68	2.67	2.80	2.97	3.26	3.64
0.50	3.21	2.89	3.37	3.50	3.56	4.20

Tabela 4.2: Pseudo-optymalne wartości współczynników zbieżności lokalnej  
 $E(C(\delta^*))$  dla algorytmu  $(1+1)ES_\alpha$  oraz trójwymiarowej funkcji celu  $\phi(\mathbf{x}) =$   
 $(\mathbf{x}^T \mathbf{x})^b$

$\alpha$	$b = 0.1$	$b = 0.25$	$b = 0.5$	$b = 1.0$	$b = 2.0$	$b = 4.0$
2.00	0.9899	0.9757	0.9547	0.9200	0.8696	0.8076
1.75	0.9906	0.9775	0.9579	0.9253	0.8774	0.8171
1.50	0.9915	0.9795	0.9616	0.9314	0.8864	0.8287
1.25	0.9924	0.9818	0.9657	0.9384	0.8971	0.8431
1.00	0.9935	0.9843	0.9704	0.9466	0.9100	0.8612
0.75	0.9947	0.9873	0.9759	0.9563	0.9259	0.8842
0.50	0.9962	0.9908	0.9825	0.9681	0.9456	0.9141

**Obserwacja 15** *Klasa izotropowych rozkładów stabilnych, w zadaniach optymalizacji lokalnej, wykazuje ciekawą regularność. Otóż, na podstawie przeprowadzonych badań, wyraźnie widać, że strategia ewolucyjna  $(1+1)ES_\alpha$  dla mniejszych indeksów stabilności, zapewnia wolniejszą zbieżność do rozwiązania. Z drugiej jednak strony, należy pamiętać, że sytuacja taka, ma miejsce jedynie wtedy, gdy w odpowiedni sposób dobrany jest parametr skali  $\sigma^* = \frac{\|\mathbf{x} - \mathbf{x}^*\|}{\delta^*}$  rozkładu eksploracyjnego - tabela 4.1. Dobór optymalnej wartości  $\sigma^*$  jest podyktowany znajomością odległości euklidesowej rozwiązania aktualnego od rozwiązania globalnego, co z praktycznego punktu widzenia, jest sytuacją na ogół niespotykaną. Niemniej jednak, wyniki przedstawione na rysunku 4.11, pozwalają odnieść się do kwestii związanej z odpornością rozważanej klasy mutacji stabilnej, na niepewność związaną z brakiem znajomości  $\|\mathbf{x} - \mathbf{x}^*\|$ . Mianowicie, dla mutacji opartych o rozkłady z większymi indeksami stabilności, zakres parametru skali, gwarantujący szybszą zbieżność, jest znacznie mniejszy niż w przypadku ich odpowiedników z cięższymi ogonami. Istnieje zatem szansa, że stosując mniej wrażliwe na właściwy dobór parametrów rozkłady ciężkoogonowe, średnia szybkość zbieżności strategii  $(1+1)ES_\alpha$  będzie większa niż w przypadku zastosowania mutacji zbliżonych do gaussowskich ( $\alpha \approx 2$ ).*

Warto zaznaczyć, iż rozważanym przypadku, zbieżność algorytmu ewolucyjnego zależy od prawidłowego doboru parametru skali rozkładu eksploracyjnego. Wybór ten, jak można się przekonać na podstawie analizy przeprowadzonej w treści rozdziału, ściśle zależy od odległości aktualnego rozwiązania od optimum globalnego. W praktyce zaś, odległość ta nie jest zwykle znana. Nie stoi to jednak na przeszkodzie, aby odległość tą można było estymować w trakcie działania algorytmu ewolucyjnego. Jedną z najwcześniejszych heurystyk adaptacyjnych tzw. reguła 1/5 sukcesu, używa do tego celu dosyć specyficznego estymatora w postaci prawdopodobieństwa sukcesu. Wielkość ta, określa prawdopodobieństwo zajścia skutecznej mutacji tj. mutacji rodzica, polegającej na otrzymaniu osobnika o lepszym dopasowaniu. Zastosowanie prawdopodobieństwa sukcesu jako estymatora odległości od rozwiązania, jest możliwe, gdyż istnieje ścisły związek pomiędzy tymi dwoma wielkościami [20].

Według najlepszej wiedzy autora problematyka ta, w kontekście rozkładów stabilnych, jak dotąd nie była przedmiotem zainteresowania żadnej publikacji naukowej. Znając optymalne konfiguracje zamieszczone w tabeli 4.1, można z łatwością wyznaczyć odpowiadające im optymalne wartości prawdopodobieństwa sukcesu. Mianowicie warto zauważyć, iż prawdopodobieństwo sukcesu może zostać wyznaczone na podstawie znanych już zależności:

$$P_s(\alpha, \delta, b) = \int_0^1 f_V(v; \alpha, \delta, b) dv, \quad (4.30)$$

gdzie  $f_V(v; \alpha, \delta, b)$  dane jest wzorem (4.27). W przypadku gdy  $n = 3$  (4.30) przyjmuje postać:

$$P_s(\alpha, \delta, b) = \frac{2}{b} \delta^3 \pi \int_0^1 \int_{-1}^1 v^{\frac{3-b}{b}} G_\alpha \left( \delta^2 (v^{2/b} + 2v^{1/b}t + 1) \right) dt dv \quad (4.31)$$

Wartości (4.31) dla optymalnie skonfigurowanych rozkładów, zamieszczone zostały w tabeli 4.3.

Tabela 4.3: Prawdopodobieństwo sukcesu  $P_s(\alpha, \delta^*, b)$  dla optymalnych parametrów  $\delta^*$  zamieszczonych w tabeli 4.1.

$\alpha$	$b = 0.1$	$b = 0.25$	$b = 0.5$	$b = 1.0$	$b = 2.0$	$b = 4.0$
2.00	0.3250	0.3265	0.3302	0.3357	0.3470	0.3650
1.75	0.3139	0.3162	0.3204	0.3277	0.3397	0.3587
1.50	0.3029	0.3056	0.3103	0.3177	0.3299	0.3508
1.25	0.2880	0.2904	0.2945	0.3022	0.3146	0.3339
1.00	0.2780	0.2827	0.2847	0.2923	0.3047	0.3216
0.75	0.2648	0.2644	0.2695	0.2759	0.2857	0.2970
0.50	0.2537	0.2453	0.2576	0.2605	0.2619	0.2748

**Obserwacja 16** *Dla mniejszych indeksu stabilności  $\alpha$ , optymalna konfiguracja rozkładów eksploracyjnych charakteryzuje się coraz mniejszym prawdopodobieństwem sukcesu. Oznacza to, że reguła 1/5 sukcesu nie będzie obowiązywać dla całej klasy rozkładów stabilnych. W celu zagwarantowania najszybszej zbieżności lokalnej strategii  $(1+1)ES_\alpha$  należy, heurystykę adaptacyjną oprzeć o parametry zamieszczone w tabeli 4.3.*

## Podsumowanie

Analiza zbieżności lokalnej jest zagadnieniem o szczególnym znaczeniu przy projektowaniu algorytmów deterministycznych. Przede wszystkim pozwala na oszacowanie zarówno czasu działania algorytmu, jak i jakości otrzymywanego rozwiązania. W przypadku algorytmów ewolucyjnych, analiza ta stwarza niepowtarzalną szansę na dokonanie analitycznej oceny wpływu rozkładu eksploracyjnego na skuteczność procedury optymalizacyjnej.

W rozdziale przedstawione zostały wyniki badań nad zbieżnością lokalną strategii ewolucyjnej  $(1+1)ES_\alpha$  z operatorem mutacji opartym o izotropowe rozkłady stabilne. Analiza wskaźnika jakości, w postaci oczekiwanej poprawy rozwiązania, wyraźnie wskazuje na wyższość rozkładu normalnego. Wyniki te, pozostają w zgodzie z dotychczasowymi spostrzeżeniami spotykanymi w literaturze [104]. Niemniej jednak, należy szczególnie wyraźnie zaznaczyć, iż sytuacja taka ma miejsce jedynie wówczas, gdy rozkład normalny zostanie optymalnie dopasowany do problemu optymalizacji. W praktyce jednak, sytuacja taka, jest najczęściej niemożliwa. Ocena efektywności poszczególnych rozkładów eksploracyjnych jest wówczas uwarunkowana przez kwestie związane z odpornością poszczególnych modeli przeszukiwań. Oceniając możliwości optymalizacji lokalnej strategii  $(1+1)ES_\alpha$ , przy wystąpieniu niepewności związanej z jej optymalną konfiguracją, należy się spodziewać drastycznego spadku skuteczności mutacji gaussowskiej. Rezultaty badań zobrazowane na rysunku 4.11 wyraźnie wskazują na prawdziwość powyższej hipotezy. Zagadnieniu odporności izotropowej mutacji stabilnej poświęcony został rozdział 4.2.4.

Ponadto wyznaczone optymalne wartości prawdopodobieństwa sukcesu dla poszczególnych modeli stabilnych, mogą posłużyć do zaprojektowania wydajnej heurystyki adaptacyjnej. Procedura, wzorowana na słynnej regule 1/5 sukcesu, zaproponowana zostanie w rozdziale 4.2.5.

#### 4.2.4 Odporność mutacji stabilnej

Pojęcie odporności, bardzo często spotykane w wielu dziedzinach nauki, w zależności od kontekstu, posiada różne definicje. W statystyce, pojęcie odporności najczęściej odnosi się do estymatorów, które mimo niespełnienia niektórych teoretycznych założeń dotyczących ich rozkładów, w dalszym ciągu mogą być podstawą wiarygodnej analizy statystycznej [70]. W teorii sterowania, kluczowym zagadnieniem jest odporność układu na zakłócenia zewnętrzne, szумы pomiarowe czy błędy modelowania [129],[60]. W przypadku technik optymalizacyjnych pojęcie odporności może oznaczać małą wrażliwość algorytmu na

- nieodpowiednie skonfigurowanie parametrów zewnętrznych [104],
- zakłócenia losowe funkcji celu [13],
- niestacjonarną naturę problemu [28].

Ogólnie rzecz biorąc, pojęcie odporności odnosić się zatem będzie do kwestii oceny, jak pewne rozwiązania teoretyczne sprawdzają się w przypadku ich praktycznych zastosowań, w warunkach niepewności lub niekompletności dostępnej wiedzy.

Jak wiadomo, dla algorytmów optymalizacji globalnej problem niekompletności wiedzy, szczególnie tej potrzebnej do optymalnego skonfigurowania procedur optymalizacyjnych, jest czymś zupełnie naturalnym, tak więc pytanie o odporność tego typu jest jak najbardziej uzasadnione.

Wyniki przedstawione w rozdziale dotyczącym analitycznej oceny zbieżności strategii ewolucyjnej z mutacją stabilną, pozwoliły na sformułowanie przypuszczenia, iż rozkłady o mniejszym indeksie stabilności powinny charakteryzować się większą odpornością pod względem niepewności związanej z wyborem ich parametru skali. Rozdział ten, poświęcony jest weryfikacji powyższego przypuszczenia, przy zastosowaniu specjalnego środowiska testowego, tzw. ogólnej przestrzeni przeszukiwań. Sam proces eksperymentalnego potwierdzenia zakładanej tezy, przebiega według następującego scenariusza. Odwołując się do modelu optymalizacji zwanego czarną skrzynką, zakłada się, że funkcja celu należy do klasy funkcji ciągłych, przyjmujących wartości dodatnie, określonych na pewnej hiperkostce. Tak postawione zadanie, należy rozwiązać dysponując narzędziem optymalizacyjnym w postaci strategii ewolucyjnej  $(1+1)ES_\alpha$  z mutacją izotropową. Przy braku jakichkolwiek dodatkowych założeń, głównym celem niniejszego rozdziału, jest wykazanie statystycznej prawidłowości, polegającej na tym, iż dobierając losowo parametr skali dla rozkładów o mniejszych indeksach stabilności, szanse na skuteczne działanie algorytmów  $(1+1)ES_\alpha$  znacząco rosną.

Generalny schemat przy wykazywaniu podobnych statystycznych zależności polega na wykorzystaniu pojęcia wartości oczekiwanej pewnego wskaźnika skuteczności algorytmów optymalizacyjnych, obliczanej względem miary prawdopodobieństwa narzucanej przez stochastyczny sposób sformułowania problemu. W tym miejscu, automatycznie nasuwają się skojarzenia z twierdzeniem *No Free Lunch Theorem* [133], które mówi o jednakowych wartościach oczekiwanych wszystkich algorytmów optymalizacyjnych stosowanych do klasy funkcji ciągłych. Należy jednak zaznaczyć, że kombinatoryczny dowód przytoczonego twierdzenia, opiera się na założeniu o równoprawdopodobnym wystąpieniu każdego problemu optymalizacyjnego. Odnosząc się do wcześniej wspomnianego schematu, odpowiada to sytuacji, w której wartości oczekiwane, wyznaczane są

względem jednorodnej miary probabilistycznej, określonej na przestrzeni funkcji ciągłych. Jest sprawą nie podlegającą żadnej dyskusji, iż ograniczając przestrzeń wszystkich funkcji do ich pewnych podklas, można zaobserwować różne oczekiwane skuteczności technik optymalizacyjnych. Na tej podstawie mówimy o specjalizacji algorytmów do rozwiązywania problemów określonych typów. Problem podziału przestrzeni funkcji ciągły i przypisanie do nich dedykowanych technik optymalizacyjnych jest zadaniem niezwykle skomplikowanym. W dalszej części rozdziału zostanie zaproponowany specjalny, stochastyczny sposób sparametryzowania przestrzeni klasy funkcji ciągłych, który umożliwi w łatwy sposób podział przestrzeni funkcyjnej na podklasy, względem których szacowana jest oczekiwana skuteczność strategii ewolucyjnych z mutacjami stabilnymi.

### Uniwersalna przestrzeń przeszukiwań

Rozważmy przestrzeń funkcji ciągłych  $\mathcal{C}[\mathbb{S}]$ , określonych na pewnym zbiorze  $\mathbb{S} \subset \mathbb{R}^n$ . Jednym z najważniejszych twierdzeń aproksymacyjnych jest tzw. twierdzenie Stone-Weierstrass'a [33], w wyniku którego możliwe jest udowodnienie twierdzenia

**Twierdzenie 12** *Niech  $\mathcal{F} = \bigcup_{N=1}^{\infty} \mathcal{F}_N$ , oznacza zbiór funkcji, gdzie  $\mathcal{F}_N$  jest postaci*

$$\mathcal{F}_N = \left\{ f(\mathbf{x}) = \sum_{i=1}^N w_i \exp(-b_i \|\mathbf{x} - \mathbf{m}_i\|^2); w_i, b_i \in \mathbb{R}_+, \mathbf{m}_i \in \mathbb{S} \right\} \quad (4.32)$$

*Wówczas  $\mathcal{F}$  jest zbiorem gęstym w  $\mathcal{C}[\mathbb{S}]$ .*

Konsekwencją twierdzenia 12 jest fakt, że każda funkcja  $g \in \mathcal{C}[\mathbb{S}]$  może zostać przybliżona przez funkcję  $f \in \mathcal{F}$  z dowolnie małym błędem, tj. dla dowolnie małej liczby  $\varepsilon > 0$  oraz każdej funkcji  $g \in \mathcal{C}[\mathbb{S}]$ , istnieje funkcja  $f \in \mathcal{F}$  taka, że zachodzi

$$|g(\mathbf{x}) - f(\mathbf{x})| < \varepsilon \quad (4.33)$$

dla wszystkich  $\mathbf{x} \in \mathbb{S}$ . Dlatego też, wydaje się, iż wnioski dotyczące skuteczności optymalizacyjnej algorytmów ewolucyjnych, sformułowane na podstawie zbioru  $\mathcal{F}$ , można bez starty ogólności, odnieść do klasy wszystkich funkcji ciągłych przyjmujących wartości dodatnie.

Idea stochastycznej parametryzacji zbioru  $\mathcal{F}$  polega na traktowaniu parametrów  $N \in \mathbb{N}$ ,  $w_i, b_i \in \mathbb{R}$ ,  $i = 1, \dots, N$ ,  $\mathbf{m}_i \in \mathbb{S} \subset \mathbb{R}^n$ ,  $i = 1, \dots, N$  jako elementów o charakterze losowym. Zakładając, że odpowiadające im rozkłady prawdopodobieństwa należą do rodzin parametrycznych, można w prosty sposób dokonać parametryzacji zbioru  $\mathcal{F}$ . W dalszej części rozdziału przyjęto następujący charakter losowy poszczególnych parametrów:

- $N \in \mathbb{N}$  posiada rozkład Poisson'a z wartością oczekiwaną  $\lambda_N$ , oraz funkcją prawdopodobieństwa

$$P(N = k; \lambda_N) = \frac{\lambda_N^k \exp(-\lambda_N)}{k!} \quad (4.34)$$

- $\mathbf{m}_i \in \mathbb{S}$ ,  $i = 1, \dots, N$  posiadają rozkład równomierny na przestrzeni  $\mathbb{S}$ . Tym samym odpowiadająca im funkcja gęstości jest postaci

$$f(\mathbf{m}_i) = \frac{1}{|\mathbb{S}|} \quad (4.35)$$

- $w_i \in \mathbb{R}$ ,  $i = 1, \dots, N$  oraz  $b_i \in \mathbb{R}$ ,  $i = 1, \dots, N$  posiadają rozkład wykładniczy ze wspólną wartością oczekiwaną  $E[w_i] = \lambda_w$ ,  $E[b_i] = \lambda_b$ ,  $i = 1, \dots, N$ , oraz gęstością

$$f(t; \lambda) = \begin{cases} \frac{1}{\lambda} \exp(-\frac{1}{\lambda}t) & , t \geq 0 \\ 0 & , t < 0 \end{cases} \quad (4.36)$$

Oznaczając  $\Omega_k \ni \boldsymbol{\theta}_k = [N, w_1, \dots, w_N, b_1, \dots, b_N, \mathbf{m}_1, \dots, \mathbf{m}_N]$ , gdzie  $\Omega_k = \mathbb{N} \times \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^{k \times n}$ , otrzymujemy przestrzeń parametryczną  $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$ , zbioru funkcji ciągłych postaci (4.32). Miara prawdopodobieństwa  $P$ , określona na przestrzeni  $\Omega$ , zdefiniowana jest poprzez (4.34)-(4.36), i charakteryzuje się następującymi rozkładami warunkowymi:

$$P(w_i \in I_w | \boldsymbol{\theta}(1) = k) = \begin{cases} \int_A f(t; \lambda) dw_i & \text{dla } i \leq k \\ 0 & \text{dla } i > k \end{cases} \quad (4.37)$$

$$P(b_i \in I_b | \boldsymbol{\theta}(1) = k) = \begin{cases} \int_A f(t; \lambda) db_i & \text{dla } i \leq k \\ 0 & \text{dla } i > k \end{cases} \quad (4.38)$$

$$P(\mathbf{m}_i \in I_m | \boldsymbol{\theta}(1) = k) = \begin{cases} |I_m| & \text{dla } i \leq k \\ 0 & \text{dla } i > k \end{cases} \quad (4.39)$$

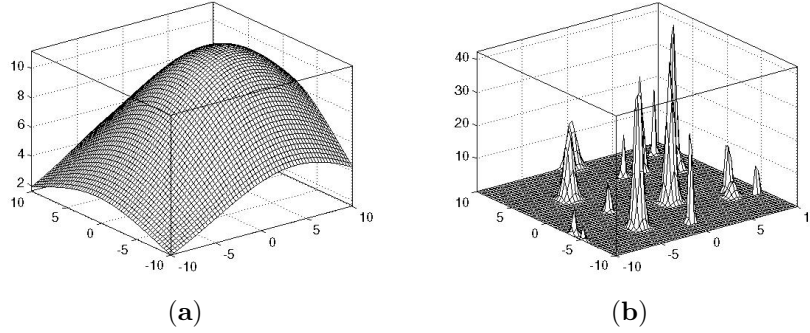
gdzie  $I_w, I_b \subset \mathbb{R}$ ,  $I_m \subset \mathbb{S}$ , natomiast  $f(\cdot; \lambda)$  jest postaci (4.36). Wybór odpowiednich rozkładów prawdopodobieństwa, dla parametrów zbioru  $\mathcal{F}$ , ma charakter arbitralny, o ile spełnione są dodatkowe założenia. Mianowicie, z teoretycznego punktu widzenia, najważniejszą sprawą jest zastosowanie rozkładów, o nośnikach pokrywających całą dziedzinę parametrów zbioru  $\mathcal{F}$ . Warunek ten jest konieczny do tego aby zachodziła poniższa zależność:

$$\forall g \in \mathcal{F}, \forall \delta > 0, \forall \mathbf{x} \in \mathbb{S} \quad P(\boldsymbol{\theta} \in \Omega : |g(\mathbf{x}) - h(\mathbf{x}; \boldsymbol{\theta})| < \delta) > 0, \quad (4.40)$$

gdzie  $h(\cdot, \boldsymbol{\theta})$  jest elementem przestrzeni probabilistycznej  $\mathcal{F}$  odpowiadającym parametrom  $\boldsymbol{\theta}$ . Nierówność (4.40) oznacza, że prawdopodobieństwo otrzymania dowolnie bliskiego przybliżenia, każdego elementu  $g \in \mathcal{F}$ , jest zawsze większe od zera. Jednocześnie oznacza to, że traktując przestrzeń jako środowisko testowe dla algorytmów ewolucyjnych możliwe jest wygenerowanie problemu, który do złudzenia przypominać będzie <sup>3</sup> dowolną funkcję ciągłą.

Należy podkreślić, że w wyniku przyjętych rozkładów (4.34)-(4.36), zbiór  $\mathcal{F}$ , który jest gęsty w zbiorze wszystkich, dodatnich funkcji ciągłych, sparametryzowany zostaje przy pomocy zaledwie trzech parametrów  $\lambda_N, \lambda_w, \lambda_b$ . Zmiany ich wartości, powodują jednocześnie zmiany miary probabilistycznej określonej na zbiorze  $\Omega$ . W ten sposób dla różnych zestawów parametrów  $\lambda_N, \lambda_w, \lambda_b$  uwypuklone zostaną różne podklasy przestrzeni funkcji ciągłych na  $\mathbb{S}$ . Aby lepiej zobrazować ten efekt na rysunku 4.12 przedstawiono przykładowe realizacje otrzymanych środowisk testowych.

<sup>3</sup>zaistniałe ewentualne różnice, nie będą miały żadnego znaczenia, z punktu widzenia porównania stochastycznych algorytmów optymalizacji



Rysunek 4.12: Przykładowe funkcje testowe należące do ogólnej przestrzeni przeszukiwań, otrzymane w wyniku realizacji losowej względem miary prawdopodobieństwa o następujących parametrach: (a) -  $\lambda_w = 10$ ,  $\lambda_b = 0.01$ ,  $\lambda_N = 0.01$ , (b) -  $\lambda_w = 20$ ,  $\lambda_b = 20$ ,  $\lambda_N = 20$

Oczywiście, różne rozkłady prawdopodobieństwa przypisane do parametrów funkcji (4.32) warunkują łatwość, czy też sposób uwypuklenia pewnych podklas problemów optymalizacyjnych, jednak ze względu na ograniczony zakres pracy, w dalszej części rozdziału ograniczymy się jedynie do zależności (4.34)-(4.36).

### Oczekiwana skuteczność stabilnej mutacji izotropowej

W niniejszej części przyjmijmy, że  $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$  będzie przestrzenią parametrów funkcji (4.32),  $P$  oznacza miarę prawdopodobieństwa określoną na  $\Omega$ . Załóżmy również, że skuteczność optymalizacyjna algorytmu stochastycznego  $A$ , z ustalonymi parametrami zewnętrznymi  $\xi$ , dla zadania  $f \in \mathcal{F}$ , jest określona przez pewną funkcję  $H_A(\xi, f)$ . Dodatkowo przyjmijmy, że wskaźnik jakości spełnia zależność  $0 \leq H \leq 1$ . W przypadku problemu oceny skuteczności optymalizacyjnej w zadaniach maksymalizacji,  $H$  może przyjmować postać:

$$H_A(\xi, \phi) = \frac{\max\{A(\xi, f)\}}{f^*}, \quad (4.41)$$

gdzie  $f^*$  jest maksimum globalnym funkcji, a  $\max\{A(\xi, f)\}$  jest najlepszym rozwiązaniem znalezionym przez algorytm  $A$  z parametrami  $\xi$ . Średnią efektywność algorytmu w zbiorze problemów  $\mathcal{F}$ , można wyznaczyć na podstawie

$$\bar{H}_A(\xi) = \int_{\Omega} H_A(\xi, \theta) d\theta. \quad (4.42)$$

Powyższy wzór, prawdziwy jest niestety tylko w przypadku algorytmów deterministycznych, które dla ustalonych warunków początkowych  $\xi$  zwracają ten sam wynik. W przypadku algorytmów stochastycznych, wskaźnik  $H$  jest pewną zmienną losową, najczęściej o nieznanym rozkładzie. Dlatego w przypadku algorytmów stochastycznych można mówić jedynie o oczekiwanej skuteczności, zdefiniowanej jako

$$E[H_A(\xi) | \theta] = \int_0^1 H_A(\xi, \theta) \Pr(dH), \quad (4.43)$$



gdzie  $\Pr(\cdot)$  jest miarą prawdopodobieństwa określoną przez rozkład wartości  $H_A(\boldsymbol{\xi}, \boldsymbol{\theta})$  dla problemu w zbiorze  $\mathcal{F}$ , określonego przez  $\boldsymbol{\theta}$ . Warto w tym miejscu zauważyć, że powszechnie stosowaną praktyką przy testowaniu różnych rozwiązań optymalizacji stochastycznej, jest przybliżanie wartości oczekiwanej (4.43) estymatorem

$$\hat{H}_A(\boldsymbol{\xi}|\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N H_A^{(i)}(\boldsymbol{\xi}, \boldsymbol{\theta}), \quad (4.44)$$

gdzie  $H_A^{(i)}(\boldsymbol{\xi}, \boldsymbol{\theta})$  są wynikami (4.41) uzyskanymi przez algorytm  $A$  w wyniku jego niezależnych uruchomień. Wzór (4.44) jest niczym innym, jak zastosowaniem dobrze znanej metody Monte-Carlo do obliczania całki (4.43). W przypadku algorytmów stochastycznych, można zatem mówić jedynie o średniej skuteczności optymalizacyjnej w zbiorze  $\mathcal{F}$

$$\bar{H}_A(\boldsymbol{\xi}) = \int_{\Omega} E[H_A(\boldsymbol{\xi})|\boldsymbol{\theta}] d\boldsymbol{\theta}. \quad (4.45)$$

Ponadto, biorąc pod uwagę probabilistyczną reprezentację uogólnionej przestrzeni przeszukiwań, należy wartość średnią (4.45) zastąpić pojęciem skuteczności oczekiwanej:

$$\hat{H}_A(\boldsymbol{\xi}) = \int_{\Omega} E[H_A(\boldsymbol{\xi})|\boldsymbol{\theta}] P(d\boldsymbol{\theta}). \quad (4.46)$$

Zawężając rozważania do przypadku strategii ewolucyjnych z izotropowymi rozkładami stabilnymi t.j.  $A \rightarrow (1+1)\text{ES}_{\alpha}$ , należy zauważyć, że dla ustalonego indeksu stabilności, otrzymujemy wektor  $\boldsymbol{\xi}$  postaci  $\boldsymbol{\xi} = [\sigma \in \mathbb{R}_+, \mathbf{x}_0 \in \mathbb{S}]^T$ . Jako, że rozpatrywany model optymalizacji nie dostarcza żadnych informacji dotyczących ewentualnego doboru  $\sigma$  oraz  $\mathbf{x}_0$  mogą one przyjmować dowolne wartości z dopuszczalnych zakresów  $\sigma \in (0, d(\mathbb{S}))^4$  oraz  $\mathbf{x}_0 \in \mathbb{S}$ . W takiej sytuacji najrozsądniejszym rozwiązaniem jest założenie ich stochastycznego charakteru w postaci rozkładów równomiernych na wspomnianych podzbiorach. Przy powyższych założeniach oczekiwana skuteczność strategii ewolucyjnej wyznaczona może zostać z poniżej zależności:

$$\hat{H}_A = \int_0^{d(\mathbb{S})} \int_{\mathbb{S}} \int_{\Omega} \int_0^1 H_A(\sigma, \mathbf{x}_0, \boldsymbol{\theta}) \Pr(dH) P(\boldsymbol{\theta}) P(d\mathbf{x}_0) P(d\sigma). \quad (4.47)$$

Miara  $\Pr(dH)$  oczywiście jest uzależniona od wielkości  $\boldsymbol{\theta}, \mathbf{x}_0, \sigma$ , co pociąga za sobą szereg komplikacji, nawet przy próbie zastosowania metody Monte-Carlo do obliczenia (4.47). Szukając niezbędnego uproszczenia problemu (4.47), posłużmy się następującym scenariuszem. Mianowicie założymy, że wynik pojedynczego uruchomienia strategii ewolucyjnej weźmiemy za pewnik, co oznacza, że miara prawdopodobieństwa  $\Pr(dH)$  redukuje się do dystrybucji Deltę-Dirac'a. Założenie takie, w przypadku zastosowania numerycznych generatorów liczb pseudolosowych nie jest zupełnie bezpodstawne. Znajomość bowiem ziarna oraz samej procedury uzyskiwania liczb pseudolosowych, czyni każdy algorytm stochastyczny, algorytmem deterministycznym. Założenie takie, jakkolwiek kontrowersyjne, umożliwia przedstawienie całki (4.47) w znacznie prostszej postaci

$$\hat{H}_A = \int_0^{d(\mathbb{S})} \int_{\mathbb{S}} \int_{\Omega} H_A(\sigma, \mathbf{x}_0, \boldsymbol{\theta}) P(\boldsymbol{\theta}) P(d\mathbf{x}_0) P(d\sigma). \quad (4.48)$$

<sup>4</sup> $d(\mathbb{S})$  oznacza średnicę zbioru  $\mathbb{S}$ , tj.  $d(A) = \sup\|x - y\| : x, y \in A$

Mimo, iż mamy do czynienia z całką potrójną, to należy zauważyć, że miary prawdopodobieństwa występujące w (4.48) są wzajemnie niezależne, co umożliwia jej przybliżenie za pomocą estymatora:

$$\hat{H}_A = \frac{1}{N} \sum_{i=1}^N H_A(\sigma^{(i)}, \mathbf{x}_0^{(i)}, \boldsymbol{\theta}^{(i)}), \quad (4.49)$$

gdzie  $\sigma^{(i)}, \mathbf{x}_0^{(i)}, \boldsymbol{\theta}^{(i)}$  są niezależnymi realizacjami losowymi miary prawdopodobieństwa  $P(\boldsymbol{\theta}) \times P(d\mathbf{x}_0) \times P(d\sigma)$ .

Jedną z największych zalet stosowania metody Monte-Carlo jest asymptotyczna zbieżność estymatora (4.49) do prawdziwej wartości całki (4.48). Z praktycznego punktu widzenia, zbieżność asymptotyczna nie odgrywa znaczącej roli, gdyż zawsze interesuje nas przybliżenie otrzymywane na podstawie skończonej liczby prób losowych. W związku z tym kluczowego znaczenia nabiera kwestia kontrolowania niepewności związanej ze zmienną losową (4.48). Mimo, iż jej rozkład przeważnie pozostaje nieznanym, można zastosować jedno z podstawowych twierdzeń uczenia statystycznego tzw. nierówność Chernoff'a [121], która w rozważanym przypadku wyraża się poprzez poniższą zależność:

$$P^k \left\{ \boldsymbol{\theta}_N \in \Theta_N : \left| E[H_A(\boldsymbol{\xi})] - \hat{E}[H_A(\boldsymbol{\xi})] \right| > \varepsilon \right\} \leq 2 \exp(-2 k \varepsilon^2), \quad (4.50)$$

gdzie  $\varepsilon > 0$ . W ten sposób, jeśli chcemy mieć pewność, iż wartość (4.48) będzie zwierać się w przedziale  $[\hat{E}[H_A(\boldsymbol{\xi})] - \varepsilon, \hat{E}[H_A(\boldsymbol{\xi})] + \varepsilon]$  na poziomie ufności  $1 - \delta$ , wówczas estymator (4.49) powinien być wyznaczony na podstawie  $k$  niezależnych próbek losowych  $\{\sigma^{(i)}, \mathbf{x}_0^{(i)}, \boldsymbol{\theta}^{(i)}\}_{i=1}^k$ , gdzie:

$$k \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right) \quad (4.51)$$

## Badania eksperymentalne

Konstrukcja ogólnej przestrzeni przeszukiwań oraz rozważania dotyczące szacowania oczekiwanej skuteczności optymalizacyjnej strategii ewolucyjnej  $(1 + 1)ES_\alpha$ , pozwalają odnieść się do kwestii odporności eksploracyjnych rozkładów stabilnych. Zdecydowano się na przeprowadzenie badań eksperymentalnych dla czterech najbardziej charakterystycznych indeksów stabilności t.j.:  $(1 + 1)ES_{2.0}$ ,  $(1 + 1)ES_{1.5}$ ,  $(1 + 1)ES_{1.0}$  and  $(1 + 1)ES_{0.5}$ . Dla porównania i zobrazowania skali problemów optymalizacyjnych utożsamianych przez uogólnioną przestrzeń przeszukiwań, zamieszczono również wyniki badań uzyskane dla standardowych metod deterministycznych zaimplementowanych w bibliotece numerycznej IMSL Fortran 90 MP Library: quasi-Newton'a, Newton'a oraz Nelder-Meada. Scenariusz przebiegu eksperymentu zakładał zdefiniowanie dwuwymiarowej uogólnionej przestrzeni przeszukiwań określonej na zbiorze  $\mathbb{S} = [-10, 10] \times [-10, 10]$ . W pierwszej kolejności, przyjęte rozkłady (4.34)-(4.36) posłużyły do zdefiniowania środowiska optymalizacyjnego. Następnie, zgodnie z przyjętymi wcześniej ustaleniami, wygenerowane zostały warunki początkowe dla rozważanych algorytmów. Należy przy tym zaznaczyć, iż wszystkie algorytmy inicjowane były z tego samego punktu przestrzeni przeszukiwań  $\mathbf{x}_0$ . Algorytmy ewolucyjne były zatrzymywane jeśli osiągnięto zadowalającą jakość rozwiązania, lub gdy zaobserwowano stagnację algorytmu, t.j. w momencie gdy w stu kolejnych iteracjach

rozwiązanie nie uległo poprawie. Warunki stopu dla algorytmów deterministycznych precyzyjnie zostały opisane w [25]. Nierówność (7.29) posłużyła do zdefiniowania przedziału ufności określonym przez  $\varepsilon = 0.005$ ,  $1 - \theta = 0.95$ , co wymusiło powtórzenie całej powyżej opisanej procedury  $k = 73778$  razy. Warto również zauważyć, że przyjęty wskaźnik jakości (4.41) wymaga znajomości maksymalnej wartości funkcji celu, która w przypadku funkcji testowych postaci (4.32), nie jest możliwa do wyznaczenia korzystając z metod analitycznych. Problem ten, w trakcie realizacji zdecydowano się rozwiązać stosując gradientowy algorytm optymalizacji lokalnej uruchamiany z każdego wylosowanego punktu  $\mathbf{m}_i$ . Największą wartość, uzyskaną w ten sposób przyjmowano jako punkt odniesienia do zdefiniowania wskaźnika jakości pozostałych algorytmów. Wartości estymatorów (4.49), uzyskanych dla porównywanych algorytmów, zamieszczone zostały w tabeli 4.4. Dodatkowo średnie liczby epok potrzebne do zakończenia pracy algorytmów odnotowane zostały w tabeli 4.5.

Tabela 4.4: Oczekiwana skuteczność optymalizacyjna wyznaczona dla strategii ewolucyjnych  $(1 + 1)ES_\alpha$  z indeksami stabilności  $\alpha = \{2.0, 1.5, 1.0, 0.5\}$  oraz algorytmów deterministycznych:  $A_1$  - metoda quasi-Newtonowska z analityczną postacią gradientu,  $A_2$  - algorytm Neldera-Meada,  $A_3$  - zmodyfikowana metoda Newton'a z analityczną definicją Hessianu. Eksperyment przeprowadzony został, dla ogólnej przestrzeni przeszukiwań zdefiniowanej za pomocą różnych konfiguracji  $\lambda_w, \lambda_b, \lambda_N$

$\lambda_w$	$\lambda_b$	$\lambda_N$	$\alpha = 2.0$	$\alpha = 1.5$	$\alpha = 1.0$	$\alpha = 0.5$	$A_1$	$A_2$	$A_3$
10	0.01	0.01	0.999	0.999	0.999	1.000	1.000	1.000	1.000
1	1	0.1	0.977	0.986	0.991	0.997	0.784	0.998	0.784
1	1	1	0.906	0.917	0.928	0.941	0.723	0.931	0.724
1	1	5	0.621	0.646	0.692	0.745	0.516	0.679	0.519
1	0.1	10	0.824	0.850	0.891	0.933	0.814	0.903	0.806
5	5	10	0.468	0.492	0.535	0.587	0.375	0.505	0.380
10	10	10	0.435	0.455	0.486	0.526	0.350	0.468	0.354
20	20	20	0.372	0.393	0.426	0.467	0.289	0.399	0.291

Tabela 4.5: Średnia liczba obliczeń funkcji celu niezbędna do osiągnięcia wyników prezentowanych w Tabeli 4.4. Wyniki zaokrąglone zostały do najbliższej liczby liczby całkowitej.

$\lambda_w$	$\lambda_b$	$\lambda_N$	$\alpha = 2.0$	$\alpha = 1.5$	$\alpha = 1.0$	$\alpha = 0.5$	$A_1$	$A_2$	$A_3$
10	0.01	0.01	352	350	291	297	30	75	141
1	1	0.1	257	259	241	273	41	84	63
1	1	1	261	265	245	275	41	84	65
1	1	5	274	279	261	295	48	86	84
1	0.1	10	295	303	274	308	47	84	149
5	5	10	250	256	245	286	43	93	68
10	10	10	239	245	239	278	42	97	61
20	20	20	231	240	233	280	42	97	59

**Obserwacja 17** *Klasyfikując rozpatrzone przypadki ogólnej przestrzeni przeszukiwań, warto zauważyć, że badania przeprowadzone zostały dla skrajnie różnych środowisk optymalizacyjnych. Mianowicie, pierwszy wiersz tabel 4.4- 4.5 odnosi się do sytuacji, w której środowisko optymalizacyjne składa się prawie wyłącznie z jednej funkcji gaussowskiej o dużej rozpiętości<sup>5</sup>. Można zatem przypuszczać, iż dla dobrze uwarunkowanych problemów optymalizacji lokalnej, otrzymamy wyniki zbliżone do tych prezentowanych w pierwszym wierszu tabel 4.4- 4.5. Nie powinno zatem budzić niczyich wątpliwości, iż dla tego obszaru funkcji ciągłych, algorytmami dedykowanymi<sup>6</sup> są rozwiązania deterministyczne bazujące na informacji w postaci gradientu lub Hessianu funkcji celu. Przewaga wspomnianych rozwiązań bardzo szybko topnieje wraz ze wzrostem prawdopodobieństwa wystąpienia więcej niż jednego optimum lokalnego. Środowiska multimodalne bowiem, wyraźnie zarysowują przewagę algorytmów stochastycznych nad zachłannymi metodami gradientowymi. Odnosząc się do wyników otrzymanych dla rozważanych strategii ewolucyjnych, rzeczą znamionną jest obserwowalny porządek jaki narzuca indeks stabilności. Okazuje się bowiem, iż czym mniejszy parametr  $\alpha$ , tym skuteczniejszy algorytm optymalizacji. Ponadto, warto zauważyć, że strategia optymalizacyjna zaburzająca rozwiązania izotropowym rozkładem stabilnym  $\alpha = 0.5$  wyraźnie dominuje nad pozostałymi algorytmami dla większości testowanych obszarów klasy funkcji ciągłych.*

## Podsumowanie

W rozdziale tym poruszona została kwestia odporności stabilnej mutacji izotropowej w zadaniach optymalizacji globalnej. Przedstawione zostało pojęcie ogólnej przestrzeni przeszukiwań, która może być postrzegana jako, ujęta w probabilistyczny sposób, klasa funkcji ciągłych. Środowiska to, zostało następnie wykorzystane do badań nad odpornością strategii ewolucyjnej na niepewność związaną z wyborem optymalnego parametru skali mutacji stabilnej. Wpływ parametrów niezbędnych do zdefiniowania ogólnej przestrzeni przeszukiwań, można w obrazowy sposób przyrównać do przemieszczania obszaru o największej intensywności próbkowania w obrębie klasy funkcji ciągłych. Zatem zmiana

<sup>5</sup>zauważmy, że wartość oczekiwana parametru  $b$  równa jest 0.01

<sup>6</sup>odznaczającymi się najwyższą skutecznością

wspomnianych parametrów, powoduje również zmianę charakteru środowiska optymalizacyjnego. W ten sposób, przedmiotem badań prezentowanych w rozdziale, stały się np. dobrze uwarunkowane problemy jednomodalne jak i również problemy o wielu optimach lokalnych z wąskimi basenami przeciągania. Generalny wniosek jaki nasuwa się po analizie wyników badań eksperymentalnych, prezentowanych w niniejszym rozdziale, może zostać sformułowany następująco: w przypadku braku jakiegokolwiek wiedzy na temat rozwiązywanego problemu, zastosowanie rozkładów izotropowych o niskich indeksach stabilności, znacznie zwiększa szansę na powodzenie procesu optymalizacji. Odnosząc się do obrazowego przykładu dobierania optymalnego parametru skali rozkładu, można przypuszczać, że właśnie taki wybór modelu przeszukiwań, pozwoli znacząco zmniejszyć liczbę uruchomień algorytmu niezbędną do otrzymania satysfakcjonującego rozwiązania.

#### 4.2.5 Adaptacja izotropowego rozkładu stabilnego

Izotropowe rozkłady eksploracyjne są jednymi z najprostszych mechanizmów, jakie znajdują zastosowanie w stochastycznych technikach optymalizacji globalnej. Biorąc pod uwagę fakt, iż klasa rozkładów izotropowych jest równoliczna z klasą dodatnich zmiennych losowych, generalnie wektory o sferycznej symetrii mogą być określone za pomocą różnej liczby parametrów. W praktyce jednakże, większość znanych generatorów [38] wykorzystuje jedynie od jednego do trzech parametrów [38], [104] (np. rozkład normalny określony jest tylko przez parametr skali  $\sigma$ , rozkłady stabilne dodatkowo wymagają zdefiniowania indeksu stabilności  $\alpha$ ). Niezależnie jednak od liczby stopni swobody rozkładu izotropowego, należy podkreślić fakt, iż model ten umożliwia reprezentację tylko jednego typu korelacji występujących między zmiennymi decyzyjnymi a funkcją celu. Mianowicie, wykorzystując model probabilistyczny o sferycznej symetrii można jedynie odzwierciedlić zależność pomiędzy wartościami funkcji celu a odległością od środka symetrii. Fakt ten znaczenie ogranicza możliwości eksploracyjne algorytmu ewolucyjnego, co szczególnie staje się widoczne w środowiskach optymalizacyjnych gdzie występują znacznie bardziej skomplikowane zależności. Niemniej jednak, jak to już zostało wielokrotnie nadmienione w treści rozprawy, rozkłady te, należy stosować w początkowej fazie działania algorytmu ewolucyjnego. Zabieg taki umożliwia zminimalizowanie ryzyka błędnego ukierunkowania procesu ewolucji w późniejszej fazie optymalizacji.

Mając na uwadze zastosowanie stabilnych rozkładów izotropowych w ewolucyjnych algorytmach optymalizacji globalnej, należy oczywiście zmierzyć się z problemem wyboru odpowiedniego parametru skali oraz indeksu stabilności. W przypadku braku wiedzy o optymalizowanym problemie, najczęściej stosowanym podejściem do tego problemu, jest metoda polegająca na wielokrotnym testowaniu różnych konfiguracji parametrów, a następnie wyborze najlepszej z nich. Należy jednak pamiętać, iż wybór, dokonywany w podobny sposób, zawsze obarczony będzie dużą niepewnością (szczególnie jeśli liczba prób jest niewielka). Ponadto, metoda prób i błędów nie zawsze jest możliwa do przeprowadzenia. Co więcej, fakt iż rozkład izotropowy może uwzględniać tylko jeden i to bardzo prosty typ korelacji, nie oznacza, że nie można poprawić jego efektywności w zadaniach optymalizacji globalnej. Przeszukiwanie przestrzeni rozwiązań za pomocą algorytmów ewolucyjnych, jak wiadomo, polega na przemieszczaniu populacji rozwiązań w kierunku obszarów o lepszym dopasowaniu. W trakcie tej

wędrówki, zmienia się krajobraz optymalizacyjny, a co za tym idzie zmieniają się zależności pomiędzy zmiennymi decyzyjnymi i funkcją celu. Można mnożyć przykłady, dla których w jednym obszarze przestrzeni, lepsze efekty dają niewielkie zaburzenia osobników, natomiast w innych obszarach do poprawienia średniego dopasowania populacji niezbędne są mutacje o większym zasięgu. Bardziej namacalnego powodu, który przemawia za zamianami rozkładu izotropowego w miarę działania algorytmu ewolucyjnego, dostarczają wyniki badań eksperymentalnych zamieszczone w rozdziale 4.2.1. Bez starty ogólności, można bowiem stwierdzić, iż jedynie parametr  $\sigma$  należący do tzw. okna ewolucyjnego, może zapewnić sukcesywną poprawę jakości rozwiązania. Z drugiej zaś strony, ustalanie nieodpowiednich wartości parametru skali mutacji, może prowadzić do dramatycznego wzrostu nakładu liczby generacji niezbędnej do rozwiązania problemu. W świetle przytoczonych faktów, niezbędnym staje się wzbogacenie procedury optymalizacyjnej o technikę adaptacji rozkładu eksploracyjnego.

### Procedura adaptacyjna strategii $(1 + 1)\text{ES}_\alpha$

Indeks stabilności odgrywa kluczową rolę w klasie rozważanych rozkładów. W zależności od wartości parametru  $\alpha$  rozkłady prawdopodobieństwa zyskują zupełnie inne właściwości statystyczne. Można się o tym przekonać, dokonując przeglądu twierdzeń oraz definicji zamieszczonych we wstępnych rozdziałach rozprawy, gdzie zwykle parametr ten ustala granicę obowiązywania odpowiednich praw statystycznych. Z tego powodu, należy się spodziewać (na co zwrócił uwagę po raz pierwszy Rudolph [104]), iż procedura adaptacyjna dedykowana klasie stabilnej, musi znacznie różnić się dla różnych indeksów stabilności. W celu uproszczenia procesu projektowania procedury adaptacyjnej, rozważania w dalszej części rozdziału, dotyczyć będą jedynie pewnych charakterystycznych wartości  $\alpha$ .

Pierwszym podejściem do problemu adaptacji skali izotropowego rozkładu eksploracyjnego, była zaproponowana przez Rechenberg'a, tzw. reguła 1/5 sukcesu [101]. Podejście to, wykorzystuje fakt istnienia bardzo silnej zależności pomiędzy współczynnikiem zbieżności lokalnej a tzw. prawdopodobieństwem sukcesu [20]. Pojęcie prawdopodobieństwa sukcesu  $P_s$ , odgrywa bardzo ważną rolę w analizie efektywności stochastycznych algorytmów optymalizacji globalnej. W ogólności, pojęcie to odnosi się do prawdopodobieństwa poprawienia jakości rozwiązania poprzez zaburzenie addytywne, t.j.:

$$P_s = P(\phi(\mathbf{x}) > \phi(\mathbf{x} + \sigma \mathbf{X}^s)), \quad (4.52)$$

gdzie  $\mathbf{x}_k$  jest punktem przestrzeni rozwiązań,  $\phi(\cdot)$  funkcją celu a  $\mathbf{X}^s$  jest izotropowym wektorem zaburzającym. Dokładne wyznaczenie (4.52) jest możliwe jedynie dla pewnych specyficznych przypadków [20, 104]. Badania analityczne dotyczące współczynnika zbieżności, przeprowadzone dla funkcji sferycznej oraz korytarzowej [18, 22, 104] wykazały, iż w obu przypadkach, strategia ewolucyjna  $(1 + 1)\text{ES}$  osiąga najszybszą zbieżność dla zbliżonych wartości prawdopodobieństwa sukcesu  $P_s \approx 0.184$  [101]. Stało się to podstawą do ustalenia reguły adaptacyjnej, w której prosta heurystyka ma za zadanie utrzymanie (4.52) na poziomie 0.2. Mimo, że przyjęta reguła znajduje gruntowane uzasadnienie w przypadku prostych środowisk optymalizacyjnych, wyniki te trudno odnieść do bardziej skomplikowanych przypadków. Ponadto, procedura ta została zaprojektowana pod kątem rozkładu normalnego, który jak wiemy z poprzednich rozdziałów,

znacznie różni się od innych przedstawicieli klasy wektorów stabilnych. Ogólnie, heurystyka adaptacyjna 1/5-sukcesu, sprowadza się do następującej reguły:

$$\sigma_{k+1} = \begin{cases} \sigma_k/a & \text{if } \hat{P}_s(\sigma_k) > p_s \\ \sigma_k a & \text{if } \hat{P}_s(\sigma_k) < p_s \\ \sigma_k & \text{if } \hat{P}_s(\sigma_k) = p_s \end{cases}, \quad (4.53)$$

gdzie  $\hat{P}_s(\sigma_k)$  jest wartością prawdopodobieństwa sukcesu, estymowaną w trakcie działania algorytmu dla skali  $\sigma_k$ ,  $a$  określa dodatkowy parametr określający wielkość zmian oraz  $P_s$  jest optymalnym prawdopodobieństwem sukcesu dla rozważanego rozkładu eksploracyjnego. Wielkość  $\hat{P}_s(\sigma_k)$  jest wyznaczana na podstawie obserwacji wyników  $m$  następujących po sobie mutacji tj.:

$$\hat{P}_s(\sigma_k) = \frac{1}{m} \sum_{i=1}^m I(\mathbf{x}_{k+i} + \sigma_k \mathbf{Z}_i), \quad (4.54)$$

gdzie  $I(\cdot)$  jest funkcją wskaźnikową, przyjmującą wartość 1 jeśli rozwiązanie w iteracji  $i$  uległo poprawie oraz 0 w przeciwnym razie.

Naturalne pytanie, jakie rodzi się pod wpływem ogólnej heurystyki adaptacyjnej (4.53), dotyczy optymalnych wartości parametrów metody, w przypadku zastosowania jej do klasy izotropowych mutacji stabilnych. Odpowiedź na postawione wyżej pytanie, nie jest oczywista, gdyż jak wiadomo [23], nawet w przypadku rozkładu normalnego, optymalne wartości parametrów  $\{a, m, p_s\}$ , ulegają zmianom dla różnych problemów optymalizacyjnych. Dlatego też, dalsza część niniejszego rozdziału, poświęcona zostanie próbie skonfigurowania procedury (4.53) dla rozkładów stabilnych, w sposób zapewniający najwyższą możliwą efektywność strategii  $(1+1)\text{ES}_\alpha$  dla najbardziej charakterystycznych środowisk optymalizacyjnych.

### Badania symulacyjne

Środowisko testowe, które posłuży do optymalnego skonfigurowania heurystyki adaptacyjnej (4.53) dla izotropowych rozkładów stabilnych, składa się z dziewięciu funkcji testowych, t.j.:

- dobrze uwarunkowanego zadania optymalizacji lokalnej:  $\phi_1$ ,
- źle uwarunkowanych zadań optymalizacji lokalnej:  $\phi_9, \phi_5$ ,
- nieróżniczkowalnej funkcji jednomodalnej:  $\phi_6$ ,
- nieograniczonej funkcji liniowej:  $\phi_8$ ,
- funkcji z wieloma optimami lokalnymi:  $\phi_{13}, \phi_{14}, \phi_{15}$ .

Rozważany zestaw problemów testowych  $\Phi = \{\phi_1, \phi_5, \phi_8, \phi_9, \phi_{13}, \phi_{14}, \phi_{15}\}$  był przedmiotem badań wielu cennych publikacjach [20, 136, 135], a ich szczegółowa definicja zamieszczona została w dodatku H. Dobór funkcji użytych w eksperymencie nie był zupełnie przypadkowy. Wymienione funkcje utożsamiają bowiem najczęściej spotykane problemy, jakie można napotkać w praktyce obliczeń inżynierskiej. Co więcej, warto zauważyć, że w zestawie funkcji testowych, zostały

umieszczone obok siebie problemy, które z jednej strony wymagają szybkiego zawężania rozkładu eksploracyjnego ( $\phi_1, \phi_5$ ), z drugiej zaś preferujących zaburzenia o dużej sile. Pamiętając o tym, że nadrzędnym celem badań jest optymalne skonfigurowanie heurystyki optymalizacyjnej, zabieg ten, ma umożliwić wybór wartości  $\{a, m, p_s\}$ , które w jak największym stopniu pozwoliłyby na ustaleniu kompromisu pomiędzy dwoma przeciwstawnymi problemami: eksploracją a eksploatacją przestrzeni rozwiązań. Do weryfikacji słuszności końcowych wniosków, posłużyć mają problemy wielomodalne, dla których sukces optymalizacyjny zagwarantowany jest jedynie przez algorytm, mogący w znacznej mierze pogodzić wspomniane dwie cechy.

### Przebieg eksperymentu

Eksperyment ma na celu wyłonienie optymalnej konfiguracji heurystyki adaptacyjnej (4.53) dla następujących indeksów stabilności:  $\alpha = 2.0, 1.75, 1.5, 1.25, 1.0, 0.75, 0.5$ . Z racji, iż stabilne rozkłady o sferycznej symetrii, posiadają skrajnie odmienne właściwości statystyczne, należy spodziewać się, iż również adaptacja ich parametru skali musi przebiegać w diametralnie różny sposób. Przypuszczenie to zaowocowało przeprowadzeniem analizy skuteczności procedur adaptacyjnych, dla dosyć szerokiego zakresu parametrów zewnętrznych  $\{a, m, p_s\} \in A \times M \times P_s$ :

$$A = \{0.5, 0.6, 0.7, 0.8, 0.9, 0.95\} \quad (4.55)$$

$$M = \{2, 4, 8, 16, 32, 64\} \quad (4.56)$$

$$P_s = \{0.1, 0.15, 0.2, 0.25, 0.30, 0.35, 0.4\} \quad (4.57)$$

Eksperyment zakładał przeprowadzenie badań dla każdej kombinacji wymienionych wartości parametrów oraz każdego indeksu stabilności. W rezultacie, zadanie to wymagało sprawdzenia skuteczności strategii  $(1+1)ES_\alpha$  dla  $7 \times 6 \times 6 \times 7 = 1764$  różnych konfiguracji algorytmu. Co więcej, każda konfiguracja testowana była osobno dla wszystkich problemów w zbiorze testowym. Pamiętając o tym, iż rezultat działania algorytmu ewolucyjnego, jest w istocie, zmienną losową, należy przyjąć jakiś wskaźnik jakości określony na rozkładzie wyników. W eksperymencie zdecydowano się na dwukryterialny wskaźnik oceny każdej konfiguracji w postaci mediany końcowej wartości funkcji celu, oraz mediany liczby generacji potrzebnych do zatrzymania algorytmu. Choć jak wiadomo, mediana jest statystyką znacznie bardziej odporną wystąpienie skrajnych wartości cechy statystycznej, to i tak zdecydowano się na przeprowadzenie 100 niezależnych uruchomień algorytmu ewolucyjnego, na podstawie których wyznaczono odpowiednie wartości median. W ten sposób, prezentowane poniżej wyniki wymagały 1411200 niezależnych uruchomień strategii ewolucyjnej  $(1+1)ES_\alpha$ .

Przeprowadzenie pojedynczego testu przebiegało w następujący sposób. Algorytmy  $(1+1)ES_{2.0}, (1+1)ES_{1.75}, (1+1)ES_{1.5}, (1+1)ES_{1.25}, (1+1)ES_{1.0}, (1+1)ES_{0.75}, (1+1)ES_{0.5}$  inicjowane były z tego samego punktu przestrzeni rozwiązań, z identycznym, początkowym parametrem skali  $\sigma = 1$  oraz zatrzymywane po spełnieniu jednego z trzech warunków stopu:

I.) znaleziono rozwiązanie o akceptowalnej jakości. Dla problemu nieograniczonego z dołu było to  $\phi_8(\mathbf{x}) < -10^{10}$ , dla pozostałych funkcji ustalono próg o wartości  $10^{-5}$ ,

II. liczba obliczeń funkcji celu przekroczyła dopuszczalną wartość  $T_{\max} = 10000$ ,



**III.)** wartość parametru skali przekroczyła próg uniemożliwiający skuteczne przeszukiwanie przestrzeni  $\sigma < 10^{-10}$ .

W celu wskazania najlepszej konfiguracji parametrów, w zbiorze  $A \times P_s \times M$ : ustalony został porządek, według następującej reguły: zestaw  $\omega^{(1)} = \{a^{(1)}, m^{(1)}, p_s^{(1)}\}$  był uznawany za lepszy niż  $\omega^{(2)} = \{a^{(2)}, m^{(2)}, p_s^{(2)}\}$  jeśli:

- a) mediana wartości funkcji celu dla  $\omega^{(1)}$  osiągała mniejszą wartość
- b) w przypadku równych median wartości funkcji celu (sytuacja w której obydwie konfiguracje  $\omega^{(1)}, \omega^{(2)}$  spowodowały, iż wielkość ta osiągnęła wartość progową) za lepszą konfigurację przyjmowano tą, dla której mediana liczby generacji osiągała mniejszą wartość.

Dla uproszczenia analizy odpowiednich zależności wprowadźmy pomocniczą funkcję  $R(\omega, \phi)$ :

$$R : A \times M \times P_s \times \Phi \rightarrow [1, \dots, 252], \quad (4.58)$$

która zwraca numer porządkowy zestawu konfiguracyjnego  $\omega = \{a, m, p_s\}$  w ciągu ustalonym przez wyniki otrzymane dla funkcji testowej  $\phi \in \Phi$ .

### **Dedykowane heurystyki adaptacyjne - $n = 3$**

W pierwszej kolejności przyjrzymy się jak wyglądają optymalne heurystyki adaptacyjne, wyspecjalizowane do optymalizacji poszczególnych problemów testowych. W tabeli 4.6 zamieszczono najlepsze zestawy parametrów dla trójwymiarowych ( $n = 3$ ) funkcji testowych. Dodatkowo przedstawiono wartości median (Me[WFC] - wynikowej funkcji celu oraz Me[LOFC] - liczby generacji), które posłużyły do wyłonienia zestawów optymalnych.

Tabela 4.6: Optymalne wartości parametrów  $\{a, m, p_s\}$  dla poszczególnych problemów testowych. Wymiar przestrzeni przeszukiwań  $n = 3$

$$\alpha = 2$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	2	0.25	$4.5414e - 006$	132
$\phi_5$	0.95	64	0.15	11.841	10000
$\phi_6$	0.5	4	0.15	0	109
$\phi_8$	0.5	2	0.2	$-1.3159e + 010$	117
$\phi_9$	0.9	32	0.1	$8.9243e - 006$	2171.5
$\phi_{13}$	0.9	32	0.15	$7.5737e - 006$	2148
$\phi_{14}$	0.5	2	0.1	19.9668	144
$\phi_{15}$	0.5	64	0.1	0.029584	2944

$$\alpha = 1.5$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	2	0.1	$6.3974e - 006$	135.5
$\phi_5$	0.95	64	0.1	13.0852	10000
$\phi_6$	0.6	4	0.1	0	116.5
$\phi_8$	0.5	2	0.35	$-1.2051e + 010$	127
$\phi_9$	0.95	16	0.2	$9.8451e - 006$	2657
$\phi_{13}$	0.9	64	0.15	$9.4678e - 006$	3595.5
$\phi_{14}$	0.5	2	0.15	19.9668	140
$\phi_{15}$	0.6	64	0.1	0.029584	3840

$$\alpha = 1$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	2	0.2	$5.3288e - 006$	143.5
$\phi_5$	0.95	64	0.15	8.9948	10000
$\phi_6$	0.6	4	0.15	0	118
$\phi_8$	0.5	2	0.3	$-1.2402e + 010$	120
$\phi_9$	0.9	64	0.4	$5.9589e - 006$	3178
$\phi_{13}$	0.9	64	0.1	$8.8239e - 006$	3289.5
$\phi_{14}$	0.9	64	0.3	$9.0356e - 006$	5921.5
$\phi_{15}$	0.5	64	0.15	0.029584	2816

$$\alpha = 0.5$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	2	0.1	$5.5879e - 006$	184.5
$\phi_5$	0.95	64	0.3	$9.374e - 006$	7131
$\phi_6$	0.8	4	0.15	0	132.5
$\phi_8$	0.5	2	0.35	$-2.7835e + 010$	103.5
$\phi_9$	0.95	32	0.1	$8.245e - 006$	3684
$\phi_{13}$	0.9	64	0.25	$6.7042e - 006$	2728.5
$\phi_{14}$	0.5	32	0.1	$8.7294e - 006$	860.5
$\phi_{15}$	0.95	64	0.15	0.0098647	10000

**Obserwacja 18** *Porównując optymalne konfiguracje otrzymane w wyniku przeprowadzonego eksperymentu (tabela 4.6), rzeczą najbardziej rzucającą się w oczy,*

jest dominacja rozkładu o indeksie stabilności  $\alpha = 0.5$ . Jakkolwiek algorytm ten wymaga niekiedy nieco większej liczby iteracji niż np.  $(1+1)ES_2$ , to w zamian otrzymujemy strategię, której poprawne skonfigurowanie, spowoduje znalezienie optimum globalnych. Dla odmiany, algorytmy z większymi indeksami  $\alpha = 2, 1.5$ , wykazują przedwczesną zbieżność do rozwiązania (np.  $\phi_{14}$ ). Duże zróżnicowanie optymalnych parametrów dla heurystyki adaptacyjnej, może świadczyć o tym, iż zestaw testowy, faktycznie zawiera problemy o zupełnie innym charakterze. Cechą znamioną jest, iż dla funkcji testowych  $\phi_1, \phi_8$ , bez względu na indeks stabilności, najlepszą okazała się heurystyka, która ustanawia identyczną regułę: gwałtowne i częste modyfikacje parametru skali okazują się najkorzystniejsze. Skuteczność strategii ewolucyjnej dla problemów multimodalnych, uwarunkowana jest przez wolniej zachodzące modyfikacje  $\sigma$ , o czym świadczą duże wartości  $m$  optymalnej heurystyki adaptacyjnej.

Obserwując skrajnie różne konfiguracje procedury adaptacyjnej, należy przypuszczać, iż w zależności od środowiska optymalizacyjnego, ewentualna zmiana poszczególnych parametrów będzie miała również różny wpływ na skuteczność strategii ewolucyjnej. Wyniki eksperymentu, mogą oczywiście posłużyć do bliżej analizy tego typu zależności. Mianowicie traktując parametry heurystyki adaptacyjnej jako zmienne objaśniające, oraz wartości funkcji (4.58) jako zmienną objaśnianą, w tabeli 4.7 zamieszczone zostały współczynniki korelacji liniowej dla poszczególnych indeksów stabilności.

Tabela 4.7: Współczynniki korelacji liniowej pomiędzy parametrami procedury adaptacyjnej  $\{a, m, p_s\}$  a pozycją w uporządkowanym rosnąco rankingu zestawów konfiguracyjnych. Wymiar przestrzeni przeszukiwań  $n = 3$ .

Funkcja celu $\phi_1$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	0.58	0.57	0.57	0.56	0.56	0.55	0.53
$m$	0.72	0.72	0.72	0.72	0.72	0.72	0.71
$p_s$	-0.07	-0.07	-0.07	-0.08	-0.10	-0.12	-0.17

Funkcja celu $\phi_5$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	-0.37	-0.34	-0.36	-0.47	-0.23	-0.26	-0.34
$m$	-0.43	-0.47	-0.40	-0.36	-0.45	-0.35	-0.37
$p_s$	0.06	0.05	0.03	0.06	0.15	0.08	0.18

Funkcja celu $\phi_6$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	-0.00	-0.03	-0.02	0.01	-0.03	-0.08	-0.21
$m$	0.20	-0.07	-0.12	-0.09	-0.10	-0.10	-0.26
$p_s$	0.57	0.63	0.60	0.62	0.62	0.62	0.53

Funkcja celu $\phi_8$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	0.64	0.64	0.64	0.64	0.64	0.64	0.64
$m$	0.66	0.66	0.66	0.66	0.66	0.66	0.66
$p_s$	0.12	0.12	0.12	0.12	0.13	0.12	0.12

Funkcja celu $\phi_9$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	-0.38	-0.36	-0.36	-0.37	-0.33	-0.33	-0.29
$m$	-0.54	-0.55	-0.53	-0.51	-0.50	-0.46	-0.40
$p_s$	0.47	0.48	0.48	0.50	0.52	0.56	0.61

Funkcja celu $\phi_{13}$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	-0.41	-0.43	-0.42	-0.45	-0.47	-0.45	-0.43
$m$	-0.73	-0.71	-0.73	-0.72	-0.73	-0.74	-0.76
$p_s$	0.23	0.23	0.23	0.20	0.19	0.17	0.16

Funkcja celu $\phi_{14}$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	0.98	0.98	0.96	0.87	0.56	0.08	-0.14
$m$	0.15	0.15	0.14	-0.01	-0.29	-0.62	-0.61
$p_s$	0.03	0.03	0.01	0.04	0.07	0.05	0.11

Funkcja celu $\phi_{15}$							
Zmienne	$\alpha = 2$	$\alpha = 1.75$	$\alpha = 1.5$	$\alpha = 1.25$	$\alpha = 1$	$\alpha = 0.75$	$\alpha = 0.5$
$a$	-0.05	-0.22	-0.17	-0.07	0.07	-0.08	-0.38
$m$	-0.70	-0.61	-0.48	-0.34	-0.21	-0.36	-0.58
$p_s$	0.24	0.34	0.52	0.63	0.76	0.70	0.47

**Obserwacja 19** Analizując wartości współczynników korelacji liniowej zaprezentowane w tabeli 4.7 wyraźnie zarysowują się dwa rodzaje zależności. Pierwszy z nich, obserwowany dla problemów  $\phi_1, \phi_8$ , polega na dodatnim określeniu współczynników dla parametrów  $a, m$ . Oznacza to, że w miarę wzrostu wartości tych parametrów, heurystyka adaptacyjna odznacza się coraz gorszą średnią skutecz-

nością. Zupelnie odwrotną sytuację możemy zaobserwować w przypadku pozostałych problemów testowych. Mianowicie, dla algorytmów które proces optymalizacji kończyły zlokalizowaniem optimum globalnego, można zaobserwować ujemne skorelowanie tych samych parametrów. Pamiętając o tym, że mniejsze wartości  $a, m$  przekładają się na gwałtowniejsze i częściej zachodzące zamiany parametru skali  $\sigma$ , tabela 4.7 w dużym stopniu tłumaczy wartości optymalnych zestawów konfiguracyjnych - tabela 4.6.

Z wyników zaprezentowanych w tabeli 4.6 jasno wynika, iż chcąc skutecznie rozwiązać problem o wielu optimach lokalnych, należałoby zastosować strategię ewolucyjną  $(1 + 1)ES_{0.5}$  z heurystyką adaptacyjną o parametrach zbliżonych do  $\{a \approx 0.9, m \approx 64, p_s \approx 0.15\}$ . Natomiast w przypadku problemów dobrze uwarunkowanych  $\phi_1, \phi_8, \phi_5$ , można z równie dobrym skutkiem, użyć strategii ewolucyjnej z dowolnym indeksem stabilności, pod warunkiem skonfigurowania procedury adaptacyjnej w sposób  $\{a \approx 0.6, m \approx 4, p_s \approx 0.15\}$ . Zaobserwowane reguły nie mogą jednak, w żaden sposób pomóc przy rozwiązywaniu problemów optymalizacji globalnej, dla których brak wiedzy a-priori, uniemożliwia przypisanie problemu do jednej z wymienionych wyżej klas. W związku z tym rodzi się pytanie o istnienie pewnej uniwersalnej konfiguracji procedury adaptacyjnej, która mogłaby zostać rekomendowana bez dodatkowych założeń odnośnie postaci funkcji celu. Jest to zarazem pytanie o ustalenie kompromisu pomiędzy dwoma skrajnie różnymi ustawieniami heurystyki adaptacyjnej. Wspomniany kompromis może zostać ustalony, przez wybór zestawu konfiguracyjnego  $\omega^*$ , który będzie odznaczał się największą średnią efektywnością na całym zbiorze testowym  $\Phi$ . W rozważanym przypadku, do wyznaczania uniwersalnych konfiguracji, posłużono się zależnością

$$\omega^* = \arg \min_{\omega} \sum_{\phi \in \Phi} R(\omega, \phi) \quad (4.59)$$

Otrzymane w ten sposób parametry heurystyki adaptacyjnej wraz z uzyskanymi dla nich wynikami przedstawiono w tabeli 4.8.

Tabela 4.8: Konfiguracje heurystyki adaptacyjnej gwarantujące najlepszą średnią skuteczność optymalizacyjną. Wymiar przestrzeni przeszukiwań  $n = 3$ .

$$\alpha = 2$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	4	0.1	$4.5329e - 006$	199.5
$\phi_5$	0.5	4	0.1	15.7917	10000
$\phi_6$	0.5	4	0.1	0	116
$\phi_8$	0.5	4	0.1	$-1.1299e + 010$	149.5
$\phi_9$	0.5	4	0.1	0.070833	10000
$\phi_{13}$	0.5	4	0.1	16.4167	380
$\phi_{14}$	0.5	4	0.1	19.9668	264
$\phi_{15}$	0.5	4	0.1	0.2663	452

$$\alpha = 1.5$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.6	4	0.15	$4.419e - 006$	231
$\phi_5$	0.6	4	0.15	17.0921	10000
$\phi_6$	0.6	4	0.15	0	124
$\phi_8$	0.6	4	0.15	$-1.1729e + 010$	201
$\phi_9$	0.6	4	0.15	0.061305	10000
$\phi_{13}$	0.6	4	0.15	7.9597	468
$\phi_{14}$	0.6	4	0.15	19.9668	308
$\phi_{15}$	0.6	4	0.15	0.2207	536

$$\alpha = 1.0$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.5	8	0.1	$4.3124e - 006$	333.5
$\phi_5$	0.5	8	0.1	16.7882	10000
$\phi_6$	0.5	8	0.1	0	155.5
$\phi_8$	0.5	8	0.1	$-1.2056e + 010$	254
$\phi_9$	0.5	8	0.1	$1.4115e - 005$	10000
$\phi_{13}$	0.5	8	0.1	1.9899	752
$\phi_{14}$	0.5	8	0.1	19.9668	512
$\phi_{15}$	0.5	8	0.1	0.12202	912

$$\alpha = 0.5$$

$\phi(\cdot)$	a	m	$p_s$	Me[WFC]	Me[LOFC]
$\phi_1$	0.7	16	0.1	$3.9677e - 006$	419
$\phi_5$	0.7	16	0.1	10.5552	10000
$\phi_6$	0.7	16	0.1	0	157.5
$\phi_8$	0.7	16	0.1	$-2.5629e + 010$	697.5
$\phi_9$	0.7	16	0.1	$9.761e - 006$	8179.5
$\phi_{13}$	0.7	16	0.1	1.4924	1344
$\phi_{14}$	0.7	16	0.1	$9.1996e - 006$	995
$\phi_{15}$	0.7	16	0.1	0.10726	1584

**Obserwacja 20** *Obserwując wyniki uzyskane dla uniwersalnych konfiguracji (tabela 4.8), znamienne jest podtrzymanie tendencji polegającej na dominacji*

algorytmu z indeksem stabilności  $\alpha = 0.5$ . Naturalnie, odejście od optymalnych konfiguracji (tabela 4.6) musiało spowodować pogorszenie rezultatów, lecz dla wspomnianej strategii, obserwowany spadek efektywności był najmniej zauważalny. Analiza heurystyk adaptacyjnych, dających najlepsze średnie wyniki, pozwala na ustalenie następującej reguły: im mniejszy indeks stabilności tym mniej gwałtowne powinny być zmiany parametru skali rozkładów eksploracyjnych. Reguła ta, znajduje swoje potwierdzenie w tabeli 4.8, gdzie uniwersalne zestawy konfiguracyjne, w miarę zmniejszania  $\alpha$ , składają się z coraz większych wartości  $a$  oraz  $m$ .

## Podsumowanie

W rozdziale tym, przedstawione zostały wyniki badań nad dostosowaniem słynnej reguły 1/5 sukcesu do izotropowych rozkładów stabilnych. Jakkolwiek klasa wektorów stabilnych zawiera rozkład normalny ( $\alpha = 2$ ), dla którego wspomniana procedura adaptacyjna została pierwotnie zaprojektowana, to nie może ona w bezpośredni sposób zostać uogólniona na inne rozkłady tej klasy. Główną przyczyną są diametralnie odmienne właściwości statystyczne uwarunkowane przez indeks stabilności  $\alpha$ . Dlatego też, w rozdziale tym zaprezentowane zostały eksperymentalne badania mające na celu wyłonienie optymalnych zestawów konfiguracyjnych heurystyki adaptacyjnej dla pewnych charakterystycznych przedstawicieli klasy rozkładów stabilnych.

Mniejsza wrażliwość na parametr skali  $\sigma$  obserwowana dla rozkładów o cięższych ogonach, przekłada się również na osłabienie związku pomiędzy parametrami heurystyki a skutecznością strategii ewolucyjnej. Umożliwia to wyznaczenie bardziej efektywnych uniwersalnych zestawów konfiguracyjnych. Wyniki badań, skłaniają do ogólnej obserwacji, iż mniejszy indeks stabilności rozkładu eksploracyjnego, wymaga mniejszej częstotliwości zmian oraz mniejszego zakresu modyfikacji parametru  $\sigma$ . Obydwie cechy, przemawiają za obraniem takiej strategii, której celem jest znacznie wolniejsze zawężanie rozkładu eksploracyjnego, niż ma to miejsce chociażby w przypadku rozkładu normalnego.

Traktując wyniki eksperymentu jako wyznacznik mogący posłużyć do konfiguracji heurystyki adaptacyjnej, należy jednak zwrócić uwagę na bardzo istotny fakt. Mianowicie, w wielu przypadkach, otrzymane optymalne konfiguracje zawierają graniczne wartości testowanych parametrów  $\{a, m, p_s\}$ . W związku z tym należy wziąć pod uwagę możliwość istnienia jeszcze lepszych konfiguracji, które niestety nie zostały uwzględnione w przeprowadzonym eksperymencie. Drugą kwestią, która do tej pory nie została poruszona jest kwestia uogólnienia wyciągniętych wniosków na przestrzenie o większej liczbie wymiarów. Wziąwszy pod uwagę nieliniową zmianę charakterystyk izotropowych rozkładów stabilnych (rozdział 4.2.2 - rysunek 4.8) nie należy się spodziewać, iż będzie to proste zadanie. Z pewnością prawdziwe okażą się ogólne hipotezy o wolniejszym zawężaniu rozkładów ciężkoogonowych, lecz parametry optymalnych konfiguracji z pewnością nie będą podlegać liniowemu przekształceniu względem  $n$ . Zagadnienia te znacznie wykraczają jednak poza ramy niniejszej rozprawy. Z pewnością jednak posłużą do wyznaczenia kierunku przyszłych badań nad tą tematyką.

### 4.3 Mutacja bazująca na dyskretnej mierze spektralnej

Dyskretna miara spektralna (DMS) pozwala na sparametryzowanie wielowymiarowego rozkładu stabilnego, a przez to na znaczne uproszczenie posługiwania się tym pojęciem w kontekście jego licznych zastosowań praktycznych [83], [82]. Definicja DMS wymaga określenia skończonego zbioru punktów, tzw. wektorów rozpinających DSM  $\boldsymbol{\xi} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{n_s}\}$ ,  $\mathbf{s}_i \in \partial S^{(d)}$  oraz skojarzonych z nim wektora wag  $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_{n_s}\}$ ;  $\gamma_i \in R_+ \cup \{0\}$ . W ten sposób, wektor losowy, określony za pomocą  $\boldsymbol{\xi}$  i  $\boldsymbol{\gamma}$  oznaczać będziemy symbolem  $\mathbf{X}_{\boldsymbol{\xi}}^{\boldsymbol{\gamma}}$ . Wykorzystując definicję DSM, miara probabilistyczna każdego podzbioru  $A \subset \partial S^{(d)}$  jest wyrażona w następujący sposób:

$$\Gamma(A; \boldsymbol{\xi}, \boldsymbol{\gamma}) = \sum_{i=1}^{n_s} \gamma_i I_A(\mathbf{s}_i) \quad (4.60)$$

gdzie  $I_A(\cdot)$  jest funkcją wskaźnikową zbioru  $A$ . Funkcja charakterystyczna stabilnego wektora losowego  $\mathbf{X}_{\boldsymbol{\xi}}^{\boldsymbol{\gamma}}$  opisanego za pomocą dyskretnej miary spektralnej (4.60), sprowadza się do następującej postaci [83]:

dla  $\alpha \neq 1$

$$\varphi(\mathbf{k}) = \exp\left(-\sum_{i=1}^{n_s} \gamma_i |\mathbf{k}^T \mathbf{s}_i|^\alpha \left(1 - j \operatorname{sign}(\mathbf{k}^T \mathbf{s}_i) \tan\left(\frac{\pi\alpha}{2}\right)\right) + j \mathbf{k}^T \boldsymbol{\mu}_0\right) \quad (4.61)$$

dla  $\alpha = 1$

$$\varphi(\mathbf{k}) = \exp\left(-\sum_{i=1}^{n_s} \gamma_i |\mathbf{k}^T \mathbf{s}_i| \left(1 - i \frac{2}{\pi} \operatorname{sign}(\mathbf{k}^T \mathbf{s}_i) \ln |\mathbf{k}^T \mathbf{s}_i|\right) + j \mathbf{k}^T \boldsymbol{\mu}_0\right) \quad (4.62)$$

#### Wybrane właściwości oraz pojęcia

Definicja dyskretnej miary spektralnymi, znacznie upraszcza posługiwanie się wielowymiarowymi rozkładami stabilnymi w algorytmach ewolucyjnych. Należy przy tym zaznaczyć, że zawężenie obszaru zainteresowania do rozkładów określonych przez DSM, teoretycznie nie powoduje utraty ogólności, o czym przekonuje twierdzenie [29]

**Twierdzenie 13** *Niech  $p(\mathbf{x})$  oznacza funkcję gęstości losowego rozkładu stabilnego z opisanego za pomocą funkcji charakterystycznej (2.29)-(2.30), a  $p^*(\mathbf{x})$  gęstość wektora losowego z funkcją charakterystyczną (4.61)-(4.62). Wówczas, dla każdego  $\varepsilon > 0$  istnieje takie  $n_s \in \mathbb{N}$  że istnieją zbiory  $\boldsymbol{\xi} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{n_s}\}$ ,  $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_{n_s}\}$ , takie że dla każdego  $\mathbf{x} \in \mathbb{R}^d$  zachodzi*

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |p(\mathbf{x}) - p^*(\mathbf{x})| < \varepsilon$$

Z twierdzenia (13) wynika, że każdy rozkład stabilny może zostać przybliżony przez rozkład z dyskretną miarą spektralną z arbitralnie dowolną dokładnością.



Szczególnie ważne, w kontekście zastosowań, omawianych rozkładów do zaburzeń rozwiązań, jest zagadnienie związane z istnieniem oraz kosztem obliczeniowym procedury generowania wektorów pseudo-losowych. Okazuje się wektor  $\mathbf{X}_\xi^\gamma$  można przedstawić wykorzystując dekompozycję stochastyczną [75]

$$\mathbf{X}_\xi^\gamma \stackrel{d}{=} \begin{cases} \sum_{i=1}^{n_s} \gamma_i^{1/\alpha} Z_i \mathbf{s}_i & \text{dla } \alpha \neq 1 \\ \sum_{i=1}^{n_s} \gamma_i (Z_i + \frac{2}{\pi} \ln(\gamma_i)) \mathbf{s}_i & \text{dla } \alpha = 1 \end{cases} \quad (4.63)$$

gdzie  $Z_i$  są niezależnymi zmiennymi losowymi o jednakowym rozkładzie stabilnym  $S_\alpha(1, 1, 0)$ . Stochastyczna dekompozycja (4.63) może, w bezpośredni sposób posłużyć do budowy generatorów wektorów pseudolosowych  $\mathbf{X}_\xi^\gamma$ .

Przy okazji omawiania wielowymiarowych rozkładów stabilnych (rozdział 2.2) dużą uwagę przywiązano do wektora losowego zawierającego niezależne elementy symetryczne t.j.  $\mathbf{X} = [X_1, \dots, X_n]^T$ ,  $X_i \sim S_\alpha S(\sigma)$ . W literaturze, rozkład ten doczekał się kilku zastosowań w operatorze mutacji algorytmów ewolucyjnych [65, 88]. Okazuje się, że za pomocą dyskretnej miary spektralnej, można również przedstawić rozkłady gdzie poszczególne elementy wektora stabilnego będą wzbogacone o dodatkowe parametry  $\beta, \mu$ , t.j.:  $\mathbf{X} = [X_1, \dots, X_n]^T$ ,  $X_i \sim S_\alpha S(\sigma, \beta, \mu)$ . Oznacza to, że każda współrzędna zyskuje dodatkowe stopnie swobody. Fakt ten ma ogromne znaczenie w kontekście zastosowania powyższego wektora losowego w zadaniach modelowania skomplikowanych zależności pomiędzy zmiennymi decyzyjnymi. Jak wiadomo, wykrywanie takich zależności oraz ich uwzględnienie w operatorze mutacji, zdecydowanie przyspiesza proces optymalizacji. Aby zobrazować zakres możliwości reprezentacji wektorów stabilnych w postaci DSM, wystarczy wspomnieć, że wektor z niezależnymi komponentami  $X_i \sim S_\alpha S(\sigma, \beta, \mu)$  posiada miarę spektralną skupioną w punktach przecięcia osi układu współrzędnych z powierzchnią kuli jednostkowej oraz różnymi wartościami wag. Uniwersalność reprezentacji rozkładu za pomocą dyskretnej miary spektralnej przejawia się w twierdzeniu [107]

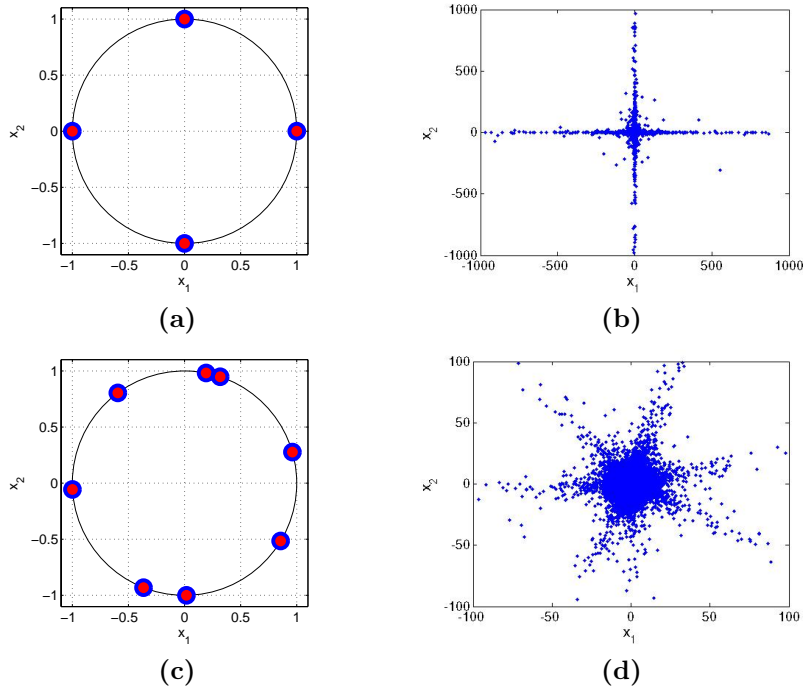
**Twierdzenie 14** *Miara spektralna stabilnego wektora  $\mathbf{X} = [X_1, X_2, \dots, X_d]^T$  jest określona za pomocą skończonej liczby wektorów rozpinających  $\mathbf{s}_i$ , wtedy i tylko wtedy, gdy wektor  $\mathbf{X}$  może zostać przedstawiony w postaci kombinacji liniowej niezależnych, stabilnych zmiennych losowych, t.j.:*

$$\mathbf{X} = \mathbf{A} \mathbf{Z}, \quad (4.64)$$

gdzie  $\mathbf{A} \in \mathbb{R}^{d \times N}$ ,  $\mathbf{Z} = [Z_1, Z_2, \dots, Z_N]^T$ ,  $Z_i \sim S_\alpha(\sigma, \beta, \mu)$ .

Z twierdzenia 14 wynika, że DSM może posłużyć także do przedstawienia wektorów, dla których poszczególne współrzędne nie tylko opisane są za pomocą trzech parametrów  $\sigma, \beta, \mu$ , lecz umożliwia wyrażenie zależności statystycznych pomiędzy nimi.

Wiele z pojęć prezentowanych w rozdziale 2.2, można bezpośrednio przełożyć na reprezentację wykorzystującą DSM. Największe znaczenie dla procedur optymalizacyjnych, zdaje się posiadać twierdzenie o asymptotycznej postaci ogonów rozkładów wielowymiarowych 10. W przełożeniu na reprezentację miary spektralnej w postaci dyskretnej zbioru punktów, oznacza to, że cała masa prawdopodobieństwa znaczenie oddalona od punktu bazowego, będzie skupiona na kierunkach wyznaczonych przez położenie wektorów  $\xi$ . Zatem makromutacje, będą występowały jedynie na kierunkach wskazywanych przez wektory rozpinające DSM. Efekt ten zobrazowany został na rysunku 4.13.



Rysunek 4.13: Rozkład wektorów rozpinających dyskretną miarę spektralną oraz odpowiadające im realizacje losowe: (a),(b) -  $\alpha = 0.75$ , (c),(d) -  $\alpha = 1.5$

Znaczenie dyskretnej miary spektralnej oraz jej wpływ na postać rozkładu prawdopodobieństwa najlepiej obrazuje wielowymiarowy rozkład normalny  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  o funkcji charakterystycznej

$$\varphi(\mathbf{k}) = \exp\left(-\frac{1}{2}\mathbf{k}^T \boldsymbol{\Sigma} \mathbf{k} + j\boldsymbol{\mu}^T \mathbf{k}\right) \quad (4.65)$$

Rozpatrzmy rozkład SVD macierzy kowariancji, t.j:  $\boldsymbol{\Sigma} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T$ . Aby funkcję (4.65) sprowadzić do postaci właściwej dla DSM (4.61) wystarczy przyjąć:

$$\mathbf{U} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n], \quad \boldsymbol{\Lambda} = \begin{bmatrix} 2\gamma_1 & 0 & \dots & 0 \\ 0 & 2\gamma_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 2\gamma_n \end{bmatrix} \quad (4.66)$$

Oznacza to, wektory własne macierzy kowariancji rozkładu normalnego są jednocześnie wektorami rozpinającymi DSM, natomiast wartości szczególne odpowiadają poszczególnym wagą DSM.

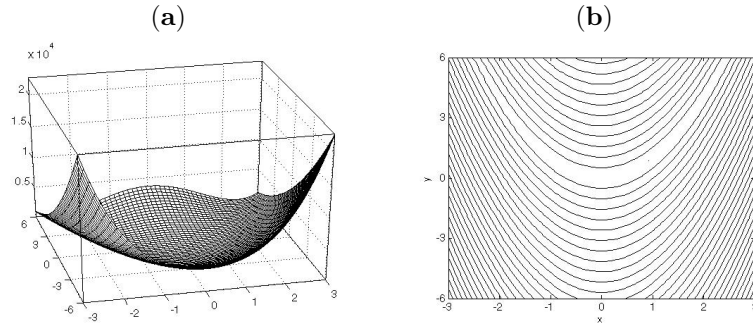
Podsumowując uwagi dotyczące wielowymiarowych rozkładów stabilnych, należałoby się zastanowić jakie rzeczywiście korzyści wynikają z ich zastosowania w algorytmach ewolucyjnych. Próbuując odnieść się do tej kwestii należy, przede wszystkim zwrócić szczególną uwagę na dwa aspekty: makromutacje - umożliwiające znacznie łatwiejsze pokonywanie szczytów ewolucyjnych oraz możliwość

modelowania złożonych zależności stochastycznych, która oznacza większą efektywność w optymalizacji lokalnej. Korzyści płynące z zastosowania rozkładów opisanych za pomocą DSM, zobrazujemy za pomocą poniższego przykładu obliczeniowego.

### Wybór optymalnego rozkładu stabilnego - przykład obliczeniowy

Załóżmy, że w niniejszym przykładzie, chcemy znaleźć optimum globalne, często stosowanej przy okazji testowania różnych technik optymalizacyjnych, dwuwymiarowej funkcji Rastrigina:

$$\phi(x, y) = (1 - x)^2 + 100(y - x^2)^2. \quad (4.67)$$



Rysunek 4.14: Funkcja Rastrigina - (a) oraz jej wykres poziomicowy - (b)

Ponadto, założymy że posiadana przez nas wiedza ogranicza się jedynie do znajomości przybliżenia rozwiązania globalnego, zlokalizowanego w punkcie  $\mathbf{x}_k = [x, y]^T = [0, -2]^T$ , o wartości funkcji celu  $\phi(\mathbf{x}_k) = 401$ . Naszym celem jest poprawienie aktualnego rozwiązania poprzez zaburzenie stabilnym wektorem losowym  $\mathbf{X}_\xi^\gamma$ . Niniejszy przykład obliczeniowy ma na celu wyłonienie optymalnego modelu stabilnego z klasy wielowymiarowych rozkładów opisanych za pomocą DSM. Na wstępie ograniczymy zbiór konkurujących modeli probabilistycznych do zbioru zawierającego cztery rozkłady stabilne  $\Omega = \{\mathbf{X}_\xi^\gamma(2.0), \mathbf{X}_\xi^\gamma(1.5), \mathbf{X}_\xi^\gamma(1.0), \mathbf{X}_\xi^\gamma(0.5)\}$ . Każdy wektor losowy  $\mathbf{X}_\xi^\gamma(\cdot)$  opisany jest za pomocą dyskretnej miary spektralnej, rozpiętej na szesnastu równomiernie rozłożonych punktach skupienia, t.j.:

$$\xi = \left\{ \begin{array}{cccccccc} [1], & [0.92], & [0.7], & [0.38], & [0], & [-0.38], & [-0.7], & [-0.92], & \dots \\ [0], & [0.38], & [0.7], & [0.92], & [1], & [0.92], & [0.7], & [0.38], & \dots \\ \dots, & [-1], & [-0.92], & [-0.38], & [-0.92], & [-1], & [-0.92], & [-0.38] \end{array} \right\}$$

W celu wyznaczenia optymalnego modelu przeszukiwań posłużymy się następującym kryterium:

$$\gamma^* = \arg \min_{\gamma \in \mathbb{R}_+^{16}} C(\gamma), \quad (4.68)$$

gdzie:

$$C(\gamma) = E \left[ \min \left\{ \frac{\phi(\mathbf{x}_k + \mathbf{X}_\xi^\gamma(\alpha))}{\phi(\mathbf{x}_k)}, 1 \right\} \right] \quad (4.69)$$

Okazuje się, że funkcję (4.69) nie sposób przedstawić za pomocą funkcji elementarnych, a co za tym idzie, zadania (4.68) nie można rozwiązać za pomocą tradycyjnych technik optymalizacji. Jednym z możliwych rozwiązań, jest zastosowanie metody Monte Carlo [57, 67]. Mianowicie, Prawo Wielkich Liczb [37] pozwala przybliżyć dokładną wartość oczekiwaną (4.69) za pomocą następującego estymatora:

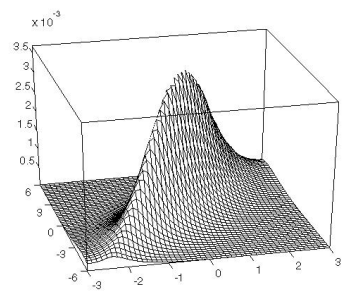
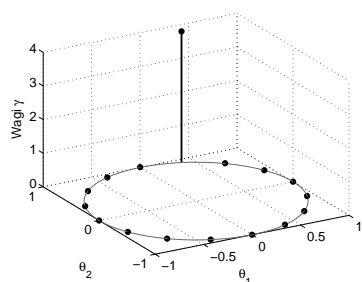
$$\hat{C}(\gamma) = \frac{1}{N} \sum_{i=1}^N \min \left\{ \frac{\phi(\mathbf{x}_k + \mathbf{X}_{i,\xi}^\gamma(\alpha))}{\phi(\mathbf{x}_k)}, 1 \right\}, \quad (4.70)$$

gdzie  $\{\mathbf{X}_{i,\xi}^\gamma(\alpha)\}_{i=1}^N$  oznacza ciąg niezależnych realizacji wektora losowego o rozkładzie  $\alpha$ -stabilnym. Zatem, wykorzystując estymator (4.70), zadanie (4.68) można przedstawić w alternatywnej formie

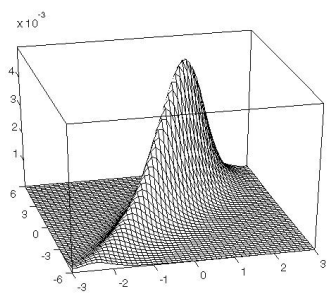
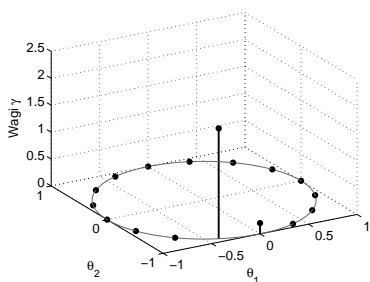
$$\gamma^* = \arg \min_{\gamma \in \mathbb{R}^{16}} \hat{C}(\gamma) \quad (4.71)$$

Kluczowym zagadnieniem przy takim przekształceniu zadań optymalizacji, jest odpowiedni wybór liczby próbek  $N$  na podstawie, których obliczana jest wartość funkcji celu. Należy przy tym podkreślić, że funkcja celu, w tym wypadku posiada charakter stochastyczny, a właśnie liczba  $N$  pozwala na kontrolowanie wariancji estymatora (4.70). Naturalnym rozwiązaniem wydaje się przyjęcie bardzo dużej liczby  $N$ , która spowoduje znaczną redukcję niepewności związanej z wyznaczaniem wskaźnikiem jakości (4.70). Z drugiej zaś strony, należy uwzględnić fakt, że proces optymalizacji zwykle wymaga wielokrotnego obliczania wartości funkcji celu, co przy dużej wartości  $N$  prowadzić może do zbyt wygórowanych nakładów obliczeniowych. W celu racjonalnego kontrolowania jakości estymatora (4.70), można zastosować, jedno z podstawowych twierdzeń teorii uczenia statystycznego, tzw. nierówność Chernoff'a [121].

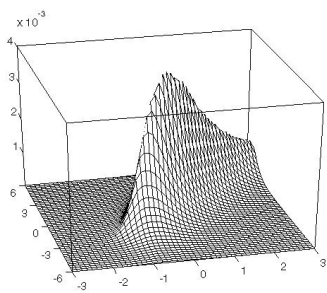
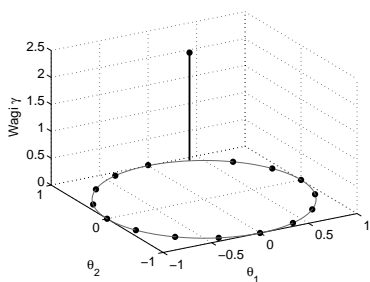
Ustalając kompromis pomiędzy jakością estymatora a nakładem obliczeniowym potrzebnym do jego wyznaczenia, w dalszym ciągu funkcja celu będzie posiadała charakter losowy. Z tego względu, standardowe techniki optymalizacji, takie jak chociażby metody oparte na numerycznym wyznaczaniu pochodnych za pomocą różnic skończonych, w rozważanym przypadku stają się bezużyteczne. Dlatego też, w celu rozwiązania zadania, posłużono się algorytmem SPSA [115, 116] który dedykowany jest do optymalizacji wielowymiarowych funkcji celu o charakterze stochastycznym. Otrzymane wyniki zaprezentowane zostały w tabeli 4.9. Dodatkowo, optymalne punkty skupienia dyskretnej masy spektralnej wraz z odpowiadającymi im funkcjami gęstości zamieszczone zostały na rysunku 4.15.



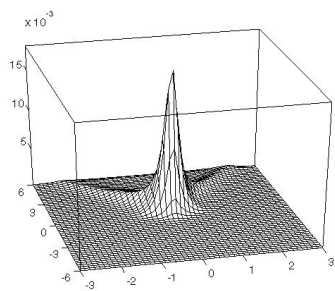
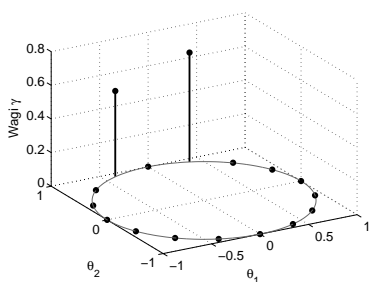
(a)



(b)



(c)



(d)

Rysunek 4.15: Pseud-optymalne rozkłady stabilne (a)  $\alpha = 2$ , (b)  $\alpha = 1.5$ , (c)  $\alpha = 1$ , (d)  $\alpha = 0.5$  dla problemu opisanego w rozdziale 4.3. Wykresy po lewej stronie przedstawiają wagi poszczególnych punktów skupienia dyskretnej miary spektralnej, wykresy po stronie prawej przedstawiają odpowiadającą im dwuwymiarową funkcję gęstości.

**Obserwacja 21** Wyniki eksperymentu w sposób jednoznaczny przemawiają na korzyść rozkładów z mniejszymi indeksami stabilności. Zaskakującym jest fakt, iż mimo ich charakteru, w znacznej mierze nastawionego na powstawanie makromutacji, mimo wszystko wykazują wyższość nad rozkładem normalnym. Główną przyczyną takiego stanu rzeczy, jest symetria rozkładu normalnego. Cecha ta powoduje, że prawdopodobieństwo sukcesu, nie może przekroczyć w większości funkcji celu wartości 0.5. Inaczej, sytuacja ta, przedstawia się w przypadku rozkładów niesymetrycznych  $\alpha < 2$ . Pozwalają one, na ukierunkowanie rozkładu eksploracyjnego na obszar o średnio mniejszej wartości funkcji celu. Ukierunkowanie rozkładu, najlepiej widoczne jest, jeśli porówna się wartości prawdopodobieństwa sukcesu zamieszczone w tabeli 4.9. Wyraźnie widać, że dla optymalnych rozkładów, prawdopodobieństwo wygenerowania rozwiązania lepszego niż  $\mathbf{x}_k = [0, -2]^T$ , wynosi ponad 0.60. Znajduje to natychmiast odzwierciedlenie w wartości przyjętego wskaźnika oceny modelu przeszukiwań.

Tabela 4.9: Wartości funkcji celu oraz odpowiadające im prawdopodobieństwa sukcesu wyznaczone dla pseudooptymalnych rozkładów stabilnych.

$\alpha$	2.0	1.5	1.0	0.5
$C(\gamma^*)$	0.6683	0.4978	0.6522	0.5385
$P_s$	0.4852	0.6445	0.5145	0.6638

### 4.3.1 Adaptacja dyskretnej miary spektralnej

Stosując algorytmy optymalizacji, nawet do najprostszych funkcji celu, bardzo szybko ujawnia się konieczność adaptacji ich parametrów konfiguracyjnych. Możliwość przeprowadzenia dowodów zbieżności globalnej dla statycznych metod optymalizacji stochastycznej, z pewnością posiada wartość teoretyczną, lecz z punktu widzenia ich praktycznego zastosowania, bardzo często pozostaje kompletnie bez znaczenia. Co więcej, statyczny rozkład eksploracyjny jak i brak mechanizmu dopasowania procedury optymalizacyjnej do zmieniającej się środowiska, zwykle rodzi wiele problemów. Z jednej strony, charakterystyka rozkładów zastosowanych w operatorze mutacji, może stworzyć sytuację, w której opuszczenie basenów przyciągania minimum lokalnego wymagać będzie wykonania nieakceptowalnej liczby prób losowych. Z drugiej strony, przyjęty eksploracyjny model stochastyczny, może spowodować, że algorytm pozostanie niewrażliwy na zbyt wąskie obszary zawierające kluczowe rozwiązania lokalne. Idealna procedura adaptacyjna, powinna zatem wykrywać sytuacje, w których następują problemy z polepszaniem jakości rozwiązania, oraz dokonać rekonfiguracji parametrów kontrolnych, w taki sposób aby przeciwdziałać stagnacji algorytmu. W rozdziale tym, zaprezentowana zostanie technika, pozwalająca na adaptację algorytmów z mutacją opartą na wielowymiarowych rozkładach stabilnych opisanych za pomocą dyskretnej miary spektralnej (4.61)-(4.62). Proponowana technika adaptacji, zaczerpnięta została ze specjalnej klasy metod optymalizacji ewolucyjnej tzw. algorytmów estymacji rozkładu eksploracyjnego (ang. *Estimation of Distribution Algorithms* - EDAs) [62]. Ogólna idea tego typu rozwiązań polega na konfiguracji przyjętego modelu probabilistycznego, przy wykorzystaniu informacji zawartej w populacji rozwiązań alternatywnych. W pierwszej kolejności operator mutacji powoduje wygenerowanie stosunkowo liczego zbioru

rozwiązań alternatywnych wokół rozwiązania bazowego. Następnie rozwiązania te, poddawane są ocenie, w wyniku której część z nich zostaje wyeliminowana poprzez zastosowanie operatora selekcji. Rozkład pozostałej części rozwiązań alternatywnych, może w intuicyjny sposób posłużyć do rekonfiguracji modelu probabilistycznego. Na przestrzeni dekady wiele rozwiązań tego typu znalazło zaprezentowanych i z powodzeniem zastosowanych do rozwiązywania zagadnień natury inżynierskiej [62]. Rozdział poświęcony rozkładom stabilnym o dyskretnej mierze spektralnej, uwidoczniał niewątpliwe zalety stosowania tej klasy modeli eksploracyjnych. Z powodów przytoczonych na wstępie rozdziału, zachodzi jednak oczywista potrzeba, adaptacji również tej miary prawdopodobieństwa. Należy jednak zaznaczyć, że sposób parametryzacji rozważanego rozkładu eksploracyjnego rodzi wiele trudności w kontekście jego optymalnej konfiguracji. W ogólności należy określić liczbę, położenie oraz wagi poszczególnych wektorów rozpinających dyskretną miarę spektralną. Uwzględnienie tych wszystkich czynników w obrębie jednej procedury adaptacyjnej jest zadaniem znaczenie wykraczającym poza zakres niniejszej rozprawy. W związku z tym, w dalszej części rozdziału, zdecydowano się przyjąć szereg uproszczeń, które mają pozwolić na uniknięcie zbyt dużego kosztu obliczeniowego związanego z adaptacją modelu. Mianowicie, zakłada się, że miara spektralna jest rozpięta na równomiernie rozłożonych punktach sfery jednostkowej. Tym samym, zadanie rekonfiguracji modelu przeszukiwań polegać będzie na dobraniu odpowiednich wag skojarzonych z ustalonymi wcześniej wektorami. Godnym zauważenia, jest fakt, iż duża liczba nośników dyskretnej miary spektralnej, w znacznej mierze zwiększa koszt otrzymywania wektorów pseudolosowych (wzór (4.63)). Fakt ten należy uwzględnić przy projektowaniu efektywnej procedury adaptacyjnej, np. poprzez nieuwzględnianie w modelu wektorów, dla których wartości wagi nie przekroczyły pewnego progu istotności.

Proponowana procedura optymalizacyjna oparta na strategii ewolucyjnej  $(1, \lambda)ES$ , wykorzystująca nakreślony schemat adaptacji, składa się następujących etapów:

Tabela 4.10: Strategia ewolucyjna  $(1, \lambda)\text{ES}_\alpha$  z mechanizmem adaptacji dyskretnej miary spektralnej

**Krok 0:** Ustal  $\lambda$ ,  $\mu = \frac{\lambda}{2}$ ,  $k = 1$ . Wybierz początkowe przybliżenie rozwiązania globalnego  $\mathbf{x}_0$ , zbiór wektorów rozpinających  $\boldsymbol{\xi} = \{\mathbf{s}_i\}_{i=1}^{n_s}$  oraz skojarzony z nim początkowy wektor wag  $\gamma_k$ .

**Krok 1:** Wykorzystując model  $\Gamma_k(\boldsymbol{\xi}, \gamma_k)$  wygeneruj zbiór rozwiązań alternatywnych.

$$\mathbf{P}_{k,\lambda} = \{\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,\lambda}\}, \quad \mathbf{x}_{k,i} = \mathbf{x}_k + \mathbf{X}\boldsymbol{\xi}^{\gamma_k}, \quad (4.72)$$

gdzie  $\mathbf{X}\boldsymbol{\xi}^{\gamma_k}$  oznaczają niezależne wektory otrzymane na podstawie (4.63).

**Krok 2:** Wybierz  $\mu$  najlepszych rozwiązań z populacji  $\mathbf{P}_{k,\lambda}$ :

$$\mathbf{P}_{k,\mu} = \{\mathbf{x}_{k,1:\lambda}, \mathbf{x}_{k,2:\lambda}, \dots, \mathbf{x}_{k,\mu:\lambda}\} \quad (4.73)$$

**Krok 3:** Na podstawie populacji  $\mathbf{P}_{k,\mu}$  dokonaj rekonfiguracji modelu przeszukiwań.

$$\gamma_k = E(\mathbf{P}_{k,\mu}) \quad (4.74)$$

gdzie  $E(\cdot)$  oznacza procedurę estymacji wag wektorów rozpinających miarę spektralną  $\gamma_k$ .

**Krok 4:** Przyjmij  $\mathbf{x}_k = \mathbf{x}_{k,1:\lambda}$ .

**Krok 5:** Jeśli nie jest spełnione kryterium stopu przejdź do kroku 1.

Dla oznaczenia algorytmu zaprezentowanego powyżej, w dalszej części rozdziału zarezerwowane zostanie oznaczenie  $(1, \lambda/\mu)\text{ES}_\alpha$ . Krokiem, który zasługuje na bardziej szczegółowy komentarz, jest procedura estymacji wagi wektorów rozpinających dyskretną miarę spektralną na podstawie rozkładu zredukowanej populacji rozwiązań.

### Estymacja dyskretnej miary spektralnej

Problem estymacji dyskretnej miary spektralnej został wielokrotnie poruszony w kontekście modelowania stochastycznego [32, 44, 59, 83]. W literaturze odnajdziemy kilka technik tworzenia modeli probabilistycznych za pomocą tej klasy rozkładów [83]. Jedną z najprostszych metod, którą zdecydowano się zastosować w strategii ewolucyjnej 4.10, zaproponowana została w [83]. Metoda opiera się na pojęciu tzw. empirycznej funkcji charakterystycznej

$$\hat{\phi}(\mathbf{k}) = \frac{1}{N} \sum_{i=1}^N \exp(j\mathbf{k}^T \mathbf{X}_i), \quad (4.75)$$

gdzie  $\mathbf{X}_i$  są realizacjami obserwowalnych zmiennych losowych, zgromadzonymi w zbiorze danych  $\{\mathbf{x}_i\}_{i=1}^N$ . Zakładając, że dyskretna miara spektralna  $\boldsymbol{\xi}$  okre-



ślona jest przez skończony zbiór wektorów  $\mathbf{S}_{n_s} = \{\mathbf{s}_i\}_{i=1}^{n_s}$  oraz skojarzonych z nimi wag  $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^{n_s}$  t.j:

$$\boldsymbol{\xi} = \left\{ \begin{array}{cccc} \mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{n_s} \\ \gamma_1 & \gamma_2 & \dots & \gamma_{n_s} \end{array} \right\}, \quad (4.76)$$

problem estymacji sprowadza się do zadania optymalizacji

$$\boldsymbol{\xi}^* = \arg \min_{\mathbf{S}_{n_s}, \boldsymbol{\gamma}} \|\hat{\phi}(\mathbf{k}) - \phi(\mathbf{k}; \mathbf{S}_{n_s}, \boldsymbol{\gamma})\| \quad (4.77)$$

Korzystając z faktu, iż dyskretna miara spektralna jest rozpięta na ustalonej siatce punktów równomiernie rozłożonych na powierzchni sfery jednostkowej, zadanie estymacji modelu (4.77), upraszcza się do następującej postaci:

$$\boldsymbol{\gamma}^* = \arg \min_{\boldsymbol{\gamma}} \|\hat{\phi}(\mathbf{k}) - \phi(\mathbf{k}; \mathbf{S}_{n_s}, \boldsymbol{\gamma})\| \quad (4.78)$$

Dokładane rozwiązanie problemu (4.78) okazuje się zadaniem na tyle skomplikowanym, że jego koszt obliczeniowy podważa sens zastosowania tej metody do wyznaczenia optymalnego zestawu wag miary spektralnej  $\boldsymbol{\gamma}^*$  w algorytmie ewolucyjnym. Dlatego też, zadanie (4.78), poprzez wprowadzenie zestawu punktów testowych  $\mathbf{K} = \{\mathbf{k}_i\}_{i=1}^{n_k}$  podlega dalszemu uproszczeniu do postaci:

$$\boldsymbol{\gamma}^* = \arg \min_{\boldsymbol{\gamma}} \sum_{i=1}^{n_k} \left( \hat{\phi}(\mathbf{k}_i) - \phi(\mathbf{k}_i; \mathbf{S}_{n_s}, \boldsymbol{\gamma}) \right)^2 \quad (4.79)$$

Wprowadzając oznaczenia:  $\mathbf{I} = -[\ln \hat{\phi}(\mathbf{k}_1), \dots, \ln \hat{\phi}(\mathbf{k}_{n_k})]^T$  oraz

$$\boldsymbol{\Psi}(\mathbf{k}_1, \dots, \mathbf{k}_{n_k}; \mathbf{s}_1, \dots, \mathbf{s}_{n_s}) = \begin{pmatrix} \psi_{\alpha}(\mathbf{k}_1^T \mathbf{s}_1) & \dots & \psi_{\alpha}(\mathbf{k}_1^T \mathbf{s}_{n_s}) \\ \vdots & \ddots & \vdots \\ \psi_{\alpha}(\mathbf{k}_{n_k}^T \mathbf{s}_1) & \dots & \psi_{\alpha}(\mathbf{k}_{n_k}^T \mathbf{s}_{n_s}) \end{pmatrix} \quad (4.80)$$

dla

$$\psi_{\alpha}(u) = \begin{cases} |u|^{\alpha} (1 - i \operatorname{sgn}(u) \tan(\frac{\pi\alpha}{2})), & \text{for } \alpha \neq 1 \\ |u| (1 - i \frac{2}{\pi} \operatorname{sgn}(u) \ln(|u|)), & \text{for } \alpha = 1 \end{cases} \quad (4.81)$$

okazuje się, że optymalny zestaw wag, jest rozwiązaniem układu równań [83]

$$\mathbf{I} = \boldsymbol{\Psi} \boldsymbol{\gamma}^*, \quad (4.82)$$

W celu poprawy uwarunkowania zadania (4.82) przyjmuje się  $n_s = n_k$  oraz  $\mathbf{s}_i = \mathbf{k}_i$ . Dodatkowo aby uniknąć sytuacji, w której optymalny zestaw wag mógłby zawierać liczby zespolone, dokonuje się podstawowych operacji na wierszach układu równań (4.82) [83]. Tym samym zadanie wyznaczenia optymalnych wag miary spektralnej można zdefiniować w formie programowania kwadratowego z ograniczeniami nierównościami [83]:

$$\boldsymbol{\gamma}^* = \arg \min_{\boldsymbol{\gamma} \geq 0} \|\mathbf{c} - \mathbf{A} \boldsymbol{\gamma}\|_2, \quad (4.83)$$

gdzie  $\mathbf{c} = [\operatorname{Re}\{I_{1:n/2}\}, \operatorname{Im}\{I_{n/2+1:n}\}]^T$  jest wektorem zawierającym najpierw  $n$  wartości rzeczywistych wektora  $\mathbf{I}$ , a następnie  $n$  jego wartości urojonych,

$\mathbf{A} = [\text{Re}\{\boldsymbol{\psi}_1^T\}, \dots, \text{Im}\{\boldsymbol{\psi}_n^T\}]^T$ , jest podobnie przeorganizowaną macierzą (4.80), gdzie  $\boldsymbol{\psi}_i = [\psi(\mathbf{s}_1^T \mathbf{s}_i), \psi(\mathbf{s}_2^T \mathbf{s}_i), \dots, \psi(\mathbf{s}_{n_s}^T \mathbf{s}_i)]^T \in \mathbb{C}^{n_s}$  są jej poszczególnymi wierszami. Problem (4.83) może z powodzeniem zostać rozwiązany przy zastosowaniu jednej z dedykowanych metod optymalizacji gradientowej. Eksperymentalnego potwierdzenia sformułowanej na początku rozdziału tezy, jakoby adaptacja rozkładu o dyskretnej mierze spektralnej miała poprawić skuteczność optymalizacyjną rozważanej strategii ewolucyjnej, postaramy się dowieść na podstawie badań symulacyjnych prezentowanych w poniższej części rozdziału.

### Badania eksperymentalne

Badania symulacyjne będące przedmiotem zainteresowania niniejszego podrozdziału, dotyczą zastosowania trzech różnych wersji strategii ewolucyjnej  $(1, \lambda/\mu)\text{ES}_\alpha$ :

- Wersja **A1** - strategia ewolucyjna  $(1, \lambda)\text{ES}_\alpha$  dla której wektor zaburzający rozwiązania określony jest przez miarę spektralną rozpiętą na wektorach

$$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

Wektor wag  $\boldsymbol{\gamma} = [\sigma/4, \sigma/4, \sigma/4, \sigma/4]^T$  pozostaje nie zmienny przez cały czas pracy algorytmu.

- Wersja **A2** - strategia ewolucyjna  $(1, \lambda)\text{ES}_\alpha$  dla której wektor zaburzający rozwiązania określony jest przez miarę spektralną rozpiętą na wektorach

$$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

Początkowy wektor wag  $\boldsymbol{\gamma} = [\sigma/4, \sigma/4, \sigma/4, \sigma/4]^T$  podlega następnie adaptacji według algorytmu (4.3.1).

- Wersja **A3** - strategia ewolucyjna  $(1, \lambda)\text{ES}_\alpha$  dla której wektor zaburzający rozwiązania określony jest przez miarę spektralną rozpiętą na wektorach

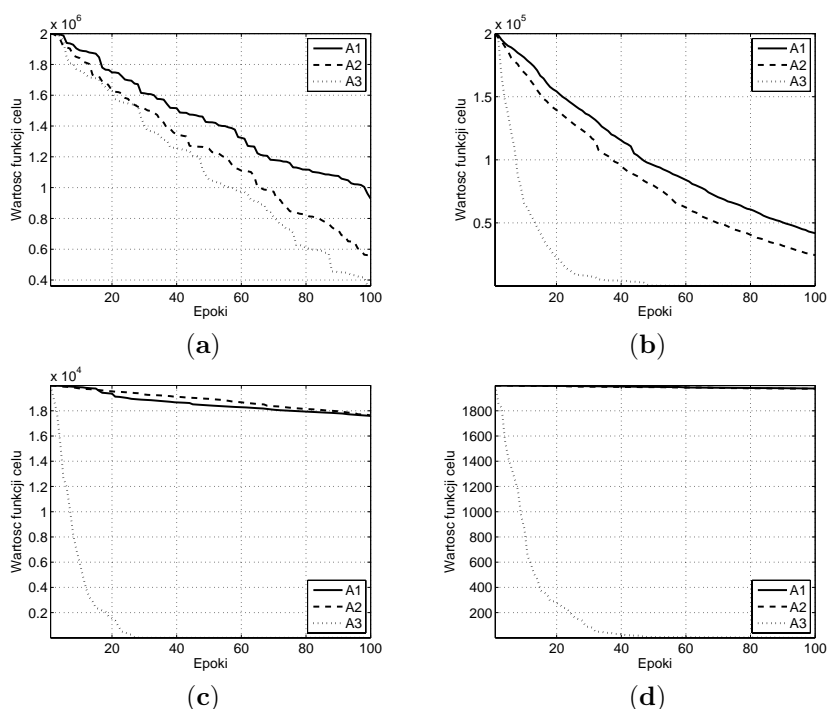
$$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \right\}$$

Początkowy wektor wag  $\boldsymbol{\gamma} = [\sigma/8, \dots, \sigma/8]^T$  podlega następnie adaptacji według algorytmu (4.3.1).

Warto zauważyć, że mutacja przypisana do algorytmu **A1** jest tożsama z wektorem zaburzającym, który składa się z dwóch niezależnych zmiennych losowych o rozkładzie symetrycznym, t.j.:  $\mathbf{Z}[S_\alpha S(\sigma/2), S_\alpha S(\sigma/2)]^T$ . Algorytm **A2** posiada operator mutacji oparty o wektor złożony z dwóch niezależnych zmiennych losowych o rozkładzie stabilnym lecz o różnych od zera współczynnikach skośności  $\mathbf{Z}[S_\alpha(\sigma/2, \beta_1)]^T$ . Pierwszą grupę eksperymentów stanowią badania przeprowadzone dla źle uwarunkowanych problemów optymalizacji lokalnej. W tym przypadku rozpatrzono cztery funkcje postaci:

$$f(x_1, x_2) = [x_1, x_2]^T \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} e_1 & 0 \\ 0 & e_2 \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}^T [x_1, x_2] \quad (4.84)$$

dla różnych wskaźników uwarunkowania zadania<sup>7</sup>, t.j.:  $(e_1, e_2) = (1, 1)$ ,  $(10, 0.1)$ ,  $(100, 0.01)$ ,  $(1000, 0.001)$ . Algorytmy **A1**, **A2**, **A3** posiadały jednakowe warunki początkowe:  $\lambda = 20$ ,  $\mu = 10$ ,  $\mathbf{x}_0 = [1000, 1000]^T$ ,  $\sigma = 1$ . Dodatkowo, aby wyniki obrazowały jedynie rolę estymacji preferowanych kierunków mutacji, zdecydowano się na zastosowanie ograniczeń równościowych narzuconych na wagi dyskretnej miary spektralnej  $\sum_{i=1}^{n_s} \gamma_i = \sigma$ . Tym samym zabezpieczono się przed ewentualnym zwiększeniem ogólnej entropii zaburzających wektorów losowych. Wyniki, w formie uśrednionych przebiegów optymalizacyjnych, dla algorytmów z indeksem stabilności  $\alpha = 0.5$ , zaprezentowane zostały na rysunku 4.16.



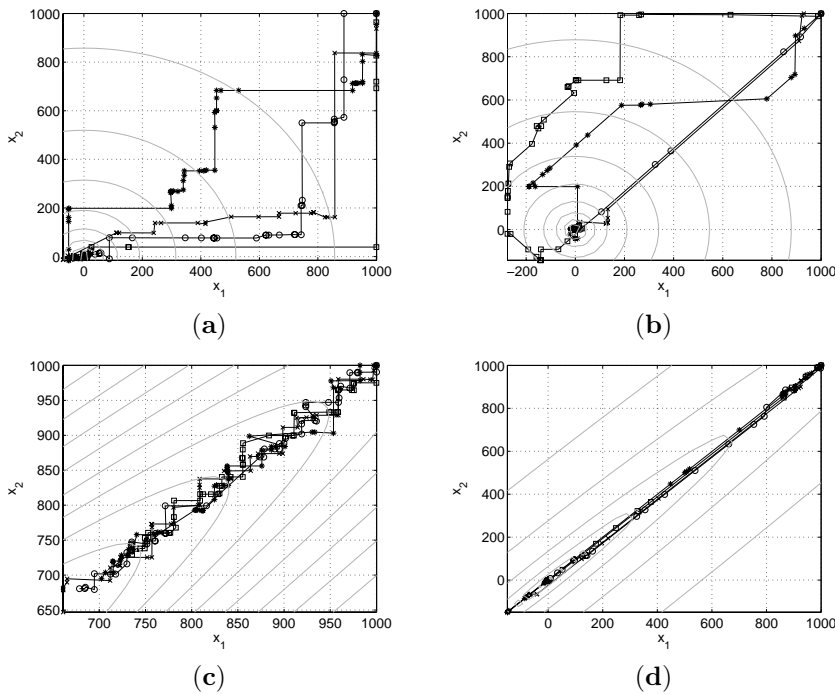
Rysunek 4.16: Średni przebieg optymalizacji funkcji kwadratowej, za pomocą algorytmów z różnymi, dyskretnymi miarami spektralnymi: **A1** - czteropunktowa bez adaptacji, **A2** - czteropunktowa z adaptacją, **A3** - ośmiopunktowa z adaptacją. Wyniki zostały uzyskane na podstawie 50 niezależnych uruchomień, każdej strategii ewolucyjnej. Parametry funkcji celu (4.84)  $(e_1, e_2)$ : (a) =  $(1, 1)$ , (b) =  $(10, 0.1)$ , (c) =  $(100, 0.01)$ , (d) =  $(1000, 0.001)$

**Obserwacja 22** Porównując uśrednione przebiegi dla funkcji na rysunku, przeważa zastosowanego mechanizmu adaptacji, nie zaznacza się nadto wyraźnie. Z pewnością jest to konsekwencją, zastosowanego indeksu stabilności  $\alpha = 0.5$ , który za pomocą mechanizmu makromutacji bardzo szybko lokalizuje rozwiązanie. Jakkolwiek najbardziej preferowany kierunek mutacji pozostaje poza zasięgiem operatora mutacji algorytmu **A1**<sup>8</sup>, to mechanizm makromutacji przeprowadzanych wzdłuż osi układu współrzędnych, może skutecznie zlokalizować rozwiązanie

<sup>7</sup>stosunek największej wartości własnej do wartości najmniejszej

<sup>8</sup>operator ten faworyzuje kierunki zgodne z osiami układu współrzędnych - tzw. efekt symetrii zaobserwowany przez Obuchowicza [85] oraz poruszany w rozdziale 4.1.

w ciągu kilku kolejnych iteracji. Algorytm adaptacyjny **A2** różni się od powyższego, jedynie możliwością adaptacji parametrów skali niezależnych zmiennych zaburzących. W prawdzie skutkuje to poprawieniem skuteczności optymalizacyjnej, ale w dalszym ciągu nie ujawnia się możliwość dopasowania rozkładu eksploracyjnego do dowolnego kierunku w przestrzeni przeszukiwań. Taka możliwość, za sprawą specjalnej konfiguracji wektorów rozpinających DSM, pojawia się dopiero w przypadku algorytmu **A3**. Obserwując trajektorie prezentowane na rysunkach 4.17 (b)-(d) wyraźnie widać wpływ przypisania stosunkowo większej wartości wagi  $\gamma_6 \approx \sigma$  wektorowi  $s_6 = [-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]^T$ . Taka konfiguracja DSM narzuca wystąpienie ewentualnych makromutacji wyłącznie na kierunku wskazywanym przez  $s_6$ . Efekt ten powoduje nieporównywalnie szybszą zbieżność algorytmu **A3** w przypadku źle uwarunkowanych problemów testowych ( wykresy na rysunkach 4.16).



Rysunek 4.17: Przykładowe trajektorie zaznaczone przez najlepsze rozwiązanie algorytmów **A1** - (a),(c) oraz **A3** - (b),(d). Rysunki (a),(b) odpowiadają funkcji celu o parametrach  $(e_1, e_2) = (1, 1)$ , (c),(d) natomiast  $(e_1, e_2) = (10, 0.1)$ .

### 4.3.2 Podsumowanie

Dotychczas, zastosowanie wielowymiarowych rozkładów stabilnych, w algorytmach optymalizacji globalnej ograniczało się jedynie do najprostszych przypadków: poszczególne składowe wektora były realizowane w postaci stabilnych, niezależnych zmiennych losowych [85, 88, 135, 136], lub wektora izotropowego [95, 104]. Zawężenie rozważań jedynie do wspomnianych przypadków, powoduje jednocześnie, że wiele właściwości rozkładów stabilnych, które mogą oka-

zać się bardzo cenne w kontekście optymalizacji stochastycznej, pozostawało niewykorzystanych. Pojęcie dyskretnej miary spektralnej otwiera drzwi przed bardzo szeroką klasą wektorów losowych, umożliwiając równocześnie, modelowanie skomplikowanych zależności stochastycznych pomiędzy zmiennymi decyzyjnymi. W rozdziale tym, zaprezentowana została metoda estymacji DSM dla standardowej strategii ewolucyjnej  $(1+1)ES_{\alpha}$ . Wielowymiarowy model stabilny, wykorzystywany do eksploracji przestrzeni przeszukiwań, dopasowywany jest do środowiska na podstawie informacji zawartej w rozkładzie najlepszych osobników w populacji. Seria prostych eksperymentów wykonanych dla prezentowanego schematu adaptacji, wykazały bardzo wyraźnie przewagę tej metody nad rozwiązaniami nie posiadającymi możliwości uwzględnienia wystąpienia silnych korelacji pomiędzy zmiennymi decyzyjnymi.

Proponowana metoda nie jest jednak wolna od wad. Należy szczególnie wyraźnie zaznaczyć, że zasadność proponowanej techniki zyskuje potwierdzenie jedynie w przypadkach, gdy koszt związany z obliczeniem funkcji celu jest stosunkowo duży. Należy bowiem zaznaczyć, że problem estymacji wag DSM wymaga wykonania, w każdej iteracji strategii ewolucyjnej, optymalizacji lokalnej. Co więcej, aby umożliwić wykrywanie złożonych zależności pomiędzy poszczególnymi zmiennymi decyzyjnymi, w przestrzeniach o większej liczbie wymiarów, należy zwiększać liczbę wektorów rozpinających DSM. W przypadku braku jakiegokolwiek wiedzy a-priori dotyczących problemu optymalizacji, liczba ta rośnie wykładniczo, co w decydujący sposób może spowodować nagły wzrost kosztu obliczeniowego związanego z estymacją DSM. Rozwiązania wspomnianego problemu, należy upatrywać w zastosowaniu podejść heurystycznych, do adaptacji wag wektorów rozpinających. W tym miejscu wystarczy wskazać na którykolwiek z istniejących schematów adaptacji dla rozkładu normalnego. Do optymalnej konfiguracji DSM można np. wykorzystać informację o położeniu wektorów własnych proponowanych macierzy kowariancji. Zagadnienia te, jednakże wykraczają poza ramy niniejszej rozprawy, i dlatego też rozdział ten należy potraktować jako wstęp wyznaczający kierunki przyszłych badań nad tą tematyką.

## Rozdział 5

# Mutacja kierunkowa

Niniejszy rozdział zainspirowany został przez chęć szczegółowego przeanalizowania efektu polegającego na spadku efektywności ewolucyjnych technik optymalizacji dla problemów wielowymiarowych - rozdział 4. W pierwszej kolejności przedstawiono twierdzenie będące specjalnym przejawem tzw. klątwy wymiarowości dla mutacji izotropowych. W drugiej części rozdziału, zaproponowano klasę rozkładów kierunkowych, które wraz z heurystykami adaptacyjnymi, mogą posłużyć do zaprojektowania efektywnych technik optymalizacji wielowymiarowej.

### 5.1 Wstęp i motywacja

W literaturze odnajdujemy wiele algorytmów ewolucyjnych, które wykorzystują izotropowe rozkłady eksploracyjne [10, 20, 41, 119]. Modele probabilistyczne o sferycznej symetrii, posiadają bez wątpienia tyle samo wad ile zalet. Wśród ich niewątpliwych zalet odnajdziemy kwestie związane z niewielką liczbą parametrów potrzebnych do ich jednoznaczego zdefiniowania. Dla algorytmów statycznych, fakt ten umożliwia znaczne uproszczenie procesu konfiguracji procedury optymalizacyjnej. W przypadku algorytmów adaptacyjnych, znacznie łatwiej jest zaproponować skuteczną heurystykę dopasowującą rozkład eksploracyjny do środowiska optymalizacyjnego. Mała liczba parametrów, ma również swoje negatywne konsekwencje, które w znacznej mierze przyczyniają się pogorszenia skuteczności algorytmów. Mając na względzie, iż efektywności stochastycznych technik optymalizacji, uwarunkowana jest w dużej mierze przez możliwości dopasowania modelu przeszukiwań do rozważanego problemu, mała liczba parametrów, musi ograniczać możliwości w modelowaniu zależności pomiędzy zmiennymi decyzyjnymi. Eksploracyjny model sferyczny może jedynie uwzględniać korelację pomiędzy wartościami funkcji celu, a odległością euklidesową od punktu bazowego. W przypadku, gdy rozwiązywany problem cechują silne korelacje pomiędzy poszczególnymi zmiennymi decyzyjnymi, zastosowanie eksploracyjnych rozkładów izotropowych musi siłą rzeczy pociągać za sobą wydłużenie czasu optymalizacji.

Niektóre z wymienionych wyżej problemów, mogą zostać zniwelowane poprzez odejście od mutacji zachodzącej pod dyktando rozkładów izotropowych. W literaturze odnajdziemy wiele technik optymalizacji ewolucyjnej bazujących na

eliptycznych rozkładach normalnych [20, 19, 49, 50, 48, 58]. Algorytmy te, są przeważnie wyposażone w procedury adaptacyjne, których nadrzędnym celem jest dopasowanie modelu przeszukiwań do środowiska, w którym przeszukiwanie to zachodzi. Mimo możliwości modelowania liniowych zależności pomiędzy zmiennymi decyzyjnymi, operator mutacji oparty o rozkład normalny  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  nie jest wolny od wad. Przede wszystkim odznacza się symetrią względem oczekiwanego wektora, przez co współczynnik zbieżności algorytmu ewolucyjnego pozostaje na stosunkowo niskim poziomie. Bardziej szczegółowe omówienie tego zagadnienia znajduje się dalszej części rozdziału.

### 5.1.1 Klątwa wymiarowości dla mutacji izotropowych

Z definicji 12 wynika, że każdy rozkład o sferycznej symetrii  $\mathbf{X}^s$  posiada następującą dekompozycję stochastyczną [38]:

$$\mathbf{X}^s = r\mathbf{u}^{(n)}, \quad (5.1)$$

gdzie  $\mathbf{u}^{(n)}$  jest wektorem losowym o rozkładzie równomiernym na powierzchni  $n$ -wymiarowej kuli jednostkowej, oraz  $r > 0$  jest niezależną od  $\mathbf{u}^{(n)}$  zmienną losową przyjmującą tylko wartości dodatnie. Związek pomiędzy  $r$  i  $\mathbf{X}^s$  wyraża się przez relacje jeden do jednego, z czego wynika, że zbiór wszystkich rozkładów izotropowych jest równoliczny ze zbiorem wszystkich rozkładów dodatnich [38]. Należy podkreślić fakt, że wektor  $\mathbf{u}^{(n)}$  posiada największą możliwą entropię losową ze wszystkich rozkładów określonych na sferze jednostkowej. Generalną zasadą jest, iż rozkład o największej entropii wnosi również największą ilość informacji o przestrzeni, na której jest określony. Zasada ta, jest również najczęściej przytaczanym powodem stosowania rozkładu normalnego w algorytmach stochastycznych dedykowanych do optymalizacji w przestrzeniach  $\mathbb{R}^n$ . Chodź cecha ta, na pierwszy rzut oka wydaje się jak najbardziej pożądana, to w rzeczywistości odnosi ona zupełnie odwrotny skutek jeśli chodzi o efektywność metod optymalizacji. Mianowicie zakładając, iż zwykle ze wszystkich elementów przestrzeni, na których określony jest rozkład prawdopodobieństwa, tylko nieliczny podzbiór posiada pożądane cechy, największa entropia określa równocześnie największą niepewność związaną z wyborem właśnie tego podzbioru. Przekładając powyższe rozważania na wybór kierunku przeszukiwań w przestrzeni rozwiązań, warto zauważyć, iż zwykle istnieje dokładnie jeden najbardziej korzystny kierunek. W przypadku problemów różniczkowalnych będzie to kierunek przeciwny do gradientu funkcji celu. Dla funkcji nieróżniczkowalnych jest to kierunek wybrany z otoczenia punktu bazowego, powodujący największą poprawę jakości rozwiązania. W większości przypadków, każde odchylenie od tego kierunku powoduje znaczne pogorszenie wyników modyfikacji aktualnego rozwiązania. Powyższe rozważania, nawiązujące bardziej do intuicji niż do ścisłego opisu matematycznego, można podsumować w formie poniższego twierdzenia

**Twierdzenie 15** *Niech  $\mathbf{X}^s$  oznacza losowy wektor o rozkładzie o charakterystycznym się sferyczną symetrią. Załóżmy, że rozwiązane alternatywnie jest generowane na podstawie addytywnych zaburzeń wektorem  $\mathbf{X}^s$  tj.:  $\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{X}^s$ . Ponadto, niech  $\boldsymbol{\mu}_t$  oznacza najbardziej korzystny kierunek mutacji punktu  $\mathbf{x}_t$ . Wówczas, prawdopodobieństwo, że rozwiązanie  $\mathbf{x}_{t+1}$ , będzie leżało na kierunku prostopadłym do  $\boldsymbol{\mu}_t$  dąży do jedynki, jeśli liczba wymiarów przestrzeni rozwiązań dąży do nieskończoności.*

**Dowód 1** Niech  $\boldsymbol{\mu} \in \mathbb{R}^n$  będzie wektorem jednostkowym  $\|\boldsymbol{\mu}\| = 1$ , oraz  $\mathbf{u}^{(n)}$  oznacza wektor o rozkładzie równomiernym na sferze jednostkowej. Wówczas rozkład zmiennej losowej  $t = \boldsymbol{\mu}^T \mathbf{u}^{(n)}$  opisany jest za pomocą następującej funkcji gęstości [38]:

$$f_n(t) = \frac{1}{\beta(\frac{n-1}{2}, \frac{1}{2})} (1-t^2)^{\frac{n-3}{2}}, \quad (5.2)$$

gdzie  $\beta(\cdot, \cdot)$  oznacza funkcję Beta [102]. Rozważmy następnie granicę ciągu funkcyjnego  $\{f_n\}$  dla  $t \in (-1, 0) \cup (0, 1)$ :

$$\lim_{n \rightarrow \infty} f_n(t) = \pi^{-1/2} \lim_{n \rightarrow \infty} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} (1-t^2)^{\frac{n-3}{2}}. \quad (5.3)$$

Podstawiając  $k = 2(n+1)$  oraz  $b = (1-t^2) \in (0, 1)$ , (5.3) przyjmuje postać:

$$\lim_{n \rightarrow \infty} f_n(t) = \lim_{k \rightarrow \infty} \frac{\Gamma(k+1)}{\Gamma(k+1/2)} b^{k-1/2}. \quad (5.4)$$

Następnie zauważmy, że prawdziwa jest nierówność

$$a_k = 0 \leq \frac{\Gamma(k+1)}{\Gamma(k+1/2)} b^{k-1/2} < (k+1/2)^{1/2} b^{k-1/2} \leq (k+1/2) b^{k-1/2} = c_k.$$

Granica ciągu  $c_k$  może zostać wyznaczona za pomocą reguły L'Hospital'a [102]:

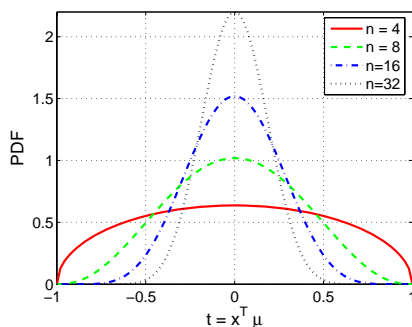
$$\lim_{k \rightarrow \infty} c_k = \lim_{k \rightarrow \infty} \frac{k+1/2}{b^{1/2-k}} = \lim_{k \rightarrow \infty} \frac{1/2}{-\ln(b)b^{1/2-k}} = -\frac{b^{-1/2}}{2 \ln(b)} \lim_{k \rightarrow \infty} b^k = 0$$

Ponieważ, oba ciągi  $a_k, c_k$  zbiegają do zera, z twierdzenia o trzech ciągach wynika [102], że również granica ciągu  $d_k$  musi wynosić 0. Tym samym, funkcja gęstości (5.2) zbiega do zera dla każdego  $t \in [-1, 0) \cup (0, 1]$  i rośnie do nieskończoności dla  $t = 0$ . Następstwem, tego faktu jest skupienie całej masy prawdopodobieństwa w punkcie  $t = 0$ , co oznacza, że ciąg funkcyjny  $\{f_n\}$  aproksymuje dystrybucję Deltę Dirac'a. Jeśli przyjmując, że wektor  $\boldsymbol{\mu}$  oznacza najkorzystniejszy kierunek przestrzeni rozwiązań wówczas natychmiast otrzymujemy potwierdzenie słuszności dowodzonego twierdzenia.  $\square$

Twierdzenie 15 pozwala na wytłumaczenie, a co więcej, na ilościowe oszacowanie spadku efektywności algorytmów ewolucyjnych dla przestrzeni o znacznej liczbie wymiarów. Zjawisko, na którym oparte jest twierdzenie 15 staje się szczególnie widoczne podczas zawężania okna ewolucyjnego w zadaniu przekraczania siodła pomiędzy dwoma wierzchołkami gaussowskimi (eksperyment przeprowadzony w rozdziale 4.2.1). Ze względu na elitarny charakter selekcji w rozpatrywanym algorytmie, możemy całkowicie wyeliminować wpływ dryftu populacji [42]. Tym samym należy się zgodzić ze stwierdzeniem, że w tym wypadku, jedynym mechanizmem powodującym przekraczanie siodła ewolucyjnych jest mechanizm makromutacji. W miarę zwiększania wymiaru przestrzeni przeszukiwań, zmniejsza stosunek kąta bryłowego zawierającego kierunek umożliwiającą poprawienie rozwiązania do powierzchni sfery jednostkowej. Tym samym coraz mniejsze jest prawdopodobieństwo otrzymania pożądanego wektora losowego we wskazanym obszarze. Za pomocą zjawiska leżącego u podstaw twierdzenia 15 można również wytłumaczyć wzrost średniej liczby generacji, niezbędnej do zlokalizowania



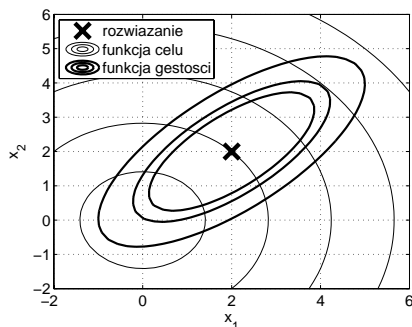
rozwiązania w rozdziale 4.2.1. Śledząc zmiany zachodzące dla liczby obliczeń funkcji celu (rysunek 4.4) w miarę rozwiązywania problemu o coraz to większym wymiarze, można zaobserwować wspólną tendencję dla wszystkich rozważanych stabilnych rozkładów izotropowych, polegającą na zwiększeniu nakładów obliczeniowych. Zjawisko to, trudno wytłumaczyć nie dysponując twierdzeniem 15, tym bardziej że odległość euklidesowa punktu startowego od rozwiązania globalnego jest stała i nie zależy od  $n$ . Twierdzenie 15 zobrazowane zostało na rysunku 5.1.



Rysunek 5.1: Funkcja gęstości zmiennej losowej  $t = \mathbf{Z}^T \boldsymbol{\mu}$  dla izotropowego wektora eksploracyjnego  $\mathbf{Z}$ , oraz najbardziej korzystnego kierunku mutacji  $\boldsymbol{\mu}$

### 5.1.2 Problem symetrii rozkładu

Na polu obliczeń ewolucyjnych, do najczęściej spotykanych nieizotropowych rozkładów eksploracyjnych, należy eliptyczny model normalny [38]. W przypadku rozkładu  $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ , termin eliptyczny nawiązuje do wykresu poziomicowego funkcji gęstości, który przyjmuje kształt elips o promieniach leżących na kierunkach wyznaczonych przez wektory własne macierzy kowariancji  $\boldsymbol{\Sigma}$ . Dużym wyzwaniem związanym z tego typu rozkładami, jest zaprojektowanie skutecznej procedury odpowiedniego doboru macierzy kowariancji podczas pracy algorytmu optymalizacji. W literaturze poświęconej ewolucyjnym technikom optymalizacji dużo miejsca poświęconego zostało temu zagadnieniu [48][58],[50], [112]. Jakkolwiek poszczególne rozwiązania różnią się pod względem zastosowanych heurystyk oraz sposobami przetwarzania informacji zbieranych podczas eksploracji przestrzeni rozwiązań, to ich wspólnym celem jest takie dopasowanie rozkładu eksploracyjnego, aby zintensyfikować próbkowanie obszaru o lepszej jakości funkcji celu. Zastosowanie wielowymiarowego rozkładu normalnego, wiąże zawsze z pewnym niebezpieczeństwem, które rzadko kiedy jest szerzej komentowane przy okazji testowania tego typu rozwiązań. Mianowicie, abstrahując od skuteczności różnych heurystyk, strojenie macierzy kowariancji skutkuje ustaleniem najdłuższej półosi elipsy w kierunku, w którym funkcja celu charakteryzuje się najmniejszymi wartościami. Taka wyidealizowana sytuacja, w dla funkcji sferycznej  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$  oraz przybliżenia minimum globalnego  $\mathbf{x}_k = [2, 2]^T$  zaprezentowana została na rysunku 5.2.



Rysunek 5.2: (a) - wykres poziomicowy funkcji celu oraz funkcji gęstości rozkładu normalnego, którego macierz kowariancji zwiększa prawdopodobieństwo poprawienia rozwiązania  $\mathbf{x}_k = [2, 2]^T$

Naszkicowana sytuacja ujawnia jednak największą wadę eliptycznych rozkładów eksploracyjnych. Symetria rozkładu sprawia bowiem, iż dokładnie taka sama masa prawdopodobieństwa, jaka zostaje umieszczona w kierunku najbardziej obiecującym, zostaje również umieszczona w kierunku do niego przeciwnym. W ogólności oznacza to, że więcej niż połowa wszystkich prób zaburzeń rozwiązania będzie kończyć się niepowodzeniem. Niewątpliwie pociąga to za sobą znaczny spadek skuteczności algorytmów stochastycznych, szczególnie wyraźnie obserwowany w przestrzeniach o dużej liczbie wymiarów.

Próba zniwelowania niekorzystnego wpływu efektu symetrii jest zaproponowana przez Obuchowicza [85], wersja algorytmu ewolucyjnego z miękką selekcją [41] zakładająca wymuszony kierunek mutacji. Idea tego typu rozwiązania, opierała się na zastosowaniu operatora mutacji, który polegał na tworzeniu populacji alternatywnych rozwiązań według schematu:

$$\mathbf{x}_{k+1}^{(i)} = \mathbf{x}_k^{(i)} + N(\boldsymbol{\mu}, \sigma \mathbf{I}_n), \quad (5.5)$$

$$\boldsymbol{\mu} = c\sigma \frac{\hat{X}_k - \hat{X}_{k-1}}{\|\hat{X}_k - \hat{X}_{k-1}\|} \quad (5.6)$$

$$\hat{X}_k = \frac{1}{\eta} \sum_{i=1}^{\eta} \mathbf{x}_k^{(i)}, \quad (5.7)$$

gdzie  $X_k = \{\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)}, \dots, \mathbf{x}_k^{(\eta)}\}$  jest zbiorem osobników w iteracji  $k$ , a  $\sigma$  oraz  $c$  arbitralnie dobraćanymi parametrami metody. Jak wynika ze wzorów (5.5)-(5.7), preferowany kierunek mutacji jest ustalany przy udziale operatora selekcji, który eliminując osobniki z obszarów o gorszej jakości dopasowania, powoduje modyfikację środków ciężkości następujących po sobie populacji osobników. Algorytm ten, początkowo zaprojektowany z myślą o niestacjonarnych problemach optymalizacyjnych [84], wykazał się również dużą skutecznością w problemach estymacji parametrycznej dynamicznych sieci neuronowych [85].

Pragnąc przezwyciężyć problemy związane z symetrią rozkładu eksploracyjnego, w rozdziale tym zaprezentowane zostanie podejście, polegające na zastosowaniu tzw. kierunkowych rozkładów eksploracyjnych [69].

## 5.2 Rozkłady kierunkowe

Ogólna idea eksploracyjnych rozkładów kierunkowych polega na zastosowaniu podobnej dekompozycji stochastycznej, jak to miało miejsce w przypadku rozkładów izotropowych (definicja 12). Największa różnica polega na zamianie rozkładu bazowego  $\mathbf{u}^{(n)}$ , przez rozkład nierównomierny na powierzchni kuli jednostkowej  $\mathbf{d}^{(n)}$ , [96]

$$Z \sim r \mathbf{d}^{(n)} \quad (5.8)$$

W literaturze można odleźć wiele przykładów wspomnianych modeli probabilistycznych [69], podczas gdy do celów optymalizacji stochastycznej, największe znaczenie wydają się mieć rozkłady o symetrii obrotowej von Mises-Fishera (vMF) [68]. Do ich jednoznacznego zdefiniowania niezbędne są dwa parametry: wyróżniony kierunek  $\boldsymbol{\mu}$ ,  $\|\boldsymbol{\mu}\| = 1$  oraz rozproszenie  $\kappa \geq 0$ . Funkcja gęstości,  $n$ -wymiarowego rozkładu (vMF) wyrażona jest poprzez [35]:

$$f_n(\mathbf{s}; \kappa, \boldsymbol{\mu}) = \frac{\kappa^{n/2-1}}{(2\pi)^{n/2} I_{\frac{n}{2}-1}(\kappa)} \exp(\kappa \boldsymbol{\mu}^T \mathbf{s}), \quad (5.9)$$

gdzie  $\mathbf{s} \in \partial S_1^{(n)}(\mathbf{0})$ . Termin symetrii obrotowej wyraża się przez niezmienniczość rozkładu względem obrotów wokół wyróżnionego kierunku. Mianowicie, przyjmując że  $g(\cdot)$  jest funkcją gęstości rozkładu  $\mathcal{M}(\boldsymbol{\mu}, \kappa)$ , zachodzi zależność:

$$g(\mathbf{x}) = g(\mathbf{y}) \Rightarrow \boldsymbol{\mu}^T \mathbf{x} = \boldsymbol{\mu}^T \mathbf{y} \quad (5.10)$$

W przypadku zastosowań rozkładów kierunkowych w algorytmach ewolucyjnych, procedura zaburzania rozwiązań może być postrzegana jako proces dwuetapowy. W pierwszej kolejności losowany jest kierunek modyfikacji rodzica - zgodnie z rozkładami vMF, a następnie za pomocą niezależnej zmiennej losowej, dokonywana jest modyfikacja rozwiązania na wcześniej obranym kierunku. Naturalnym pytaniem, jakie powstaje w perspektywie zastosowań rozkładów kierunkowych vMF, dotyczy sposobu ustawień parametrów  $\boldsymbol{\mu}$  i  $\kappa$ . Na pytanie to można odpowiedzieć w świetle już istniejących rozwiązań heurystycznych. Mianowicie techniki, które wcześniej wykorzystywane były do ustalenia kierunku najdłuższej półosi elipsoidy rozkładu normalnego, mogą zostać bezpośrednio zastosowane do ustalania parametru  $\boldsymbol{\mu}$ . W ten sposób, otrzymujemy model probabilistyczny, który intensyfikuje próbkowanie obszaru, w centrum którego leży najbardziej obiecujący kierunek przestrzeni rozwiązań. Stopień nasilenia próbkowania kierunku  $\boldsymbol{\mu}$  ustalany jest przez parametr rozproszenia  $\kappa$  i może zostać dobrany arbitralnie. Cecha, która z pewnością przemawia na korzyść stosowania rozkładów kierunkowych, przejawia się tym, iż prawdopodobieństwo otrzymania rozwiązania alternatywnego leżącego na dowolnym kierunku  $\hat{\boldsymbol{\mu}}$ , maleje wraz ze wzrostem kąta pomiędzy  $\boldsymbol{\mu}$  i  $\hat{\boldsymbol{\mu}}$ . W ten sposób, można skutecznie rozwiązać problem, wspólny dla wszystkich symetrycznych rozkładów eksploracyjnych, polegający na zbyt dużej liczbie mutacji kończących się niepowodzeniem (zobacz: komentarz do rysunku 5.2).

Niestety rozkłady von Mises-Fisher'a, nie są wolne od wad, które trzeba wziąć pod uwagę przed ich zastosowaniem w operatorze mutacji algorytmów ewolucyjnych. Aby zobrazować największy z nich, rozpatrzmy następujący przykład numeryczny:

**Przykład 1** Rozważmy problem minimalizacji funkcji sferycznej  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ , za pomocą najprostszej strategii ewolucyjnej  $(1+1)ES$ . Tak postawione zadanie, posłuży nam do przedstawienia korzyści płynących z zastosowania rozkładów kierunkowych na tle ich izotropowych odpowiedników. Na potrzeby niniejszego przykładu, założymy istnienie idealnej heurystyki adaptacyjnej pozwalającej, która w każdej iteracji pozwala dokładnie wskazać najbardziej preferowany kierunek mutacji. Korzyści płynące z istnienia takiej wyidealizowanej heurystyki łatwo zasymulować umożliwiając algorytmowi ewolucyjnemu korzystanie z analitycznej postaci gradientu funkcji celu. Aby uczynić porównanie wiarygodnym, zauważmy iż wektor losowy  $\mathbf{X}$  o izotropowym rozkładzie normalny  $N(\mathbf{0}, \sigma \mathbf{I}_n)$  posiada następującą dekompozycję stochastyczną:

$$\mathbf{X} \stackrel{d}{=} \mathbf{u}^{(n)} \sigma \chi_n, \quad (5.11)$$

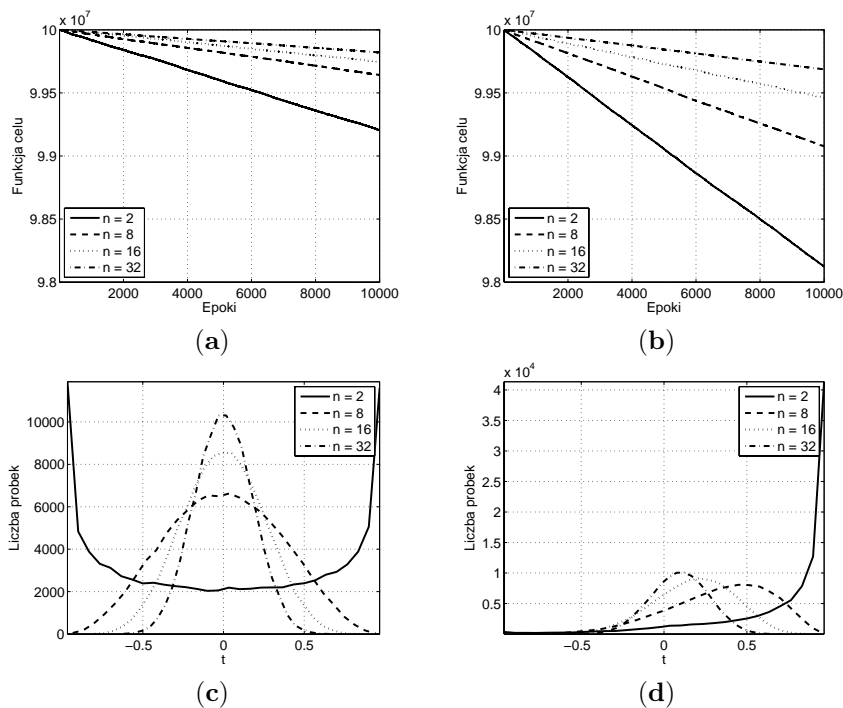
gdzie  $\chi_n$  jest zmienną losową o  $n$  stopniach swobody. Dlatego też, w strategii ewolucyjnej z rozkładem kierunkowym zastosujemy następujący wektor zaburzający:

$$\mathbf{X} \stackrel{d}{=} \mathbf{d}^{(n)} \sigma \chi_n, \quad (5.12)$$

gdzie  $\mathbf{d}^{(n)} \sim \mathcal{M}(\boldsymbol{\mu}, \kappa)$ . Porównując rozpatrywane rozkłady (5.11) i (5.12), różnica pomiędzy nimi polega jedynie na zastosowaniu różnych rozkładów bazowych, których zadaniem jest wybór kierunku mutacji. Warto zauważyć, że znajomość gradientu funkcji celu, jest zupełnie bez znaczenia w przypadku rozkładu izotropowego, natomiast w przypadku rozkładu kierunkowego, wiedza ta, zostanie wykorzystana do ustalenia kierunku mutacji  $\boldsymbol{\mu}$ . Mianowicie najbardziej preferowany kierunek będzie utożsamiany z kierunkiem najszybszego spadku, t.j.:

$$\boldsymbol{\mu}^* = -\nabla \phi(\mathbf{x}_k) = -\left[ \frac{\partial \phi(\mathbf{x})}{\partial x_1}, \frac{\partial \phi(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial \phi(\mathbf{x})}{\partial x_n} \right]^T \quad (5.13)$$

Porównanie skuteczności optymalizacyjnej obu algorytmów ewolucyjnych, wykonane zostało przy założeniu następujących warunków początkowych: punkt startowy  $\mathbf{x}_0 = [10000, 0, 0, \dots, 0]^T$ , skala mutacji  $\sigma = 0.01$ , maksymalna liczba iteracji  $T_{max} = 10000$ , rozproszenie rozkładów kierunkowych vMF  $\kappa = 3$ . Przebieg procesu optymalizacji, pojedynczego uruchomienia strategii ewolucyjnych zamieszczony został na rysunkach 5.3 - (a) i (b). Dodatkowo, na rysunkach 5.3 - (c) i (d) przedstawiono histogramy wartości  $\mathbf{X}^T \boldsymbol{\mu}^*$ , gdzie  $\mathbf{X}$  jest wektorem zaburzającym (w jednym przypadku izotropowym, w drugim kierunkowym) a  $\boldsymbol{\mu}^*$  jest kierunkiem gwarantującym najszybszą zbieżność algorytmów (5.13).



Rysunek 5.3: Przebieg procesu minimalizacji funkcji sferycznej  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ ,  $\mathbf{x} \in \mathbb{R}^n$  dla algorytmu z mutacją izotropową - (a), oraz kierunkową - (b). Rysunki (c), (d) obrazują histogramy zaobserwowanych realizacji zmiennej losowej  $t = \mathbf{X}^T \boldsymbol{\mu}^*$  dla wyników zamieszczonych odpowiednio na wykresach (a) i (b)

Wyniki zaprezentowane na rysunku 5.3, zdecydowanie ukazują przewagę rozkładów kierunkowych. Niestety przewaga ta zdaje się zacierać dla problemów o większej liczbie wymiarów przestrzeni rozwiązań. Zjawisko to, można wytłumaczyć obserwując histogramy na rysunku 5.3 (d). Otóż, wynika z nich, iż wraz ze wzrostem wymiaru przestrzeni, mimo stałego parametru rozproszenia, rozkłady te generują coraz więcej wektorów kierunkowych, znacznie różniących się od optymalnego kierunku. W najgorszym przypadku  $n = 32$  sytuacja przybiera paradoksalny obrót, gdyż większość mutacji zachodzi na kierunkach niemalże prostopadłych do (5.13). Fakt ten, jest tym bardziej zaskakujący, iż funkcja gęstości rozkładów vMF osiąga największą wartość dla kierunku  $\boldsymbol{\mu}^*$ .

Problemy z mutacjami kierunkowymi, ujawnione na podstawie prostego przykładu 1, stawiają pod znakiem zapytania przydatność rozkładów vMF do optymalizacji problemów o znacznej liczbie zmiennych decyzyjnych. Mianowicie, nawet w przypadku posiadania idealnej heurystyki mogącej wskazać dokładny kierunek w którym powinna przebiegać mutacja, rozkłady vMF, nie pozwalają w pełni wykorzystać tej cennej informacji. Dlatego też, w dalszej części rozdziału, zaprezentowana zostanie koncepcja rozkładów kierunkowych, które w dużej mierze umożliwiają ominięcie zaobserwowanych wad, a tym samym dalszą poprawę skuteczności optymalizacyjnej algorytmów ewolucyjnych.

W pierwszej kolejności wskaźmy na źródło problemów związanych z rozkładami

vMF. Mianowicie, zauważmy, że funkcja gęstości zmiennej losowej  $t = \hat{\boldsymbol{\mu}}^T \mathbf{d}^{(n)}$ , gdzie  $\mathbf{d}^{(n)} \sim \mathcal{M}(\boldsymbol{\mu}, \kappa)$ , a  $\hat{\boldsymbol{\mu}}$  jest dowolnym wektorem  $\|\hat{\boldsymbol{\mu}}\| = 1$ , dana jest wzorem [69]:

$$f_n(t; \kappa) = \frac{\left(\frac{\kappa}{2}\right)^{n/2-1}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma(n/2)I_{\frac{n-1}{2}}(\kappa)} \exp(\kappa t)(1-t^2)^{\frac{n-3}{2}}, \quad (5.14)$$

gdzie  $\|\boldsymbol{\mu}\|_2 = 1$ ,  $I_n(\cdot)$  oznacza zmodyfikowaną funkcję Bessela pierwszego rodzaju. Jasnym jest, że wartość oczekiwana  $E[t] = E[\hat{\boldsymbol{\mu}}^T \mathbf{d}^{(n)}]$  jest funkcją malejącą parametru  $n$  [96]. Fakt ten skutkuje dosyć zadziwiającą obserwacją. Otóż, mimo wyróżnienia przez rozkład vMF kierunku  $\boldsymbol{\mu}$ , w miarę wzrostu wymiaru przestrzeni przeszukiwań, wektory pseudolosowe generowane zgodnie rozkładem coraz bardziej będą różniły się od kierunku  $\boldsymbol{\mu}$ . Oznacza to, że mutacje algorytmu ewolucyjnego wcale nie będą zachodziły najczęściej na ustalonym kierunku  $\boldsymbol{\mu}$ . W ten sposób powstaje pytanie: czy możliwe jest skonstruowanie rozkładu prawdopodobieństwa o symetrii obrotowej, ale posiadającego dowolny rozkład brzegowy  $t = \boldsymbol{\mu}^T \mathbf{X}$ ? Szukając odpowiedzi na wyżej postawione pytanie, zauważmy, że każdy wektor losowy o symetrii obrotowej, może zostać przedstawiony w postaci następującej dekompozycji stochastycznej [69]:

$$\mathbf{X} = t\boldsymbol{\theta} + \sqrt{1-t^2}\boldsymbol{\xi}, \quad (5.15)$$

gdzie  $t$  jest zmienną losową niezmienniczą względem obrotów wokół  $\boldsymbol{\theta}$ , a  $\boldsymbol{\xi}$  posiada rozkład równomierny na powierzchni sfery jednostkowej  $\partial S_1^{(n-2)}(\mathbf{0})$ . Ponadto  $\boldsymbol{\xi}$  oraz  $t$  są wzajemnie niezależne. Dekompozycja (5.15) może zatem posłużyć do konstrukcji rozkładu o sferycznej symetrii z dowolnym rozkładem brzegowym. W tym celu podstawmy  $\boldsymbol{\theta} = [0, 0, \dots, 1]^T \in \mathbb{R}^n$  w (5.15) oraz rozważmy przekształcenie ortogonalne  $\mathbf{Q}\mathbf{X}$ , takie że  $\mathbf{Q}\boldsymbol{\theta} = \boldsymbol{\mu}$ . Można w prosty sposób pokazać, że wektor losowy  $\mathbf{X}$  otrzymany w ten sposób, posiada rozkład brzegowy  $\mathbf{X}^T \boldsymbol{\theta} \stackrel{d}{=} t$ . Wybierając w odpowiedni sposób, rozkład dla zmiennej losowej  $t$  można w dowolny sposób kontrolować ilość masy prawdopodobieństwa skupionej wokół oczekiwanego kierunku  $\boldsymbol{\mu}$  [96]. W dalszej treści tego rozdziału, rekomendowane jest zastosowanie rozkładu brzegowego postaci  $t = 2X - 1$ , gdzie  $X$  jest zmienną losową o rozkładzie Beta  $\beta(a, b)$  [113]. W ten sposób, gęstość  $t$  jest określona przez [96]

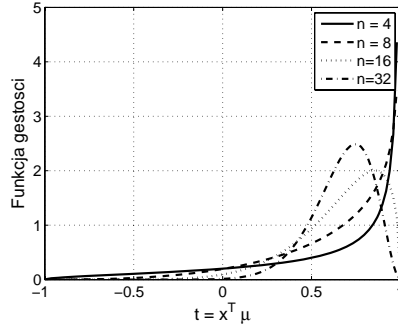
$$f(t|a, b) = \frac{2^{1-a-b}}{\beta(a, b)} (1-t)^{b-1} (1+t)^{a-1}, \quad (5.16)$$

gdzie  $a$  oraz  $b$  są parametrami rozkładu Beta. Parametr rozproszenia, który występował w rozkładach vMF, może zostać również uwzględniony w proponowanych rozkładach, poprzez przyjęcie odpowiednich wartości dla parametrów rozkładu (5.16) t.j.:  $a = \frac{n-1}{2}$  i  $b = \kappa \frac{n-1}{2}$ . Wartość oczekiwana zmiennej losowej  $T = \boldsymbol{\theta}^T \mathbf{X}$  oraz jej wariancja równe są:

$$E[T] = \frac{1-\kappa}{1+\kappa}, \quad \text{Var}(T) = \frac{8\kappa}{n(1+\kappa)^2(1-\kappa^2)}. \quad (5.17)$$

Jak wynika ze wzorów (5.17) wartość oczekiwana rozkładu brzegowego  $\boldsymbol{\theta}^T \mathbf{X}$ , nie zależy od wymiaru przestrzeni  $n$ . Tym samym wyeliminowana została główna wada rozkładów vMF, która skutecznie podważała ich użyteczność w wielowymiarowej optymalizacji ewolucyjnej. Wymiar przestrzeni przeszukiwań, ujawnia

się natomiast w definicji wariancji, co sprawia, że dla proponowanych rozkładów kierunkowych, wraz ze wzrostem wymiaru przestrzeni przeszukiwań, powinniśmy obserwować coraz mocniejsze zagęszczenie realizacji pseudolosowych wokół kierunku oczekiwanego. Efekt ten jest szczególnie widoczny na rysunku 5.4.



Rysunek 5.4: Wykres funkcji gęstości rozkładu brzegowego  $t = \mathbf{X}\boldsymbol{\mu}$  dla zaproponowanego rozkładu kierunkowego  $\mathbf{X} \sim \mathcal{M}(\boldsymbol{\mu}, \kappa)$

Z drugiej strony, przyjęty sposób parametryzacji proponowanych rozkładów kierunkowych, umożliwia otrzymanie rozkładu równomiernego na powierzchni kuli jednostkowej  $\kappa = 1$ , z drugiej zaś strony, rozkładu zdegenerowanego do wektora  $\boldsymbol{\mu}$  dla  $\kappa = 0$  [96]. Dużą zaletą, proponowanych rozkładów kierunkowych, jest prosty algorytm służący do generowania wektorów pseudolosowych - zaprezentowany w tabeli 5.1.

Tabela 5.1: Algorytm generowania wektorów pseudolosowych z rozkładu kierunkowego  $\mathcal{M}(\boldsymbol{\mu}, \kappa)$

---

*Dane wejściowe*

$\boldsymbol{\mu} \in \mathbb{R}^n$  – oczekiwany kierunek

$\kappa \in (0, 1]$  – parametr rozproszenia

*Dane wyjściowe*

$Y$  – wektor pseudolosowy rozkładu  $\mathcal{M}(\boldsymbol{\mu}, \kappa)$

*Algorytm*

$t = 2\beta\left(\frac{n-1}{2}, \frac{\kappa(n-1)}{2}\right) - 1$ , gdzie  $\beta(a, b)$  zwraca liczbę losową z rozkładu Beta

$X \leftarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_{n-1})$

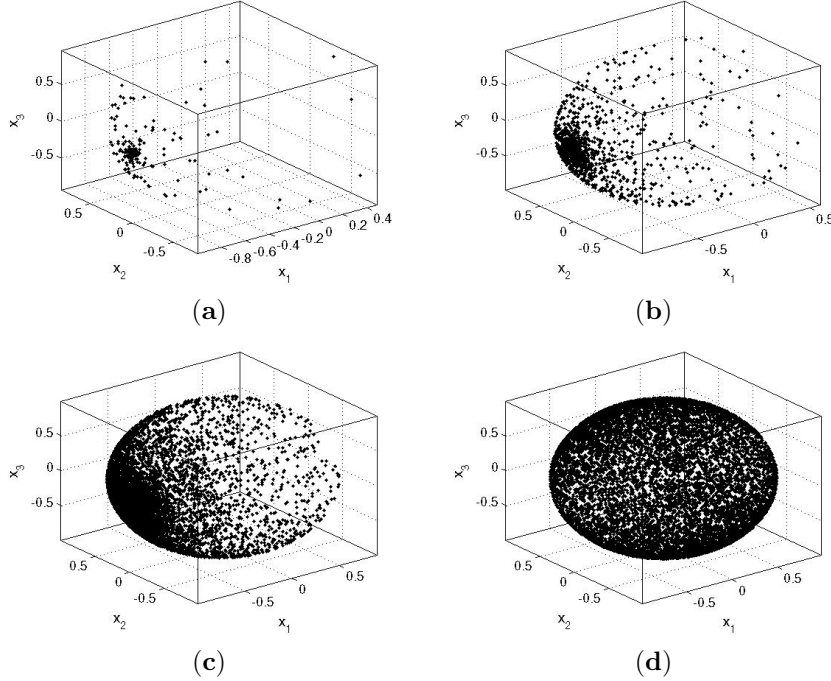
$Z \leftarrow X/\|X\|_2$

$Y \leftarrow [\sqrt{1-t^2}Z^T, t]^T$

$Y \leftarrow [\mathbf{I}_n - 2\mathbf{v}\mathbf{v}^T]Y$  gdzie  $\mathbf{v} = \frac{[0,0,\dots,1]^T - \boldsymbol{\mu}}{\|[0,0,\dots,1]^T - \boldsymbol{\mu}\|_2}$

---

Wykorzystując algorytm 5.1, na wykresie (5.5) zamieszczone zostały realizacje proponowanych rozkładów kierunkowych, dla różnych wartości parametru rozproszenia.



Rysunek 5.5: Realizacja 10000 wektorów pseudolosowych o rozkładzie kierunkowym  $M(\boldsymbol{\mu}, \kappa)$  dla  $\boldsymbol{\mu} = [-1, 0, 0]^T$  oraz : **(a)** -  $\kappa = 0.001$ , **(b)** -  $\kappa = 0.01$ , **(c)** -  $\kappa = 0.1$ , **(d)** -  $\kappa = 1$

Osobnym zagadnieniem, związanym bezpośrednio z zastosowaniem rozkładów kierunkowych w operatorze mutacji, jest kwestia doboru zmiennej mutującej  $r > 0$  w (5.12). Biorąc pod uwagę fakt, iż  $r$  może przyjmować postać dowolnej zmiennej losowej przyjmującej wartości dodatnie, dyskusja tego zagadnienia w ogólnym przypadku zdecydowanie wykracza poza zakres niniejszej rozprawy. W tym miejscu natomiast ograniczymy się jedynie do podkreślenia, że wybór ten może mieć decydujące znaczenie w świetle tzw. efektu otoczenia, dyskutowanego szerzej w rozdziale 4.2.2. Wystarczy wspomnieć, że efekt ten, odpowiadający za pogorszenie współczynnika zbieżności jak i powiększenie martwego obszaru wokół rozwiązań lokalnych, był konsekwencją zmian rozkładu normy wektora zaburzającego. Warto zauważyć, że w przypadku dekompozycji stochastycznej właściwej dla rozkładów kierunkowych  $\mathbf{Z} \sim \mathcal{M}(\boldsymbol{\mu}, \kappa)$ , rozkład normy wynosi  $\|\mathbf{Z}\| \stackrel{d}{=} r$ . Oznacza to, że w przypadku uniezależnienia, zmiennej mutującej  $r > 0$  od wymiaru przestrzeni przeszukiwań  $n$ , można w prosty sposób wyeliminować uciążliwy efekt otoczenia. Eksperymentalna część rozdziału dotyczyć będzie wyłącznie zastosowania rozkładów stabilnych jako zmiennych mutujących w kierunkowym operatorze mutacji.



### 5.3 Badania eksperymentalne

Badania eksperymentalne, których celem jest sprawdzenie skuteczności optymalizacyjnej kierunkowego operatora mutacji, przeprowadzone zostały dla algorytmu przeszukiwań ewolucyjnych z miękką selekcją 5.2. Schemat algorytmu zamieszczony został w tabeli 5.2. Jedyna różnica, pomiędzy wersją oryginalną algorytmu [41], a wersją proponowaną, polega na zastosowaniu rozkładów kierunkowych w procesie tworzenia populacji potomnej. W związku, przyjętą parametryzacją rozkładów kierunkowych, rodzi się pytanie, o sposób wyboru oczekiwanego kierunku mutacji  $\boldsymbol{\mu}$ . Przyjmując model optymalizacji, w którym niemożliwe jest korzystanie z analitycznej postaci gradientu funkcji celu, ani z jego numerycznego przybliżenia, zdecydowano się na porównanie trzech heurystyk mających na celu estymację najbardziej preferowanego kierunku przeszukiwania przestrzeni rozwiązań:

- Heurystyka nr 1:

$$\boldsymbol{\mu}^k = -\frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \quad \text{gdzie} \quad \mathbf{z} = [\mathbf{P}_k^T \mathbf{P}_k]^{-1} \mathbf{P}_k^T \boldsymbol{\Phi}_k, \quad (5.18)$$

gdzie  $\mathbf{P}_k \in R^{\eta \times n}$  oznacza macierz, której w poszczególne wiersze zawierają fenotypy osobników populacji w iteracji  $k$ , oraz odpowiadający im wektor dopasowania  $\boldsymbol{\Phi}_k \in R^\eta$ :

$$\mathbf{P}_t = [\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_\eta^t]^T \quad \boldsymbol{\Phi}_t = [\phi(\mathbf{x}_1^t), \phi(\mathbf{x}_2^t), \dots, \phi(\mathbf{x}_\eta^t)]^T$$

- Heurystyka nr 2:

$$\boldsymbol{\mu}^t = \frac{\mathbf{z}}{\|\mathbf{z}\|_2}, \quad \text{gdzie} \quad \mathbf{z} = \sum_{k=1}^{\eta} \frac{\phi(\mathbf{x}_k^{t-1}) - \phi(\mathbf{x}_k^t)}{\phi(\mathbf{x}_k^{t-1})} \frac{\mathbf{x}_k^t - \mathbf{x}_k^{t-1}}{\|\mathbf{x}_k^t - \mathbf{x}_k^{t-1}\|_2}. \quad (5.19)$$

- Heurystyka nr 3:

$$\boldsymbol{\mu}^t = \frac{\langle \mathbf{x}^t \rangle - \langle \mathbf{x}^{t-1} \rangle}{\|\langle \mathbf{x}^t \rangle - \langle \mathbf{x}^{t-1} \rangle\|}, \quad \text{where} \quad \langle \mathbf{x}^t \rangle = \frac{1}{\eta} \sum_{k=1}^{\eta} \mathbf{x}_k^t. \quad (5.20)$$

Heurystyka (5.18) traktuje populację osobników otrzymaną w wyniku selekcji jako zbiór węzłów aproksymacji, które w dalszej kolejności służą do odnalezienia najlepiej dopasowanej (w sensie średnio-kwadratowym) wielowymiarowej funkcji liniowej. Kierunek  $\boldsymbol{\mu}$  jest utożsamiany z kierunkiem najszybszego spadku, wskazywany przez liniową aproksymację rozkładu populacji. Do wad wspomnianej metody, niewątpliwie należy zaliczyć konieczność wykonywania operacji odwracania macierzy. W związku z tym, należy zagwarantować odpowiednią liczebność populacji, aby operacja ta mogła w ogóle zostać wykonana. Heurystyka (5.19) zaproponowana została po raz pierwszy w [106], dla algorytmów znanych jako ewolucyjne przeszukiwanie gradientowe (ang. *Evolutionary Gradient Search*). Dużą zaletą, proponowanej metod estymacji kierunku  $\boldsymbol{\mu}$  jest kumulowanie doświadczeń z każdej próby losowego poprawienia rozwiązania. Heurystyka (5.20) natomiast, zaczerpnięta została z wcześniej wspomnianego algorytmu przeszukiwań ewolucyjnych z miękką selekcją zaproponowanego przez Obuchowicza [85].

Tabela 5.2: Algorytm przeszukiwań ewolucyjnych z miękką selekcją oraz wymuszonym kierunkiem mutacji.  $ESSS_\alpha - DM$

---

**Dane wejściowe**

- $\eta$  – rozmiar populacji;
- $T_{\max}$  – maksymalna liczba generacji;
- $\sigma, \alpha, \kappa$  – parametry mutacji: skala, indeks stabilności, rozproszenie;
- $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  – funkcja celu;
- $\mathbf{x}_0^0$  – punkt startowy.

**Algorytm**

1. Inicjacja

$$P(0) = (\mathbf{x}_1^0, \mathbf{x}_2^0, \dots, \mathbf{x}_\eta^0), \quad \mathbf{x}_k^0 = \mathbf{x}_0^0 + \mathbf{Z},$$

gdzie  $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_n)$ ,  $k = 1, 2, \dots, \eta$

2. Powtarzaj

(a) Ocena

$$\Phi(P(t)) = (q_1^t, q_2^t, \dots, q_\eta^t), \quad \text{gdzie } q_k^t = \phi(\mathbf{x}_k^t), \quad k = 1, 2, \dots, \eta.$$

(b) Miękka selekcja

$$P(t) \longrightarrow P(t)' = (\mathbf{x}_{h_1}^t, \mathbf{x}_{h_2}^t, \dots, \mathbf{x}_{h_\eta}^t),$$

(c) Estymacja kierunku mutacji przy użyciu heurystyk (5.18)-(5.20)

$$\boldsymbol{\mu}(t) \longleftarrow H(P'(t), P'(t-1)),$$

(d) Mutacja

$$P(t)' \longrightarrow P(t+1);$$

$$\mathbf{x}_k^{t+1} = \mathbf{x}_{h_k}^t + \chi_{\alpha, \sigma} \mathbf{d}^{(n)}, \quad \mathbf{d}^{(n)} \sim \mathcal{M}(\boldsymbol{\mu}(t), \kappa), \quad \chi_{\alpha, \sigma} \sim |S_\alpha S(\sigma)| \quad k = 1, 2, \dots, \eta.$$

Dopóki  $t > T_{\max}$ .

---

**Przebieg eksperymentu**

Eksperyment zakładał porównanie algorytmu ewolucyjnego 5.2, opartego o mutacje kierunkowe oraz wspomagające je heurystyki (5.18)-(5.20). Badania skupiały się na czterech najbardziej popularnych funkcjach testowych  $\phi_1, \phi_9, \phi_{13}, \phi_{14}$ . Mimo że oryginalny algorytm przeszukiwań ewolucyjnych z miękką selekcją [41], zakłada zastosowanie selekcji proporcjonalnej (Tabela 3.3), zdecydowano się na przeprowadzenie testów także dla selekcji turniejowej (Tabela 3.2). Źródła takiej decyzji należy upatrywać w fakcie zastosowania rozkładów stabilnych do zaburzania rozwiązań na odpowiednich kierunkach. Istnieje bowiem podejrzenie, że makromutacje, które są właściwe zwłaszcza dla rozkładów o mniejszym indeksie stabilności  $\alpha$ , mogą w przypadku selekcji proporcjonalnej znacząco przyczynić się do spadku efektywności całego algorytmu. Wspomniane przypuszczenie, nie jest całkiem bezpodstawne, jeśli wziąć pod uwagę sposób w jaki wartości dopa-

sowania poszczególnych osobników, przekształcane są w celu przeprowadzenie selekcji, t.j:

$$\phi_i = \phi^* - \phi(\mathbf{x}_i) + \frac{1}{\eta^2} \quad (5.21)$$

gdzie  $\{\mathbf{x}_i\}_{i=1}^{\eta}$  jest populacją fenotypów,  $\phi(\cdot)$  funkcją celu poddawaną minimalizacji,  $\frac{1}{\eta^2}$  dodatkowym składnikiem, umożliwiającym wybór także najgorszego osobnika w populacji, oraz  $\phi^* = \max_i \phi(\mathbf{x}_i)$ . Kłopoty z selekcją proporcjonalną, opartą na powszechnie stosowanym przekształceniu (5.21), polegają na częstym zaburzeniu selekcji w wyniku zastosowania rozkładu stabilnego. Mianowicie, istnieje duża szansa<sup>1</sup>, iż osobnik powstały w wyniku makromutacji będzie posiadał znacząco gorsze dopasowanie od reszty populacji. Przekształcenie (5.21) powoduje w tym wypadku, że prawdopodobieństwa przeżycia pozostałych osobników w dużej mierze wyrównują się, co w skrajnych przypadkach może prowadzić nawet do efektu podobnego do błędzenia losowego. Przeprowadzenie badań dla algorytmu ESSS z selekcją turniejową, która jest pozbawiona wspomnianej wady, mają również na celu eksperymentalne zweryfikowanie wpływu typu zastosowanej selekcji na skuteczność algorytmu.

Warto zauważyć, że zastosowanie algorytmu (5.2) wymaga skonfigurowania zestawu zewnętrznych parametrów kontrolnych. W celu bliższej analizy wzajemnych korelacji pomiędzy parametrami  $\alpha, \sigma, \kappa$  oraz typem zastosowanej heurystyki, zdecydowano się na przeprowadzenie całej serii eksperymentów, uwzględniających wszystkie możliwe kombinacje następujących wartości:

$$\begin{aligned} \alpha &= \{2.0, 1.75, 1.5, 1.25, 1.0, 0.75, 0.5\}, \\ \sigma &= \{0.01, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10\}, \\ \kappa &= \{0.01, 0.1, 0.25, 0.5, 0.75, 1\}. \end{aligned}$$

Każda konfiguracja parametrów  $\{\alpha, \sigma, \kappa\}$ , wraz z jedną z trzech rozważanych heurystyk, oceniana była na podstawie 100 niezależnych uruchomień algorytmu 5.2. Populacje początkowe, identycznie skonfigurowanych algorytmów ESSS, otrzymywane były przez zaburzenie rozkładem normalnym  $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$  wspólnego punktu  $\mathbf{x}_0$ . Miękka selekcja oraz brak mechanizmu adaptacji parametru skali rozkładu mutującego, powodują że algorytmy te nie są zbieżne do rozwiązań globalnych, ale dążą do ustalenia równowagi selekcyjno-mutacyjnej w pewnej odległości od optimum [56]. Dlatego też zastosowano specyficzne kryteria stopu. Przede wszystkim ustalono maksymalną liczbę obliczeń funkcji celu:  $T_{\max} = 10000$  dla  $\phi_1, \phi_9$  oraz  $T_{\max} = 100000$  dla  $\phi_{13}, \phi_{14}$ . Oprócz tego, algorytmy zostawały zatrzymywane w chwili, gdy najlepsze rozwiązanie nie uległo poprawie: dla funkcji jednomodalnych - w przeciągu dwustu następujących po sobie mutacji oraz dla funkcji wielomodalnych - w przeciągu 500 mutacji. Po zakończeniu działania algorytmów, zapamiętywano jednocześnie liczbę obliczeń funkcji celu jak i jakość otrzymanego rozwiązania. Zastosowany rozmiar populacji  $\eta = 50$  został dobrany, głównie z myślą o uniknięciu problemów z operacją odwracania macierzy, właściwej dla heurystyki nr 1.

## Wyniki eksperymentu

Analizę wyników uzyskanych w powyżej opisanym eksperymencie, zacznijmy od ustalenia najlepszych konfiguracji parametrów  $\{\alpha, \sigma, \kappa\}$  dla zadanych proble-

<sup>1</sup>w zależności od postaci funkcji celu

mów testowych. W tym celu dla każdego zestawu parametrów oraz każdej heurystyki adaptacyjnej, wyznaczono medianę wartości funkcji celu zwracaną przez każdy z algorytmów *ESSS*. Naturalnie, za najlepszą konfigurację przyjęto tą, dla której wartość ta była najmniejsza. W tabeli 5.3 zamieszczono optymalne zestawy  $\{\alpha, \sigma, \kappa\}$  dla dwuwymiarowych funkcji celu. Tabele 5.6, 5.9 zawierają rezultaty dla trzydziestowymiarowych wersji problemów testowych oraz selekcji proporcjonalnej i turniejowej.

Tabela 5.3: Optymalne wartości parametrów  $\alpha^*, \kappa^*, \sigma^*$  otrzymane dla algorytmu *ESSS* z mutacją kierunkową oraz różnymi heurystykami wyznaczania kierunku mutacji. *SLOFC* - Średnia Liczba Obliczeń Funkcji Celu, *SWFC* - Średnia Wartość Funkcji Celu. Wymiar przestrzeni przeszukiwań  $n = 2$ , selekcja turniejowa

**ESSS + Heurystyka nr 1**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	Me[SWFC]	Me[SLOFC]
$\phi_1$	0.5	1	0.01	$2.4487e - 007$	2495
$\phi_9$	1	0.75	0.1	0.24094	328.5
$\phi_{13}$	1.5	0.1	0.5	1.0544	373.5
$\phi_{14}$	0.5	0.25	0.1	0.033006	917.5

**ESSS + Heurystyka nr 2**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	0.5	1	0.01	$3.0264e - 007$	2455
$\phi_9$	1.99	0.75	0.1	0.1708	302
$\phi_{13}$	0.5	0.5	0.1	1.0176	567.5
$\phi_{14}$	0.5	0.75	0.1	0.067061	692.5

**ESSS + Heurystyka nr 3**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	0.5	1	0.01	$2.0941e - 007$	2517
$\phi_9$	1.99	0.1	0.01	$1.3824e - 005$	2355
$\phi_{13}$	0.5	0.75	0.1	1.0094	541
$\phi_{14}$	0.5	0.5	1	0.37435	533.5

Tabela 5.4: Współczynniki korelacji dla zmiennych niezależnych  $\alpha, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$ . Wymiar przestrzeni przeszukiwań  $n = 2$ , selekcja turniejowa.

**Funkcja celu  $\phi_1$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	0.31	0.31	0.30	0.33	0.34	0.33
$\sigma$	-0.23	-0.21	-0.23	-0.25	-0.24	-0.25

**Funkcja celu  $\phi_9$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	0.12	0.09	0.14	0.08	0.09	0.08
$\sigma$	0.95	0.87	0.93	0.95	0.96	0.92

**Funkcja celu  $\phi_{13}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	0.26	0.33	0.33	0.37	0.38	0.38
$\sigma$	-0.26	-0.25	-0.28	-0.21	-0.21	-0.21

**Funkcja celu  $\phi_{14}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	0.50	0.30	0.27	0.45	0.38	0.39
$\sigma$	-0.51	0.73	0.46	-0.64	-0.68	-0.67

Tabela 5.5: Współczynniki korelacji liniowej dla zmiennych niezależnych  $\kappa, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$  - mediany wartości funkcji celu wyznaczone na podstawie 100 niezależnych uruchomień algorytmu *ESSS* z heurystykami:  $H_1, H_2, H_3$ . Wymiar przestrzeni przeszukiwań  $n = 2$ , selekcja turniejowa

**Funkcja celu  $\phi_1$**

Zmienne	$H_1, 1.99$	$H_2, 1.99$	$H_3, 1.99$	$H_1, 0.50$	$H_2, 0.50$	$H_3, 0.50$
$\kappa$	0.03	0.05	0.04	-0.33	-0.25	-0.28
$\sigma$	-0.39	-0.39	-0.39	-0.11	-0.13	-0.14

**Funkcja celu  $\phi_9$**

Zmienne	$H_1, 1.99$	$H_2, 1.99$	$H_3, 1.99$	$H_1, 0.50$	$H_2, 0.50$	$H_3, 0.50$
$\kappa$	-0.05	-0.09	-0.14	0.01	-0.04	-0.06
$\sigma$	0.95	0.90	0.93	0.94	0.90	0.92

**Funkcja celu  $\phi_{13}$**

Zmienne	$H_1, 1.99$	$H_2, 1.99$	$H_3, 1.99$	$H_1, 0.50$	$H_2, 0.50$	$H_3, 0.50$
$\kappa$	0.03	0.04	0.01	-0.20	0.04	-0.09
$\sigma$	-0.42	-0.41	-0.41	0.66	0.76	0.59

**Funkcja celu  $\phi_{14}$**

Zmienne	$H_1, 1.99$	$H_2, 1.99$	$H_3, 1.99$	$H_1, 0.50$	$H_2, 0.50$	$H_3, 0.50$
$\kappa$	0.14	-0.16	-0.14	0.20	-0.58	-0.55
$\sigma$	-0.88	-0.71	-0.76	-0.21	-0.13	-0.30

**Obserwacja 23** *Obserwując optymalne zestawy parametrów otrzymane dla dwuwymiarowych problemów testowych, pierwszą rzeczą która przykuwa uwagę jest stosunkowo duża wartość parametru rozproszenia rozkładów kierunkowych. Zastosowanie licznej  $\eta = 50$  (jak na ten wymiar przestrzeni przeszukiwań) populacji powoduje, że kwestia modelowania najbardziej korzystnego kierunku zostaje zepchnięta na drugi plan. Decydujący wpływ na skuteczność algorytmu *ESSS* ma w tym przypadku mechanizm makromutacji gwarantowany przez rozkłady stabilne z niskim indeksem  $\alpha$ . Potwierdzenie tej obserwacji znajdziemy również w tabeli 5.4, gdzie dodatnie współczynniki korelacji indeksu stabilności i mediany funkcji celu  $\phi_1, \phi_{13}, \phi_{14}$  wyraźnie wskazują na wyższość rozkładów o cięższych ogonach. Warto również podkreślić brak wyraźnej korelacji pomiędzy współczynnikiem rozproszenia  $\kappa$  a otrzymywanymi wynikami (Tabela 5.5). Biorąc pod uwagę dużą liczebność populacji w stosunku do wymiaru przestrzeni przeszukiwań, należy przypuszczać, że w każdej iteracji, bez względu na wyróżniony kierunek, co najmniej kilka mutacji przeprowadzonych zostaje w poprawnym kierunku. W takiej sytuacji, przewaga mechanizmu makromutacji, który powoduje szybsze przemieszczenie w pobliże optimum globalnego, nie jest niczym dziwnym. Z tego też względu nie sposób doszukać się znaczącej różnicy pomiędzy wynikami uzyskiwanymi dla różnych heurystyk wyznaczających kierunek mutacji. Odstępstwem od wyżej naszkicowanej reguły, są wyniki uzyskane dla źle uwarunkowanego problemu  $\phi_9$ . Aby wytłumaczyć zjawisko jakie odgrywa w tym przypadku dominującą*

rolę, warto zauważyć, iż prawdopodobieństwo sukcesu dąży do  $P_s \rightarrow 0.5$ , jeśli tylko skala rozkładu dąży do  $\sigma \rightarrow 0$ . W przypadku dużych populacji, okazuje się, że najefektywniejszym rozwiązaniem jest zastosowanie operatora mutacji, który będzie powodował niewielkie, ale często kończące się sukcesami, przemieszczenia w przestrzeni rozwiązań. Potwierdzenie tej obserwacji odnajdziemy w niezwykle silnym skorelowaniu parametru skali rozkładów oraz wynikami uzyskanymi dla funkcji  $\phi_9$  - tabele 5.4 oraz 5.5.

Tabela 5.6: Optymalne wartości parametrów  $\alpha^*$ ,  $\kappa^*$ ,  $\sigma^*$  otrzymane dla algorytmu *ESSS* z mutacją kierunkową oraz różnymi heurystykami wyznaczania kierunku mutacji. *SLOFC* - Średnia Liczba Obliczeń Funkcji Celu, *SWFC* - Średnia Wartość Funkcji Celu. Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja proporcjonalna

**ESSS + Heurystyka nr 1**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	2.0	0.1	5	505.4189	6794.5
$\phi_9$	2.0	0.25	0.1	3950.4333	2572.5
$\phi_{13}$	2.0	0.1	2	244.6408	2782.5
$\phi_{14}$	2.0	0.1	10	11.6101	3184.5

**ESSS + Heurystyka nr 2**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	2.0	0.1	5	7596.2018	6924.5
$\phi_9$	2.0	0.1	1	18778.1374	341
$\phi_{13}$	2.0	0.5	2	406.4481	2832
$\phi_{14}$	2.0	0.5	0.01	19.9727	19607

**ESSS + Heurystyka nr 3**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	2.0	0.1	5	31618.8829	6651
$\phi_9$	2.0	0.25	1	24163.9718	305
$\phi_{13}$	2.0	0.75	2	416.7339	3182
$\phi_{14}$	2.0	0.75	0.01	19.9731	26814

Tabela 5.7: Współczynniki korelacji liniowej dla zmiennych niezależnych  $\kappa, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$  - mediany wartości funkcji celu wyznaczone na podstawie 100 niezależnych uruchomień algorytmu *ESSS* z heurystykami:  $H_1, H_2, H_3$ . Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja proporcjonalna

**Funkcja celu  $\phi_1$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.57	0.63	0.57	0.89	0.94	0.82
$\sigma$	-0.63	-0.51	-0.58	-0.28	-0.17	-0.36

**Funkcja celu  $\phi_9$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.53	0.59	0.08	0.85	0.94	0.38
$\sigma$	0.39	0.04	0.33	-0.16	0.05	-0.18

**Funkcja celu  $\phi_{13}$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.11	-0.08	-0.19	0.86	0.80	0.56
$\sigma$	-0.04	0.30	0.31	-0.15	0.26	0.31

**Funkcja celu  $\phi_{14}$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.24	-0.13	-0.13	0.11	-0.26	-0.29
$\sigma$	-0.49	0.47	0.46	0.46	0.62	0.66



Tabela 5.8: Współczynniki korelacji liniowej dla zmiennych niezależnych  $\alpha, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$ . Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja proporcjonalna.

**Funkcja celu  $\phi_1$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.36	0.24	-0.40	0.34	0.32	0.32
$\sigma$	-0.47	-0.57	-0.32	-0.82	-0.87	-0.85

**Funkcja celu  $\phi_9$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.29	0.00	-0.37	-0.45	-0.52	-0.46
$\sigma$	-0.06	-0.04	0.31	0.28	0.23	0.08

**Funkcja celu  $\phi_{13}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.39	-0.32	-0.53	-0.66	-0.67	-0.67
$\sigma$	-0.50	0.43	0.38	0.35	0.38	0.38

**Funkcja celu  $\phi_{14}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.23	-0.25	-0.25	-0.33	-0.36	-0.35
$\sigma$	-0.70	0.22	0.21	0.45	0.45	0.44

**Obserwacja 24** Makromutacje które przyczyniają się do obniżenia presji selekcyjnej, spowodowały, że obserwowane w tabeli 5.6 optymalne zbiory parametrów  $\{\alpha, \sigma, \kappa\}$  całkowicie zdominowane zostały przez rozkład normalny. Obserwując współczynniki korelacji zamieszczone w tabeli 5.7 wyraźnie widać, że wraz ze wzrostem wymiaru przestrzeni przeszukiwań, coraz większego znaczenia nabiera parametr rozproszenia  $\kappa$ . Szczególnie silne zależności odnotowane zostały dla rozkładów mutujących o niewielkich indeksach stabilności  $\alpha$ . Warto również zauważyć, iż zaproponowana w pracy heurystyka nr 1, okazała się znacznie skuteczniejsza od dwóch rozwiązań zaproponowanych przez Obuchowicza [85] i Salomona [106]. Dla funkcji multimodalnych  $\phi_{13}, \phi_{14}$  można zaobserwować, wspólną dla wszystkich heurystyk, przedwczesną zbieżność algorytmu. Niemniej jednak duża liczba iteracji wykonywana przed spełnieniem kryterium stopu, pozwala wyznaczyć wiarygodne zależności statystyczne pomiędzy poszczególnymi parametrami. Współczynniki korelacji zamieszczone w Tabeli 5.8, pokazują wyraźnie, że tylko w przypadku modelu sferycznego  $\phi_1$  nastawienie operatora mutacji na wykonywanie znaczących zaburzeń powoduje polepszenie rezultatu. Dla mutacji izotropowych  $\kappa = 1$  świadczą o tym dodatnie współczynniki korelacji dla indeksu stabilności oraz ujemne wartości dla parametru skali. Zaskakujące jest, iż zupełnie odwrotne korelacje można zaobserwować dla pozostałych funkcji testowych. Skonfigurowanie rozkładu mutującego pod kątem wykonywania zaburzeń o znacznej sile, co zwykle przemawia za ułatwieniem w omijaniu optimum lokalnych, w tym wypadku powoduje pogorszenie wyników. Obserwując dwukrotnie słabszą zależność dla indeksu stabilności dla funkcji nasyczonej  $\phi_{14}$ , winę za ta-

ki stan rzeczy ponosi operator selekcji, który pod wpływem makromutacji traci zdolności ukierunkowywania procesu ewolucji.

Tabela 5.9: Optymalne wartości parametrów  $\alpha^*$ ,  $\kappa^*$ ,  $\sigma^*$  otrzymane dla algorytmu *ESSS* z mutacją kierunkową oraz różnymi heurystykami wyznaczania kierunku mutacji. *SLOFC* - Średnia Liczba Obliczeń Funkcji Celu, *SWFC* - Średnia Wartość Funkcji Celu. Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja turniejowa

**ESSS + Heurystyka nr 1**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	1.25	0.25	1	4.5129	9773
$\phi_9$	1.99	0.5	0.1	403.3957	6893.5
$\phi_{13}$	1.99	0.1	5	224.9375	2021.5
$\phi_{14}$	1.75	0.1	10	3.546	3648.5

**ESSS + Heurystyka nr 2**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	0.75	0.5	1	15.1904	9890.5
$\phi_9$	1.75	0.25	0.1	304.8702	5975
$\phi_{13}$	1.75	1	5	305.6067	2474.5
$\phi_{14}$	1.0	0.75	0.01	19.9664	11061.5

**ESSS + Heurystyka nr 3**

$\phi(\cdot)$	$\alpha^*$	$\kappa^*$	$\sigma^*$	SWFC	SLOFC
$\phi_1$	1	0.5	2	25.381	9302
$\phi_9$	1.99	0.25	0.1	236.0726	6946.5
$\phi_{13}$	1.5	1	5	306.8533	2481.5
$\phi_{14}$	1.25	0.5	0.01	19.9665	10580.5

Tabela 5.10: Współczynniki korelacji liniowej dla zmiennych niezależnych  $\alpha, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$ . Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja turniejowa.

**Funkcja celu  $\phi_1$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.01	-0.01	-0.07	0.57	0.57	0.57
$\sigma$	-0.17	-0.55	-0.75	-0.53	-0.53	-0.53

**Funkcja celu  $\phi_9$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.54	-0.34	0.04	-0.02	-0.01	-0.02
$\sigma$	0.56	0.44	0.23	0.99	0.99	0.99

**Funkcja celu  $\phi_{13}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	0.67	0.60	0.68	0.11	0.10	0.12
$\sigma$	0.08	-0.04	0.13	-0.75	-0.76	-0.75

**Funkcja celu  $\phi_{14}$**

Zmienne	H <sub>1</sub> ,0.01	H <sub>2</sub> ,0.01	H <sub>3</sub> ,0.01	H <sub>1</sub> ,1.00	H <sub>2</sub> ,1.00	H <sub>3</sub> ,1.00
$\alpha$	-0.05	0.01	0.07	0.05	0.05	0.04
$\sigma$	0.68	0.70	0.58	0.81	0.81	0.81

Tabela 5.11: Współczynniki korelacji liniowej dla zmiennych niezależnych  $\kappa, \sigma$  oraz trzech zmiennych objaśnianych  $Me[H_1], Me[H_2], Me[H_3]$ . Wymiar przestrzeni przeszukiwań  $n = 30$ , selekcja turniejowa.

**Funkcja celu  $\phi_1$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.15	0.22	0.14	-0.55	-0.65	-0.58
$\sigma$	-0.77	-0.75	-0.80	0.10	0.01	-0.06

**Funkcja celu  $\phi_9$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	-0.16	-0.29	-0.37	-0.54	-0.52	-0.46
$\sigma$	0.93	0.81	0.69	0.67	0.63	0.69

**Funkcja celu  $\phi_{13}$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	-0.11	-0.12	-0.26	-0.40	0.01	-0.22
$\sigma$	-0.71	-0.69	-0.60	-0.39	-0.48	-0.46

**Funkcja celu  $\phi_{14}$**

Zmienne	H <sub>1,2.0</sub>	H <sub>2,2.0</sub>	H <sub>3,2.0</sub>	H <sub>1,0.50</sub>	H <sub>2,0.50</sub>	H <sub>3,0.50</sub>
$\kappa$	0.16	-0.05	-0.16	0.12	-0.08	-0.14
$\sigma$	-0.38	0.76	0.71	-0.19	0.84	0.80

**Obserwacja 25** Zastosowanie selekcji turniejowej o względnie dużym rozmiarze grupy turniejowej, powoduje wzrost ciśnienia selekcyjnego, który powoduje, że szanse przeżycia mają tylko osobniki najlepsze. Ma to swoje odzwierciedlenie w otrzymanych najlepszych konfiguracjach parametrów (Tabela 5.9). Można zaobserwować, że w większości przypadków, optymalny parametr  $\kappa$  wcale nie nakłada silnej presji na wybór najbardziej korzystnego kierunku mutacji. W dwóch przypadkach, najlepszymi rozkładami eksploracyjnymi okazały się nawet rozkłady izotropowe, co czyni wątpliwym stosowanie mutacji kierunkowej dla selekcji elitarniej. Warto również zauważyć, że w odróżnieniu od algorytmów z selekcją proporcjonalną, optymalne rozkłady eksploracyjne oparte są o indeksy stabilności  $\alpha < 2$ . Dla modelu sferycznego  $\phi_1$ , godnym odnotowania jest znacznie słabszy wpływ parametru skali dla rozkładów  $\alpha = 0.5$  niż to ma miejsce dla rozkładu normalnego  $\alpha = 2.0$ . Omawiany przy okazji obserwacji poczynionych dla dwuwymiarowej  $\phi_9$ , mechanizm zaburzeń o małej sile, znajduje również potwierdzenie w Tabeli (5.10). Mianowicie, dla rozkładu izotropowego można zaobserwować bardzo silną zależność (współczynnik korelacji liniowej bliski 1) skuteczności algorytmu ESSS od parametru skali  $\sigma$ .

## 5.4 Podsumowanie

W rozdziale tym zaprezentowane zostało pojęcie tzw. mutacji kierunkowej. W pierwszej kolejności, na podstawie prostego przykładu obliczeniowego, uwypuklone zostały największe wady mutacji: izotropowych, symetrycznych, a także

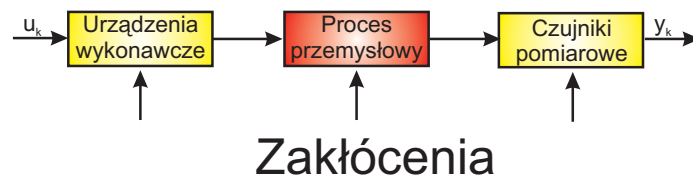
mutacji opartych o popularne rozkłady kierunkowe von Misess-Fisher'a (vMF). Pogorszenie skuteczności optymalizacyjnej wraz ze wzrostem wymiaru przestrzeni przeszukiwań, w przypadku tych pierwszych jest nader oczywiste. W przypadku rozkładów kierunkowych vMF, pogorszenie efektywności algorytmów ewolucyjnych, jest spowodowane efektem podobnym do efektu otoczenia obserwowanego dla normalnego rozkładu zaburzającego. Przypomnijmy, iż mimo iż wielowymiarowa funkcja gęstości posiada wokół punktu bazowego, to zwiększenie liczby stopni swobody powoduje, że obszar najintensywniejszego próbkowania oddala się od punktu bazowego. Identyczny efekt obserwujemy dla rozkładów mVF. Mimo, iż ich funkcja gęstości zawsze posiada największą wartość dla najkorzystniejszego kierunku mutacji  $\mu$ , to w miarę wzrostu wymiaru przestrzeni jest on coraz rzadziej wybierany. Aby zniwelować ten niepożądany efekt, w rozdziale tym, zaproponowane zostały rozkłady kierunkowe, dla których oczekiwany kierunek mutacji nie zależy od wymiaru przestrzeni przeszukiwań. Proponowany rozkład kierunkowy w połączeniu z mutującą stabilną zmienną losową, został następnie wykorzystany w operatorze mutacji algorytmu przeszukiwań ewolucyjnych z miękką selekcją. W procedurze adaptacyjnej zastosowano trzy heurystyki służące do estymacji najbardziej korzystnego kierunku mutacji. Szereg, przeprowadzonych w systematyczny sposób, eksperymentów, pozwolił na wyznaczenie statystycznych liniowych zależności pomiędzy poszczególnymi parametrami  $\{\alpha, \sigma, \kappa\}$  oraz heurystykami adaptacyjnymi. Na podstawie otrzymanych wyników, oczywistym jest, iż stosowanie rozkładów kierunkowych znajduje gruntowane uzasadnienie tylko w przypadku problemów o znacznej liczbie wymiarów oraz gdy specyfika problemu nie pozwala na wykonywanie dużej liczby obliczeń wartości funkcji celu. Warto również zaznaczyć, że w takich znacznie skuteczniejsze okazują się heurystyki  $H_1, H_2$ , biorące pod uwagę nie tylko względną zmianę położenia populacji, ale również wartości dopasowania poszczególnych osobników.

Znamienną cechą przy stosowaniu rozkładów ciężkoogonowych w stochastycznych algorytmach optymalizacji globalnej, jest fakt, iż wiele z dotychczas obserwowanych, dla mutacji gaussowskiej, zależności musi zostać zweryfikowanych. Nie inaczej ma się kwestia związana z selekcją proporcjonalną stosowaną w algorytmie ESSS. Wyniki badań eksperymentalnych, przedstawione w powyższym rozdziale, pokazują wyraźnie, że w przypadku makromutacji, bardzo często ten schemat selekcji, okazuje się skrajnie nieefektywny. Dlatego też w kontekście stosowania rozkładów stabilnych rekomenduje się wykorzystanie innych miękkich procedur selekcji np. selekcję turniejową.

## Rozdział 6

# Algorytmy ewolucyjne w zadaniach projektowania systemów wykrywania uszkodzeń

W dzisiejszych czasach, gdy komputerowe techniki sterowania, planowania oraz wspomaganie procesu decyzyjnego, z powodzeniem znajdują zastosowanie w różnych gałęziach przemysłu, podwyższeniu ulegają standardy związane z kwestiami bezpieczeństwa i niezawodności systemów przemysłowych. Jedną z metod wychodzących na przeciw stale rosnącym wymaganiom w tej kwestii jest zastosowanie tzw. systemów wykrywania i identyfikacji uszkodzeń procesów przemysłowych (ang. *Fault Detection and Isolation Systems - FDI*). Głównym celem wspomnianych systemów, jest nadzorowanie przebiegu pewnego procesu przemysłowego. W przypadku, gdy proces ten przebiega w sposób inny od zamierzonego, systemy FDI mają za zadanie wykryć taką sytuację, zdefiniować ich przyczynę oraz przedłożyć stosowny raport systemowi wsparcia decyzyjnego. Z reguły systemy przemysłowe są obiektami o dużym stopniu złożoności, w skład których wchodzi następujące podsystemy: urządzeń wykonawczych, procesu przemysłowego oraz czujników pomiarowych.



Rysunek 6.1: Komponenty systemu przemysłowego

Klasyczne podejście do problemu wykrywania uszkodzeń polega na obserwowaniu pojedynczej zmiennej procesowej oraz bezustannym konfrontowaniu jej wielkości z dopuszczalnym zakresem (ustalonym na podstawie wiedzy eksper-

kiej). Złożoność współczesnych systemów przemysłowych, wymusza odejście od wspomnianej klasycznej metody, w stronę skuteczniejszych oraz bardziej zaawansowanych technik. Jedną z metod podniesienia niezawodności systemów FDI jest wykorzystanie wiedzy pozyskiwanej ze wszystkich możliwych czujników pomiarowych [129], w które rozważany system jest zaopatrzony. Wykorzystanie całej wiedzy zawartej w charakterystykach pomiarowych jest możliwe, przy zastosowaniu tzw. systemów wykrywania uszkodzeń na podstawie modelu systemu (ang. Model-based FDI system). Istnieją dwa podejścia do budowy modelu [122]:

- modelowanie behawioralne
- modelowanie fenomenologiczne

Pierwsze z nich polega na wykorzystaniu dalece specyficznej wiedzy dziedzinowej o systemie podlegającym modelowaniu. Modele te tworzone są w oparciu o podstawowe prawa fizyczne, a ich struktura niejednokrotnie sprowadza się do pokaźnych układów nieliniowych równań różniczkowych. W konsekwencji stosowania tego podejścia, otrzymuje się modele o dobrych właściwościach predykcyjnych, a ścisła interpretacja fizyczna parametrów pozwala na przeprowadzenie wiarygodnej analizy procesu. Z drugiej zaś strony, często modele te są na tyle skomplikowane, że zastosowanie klasycznego aparatu matematycznego do wiarygodnej analizy ich działania, nie jest możliwe w trybie on-line. W związku z tym, często obserwowalną praktyką jest upraszczanie oryginalnych modeli np. za pomocą linearyzacji.

Modelowanie behawioralne w całości odbywa się na podstawie danych, zgromadzonych podczas obserwacji systemu. Struktura modelu, jak i jego parametryzacja, w żaden sposób nie muszą odwzorowywać rzeczywistych praw rządzących modelowanym systemem. Mimo, że otrzymane modele zwykle charakteryzują się stosunkowo małym stopniem złożoności, to ich zdolności predykcyjne są z reguły znacznie gorsze niż w przypadku ich fenomenologicznych odpowiedników. Wybór jednej z dwóch wymienionych wyżej metodologii, nie jest oczywisty. Co więcej, brak jest ogólnych reguł mogących posłużyć za wyznacznik w tej kwestii.

## 6.1 Statyczne systemy FDI oparte o neuronowe generatory residuum

### 6.1.1 Sztuczne sieci neuronowe

Skuteczność systemów FDI opartych o model procesu, w dużej mierze zależy od precyzji z jaką model odzwierciedla zachowanie nadzorowanego systemu. Z tego też względu, etap identyfikacji procesu odgrywa kluczową rolę przy projektowaniu systemów wykrywania uszkodzeń. Na przestrzeni ostatniego dziesięciolecia, w literaturze poświęconej metodą identyfikacji [60],[129], bardzo dużo miejsca poświęcono tzw. sztucznych sieciom neuronowym - SSE. Sieć neuronowa może być postrzegana jako zbiór połączonych ze sobą pojedynczych jednostek przetwarzających tzw. neuronów, które z matematycznego punktu widzenia opisuje równanie:

$$y = g(\mathbf{u}^T \boldsymbol{\theta}) \quad (6.1)$$

W literaturze poświęconej sieciom neuronowym,  $g(\cdot)$  jest określane mianem funkcji aktywacji,  $\theta = [\theta_1, \theta_1, \cdot, \theta_{n_u}, b]^T$  jest wektorem parametrów ( $\theta$  - wagi połączeń,  $b$  - bias),  $\mathbf{u}^T = [u_1, u_2, \cdot, u_{n_u}, 1]^T$  jest wektorem sygnałów docierającym do neuronu,  $y$  jest jego odpowiedzią. W kontekście zastosowań sztucznych sieci neuronowych w zadaniach identyfikacji systemów nieliniowych, najczęściej stosuje się dwa typy funkcji aktywacji:

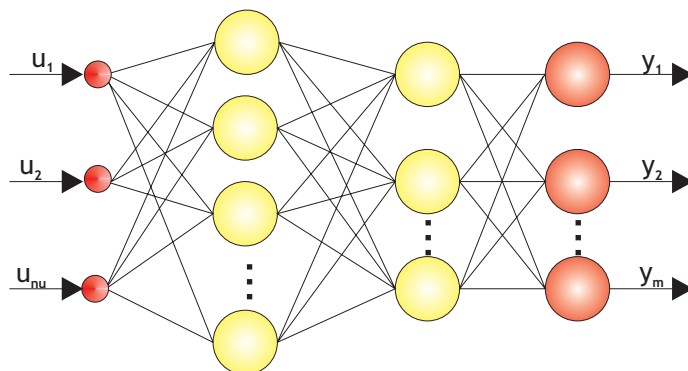
- sigmoidalną

$$g(x) = \frac{1}{1 + \exp(-x)} \quad (6.2)$$

- tangensoidalną

$$g(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)} \quad (6.3)$$

Neurony zwykle zorganizowane są struktury, które określają topologię połączeń między nimi. Do najpopularniejszych struktur sieci neuronowych, należy tzw. perceptron wielowarstwowy. Dla tego typu sieci, neurony uporządkowane są w warstwy, w taki sposób, aby pojedynczy neuron posiadał połączenia ze wszystkimi neuronami warstw sąsiednich. Brak jest natomiast połączeń pomiędzy neuronami leżącymi w tej samej warstwie. Hierarchiczna struktura perceptronu wielowarstwowego narzuca także kierunek przepływu sygnałów: od warstwy wejściowej, poprzez warstwy ukryte, do warstwy wyjściowej. Przykład tego typu sieci zamieszczony został na rysunku 6.2.



Rysunek 6.2: Perceptron wielowarstwowy złożony z warstwy wejściowej, dwóch warstw ukrytych oraz warstwy wyjściowej

Popularność sztucznych sieci neuronowych wynika przede wszystkim z faktu, iż mogą one być postrzegane jako tzw. uniwersalne aproksymatory. Na mocy twierdzenia Stone-Weierstrass'a, można mianowicie pokazać, iż nawet jednowarstwowa sieć z nieliniowymi jednostkami przetwarzającymi typu tangensoidalnego, może przybliżyć każdą funkcję ciągłą z dowolnie małym błędem. Innym, ale również nie bez znaczenia powodem popularności SSN, jest rozpowszechnienie wielu doskonałych pakietów numerycznych, umożliwiających ich zastosowanie np. Matlab, Statistica, SAS.

Proces identyfikacji statycznego sytemu nieliniowego za pomocą perceptronu



wielowarstwowego przebiega zwykle według następującego schematu. W pierwszej kolejności następuje zgromadzenie tzw. zbioru uczącego, na który składają się dane wejściowe modelowanego systemu oraz odpowiedzi na zadane wymuszenia. Następnie, dla określonej struktury sieci neuronowej, następuje proces strojenia parametrów poszczególnych jednostek przetwarzających. W tym celu opracowano wiele wydajnych algorytmów optymalizacji, które wykorzystując technikę wstecznej propagacji błędu [117], pozwalają w efektywny sposób na uczenie sieci nawet z setkami jednostek przetwarzających. W przypadku spełnienia przez wynikową sieć odpowiednich założeń dotyczących błędu modelowania, można ona być podstawą do budowy generatora residuum w systemach wrywania uszkodzeń.

Z zastosowaniem tym, wiąże się jednak wiele problemów, które mają decydujący wpływ na skuteczność systemów FDI. Należy bowiem, pamiętać iż w rzeczywistych systemach przemysłowych dane potrzebne do utworzenia zbioru uczącego, pochodzą z różnego rodzaju czujników pomiarowych o pewnej tolerancji błędu pomiaru. W związku z tym, dane zawsze będą charakteryzować się pewnymi odchyleniami od rzeczywistej wartości, które z matematycznego punktu widzenia, można zamodelować jako pewien dodatkowy składnik losowy, wpływający na wartości zmiennych procesowych. Jest rzeczą niezwykle naiwną sądzić, że wskazany przez algorytm uczenia model neuronowy, jest jedyną i najlepszą siecią, mogącą w adekwatny sposób odzwierciedlać zależności wejściowo-wyjściowe rozważanego systemu. Mianowicie, należy uwzględnić fakt, iż proces strojenia parametrów modelu neuronowego przebiega na podstawie danych o charakterze losowym. W konsekwencji, otrzymane parametry również posiadać będą charakter stochastyczny. Zatem istnieje będzie cała rodzina sieci neuronowych, które, zakładając pewien poziom prawdopodobieństwa, mogą z powodzeniem odzwierciedlać zależności obserwowane w zbiorze uczącym. Z matematycznego punktu widzenia, zbiór parametrów wspomnianej rodziny, nazywany jest obszarem ufności, a jego charakterystyka wykorzystywana jest do zdefiniowania tzw. niepewności parametrycznej modelu neuronowego. Oczywiście czym mniejsza rodzina modeli akceptowalnych, tym większa pewność, że otrzymany w procesie uczenia model neuronowy, jest odzwierciedleniem obserwowanego procesu. Tym samym, usprawiedliwione staje się przypuszczenie, że model o mniejszej niepewności parametrycznej, może posłużyć do budowy bardziej niezawodnych i skuteczniejszych systemów FDI. Zagadnienie minimalizacji niepewności parametrycznej modeli neuronowych, za pomocą techniki planowania eksperymentu oraz metod ewolucyjnej optymalizacji globalnej, leży w centrum zainteresowania kolejnych rozdziałów rozprawy.

### 6.1.2 Niepewność parametryczna modelu neuronowego

Niniejsze rozważania, rozpoczniemy od precyzyjnego scharakteryzowania niepewności parametrycznej modelu neuronowego. Załóżmy, że odpowiedź systemu  $y_s$  w chwili  $k$  wyrażona jest zależnością:

$$y_{s,k} = y_m(\boldsymbol{\theta}^*, \mathbf{x}_k) + \epsilon_k \quad (6.4)$$

gdzie  $y_{m,k}(\boldsymbol{\theta}^*, \mathbf{x}_k)$  jest odpowiedzią modelu dla pewnego wektora parametrów  $\boldsymbol{\theta}^* \in \mathbb{R}^{n_p}$  i wejścia  $\mathbf{x}_k \in \mathbb{R}^{n_u}$ , a  $\{\epsilon_k\}$  jest ciągiem niezależnych zmiennych losowych o rozkładzie normalnym tj.  $\epsilon_k \sim \mathcal{N}(0, \sigma)$ . W dalszym ciągu założymy, że znajomość systemu rzeczywistego, ograniczona jest do znajomości jego

struktury, natomiast, zarówno wektor parametrów  $\boldsymbol{\theta}^*$  jak i odchylenie standardowe  $\sigma$  pozostają nieznanymi. Pierwszy etap zadania estymacji parametrycznej polega na zgromadzeniu zbioru uczącego tj. zestawu danych wejściowych oraz odpowiadających im wyjść systemu  $\Xi = \{\mathbf{x}_k, y_{s,k}\}_{k=1}^N = [\mathbf{X}, \mathbf{Y}_s]$ ,  $\mathbf{X} \in \mathbb{R}^{N \times n_u}$ ,  $\mathbf{Y}_s \in \mathbb{R}^N$ . Następnie, zadanie estymacji sprowadza się do poszukiwania wektora parametrów  $\boldsymbol{\theta}$ , który w najadekwatniejszy sposób będzie tłumaczył zależność pomiędzy danymi wejściowymi i odpowiedziami systemu obserwowane w zbiorze uczącym  $\Xi$ . Warto zauważyć, że stochastyczny charakter wielkości  $y_{s,k} \sim \mathcal{N}(y_m(\boldsymbol{\theta}^*, \mathbf{x}_k), \sigma)$  zgromadzonych w zbiorze uczącym (wzór 6.4), powoduje, że wyznaczony na ich podstawie wektor parametrów  $\hat{\boldsymbol{\theta}}$  również będzie posiadał charakter losowy. Jedną z najpopularniejszych metod wyznaczania wektora  $\hat{\boldsymbol{\theta}}$  jest tzw. Metoda Największej Wiarygodności - MNW (ang. *Maximum Likelihood Estimation*) [122]. Estymatę największej wiarygodności  $\hat{\boldsymbol{\theta}}$  otrzymuje się poprzez rozwiązanie zadania maksymalizacji

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \mathbb{R}^{n_p}} \pi(\mathbf{Y}|\boldsymbol{\theta}), \quad (6.5)$$

gdzie  $\pi(\mathbf{Y}|\boldsymbol{\theta})$  jest tzw. funkcją wiarygodności. Funkcja ta, dla ustalonego  $\boldsymbol{\theta}$  jest funkcją gęstości rozkładu odpowiedzi systemu. W MNW zakłada się znajomość  $\mathbf{Y} = \mathbf{Y}_s$ , co powoduje, że (6.5) staje się automatycznie funkcją wektora  $\boldsymbol{\theta}$ . Estymatory otrzymane w wyniku (6.5) posiadają wiele cennych właściwości, które decydują o ich szerokim stosowaniu. Przede wszystkim,  $\hat{\boldsymbol{\theta}}$  jest estymatorem zgodnym, najefektywniejszym, mającym asymptotyczny rozkład normalny  $\mathcal{N}(\boldsymbol{\theta}^*, \mathbf{F}^{-1}(\boldsymbol{\theta}^*))$  [122]<sup>1</sup>. Wykorzystując, założenia poczynione przy zależności (6.4) oraz monotoniczność funkcji  $\log(\cdot)$ , zadanie (6.5) można przedstawić w alternatywnej postaci [122]:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{n_p}} - \sum_{k=1}^N \log(\pi(\mathbf{Y}_s|\boldsymbol{\theta})) = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{n_p}} S(\boldsymbol{\theta}) \quad (6.6)$$

gdzie

$$S(\boldsymbol{\theta}) = \sum_{k=1}^N (y_{s,k} - y_m(\mathbf{x}_k, \boldsymbol{\theta}))^2 = \|\mathbf{Y}_s - \mathbf{Y}_m(\boldsymbol{\theta})\|^2 \quad (6.7)$$

oraz  $\mathbf{Y}_s = [y_{s,1}, y_{s,2}, \dots, y_{s,N}]^T$ ,  $\mathbf{Y}_m(\boldsymbol{\theta}) = [y_m(\mathbf{x}_1, \boldsymbol{\theta}), y_m(\mathbf{x}_2, \boldsymbol{\theta}), \dots, y_m(\mathbf{x}_N, \boldsymbol{\theta})]^T$ , a  $\|\cdot\|$  oznacza normę euklidesową.

Z praktycznego punktu widzenia asymptotyczne właściwości estymatora (6.6) są bez większego znaczenia, gdyż zbiór  $\Xi$  zawsze zawiera zawsze skończoną liczbę par uczących. Należy przy tym zaznaczyć, że w przypadku modeli nieliniowych, w tym rozważanych sieci neuronowych, rozkład prawdopodobieństwa (6.6) pozostaje nieznanym, co niezwykle komplikuje testowanie hipotez statystycznych. Jedną z najpopularniejszych technik, mających na celu umożliwienie przeprowadzenia wnioskowania statystycznego dotyczącego estymatora  $\hat{\boldsymbol{\theta}}$ , jest przybliżenie modelu nieliniowego, jego liniowym odpowiednikiem. W ten sposób wyniki testowania hipotez statystycznych otrzymane dla uproszczonego modelu liniowego modelu uogólnia się na model nieliniowy. Z tego względu, usprawiedliwione wydaje się przytoczenie poniżej najważniejszych faktów dotyczących wnioskowania

<sup>1</sup> $\mathbf{F}$  jest tzw. macierzą informacyjną Fisher'a

statystycznego dla modeli liniowych postaci:

$$y_m = \mathbf{x}_k^T \boldsymbol{\theta}. \quad (6.8)$$

Przed wszystkim zadanie minimalizacji (6.6) dla modelu (6.8) posiada analityczne rozwiązanie postaci:

$$\hat{\boldsymbol{\theta}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{Y}_s, \quad (6.9)$$

gdzie  $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T]^T \in \mathbb{R}^{N \times n_u}$ . Ważną konsekwencją rozwiązania (6.9) jest znajomość rozkładu estymatora  $\hat{\boldsymbol{\theta}}$ , tj. uwzględniając deterministyczny charakter  $\mathbf{X}$  oraz rozkład wektora  $\mathbf{Y}_s = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ ,  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ , można pokazać, że

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}^*, \sigma^2 \mathbf{F}^{-1}), \quad (6.10)$$

gdzie

$$\mathbf{F} = \mathbf{X}^T \mathbf{X} = \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \quad (6.11)$$

jest macierzą informacyjną Fisher'a. Znajomość rozkładu estymatora, umożliwia ponadto wyznaczenie:

- obszaru ufności  $\Theta_{c,L}$  wyznaczony na poziomie  $1 - \alpha$  dla parametrów  $\boldsymbol{\theta}$  [17]:

$$\Theta_{c,L} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{n_p} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{F} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \hat{\sigma}^2 n_p F(n_p, N - n_p; \alpha) \right\} \quad (6.12)$$

gdzie

$$\hat{\sigma}^2 = \frac{S(\hat{\boldsymbol{\theta}})}{N - n_p} \quad (6.13)$$

jest estymatą wariancji obliczoną na podstawie zbioru uczącego, a  $F(n_p, N - n_p; \alpha)$  jest górnym  $\alpha$  kwantylem rozkładu Fisher'a o  $n_p$  i  $N - n_p$  stopniach swobody.

- Przedziału ufności odpowiedzi modelu dla  $\mathbf{x}_k$ , wyznaczony na poziomie  $1 - \alpha$ :

$$Y_{c,L}(\mathbf{x}_k) = \left\{ y \in \mathbb{R} : |y_m(\mathbf{x}_k, \hat{\boldsymbol{\theta}}) - y| \leq \hat{\sigma} \sqrt{\mathbf{x}_k^T \mathbf{F}^{-1} \mathbf{x}_k} t(N - n_p; \alpha/2) \right\}, \quad (6.14)$$

gdzie  $t(N - n_p; \alpha/2)$  jest górnym  $\alpha/2$  kwantylem rozkładu studenta o  $N - n_p$  stopniach swobody.

W ogólności, dla modeli nieliniowych, (6.6) nie posiada analitycznego rozwiązania. Niemniej jednak, w literaturze odnajdziemy wiele skutecznych metod numerycznych mogących posłużyć do rozwiązania zadania optymalizacji (6.6). W środowisku obliczeń neuronowych jedną z najczęściej stosowanych metod jest algorytm Levenberg'a-Marquardt'a. Obszar ufności, dla parametrów modeli nieliniowych, może przyjmować skomplikowany kształt i ogólnie może być wyrażony w następującej postaci:

$$\Theta_{c,N} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{n_p} : S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) \leq \hat{\sigma}^2 n_p F(n_p, N - n_p; \alpha) \right\} \quad (6.15)$$

Podobnie, można sformułować ogólną postać przedziałów ufności:

$$Y_{c,N}(\mathbf{x}_k) = \left\{ y \in \mathbb{R} : y = y_m(\mathbf{x}_k, \boldsymbol{\theta}_0) \mid \boldsymbol{\theta}_0 \in \Theta \right\} \quad (6.16)$$

Brak metod służących do dokładnego wyznaczania postaci (6.15) oraz (6.16), jest głównym powodem przybliżania modeli nieliniowych, poprzez rozwinięcie funkcji  $y_m(\cdot)$  w szereg Taylora wokół najlepszej estymaty  $\hat{\boldsymbol{\theta}}$ , tj.:

$$y_m^L(\mathbf{x}_k, \boldsymbol{\theta}) \approx y_m(\mathbf{x}_k, \hat{\boldsymbol{\theta}}) + [\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k)]^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \quad (6.17)$$

gdzie

$$\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k) = \left. \frac{\partial y_m(\mathbf{x}_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (6.18)$$

Informacyjna macierz Fishera dla (6.17) przyjmuje postać:

$$\mathbf{F}(\hat{\boldsymbol{\theta}}) = \sum_{k=1}^N [\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k)] [\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k)]^T \quad (6.19)$$

Wykorzystując (6.17), można wyznaczyć obszar oraz przedział ufności dla modelu nieliniowego

- Obszar (6.15) zostaje przybliżony elipsoidą:

$$\hat{\Theta}_{c,N} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^{n_p} : (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{F}(\hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \hat{\sigma}^2 n_p F(n_p, N - n_p; \alpha) \right\}, \quad (6.20)$$

- Zbiór (6.16) przybliża się przedziałem ufności:

$$\hat{Y}_{c,N}(\mathbf{x}_k) = \left\{ y \in \mathbb{R} : |y_m(\mathbf{x}_k, \hat{\boldsymbol{\theta}}) - y| \leq \hat{\sigma} t(N - n_p; \alpha/2) \cdot \sqrt{[\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k)]^T \mathbf{F}(\hat{\boldsymbol{\theta}})^{-1} [\mathbf{v}(\hat{\boldsymbol{\theta}}, \mathbf{x}_k)]} \right\}. \quad (6.21)$$

### 6.1.3 Minimalizacja niepewność parametrycznej modelu neuronowego

Wzory (6.15) oraz (6.16) ukazują, że zarówno wielkość obszaru jak i przedziału ufności dla modelu (6.17) zależą od sekwencji wejść zgromadzonych w zbiorze uczącym [99, 130]. Z tego też względu, dokonując odpowiedniego wyboru przykładów uczących, można minimalizować niepewność parametryczną modeli nieliniowych, a co za tym idzie zwiększać wiarygodność systemów FDI budowanych na ich podstawie [97, 99, 130]. Dobór wejść systemu dokonuje się poprzez sformułowanie tzw. ciągłego planu eksperymentu:

$$\boldsymbol{\xi} = \left\{ \begin{array}{cccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n_e} \\ \mu_1 & \mu_2 & \cdots & \mu_{n_e} \end{array} \right\} \quad (6.22)$$

W teorii planowania eksperymentu, (6.22) jest utożsamiany z miarą probabilistyczną o nośniku  $\{\mathbf{x}_k\}_{k=1}^{n_e}$ , i prawdopodobieństwach  $\sum_{k=1}^{n_e} \mu_k = 1$ . Z praktycznego punktu widzenia, waga  $\mu_k$  w bezpośredni sposób przekłada się na liczbę pomiarów wyjścia systemu, jakie należy wykonać dla sygnału wejściowego. Na

podstawie ciągłego planu eksperymentu (6.22) tworzona jest macierz informacyjna Fisher'a, która w dalszej kolejności wykorzystywana jest do wyznaczenia zbiorów ufności (6.15), (6.16):

$$\mathbf{F}(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}}) = \sum_{k=1}^{n_e} \mu_k \left[ \frac{\partial y_m(\mathbf{u}_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right] \left[ \frac{\partial y_m(\mathbf{u}_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right]^T \quad (6.23)$$

Minimalizacja niepewności parametrycznej z wykorzystaniem ciągłych planów eksperymentu sprowadza się do minimalizacji pewnej funkcji skalarnej  $\phi(\cdot)$  określonej na macierzy (6.23). Funkcja  $\phi(\cdot)$  w zależności od kontekstu, może przyjmować różną postać:

- Kryterium  $D$ -optymalności:

$$\phi(\mathbf{F}) = \det(\mathbf{F}(\hat{\boldsymbol{\theta}})), \quad (6.24)$$

- Kryterium  $E$ -optymalności:

$$\phi(\mathbf{F}) = \lambda_{\max}(\mathbf{F}(\hat{\boldsymbol{\theta}})), \quad (6.25)$$

gdzie  $(\lambda_{\max}(\mathbf{A}))$  oznacza największą wartość własną macierzy  $\mathbf{A}$ .

- Kryterium  $A$ -optymalności:

$$\phi(\mathbf{F}) = \text{trace } \mathbf{F}(\hat{\boldsymbol{\theta}}), \quad (6.26)$$

- Kryterium  $G$ -optymalności:

$$\phi(\mathbf{F}) = \max_{\mathbf{x}} [\mathbf{v}(\mathbf{x}, \hat{\boldsymbol{\theta}})]^T \mathbf{F}(\hat{\boldsymbol{\theta}}) [\mathbf{v}(\mathbf{x}, \hat{\boldsymbol{\theta}})]. \quad (6.27)$$

Kryterium  $D$ - oraz  $E$ -optymalności powoduje odpowiednio zminimalizowanie objętości elipsoidy ufności (6.20) oraz minimalizację jej najdłuższego promienia. Przyjęcie kryterium  $A$ -optymalności oznacza minimalizację średniej wariancji estymatora największej wiarygodności, natomiast dla planu  $G$ -optymalnego otrzymujemy minimalizację wariancji odpowiedzi modelu neuronowego. Mając na względzie, że nadrzędnym celem niniejszego rozdziału jest budowa modelu neuronowego o najmniejszej niepewności związanej z jego wyjściem, dobór ciągłego planu eksperymentu powinien odbywać się w oparciu o kryterium  $G$ -optymalności. Jak powszechnie wiadomo, problemy typu min-max, należą do jednych z najgorzej uwarunkowanych zadań optymalizacji. Na szczęście, na mocy twierdzenia Kiefer-Wolfowitz'a, mówiącym o równoznaczności kryteriów  $D$ - i  $G$ -optymalności, możemy skutecznie ominąć trudności związane z optymalizacją min-max, skupiając się na poszukiwaniu planu eksperymentu rozwiązując poniższe zadanie:

$$\boldsymbol{\xi}^* = \arg \min_{\boldsymbol{\xi}} \det(\mathbf{F}(\boldsymbol{\xi}, \hat{\boldsymbol{\theta}})), \quad (6.28)$$

gdzie macierz informacyjna  $\mathbf{F}$  jest wyznaczana na podstawie (6.23). Jednym z najczęściej stosowanych algorytmów służących do znajdowania planów  $D$ -optymalnych jest metoda zaproponowana przez Wynn'a i Fedorov'a. Algorytm Wynn'a-Fedorov'a znajdowania  $D$ -optymalnych planów eksperymentu dla nieliniowych modeli parametrycznych składa się z kilku etapów [130, 129, 131]:

**krok 0)** Dokonaj wstępnej estymaty parametrów sieci neuronowej  $\hat{\theta}$ . Przyjmij  $k = 1$  oraz wybierz początkowy plan eksperymentu  $\xi_k$ , dla którego  $\det(\mathbf{F}(\xi_k, \hat{\theta})) > 0$ . Ustal maksymalną liczbę iteracji  $n_{\max}$  oraz dokładność wyznaczenia planu  $\epsilon > 0$

**krok 1)** Wyznacz  $\mathbf{u}_k$  według:

$$\mathbf{u}_k = \arg \max_{\mathbf{u}} \phi(\mathbf{u}, \xi_k) \quad (6.29)$$

gdzie

$$\phi(\mathbf{u}, \xi_k) = [\mathbf{v}(\hat{\theta}, \mathbf{u})]^T \mathbf{F}(\xi_k, \hat{\theta}) [\mathbf{v}(\hat{\theta}, \mathbf{u})] \quad (6.30)$$

**krok 2)** Jeśli  $\phi(\mathbf{u}_k, \xi_k)/n_p < 1 + \epsilon$  to przerwij obliczenia i przyjmij  $\xi_k$  jako plan  $D$ -optymalny.

**krok 3)** Oblicz wagę skojarzoną z nowym punktem skupienia planu  $\mathbf{u}_k$ :

$$\mu_k = \frac{\phi(\mathbf{u}_k, \xi_k) - n_p}{(\phi(\mathbf{u}_k, \xi_k) - 1)n_p} \quad (6.31)$$

**krok 4)** Utwórz nowy plan eksperymentu  $\xi_{k+1}$ , dokonując kombinacji wypukłej planów  $\xi_k$  i  $\{\mathbf{u}_k\}$ :

$$\xi_{k+1} = \left\{ \begin{array}{cccccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n_e} & \mathbf{u}_k \\ (1 - \mu_k)\mu_1 & (1 - \mu_k)\mu_2 & \cdots & (1 - \mu_k)\mu_{n_e} & \mu_k \end{array} \right\} \quad (6.32)$$

**krok 5)** Jeśli  $k = n_{\max}$  zakończ algorytm, w przeciwnym razie zwiększ  $k = k + 1$  i przejdź do kroku 1).

Skuteczność zaprezentowanego algorytmu, zależy w dużym stopniu od pierwszego kroku, który ze względu na multimodalność funkcji wariancji  $\phi(\mathbf{u}, \xi_k)$  wymaga zastosowania technik optymalizacji globalnej. Warto w tym miejscu podkreślić korzyści jakie wynikają ze skuteczniejszego rozwiązywania problemu optymalizacji globalnej w algorytmie Wynn'a-Fedorov'a. Krok pierwszy wspomnianego algorytmu ma za zadanie wytypowanie punktu w przestrzeni wejściowej, który należy dodać do aktualnego planu. W związku z tym, naturalnym jest, iż większa precyzja w lokalizowaniu nowego punktu skutkuje mniejszym rozmiarem planu optymalnego [130]. Pamiętając o tym, że plany ciągłe, konstruowane są z myślą o gromadzeniu zbioru uczącego, oczywistym jest, że skuteczne techniki optymalizacji globalnej mogą w znacznej mierze przyczynić się do poprawienia aspektów ekonomicznych problemu tworzenia modeli neuronowych.

W literaturze jedną z powszechnie stosowanych metod do optymalizacji funkcji wariancji, jest stochastyczny algorytm Adaptacyjnych Przeszukiwań Losowych (ang. Adaptive Random Search) [93, 122, 128], (zaprezentowany w tabeli 6.1). Znamienne jest, że technika ta bazuje jedynie na rozkładzie normalnym. Z treści poprzednich rozdziałów niniejszej rozprawy, jasno wynika, iż rozkład normalny nie jest na ogół optymalnym rozkładem stabilnym. Tym samym, usprawiedliwione jest przypuszczenie mówiące o tym, iż zadanie optymalizacji funkcji wariancji, w algorytmie Wynn'a-Fedorova, może być skuteczniej rozwiązane za

pomocą innych modeli stabilnych aniżeli wspomniany rozkład. Ewentualne korzyści z podniesienia efektywności algorytmu w rozwiązywaniu zadania (6.29) są oczywiste. Z jednej strony nastąpiłoby przyspieszenie procesu wyznaczania planu D-optimalnego, z drugiej zaś, wynikowy plan, mógłby zawierać zdecydowanie mniej punktów skupienia, co oznaczałoby zmniejszenie liczby pomiarów niezbędnych w procesie estymacji parametrów modelu neuronowego. W celu weryfikacji wspomnianej hipotezy, zdecydowano się przeprowadzić analizę skuteczności algorytmu ARS wzbogaconego o rozkłady stabilne -  $ARS_{\alpha}$ , w zadaniach maksymalizacji funkcji wariancji (6.29)

Tabela 6.1: Algorytm adaptacyjnych przeszukiwań losowych  $ARS_\alpha$

---

**Dane wejściowe**

$\mathbf{x}_0 \in \mathbb{R}^n$  – początkowe przybliżenie rozwiązania

$\mathbf{L}_B, \mathbf{U}_B \in \mathbb{R}^n$  – wektory definiujące obszar przeszukiwań, tj.:  $L_B(i) \leq x(i) \leq U_B(i)$  ;

$j_{\max}$  – maksymalna liczba iteracji etapu adaptacji rozkładu eksploracyjnego;

$i_{\max}$  – liczba różnych wariantów rozkładu eksploracyjnego;

$A$  – współczynnik skalujący ;

$\alpha$  – indeks stabilności ;

$iter_{\max}$  – maksymalna liczba iteracji ;

**Dane wyjściowe**

$\mathbf{x}^*$  – przybliżenie rozwiązania globalnego ;

**Inicjacja**

$$\sigma_{i,j} = \frac{A}{10^{i-1}} (U_B(j) - L_B(j)); \quad i = 1, \dots, i_{\max}; \quad j = 1, \dots, n$$

$$e_i = \lceil \frac{j_{\max}}{i} \rceil; \quad i = 1, \dots, i_{\max}$$

**Algorytm**

Ustal  $iter = 1$

Powtarzaj

Ustal  $i = 1$

Powtarzaj

Ustal  $j = 1, i_{best} = 1$

Powtarzaj

$$\mathbf{x}' = \prod_{\mathbf{X}} (\mathbf{x}^* + \mathbf{Z}); \quad \mathbf{Z} = [S_\alpha S(\sigma_{i,1}), S_\alpha S(\sigma_{i,2}), \dots, S_\alpha S(\sigma_{i,n})]^T$$

Jeśli  $\phi(\mathbf{x}') > \phi(\mathbf{x}^*)$  to ustal  $\mathbf{x}^* = \mathbf{x}', i_{best} = i$

Zwiększ  $j = j + 1$

Dopóki  $j < e_i$

Zwiększ  $i = i + 1$

Dopóki  $i < i_{\max}$

Ustal  $i = 1$

Powtarzaj

$$\mathbf{x}' = \prod_{\mathbf{X}} (\mathbf{x}^* + \mathbf{Z}); \quad \mathbf{Z} = [S_\alpha S(\sigma_{i_{best},1}), S_\alpha S(\sigma_{i_{best},2}), \dots, S_\alpha S(\sigma_{i_{best},n})]^T$$

Jeśli  $\phi(\mathbf{x}') > \phi(\mathbf{x}^*)$  to ustal  $\mathbf{x}^* = \mathbf{x}'$

Zwiększ  $i = i + 1$

Dopóki  $i < e_1$

Ustal  $iter = iter + 1$

Dopóki  $iter \leq iter_{\max}$

---



### Optymalny algorytm $ARS_\alpha$

Głównym celem badań eksperymentalnych prezentowanych w niniejszym rozdziale jest znalezienie optymalnej konfiguracji algorytmu adaptacyjnych przeszukiwań losowych - ARS, opartego o zaburzone rozkłady stabilne, w zadaniach maksymalizacji funkcji wariancji modeli neuronowych. Przedmiotem eksperymentu jest sieć neuronowa o dwuwymiarowej przestrzeni wejściowej  $n_u = 2$ , czterech tangensoidalnych jednostkach w warstwie ukrytej  $n_h = 4$  oraz jednym liniowym neuronie wyjściowym. Całkowita liczba parametrów modelu wynosi  $n_p = 17$ . W celu precyzyjnego dopasowania algorytmu ARS do problemu (6.29), testom poddano następujące konfiguracje algorytmu  $ARS_\alpha$ :

$$\begin{aligned}\alpha &= [2.0, 1.5, 1, 0.5], \\ A &= [0.1, 0.25, 0.5, 0.75, 1, 4/3, 2, 4, 10], \\ i_{\max} &= [2, 3, 4, 5] \\ j_{\max} &= [20, 40, 80, 100]\end{aligned}$$

Każda kombinacja powyższych parametrów, oceniana była na podstawie liczby obliczeń funkcji celu, niezbędnej do rozwiązania problemu (6.29). Oczywiście rzeczywiste optimum funkcji wariancji nie jest znane, i dlatego zanim zostało użyte jako punkt odniesienia do oceny skuteczności różnie skonfigurowanych algorytmów ARS, należało je bardzo starannie oszacować. W tym celu, wykonywano wstępną optymalizację polegającą na szczelnym zaścieleniu, zbioru rozwiązań siatką punktów testowych. Dla każdego z punktów wyznaczana była wartość funkcji wariancji, a następnie z punktu o największej wartości uruchamiana była procedura optymalizacji gradientowej. Zastosowanie siatki o rozmiarze  $200 \times 200$  pozwala przypuszczać, że rezultatem przedstawionej procedury było rzeczywiste optimum globalne. W rezultacie, wstępna faza optymalizacji charakteryzowała się nieporównywalnie większą złożonością obliczeniową niż testowane algorytmy ARS.

Każda konfiguracja algorytmu ARS oceniana była poprzez wyznaczenie średniej liczby epok niezbędnych do zlokalizowania maksimum globalnego  $\mathbf{x}^*$ . Zadanie to przyjęto uważać za rozwiązane, w momencie osiągnięcia zadowalającego przybliżenia optimum globalnego  $\hat{\mathbf{x}}$  t.j:  $\|\hat{\mathbf{x}} - \mathbf{x}^*\| < 0.01$ . Wyznaczenie wskaźnika jakości następowało na podstawie uśrednienia wyników eksperymentu przeprowadzonych dla stu niezależnych modeli neuronowych. Poszczególne modele otrzymywane były poprzez wylosowanie wektora parametrów sieci z rozkładem  $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\mathbf{0}, 2\mathbf{I}_{17})$ . Następnie, dla otrzymanego modelu  $y_m(\mathbf{x}, \hat{\boldsymbol{\theta}})$ , konstruowany był pierwszy plan eksperymentu. W celu uniknięcia sytuacji, w której początkowa macierz informacyjna Fisher'a, byłaby macierzą osobliwą, zdecydowano się na równomierne rozmieszczenie punktów skupienia początkowego planu. Do tego celu wykorzystano tzw. pseudo-losowe sekwencje Haltona [76], które eliminują redundancje przykładów uczących. Ponadto, aby numerycznie poprawić uwarunkowanie zadania odwracania macierzy Fisher'a, początkowy plan eksperymentu zawierał większą liczbę punktów skupienia, niż to teoretycznie konieczne [4] t.j.  $n_e = n_p + 5$ . Następnie, dla każdego modelu, wykonywano dokładnie 10 iteracji algorytmu Wynn'a-Fedorov'a. W ten sposób dla każdej konfiguracji algorytmu ARS otrzymano 1000 wyników optymalizacji, co może świadczyć o dużej wiarygodności otrzymanych wskaźników jakości. Optymalne konfiguracje, wyznaczone na podstawie najmniejszej liczby obliczeń funkcji celu zaprezen-

towane zostały w tabeli 6.2. Ponadto, aby odnieść się do kwestii odporności rozkładów stabilnych na wybór poszczególnych parametrów konfiguracyjnych algorytmu ARS, w tabeli 6.3 przedstawiono, najgorsze ustawienia otrzymane w wyniku przeprowadzonego eksperymentu.

Tabela 6.2: Optymalne konfiguracje algorytmu adaptacyjnych przeszukiwań losowych dla rozkładów stabilnych - **ARS** $_{\alpha}$ , **SLOFC** - Średnia Liczba Obliczeń Funkcji Celu. Ostatni wiersz tabeli odnosi się do algorytmu skonfigurowanego w najczęściej rekomendowany sposób [122], [128].

Algorytm	$A$	$i_{\max}$	$j_{\max}$	<b>SLOFC</b>
<b>ARS</b> $_{0.5}$	0.5	2	20	338.91
<b>ARS</b> $_1$	4	4	40	263.20
<b>ARS</b> $_{1.5}$	1	3	80	254.27
<b>ARS</b> $_2$	2	3	20	216.03
<b>ARS</b> $_2 - stand.$	1	5	100	321.59

Tabela 6.3: Najgorsze konfiguracje algorytmu adaptacyjnych przeszukiwań losowych dla rozkładów stabilnych - **ARS** $_{\alpha}$ , **SLOFC** - Średnia Liczba Obliczeń Funkcji Celu.

Algorytm	$A$	$i_{\max}$	$j_{\max}$	<b>SLOFC</b>
<b>ARS</b> $_{0.5}$	0.1	5	20	797.71
<b>ARS</b> $_1$	0.1	5	40	1103.44
<b>ARS</b> $_{1.5}$	0.1	5	40	1582.60
<b>ARS</b> $_2$	0.1	2	20	2718.34

**Obserwacja 26** Wyniki zaprezentowane w tabeli 6.2 wyraźnie pokazują przewagę rozkładu normalnego w zadaniach optymalizacji funkcji wariancji dla modeli neuronowych. Warto jednak zwrócić uwagę na istotną poprawę skuteczności algorytmów, których optymalne konfiguracje wyłonione zostały na podstawie eksperymentu. Okazuje się, że najbardziej wydajny algorytm **ARS** $_2$  korzysta z mniejszej liczby konkurujących modeli probabilistycznych  $i_{\max} = 3$  oraz wykorzystuje mniejszą liczbę testów  $j_{\max} = 20$  w fazie wyboru rozkładu eksploracyjnego, niż jego powszechnie stosowana wersja [128], [122], [93]. Wyniki te, pozwalają również przypuszczać, iż w przypadku ograniczonej przestrzeni przeszukiwań, rozkłady o mniejszych indeksach stabilności mogą charakteryzować się nieco gorszą skutecznością. Źródła takiej sytuacji, należy upatrywać w zbyt dużej liczbie makromutacji właściwych dla rozkładów ciężkoogonowych, które zwykle powodują wykroczenie poza obszar rozwiązań dopuszczalnych.

**Obserwacja 27** Zestawienie najgorszych możliwych konfiguracji (tabela 6.3) z wynikami w tabeli 6.2 jest kolejnym niezbitym dowodem na większą odporność rozkładów ciężkoogonowych. Kwestii tej, w przypadku nieograniczonej przestrzeni przeszukiwań oraz strategii ewolucyjnej, poświęcony został rozdział 4.2.4. Zadanie maksymalizacji funkcji wariancji, pokazuje, że także dla ograniczonego

zbioru rozwiązań dopuszczalnych oraz algorytmu  $ARS_\alpha$ , można wyciągnąć bardzo podobne wnioski. Mianowicie, rozkłady o większych indeksach stabilności charakteryzują się większą efektywnością, ale tylko w przypadku precyzyjnego dopasowania algorytmów do rozpatrywanego zadania. W momencie zastosowania dalekiej od optymalnej konfiguracji algorytmu można zaobserwować drastyczne pogorszenie skuteczności metody optymalizacyjnej. Inaczej przedstawia się ta kwestia dla rozkładów o mniejszym indeksie stabilności. Porównując wyniki otrzymane dla najlepszej oraz najgorszej konfiguracji dla indeksu stabilności  $\alpha = 0.5$  spadek wydajności algorytmu jest najmniej dotkliwy ze wszystkich modeli stabilnych.

#### 6.1.4 Ocena dokładności linearyzacji modelu neuronowego w zadaniach wnioskowania statystycznego

Metoda linearyzacji, przedstawiona w poprzednim rozdziale, z jednej strony pozwala w prosty sposób wyznaczyć zarówno przedziały jak i obszary ufności, w sytuacji gdy parametry modelu nieliniowego wyznaczane są na podstawie danych mających losowy charakter. Mimo, iż metoda ta umożliwia w łatwy sposób przeprowadzenie analizy statystycznej nawet dla modeli o znacznej liczbie parametrów, to należy pamiętać o tym, że wnioski wysuwane na jej podstawie będą wiarygodne wówczas gdy, linearyzacja nie będzie wprowadzać dużej rozbieżności pomiędzy modelem nieliniowym a jego przybliżeniem. Oznacza to, że stosując wspomnianą metodę, należy pogodzić się z cichym założeniem mówiącym o tym, że zlinearyzowany model w pewnym otoczeniu estymaty  $\hat{\theta}$  posiada charakterystyki niemalże identycznie jak jego nieliniowy odpowiednik. W wielu przypadkach, nie zawsze jednak można bez przystać na takie założenie, o czym świadczą chociażby liczne przykłady przytaczane w literaturze [17, 98]. W ich świetle, szczególnego znaczenia nabiera potrzeba dysponowania dodatkowym narzędziem, służącym do oceny precyzji przybliżenia liniowego [98]. Okazuje się, że dokładność liniowej aproksymacji nieliniowych modeli regresji, można w sposób ilościowy scharakteryzować za pomocą tzw. średnich miar krzywizny zaproponowanych przez Wattesa i Bettsa [17]:

**średnio-kwadratowa krzywizna parametryczna**

$$c_P = \frac{\Gamma(n_p/2)}{2\pi^{n_p/2}} \sum_{i=1}^{n_p} \int_{S:d^T d=1} \frac{\|\mathbf{d}^T \mathbf{A}_\xi \mathbf{d}\|}{\|\mathbf{V}_\xi \mathbf{d}\|^2} \hat{\sigma} \sqrt{n_p} \, dS \quad (6.33)$$

**średnio-kwadratowa krzywizna wewnętrzna**

$$c_I = \frac{\Gamma(n_p/2)}{2\pi^{n_p/2}} \sum_{i=n_p+1}^{n_t} \int_{S:d^T d=1} \frac{\|\mathbf{d}^T \mathbf{A}_\xi \mathbf{d}\|}{\|\mathbf{V}_\xi \mathbf{d}\|^2} \hat{\sigma} \sqrt{n_p} \, dS \quad (6.34)$$

gdzie  $\mathbf{V}_\xi \in \mathbb{R}^{n_t \times n_p}$  w przypadku rozważanych modeli neuronowych równy jest:

$$\mathbf{V}_\xi = [\mathbf{v}(\hat{\theta}, \mathbf{u}_1), \mathbf{v}(\hat{\theta}, \mathbf{u}_2), \dots, \mathbf{v}(\hat{\theta}, \mathbf{u}_{n_\epsilon})]^T, \quad (6.35)$$

oraz  $\mathbf{A}_\xi \in \mathbb{R}^{n_t \times n_p \times n_p}$  jest trójwymiarową macierzą, której poszczególne warstwy posiadają postać  $\{\mathbf{A}_\xi\}_k$

$$\{\mathbf{A}_\xi^{(i)}\}_k = \left( \frac{\partial^2 y_m(\mathbf{u}_k, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)_{\boldsymbol{\theta}=\hat{\theta}}. \quad (6.36)$$

$dS$  określa element powierzchniowy  $n_p$ -wymiarowej kuli jednostkowej. Analiza licznych przypadków modeli nieliniowych, rozpatrywanych przez autorów, doprowadziła do ustalenia dopuszczalnych wartości średnich miar krzywizny:

$$\bar{c}_I \sqrt{\chi_{n_t, \alpha}^2} < 0.3 \quad (6.37)$$

$$\bar{c}_P \sqrt{\chi_{n_t, \alpha}^2} < 0.3, \quad (6.38)$$

gdzie  $\chi_{n_t, \alpha}^2$  jest górnym  $\alpha$  kwantylem rozkładu Chi-kwadrat o  $n_t$  stopniach swobody. Według sugestii autorów, niespełnienie nierówności (6.37), (6.38) prowadzi do dalece błędnego wyznaczenia przedziałów oraz obszarów ufności dla modelu nieliniowego [17]). Zastosowanie oraz wpływ obu wskaźników jakości liniowej aproksymacji zobrazujemy poniższym przykładem obliczeniowym.

### Przykład numeryczny

Rozważ system w postaci modelu pojedynczego neuronu, o dwóch wejściach :

$$y_{s,k} = \tanh(\theta_1^* x_k + \theta_2^*) + \epsilon_k, \quad (6.39)$$

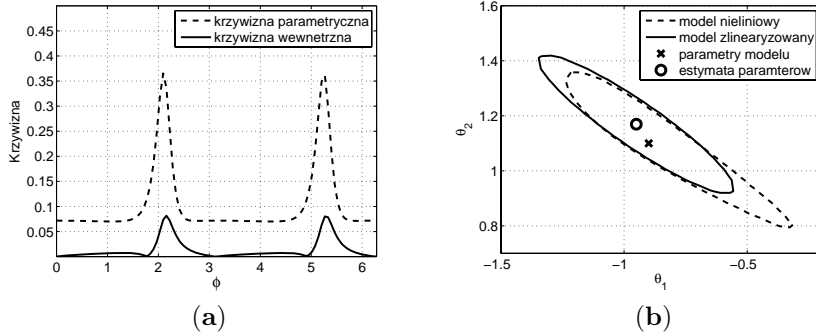
gdzie  $\epsilon_k$  oznacza sekwencje niezależnych, zmiennych losowych o rozkładzie  $\mathcal{N}(0, 0.05)$ , oraz  $\theta^* = [\theta_1^*, \theta_2^*]^T = [-0.9, 1.1]^T$ . Ponadto założymy, że parametry modelu  $y_m(x_k, \theta) = \tanh(\theta_1 x_k + \theta_2)$ , estymowane są na podstawie trójelementowego  $n_e = 3$  zbioru uczącego:  $\xi = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] = [(1.7, -0.41), (0, 0.8), (-0.7, 0.98)]$ , gdzie wyjścia systemu zostały zaburzone zgodnie z równaniem (6.39). W wyniku zastosowania algorytmu Levenberg'a-Marquard'a, otrzymano przybliżenie rzeczywistych wartości parametrów  $\hat{\theta} = [-0.95, 1.16]^T$ , dla których macierz otrzymano:

$$\mathbf{V} = \left[ \mathbf{v}(\hat{\theta}, \mathbf{u}_1) \quad \mathbf{v}(\hat{\theta}, \mathbf{u}_3) \right]^T = \begin{bmatrix} 0.0396 & 0.0233 \\ 0 & 0.4523 \\ -0.6460 & 0.9229 \end{bmatrix},$$

oraz :

$$\mathbf{A} = \left[ \mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \right] = \begin{bmatrix} 0.2425 & -0.3629 & 0.5088 \\ -0.0258 & -0.0098 & -0.6727 \\ 0.1460 & 0.0554 & -0.0401 \end{bmatrix}.$$

Dla rozważanego przykładu, miary krzywizny, obecne pod znakiem całki we wzorach (6.33),(6.34), zostały zaprezentowane na wykresie 6.3 (po ówczesnym przekształceniu do współrzędnych sferycznych tj.:  $\theta_1 = \cos(\phi), \theta_2 = \sin(\phi)$ ). Średnie miary krzywizny (6.33),(6.34) wyniosły odpowiednio 0.96 oraz 0.47. Jako, że obydwie wartości przekroczyły akceptowalne progi (6.37), (6.38), należy się spodziewać, że elipsoida ufności dla parametrów, wyznaczona przy pomocy linearyzacji modelu, będzie w znaczący sposób różnić się od prawdziwego obszaru ufności. Na rysunku 6.3 zestawiono oba obszary wyznaczone na poziomie ufności 95%.



Rysunek 6.3: (a) - kierunkowe miary krzywizny oraz (b) - obszary ufności dla modelu rozważanego w przykładzie obliczeniowym 6.1.4.

**Obserwacja 28** *Wzajemne ułożenie konturów obu obszarów ufności na rysunku 6.3 - (b), doskonale obrazuje wpływ miar (rysunek 6.3 - (a)) na precyzję metody linearyzacji. Warto zauważyć, że na kierunkach, na których obie miary krzywizny przybierają stosunkowo duże wartości, rozbieżność pomiędzy brzegami obu zbiorów ufności stają się wyraźnie widoczne.*

Efekt zobrazowany powyższym przykładem obliczeniowym ma swoje daleko idące negatywne konsekwencje, w sytuacji wyznaczania przedziału ufności dla odpowiedzi modelu nieliniowego. Mianowicie, w sytuacji dla której wskaźniki (6.33) - (6.34), osiągają znaczne wartości, należy się spodziewać dużych błędów przy wyznaczaniu przedziałów ufności metodą linearyzacji. W związku z tym, należy być bardzo ostrożnym formułując ostateczne wnioski związane z oczekiwaną odpowiedzią modelu [98].

### 6.1.5 Ewolucyjne przedziały ufności

Przykład obliczeniowy zaprezentowany w powyższym punkcie, pokazał, że nawet w przypadku pojedynczego neuronu, metoda linearyzacji, powszechnie stosowana w procesie wnioskowania statystycznego dla nieliniowych modeli regresji, prowadzić może do powstania dalece błędnych konkluzji [17, 98]. Sytuacja, w której dla otrzymanego modelu neuronowego, średnie miary krzywizny przekroczą akceptowalne progi (6.37)-(6.38), powinna stanowić przesłankę do zaniechania stosowania systemu FDI opartego o progi adaptacyjne wyznaczone na podstawie zlinearyzowanej wersji sieci neuronowej [129]. Alternatywnym rozwiązaniem problemu wyznaczania przedziałów ufności dla odpowiedzi modeli nieliniowych, jest zastosowanie ewolucyjnych technik optymalizacyjnych [87, 98].

Dla nieliniowych modeli regresji, w tym oczywiście sieci neuronowych, obszar ufności jest określony zależnością (6.15). Przedział ufności dla odpowiedzi modelu nieliniowego, może zostać wyznaczony przez podanie dwóch skrajnych wartości [97]

$$y_{m,k}^N = \arg \max_{\theta \in \Theta_{c,N}} y_m(\mathbf{x}_k, \theta) \quad (6.40)$$

$$y_{m,k}^M = \arg \min_{\theta \in \Theta_{c,N}} y_m(\mathbf{x}_k, \theta) \quad (6.41)$$

Rozwiązanie zadań optymalizacyjnych (6.40) - (6.41) sprowadza się zatem do odnalezienia najmniejszej oraz największej odpowiedzi modelu nieliniowego wewnątrz obszaru ufności. Ze względu na skomplikowany kształt obszaru ufności (6.15) niezwykle trudnym zadaniem, byłoby zaprojektowanie takich operatorów mutacji, które gwarantowałyby powstawanie osobników leżących wyłącznie w obszarze ufności. Z tego powodu, stosując tradycyjny operator mutacji, należy się liczyć z koniecznością wzbogacenia algorytmu optymalizacyjnego o mechanizm umożliwiający przetrzymywanie algorytmu wewnątrz zbioru parametrów dopuszczalnych. W literaturze można odnaleźć wiele rozwiązań umożliwiających uwzględnienie w zdaniu optymalizacji różnego rodzaju ograniczeń narzucanych na przestrzeń rozwiązań [74]. Wiele z nich, zakłada jednak regularny kształt zbioru rozwiązań, bądź dopuszcza jego nieznaczne przekroczenie. Mając na względzie, podstawowy cel podczas tworzenia modeli neuronowych, którym jest ich dalsze zastosowanie w wysokowydajnych systemach FDI, zdecydowano się zastosować algorytm rzutowania na brzeg obszaru dopuszczalnego [97] (tabela 6.4).

Tabela 6.4: Algorytm rzutowania rozwiązania niedopuszczalnego na brzeg obszaru ufności

---

**Dane wejściowe**

$Eps$  – bezwzględna dokładność w lokalizowaniu  
brzegu obszaru dopuszczalnego;

$\theta, \theta'$  – rozwiązanie oraz jego modyfikacja ;

$R = s^2 n_p F_{(n_p, n_t - n_p; \alpha)}$  – promień wyznaczający obszar ufności;

**Dane wyjściowe**

$\theta^*$  – rozwiązanie zlokalizowane na brzegu obszaru ufności ;

**Algorytm**

$\theta^* = \theta$

$h = \|\theta^* - \theta'\|_2$

Powtarzaj

Jeśli  $S(\frac{1}{2}(\theta^* + \theta')) - S(\theta) > R$  to  $\theta' = \frac{1}{2}(\theta^* + \theta')$

w przeciwnym razie  $\theta^* = \frac{1}{2}(\theta^* + \theta')$

$h = h/2$

Dopóki ( $h > Eps$ )

---

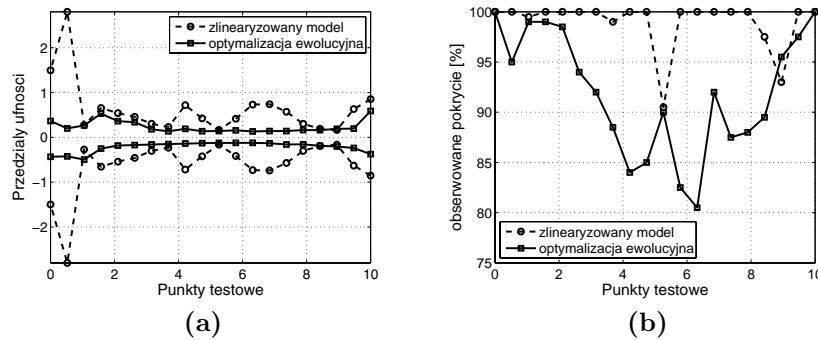
Algorytm składa się z kilku prostych kroków i polega na testowaniu kolejnych rozwiązań leżących na odcinku łączącym punkt leżący wewnątrz obszaru ufności  $\theta$  oraz rozwiązanie niedopuszczalne  $\theta'$ . Do testowania kolejnych punktów przestrzeni rozwiązań stosowana jest metoda połowienia, która w każdej iteracji zmniejsza odległość pomiędzy punktami rozgraniczającym przestrzeń

parametrów. Mimo, że metoda połowienia zapewnia stosunkowo szybką lokalizację brzegu obszaru dopuszczalnego, to jej stosowanie w przypadku operatora mutacji opartego na niskich indeksach stabilności, może okazać się skrajnie nieefektywnie. Wiadomo bowiem, że rozkłady stabilne  $\alpha < 1$  cechują się coraz częstszym występowaniem makromutacji, co może powodować otrzymanie rozwiązania ekstremalnie odległego od zbioru dopuszczalnego. W zależności od przyjętej precyzji zlokalizowania brzegu może być niezbędne wykonanie nawet kilkudziesięciu algorytmu zaprezentowanego w tabeli 6.4. Uwzględniając fakt, iż każda iteracja wymaga pojedynczego obliczenia funkcji  $S(\cdot)$ , która z kolei polega na otrzymaniu odpowiedzi modelu neuronowego dla wszystkich wzorców uczących, metoda rzutowania na brzeg obszaru może powodować znaczny wzrost nakładów obliczeniowych samego algorytmu optymalizacji [97].

### Przykład numeryczny

Aby potwierdzić zasadność proponowanej metody wyznaczania przedziałów ufności dla nieliniowych modeli regresji, rozważmy następujący eksperyment. Podobnie jak w poprzednim przykładzie numerycznym, założmy że deterministyczna część systemu (6.4) jest siecią neuronową o jednym wejściu, trzech nieliniowych jednostkach przetwarzających w warstwie ukrytej oraz jednym liniowym neuronie wyjściowym. W ten sposób całkowita liczba parametrów modelu, wliczając wszystkie wagi połączeń międzyneuronowych oraz biasy jest równa  $n_p = 14$ . Wektor parametrów  $\theta^* \in \mathbb{R}^{n_p}$  wygenerowany został zgodnie z rozkładem normalnym  $\mathcal{N}(\mathbf{0}, 5\mathbf{I}_{n_p})$ . Proces estymacji parametrów modelu  $\theta$  przebiegał na podstawie 18-elementowego zbioru uczącego równomiernie rozłożonego na odcinku  $[0, 10]$  oraz metody Levenberg'a-Marquard'a. Odpowiedzi systemu, były dodatkowo modyfikowane poprzez dodanie liczb pseudolosowych pochodzących z rozkładu normalnego  $N(0, 0.1^2)$ . Przedziały ufności dla odpowiedzi modelu neuronowego zostały wyznaczone dla dwudziestu równomiernie rozłożonych punktów testowych na przedziale  $[0, 10]$ .

Na rysunku 6.4-(a) zamieszczono porównanie przedziałów ufności (wyznaczonych na poziomie  $1 - \alpha = 0.9$ ) dla sygnału residuum. Przedziały wyznaczone zostały z wykorzystaniem linearyzacji modelu neuronowego (6.20) oraz poprzez rozwiązanie zadania optymalizacji (6.40), (6.41) przy zastosowaniu strategii ewolucyjnej  $(1 + 1)ES_{1.5}$  z algorytmem korekty rozwiązania 6.4. W celu weryfikacji wiarygodności otrzymanych przedziałów ufności, dla każdego z punktów testowych wygenerowano 200 odpowiedzi systemu jak i jego modelu. Procent obserwacji które zawierały się w wyznaczonych przedziałach ufności przedstawiony został na rysunku 6.4-(b).



Rysunek 6.4: (a) - przedziały ufności na poziomie  $1 - \alpha = 0.9$  dla sygnału residuum modelu neuronowego, oraz (b) - procent pokrycia sygnału residuum przez przedziały ufności

**Obserwacja 29** *Metoda linearyzacji modelu neuronowego wokół estymaty parametrów może prowadzić do znacznych rozbieżności pomiędzy rzeczywistymi przedziałami ufności, a tymi wyznaczonymi dla uproszczonego modelu liniowego. Na rysunku 6.4-(b) można zaobserwować, że wyznaczone przedziały ufności zawierają niemalże 100 % wszystkich obserwowanych sygnałów residuum. Biorąc pod uwagę, iż w rzeczywistości przedziały te powinny zawierać jedynie 90 % wszystkich obserwacji, jest oczywiste, że metoda linearyzacji dostarczyła, w tym przypadku, zbyt szerokie przedziały ufności. Sytuacja taka może mieć dalekosiężne konsekwencje przy projektowaniu systemów wykrywania uszkodzeń FDI. Zbyt optymistyczne wyznaczenie przedziałów ufności może powodować problemy z wykrywaniem uszkodzeń, które skutkują tylko nieznacznym odchyleniem od nominalnego stanu pracy systemu [60], [98].*

## 6.2 System wykrywania uszkodzeń systemu przemysłowego

Międzynarodowy projekt DAMADICS (ang. *Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems*), zorientowany był na rozwój oraz implementację nowoczesnych metod diagnostyki przemysłowej. W ramach wspomnianego projektu opracowano model zaworu sterującego przepływem soku owocowego przez system podzespołów (bardziej szczegółowe omówienie tego zagadnienia znajduje się w [1]). Obiekt udostępniał pomiary następujących zmiennych procesowych: sygnał sterujący  $CV$ , ciśnienie soku na wlocie zaworu  $P1$ , ciśnienie soku po opuszczeniu zaworu, temperatura soku mierzona przed zaworem, położenie serwowatora  $X$  oraz ilość soku przepływająca przez zawór. Biorąc pod uwagę, fakt, iż rzeczywiste uszkodzenie urządzenia przemysłowego nie wchodzi w rachubę, jako środowisko testowe, wykorzystano model matematyczny zaworu cukrowni Lublin (dostępny na stronie internetowej [1]). Model ten udostępnia szereg scenariuszy polegających na wystąpieniu uszkodzeń podzespołów wykonawczych o różnym stopniu nasilenia. Wybór środowiska testowego jest podyktowany przede wszystkim możliwością porównania skuteczności proponowanych systemów FDI, z tymi opracowanymi w ramach



projektu DAMADICS [73, 2, 127].

Głównym celem niniejszego rozdziału, jest eksperymentalna weryfikacja skuteczności systemów FDI opartych na neuronowych generatorach sygnału residuum. W szczególności skupiono się na zastosowaniu procedury estymacji parametrycznej modelu neuronowego z wykorzystaniem  $D$ -optymalnych zbiorów uczących. Metoda ta jakkolwiek skutkuje otrzymaniem modelu o mniejszej niepewności parametrycznej, to zastosowana technika linearyzacji zwykle wypacza wyniki analizy statystycznej. Dlatego też, w badaniach przedstawiono również wyniki wykrywania uszkodzeń dla rozważanego systemu przemysłowego, z wykorzystaniem metod optymalizacji ewolucyjnej do wyznaczania przedziałów ufności dla sygnału residuum. Bazując na doświadczeniach zdobytych w ramach realizacji projektu DAMADICS, okazało się, że zmienną procesową  $F$  można z akceptowalną dokładnością przybliżyć pewną funkcją zmiennych procesowych  $(CV, P1)$ . Dlatego też, w celu zbudowania modelu zaworu, zdecydowano się na zastosowanie statycznej sieci neuronowej o czterech tangensoidalnych jednostkach przetwarzających w warstwie ukrytej  $n_h = 4$  oraz jednym liniowym neuronie wyjściowym. Przyjęta struktura modelu neuronowego była przedmiotem eksperymentu, na przebiegał z uwzględnieniem następujących etapów:

**Krok 1:** Wstępna estymacja parametrów sieci neuronowej na podstawie niewielkiego zbioru uczącego,

**Krok 2:** Planowanie eksperymentu dla modelu otrzymanego w kroku 1), z wykorzystaniem algorytmu Wynn'a-Fedorov'a, wzbogaconego o procedurę optymalizacji globalnej ARS.

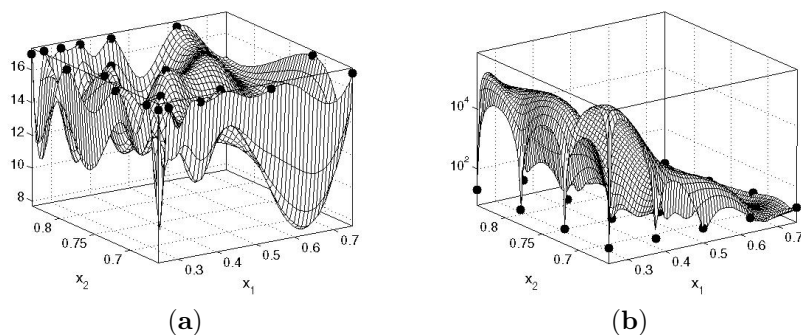
**Krok 3 (a):** Estymacja parametrów modelu neuronowego otrzymanego w kroku 1) na podstawie  $D$ -optymalnego zbioru uczącego z kroku 2).

**Krok 3 (b):** Estymacja parametrów modelu neuronowego otrzymanego w kroku 1) na podstawie pseudolosowych sekwencji Haltona.

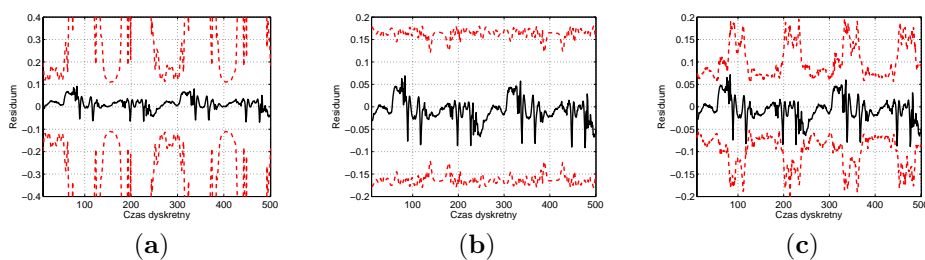
**Krok 4:** Wyznaczenie przedziałów ufności dla modelu neuronowego otrzymanego w krokach 3 (a) oraz 3 (b), oraz dodatkowo wyznaczenie odpowiednich przedziałów metodą optymalizacji ewolucyjnej dla modelu z kroku 3 (a).

**Krok 5:** Przeprowadzenie skuteczności systemu FDI opartego o neuronowe generatory residuum oraz trzy różne adaptacyjne progi decyzyjne otrzymane w kroku 4.

Tabela 6.5 zawiera porównanie skuteczności systemów FDI, zbudowanych o rozważany typ sieci neuronowej. Model będący przedmiotem eksperymentu, umożliwiał symulację 14 różnych scenariuszy wystąpienia uszkodzenia o różnej szybkości narastania (symbol S - mała, M - średnia oraz B - duża). Jak można zaobserwować (tabela 6.5), istnieją takie uszkodzenia, których wykrycie możliwe jest jedynie przez system bazujący na progach adaptacyjnych wyznaczonych za pomocą optymalizacji ewolucyjnej ( $f_1$ -S,  $f_{14}$ -B). Oznacza, to że progi adaptacyjne wyznaczone na podstawie planów eksperymentu oraz linearyzacji modelu neuronowego nie mogą być traktowane jako wiarygodne narzędzie diagnostyczne.



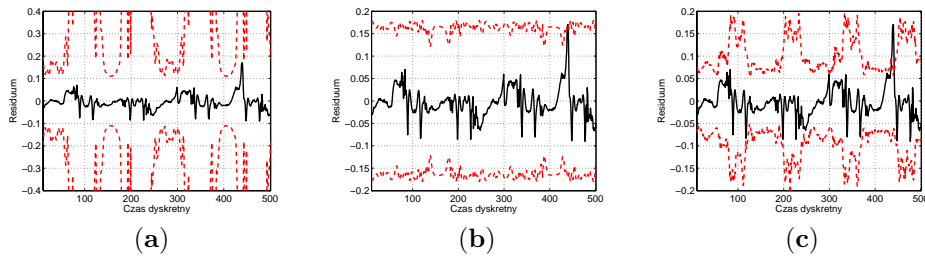
Rysunek 6.5: Funkcja wariancji rozpięta nad obszarem planowania eksperymentu: (a) - dla planu  $D$ -optymalnego, (b) - dla równomiernie rozłożonych wzorców uczących



Rysunek 6.6: Sygnał residuum oraz progi decyzyjne dla nominalnego stanu pracy zaworu. Progi decyzyjne wyznaczone dla modelu neuronowego : (a) - uczonego na podstawie sekwencji Haltona, (b) - uczonego z wykorzystaniem zbioru  $D$ -optymalnego i metody linearyzacji, (c) - uczonego z wykorzystaniem zbioru  $D$ -optymalnego oraz optymalizacji ewolucyjnej.

Tabela 6.5: Zestawienie wyników procesu wykrywania uszkodzeń zaworu cukrowni Lublin, dla dwóch systemów FDI:  $i = 1$  system oparty o progi adaptacyjne wprowadzane przez model neuronowy otrzymany poprzez proces uczenia na podstawie równomiernie rozmieszczonych wzorców uczących, oraz  $i = 2$  system oparty o progi adaptacyjne dostarczane przez model neuronowy otrzymany na podstawie  $D$ -optymalnego zbioru uczącego 3 - system oparty o progi wyznaczone za pomocą optymalizacji ewolucyjnej,  $D_i$  - oznacza wykrycie uszkodzenia przez system  $i$ ,  $N_i$  - oznacza brak decyzji o wystąpieniu uszkodzenia raportowany przez system  $i$ ,  $X$  - nie dotyczy rozważanego systemu.

Fault	S	M	B
$f_1$	$(N_1, N_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$
$f_2$	X	X	$(D_1, D_2, D_3)$
$f_7$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$
$f_8$	$(N_1, N_2, N_3)$	$(N_1, N_2, N_3)$	$(N_1, N_2, N_3)$
$f_{10}$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$
$f_{11}$	X	X	$(D_1, D_2, D_3)$
$f_{12}$	$(N_1, N_2, N_3)$	$(N_1, N_2, N_3)$	$(N_1, D_2, D_3)$
$f_{13}$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$
$f_{14}$	$(N_1, N_2, N_3)$	$(N_1, N_2, N_3)$	$(N_1, N_2, D_3)$
$f_{15}$	X	X	$(D_1, D_2, D_3)$
$f_{16}$	$(N_1, N_2, N_3)$	$(N_1, N_2, N_3)$	$(D_1, D_2, D_3)$
$f_{17}$	X	X	$(D_1, D_2, D_3)$
$f_{18}$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$
$f_{19}$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$	$(D_1, D_2, D_3)$



Rysunek 6.7: Sygnal residuum oraz progi decyzyjne dla scenariusza polegające na wystąpieniu uszkodzenia  $f_1$  o małym stopniu natężenia. Progi decyzyjne wyznaczone dla modelu neuronowego : (a) - uczonego na podstawie sekwencji Haltona, (b) - uczonego z wykorzystaniem zbioru  $D$ -optymalnego i metody linearyzacji, (c) - uczonego z wykorzystaniem zbioru  $D$ -optymalnego oraz optymalizacji ewolucyjnej.

**Obserwacja 30** *Różnice pomiędzy progami decyzyjnymi wyznaczonymi na podstawie równomiernie rozłożonej sekwencji uczącej oraz  $D$ -optymalnego planu eksperymentu, stają się oczywiste jeśli porównać wykresy funkcji wariancji (6.5), otrzymanych dla obu metod. Znaczące różnice w rozmiarze przedziałów*

można również zaobserwować w przypadku metody linearyzacji sieci neuronowej oraz zastosowania algorytmów ewolucyjnych do wyznaczenia granicznych wartości dla oczekiwanej odpowiedzi modelu. Pokładając wiarę, w to że ewolucyjny algorytm optymalizacji, pozwolił na skuteczne rozwiązanie (6.40 - 6.41), obserwowane różnice mogą zostać wytłumaczone brakiem liniowej charakterystyki modelu, w otoczeniu aktualnej estymaty  $\hat{\theta}$ .

### 6.3 Podsumowanie

W rozdziale tym przedstawiony został problem projektowania odpornych systemów wykrywania uszkodzeń opartych o neuronowe generatory residuum. W pierwszej kolejności przedstawiono strukturę perceptronu wielowarstwowego - najczęściej wykorzystywanego modelu neuronowego, w zadaniach modelowania nieliniowych systemów statycznych. Ponadto zdefiniowano pojęcie niepewności parametrycznej, która odgrywa kluczową rolę w procesie wnioskowania statycznego. W celu minimalizacji parametrycznej niepewności sieci neuronowych, w niniejszym rozdziale, zaprezentowane zostało podejście oparte na zastosowaniu zbiorów uczących pozyskiwanych metodą planowania eksperymentu. Najpoważniejszym problemem związanym konstrukcją ciągłych planów eksperymentu, polega na konieczności rozwiązywania zadania optymalizacji globalnej, w każdej iteracji algorytmu Wynn'a-Fedorova. W związku z tym, problem maksymalizacji funkcji wariancji modeli neuronowych, posłużyło jako specyficzne środowisko testowe dla proponowanych algorytmów z rozkładami stabilnymi. Na podstawie przeprowadzonych badań, dokonano optymalnego doboru parametrów algorytmu  $ARS_{\alpha}$ . W wyniku zastosowania proponowanych rozwiązań, otrzymuje się plany eksperymentu o mniejszej liczbie punktów skupienia, co bezpośrednio przekłada się na mniejszą liczbę niezbędnych do wykonania pomiarów. Pozwala to również na obniżenie kosztów związanych z przeprowadzeniem całego procesu modelowania systemu nieliniowego. Co więcej algorytm Wynn'a-Fedorov'a wzbogacony o optymalizację ewolucyjną, zapewnia zmniejszenie niepewności parametrycznej, a co za tym idzie, możliwe jest projektowanie odpornych i bardziej niezawodnych systemów FDI.

W rozdziale dużo miejsca poświęcone zostało problemowi wnioskowania statycznego dokonywanego na podstawie linearyzacji modelu neuronowego. Przedstawiono definicje miar krzywizny, mogących posłużyć jako narzędzie diagnostyczne, umożliwiające wskazanie sytuacji, w której należy zachować szczególną ostrożność podejmując decyzję w oparciu o metodę linearyzacji. Zaproponowano również alternatywną metodę wyznaczania przedziałów ufności dla oczekiwanej odpowiedzi modeli neuronowych, która pozawala ominąć etap upraszczania sieci neuronowej do jej liniowego odpowiednika. Wyniki badań dotyczące systemu wykrywania uszkodzeń dla zaworu Cukrowni Lublin, potwierdziły zasadność stosowania ewolucyjnych progów adaptacyjnych.

Jakkolwiek przedstawione wyniki, wyglądają obiecująco, to należy zaznaczyć, że ich weryfikacja eksperymentalna nastąpiła na podstawie stosunkowo prostego systemu przemysłowego. Z doświadczenia zdobytego podczas pracy z rozważanymi rozwiązaniami, wynika, że zarówno planowanie eksperymentu dla sztucznych sieci neuronowych jak i wyznaczanie przedziałów ufności za pomocą algorytmów ewolucyjnych może być możliwe i skuteczne jedynie dla modeli o niewielkiej liczbie parametrów. Jakkolwiek teoria leżąca u podstaw zastosowań

wspomnianych rozwiązań, nie ogranicza liczby parametrów, to w praktyce należy liczyć się z problemami natury numerycznej - w przypadku planowania eksperymentu, jak i złożoności obliczeniowej - w przypadku zastosowania algorytmów ewolucyjnych. Najwięcej kłopotów sprawia procedura wyznaczania macierzy odwrotnej Fisher'a, która dla sieci neuronowych, nawet w przypadku planów  $D$ -optymalnych jest zwykle problemem źle uwarunkowanym.

## Rozdział 7

# Algorytmy ewolucyjne w zadaniach projektowania nieliniowych obserwatorów stanu

### 7.1 Wstęp i podstawowe pojęcia

Rozważmy nieliniowy, dyskretny system dynamiczny opisany równaniami:

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k) + \mathbf{h}(\mathbf{u}_k) + \mathbf{E}_k \mathbf{d}_k, \quad (7.1)$$

$$\mathbf{y}_{k+1} = \mathbf{C}_{k+1} \mathbf{x}_{k+1}. \quad (7.2)$$

gdzie  $\mathbf{x}_k \in \mathbb{R}^n$  oznacza wektor stanu,  $\mathbf{y} \in \mathbb{R}^m$  jest wyjściem systemu,  $\mathbf{u}_k \in \mathbb{R}^r$  jego wejściem,  $\mathbf{d}_k \in \mathbb{R}^q$  jest tzw. nieznanym wejściem, natomiast  $\mathbf{g}(\cdot), \mathbf{h}(\cdot)$  są parą, różniczkowalnych funkcji nieliniowych.

Podstawowym problemem, związanym z systemem (7.1) - (7.2), jest zaprojektowanie takiego obserwatora stanu, który bez względu na charakter nieznanego wejścia i jego wpływ na równanie stanu, byłby w stanie precyzyjnie estymować stan systemu [105, 129]. Okazuje się, że przy spełnieniu warunku [31, p. 72, Lemma 3.1]:

$$\text{rank}(\mathbf{C}_{k+1} \mathbf{E}_k) = \text{rank}(\mathbf{E}_k) = q, \quad (7.3)$$

można całkowicie wyeliminować wpływ sygnału  $\mathbf{d}_k$  na proces estymacji stanu systemu [129, 132]. Mianowicie, równanie (7.3) umożliwia obliczenie macierzy pseudoodwrotnej w sensie Moore-Penrose [132]

$$\mathbf{H}_{k+1} = (\mathbf{C}_{k+1} \mathbf{E}_k)^+ = [(\mathbf{C}_{k+1} \mathbf{E}_k)^T (\mathbf{C}_{k+1} \mathbf{E}_k)]^{-1} (\mathbf{C}_{k+1} \mathbf{E}_k)^T. \quad (7.4)$$

Mnożąc równanie stanu przez macierz (7.4) a następnie podstawiając równanie stanu (7.1), nieznanne wejście można przedstawić w postaci:

$$\mathbf{d}_k = \mathbf{H}_{k+1} \left[ \mathbf{y}_{k+1} - \mathbf{C}_{k+1} [g(\mathbf{x}_k) + h(\mathbf{u}_k)] \right] \quad (7.5)$$

W konsekwencji powyższych operacji, można zastosować algorytm estymacji stanu [126, 129, 132]

$$\hat{\mathbf{x}}_{k+1/k} = \bar{\mathbf{g}}(\hat{\mathbf{x}}_k) + \bar{\mathbf{h}}(\mathbf{u}_k) + \bar{\mathbf{E}}_k \mathbf{y}_{k+1}, \quad (7.6)$$

$$\mathbf{P}_{k+1/k} = \bar{\mathbf{A}}_k \mathbf{P}_k \bar{\mathbf{A}}_k^T + \mathbf{Q}_k, \quad (7.7)$$

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1/k} \mathbf{C}_{k+1}^T \left( \mathbf{C}_{k+1} \mathbf{P}_{k+1/k} \mathbf{C}_{k+1}^T + \mathbf{R}_{k+1} \right)^{-1}, \quad (7.8)$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_{k+1/k} + \mathbf{K}_{k+1} (\mathbf{y}_{k+1} - \mathbf{C}_{k+1} \hat{\mathbf{x}}_{k+1/k}), \quad (7.9)$$

$$\mathbf{P}_{k+1} = [\mathbf{I} - \mathbf{K}_{k+1} \mathbf{C}_{k+1}] \mathbf{P}_{k+1/k}, \quad (7.10)$$

gdzie

$$\bar{\mathbf{A}}_k = \left. \frac{\partial \bar{\mathbf{g}}(\mathbf{x}_k)}{\partial \mathbf{x}_k} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_k} = \bar{\mathbf{G}}_k \left. \frac{\partial \mathbf{g}(\mathbf{x}_k)}{\partial \mathbf{x}_k} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_k} = \bar{\mathbf{G}}_k \mathbf{A}_k. \quad (7.11)$$

oraz

$$\begin{aligned} \bar{\mathbf{g}}(\cdot) &= \bar{\mathbf{G}}_k \mathbf{g}(\cdot), \quad \bar{\mathbf{h}}(\cdot) = \bar{\mathbf{G}}_k \hat{\mathbf{h}}(\cdot) \\ \bar{\mathbf{G}}_k &= \mathbf{I} - \mathbf{E}_k \mathbf{H}_{k+1} \mathbf{C}_{k+1}, \quad \bar{\mathbf{E}}_k = \mathbf{E}_k \mathbf{H}_{k+1} \end{aligned} \quad (7.12)$$

Algorytm (7.6)–(7.10) jest w rzeczywistości Rozszerzonym Filtrem Kalmana (ang. *Extended Kalman Filter*) dla nieliniowych systemów deterministycznych z nieznanym wejściem. Fakt przystosowania algorytm Kalmana, do pracy z systemami deterministycznymi, ma swoje daleko idące konsekwencje. Przede wszystkim należy zaznaczyć, że w algorytmie (7.6)–(7.10) występują macierze  $\mathbf{Q}_k$  oraz  $\mathbf{R}_k$ , które w oryginalnym algorytmie Kalmana oznaczały macierze kowariancji sygnałów losowych wpływających na stan oraz wyjście systemu. Jako, że system (7.1)–(7.2) nie uwzględnia żadnych składników losowych, istnieje pewien stopień dowolność w wyborze wartości macierzy  $\mathbf{Q}_k$  i  $\mathbf{R}_k$  [129, 132]. Oczywiście wybór taki, nie pozostaje bez wpływu na zbieżność algorytmu (7.6)–(7.10). Na potrzeby niniejszego rozdziału, przytoczymy tylko najważniejsze fakty dotyczące tego zagadnienia (szczegółowe omówienie zawarte jest w monografii [129]). Zauważmy, że błąd estymacji stanu może zostać przedstawiony jako

$$\mathbf{e}_{k+1} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1} = [\mathbf{I} - \mathbf{K}_{k+1} \mathbf{C}_{k+1}] \mathbf{e}_{k+1/k}, \quad (7.13)$$

gdzie

$$\mathbf{e}_{k+1/k} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1/k} = \bar{\mathbf{g}}(\mathbf{x}_k) - \bar{\mathbf{g}}(\hat{\mathbf{x}}_k) = \boldsymbol{\alpha}_k \bar{\mathbf{A}}_k \mathbf{e}_k, \quad (7.14)$$

gdzie  $\boldsymbol{\alpha}_k = \text{diag}(\alpha_{1,k}, \dots, \alpha_{n,k})$  jest nieznaną macierzą diagonalną. W ten sposób, wykorzystując (7.14), równanie equation (7.13) przybiera postać:

$$\mathbf{e}_{k+1} = [\mathbf{I} - \mathbf{K}_{k+1} \mathbf{C}_{k+1}] \boldsymbol{\alpha}_k \bar{\mathbf{A}}_k \mathbf{e}_k. \quad (7.15)$$

Z równania (7.14) wynika, że  $\boldsymbol{\alpha}_k$  reprezentuje błąd linearyzacji. Stosując oznaczenia

$$\bar{\alpha}_k = \max_{j=1, \dots, n} |\alpha_{j,k}|, \quad \underline{\alpha}_k = \min_{j=1, \dots, n} |\alpha_{j,k}|. \quad (7.16)$$

można udowodnić następujące twierdzenie [129]:

**Twierdzenie 16** *Rozszerzony obserwator o nieznanym wejściu (7.6)–(7.10) jest loklanie asymptotycznie zbieżny, jeśli spełniona jest nierówność*

$$\bar{\alpha}_k \leq \left[ \frac{\alpha_k^2 \frac{\underline{\sigma}(\bar{\mathbf{A}}_k)^2 \underline{\sigma}(\mathbf{C}_{k+1})^2 \underline{\sigma}(\bar{\mathbf{A}}_k \mathbf{P}_k \bar{\mathbf{A}}_k^T + \mathbf{Q}_k)}{\bar{\sigma}(\mathbf{C}_{k+1} \mathbf{P}_{k+1/k} \mathbf{C}_{k+1}^T + \mathbf{R}_{k+1})} + \frac{(1-\zeta) \underline{\sigma}(\bar{\mathbf{A}}_k \mathbf{P}_k \bar{\mathbf{A}}_k^T + \mathbf{Q}_k)}{\bar{\sigma}(\bar{\mathbf{A}}_k)^2 \bar{\sigma}(\mathbf{P}_k)} \right]^{\frac{1}{2}}, \quad (7.17)$$

gdzie  $0 < \zeta < 1$

na podstawie twierdzenia 7.17, staje się oczywiste, że zbieżność proponowanego obserwatora, silnie zależy od granicznych wartości elementów macierzy  $\alpha_k$ . Co więcej, wartości te, mogą być kontrolowane poprzez odpowiednie ustalenia macierzy instrumentalnych  $\mathbf{Q}_k$  i  $\mathbf{R}_k$ . Z jednej strony, aby poszerzyć obszar zbieżności obserwatora, należy tak dobrać macierze instrumentalne, aby maksymalizować:

$$\underline{\sigma}(\bar{\mathbf{A}}_k \mathbf{P}_k \bar{\mathbf{A}}_k^T + \mathbf{Q}_k), \quad (7.18)$$

oraz minimalizować

$$\bar{\sigma}(\mathbf{C}_{k+1} \mathbf{P}_{k+1/k} \mathbf{C}_{k+1}^T + \mathbf{R}_{k+1}). \quad (7.19)$$

Z drugiej jednak strony, odwołując się do tradycyjnego filtra Klamana, wiadomo, że współczynnik zbieżności jest tym większy, im przyjęte wartości macierzy kowariancji  $\mathbf{Q}_k$  i  $\mathbf{R}_k$  odpowiadają rzeczywistemu charakterowi sygnałów losowych. Podążając tym śladem, należałoby przyjąć zerowe macierze instrumentalne dla obserwatora (7.6)–(7.10). Podejście takie, powoduje jednak bardzo często rozbieżność algorytmu oraz szereg problemów związanych ze stabilnością obliczeń numerycznych [129]. Z tego powodu, niezbędne jest ustawienie macierzy instrumentalnych, w taki sposób, aby ustalić kompromis pomiędzy zakresem a szybkością zbieżności obserwatora. W literaturze można znaleźć rozwiązania, które w różny sposób starają się sprostać temu zadaniu:

- **Ustawienie 1** - podejście klasyczne [26, 128]

$$\begin{aligned} \mathbf{Q}_k &= 0.1 \mathbf{I}_n, \\ \mathbf{R}_{k+1} &= 0.1 \mathbf{I}_m, \end{aligned}$$

- **Ustawienie 2** - podejście zaproponowane w [26]

$$\begin{aligned} \mathbf{Q}_k &= 10^3 \varepsilon_k^T \varepsilon_k \mathbf{I}_n + 0.01 \mathbf{I}_n, \\ \mathbf{R}_{k+1} &= 10 \varepsilon_k^T \varepsilon_k \mathbf{I}_m + 0.01 \mathbf{I}_m, \end{aligned}$$

gdzie  $\varepsilon_k = \mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k$  oznacza obserwowany błąd wyjścia,

- **Ustawienie 3** - podejście zaproponowane w [128]

$$\begin{aligned} \mathbf{Q}_k &= q^2(\varepsilon_k) \mathbf{I}_n + 0.1 \mathbf{I}_n, \\ \mathbf{R}_{k+1} &= r^2(\varepsilon_k) \mathbf{I}_m + 0.01 \mathbf{I}_m. \end{aligned}$$

gdzie  $q^2(\cdot)$  i  $r^2(\cdot)$  są parą funkcji, przyjmujących tylko wartości dodatnie, dobierane na drodze optymalizacji za pomocą programowania genetycznego.



Z badań eksperymentalnych zamieszczonych w [128], wynika, że podejście klasyczne do ustalenia macierzy instrumentalnych, prowadzi do najgorszych wyników. Uzależnienie wartości obu macierzy od wielkości błędu wyjścia, co staje się udziałem podejścia nr 2, znacznie podnosi efektywność algorytmu estymacji stanu. Dalsze polepszenie działania obserwatora może zostać osiągnięte poprzez zastosowanie rozwiązania nr 3 [128]. Zupełnie naturalne bowiem wydaje się, że rozwiązanie w pewnym stopniu dedykowane do konkretnego systemu (dopasowanie następuje na drodze programowania genetycznego), musi być lepsze niż ogólna postać proponowana w [26].

## 7.2 Projektowanie odpornych obserwatorów nieliniowych

Wykorzystując pewną dowolność przy ustawieniach macierzy instrumentalnych, możliwe jest polepszenie zbieżności obserwatorów nieliniowych. Ustawienia prezentowane w poprzednim rozdziale mogą posłużyć jako punkt wyjścia, do głębszych rozważań dotyczących zbieżności proponowanego algorytmu estymacji stanu. Z jednej strony przyjęcie kwadratowej funkcji błędu wyjścia  $\varepsilon_k^T \varepsilon_k$  nie może być jedynym słusznym ustawieniem dla wszystkich systemów dynamicznych postaci (7.1)-(7.2). Z drugiej strony ustawienie nr 3, może okazać się skrajnie nieefektywnym, jeśli stan początkowy systemu będzie różny od tego, dla którego przeprowadzono optymalizację funkcji  $q^2(\cdot)$  i  $r^2(\cdot)$ . Dlatego też, w rozdziale tym, przedstawione zostanie ogólne rozwiązanie, które ma na celu zarówno poprawę współczynnika zbieżności obserwatora nieliniowego jaki i zwiększenie odporności algorytmu (7.6)-(7.10) na niepewność związaną z początkowym stanem systemu.

Wzorując się na [128] przyjmijmy następujące ustawienia macierzy instrumentalnych:

$$\mathbf{Q}_k = q(\varepsilon_k; \boldsymbol{\theta}) \mathbf{I}_n + \delta_1 \mathbf{I}_n, \quad (7.20)$$

$$\mathbf{R}_k = \delta_2 \mathbf{I}, \quad (7.21)$$

gdzie  $q(\varepsilon_k; \boldsymbol{\theta})$  jest pewną funkcją, przyjmującą tylko wartości dodatnie, określoną za pomocą  $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$ . W proponowanym rozwiązaniu, zrezygnowano ze złożonej obliczeniowo techniki programowania genetycznego, na korzyść przedstawienia  $q(\varepsilon_k; \boldsymbol{\theta})$  w postaci wielowymiarowych funkcji sklepanych. W literaturze istnieją, dwa najbardziej rozpowszechnione sposoby reprezentowania funkcji sklepanych: typu-**pp** (ang. *piecewise polynomial function*) oraz typu-**B** (ang. *basis-splines*) [34]. Dysponując ciągiem węzłów  $t_x = \{t_1, t_2, \dots, t_n\}$ , funkcja sklejana  $f(x)$  jest wyrażona w postaci sumy:

$$f(x) = \sum_{i=1}^n c_i B_{i,k}(x), \quad (7.22)$$

gdzie  $B_{i,k}(x)$  jest dodatnim wielomianem stopnia  $k$ , określonym na przedziale  $[t_i, t_{i+k}]$ . Wielowymiarowe funkcje sklepane otrzymuje się poprzez zastosowanie

iloczynu tensorowego dla jednowymiarowych funkcji (7.22) [34]:

$$f(\mathbf{x}) = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_n=1}^{N_n} c_{i_1, i_2, \dots, i_n} B_{i_1, k_1, t_1}(x_1) B_{i_2, k_2, t_2}(x_2) \cdots B_{i_n, k_n, t_n}(x_n), \quad (7.23)$$

gdzie każda zmienna  $x_i$  posiada własną sekwencję  $t_i$  złożoną z  $N_i$  węzłów oraz stopień wielomianu  $k_i$ . Problem doboru współczynników  $c_{i_1, i_2, \dots, i_n}$  w zadaniach interpolacji, może być skutecznie rozwiązany poprzez sekwencyjne rozwiązywanie problemu interpolacji jednowymiarowej [34]. Wykorzystując funkcje skleja-  
ne postaci (7.23) do ustalania macierzy instrumentalnej  $\mathbf{Q}_k$  (7.20) przyjmijmy, że wektor  $\boldsymbol{\theta}$  funkcji  $q(\cdot) = f(\cdot)$ , zawiera wszystkie współczynniki  $c_{i_1, i_2, \dots, i_n}$ ,  $i_1 = 1, \dots, N_1, \dots, i_n = 1, \dots, N_n$ .

W celu uzyskania obserwatora, którego skuteczność będzie w jak najmniejszym stopniu uzależniona od stanu początkowego systemu, zaproponowano schemat projektowania obserwatora bazujący na idei odporności stochastycznej [71].

Wspomniana technika, w przypadku rozważanego obserwatora nieliniowego, sprowadza się do ustalenia pewnej miary prawdopodobieństwa  $P$  na przestrzeni stanów systemu dynamicznego  $\mathbb{X}$ , tj:

$$P(\mathbb{X}) = 1 \quad (7.24)$$

Wybór miary może zostać dokonany bądź to na podstawie wiedzy eksperckiej, bądź na podstawie historycznej bazy pomiarów wykonanych dla rozważanego systemu. Następnie, niezbędne jest wprowadzenie funkcji oceny obserwatora. Spełnienie warunku zawartego w twierdzeniu (7.17) umożliwia zdefiniowanie funkcji jakości  $a : \mathbb{R}^{n_p + n_u} \rightarrow [0, 1]$  :

$$a(\boldsymbol{\theta}, \mathbf{x}_0) = \frac{\|\mathbf{e}_T\|_2}{\|\mathbf{e}_0\|_2}, \quad (7.25)$$

gdzie  $T$  określa dowolny horyzont czasowy,  $\mathbf{x}_0$  określa początkowy stan systemu,  $\boldsymbol{\theta}$  jest wektorem złożonym z parametrów wielowymiarowej funkcji skleja-  
nej  $q(\boldsymbol{\varepsilon}_k; \boldsymbol{\theta})$ . Funkcja (7.25) posiada bardzo prostą interpretację. Mianowicie dla obserwatora zdefiniowanego za pomocą wektora parametrów  $\boldsymbol{\theta}$ , wyraża ona stosunek wielkości błędu wyjścia po wykonaniu  $T$  iteracji obserwatora (7.6)-(7.10), do wielkości błędu początkowego. Należy podkreślić, że ewentualna optymalizacja oparta o funkcję celu (7.25) spowodowałaby dopasowanie obserwatora tylko do pojedynczego stanu systemu  $\mathbf{x}_0$ . Uwzględniając fakt, iż  $\mathbf{x}_0$  jest wektorem losowym indukowanym przez miarę prawdopodobieństwa (7.24), określmy zatem wartość oczekiwaną funkcji (7.25):

$$E_P(\boldsymbol{\theta}) = \int_{\mathbb{X}} a(\boldsymbol{\theta}, \mathbf{x}_0) P(d\mathbf{x}_0), \quad (7.26)$$

W ten sposób, problem projektowania odpornych obserwatorów nieliniowych, może zostać sprowadzony do zadania minimalizacji:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \in \Omega} E_P(\boldsymbol{\theta}), \quad (7.27)$$

gdzie  $\boldsymbol{\theta}^*$  jest wektorem parametrów definiującym optymalny obserwator dla systemu (7.1)-(7.2).

### 7.3 Metoda Monte-Carlo

Przyjęta w poprzednim punkcie, postać funkcji celu (7.27) skutecznie uniemożliwia otrzymanie rozwiązania analitycznego. Jeśli wziąć pod uwagę fakt, iż obliczenie pojedynczej wartości funkcji podcałkowej (7.25) wymaga wykonania  $T$  iteracji algorytmu (7.6)-(7.10), a obliczenie funkcji celu wymaga wykonania całkowania po przestrzeni o znacznej liczbie wymiarów, problem optymalizacji (7.27) wydaje się niezwykle trudnym do rozwiązania. Aby jednak umożliwić rozwiązanie problemu w akceptowalnym czasie, w rozdziale tym proponuje się zastosowanie całkowania metodą Monte-Carlo. Metoda ta, zakłada zastąpienie funkcji celu (7.27) jej przybliżeniem:

$$\hat{E}_P(\boldsymbol{\theta}, \mathbf{X}^{(k)}) = \frac{1}{k} \sum_{i=1}^k a(\boldsymbol{\theta}, \mathbf{x}_0^{(i)}), \quad (7.28)$$

gdzie  $\mathbf{X}^{(k)} = \{\mathbf{x}_0^{(i)}\}_{i=1}^k$  oznacza sekwencję niezależnych zmiennych losowych pochodzących z rozkładu prawdopodobieństwa określonego przez miarę  $P$ . Należy przy tym podkreślić, że estymator (7.28) posiada charakter stochastyczny. Pomijając fakt, iż prawdziwy rozkład rozważanego estymatora zwykle pozostaje nieznanym, można kontrolować jego dokładność wykorzystując tzw. nierówność Chernoff'a:

$$P^k \left\{ \mathbf{x}_0 \in \mathbb{X} : |\hat{E}_P(\boldsymbol{\theta}, \mathbf{X}^{(k)}) - E_P(\boldsymbol{\theta})| > \varepsilon \right\} \leq 2 \exp(-2 k \varepsilon^2), \quad (7.29)$$

gdzie  $\varepsilon > 0$ . W ten sposób, jeśli prawdziwa wartość oczekiwana  $E_P(\boldsymbol{\theta})$  ma należeć do przedziału ufności  $[\hat{E}_P(\boldsymbol{\theta}, \mathbf{X}^{(k)}) - \varepsilon; \hat{E}_P(\boldsymbol{\theta}, \mathbf{X}^{(k)}) + \varepsilon]$  z prawdopodobieństwem  $1 - \delta$ , to estymator (7.28) musi zostać wyznaczony na podstawie próbki losowej o wielkości  $k$ , gdzie

$$k \geq \frac{1}{2\varepsilon^2} \ln\left(\frac{2}{\delta}\right) \quad (7.30)$$

Największą zaletą wspomnianej metody Monte-Carlo, jest to, iż liczebność sekwencji losowych, nie zależy od wymiaru przestrzeni  $\mathbb{X}$ , co dodatkowo przemawia na korzyść proponowanego podejścia.

### 7.4 Aspekty optymalizacyjne

Największy problem związany z projektowaniem odpornych obserwatorów przejawia się w tym, iż funkcja celu bazuje na całkowaniu metodą Monte-Carlo, a przez to jej wartości obarczone są pewnym stopniem niepewności. Zastosowanie sekwencji losowych o długościach wyznaczonych przez nierówność (7.30), pozwala skutecznie zredukować rozmiar przedziału ufności dla wartości oczekiwanej (7.28). Analizując dokładnie wzór (7.30), należy jednak zwrócić uwagę na fakt, iż zmniejszając przedział ufności, liczba niezbędnych realizacji zmiennej losowej gwałtownie rośnie. Dlatego też, chcąc utrzymać złożoność obliczeniową procesu optymalizacji parametrycznej (7.27), na możliwym do zaakceptowania poziomie, nie należy dążyć do bezwarunkowego zredukowania czynnika losowego. Ponadto, należy uwzględnić fakt, iż w trakcie procesu minimalizacji, wartość oczekiwana (7.26) będzie ulegać zmianą. Dlatego też, koniecznym

wyduje się adekwatnie dostosowywanie poziom szumu do aktualnych wartości funkcji celu, tak aby umożliwić dalszy, wiarygodny, postęp procesu optymalizacji. Chcąc uwzględnić wspomniane uwagi, rozważmy granice przedziału ufności postaci  $\varepsilon = r_w \hat{E}_P(\boldsymbol{\theta}, X^{(k)})$  w (7.29), dla pewnego  $r_w \in (0, 1)$ . W ten sposób można otrzymać alternatywną definicję (7.29), dla której stosunek poziomu szumu do rzeczywistej wartości funkcji celu, jest stały. Konsekwencją powyższego zabiegu jest modyfikacja (7.30) do postaci:

$$k = \left\lceil \frac{1}{2(r_w \hat{E}_P(\boldsymbol{\theta}, X^{(k)}))^2} \ln\left(\frac{2}{\delta}\right) \right\rceil, \quad (7.31)$$

gdzie  $\lceil \cdot \rceil$  oznacza operator zaokrąglenia do najbliższej liczby całkowitej, która jest nie mniejsza niż jego argument.

Przedstawiona powyżej metodologia pozwoliła sprowadzić problem projektowania nieliniowych obserwatorów stanu, do problemu optymalizacji parametrycznej. Jest rzeczą zmienną, iż mimo kontrolowania czynnika losowego (7.31), pojawiającego się w funkcji celu (7.28), nie jest możliwe zastosowanie klasycznych algorytmów optymalizacji gradientowej. Ponadto, stając przed wyborem techniki optymalizacji, mogącej posłużyć do minimalizacji (7.28), należy wziąć pod uwagę fakt, iż problem (7.27) może okazać się w rzeczywistości wielomodalny. Dlatego też, metody optymalizacji lokalnej nie powinny, w tym przypadku, znaleźć zastosowania. Co więcej, sposób parametryzacji rozważanych obserwatorów, oraz zastosowanie funkcji sklepanych o dużej liczbie współczynników, znacznie podnosi stopień skomplikowania problemu. Pamiętając o tym, że funkcja  $q(\varepsilon_k; \boldsymbol{\theta})$ , będąca wynikiem optymalizacji, musi przyjmować wartości dodatnie, należy dodatkowo uwzględnić ograniczenia w przestrzeni rozwiązań. Dokonując wstępnego przeglądu metod, mogących posłużyć do optymalizacji tak złożonego problemu, należy również wziąć pod uwagę niezwykle duży nakład obliczeniowy potrzebny do jednorazowego wyznaczenia funkcji celu. Wystarczy bowiem zauważyć, że wartość (7.28) zostaje obliczona po wcześniejszym przeprowadzeniu  $k$  (7.31) symulacji działania systemu oraz obserwatora w przeciągu  $T$  iteracji. Wszystkie przytoczone wyżej uwagi, skutecznie zawężają krąg algorytmów mogących posłużyć do skutecznego projektowania obserwatorów nieliniowych. W następnym punkcie, zaprezentowana została technika, oparta na strategii ewolucyjnej  $(1 + \lambda)ES_\alpha$  z mutacją kierunkową (rozdział 5.2) oraz operatorem selekcji, uwzględniającym niepewność związaną z obecnością czynnika losowego w funkcji celu.

### 7.4.1 Dedykowana strategia ewolucyjna

Biorąc pod uwagę, iż liczba symulacji obserwatora, jaką trzeba wykonać w celu wyznaczenia pojedynczej wartości funkcji celu, może w rzeczywistości okazać się bardzo duża, niezbędnym wydaje się poszukiwanie sposobu na zredukowanie liczby próbek losowych (7.31). W pierwszej kolejności zauważmy, iż zbiór warunków początkowych, który jest podstawą oszacowania wartości funkcji celu  $\mathbf{X}^{(k)} = \{\mathbf{x}_0^{(i)}\}_{i=1}^k$  może zawierać wiele nieistotnych elementów. Zakładając ograniczoną zmienność funkcji celu (7.25), można przypuszczać, iż zastąpienie pewnego skupiska stanów początkowych systemu ich wartością średnią nie spowoduje znacznych różnic przy wyznaczaniu (7.28). Przypuszczenie to, otwiera drzwi przed możliwością zastosowania bardziej lub mniej skomplikowanych me-

tod klasteryzacji do redukcji zbioru losowego. Innym sposobem na wyeliminowanie nadmiarowości w zbiorze warunków początkowych systemu, jest zastosowanie sekwencji pseudo-losowych [76]. Jest to jednak możliwe tylko i wyłącznie w przypadku, gdy mamy do czynienia z jednorodną miarą prawdopodobieństwa (7.24) oraz przestrzenią stanów w formie hiperkostki. Niemniej jednak w rozdziale tym zaproponowane zostało bardziej intuicyjne podejście. Mianowicie zauważono, iż przestrzeń początkowych stanów systemu jest podzielona w pewien charakterystyczny sposób. W jej pewnych obszarach, funkcja celu przyjmuje wartości na tyle małe, że w porównaniu do wartości z innych podzbiorów, stają się one nie istotne dla minimalizacji (7.28). Dlatego też, zdecydowano się zastosować dwuetapowy proces optymalizacji. W pierwszym, przyjmując (7.31) następuje wygenerowanie zbioru warunków początkowych służących do wiarygodnej estymacji funkcji celu. Pozwala to również na stosunkowo gęste zaścielenie przestrzeni stanów początkowych systemu. Po wyznaczeniu (7.28) następuje zredukowanie wcześniej wylosowanego zbioru, poprzez wybór jedynie najgorszych przypadków t.j. warunków początkowych, dla których obserwator odznacza się najwolniejszą zbieżnością. Etap drugi, uwzględnia uruchomienie procedury optymalizacji bazującej już na zredukowanym zbiorze warunków początkowych. Po jej zakończeniu, znów następuje szczelne zaścielenie przestrzeni z wykorzystaniem sekwencji losowej o (7.31) elementach. W ten sposób cały proces powtarza się cyklicznie, aż do momentu spełnienia, wcześniej przyjętego kryterium stopu.

W celu uwzględnienia ograniczeń narzuconych na zmienne decyzyjne zastosowano prosty algorytm, rzutowania na brzeg obszaru dopuszczalnego. Uwzględniając dużą liczbę współczynników funkcji sklejaney, niezbędna do zdefiniowania obserwatora, zaproponowano rozwiązanie specjalnie dedykowane do problemów o znacznej liczbie wymiarów - mutacje kierunkowe oparte o rozkłady stabilne (rozdział 5.2). Pamiętając o tym, iż funkcja celu ma charakter losowy, heurystyka bazująca tylko i wyłącznie na zmianach położenia populacji, wydaje się być najlepszym rozwiązaniem (heurystyka nr 3, rozdział 5.2). Szkic strategii ewolucyjnej  $(1 + 1)ES_{\alpha}$ , uwzględniającej wszystkie wymienione wyżej aspekty, zamieszczony został w tabeli 7.1.

Tabela 7.1: Schemat strategii ewolucyjnej  $(1 + \lambda)ES_\alpha$  dedykowanej do optymalizacji funkcji celu definiującej odporne obserwatory nieliniowe EUIO

**1. Wybór parametrów metody:**

- $T_{\max}^{(\text{in})}, T_{\max}^{(\text{out})}$  - maksymalna liczba iteracji
- $\alpha, \sigma$  - parametry rozkładu stabilnego
- $r_w$  - względna szerokość przedziału ufności
- $\theta_0$  - rozwiązanie początkowe
- $\lambda$  - rozmiar populacji
- $1 - \delta$  - poziom ufności
- $\gamma$  - poziom istotności

**2. Inicjacja:**

- a) Ustal  $k_{\text{out}} = 1$ ,  $\theta^* = \theta_0$ , oraz  $N = 100000$

**3. Powtarzaj:**

- a) Utwórz zbiór  $\mathbf{X}^{(N)} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$  zgodnie z miarą praw.  $P$  (7.24)
- b) Utwórz zbiór  $A_k = \{a(\theta^*, \mathbf{x}^{(i)})\}_{i=1}^N$
- c) Oszacuj przybliżenie rozwiązania globalnego  $\omega^* = \hat{E}_P(\theta^*, \mathbf{X}^{(N)})$
- d) Utwórz zredukowany zbiór warunków początkowych  $\mathbf{X}^{(r)} = \{\mathbf{x}_i\}_{i=1}^r$   
gdzie  $\mathbf{x}_i \in \mathbf{X}^{(r)} \iff a(\theta^*, \mathbf{x}_0^{(i)}) > \gamma \omega^*$
- e) Ustal  $k_{\text{in}} = 1$

**4. Powtarzaj**

- a) Utwórz populację rozwiązań alternatywnych  $P_\lambda$   
 $P_\lambda = \{\theta_1, \theta_1, \dots, \theta_\lambda\}$ , gdzie  $\theta_i = \prod (\theta^* + \sigma \mathbf{Z}_i)$ ,  $i = 1, 2, \dots, \lambda$   
gdzie  $\mathbf{Z}_i \sim \chi_{\alpha, \sigma} \mathbf{d}^{(n)}$ ,  $\mathbf{d}^{(n)} \sim \mathcal{M}(\boldsymbol{\mu}, \kappa)$ ,  $\chi_{\alpha, \sigma} \sim |S_\alpha S(\sigma)|$
- b) Estymacja najkorzystniejszego kierunku mutacji  $\boldsymbol{\mu}$  (heurystyka nr 3 - rozdział 5)
- c) Wyznacz  $\Omega = \{\omega_i\}_{i=1}^\lambda$ , gdzie  $\omega_i = \hat{E}_P(\theta_i, \mathbf{X}^{(r)})$
- d) Modyfikuj rozwiązanie według reguły:  
jeśli  $\omega^* - \omega_{\min} > 2r_w \omega^*$  to  $\theta^* = \theta_{\min}$ ,  $\omega^* = \omega_{\min}$   
gdzie  $\theta_{\min} \in P_\lambda$  oznacza rozwiązanie o najmniejszej funkcji celu
- e) Ustal  $k_{\text{in}} = k_{\text{in}} + 1$

**5. Jeśli  $(k_{\text{in}} < T_{\max}^{(\text{in})})$  przejdź do kroku 4**

- a) Ustal  $k_{\text{out}} = k_{\text{out}} + 1$  oraz  $N = \left\lceil \ln\left(\frac{2}{\delta}\right) / (2(r_w \omega^*)^2) \right\rceil$

**6. Jeśli  $(k_{\text{out}} < T_{\max}^{(\text{out})})$  przejdź do kroku 3**

## 7.5 Badania symulacyjne

W celu zademonstrowania, skuteczności proponowanego podejścia, rozważmy zadanie estymacji stanu trójfazowego silnika indukcyjnego [91]. Wspomniany sys-

tem, był przedmiotem badań wielu wspaniałych opracowań, przy okazji testowania różnych metod estymacji stanu [26],[6]. Po rozłożeniu wektorów prądu i napięcia stojana oraz strumieni wirnika na składowe  $(\alpha - \beta)$  w układzie współrzędnych nieruchomym względem stojana, rozważany silnik indukcyjny może zostać przedstawiony w postaci układu równań:

$$\begin{aligned}\dot{x}_1 &= -\gamma \frac{K}{T_r} x_3 + K p x_5 x_4 + \frac{1}{\sigma L_s} u_1 \\ \dot{x}_2 &= -\gamma x_2 - K p x_5 x_3 + \frac{K}{T_r} x_4 + \frac{1}{\sigma L_s} u_2 \\ \dot{x}_3 &= \frac{M}{T_r} x_1 - \frac{1}{T_r} x_3 - p x_5 x_4 \\ \dot{x}_4 &= \frac{M}{T_r} x_2 + p x_5 x_3 - \frac{1}{T_r} x_4 \\ \dot{x}_5 &= \frac{pM}{J L_r} (x_3 x_2 - x_4 x_1) - \frac{T_L}{J} \\ y_1 &= x_1 \quad y_2 = x_2\end{aligned}$$

gdzie  $\mathbf{x} = [x_1, \dots, x_n]^T = [i_{s\alpha}, i_{s\beta}, \psi_{r\alpha}, \psi_{r\beta}, \omega]^T$  reprezentuje prądy stojana, strumienie magnetyczne wirnika, i prędkość kątową,  $\mathbf{u} = (u_{s\alpha}, u_{s\beta})$  wektorem napięć stojana,  $p$  jest liczbą par biegunów,  $T_L$  jest momentem obciążenia. Parametry  $T_r$ ,  $\sigma$ ,  $K$  i  $\gamma$  są zdefiniowane następująco:

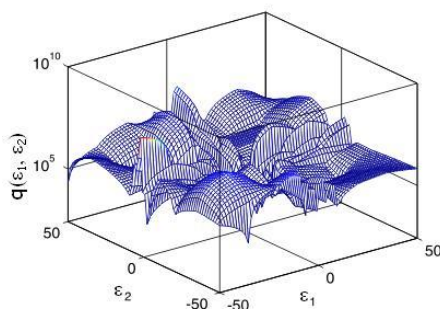
$$T_r = \frac{L_r}{R_r}, \quad \sigma = 1 - \frac{M^2}{L_s L_r}, \quad K = \frac{M}{\sigma L_s L_r^2}, \quad \gamma = \frac{R_s}{\sigma L_s} + \frac{R_r M^2}{\sigma L_s L_r^2},$$

gdzie  $R_s$ ,  $R_r$  i  $L_s$ ,  $L_r$  oznaczają odpowiednio rezystancje i indukcyjności fazowe stojana oraz wirnika,  $J$  jest momentem bezwładności wirnika.

Rozważmy przestrzeń początkowych stanów systemu:

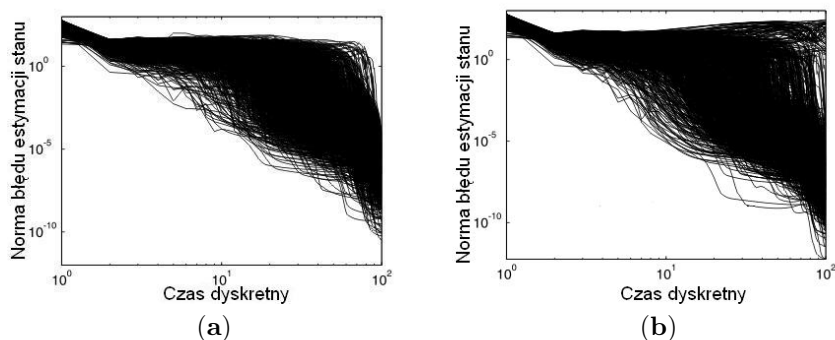
$$\mathbb{X} = \{[-276, 279] \times [-243, 369] \times S_{15}^{(2)}(\mathbf{0}) \times [-11, 56]\} \subset \mathbb{R}^5, \quad (7.32)$$

gdzie  $S_r^{(n)}(\mathbf{c}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\|_2 \leq r\}$ . Załóżmy, że każdy początkowy stan systemu jest równoprawdopodobny. W celu dokonania porównania skuteczności proponowanej metody, z metodami proponowanymi w literaturze, załóżmy brak sygnału zakłócającego proces estymacji stanu t.j.:  $\mathbf{d}_k = 0$ . Określmy również, horyzont czasowy  $T = 30$  niezbędny do pełnego określenia funkcji (7.25). Rozważany silnik indukcyjny charakteryzuje się dwoma bezpośrednio mierzalnymi zmiennymi stanu, co jednocześnie określa przestrzeń wektora błędu wyjścia  $\boldsymbol{\varepsilon}_k \in \mathbb{R}^2$ . W ten sposób, do określenia funkcji  $q(\boldsymbol{\varepsilon}_k)$ , w eksperymencie została wykorzystana dwuwymiarowa funkcja sklejana. Węzły funkcji rozłożone zostały na siatce  $20 \times 20$  równomiernie rozłożonych punktów na przestrzeni  $[-200, 200] \times [-200, 200]$ . W celu znalezienia optymalnej postaci funkcji  $q(\boldsymbol{\varepsilon}_k)$   $\boldsymbol{\theta}^* \subset \mathbb{R}^{400}$ , został zastosowany strategia ewolucyjna z mutacją stabilną (Tabela 7.1). Funkcja otrzymana w wyniku wspomnianej optymalizacji  $q(\boldsymbol{\varepsilon}_k)$  zaprezentowana została na Rysunku 7.1.



Rysunek 7.1: Funkcja  $q(\varepsilon_k)$  definiująca optymalny obserwator nieliniowy dla silnika indukcyjnego

Warto zauważyć, iż ustawienie macierzy instrumentalnych proponowanych w [26] skutkuje przyjęciem kwadratowej postaci funkcji  $q(\varepsilon_k)$ . Jak można zaobserwować na Rysunku 7.1, funkcja  $q(\varepsilon_k)$ , specjalnie dopasowana do rozważanego systemu na drodze optymalizacji ewolucyjnej, posiada znacznie bardziej skomplikowany kształt. Aby się przekonać, o tym, co w rzeczywistości zyskujemy stosując proponowane podejście, porównajmy skuteczność dwóch obserwatorów: otrzymanego w eksperymencie oraz tego proponowanego w [26]. W tym celu, dla losowo wybranych 1000 stanów początkowych silnika indukcyjnego, wyznaczono wartość estymatora (7.28) dla dwóch obserwatorów. Uzyskane wyniki  $3.38 \times 10^{-8}$  - dla proponowanego rozwiązania, oraz 0.016 dla obserwatora [26], wyraźnie świadczą o słuszności przyjętej metodologii. Ponadto na wykresie 7.2 zamieszczono, wykresy trajektorii normy błęd estymacji stanu dla warunków początkowych biorących udział w porównaniu.



Rysunek 7.2: Trajektorie  $\|\varepsilon_k\|_2$  dla 1000 losowo wybranych stanów początkowych silnika indukcyjnego. (a) - proponowane rozwiązanie ( $\hat{E}_P = 3.38 \times 10^{-8}$ ), oraz (b) - obserwator opisany w [26] ( $\hat{E}_P = 0.016$ )

## 7.6 Podsumowanie

Problem estymacji stanu systemów nieliniowych, nie znalazł dotychczas rozwiązania, które mogłoby być z powodzeniem stosowane do wszystkich systemów nieliniowych.



W rozdziale tym, zaprezentowano metodę estymacji deterministycznego systemu nieliniowego bazującą na Rozszerzonym Filtrze Kalmana. Wykorzystanie algorytmu pierwotnie przeznaczonego do pracy z systemami stochastycznymi, powoduje, iż istnieje pewna dowolność w określeniu macierzy instrumentalnych  $\mathbf{Q}_k$  oraz  $\mathbf{R}_k$ . Witczak [129] pokazał, że istnieje ścisły związek pomiędzy wartościami wspomnianych macierzy a współczynnikiem zbieżności proponowanego obserwatora.

W rozdziale tym wykorzystuje się powyższy fakt w celu uzyskania odpornych obserwatorów stanu. Między innymi, wykorzystując technikę odporności stochastycznej, pokazano jak można sprowadzić zadanie projektowania odpornych obserwatorów nieliniowych do problemu optymalizacji globalnej. Stopień złożoności zadania, przede wszystkim stochastyczny charakter oraz niezwykle duży wymiar przestrzeni rozwiązań, sprawia, iż nie istnieje wiele algorytmów mogących posłużyć do jego skutecznego rozwiązania. Dlatego też, zaproponowano, specjalnie do takich celów przystosowaną, strategię ewolucyjną z mutacją kierunkową. Ponadto, algorytm wzbogacony został o mechanizm selekcji bazujący na nierówności Chernoff'a, której celem jest ustalenie reguł akceptacji lepszych rozwiązań.

Przedstawiony przykład obliczeniowy polegający na estymacji stanu modelu silnika indukcyjnego, wyraźnie pokazał przewagę zaproponowanej procedury projektowania obserwatorów nieliniowych w stosunku do już istniejących rozwiązań.

## Rozdział 8

# Wnioski końcowe i kierunki dalszych badań

### 8.1 Podsumowanie wyników pracy

Niniejsza rozprawa skupia się na procesie projektowania efektywnych, ewolucyjnych algorytmów optymalizacji globalnej. W tym celu, zaproponowano modyfikację klasycznych procedur optymalizacji, polegającą na zastąpieniu najczęściej wykorzystywanego, w operatorze mutacji, rozkładu normalnego, przez rozkłady należące do klasy stabilnej. Ta na pozór niewielka zmiana, diametralnie zmienia prawa rządzące procesem ewolucji, a tym samym, wpływa na skuteczność algorytmów optymalizacji.

Przeprowadzone badania doprowadziły do uzyskania wyników, które można podzielić na metodyczne, teoretyczne oraz empiryczne. Do najistotniejszych z uzyskanych wyników metodycznych należą:

- zintegrowanie procedury estymacji dyskretnej miary spektralnej z adaptacyjnym algorytmem ewolucyjnym bazującym na rozkładzie populacji,
- opracowanie alternatywnej techniki wyznaczania przedziałów ufności dla oczekiwanej odpowiedzi nieliniowych modeli regresji,
- połączenie technik: aproksymacji wielowymiarowymi funkcjami sklejanymi, odporności stochastycznej oraz optymalizacji ewolucyjnej, w zadaniu projektowania odpornych obserwatorów nieliniowych,

Zastosowanie rozkładów stabilnych w zupełnie nowym dla nich obszarze, jakim jest optymalizacja globalna, wymusiło spojrzenie na tę klasę modeli probabilistycznych z zupełnie innej strony. Między innymi, żadne z dotychczas znanych autorowi rozprawy, opracowań statystycznych nie poruszało pojęcia normy wektora stabilnego, która w ramach obliczeń ewolucyjnych odgrywa kluczową rolę. Do najważniejszych wyników teoretycznych uzyskanych w pracy, zaliczyć zatem można:

- wyznaczenie współczynników zbieżności liniowej dla strategii ewolucyjnej  $(1 + 1)ES_{\alpha}$  z izotropową mutacją stabilną - rozdział 4.2.3,
- zdefiniowanie postaci całkowitej generatora gęstości izotropowych rozkładów stabilnych - rozdział ?? oraz dodatek A,
- dostarczenie wzoru aproksymacyjnego na położenie mediany rozkładu normy izotropowego rozkładu stabilnego - rozdział A,
- opracowanie oraz weryfikacja skuteczności mutacji opartych o rozkłady kierunkowe - 5,

- opracowanie mutacji wolnej od efektu otoczenia - rozdział 4.2.2,
- udowodnienie twierdzenia o asymptotycznym zdegenerowaniu rozkładu brzegowego losowego wektora izotropowego - rozdział 5.1.1,
- udowodnienie twierdzenia o istnieniu momentów uporządkowanej, stabilnej zmiennej losowej - rozdział 2.1,

Każda nowa metoda optymalizacji globalnej, wymaga eksperymentalnego potwierdzenia skuteczności proponowanego rozwiązania. Dlatego też, w treści rozprawy przedstawiono kompleksowe badania symulacyjne z wykorzystaniem standardowych, jak i wynikających z zastosowań inżynierskich, środowisk testowych. Do najważniejszych osiągnięć tego typu należą:

- dostarczenie optymalnej konfiguracji strategii adaptacyjnej dla izotropowego rozkładu stabilnego - rozdział 4.2.5,
- eksperymentalna analiza efektu symetrii oraz efektu otoczenia odniesieniu do wektora stabilnego z niezależnymi komponentami losowymi oraz wektora o sferycznej symetrii - rozdział ??,4.2.2,
- eksperymentalna analiza odporności izotropowego rozkładu stabilnego na niepewność związaną z doбором parametru skali w zadaniach optymalizacji globalnej - rozdział 4.2.4,
- dostarczenie optymalnej konfiguracji algorytmu ARS z pulą konkurujących rozkładów stabilnych w zadaniach optymalizacji funkcji wariacji nieliniowych modeli neuronowych - rozdział 7,

## 8.2 Plany dalszych badań

Prace prowadzone nad rozprawą zaowocowały szeregiem spostrzeżeń oraz nowych pomysłów, które ze względu na ograniczony charakter opracowania, nie zostały ujęte w jego treści.

1. Wyznaczone optymalne konfiguracje algorytmów ewolucyjnych w rozdziałach 4.2.5 oraz 7 odnoszą się jedynie do pewnych szczególnych przypadków jeśli chodzi o wymiar przestrzeni przeszukiwań. Z teoretycznego punktu widzenia, cennym wynikiem byłoby wyznaczenie trajektorii w przestrzeni parametrów konfiguracyjnych algorytmów, które gwarantowałyby ich optymalną skuteczność.
2. Zaprezentowany w rozdziale 4.3.1 schemat adaptacji dyskretnej miary spektralnej, choć wykazał się znaczną poprawą skuteczności optymalizacyjnej, to należy pamiętać, iż jest on dedykowany jedynie do przestrzeni dwuwymiarowych. Uogólnienie tej procedury dla problemów o dowolnym rozmiarze, stanowiłoby z pewnością silną konkurencję dla innych rozwiązań tej klasy.

# Bibliografia

- [1] Website of damadics: *Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems*. <http://diag.mchtr.pw.edu.pl/damadics/>.
- [2] Papers of the special sessions (2003): *DAMADICS I, II, III*, 2003.
- [3] Husein Abdul-Hamid and John P. Nolan. Multivariate stable densities as functions of one dimensional projections. *J. Multivar. Anal.*, 67(1):80–89, 1998.
- [4] Atkinson A.C. and Donev A.N. *Optimum Experimental Designs*. New York: Oxford University Press, 1992.
- [5] A. Achim, P. Tsakalides, and A. Bezerianos. Sar image denoising via bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 41:1773–1784, 2003.
- [6] A.Mansouri, M.Chenafa, A.Bouhenna, and E.Etien. Powerful nonlinear observer associated with field-oriented control of an induction motor. *Int. J. Appl. Math. Comput. Sci.*, 14(2):209–220, 2004.
- [7] T. W. Anderson. *The Statistical Analysis of Time Series*. Wiley and Sons, 1971.
- [8] R. S. Anderssen and P. Bloomfield. Properties of the random search in global optimization. *Journal Journal of Optimization Theory and Applications*, 16(5-6):383–398, 1975.
- [9] E. F. Andrea, Clementi, D.P. Jose, Rolim, and E. Urland. Randomized parallel algorithms. In *Solving Combinatorial Optimization Problems in Parallel*, pages 25–50, 1996.
- [10] J. Arabas. *Wykłady z algorytmów ewolucyjnych*. Wydawnictwa Naukowo-Techniczne, Warszawa, Polska, 2001.
- [11] A. Araujo and E. Gine. *The central limit theorem for real and Banach valued random variables*. John Wiley, New York, USA, 1980.
- [12] D. V. Arnold and H.-G. Beyer. On the benefits of populations for noisy optimization. *Evol. Comput.*, 11(2):111–127, 2003.
- [13] D. V. Arnold and H.G. Beyer. *Noisy Local Optimization with Evolution Strategies*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.
- [14] D. V. Arnold and H.G. Beyer. A comparison of evolution strategies with other direct search methods in the presence of noise. *Comput. Optim. Appl.*, 24(1):135–159, 2003.
- [15] A. Auger and N. Hansen. Reconsidering the progress rate theory for evolution strategies in finite dimensions. In *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 445–452, New York, NY, USA, 2006. ACM.
- [16] A. Banerjee, P. Burlina, and R. Chellappa. Adaptive target detection in foliage-penetrating sar images using alpha-stable models. *IEEE Transactions on Image Processing*, 8:1823–1831, 1999.

- [17] D.M. Bates and D.G. Watts. *Nonlinear Regression Analysis and Its Application*. Willey & Sons, New York, 1988.
- [18] H. G. Beyer. Toward a Theory of Evolution Strategies: Self-Adaptation. *Evolutionary Computation*, 3(3):311–347, 1996.
- [19] H.-G. Beyer and D. V. Arnold. Qualms regarding the optimality of cumulative path length control in csa/cma-evolution strategies. *Evol. Comput.*, 11(1):19–28, 2003.
- [20] H. G. Beyer and H. P. Schwefel. Evolution strategies - a comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.
- [21] H. G. Beyer and H. P. Schwefel. Evolution strategies – a comprehensive introduction. *Journal Natural Computing, Publisher Springer Netherlands*, 1(1):3–52, 2002.
- [22] Hans Georg Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.
- [23] H.G. Beyer. *The theory of evolution strategies*. Springer-Verlag New York, Inc., New York, NY, USA, 2001.
- [24] A. Bienvenue and O. Francois. Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties. *Theor. Comput. Sci.*, 306(1-3):269–289, 2003.
- [25] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization*. Springer Verlag, 2003.
- [26] M. Boutayeb and D. Aubry. A strong tracking extended kalman observer for nonlinear discrete-time systems. *IEEE Trans. Automat. Contr.*, 44:1550–1556, 1999.
- [27] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [28] J. Branke. *Evolutionary Optimization in Dynamic Environments*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
- [29] T. Byczkowski, J.P. Nolan, and B. Rajput. Approximation of multidimensional stable densities. *J. of Mult. Anal.*, 46(1):13–31, 1993.
- [30] K. Chellapilla. Combining mutation operators in evolutionary programming. *IEEE Trans. Evolutionary Computation*, 2(3):91–96, 1998.
- [31] J. Chen and R. J. Patton. *Robust Model-based Fault Diagnosis for Dynamic Systems*. London: Kluwer Academic Publishers, London, 1999.
- [32] L. N. Chrysostomos and M. Shao. *Signal Processing with Alpha-Stable Distribution and Applications*. Wiley, Chichester, England., 1981.
- [33] H.F. Cullen. The stone-weierstrass theorem and complex stone-weierstrass theorem. In: *Introduction to General Topology*, pages 286–293. Boston, MA: Heath, 1968.
- [34] C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.
- [35] I. Dhillon and S. Sra. Modeling data using directional distributions, 2003.
- [36] H. Dong, J. He, H. Huang, and W. Hou. Evolutionary programming using a mixed mutation strategy. *Information Sciences*, 177:312–327, 2007.
- [37] R. Durrett. *Probability: Theory and Examples, 2nd Edition*. Duxbury Press, 1995.
- [38] Kai Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London, 1990.

- [39] L.J. Fogel, A.J. Owens, and M.J. Walsh. *Artificial Intelligence through Simulated Evolution*. John Wiley, New York, 1966.
- [40] H. Follmer and M. Schweizer. A microeconomic approach to diffusion models for stock prices. *Math. Finance*, 3:1–23, 1993.
- [41] R. Galar. Evolutionary search with soft selection. *Biological Cybernetics*, 60:357–364, 1989.
- [42] R. Galar. Simulation of local evolutionary dynamics of small populations. *Biological Cybernetics*, 65:37–45, 2004.
- [43] R. Galar. Some evolutionary algorithms inspired insights into evolutionary drive of development processes. In *Algorytmy ewolucyjne i optymalizacja globalna : materiały VIII krajowej konferencji; ISBN: 83-9145580-7-5*, pages 79–87, Korbiewów, Polska, 2005. Wydział Elektroniki i Technik Informacyjnych Politechniki Warszawskiej, Warszawa, Oficyna Wydaw. Politechniki Warszawskiej.
- [44] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis. Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Transactions on Multimedia*, 1(3):291–301, 1999.
- [45] M. Greene and B. Fielitz. The effect of long-term dependence on risk return models of common stocks. *Operations Research*, 27:944–951, 1979.
- [46] M. Gutowski. Lévy flights as an underlying mechanism for a global optimization algorithm. In *5th Conf. Evolutionary Algorithms and Global Optimization*, Jastrzebia Gora, Poland, 2001. Wydział Elektroniki i Technik Informacyjnych Politechniki Warszawskiej, Warsaw University of Technology Press.
- [47] M. Gutowski. Levy sampling made easy. In *Algorytmy ewolucyjne i optymalizacja globalna : materiały VII krajowej konferencji; ISBN: 83-914580-6-7*, pages 63–67, Kazimierz Dolny, Polska, 2004. Politechnika Warszawska, Warszawa.
- [48] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [49] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [50] Nikolaus Hansen and Andreas Ostermeier. Convergence properties of evolution strategies with the derandomized covariance matrix adaption . . . In Hans-Jürgen Zimmermann, editor, *Eufit '97 – 5th European Congress on Intelligent Techniques and Soft Computing*, pages 650–654, Aachen, 1997. Verlag Mainz, Wissenschaftsverlag.
- [51] William E. Hart. *Adaptive Global Optimization with Local Search*. PhD thesis, San Diego, CA, 1994.
- [52] F. Hiai and D. Petz. *The Semicircle Law, Free Random Variables and Entropy (Mathematical Surveys & Monographs)*. American Mathematical Society, Boston, MA, USA, 2006.
- [53] J.B. Hiriart-Urruty and C. Lemaréchal . *Convex Analysis and Minimization Algorithms*. Springer Verlag, Heidelberg, 1996. Two volumes - 2nd printing.
- [54] J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor., 1975.
- [55] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *J. of Global Optimization*, 13(4):455–492, 1998.
- [56] I. Karcz-Dulęba. Asymptotic behaviour of a discrete dynamical system generated by a simple evolutionary process. *Int. J. Appl. Math. Comput. Sci.*, 14(1):79–90, 2004.

- [57] F. Kemp. An introduction to sequential monte carlo methods. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(4):694–695, December 2003.
- [58] S. Kern, S. Uller, D. Uche, N. Hansen, and P. Koumoutsakos. Learning probability distributions in continuous evolutionary algorithms. In *Workshop on Fundamentals in Evolutionary Algorithms, Thirtieth International Colloquium on Automata, Languages and Programming*, Eindhoven, 2004.
- [59] P. Kidmose. Alpha-stable distributions in signal processing of audio signals. In *41st Conference on Simulation and Modelling, Scandinavian Simulation Society*, pages 87–94, 2000.
- [60] J. Korbicz, J. Kościelny, Z. Kowalczyk, and W. Cholewa. *Fault diagnosis : models, artificial intelligence, applications*. Springer-Verlag, Berlin, 2004.
- [61] J. Korbicz, A. Obuchowicz, and P. Prętki. Some aspects of multidimensional mutations in evolutionary algorithms. In S. Zadrożny ed. P. Grzegorzewski, M. Krawczak, editor, *Issues in soft computing theory and applications, ISBN: 83-87674-98-2, Problemy Współczesnej Nauki - Teoria i Zastosowania : Informatyka*, pages 3–45. Akademicka Oficyna Wydaw. EXIT, Warszawa, 2005.
- [62] P. Larranaga and J.A. Lozano. *Estimation of Distribution Algorithms: A New Tool for Evolutionary Optimization*. Kluwer Academic Publishers, Boston, Boston, USA, 2001.
- [63] C. Y. Lee and X. Yao. Evolutionary programming using the mutations based on the levy probability distribution. *IEEE Transactions on Evolutionary Computation*, 8:1–13, 2004.
- [64] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of ethernet traffic. *IEEE/ACM Transactions on Networking*, 2:1–15, 1994.
- [65] X. Liu and W. Xu. A new filled function applied to global optimization. *Comput. Oper. Res.*, 31(1):61–80, 2004.
- [66] H. Lu. *State-of-the-art Multiobjective Evolutionary Algorithms—Pareto Ranking, Density Estimation and Dynamic Population*. PhD thesis, Oklahoma State University, Stillwater, Oklahoma, August 2002.
- [67] D. J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*, NATO Science Series, pages 175–204. Kluwer Academic Press, 1998.
- [68] K. V. Mardia. Directional statistics and shape analysis. *Journal of Applied Statistics*, 26(8):949–957, December 1999. available at <http://ideas.repec.org/a/taf/japsta/v26y1999i8p949-957.html>.
- [69] K.V. Mardia and P.E. Jupp. *Directional statistics*. John Wiley and Sons Ltd, 2000.
- [70] R. Maronna, D. Martin, and V. Yohai. *Robust Statistics - Theory and Methods*. Wiley, 2006.
- [71] Ch. I. Marrison and R. F. Stengel. Robust control system design using random search and genetic algorithms. *IEEE Transactions on Automatic Control*, 42(6):835–839, 1997.
- [72] J. H. McCulloch. *Financial applications of stable distributions*. In G. S. Maddala and C. R. Rao (Eds.), *Handbook of Statistics*, New York: North-Holland, 1996.
- [73] M. Metenidis, M. Witczak, and J. Korbicz. A novel genetic programming approach to nonlinear system modelling : application to the damadics benchmark problem. *Engineering Applications of Artificial Intelligence*, Vol. 17:363–370, 2004.

- [74] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs (3rd ed.)*. Springer-Verlag, London, UK, 1996.
- [75] R. Modarres and J.P. Nolan. A method for simulating stable random vectors. *Computational Statistics*, 9:11–19, 1994.
- [76] William J. Morokoff and Russel E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251–1279, 1994.
- [77] R. W. Morrison. *Designing Evolutionary Algorithms for Dynamic Environments*. SpringerVerlag, 2004.
- [78] H. Muehlenbein and G. Paaß. From recombination of genes to the estimation of distribution. *Lecture Notes in Computer Science 1141: Parallel Problems Solving from Nature- PPSN IV*, pages 178–187, 1996.
- [79] H. Narihisa, T. Taniguchi, and M. Thuda. Efficiency of parallel exponential evolutionary programming. In *ICPPW '05: Proceedings of the 2005 International Conference on Parallel Processing Workshops (ICPPW'05)*, pages 588–595, Washington, DC, USA, 2005. IEEE Computer Society.
- [80] V. Nissen and J. Propach. Optimization with noisy function evaluations. In *PPSN V: Proceedings of the 5th International Conference on Parallel Problem Solving from Nature*, pages 159–168, London, UK, 1998. Springer-Verlag.
- [81] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer-Verlag New York, 1999.
- [82] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston, 2007. In progress, Chapter 1 online at [academic2.american.edu/~jpnolan](http://academic2.american.edu/~jpnolan).
- [83] J.P. Nolan, A.K. Panorska, and J.H. McCulloch. Estimation of stable spectral measures. stable non-gaussian models in finance and econometrics. *Math. Comput. Modelling*, 34(9):1113–1122, 2001.
- [84] A. Obuchowicz. Adaptation in a time-varying landscape using an evolutionary search with soft selection. In *Algorytmy ewolucyjne i optymalizacja globalna : Materiały III Krajowej Konferencji*, pages 245–251, Potok Złoty, Polska, 1999. Instytut Systemów Elektronicznych, Instytut Automatyki i Informatyki Stosowanej Politechniki Warszawskiej, Warszawa, Oficyna Wydawnicza Politechniki Warszawskiej.
- [85] A. Obuchowicz. *Evolutionary Algorithms in Global Optimization and Dynamic System Diagnosis*. Lubuskie Scientific Society, Zielona Góra, Zielona Góra, Poland, 2003.
- [86] A. Obuchowicz. Multidimensional mutations in evolutionary algorithms based on real-valued representation. *International Journal of Systems Science*, Vol. 34(no 7):469–483, 2003.
- [87] A. Obuchowicz, A. Pieczyński, M. Kowal, and P. Prętki. Solving optimization tasks in the construction of diagnostic systems. In ed. by J. Korbicz, editor, *Measurements models systems and design*, ISBN: 978-83-206-1644-6, pages 204–239. Wydaw. Komunikacji i Łączności, Warszawa, 2007.
- [88] A. Obuchowicz and P. Prętki. Phenotypic evolution with a mutation based on symmetric  $\alpha$ -stable distributions. *International Journal of Applied Mathematics and Computer Science*, Vol. 14(no 3):289–316, 2004.
- [89] A. Obuchowicz and D. Wawrzyniak. Evolutionary adaptation in non-stationary environments: a case study. *Lecture Notes in Computer Science : Parallel processing and applied mathematics - 6th international conference*, Vol. 3911:439–446, 2006.



- [90] Esin Onbasoglu and Linet Ozdamar. Parallel simulated annealing algorithms in global optimization. *J. of Global Optimization*, 19(1):27–50, 2001.
- [91] T. Orłowska-Kowalska. *Bezczujnikowe układy napędowe z silnikami indukcyjnymi*. Oficyna Wydawnicza Politechniki Wrocławskiej, Seria Wydawnicza Komitetu Elektrotechniki PAN Postępy Napędu Elektrycznego i Energoelektroniki, T. 48, Wrocław, 2003.
- [92] K. Patan. Simultaneous perturbation stochastic approximation in global optimisation. *Prace Naukowe Politechniki Warszawskiej : Elektronika : Evolutionary computation and global optimization 2006*, (156):343–350, 2006. Oficyna Wydaw. Politechniki Warszawskiej, Warszawa.
- [93] M. Patan. *Optimal observation strategies for parameter estimation of distributed systems*, ISBN: 83-89712-03-2. Lecture Notes in Control and Computer Science, Vol. 5. University of Zielona Góra Press, Zielona Góra, 2004.
- [94] Adam Piskorek. *Równania całkowe. Elementy teorii i zastosowania*. WNT, 3 edition, 1997.
- [95] P. Prętki and A. Obuchowicz. The progress rate of simple evolutionary strategies with an isotropic  $\alpha$ -stable mutation. In *Selected problems of computer science : proceedings of the 1st Polish and international PD forum-conference on computer science*, ISBN: 83-60434-25-5, pages 93–100. Academic Publishing House EXIT, Warsaw, 2005.
- [96] P. Prętki and A. Obuchowicz. Directional distributions and their application to evolutionary algorithms. *Lecture Notes in Artificial Intelligence : Artificial Intelligence and Soft Computing - ICAISC 2006*, Vol. 4029:440–449, 2006.
- [97] P. Prętki and A. Obuchowicz. Evolutionary bands for the expected response in nonlinear regression. *Prace Naukowe Politechniki Warszawskiej : Elektronika : Evolutionary computation and global optimization 2006*, (156):359–364, 2006. Oficyna Wydaw. Politechniki Warszawskiej, Warszawa.
- [98] P. Prętki and M. Witczak. Assessment and minimization of parametric uncertainty for multi-output neural networks - application to fault diagnosis. In *Recent Developments in Artificial Intelligence Methods - AI-METH 2005*; ISBN: 83-921605-8-4, pages 155–160, Gliwice, Polska, 2005. Centre of Excellence AI-METH, Silesian University of Technology, Polish Association for Computational Mechanics, Gliwice, Silesian University of Technology.
- [99] P. Prętki and M. Witczak. Developing measurement selection strategy for neural network models. *Lecture Notes in Computer Science : Artificial neural networks: Formal Models and Their Applications - ICANN 2005, part 2*, Vol. 3697:79–84, 2005. + CD-ROM.
- [100] I. Rechenberg. *Cybernetic solution path of an experimental problem*. Roy. Aircr. Establ., libr. transl. 1122. Farnborough, Hants, UK., 1965.
- [101] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag, Stuttgart, 1973.
- [102] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill: New York, 1976.
- [103] G. Rudolph. Convergence of evolutionary algorithms in general search spaces. In *Proc. ICEC'96, Nagoya*, pages 50–54, 1996.
- [104] G. Rudolph. Local convergence rates of simple evolutionary algorithms with cauchy mutations. *IEEE Transactions on Evolutionary Computation*, 1(4):249–258, 1997.
- [105] Hui S. and S.H. Zak. Observer design for systems with unknown input. *Int. J. Appl. Math. Comput. Sci.*, 15(5):431–446, 2005.

- [106] R. Salomon. The evolutionary-gradient-search procedure. In John R. Koza, Wolfgang Banzhaf, Kumar Chellapilla, Kalyanmoy Deb, Marco Dorigo, David B. Fogel, Max H. Garzon, David E. Goldberg, Hitoshi Iba, and Rick Riolo, editors, *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pages 852–862, University of Wisconsin, Madison, Wisconsin, USA, 22–25 1998. Morgan Kaufmann.
- [107] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York, 1994.
- [108] M. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization, 2001.
- [109] R. Schaefer. *Podstawy genetycznej optymalizacji globalnej*. Wydawnictwo Uniwersytetu Jagiellońskiego, 2002.
- [110] J. Schumpeter. *Depressions: Can we learn from past experience? In Economics of the Recovery Program*. 1934.
- [111] Hanss Paul Schwefel. *Evolution and Optimum Seeking*. Wiley, New York, 1995.
- [112] H.P. Schwefel. *Numerical Optimization of Computer Models*. Wiley, Chichester, England., 1981.
- [113] J. Shao. *Mathematical Statistics*. Springer-Verlag, New York, USA, 1999.
- [114] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37:332–341, 1992.
- [115] J. C. Spall. Implementation of simultaneous perturbation algorithm for stochastic optimization. *IEEE Trans. Aerosp. Electron. Syst.*, 34:817–823, 1998.
- [116] J. C. Spall. *Introduction to Stochastic Search and Optimization*. Wiley, Hoboken, NJ, 2003.
- [117] J. G. Taylor, editor. *Mathematical Approaches to Neural Networks*. Elsevier Science Publishers, 1993.
- [118] P. B. Thanedar, J. S. Arora, G. Y. Li, and T. C. Lin. Robustness, generality and efficiency of optimization algorithms for practical applications. *Journal Structural and Multidisciplinary Optimization*, 2:203–212, 1990.
- [119] A. Törn and A. Zilinskas. *Global optimization*. Springer-Verlag New York, Inc., New York, NY, USA, 1989.
- [120] K. Trojanowski and Z. Michalewicz. Evolutionary algorithms for non-stationary environments. In *Proc. of 8th Workshop: Intelligent Information systems*, pages 229–240, ?, 1999. ICS PAS Press.
- [121] M. Vidyasagar. Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, 37(10):1515–1528, 2001.
- [122] E. Walter and L. Pronzato. *Identification Of Parametric Models From Experimental Data. Communication and Control Engineering*. Berlin: Springer-Verlag, 1997.
- [123] G. Wang, E.D. Goodman, and W.F. Punch. Toward the optimization of a class of black box optimization algorithms. In *ICTAI '97: Proceedings of the 9th International Conference on Tools with Artificial Intelligence*, page 348, Washington, DC, USA, 1997. IEEE Computer Society.
- [124] D. Wawrzyniak and A. Obuchowicz. New approach to optimization with evolutionary algorithms in non-stationary environments. *Proceedings of Artificial Intelligence Studies*, Vol. 3:187–196, 2006.

- [125] D. Wawrzyniak and A. Obuchowicz. New evolutionary algorithm modifications in non-stationary environments. *Prace Naukowe Politechniki Warszawskiej : Elektronika : Evolutionary computation and global optimization 2007*, (160):285–292, 2007. Oficyna Wydaw. Politechniki Warszawskiej, Warszawa.
- [126] M. Witczak. *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems. From Analytical to Soft Computing Approaches*. Springer Berlin. Heidelberg, Berlin, 2007.
- [127] M. Witczak and J. Korbicz. Observers and genetic programming in the identification and fault diagnosis of non-linear dynamic systems. In *Fault diagnosis : models, artificial intelligence, applications, ISBN: 3-540-40767-7*, pages 457–509. Springer-Verlag, Berlin, 2004.
- [128] M. Witczak, A. Obuchowicz, and J. Korbicz. Genetic programming based approaches to identification and fault diagnosis of non-linear dynamic systems. *International Journal of Control*, 75(13):1012–1031, 2002.
- [129] M. Witczak and P. Prętki. Design of an extended unknown input observer with stochastic robustness techniques and evolutionary algorithms. *International Journal of Control*, 80:749–762, 2007.
- [130] M. Witczak and P. Prętki. Designing neural-network-based fault detection systems with d-optimum experimental conditions. *Computer Assisted Mechanics and Engineering Sciences*, (no 12):279–291, 2005.
- [131] M. Witczak and P. Prętki. Design of an extended unknown input observer with stochastic robustness techniques and evolutionary algorithms. *International Journal of Control*, Vol. 80(no 5):749–762, 2007.
- [132] M. Witczak, P. Prętki, J. Korbicz, and V. Puig. Design of an extended unknown input observer. In *IAR Annual Meeting 2006 and Workshop on Advanced Control Diagnosis*, pages [6] CD-ROM, Nancy, France, 2006. Nancy.
- [133] D. Wolpert and W. Macready. *The mathematics of search*, 1994.
- [134] D. Wolpert and W. MacReady. No free lunch theorems for optimization, 1996.
- [135] X. Yao and Y. Liu. Fast evolution strategies. In Peter J. Angeline, Robert G. Reynolds, John R. McDonnell, and Russ Eberhart, editors, *Evolutionary Programming VI*, pages 151–161, Berlin, 1997. Springer.
- [136] X. Yao, Y. Liu, and G. Liu. Evolutionary programming made faster. *IEEE Trans. on Evolutionary Computation*, 3(2):82–102, 1999.
- [137] Y. Yuan. Global convergence of trust region algorithms for nonsmooth optimization. Technical Report DAMTP 1983/NA13, Cambridge, England, 1983.
- [138] Y. Zhang. General robust-optimization formulation for nonlinear programming. *Journal of optimization theory and applications*, 132:111–124, 2007.

## Dodatek A

# Funkcja gęstości izotropowego rozkładu stabilnego

Wykorzystajmy definicję odwrotnej transformaty Fouriera do funkcji charakterystycznej rozkładu izotropowego:

$$\begin{aligned}g(\mathbf{x}) &= (2\pi)^{-n} \int_{\mathbb{R}^n} \exp(-j\mathbf{x}^T \mathbf{k}) \varphi(\mathbf{k}) d\mathbf{k} \\ &= (2\pi)^{-n} \int_{\mathbb{R}^n} \cos(\mathbf{x}^T \mathbf{k}) \varphi(\mathbf{k}) d\mathbf{k} - j(2\pi)^{-n} \int_{\mathbb{R}^n} \sin(\mathbf{x}^T \mathbf{k}) \varphi(\mathbf{k}) d\mathbf{k}\end{aligned}$$

Pierwsza z powyższych całek, może zostać uproszczona w następujący sposób:

$$\begin{aligned}\int_{\mathbb{R}^n} \cos(\mathbf{x}^T \mathbf{k}) \varphi(\mathbf{k}) d\mathbf{k} &= \int_0^\infty \int_{S_r^n} \cos(\mathbf{x}^T \mathbf{s}) \varphi(\mathbf{s}) d\mathbf{s} dr \\ &= \int_0^\infty \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \int_{S_r^n} \cos(\mathbf{x}^T \mathbf{s}) d\mathbf{s} dr \\ &= \int_0^\infty \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \frac{r^{n-1} 2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_{-1}^1 \cos(r\|\mathbf{x}\|t) (1-t^2)^{\frac{n-3}{2}} dt dr \\ &= \int_0^\infty \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \frac{r^{n-1} 4\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_0^1 \cos(r\|\mathbf{x}\|t) (1-t^2)^{\frac{n-3}{2}} dt dr\end{aligned}$$

W podobny sposób należy postąpić z drugą całką w (??), otrzymując:

$$\int_{\mathbb{R}^n} \sin(\mathbf{x}^T \mathbf{k}) \varphi(\mathbf{k}) d\mathbf{k} = 0$$

W ten sposób, otrzymuje się ostateczną postać funkcji gęstości  $n$ -wymiarowego wektora stabilnego:

$$g(\mathbf{x}) = \frac{2^{2-n}}{\pi^{\frac{n+1}{2}} \Gamma(\frac{n-1}{2})} \int_0^\infty r^{n-1} \exp(-2^{-\alpha/2} \sigma^\alpha r^\alpha) \int_0^1 \cos(r\|\mathbf{x}\|_2 t) (1-t^2)^{\frac{n-3}{2}} dt dr$$

## Dodatek B

# Twierdzenie o istnieniu momentów stabilnej statystyki rankingowej

Rozpatrzmy ciąg zmiennych losowych  $Z_1, Z_2, \dots, Z_\lambda$ , gdzie  $Z_i \sim |S_\alpha S(1)|$ ,  $i = 1, \dots, \lambda$  oraz jego uporządkowaną wersję  $Z_{1:\lambda} < Z_{2:\lambda} < \dots < Z_{\lambda:\lambda}$ . W ten sposób zmienną losową  $Z_{1:\lambda}$  można utożsamiać z rozkładem minimalnym próby  $\lambda$ -elementowej t.j.:  $Z_{1:\lambda} = \min\{Z_1, Z_2, \dots, Z_\lambda\}$ , natomiast  $Z_{\lambda:\lambda}$  z rozkładem maksymalnym, t.j.:  $Z_{\lambda:\lambda} = \max\{Z_1, Z_2, \dots, Z_\lambda\}$ . Wówczas  $k$ -ty moment zmiennej losowej  $Z_{i:\lambda}$  można obliczyć stosując wzór [113]:

$$E[Z_{i:\lambda}^k] = \int_0^\infty z^k f_{\alpha,i:\lambda}(z) dz, \quad (\text{B.1})$$

gdzie  $f_{\alpha,i:\lambda}(\cdot)$  jest funkcją gęstości zmiennej losowej  $Z_{i:\lambda}$ , daną wzorem:

$$f_{\alpha,i:\lambda}(z) = \frac{\lambda!}{(i-1)!(\lambda-i)!} [F_\alpha(z)]^{i-1} [1 - F_\alpha(z)]^{\lambda-i} f_\alpha(z), \quad (\text{B.2})$$

gdzie  $f_\alpha(z)$  funkcją gęstości zmiennej losowej  $|S_\alpha S(1)|$ , a  $F_\alpha(z)$  jej dystrybuantą. Oczywiście zachodzi

$$f_\alpha(z) = 2f_{\alpha,1}(z)I_{0,\infty} \quad (\text{B.3})$$

$$F_\alpha(z) = 2F_{\alpha,1}(z)I_{0,\infty} \quad (\text{B.4})$$

gdzie  $f_{\alpha,1}$  oraz  $F_{\alpha,1}$  są odpowiednio funkcją gęstości oraz dystrybuantą symetrycznej, stabilnej zmiennej losowej  $S_\alpha S(\sigma = 1)$ , a  $I_A$  jest funkcją wskaźnikową zbioru  $A$ . Z teorii funkcji wolno-zmiennych wiemy że prawdziwa jest zależność:

$$\int_0^\infty g(z) dz < \infty \Leftrightarrow \forall z > 0 : \lim_{h \rightarrow \infty} \frac{g(hz)}{g(h)} = z^a$$

gdzie  $a < -1$ . Rozważmy zatem następującą granicę

$$\begin{aligned} \lim_{h \rightarrow \infty} \frac{(zh)^k f_{i:\lambda}(zh)}{h^k f_{i:\lambda}(z)} &= z^k \lim_{h \rightarrow \infty} \frac{f_{i:\lambda}(zh)}{f_{i:\lambda}(z)} \\ &= z^k \lim_{h \rightarrow \infty} \left[ \frac{F_\alpha(zh)}{F_\alpha(h)} \right]^{i-1} \left[ \frac{1 - F_\alpha(zh)}{1 - F_\alpha(h)} \right]^{\lambda-i} \frac{f_\alpha(zh)}{f_\alpha(h)} \end{aligned} \quad (\text{B.5})$$

Zauważmy, że:

$$\lim_{h \rightarrow \infty} \frac{F_\alpha(zh)}{F_\alpha(h)} = 1 \quad (\text{B.6})$$

Stosując regułę l'Hospital'a otrzymujemy

$$\lim_{h \rightarrow \infty} \frac{1 - F_\alpha(zh)}{1 - F_\alpha(h)} = \lim_{h \rightarrow \infty} \frac{z f_\alpha(zh)}{f_\alpha(h)} \quad (\text{B.7})$$

Biorąc pod uwagę twierdzenie o asymptotycznej postaci ogonów symetrycznego rozkładu stabilnego 5 otrzymujemy:

$$\lim_{z \rightarrow \infty} f_\alpha(z) \sim \alpha C_\alpha z^{-(1+\alpha)},$$

gdzie  $C_\alpha$  jest stałą określoną w twierdzeniu 5.

$$\lim_{h \rightarrow \infty} \frac{f_\alpha(zh)}{f_\alpha(h)} = z^{-\alpha-1}$$

Uwzględniając powyższe, ostatecznie otrzymujemy

$$\lim_{h \rightarrow \infty} \frac{(zh)^k f_{i;\lambda}(zh)}{h^k f_{i;\lambda}(h)} = z^{k-\alpha(\lambda-i+1)-1} \quad (\text{B.8})$$

co kończy dowód  $\square$ .

## Dodatek C

# Parametr skali w generatorze gęstości

Na wstępie rozważmy ogólny przypadek funkcji gęstości  $f(\mathbf{x}, \sigma)$ , wektora losowego  $\mathbf{X}$  z parametrem skali  $\sigma$ . Zauważmy, że zachodzi równość:

$$f(\mathbf{x}, \sigma) = \frac{1}{\sigma^n} f\left(\frac{\mathbf{x}}{\sigma}, 1\right) \quad (\text{C.1})$$

Aby przekonać się o prawdziwości (C.1) rozważmy rozkład wektora  $\sigma\mathbf{X}$ :

$$P(\sigma\mathbf{X} < \mathbf{e}) = P\left(\mathbf{X} < \frac{1}{\sigma}\mathbf{e}\right) \int_{-\infty}^{\frac{e_n}{\sigma}} \dots \int_{-\infty}^{\frac{e_1}{\sigma}} f(x_1, \dots, x_n) dx_1, \dots, dx_n \quad (\text{C.2})$$

Stosując dobrze znaną metodę zamiany współrzędnych  $y_i = \sigma x_i$  otrzymujemy:

$$P(\sigma\mathbf{X} < \mathbf{e}) = \frac{1}{\sigma^n} \int_{-\infty}^{e_n} \dots \int_{-\infty}^{e_1} f\left(\frac{y_1}{\sigma}, \dots, \frac{y_n}{\sigma}\right) dy_1, \dots, dy_n, \quad (\text{C.3})$$

co ostatecznie potwierdza prawdziwość (C.1). Przyjmijmy zatem, że  $f(\mathbf{x}, 1)$  jest funkcją gęstości wektora losowego o sferycznej symetrii, której odpowiada generator gęstości  $G(t, 1)$  t.j.:

$$f(\mathbf{x}, 1) = cG(\mathbf{x}^T \mathbf{x}, 1), \quad \forall \mathbf{x} \in \mathbb{R}^n, \quad (\text{C.4})$$

gdzie  $c$  jest stałą normalizującą. Wykonując podstawienie  $\mathbf{x} = \frac{\mathbf{y}}{\sigma}$ , oraz mnożąc obie strony równania (C.4) przez  $\sigma^{-n}$ , otrzymujemy:

$$\sigma^{-n} f\left(\frac{\mathbf{y}}{\sigma}, 1\right) = \sigma^{-n} cG\left(\frac{\mathbf{y}^T \mathbf{y}}{\sigma^2}, 1\right), \quad \forall \mathbf{y} \in \mathbb{R}^n, \quad (\text{C.5})$$

co umożliwia sformułowanie następującego wniosku:

$$g(t, \sigma) = \frac{1}{\sigma^n} g\left(\frac{t}{\sigma^2}, 1\right) \quad \forall t \geq 0 \quad (\text{C.6})$$

## Dodatek D

# Rozkład normy izotropowego wektora stabilnego

Posługując się pojęciem generatorem gęstości  $G(\cdot)$ , rozkład zmiennej losowej  $\|\mathbf{X}_\alpha^s\|$  można wyrazić w następujący sposób:

$$P(\|\mathbf{X}_\alpha^s\| \leq R) = \int_{S_R^{(n)}} g(\mathbf{x}) d\mathbf{x} = c \int_{S_R^{(n)}} G(\mathbf{x}^T \mathbf{x}) d\mathbf{x} \quad (\text{D.1})$$

Rozważmy przejście do układu współrzędnych sferycznych poprzez poniższe przekształcenia:

$$x_j = r \left( \prod_{k=1}^{j-1} \sin(\alpha_k) \right) \cos(\alpha_j) \quad \text{for } 1 \leq j \leq n-1 \quad (\text{D.2})$$

$$x_n = r \left( \prod_{k=1}^{n-2} \sin(\alpha_k) \right) \sin(\alpha_{n-1}), \quad (\text{D.3})$$

gdzie  $\alpha_i \in [0, \pi]$  for  $i = 1, 2, \dots, n-2$  oraz  $\alpha_{n-1} \in [0, 2\pi)$ . Jako, że Jakobian powyższego przekształcenia współrzędnych wynosi  $r^{n-1} \prod_{k=1}^{n-2} \sin(\alpha_k)^{n-k-1}$ , wzór (D.1) przyjmuje postać:

$$c \int_0^R \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi r^{n-1} G(r^2) \prod_{k=1}^{n-2} \sin(\alpha_k)^{n-k-1} dr d\alpha_1 d\alpha_2 \dots d\alpha_{n-1} \quad (\text{D.4})$$

Biorąc pod uwagę, że

$$\int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \prod_{k=1}^{n-2} \sin(\alpha_k)^{n-k-1} d\alpha_1 d\alpha_2 \dots d\alpha_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}, \quad (\text{D.5})$$

ostatecznie otrzymujemy

$$P(\|\mathbf{Z}\| \leq R) = c \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^R r^{n-1} G(r^2) dr \quad (\text{D.6})$$



## Dodatek E

# Rozkład zmiennej losowej

$$\|\mathbf{Z} + \mathbf{z}_0\|, \mathbf{Z} \sim \mathbf{X}_\alpha^S$$

Rozważmy izotropowy wektor losowy  $\mathbf{Z}$  oraz pewien element  $\mathbf{z}_0 \in \mathbb{R}^n$

$$P(\|\mathbf{Z} + \mathbf{z}_0\| < e) = \int_{S_{\mathbf{z}_0}^n(e)} f(\mathbf{z}) d\mathbf{z} = \int_{S_{\mathbf{z}_0}^n(e)} G(\mathbf{z}^T \mathbf{z}) d\mathbf{z}, \quad (\text{E.1})$$

gdzie  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  oznacza funkcję gęstości  $\mathbf{Z}$ ,  $S_{\mathbf{z}_0}^n(r) = \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z} - \mathbf{z}_0\| \leq r\}$ , a  $G : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  oznacza generator gęstości. Wprowadźmy następujące liniowe przekształcenie oryginalnego układu współrzędnych  $\mathbf{z}$  według wzoru  $\hat{\mathbf{z}} = H[\mathbf{z} - \mathbf{z}_0]$ , gdzie  $H$  jest ortogonalną macierzą obrotu, dobraną tak, aby  $H\mathbf{z}_0 = \mathbf{e}_n \|\mathbf{z}_0\| = [0, 0, \dots, \|\mathbf{z}_0\|]^T$ . Wówczas (E.1) przybiera postać:

$$P(\|\mathbf{Z} + \mathbf{z}_0\| < e) = \int_{S_{\mathbf{0}}^n(e)} G(\|\hat{\mathbf{z}}\|^2 + 2\hat{z}_n \|\mathbf{z}_0\| + \|\mathbf{z}_0\|) d\hat{\mathbf{z}} \quad (\text{E.2})$$

Całka (E.2) może zostać wyrażona w następującej równoważnej postaci:

$$P(\|\mathbf{Z} + \mathbf{z}_0\|_2 < e) = \int_0^e \left[ \int_{\partial S_{\mathbf{0}}^{n-1}(r)} G(\|\hat{\mathbf{z}}\|^2 + 2\hat{z}_n \|\mathbf{z}_0\| + \|\mathbf{z}_0\|) dS \right] dr, \quad (\text{E.3})$$

gdzie  $\partial S_{\mathbf{0}}^{n-1}(r)$  oznacza powierzchnie  $n$ -wymiarowej kuli o promieniu  $r$ , tj.:  $\partial S_{\mathbf{0}}^{n-1}(r) = \{\hat{\mathbf{z}} \in \mathbb{R}^n : \|\hat{\mathbf{z}}\| = r\}$ . Niech wolno nam będzie zauważyć, że funkcja  $G(\cdot)$  zależy jedynie od  $\hat{z}_n$ , gdyż wszystkie pozostałe składniki funkcji wewnętrznej dla  $G(\cdot)$  przyjmują wartości stałe na powierzchni  $S_{\mathbf{0}}^{n-1}(r)$ . Co więcej, dla każdego ustalonego  $\hat{z}_n \in [-r, r]$  zbiór  $S_{\mathbf{0}}^{n-1}(r)$  formuje  $n-2$ -wymiarową powierzchnię kuli o promieniu  $q = \sqrt{r^2 - \hat{z}_n^2}$  i środku  $\mathbf{c} = [0, 0, \dots, \hat{z}_n]$ . Biorąc pod uwagę, równość ([94] strona 314)

$$dS = (1 - t^2)^{\frac{n-3}{2}} dS' dt \quad (\text{E.4})$$

gdzie  $dS$  i  $dS'$  są elementami powierzchniowymi sfer  $S_{\mathbf{0}}^{n-1}(1)$  i  $S_{\mathbf{0}}^{n-2}(1)$ , a  $t = \hat{z}_n/r$ . Całka (E.3) może być zatem przedstawiona jako:

$$P(\|\mathbf{Z} + \mathbf{z}_0\|_2 < e) = \int_0^e \int_{-1}^1 r^{n-1} (1 - t^2)^{\frac{n-3}{2}} \left[ \int_{\partial S_{\mathbf{0}}^{n-2}(1)} G(r^2 + 2r \|\mathbf{z}_0\| t + \|\mathbf{z}_0\|^2) dS' \right] dt dr \quad (\text{E.5})$$

co daje ostateczną postać

$$P(\|\mathbf{Z} + \mathbf{z}_0\|_2 < e) = \int_0^e \frac{r^{n-1} 2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \cdot \int_{-1}^1 G(r^2 + 2r\|\mathbf{z}_0\|t + \|\mathbf{z}_0\|^2)(1-t^2)^{\frac{n-3}{2}} dt dr. \quad (\text{E.6})$$

## Dodatek F

# Rozkład zmiennej losowej

$$V = \|\mathbf{Z} + \mathbf{z}_0\|^2 / \|\mathbf{z}_0\|^2$$

W pierwszej kolejności zauważmy, że gdy zmienna losowa  $X$  posiada gęstość  $f_X(t)$ , wówczas gęstość  $\frac{X^2}{\|\mathbf{z}_0\|^2}$  może być wyznaczona z zależności:

$$P\left(\frac{X^2}{\|\mathbf{z}_0\|^2} < e\right) = \frac{1}{2} \int_0^e f_X(\sqrt{y}\|\mathbf{z}_0\|) \frac{\|\mathbf{z}_0\|}{\sqrt{y}} dy. \quad (\text{F.1})$$

Przyjmując za  $X = \|\mathbf{Z} + \mathbf{z}_0\|$  o gęstości (E.6) i podstawiając do (F.1) otrzymujemy wzór na gęstość zmiennej losowej  $V = \frac{\|\mathbf{Z} + \mathbf{z}_0\|^2}{\|\mathbf{z}_0\|^2}$ :

$$f_V(v) = \frac{\|\mathbf{z}_0\|^n v^{\frac{n-2}{2}} \pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_{-1}^1 g\left(\|\mathbf{z}_0\|^2 (v + 2\sqrt{v}t + 1)\right) (1-t^2)^{(n-3)/2} dt, \quad (\text{F.2})$$

Dodatkowo uwzględniając parametr skali generatora gęstości  $\sigma$  i zależność (C.6) otrzymujemy:

$$f_V(v, \delta) = \frac{\delta^n v^{\frac{n-2}{2}} \pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_{-1}^1 g\left(\delta^2 (v + 2\sqrt{v}t + 1)\right) (1-t^2)^{(n-3)/2} dt, \quad (\text{F.3})$$

gdzie  $\delta = \frac{\|\mathbf{z}_0\|}{\sigma}$

## Dodatek G

# Rozkład zmiennej losowej $V^{b/2}$

Generalnie, jeśli zmienna losowa  $V \geq 0$  posiada funkcję gęstości  $f(v)$ , wówczas rozkład zmiennej losowej  $V^{b/2}$  równy jest:

$$P(V^{b/2} < e) = P(V < e^{2/b}) = \int_0^{e^{2/b}} f(v) dv \quad (\text{G.1})$$

Wykorzystując przekształcenie  $t = v^{b/2}$ , całka (G.1) przybiera postać:

$$\int_0^{e^{2/b}} f(v) dv = \int_0^e \frac{2}{b} t^{\frac{2}{b}-1} f(t^{2/b}) \quad (\text{G.2})$$

Podstawiając za  $f(\cdot)$  wyrażenie (F.3), otrzymujemy:

$$f_{V^{b/2}}(v; \delta, b) = \frac{2\delta^n v^{\frac{n-b}{b}} \pi^{\frac{n-1}{2}}}{b\Gamma(\frac{n-1}{2})} \int_{-1}^1 g\left(\delta^2(v^{2/b} + 2v^{1/b}t + 1)\right) (1-t^2)^{(n-3)/2} dt, \quad (\text{G.3})$$

dla  $\delta = \frac{\|z_0\|}{\sigma}$ .

## Dodatek H

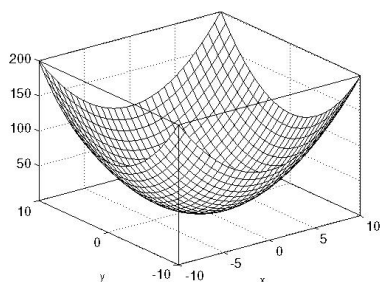
# Funkcje testowe

### Model sferyczny - $\phi_1$

DEFINICJA:

$$\phi_1(\mathbf{x}) = \sum_{i=1}^n x_i^2 \quad (\text{H.1})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_1(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-100, 100] \times [-100, 100] \times \dots \times [-100, 100]$$

*Warunek stopu:*

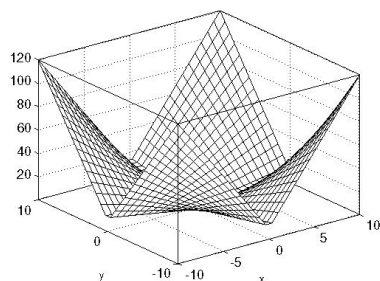
$$\phi_1(\mathbf{x}) < 0.0001$$

### Problem Schwefel'a 2.22 - $\phi_2$

DEFINICJA:

$$\phi_2(\mathbf{x}) = \sum_{i=1}^n |x_i| \prod_{i=1}^n |x_i| \quad (\text{H.2})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_2(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-10, 10] \times [-10, 10] \times \dots \times [-10, 10]$$

*Warunek stopu:*

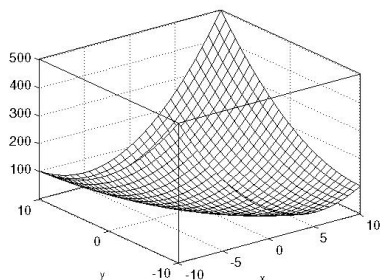
$$\phi_2(\mathbf{x}) < 0.0001$$

### Problem Schwefel'a 1.2 - $\phi_3$

DEFINICJA:

$$\phi_3(\mathbf{x}) = \sum_{i=1}^n \left( \sum_{j=1}^i x_j \right)^2 \quad (\text{H.3})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_3(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-100, 100] \times [-100, 100] \times \dots \times [-100, 100]$$

*Warunek stopu:*

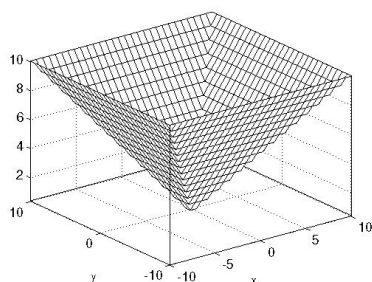
$$\phi_3(\mathbf{x}) < 0.0001$$

### Problem Schwefel'a 2.21 - $\phi_4$

DEFINICJA:

$$\phi_4(\mathbf{x}) = \max_i \{|x_i| | i = 1, \dots, n\} \quad (\text{H.4})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_4(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-100, 100] \times [-100, 100] \times \dots \times [-100, 100]$$

*Warunek stopu:*

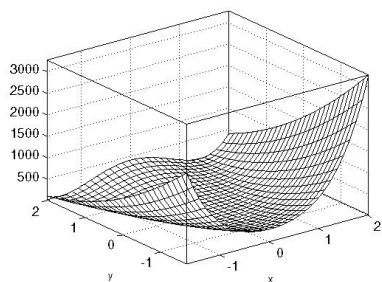
$$\phi_4(\mathbf{x}) < 0.0001$$

### Funkcja Rosenbrock'a - $\phi_5$

DEFINICJA:

$$\phi_5(\mathbf{x}) = \sum_{i=1}^{n-1} \left( 100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2 \right) \quad (\text{H.5})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_5(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [1, 1, \dots, 1]^T$$

*Obszar inicjacji:*

$$[-30, 30] \times [-30, 30] \times \dots \times [-30, 30]$$

*Warunek stopu:*

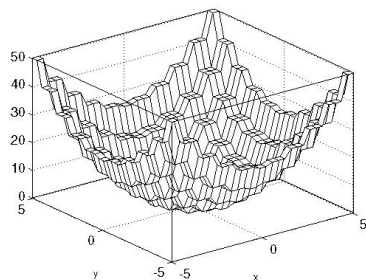
$$\phi_5(\mathbf{x}) < 0.0001$$

### Funkcja schodkowa - $\phi_6$

DEFINICJA:

$$\phi_6(\mathbf{x}) = \sum_{i=1}^n \left( \lfloor x_i + 0.5 \rfloor \right)^2 \quad (\text{H.6})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_6(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-100, 100] \times [-100, 100] \times \dots \times [-100, 100]$$

*Warunek stopu:*

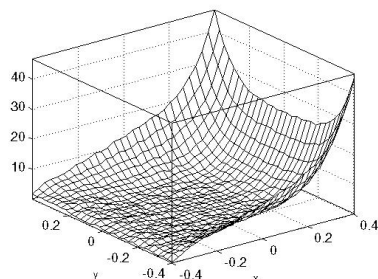
$$\phi_6(\mathbf{x}) < 0.0001$$

### Funkcja kwadratowa z zakłóceniami - $\phi_7$

DEFINICJA:

$$\phi_7(\mathbf{x}) = \sum_{i=1}^n ix_i^4 + \text{random}[0, 1] \quad (\text{H.7})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_7(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-1.28, 1.28] \times [-1.28, 1.28] \times \dots \times [-1.28, 1.28]$$

*Warunek stopu:*

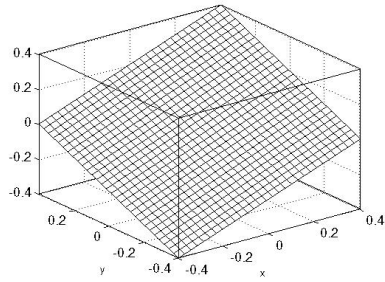
$$\phi_7(\mathbf{x}) < 0.0001$$

### Diagonal plane - $\phi_8$

DEFINICJA:

$$\phi_8(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{H.8})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

nie istnieje

*Obszar inicjacji:*

$$[0.5, 1.5] \times [0.5, 1.5] \times \dots \times [0.5, 1.5]$$

*Warunek stopu:*

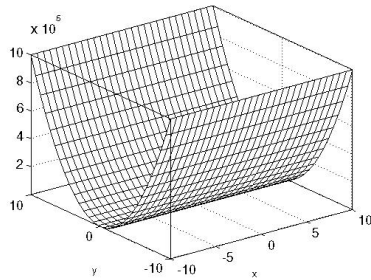
$$\phi_8(\mathbf{x}) < 0.0001$$

### Funkcja elipsoidalna - $\phi_9$

DEFINICJA:

$$\phi_9(\mathbf{x}) = \sum_{i=1}^n \left( 100^{\frac{i-1}{n-1}} x_i \right)^2 \quad (\text{H.9})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_9(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[0.5, 1.5] \times [0.5, 1.5] \times \dots \times [0.5, 1.5]$$

*Warunek stopu:*

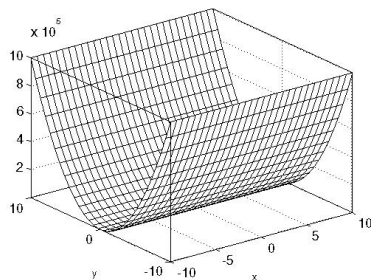
$$\phi_9(\mathbf{x}) < 0.0001$$

### Cigar function - $\phi_{10}$

DEFINICJA:

$$\phi_{10}(\mathbf{x}) = x_1^2 + 10^4 \sum_{i=2}^n x_i^2 \quad (\text{H.10})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{10}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-3, 7] \times [-3, 7] \times \dots \times [-3, 7]$$

*Warunek stopu:*

$$\phi_{10}(\mathbf{x}) < 0.0001$$

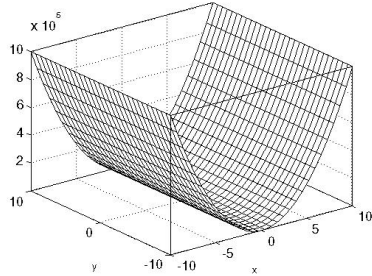
### Tablet function - $\phi_{11}$

DEFINICJA:

$$\phi_{11}(\mathbf{x}) = 10^4 x_1^2 + \sum_{i=2}^n x_i^2 \quad (\text{H.11})$$



CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{11}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-3, 7] \times [-3, 7] \times \dots \times [-3, 7]$$

*Warunek stopu:*

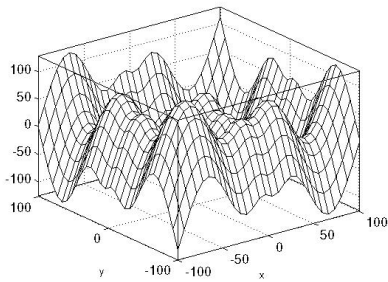
$$\phi_{11}(\mathbf{x}) < 0.0001$$

### Problem Schwefel'a 2.26 - $\phi_{12}$

DEFINICJA:

$$\phi_{12}(\mathbf{x}) = - \sum_{i=1}^n \left( x_i \sin(\sqrt{|x_i|}) \right) \quad (\text{H.12})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

nie istnieje

*Obszar inicjacji:*

$$[-500, 500] \times [-500, 500] \times \dots \times [-500, 500]$$

*Warunek stopu:*

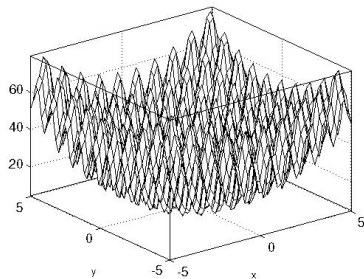
$$\phi_{12}(\mathbf{x}) < -10E10$$

### Funkcja Rastringin'a - $\phi_{13}$

DEFINICJA:

$$\phi_{13}(\mathbf{x}) = 10n + \sum_{i=1}^n \left( x_i^2 - 10 \cos(2\pi x_i) \right) \quad (\text{H.13})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{13}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-5.12, 5.12] \times [-5.12, 5.12] \times \dots \times [-5.12, 5.12]$$

*Warunek stopu:*

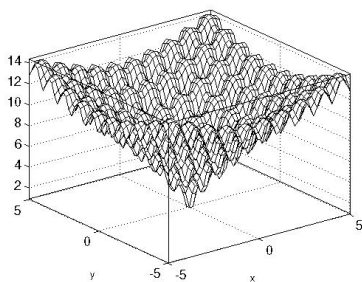
$$\phi_{13}(\mathbf{x}) < 0.0001$$

### Funkcja Ackley'a - $\phi_{14}$

DEFINICJA:

$$\phi_{14}(\mathbf{x}) = -20 \exp\left(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e \quad (\text{H.14})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{14}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-32, 32] \times [-32, 32] \times \dots \times [-32, 32]$$

*Warunek stopu:*

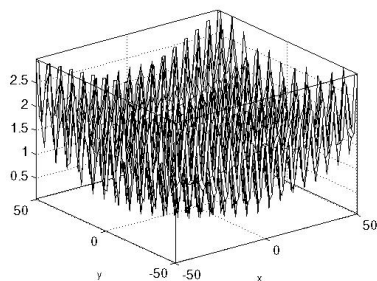
$$\phi_{14}(\mathbf{x}) < 0.0001$$

### Funkcja Griewank'a - $\phi_{15}$

DEFINICJA:

$$\phi_{15}(\mathbf{x}) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \quad (\text{H.15})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{15}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [0, 0, \dots, 0]^T$$

*Obszar inicjacji:*

$$[-600, 600] \times [-600, 600] \times \dots \times [-600, 600]$$

*Warunek stopu:*

$$\phi_{15}(\mathbf{x}) < 0.0001$$

### Generalised Penalised Function 1 - $\phi_{16}$

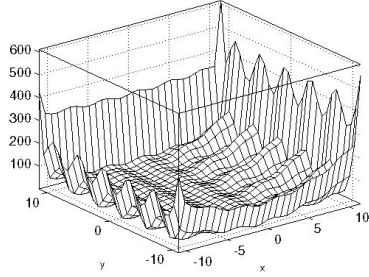
DEFINICJA:

$$\begin{aligned} \phi_{16}(\mathbf{x}) = & \frac{\pi}{30} \left( 10 \sin^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 (1 + 10 \sin^2(\pi y_{i+1})) \right) \\ & + (y_n - 1)^2 + \sum_{i=1}^n u(x_i, 10, 100, 4) \end{aligned} \quad (\text{H.16})$$

gdzie  $y_i = 1 + \frac{1}{4}(x_i + 1)$  oraz

$$u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, & x_i > a \\ 0, & -a \leq x_i \leq a \\ k(-x_i - a)^m, & x_i < -a \end{cases}$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{16}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [1, 1, \dots, 1]^T$$

*Obszar inicjacji:*

$$[-50, 50] \times [-50, 50] \times \dots \times [-50, 50]$$

*Warunek stopu:*

$$\phi_{16}(\mathbf{x}) < 0.0001$$

### Generalised Penalised Function 2 - $\phi_{17}$

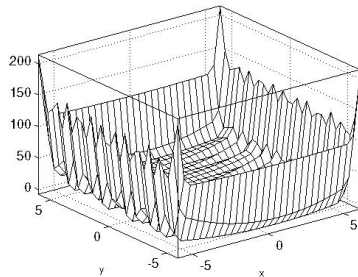
DEFINICJA:

$$\begin{aligned} \phi_{17}(\mathbf{x}) = & 0.1 \left( \sin^2(\pi 3x_1) + \sum_{i=1}^{n-1} (x_i - 1)^2 (1 + \sin^2(3\pi x_{i+1})) \right. \\ & \left. + (x_n - 1)(1 + \sin^2(2\pi x_n)) \right) + \sum_{i=1}^n u(x_i, 5, 100, 4) \end{aligned} \quad (\text{H.17})$$

gdzie

$$u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, & x_i > a \\ 0, & -a \leq x_i \leq a \\ k(-x_i - a)^m, & x_i < -a \end{cases}$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

$$\phi_{17}(\mathbf{x}^*) = 0, \quad \mathbf{x}^* = [1, 1, \dots, 1]^T$$

*Obszar inicjacji:*

$$[-50, 50] \times [-50, 50] \times \dots \times [-50, 50]$$

*Warunek stopu:*

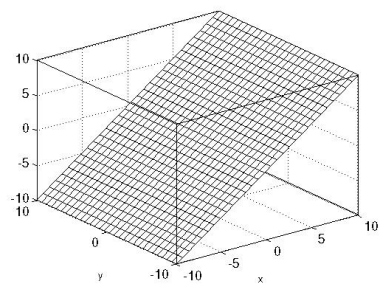
$$\phi_{17}(\mathbf{x}) < 0.0001$$

### Plane function - $\phi_{18}$

DEFINICJA:

$$\phi_{18}(\mathbf{x}) = x_1 \quad (\text{H.18})$$

CHARAKTERYSTYKA ZDANIA:



*Minimum globalne:*

nie istnieje

*Obszar inicjacji:*

$$[0.5, 1.5] \times [0.5, 1.5] \times \dots \times [0.5, 1.5]$$

*Warunek stopu:*

$$\phi_{18}(\mathbf{x}) < -10E10$$