

Instytut Informatyki, Automatyki i Robotyki

Seria: PREPRINTY nr 03/2010

**Wspomaganie procesów
optymalizacyjnych za pomocą sieci
neuronowych z radialnymi funkcjami
aktywacji**

(rozprawa doktorska)

Marek Bazan

Promotor: dr hab. inż. Ewa Skubalska-Rafajłowicz
prof. PWr

Słowa kluczowe:

- optymalizacja lokalna;
- przyspieszanie optymalizacji lokalnej;
- sieci RBF;
- regularyzacja Tikhonova.

Wspomaganie procesów optymalizacyjnych za
pomocą sieci neuronowych z radialnymi funkcjami
aktywacji

(rozprawa doktorska)

Marek Bazan

promotor: dr hab. inż. Ewa Skubalska-Rafajłowicz,
prof. PWr

Instytut Informatyki, Automatyki i Robotyki,
Wydział Elektroniki,
Politechnika Wrocławska

Styczeń 2010



346588 4/1

Alleluja.

Chwalcie Pana z niebios,
chwalcie Go na wysokościach!
Chwalcie Go, wszyscy Jego aniołowie,
chwalcie Go, wszystkie Jego zastępy!
Chwalcie Go, słońce i księżycu,
chwalcie Go, wszystkie gwiazdy świecące.
Chwalcie Go, nieba najwyższe
i wody, co są ponad niebem:
niech imię Pana wychwalają,
On bowiem nakazał i zostały stworzone,
utwierdził je na zawsze, na wieki;
nadał im prawo, które nie przeminie.
Chwalcie Pana z ziemi,
potwory i wszystkie morskie głębiny,
ogniu i gradzie, śniegu i mgło,
gwałtowny huraganie, co pełnisz Jego słowo,
góry i wszelkie pagórki,
drzewa rodzące owoc i wszystkie cedry,
dzikie zwierzęta i bydło wszelakie,
to, co się roi na ziemi, i ptactwo skrzydlate,
królowie ziemscy i wszystkie narody,
władcy i wszyscy sędziowie na ziemi,
młodzieńcy, a także dziewice,
starcy wraz z młodzieżą
niech imię Pana wychwalają,
bo tylko Jego imię jest wzniosłe,
majestat Jego góruje nad ziemią i niebem
i pomnaża moc swojego ludu.
Pieśń pochwalana dla wszystkich Jego świętych,
synów Izraela - ludu, który Mu jest bliski.
Alleluja.

Psalm 148

Mojej ukochanej żonie i dzieciom
oraz moim rodzicom
w podziękowaniu za wszystko co mi dali

Podziękowania

Przedkładając niniejszą pracę jako rozprawę doktorską chciałbym podziękować wszystkim, którzy przyczynili się do jej powstania. Przede wszystkim chciałbym złożyć podziękowania na ręce pani Profesor Ewy Skubalskiej-Rafajłowicz – mojej promotor. Dziękuję pani Profesor przede wszystkim za życzliwość i cierpliwość oraz za to, że pokierowała moją pracą naukową w ten sposób, że doprowadziła ona do powstania niniejszej rozprawy. Dziękuję za wszystko czego nauczyłem się od Pani Profesor z automatyki i robotyki oraz sieci neuronowych jak również z metodologii pracy dydaktycznej ze studentami. Dziękuję również panu Profesorowi Ewarystowi Rafajłowiczowi za wiele bardzo cennych uwag dotyczących tej pracy.

Dziękuję również wszystkim kolegom z Zakładu Sterowania Jakością Procesów Wytwórczych w Instytucie Informatyki, Automatyki i Robotyki Politechniki Wrocławskiej, z którymi dyskusje pomogły mi nadać tej pracy obecną formę. Dziękuję również Fundacji na rzecz Nauki Polskiej, która częściowo finansowała badania zawarte w niniejszej pracy.

Chciałbym również podziękować panu Profesorowi Stephanowi Russenschukowi z Europejskiego Laboratorium Fizyki Nuklearnej (CERN) z działu LHC/ICP, który był moim promotorem podczas mojego dwuletniego stażu doktoranckiego w CERNie. Dziękuję mu za wszystko czego nauczyłem się od niego z fizyki akceleratorów oraz na temat optymalizacji magnesów nadprzewodzących oraz za umożliwienie mi zastosowania tej wiedzy do optymalizacji głównego magnesu zakrzywiającego zderzacza cząstek elementarnych LHC (Large Hadron Collider) w CERNie za pomocą zaprezentowanej w tej pracy metody. Dziękuję również Doktorowi Martinowi Aleksie z CERNu z działu LHC/ICP za dyskusje w czasie mojego pobytu na stażu doktoranckim w CERNie, które zaowocowały pomysłem wprowadzonego w tej pracy wskaźnika mierzącego jakość otoczenia punktu przez dane.

Dziękuję również panu prof. Stanisławowi Lewanowiczowi; pani prof. Krystynie Ziętak oraz prof. Annie Bartkowiak z Instytutu Informatyki Uniwersytetu Wrocławskiego za wszystko czego nauczyłem się od nich z metod numerycznych, statystyki oraz sieci neuronowych oraz za umożliwienie mi odbycia stażu doktoranckiego w CERNie.

Dziękuję również panu Doktorowi Joachimowi Tückmantelowi oraz panu Doktorowi Ernstowi Haeblowi z działu SPS/LEP/RF w CERNie – moim opiekunom na stażu studenckim w CERNie – za wszystko czego nauczyłem się od nich z elektromagnetyzmu oraz za umożliwienie mi wzięcia udziału w pracach związanych ze zderzaczem cząstek elementarnych LHC.

Dziękuję również moim szefom w firmie Microgen panu Doktorowi Neilowi Thompsonowi, panu Doktorowi Grzegorzowi Puszwowi oraz mojemu bezpośredniemu szefowi Doktorowi Piotrowi Kozaczewskiemu z grupy Aptitude/IO za tworzenie

firmy, w której ze względu na bardzo dobrą atmosferę oraz na elastyczność czasu pracy mam możliwość pracy dydaktycznej na Politechnice Wrocławskiej.

Największe podziękowania chciałbym złożyć Panu Bogu i mojej rodzinie. Panu Bogu dziękuję za wiarę i za to, że dał mi wspaniałą rodzinę oraz za to, że postawił na mojej drodze takich nauczycieli i wychowawców jak również za to, że pozwolił mi odkryć Regułę Św. Benedykta w najtrudniejszym momencie pracy naukowej.

Moim Rodzicom dziękuję za dar życia oraz za dobre wychowanie. Mojej Żonie dziękuję za naszą miłość i małżeństwo, którego owocem są wspaniałe dzieci, jak również, za motywowanie mnie do pracy naukowej.

Marek Bazan
Wrocław, styczeń 2010 r.

Spis treści

1	Wstęp	13
2	Optymalizacja	19
2.1	Metoda Zannilla	20
2.2	Główne twierdzenie	22
2.3	Domknięcie przekształcenia algorytmicznego M^*	23
2.4	Zaburzenie w algorytmie minimalizacji kierunkowej	25
2.5	Procedura ε -check	38
3	Przybliżanie funkcji ciągłych	41
3.1	Interpolacja oraz aproksymacja za pomocą regularyzacji Tikhonova	41
3.2	Radialne funkcje bazowe	42
3.3	Przestrzeń natywna	44
3.3.1	Przestrzeń Hilberta z samoreprodukującą funkcją jądrową	44
3.3.2	Przestrzeń natywna	45
3.3.3	Przestrzeń Sobolewa	46
3.3.4	Przestrzeń natywna dla funkcji Gaussa	47
3.4	Regularyzacja Tikhonova	47
3.4.1	Metoda uogólnionej walidacji krzyżowej	49
3.4.2	Metoda L-krzywej	50
3.4.3	Metoda ważonej wariancji gradientu	52
3.5	Stabilna metoda wyznaczania przybliżenia	55
3.6	Ograniczenia błędu przybliżenia	57
3.6.1	Wskaźniki gęstości danych	58
	Norma siatki	58
	Wskaźnik otoczenia $\gamma_{\mathbf{x}}(\mathbf{x})$	59
3.6.2	Odtwarzanie wielomianów	60
3.6.3	Funkcjonał potęgowy	63
3.6.4	Funkcja Lebesgue’a	64
3.6.5	Ograniczenia błędu z założeniem o gęstości zbioru danych	66
	Ograniczenia błędu wykorzystujące funkcjonal potęgowy	66
	Ograniczenia błędu dla funkcji z przestrzeni Sobolewa	68
	Ograniczenia błędu dla zagadnienia aproksymacji	71
3.6.6	Ograniczenie błędu bez założenia o gęstości zbioru danych	72
3.6.7	Ograniczenie błędu dla metody wariancji ważonej	73

4	Zastosowania metody SPELROA	81
4.1	Optymalizacja funkcji testowych	81
4.1.1	Wyniki numeryczne	84
	Funkcja kary II czterech zmiennych	84
	Rozszerzona funkcja Rosenbrocka sześciu zmiennych	86
	Rozszerzona funkcja Rosenbrocka ośmiu zmiennych	88
	Funkcja „Chebquad” ośmiu zmiennych	91
	Pasmowa funkcja Broydena dziesięciu zmiennych	93
	Dyskretna funkcja brzegowa dziesięciu zmiennych	95
	Funkcja kary I dziesięciu zmiennych	97
	Funkcja Osborne’a II jedenastu zmiennych	99
	Funkcja dyskretnego równania całkowego piętnastu zmiennych	101
4.1.2	Synteza wyników numerycznych dla funkcji testowych	103
4.2	Projektowanie magnesów nadprzewodzących	105
4.2.1	Optymalizacja elektromagnesu LHC	105
4.2.2	Określenie funkcji celu	105
	Rozwinięcie pola magnetycznego w szereg Fouriera w aperturze magnesu	105
	Składowa pola generowana przez cewkę nadprzewodzącą	106
	Składowa pola generowana przez jarzmo magnesu	106
	Optymalizacja wielocelowa – optymalność w sensie Pareto	107
	Funkcja celu w optymalizacji magnesów zakrzywiających zderzacza cząstek elementarnych LHC	109
4.2.3	Wyniki numeryczne	110
	LHC coil test facility, przykład optymalizacji funkcji dwóch zmiennych	110
	LHC Main Dipole, przykład optymalizacji funkcji trzech i pięciu zmiennych	112
4.3	Strojenie regulatora PID - BFGS quasi Newton	114
5	Zakończenie	119
5.1	Podsumowanie	119
5.2	Proponowane dalsze kierunki badań	120
6	Dodatki	121
6.1	Dodatek A	121
6.2	Dodatek B	122
6.3	Dodatek C	128
6.3.1	Wyrażenie $\Delta_l(\varepsilon; \zeta)$	128
6.3.2	Główne warunki	129

Spis rysunków

3.1	(a) Warunek Picarda dla równania całkowego Fredholma [17], (b) Warunek Picarda dla aproksymacji bazowymi funkcjami radialnymi funkcji Rosenbrocka dwóch zmiennych.	49
3.2	Wykres wskaźnika $GCV(\lambda)$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.	51
3.3	L-krzywa dla macierzy interpolacyjnej dla funkcji Gaussa dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.	52
3.4	(a) Zbiór 30 punktów otrzymanych ze ścieżki optymalizacyjnej dla funkcji Rosenbrocka dwóch zmiennych, (b) Lokalne odtwarzanie zbioru treningowego w otoczeniu punktu A oraz punktu B.	54
3.5	$WGV_{\lambda, \mathbf{z}}(\mathbf{x})$ dla punktu A oraz punktu B z rysunku 3.4a). Symbolem * na wykresie oznaczona jest wartość λ wskazane przez metodę, dla $NLMSE_{\lambda, \mathbf{z}}(\mathbf{x}) \leq 10^{-4}$	54
3.6	a) Błąd aproksymacji dla λ wybranego przez wskaźnik WGV dla $NLMSE = 5.0 \cdot 10^{-6}$, b) Wybrana wartość λ – jak widać w regionie uboższym w dane wskaźnik sugeruje większą wartość parametru λ . . .	55
3.7	a) Błąd aproksymacji dla $\lambda \approx 10^{-13}$ wybranego przez wskaźnik GCV, b) Błąd aproksymacji dla $\lambda \approx 10^{-8.5}$ wybranego za pomocą L-krzywej. . .	55
3.8	$h_{\rho, \mathbf{x}, \Omega}(\mathbf{x})$ dla $\rho = 10$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.	59
3.9	Wykres funkcji a_{ij} . Maksymalna wartość przyjmowana jest na odcinku $\mathbf{x}_i, \mathbf{x}_j$. Wykres funkcji wagowej W_{ij} jest analogiczny. Maksymalna wartość funkcji wagowej również jest przyjmowana na odcinku $\mathbf{x}_i, \mathbf{x}_j$ i wynosi $1/d_{ij}$	60
3.10	Wykres wskaźnika $\gamma_{\mathbf{x}}(\mathbf{x})$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.	60
3.11	Zależność wskaźnika h w (3.30) oraz h_0 w (3.33) od wymaganego stopnia odtwarzanych wielomianów w kuli jednostkowej.	63

- 3.12 Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Rosenbrocka ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej (GCV), b) Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ 74
- 3.13 Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Rosenbrocka ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$ 75
- 3.14 Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji „Chebquad” ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej, b) Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ 76
- 3.15 Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji „Chebquad” ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$ 77
- 3.16 Błąd aproksymacji w punktach w początkowych tysiącu punktów ścieżki optymalizacyjnej funkcji Osborne 2 jedenastu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej, b) Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ 78
- 3.17 Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Osborne 2 jedenastu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$ 79
- 4.1 Model Coil Test Facility. Optymalizowane są rozmiary półośi elipsy określającej kształt kołnierza magnesu. 111

- 4.2 Ścieżka optymalizacyjna dla metody SPELROA połączonej z algorytmem EXTREM dla modelu CTF (c.f. [2]). Proces wystartowany został z punktu $[130.0, 80.0]$. Algorytm „zbiegł” do minimum funkcji celu znajdującego się w punkcie $[68.55, 75.87]$ po 111 bezpośrednich wartościowaniach funkcji i konstrukcji aproksymacji RBF w 31 punktach (kółka z krzyżykiem). Algorytm EXTREM potrzebował 167 wartościowań funkcji. Uzyskano więc przyspieszenie rzędu 30%. 112
- 4.3 Parametryczna reprezentacja optymalizowanego jarzma magnesu LHC. 114
- 6.1 Obszar w przestrzeni $(\Delta', \zeta_{(r)})$, w którym dobrze określona est wielkość $K_{(a),i}$ zdefiniowany w (6.2). 123
- 6.2 W regionie $\mathbf{P}_1$ wyrażenie (6.5) jest większe od zera. W regionie $\mathbf{P}_2$ wyrażenie (6.5) jest mniejsze od zera. W regionie $\mathbf{P}_3$ wyrażenie (6.5) nie jest określone. 125
- 6.3 Wyrażenie (6.6) w obszarze $\mathbf{P}_1$ jest ujemne, natomiast w obszarze $\mathbf{P}_2$ jest dodatnie. 126
- 6.4 W regionie $\mathbf{P}_1$ wyrażenie (6.7) jest mniejsze od zera. W regionie $\mathbf{P}_2$ wyrażenie (6.7) jest większe od zera. W regionie $\mathbf{P}_3$ wyrażenie (6.7) nie jest określone. 128
- 6.5 Ograniczenie na ε dla $(\zeta_{(r)}, \lambda^*) \in (0, 1) \times (0, 0.5)$ (zauważmy, że $\zeta_{(r)} > \lambda^*$) a) dla $l = 1$ określone przez warunki (6.12) i (6.16) b) dla $l = 2$ określone przez warunki (6.14) i (6.17) (zauważmy, że dla $\zeta_{(r)} = 1/2$ mamy osobliwość) c) dla $l = 3$ określone przez warunki (6.15) i (6.18). 132

Oznaczenia

d	– wymiar przestrzeni danych,
N	– liczba danych w zbiorze treningowym,
$\mathbf{X}$	– zbiór danych w przestrzeni parametrów,
f	– funkcja celu,
$\mathbf{Z}$	– zbiór danych w przestrzeni parametrów i przestrzeni funkcji celu,
ε	– dokładność aproksymacji,
I_s	– liczba danych w fazie początkowej metody SPELROA,
$\mathbf{x}_k$	– punkt w przestrzeni danych w k -tej iteracji metody SPELROA,
$A(\cdot)$	– przekształcenie algorytmiczne,
$\tilde{A}^l(\varepsilon; \zeta)$	– przekształcenie dla danych zaburzonych w punkcie l w algorytmie aproksymacji parabolicznej,
V	– przestrzeń parametrów optymalizacji w teorii Zangwilla,
u_i	– i -ty lewy wektor szczególny macierzy A ,
σ_i	– i -ta wartość szczególna macierzy A ,
D	– Operator różniczkowania kierunkowego,
$D^i(\cdot)$	– przekształcenie wyboru i -tego kierunku z bazy kierunków ortogonalnych (sprzężonych),
$M^*(\cdot), M^1(\cdot)$	– przekształcenia minimalizacji wzdłuż kierunku,
$\mathbf{T}$	– zbiór osiągalnych trójek w kolejnym kroku algorytmu optymalizacji parabolicznej,
$\mathbf{T}^l$	– zbiór osiągalnych trójek w kolejnym kroku algorytmu optymalizacji parabolicznej zaburzonego w punkcie $l = \{1, 2, 3\}$,
$\zeta = (\zeta^1, \zeta^2, \zeta^3)$	– trójka punktów z których konstruowana jest parabola w kolejnych krokach algorytmu aproksymacji parabolicznej,
$\lambda^*(\zeta)$	– minimum paraboli skonstruowanej dla trójki ζ ,
$\tilde{\lambda}_l^*(\varepsilon, \zeta)$	– minimum paraboli skonstruowanej dla trójki ζ zaburzonej w punkcie $l = \{1, 2, 3\}$ błędem względnym na poziomie ε
$\Delta_l(\varepsilon; \zeta)$	– odległość minimum paraboli zaburzonej od minimum paraboli dla danych niezaburzonych w algorytmie aproksymacji parabolicznej
$u_n(\zeta) \ n = (1, \dots, 4)$	– trójka w następnym kroku algorytmu aproksymacji parabolicznej,
$c(\cdot)$	– funkcja celu w algorytmie aproksymacji parabolicznej,
$\hat{q}$	– transformacja rodziny parabol w jedną parabolę,
A, B, C	– wyrażenia w definicji transformacji $\hat{q}$,
$\mathbf{S}$	– zbiór punktów krytycznych pierwszego rzędu funkcji f ,
$h_{\mathbf{x}, \Omega}$	– norma siatki,
$\gamma_{\mathbf{x}}(\mathbf{x})$	– wskaźnik mierzący jakość otoczenia punktu $\mathbf{x}$ przez punkty ze zbioru $\mathbf{X}$,
ϵ	– błąd reprezentacji zmiennopozycyjnej,
D^α	– operator różniczkowania stopnia α ,
$\mathcal{N}_\phi$	– przestrzeń natywna generowania przez funkcję ϕ .

Rozdział 1

Wstęp

Wprowadzenie

Zadanie optymalizacji, w którym obliczenie funkcji celu jest bardzo czasochłonne pojawia się w wielu zagadnieniach projektowania oraz optymalizacji różnego rodzaju procesów w różnych dziedzinach techniki. Obliczanie wartości funkcji w takim zadaniu wiąże się z zazwyczaj z przeprowadzeniem złożonej i czasochłonnej operacji w zależności od zmiennych projektowych.

W dziedzinie automatyki i robotyki metody optymalizacji stanowią jedno z podstawowych narzędzi projektowania układów regulacji i konstrukcji algorytmów sterowania. Z optymalizacją układów regulacji mamy do czynienia w zadaniach

- sterowania optymalnego [16], [47],
- sterowania ekstremalnego [45],
- sterowania predykcyjnego [37].

W sterowaniu procesami wytwórczymi z przykładami funkcji, których obliczenie jest czasochłonne spotykamy się np. w sterowaniu *run-to-run*. Procesy sterowane *run-to-run* odgrywają istotną rolę w produkcji związków chemicznych czy w produkcji półprzewodników [36]. Typowym przykładem procesu sterowanego *run-to-run* jest modyfikacja czasu reakcji stechiometrią produkowanego związku lub temperaturą reaktora chemicznego. Podejście *run-to-run* motywowane jest brakiem bezpośredniego dostępu do sygnałów wpływających na jakość produktu w trakcie wykonywania danego podprocesu. Produkt może być zanalizowany pod względem jakości dopiero po wykonaniu całego długotrwałego podprocesu. Zadaniem sterownika *run-to-run* jest więc zmodyfikowanie parametrów schematu produkcji tak, aby minimalizować odchylenie produktu od wstępnej specyfikacji w kolejnych wykonaniach podprocesu.

Z optymalizacją parametrów sterowania procesów wytwórczych spotykamy się również w praktyce systemów procesów multiskalowych. Systemy procesów multiskalowych są to systemy, które łączą kilka silnie powiązanych ze sobą procesów sterowania w przestrzeniach stanów o różnych skalach. Przykładem z przemysłu

chemicznego może być nakładanie warstw związków chemicznych na powierzchnię metalu. Modelowanie takiego systemu wymaga sterowania fazą gazową w skali makro jak i rozpatrzenie fazy mikro jaką są zjawiska zachodzące na powierzchni metalu. W skali mikro nie jest zazwyczaj możliwe określenie modelu ciągłego, analogicznego do modelu ciągłego, jaki występuje w skali makro. Optymalizacja w skali mikro wymaga przybliżonego rozwiązania równania różniczkowego za pomocą metody Monte-Carlo. Typowym problemem sterowania w tym zagadnieniu jest regulacja szybkości wzrostu warstwy związku na powierzchni metalu za pomocą składu gazu oraz temperatury substratów [10]. Różne metody optymalizacji parametrów sterowania procesami chemicznymi w skali całej fabryki produkcyjnej na przykładzie procesów destylacji jest omawiana w monografii [35].

Motywacja do podjęcia badań

Z przykładami funkcji celu, których obliczenie wymaga rozwiązania równań Maxwella metodą elementów skończonych lub rozwiązania dużego zagadnienia własnego [8] dla obiektów dyskretyzowanych metodą różnic skończonych dla bardzo gęstych siatek, autor zetknął się w Europejskim Laboratorium Fizyki Cząstek CERN w procesie projektowania głównego magnesu zakrzywiającego dla zderzacza cząstek elementarnych LHC (Large Hadron Collider) [4] oraz w procesie projektowania wnęk przyspieszających dla akceleratora SPS (Super Proton Synchrotron) [6], [7] odpowiednio. Główną motywacją do podjęcia próby skonstruowania metody eksploracji funkcji celu wielu zmiennych był czas obliczania funkcji celu w procesie projektowania głównego magnesu zakrzywiającego dla zderzacza cząstek elementarnych LHC (patrz rozdział 4.2). Obliczenie funkcji celu dla procesu projektowania elektromagnesów nadprzewodzących umożliwia program ROXIE (Routine for the Optimization of magnet X-section, Inverse field calculation and coil End design) stworzony w CERNie przez profesora S. Russenschucka [28]. Funkcja celu w procesie projektowania elektromagnesów dla akceleratorów opisuje jakość pola magnetycznego w aperturze, w której poruszają się wiązka cząstek elementarnych, podczas procesu pobudzania magnesu począwszy od poziomu natężenia pola przy wstrzyknięciu wiązki do akceleratora do natężenia nominalnego. Funkcja celu zawiera również składową opisującą tzw. prądy persystentne powstałe w jarzmie magnesu na skutek pobudzania prądem [4]. Obliczenie wartości funkcji celu w procesie projektowania elektromagnesów nadprzewodzących może trwać nawet do 2h na komputerze DEC Alpha XP 1000. Mając na uwadze to, że proces optymalizacji elektro-magnesu nadprzewodzącego może wymagać obliczenia wartości funkcji celu dla nawet kilkuset zestawów wartości parametrów projektowych, wartościowe wydały się poszukiwania metody na skrócenie pełnego czasu procesu optymalizacji. Od początku tych poszukiwań autor rozważał tylko podejście ogólne - tzn. poszukiwał takiej metody, która nie będzie wykorzystywała żadnych specyficznych właściwości problemu lecz pozwoli przyspieszyć dowolny proces optymalizacyjny wykorzystujący jakiś algorytm optymalizacji lokalnej przy możliwie małych dodatkowych założeniach o funkcji

celu. Na gruncie automatyki i robotyki, metoda taka pozwoliłaby przyspieszyć zaawansowane warstwowe algorytmy sterowania procesami produkcyjnymi w skali całej linii technologicznej czy sterujących procesami zarządzania produkcją w skali fabryki. Algorytmy sterowania o strukturze warstwowej, w których występuje faza optymalizacji, która zależy nie tylko od sygnałów wejściowych i sterujących ale może zależeć również od innych parametrów procesu omawiane są szczegółowo np. w monografii [37].

Do optymalizacji funkcji celu, których obliczenie jest czasochłonne zazwyczaj stosuje się algorytmy bezgradientowe ze względu na to, że gradient nie jest dostępny a aproksymacja gradientu jest zbyt czasochłonna lub niemożliwa.

Reasumując można powiedzieć, że każda zbieżna metoda, które pozwoli przyspieszyć proces optymalizacji jest pożądana. Przyspieszenie procesu optymalizacji rozumiemy tutaj jako skrócenie czasu działania algorytmu optymalizacyjnego, tzn. zmniejszenie liczby obliczeń bezpośrednich wartości funkcji.

Najpowszechniejszymi metodami przyspieszania algorytmów bezgradientowych dla funkcji, dla których obliczenie wartości jest czasochłonne, są metody regionu wiarygodności [11]. Metoda regionu wiarygodności polegają na konstrukcji modelu aproksymującego funkcję korzystając z wartości funkcji w punktach, w których funkcja została obliczona w sposób bezpośredni, w trakcie optymalizacji funkcji przez algorytm podstawowy. Zakłada się, że skonstruowany model dostatecznie wiarygodnie aproksymuje funkcję celu w najbliższym otoczeniu punktu ostatniego bezpośredniego wartościowania funkcji celu. Skonstruowany model jest następnie optymalizowany jedną z metod gradientowych w regionie wiarygodności. Jeśli po zakończeniu tej optymalizacji spełniony jest warunek dostatecznego zmniejszenia wartości funkcji to znalezione minimum jest punktem startowym do dalszej optymalizacji algorytmem podstawowym. W tym wypadku może nastąpić również zwiększenie promienia regionu wiarygodności. Jeśli natomiast nie jest spełniony warunek dostatecznego zmniejszenia wartości funkcji to następuje zmniejszenie promienia regionu wiarygodności i optymalizacja modelu algorytmem gradientowym jest powtarzana.

W niniejszej rozprawie prezentujemy metodę służącą do przyspieszania procesów optymalizacji funkcji ciągłych, która może być alternatywą dla metod regionu wiarygodności [11]. Prezentowana metoda, ogólnie rzecz biorąc, polega na konstrukcji modelu aproksymującego korzystając z wartości funkcji w części punktów ścieżki optymalizacyjnej algorytmu podstawowego w otoczeniu punktu, w którym wartość funkcji celu ma być obliczona w następnym kroku i na nieobliczaniu wartości funkcji w sposób bezpośredni lecz, jeśli to możliwe, zastąpieniu jej skonstruowanym modelem. Aproksymacja funkcji celu jest więc wiarygodna w swoistym oknie, które przesuwają się w kierunku minimum funkcji celu wraz z kolejnymi krokami algorytmu podstawowego. W zaproponowanej metodzie modelem aproksymującym jest sieć z radialnymi funkcjami aktywacji trenowana za pomocą metody regularyzacji Tikhonova, stąd nazwana została Search Procedure Exploiting Locally Regularized Objective Approximation (SPELROA). Istotnym założeniem

zapropionowanej metody jest, że czas konstrukcji aproksymacji oraz obliczenia jej wartości w danym punkcie jest zaniebbywalnie mały w porównaniu z czasem potrzebnym na obliczenie funkcji celu w sposób bezpośredni. Omawiana metoda po raz pierwszy została zaprezentowana w pracy [2], natomiast w takiej formie w jakiej jest rozważana w niniejszej rozprawie została ona opisana w pracy [3] oraz [5] a dowód jej zbieżności został podany w pracy [1].

Charakterystyka problemu

Podsumowując stwierdzić możemy, że

- w wielu praktycznych interesujących zadaniach optymalizacyjnych obliczenie funkcji celu jest bardzo czasochłonne,
- do optymalizacji takich funkcji zazwyczaj stosuje się algorytmy bezgradientowe, chociaż dla małej liczby zmiennych możliwe jest stosowanie algorytmów gradientowych aproksymujących gradient metodami różnicowymi,
- procesy optymalizacji w takich zadaniach są bardzo długotrwałe.

Sformułowanie zadania

Dla tak scharakteryzowanego problemu długotrwałej optymalizacji w niniejszej rozprawie postawiliśmy sobie następujące zadanie: opracować zbieżną metodę przyspieszenie procesu optymalizacji.

Przyjęte założenia

W rozwiązaniu wyżej postawionego zadania przyjęliśmy następujące założenia:

- funkcja celu jest funkcją ciągłą, dla dowodu zaproponowanej metody potrzebna jest różniczkowalność funkcji,
- wartość funkcji celu nie powinna zawierać zakłóceń chociaż metoda powinna być odporna na małe zakłócenia,
- obliczanie wartości funkcji na tyle czasochłonne, że czas konstrukcji aproksymacji jest zaniebbywalnie mały,
- w przyspieszonym algorytmie krok poszukiwania minimum wzdłuż kierunku powinien być realizowany przez sekwencyjną interpolację paraboliczną (założenie wykorzystywane w dowodzie zbieżności),
- zaproponowana metoda powinna współpracować z różnymi algorytmami optymalizacji lokalnej.

Elementy oryginalne w pracy

Nowymi elementami w pracy są:

- Metoda Search Procedure Exploiting Regularized Objective Approximation (SPELROA) przyspieszania procesów optymalizacji długotrwałej,
- dowód zbieżności metody SPELROA dla algorytmów bezgradientowych,
- wskaźnik określający czy dany punkt znajduje się w regionie dziedziny zawierającym dane dobrze go otaczające,
- metoda wyboru parametru regularyzacji w konstrukcji aproksymacji radialnymi funkcjami bazowymi.
- Testy metody SPELROA
 - w połączeniu z algorytmem Gaussa-Seidela dla dziewięciu problemów testowych,
 - w połączeniu z algorytmem Gaussa-Seidela dla problemów optymalizacji głównego magnesu zakrzywiającego zderzacza cząstek elementarnych LHC,
 - w połączeniu z algorytmem BFGS quasi-Newton dla procesu strojenia regulatora PID.

Układ pracy

Rozprawa, oprócz niniejszego wstępu, zawiera pięć części.

• Rozdział drugi

W rozdziale tym wprowadzamy metodę SPELROA oraz podajemy twierdzenie o jej zbieżności. Do dowodu twierdzenia o zbieżności metody wykorzystujemy Twierdzenie Zangwilla o zbieżności algorytmów optymalizacji lokalnej oparte na teorii przekształceń domkniętych.

• Rozdział trzeci

W rozdziale trzecim omawiamy zagadnienie konstrukcji sieci neuronowej ze ściśle dodatnio określonymi radialnymi funkcjami aktywacji (RBF), która jest modelem aproksymującym funkcję celu w pewnym otoczeniu punktu, w którym ma być obliczana funkcja celu. Wprowadzamy wskaźnik, który skutecznie pozwala określić czy położenie punktu obliczania wartości funkcji względem punktów z poprzednich iteracji zapewni wystarczającą jakość aproksymacji. Sieć aproksymującą trenujemy za pomocą metody regularyzacji Tikhonova z pojedynczym parametrem regularyzacji λ . Omawiamy dwie istniejące metody wyboru wartości tego parametru, mianowicie metodę Uogólnionej Walidacji Krzyżowej [40] oraz metodę L-krzywej [17]. Proponujemy również nową metodę, która lepiej nadaje się do lokalnej aproksymacji niż dwie wcześniej wymienione. Jej przewagę pokazujemy na przykładzie trzech funkcji testowych z pracy [1].

- **Rozdział czwarty**

W czwarty rozdziale przedstawiamy zastosowanie metody do rozwiązania ośmiu problemów testowym z pracy [23]. Te osiem wybranych funkcji jest podzbiorem zbioru funkcji zaproponowanych do testowania algorytmów optymalizacji lokalnej w [23]. Ponadto w rozdziale tym prezentujemy zastosowanie metody SPELROA do przyspieszenia procesu optymalizacji głównego magnesu dipolowego dla zderzacza cząstek elementarnych LHC zaimplementowanego w programie ROXIE [4].

- **Rozdział piąty**

Piąty rozdział pracy zawiera podsumowanie oraz określa kierunki dalszych badań.

- **Dodatek**

Ostatni rozdział stanowi dodatek, do którego przeniesiona została większość żmudnych wyprowadzeń wzorów potrzebnych w dowodzie zbieżności metody SPELROA omówionych w rozdziale drugim.

Rozdział 2

Optymalizacja

Niech dana będzie funkcja celu $f : \mathbb{R}^d \rightarrow \mathbb{R}$ oraz algorytm A minimalizujący funkcję f . Metoda SPELROA (Search Procedure exploiting Locally Regularized Objective Approximation) połączona z algorytmem A zapisuje się w formie algorytmicznej w sposób następujący (patrz [2]):

Algorytm 1.

Metoda SPELROA

Wejście : $f : \mathbb{R}^d \rightarrow \mathbb{R}$ – funkcja celu,
 $\mathbf{x}_0 \in \mathbb{R}^d$ – punkt startowy,
 $\varepsilon > 0$ – dokładność aproksymacji,
 $\tilde{f}(\cdot)$ – aproksymacja funkcji $f(\cdot)$ za pomocą radialnych funkcji bazowych,
 I_s – liczba bezpośrednich obliczeń wartości funkcji w fazie początkowej w algorytmie A ,
 $N < I_s$ – rozmiar zbioru danych używanych do konstrukcji aproksymacji funkcji $\tilde{f}(\cdot)$,
 ε -check – procedura sprawdzająca warunki zbieżności wymagane w twierdzeniu o zbieżności.

0. Wykonać I_s obliczeń wartości funkcji celu $f(\cdot)$ w początkowych punktach wygenerowanych przez algorytmem A .
1. W k -tym kroku wygenerować za pomocą algorytmu A punkt $\mathbf{x}_k$, w którym ma być obliczana funkcja celu.
2. Wygenerować zbiór $\mathbf{X}$ z N najbliższych punktowi $\mathbf{x}_k$ punktów, w których funkcja $f(\cdot)$ była obliczona bezpośrednio.
3. Ocenic czy dla punktu $\mathbf{x}_k$ można skonstruować wiarygodną aproksymację $f(\mathbf{x}_k)$.
 - (a) Jeśli punkt $\mathbf{x}_k$ leży we wiarygodnym regionie dziedziny to skonstruować aproksymację $\tilde{f}$ oraz obliczyć $\tilde{f}(\mathbf{x}_k)$.

(b) Jeśli aproksymacja $\tilde{f}(\mathbf{x}_k)$ była poprawnie skonstruowana to wykonać procedurę ε -check (patrz paragraf 2.5).

(c) Jeśli wynik procedury ε -check jest pozytywny to podstaw

$$f(\mathbf{x}_k) \leftarrow \tilde{f}(\mathbf{x}_k)$$

do algorytmu A .

(d) W przeciwnym razie oblicz $f(\mathbf{x}_k)$ w sposób bezpośredni.

4. Jeśli spełnione jest kryterium stopu algorytmu A to koniec.

W przeciwnym razie podstaw $k := k + 1$ i przejdź do kroku 1.

Podczas tworzenia zbioru $\mathbf{X}$ należy zadbać o to, aby punkty do niego włączone były dostatecznie od siebie oddalone. Jeśli dwa punkty leżą zbyt blisko siebie – gdzie minimalna dopuszczalna odległość jest kontrolowana przez parametr wprowadzany przez użytkownika, którego wartość jest proporcjonalna do średnicy zbioru $\mathbf{X}$ – jeden z punktów powinien zostać zamieniony na inny punkt do tej pory nie włączony do konstruowanego zbioru. Taka procedura konstrukcji zbioru $\mathbf{X}$ gwarantuje, że odległości separująca (*ang. separation distance* – patrz [31]) jest większa od wartości parametru wprowadzanego przez użytkownika, co zaś zapewnia, że macierz interpolacyjna nie jest osobliwa (patrz [22]). Kluczowym krokiem powyższego schematu jest punkt 3. zawierający trójetapowe sprawdzenie czy aproksymacja $\tilde{f}(\mathbf{x}_k)$ może być użyta w algorytmie A zamiast wartości obliczonej bezpośrednio. Warunki sprawdzane w krokach 3.a) i 3.b) są związane z konstrukcją aproksymacji radialnymi funkcjami bazowymi i będą omówione w rozdziale 3, natomiast procedura ε -check występująca w kroku 3.c) jest związana z założeniami twierdzenia o zbieżności i zostanie omówiona w paragrafie 2.2 niniejszego rozdziału.

2.1 Metoda Zangwilla dowodzenia zbieżności algorytmów optymalizacji

Algorytmy optymalizacji nieliniowej kierunków dopuszczalnych mają postać

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \tau_k \mathbf{d}^k$$

gdzie $\mathbf{d}^k$ jest kierunkiem poszukiwania a τ_k jest krokiem w k -tej iteracji. Jedną iteracją w algorytmach kierunków dopuszczalnych jest złożenie przekształceń algorytmicznych $A = M^1 D$ gdzie $D : \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d$ jest przekształceniem generującym kierunek

$$D(\mathbf{x}) = (\mathbf{x}, \mathbf{d})$$

gdzie $\mathbf{d} \in \mathbb{R}^d$ jest wektorem kierunku, natomiast $M^1 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ jest przekształceniem minimalizującym funkcję celu $f(\cdot)$ wzdłuż kierunku $\mathbf{d}$ tzn.

$$M^1(\mathbf{x}, \mathbf{d}) = \{\mathbf{y} : f(\mathbf{y}) = \min_{\tau \in J} f(\mathbf{x} + \tau \mathbf{d})\}$$

gdzie $(\mathbf{x}, \mathbf{d}) \in \mathbb{R}^d \times \mathbb{R}^d$ a J jest przedziałem zmienności parametru τ .

W swojej monografii [46] Zangwill zaproponował metodę dowodzenia zbieżności algorytmów optymalizacyjnych kierunków dopuszczalnych bazującą na własnościach przekształcenia algorytmicznego A . W paragrafie tym omawiamy główne elementy teorii Zangwilla, za pomocą których będziemy mogli sformułować i udowodnić twierdzenie o zbieżności metody SPELROA połączonej z algorytmami poszukiwań bezpośrednich takimi jak metoda Gaussa-Seidela. Po drobnych modyfikacjach zaprezentowana metoda dowodu może być również zastosowana dla algorytmów kierunków sprzężonych.

Definicja 1. ([46], str. 92)

Przez przekształcenie $A : V \rightarrow V$ punktu w zbiór rozumiemy przyporządkowanie każdemu punktowi $\mathbf{x} \in V$ zbioru $A(\mathbf{x}) \subseteq V$. Działanie algorytmu $A(\cdot)$ będącego odwzorowaniem punktu w zbiór polega na przyporządkowaniu punktowi $\mathbf{x}_k$ zbioru $A(\mathbf{x}^k)$ i wybraniu dowolnego

$$\mathbf{x}^{k+1} \in A(\mathbf{x}^k).$$

Przekształcenie $A = M^1 D$ określające algorytm optymalizacji kierunków dopuszczalnych jest przekształceniem punktu w zbiór.

Definicja 2. ([46], str. 96)

Mówimy, że przekształcenie $A : V \rightarrow V$ jest domknięte w punkcie $\mathbf{x}^\infty$, jeśli zachodzi następująca implikacja:

1. $\mathbf{x}^k \rightarrow \mathbf{x}^\infty, k \in \mathcal{K}$,
2. $\mathbf{y}^k \in A(\mathbf{x}^k), k \in \mathcal{K}$,
3. $\mathbf{y}^k \rightarrow \mathbf{y}^\infty$,

implikuje

4. $\mathbf{y}^\infty \in A(\mathbf{x}^\infty)$,

gdzie $\mathcal{K}$ jest ciągiem liczb naturalnych.

Mówimy, że przekształcenie A jest domknięte na zbiorze $X \subseteq V$, jeśli jest domknięte w dowolnym punkcie $\mathbf{x} \in X$.

Własność domkniętości przekształcenia algorytmicznego jest analogią własności ciągłości dla „zwykłych” funkcji.

Twierdzenie 1. (patrz [46], str. 99)

Niech przekształcenie $A : V \rightarrow V$ punktu w zbiór wyznacza algorytm, który dla danego punktu $\mathbf{x}_1$, generuje ciąg $\{\mathbf{x}^k\}_{k=1}^\infty$. Niech będzie dany zbiór rozwiązań $S \subset V$.

Założmy, że

1. Wszystkie punkty $\mathbf{x}^k$ należą do zbioru zwartego $X \subset V$.

2. Istnieje funkcja $Z : V \rightarrow \mathbb{R}$ taka, że

(a) jeśli punkt $\mathbf{x}$ nie jest rozwiązaniem, to dla dowolnego $\mathbf{y} \in A(\mathbf{x})$ zachodzi

$$Z(\mathbf{y}) < Z(\mathbf{x})$$

(b) jeśli punkt $\mathbf{x}$ jest rozwiązaniem, to albo algorytm kończy działanie, albo dla dowolnego $\mathbf{y} \in A(\mathbf{x})$ zachodzi

$$Z(\mathbf{y}) \leq Z(\mathbf{x}).$$

3. Przekształcenie A jest domknięte w punkcie $\mathbf{x}$, jeśli punkt ten nie jest rozwiązaniem..

Wówczas, albo algorytm kończy działanie w punkcie będącym rozwiązaniem, albo każdy zbieżny podciąg ciągu generowanego przez algorytm ma granicę w zbiorze rozwiązań S .

Dodatkowo potrzebujemy jeszcze dwóch lematów z [46] mówiących o domkniętości złożenia przekształceń domkniętych oraz o domkniętości przekształcenia minimalizacji wzdłuż kierunku M^1 .

Lemat 1. (patrz [46], str. 104)

Niech $C : W \rightarrow X \subset \mathbb{R}^d$ będzie daną funkcją oraz $B : X \rightarrow Y$ będzie przekształceniem punktu w zbiór. Załóżmy, że funkcja C jest ciągłą w punkcie $\mathbf{w}^\infty$ a B jest domknięte w $C(\mathbf{w}^\infty)$. Wówczas odwzorowanie złożone $A = BC : W \rightarrow Y$ jest domknięte w $\mathbf{w}^\infty$.

Lemat 2. (patrz [46], str. 112)

Niech f będzie funkcją ciągłą. Wówczas przekształcenie M^1 , jest domknięte jeśli zbiór J jest przedziałem domkniętym i ograniczonym.

2.2 Główne twierdzenie

W praktyce stosuje się inny operator poszukiwania minimum wzdłuż kierunku niż przekształcenie M^1 . Wiele różnych implementacji operatora M^1 można znaleźć w [46]. Niezależnie od implementacji wykonanie przekształcenia $M^1(\cdot, \cdot)$ jest kosztowne.

Rozważmy operator poszukiwania minimum wzdłuż kierunku zdefiniowany następująco

$$M^*(\mathbf{x}, \mathbf{d}) = M^1(\mathbf{x}, \mathbf{d}) \cup \{\mathbf{y} = \mathbf{x} + \tau \mathbf{d} : f(\mathbf{y}) \leq f(\mathbf{x}) - \Delta, \tau \in J\}. \quad (2.1)$$

Jest to przekształcenie punktu w zbiór, którego wartością jest zbiór punktów dla których funkcja f zmniejsza wartość o $\Delta \in \mathbb{R}_+$ wzdłuż kierunku $\mathbf{d}$ począwszy od

punktu $\mathbf{x}$ lub gdy ten zbiór jest pusty przyjmuje on wartość minimalną funkcji f wzdłuż tego kierunku.

Sugestię praktycznego wykorzystania operatora M^* znaleźć można w jednym z ćwiczeń w [46]. Poniżej prezentujemy dowód domkniętości operatora M^* co pokazuje, że może on być użyty w algorytmach optymalizacyjnych kierunków dopuszczalnych zamiast operatora M^1 . Zastosowanie operatora M^* , który w końcowych fazach optymalizacji przechodzi w operator M^1 jest praktyczniejsze od zastosowania tylko operatora M^1 . Wynika to stąd, że w początkowej części poszukiwania minimum wzdłuż kierunku niezależnie od implementacji, występują najbardziej znaczące zmniejszenie wartości funkcji, natomiast kolejne kroki powodują znacznie wolniejsze zmniejszanie wartości funkcji.

W niniejszym paragrafie omówimy implementację operatora M^* za pomocą zaburzonej interpolacji parabolicznej. Przez zaburzoną interpolację paraboliczną rozumiemy, algorytm Luenbergera [25], w którym przy pewnych warunkach wprowadzamy zaburzenie wartości funkcji z zadaniem błędem ε . Te wyniki pozwolą nam dowieść prawdziwości następującego twierdzenia:

Twierdzenie 2.

Załóżmy, że $f : \Omega \rightarrow \mathbb{R}; \Omega \subset \mathbb{R}^d$ jest funkcją celu w zagadnieniu optymalizacyjnym. Załóżmy, że mamy daną metodę aproksymacji funkcji celu f w pewnych punktach dziedziny Ω z błędem względnym $\varepsilon > 0$. Jeśli funkcja f jest ściśle wypukła to metoda SPELROA w połączeniu z algorytmem Gaussa-Seidela z obrotem bazy kierunków poszukiwań, zbiega do punktu krytycznego pierwszego rzędu $x_0 \in \mathbf{S}$, tzn. $\nabla f(x_0) = 0$.

Dowód Twierdzenia 2 będzie polegał na sprawdzeniu założeń Twierdzenia 1, gdzie przekształcenie A jest postaci

$$A = RM^*D^dM^*D^{d-1} \dots M^*D^2M^*D^1$$

gdzie D^i wybiera i -ty kierunek z bazy kierunków ortogonalnych w k -tej iteracji a R jest przekształceniem ortogonalizującym wektory bazowe wzdłuż kierunku $\mathbf{x}_0^{k-1} \mathbf{x}_d^{k-1}$ z $(k-1)$ -szej iteracji.

2.3 Domkniętość przekształcenia algorytmicznego M^*

Aby wykazać, że założenie 3. z twierdzenia 1 jest spełnione musimy najpierw pokazać, że transformacja M^* zdefiniowana przez (2.1) jest domknięta. Dowiedzimy następującego lematu:

Lemat 3.

Niech f będzie funkcją ciągłą. Wówczas przekształcenie M^ jest domknięte, jeśli J jest przedziałem domkniętym i ograniczonym.*

Poniżej prezentujemy dowód, który jest modyfikacją dowodu Lematu 2 podanego w [46].

Dowód

Zgodnie z definicją 2 rozważmy ciąg $\{(\mathbf{x}^k, \mathbf{d}^k)\}_{k=1}^{\infty}$ oraz $\{\mathbf{y}^k\}_{k=1}^{\infty}$. Przy założeniu, że

1. $(\mathbf{x}^k, \mathbf{d}^k) \rightarrow (\mathbf{x}^{\infty}, \mathbf{d}^{\infty}), \quad k \in \mathcal{K},$
2. $\mathbf{y}^k \in M^*(\mathbf{x}^k, \mathbf{d}^k), \quad k \in \mathcal{K},$
3. $\mathbf{y} \rightarrow \mathbf{y}^{\infty}, \quad k \in \mathcal{K},$

należy pokazać, że

$$\mathbf{y}^{\infty} \in A(\mathbf{x}^{\infty}).$$

Z definicji operatora M^* mamy

$$\mathbf{y}^k = \mathbf{x}^k + \tau^k \mathbf{d}^k$$

gdzie $\tau_k \in J$ jest takie, że

$$f(\mathbf{x}^k + \tau^k \mathbf{d}^k) \leq f(\mathbf{x}^k) - \Delta, \quad (2.2)$$

gdy możliwe jest zmniejszenie wartości funkcji o Δ lub

$$\tau_k = \arg \min_{\tau \in J} \{f(\mathbf{x}_k + \tau \mathbf{d}_k)\} \quad (2.3)$$

w przeciwnym razie.

Ponieważ J jest domknięty i ograniczony jest również zwarty. Ponieważ $\tau^k \in J$ dla $k \in \mathcal{K}$ a J jest zwarty, to istnieje zbieżny podciąg

$$\tau^k \rightarrow \tau^{\infty}, \quad k \in \mathcal{K}^1,$$

gdzie $\mathcal{K}^1 \subset \mathcal{K}$ oraz $\tau^{\infty} \in J$.

Dla ustalonego $\tau \in J$ z monotoniczności ciągu $\mathbf{y}^k$ wynika, że istnieje takie k_0 , że dla $k > k_0$ nie może być użyta już składowa zminiejszająca wartość funkcji o Δ , lecz tylko składowa odpowiadająca operatorowi M^1 , mamy wówczas

$$f(\mathbf{y}^k) \leq f(\mathbf{x}^k + \tau \mathbf{d}^k) \quad \text{dla } k > k_0. \quad (2.4)$$

Sytuacja taka występuje dostatecznie blisko minimum lokalnego funkcji f .

Ponieważ funkcja f jest ciągła, to w granicy otrzymujemy

$$f(\mathbf{y}^{\infty}) = \lim_{k \in \mathcal{K}^1} f(\mathbf{y}^k) \leq \lim_{k \in \mathcal{K}^1} f(\mathbf{x}^k + \tau \mathbf{d}^k) = f(\mathbf{x}^{\infty} + \tau \mathbf{d}^{\infty}). \quad (2.5)$$

Ponieważ (2.5) zachodzi dla dowolnego τ dla $k > k_0$, więc i dla dowolnego punktu $\mathbf{y}^* \in M^*(\mathbf{x}^{\infty}, \mathbf{d}^{\infty})$, mamy

$$f(\mathbf{y}^{\infty}) \leq f(\mathbf{y}^*). \quad (2.6)$$

Z drugiej strony w punkcie $\mathbf{y}^* \in M^*(\mathbf{x}^\infty, \mathbf{d}^\infty)$ funkcja f dla $\tau \in J$ osiąga najmniejszą wartość

$$\mathbf{y}^\infty = \mathbf{x}^\infty + \tau^\infty \mathbf{d}^\infty, \quad \tau^\infty \in J.$$

więc

$$f(\mathbf{y}^*) \leq f(\mathbf{y}^\infty). \quad (2.7)$$

Porównując (2.7) z (2.6) otrzymujemy wynik

$$\mathbf{y}^\infty \in M^*(\mathbf{x}^\infty, \mathbf{d}^\infty).$$

□

Zauważmy, że dowód będzie również prawdziwy, jeśli funkcja f zawierać będzie pewną skończoną liczbę izolowanych punktów nieciągłości pierwszego rzędu znajdujących się w dostatecznie dużej odległości od minimum lokalnego.

Aby pokazać domkniętość przekształcenia A robiąc użytek z Lematu 1 musimy również zauważyć, że przekształcenia $D^i(\mathbf{x}) = (\mathbf{x}, \mathbf{d}_i)$ ($i = 1, \dots, d$) są funkcjami ciągłymi. Przekształcenie R , które generuje bazę kierunków ortogonalnych $[\mathbf{d}_0^{\text{new}}, \dots, \mathbf{d}_{d-1}^{\text{new}}]$ dla kroku o indeksie k jest określone jako

$$R(\mathbf{x}_0^{k-1}, [\mathbf{d}_0, \dots, \mathbf{d}_{d-1}]) = (\mathbf{x}_0^k, [\mathbf{d}_0^{\text{new}}, \dots, \mathbf{d}_{d-1}^{\text{new}}]),$$

gdzie ciąg nowych wektorów ortogonalnych jest określony jednoznacznie przez proces ortogonalizacji wektorów $\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{d-1}$

$$\begin{aligned} \mathbf{w}_0 &= s_0 \mathbf{d}_0 + s_1 \mathbf{d}_1 + \dots + s_{d-1} \mathbf{d}_{d-1} \\ \mathbf{w}_1 &= \quad \quad + s_1 \mathbf{d}_1 + \dots + s_{d-1} \mathbf{d}_{d-1} \\ &\dots \\ \mathbf{w}_{d-1} &= \quad \quad \quad \quad + s_{d-1} \mathbf{d}_{d-1} \end{aligned}$$

gdzie skalary $s_0, s_1, \dots, s_{d-1}$ odpowiadają rozmiarom kroków we wszystkich kierunkach z kroku $k - 1$. Przekształcenie R jest jednoznacznie określone bez żadnych dodatkowych warunków spełnianych przez skalary $s_0, s_1, \dots, s_{d-1}$ jeśli proces ortogonalizacji jest wykonany za pomocą algorytmu zaprezentowanego w [26]. W tym wypadku przekształcenie R jest również funkcją ciągłą. Ponieważ przekształcenie A jest złożeniem przekształceń M^* z funkcjami ciągłymi D_i ($i = 0, \dots, d - 1$) oraz R , spełnione są założenia Lematu 1 i stąd otrzymujemy, że przekształcenie A jest domknięte. To dowodzi, że spełnione jest założenie 3. Twierdzenia 1 dla przekształcenia M^* bez zaburzeń. W następnym paragrafie wykazemy, że przekształcenie M^* może zostać zrealizowane za pomocą przekształcenia M^1 , do którego w każdym kroku może zostać wprowadzone zaburzenie.

2.4 Zaburzenie w algorytmie poszukiwania minimum wzdłuż kierunku

Dla algorytmów bezgradientowych z bazą wektorów ortogonalnych jako zbiorem kierunków poszukiwań przekształcenie M^* jest jedynym miejscem w algorytmie

gdzie wprowadzane jest zaburzenie powstałe z aproksymacji funkcji celu w metodzie SPELROA. Tak więc aby pokazać, że założenie 2. Twierdzenia 1 są spełnione wystarczy, że implementacją M^* z zaburzeniem wartości funkcji na poziomie $\varepsilon > 0$ minimalizuje funkcję f wzdłuż kierunku $\mathbf{d}$.

Dowód zbieżności algorytmu poszukiwania minimum wzdłuż kierunku za pomocą interpolacji parabolicznej można znaleźć w pracach [21] lub [25]. W paragrafie tym podamy warunki jakie musi spełniać zaburzenie wartości funkcji wprowadzane do algorytmu, aby dowód podany w [25] pozostawał w mocy. W paragrafie tym staramy się aby notacja była możliwie najbardziej zbliżona do notacji dowodu w [25].

Niech funkcja $f : \mathbb{R} \rightarrow \mathbb{R}$ będzie unimodalna. Dla trójki $\zeta_i = (\zeta_i^1, \zeta_i^2, \zeta_i^3)$ zachodzi $f(\zeta_i^2) \leq \min\{f(\zeta_i^1), f(\zeta_i^3)\}$ tzn. przedział $[\zeta_i^1, \zeta_i^3]$ zawiera unikalne minimum funkcji f . Parabola interpolująca funkcję w punktach $(\zeta_i^1, f(\zeta_i^1))$, $(\zeta_i^2, f(\zeta_i^2))$, $(\zeta_i^3, f(\zeta_i^3))$ w formie Lagrange'a zapisuje się

$$\begin{aligned} q(x, \zeta_i) = & f(\zeta_i^1) \frac{(x - \zeta_i^2)(x - \zeta_i^3)}{(\zeta_i^1 - \zeta_i^2)(\zeta_i^1 - \zeta_i^3)} + f(\zeta_i^2) \frac{(x - \zeta_i^1)(x - \zeta_i^3)}{(\zeta_i^2 - \zeta_i^1)(\zeta_i^2 - \zeta_i^3)} + \\ & + f(\zeta_i^3) \frac{(x - \zeta_i^1)(x - \zeta_i^2)}{(\zeta_i^3 - \zeta_i^1)(\zeta_i^3 - \zeta_i^2)}. \end{aligned} \quad (2.8)$$

Dla niezaburzonej funkcji celu definiujemy zbiór osiągalnych trójek $\mathbf{T} \subset \mathbb{R}^3$ definiujących przedział $[\zeta^1, \zeta^3]$, który zawiera minimum $\lambda^*(\zeta)$ jako

$$\begin{aligned} \mathbf{T} := & \{ \zeta \in \mathbb{R}^3 : \zeta^1 < \zeta^2 < \zeta^3, f(\zeta^2) \leq \min\{f(\zeta^1), f(\zeta^3)\} \\ & \cup \{ \zeta \in \mathbb{R}^3 : \zeta^1 = \zeta^2 < \zeta^3, f'(\zeta^1) \leq 0, f(\zeta^3) \geq f(\zeta^1) \} \\ & \cup \{ \zeta \in \mathbb{R}^3 : \zeta^1 < \zeta^2 = \zeta^3, f'(\zeta^3) \geq 0, f(\zeta^1) \geq f(\zeta^3) \} \\ & \cup \{ \zeta \in \mathbb{R}^3 : \zeta^1 = \zeta^2 = \zeta^3 = \hat{\lambda} \}. \end{aligned}$$

Dla $\zeta \in \mathbf{T}$ z $\zeta^1 < \zeta^2 < \zeta^3$ minimum paraboli (2.8) zapisuje się

$$\lambda^*(\zeta) = \frac{1}{2} \frac{((\zeta^3)^2 - (\zeta^2)^2)f(\zeta^1) + ((\zeta^3)^2 - (\zeta^1)^2)f(\zeta^2) + ((\zeta^2)^2 - (\zeta^1)^2)f(\zeta^3)}{(\zeta^3 - \zeta^2)f(\zeta^1) + (\zeta^3 - \zeta^1)f(\zeta^2) + (\zeta^2 - \zeta^1)f(\zeta^3)}.$$

Zbiór osiągalnych trójek $A(\zeta)$ dla następnego kroku jest zbiorem trójek kandydatów, które mogą zamienić $\zeta \in \mathbf{T}$ określając mniejszy przedział zawierający $\hat{\lambda}$ w następnej iteracji algorytmu. Zdefiniujmy A^0 jako

$$A^0(\zeta) := \mathbf{T} \cap \{u_1(\zeta), u_2(\zeta), u_3(\zeta), u_4(\zeta)\} \quad (2.9)$$

gdzie

$$\begin{aligned} u_1(\zeta) &= (\zeta^1, \lambda^*(\zeta), \zeta^2), \\ u_2(\zeta) &= (\zeta^2, \lambda^*(\zeta), \zeta^3), \\ u_3(\zeta) &= (\lambda^*(\zeta), \zeta^2, \zeta^3), \\ u_4(\zeta) &= (\zeta^1, \zeta^2, \lambda^*(\zeta)). \end{aligned}$$

Założmy, że wartość funkcji jest zaburzona tylko w jednym z trzech punktów w trójce $\zeta = (\zeta^1, \zeta^2, \zeta^3)$ użytej do konstrukcji paraboli interpolującej. Dla zaburzenia wartości funkcji f na poziomie $\varepsilon \neq 0$ definiujemy trzy zbiory trójek:

$$\mathbf{T}^1(\varepsilon) := \{\zeta \in \mathbb{R}^3 : \zeta^1 < \zeta^2 < \zeta^3, f(\zeta^2) \leq \min\{f(\zeta^1)(1 - |\varepsilon|), f(\zeta^3)\}\} \quad (2.10)$$

z minimum

$$\tilde{\lambda}_1^*(\varepsilon; \zeta) = \frac{1}{2} \frac{[(\zeta^3)^2 - (\zeta^2)^2]f(\zeta^1)(1 + |\varepsilon|) + [(\zeta^3)^2 - (\zeta^1)^2]f(\zeta^2) + [(\zeta^2)^2 - (\zeta^1)^2]f(\zeta^3)}{(\zeta^3 - \zeta^2)f(\zeta^1)(1 + |\varepsilon|) + (\zeta^3 - \zeta^1)f(\zeta^2) + (\zeta^2 - \zeta^1)f(\zeta^3)},$$

dla zaburzenia funkcji f w punkcie ζ_1 ,

$$\mathbf{T}^2(\varepsilon) := \{\zeta \in \mathbb{R}^3 : \zeta^1 < \zeta^2 < \zeta^3, f(\zeta^2)(1 + |\varepsilon|) \leq \min\{f(\zeta^1), f(\zeta^3)\}\} \quad (2.11)$$

z minimum

$$\tilde{\lambda}_2^*(\varepsilon; \zeta) = \frac{1}{2} \frac{[(\zeta^3)^2 - (\zeta^2)^2]f(\zeta^1) + [(\zeta^3)^2 - (\zeta^1)^2]f(\zeta^2)(1 + |\varepsilon|) + [(\zeta^2)^2 - (\zeta^1)^2]f(\zeta^3)}{(\zeta^3 - \zeta^2)f(\zeta^1) + (\zeta^3 - \zeta^1)f(\zeta^2)(1 + |\varepsilon|) + (\zeta^2 - \zeta^1)f(\zeta^3)},$$

dla zaburzenia funkcji f w punkcie ζ_2 oraz

$$\mathbf{T}^3(\varepsilon) := \{\zeta \in \mathbb{R}^3 : \zeta^1 < \zeta^2 < \zeta^3, f(\zeta^2) \leq \min\{f(\zeta^1), f(\zeta^3)(1 - |\varepsilon|)\}\} \quad (2.12)$$

z minimum

$$\tilde{\lambda}_3^*(\varepsilon; \zeta) = \frac{1}{2} \frac{[(\zeta^3)^2 - (\zeta^2)^2]f(\zeta^1) + [(\zeta^3)^2 - (\zeta^1)^2]f(\zeta^2) + [(\zeta^2)^2 - (\zeta^1)^2]f(\zeta^3)(1 + |\varepsilon|)}{(\zeta^3 - \zeta^2)f(\zeta^1) + (\zeta^3 - \zeta^1)f(\zeta^2) + (\zeta^2 - \zeta^1)f(\zeta^3)(1 + |\varepsilon|)},$$

dla zaburzenia funkcji f w punkcie ζ_3 . Takie definicje zbiorów $\mathbf{T}^l(\varepsilon)$ ($l \in \{1, 2, 3\}$) zapewniają, że zaburzone minima znajdują się w przedziale $[\zeta^1, \zeta^3]$. Dla $l \in \{1, 2, 3\}$ odpowiadający zbiór trójek osiągalnych definiujemy jako

$$\tilde{A}^l(\varepsilon; \zeta) := \mathbf{T}^l(\varepsilon) \cap \{\tilde{u}_1^l(\varepsilon; \zeta), \tilde{u}_2^l(\varepsilon; \zeta), \tilde{u}_3^l(\varepsilon; \zeta), \tilde{u}_4^l(\varepsilon; \zeta)\} \quad (2.13)$$

gdzie

$$\begin{aligned} \tilde{u}_1^l(\varepsilon; \zeta) &= (\zeta^1, \tilde{\lambda}_l^*(\varepsilon; \zeta), \zeta^2), \\ \tilde{u}_2^l(\varepsilon; \zeta) &= (\zeta^2, \tilde{\lambda}_l^*(\varepsilon; \zeta), \zeta^3), \\ \tilde{u}_3^l(\varepsilon; \zeta) &= (\tilde{\lambda}_l^*(\varepsilon; \zeta), \zeta^2, \zeta^3), \\ \tilde{u}_4^l(\varepsilon; \zeta) &= (\zeta^1, \zeta^2, \tilde{\lambda}_l^*(\varepsilon; \zeta)). \end{aligned} \quad (2.14)$$

W końcu zdefiniujemy zbiór punktów stacjonarnych $\mathbf{S}$ jako

$$\mathbf{S} := \{\zeta \in \mathbf{T} : f'(\zeta^1) = 0 \text{ or } f'(\zeta^2) = 0 \text{ or } f'(\zeta^3) = 0\}.$$

Jest to zbiór punktów krytycznych pierwszego rzędu funkcji f .

Kluczowym elementem dowodu zbieżności algorytmu poszukiwania minimum jednej zmiennej podanym w [25] jest lemat (Lemat 1.7.9, str. 153) o domkniętości niezaburzonego przekształcenia $A^0(\cdot)$. Podamy teraz analogiczny lemat dla przekształcenia, które może zawierać zaburzenie wartości funkcji na poziomie błędu ε w punktach, na których skonstruowana jest parabola. W dowodzie podamy jakie warunki musi spełniać ε , aby zachowana była zbieżność algorytmu.

Lemat 4.

1. Dla każdego $\zeta \in \mathbf{T}$, zbiór $A(\zeta) = A^0(\zeta) \cup \tilde{A}^1(\varepsilon; \zeta) \cup \tilde{A}^2(\varepsilon; \zeta) \cup \tilde{A}^3(\varepsilon; \zeta)$ jest niepusty.
2. Przekształcenie punktu w zbiór $A(\cdot)$ jest domknięte.
3. Dla każdego $\zeta \in \mathbf{T} \setminus \mathbf{S} := \{\zeta \in \mathbf{T} : \zeta \notin \mathbf{S}\}$ takiego, że $\zeta^1 < \zeta^2 < \zeta^3$ zachodzi $c(y) < c(\zeta)$ dla każdego $y \in A(\zeta)$ gdzie

$$c(\zeta) = \tilde{f}_1(\zeta^1) + \tilde{f}_2(\zeta^2) + \tilde{f}_3(\zeta^3) \quad (2.15)$$

gdzie $\tilde{f}_l \equiv f$ lub $\tilde{f}_l \equiv f(\cdot)(1 + |\varepsilon|)$ w zależności od tego, czy wartość w punkcie ζ^l jest „dokładna” czy zaburzona.

Zauważmy, że „dokładna” znaczy tutaj zaburzona co najwyżej na poziomie błędu reprezentacji zmiennoprzecinkowej $\epsilon \ll \varepsilon$, tzn. obliczona w sposób bezpośredni nie zaś aproksymowana.

Dowód

Wprowadźmy następującą notację $\tilde{f}_l(\cdot) = f(\cdot)$, gdy funkcja f jest wartościowana w punktach $\zeta^0 \neq \zeta^l$ lub $\tilde{f}_l(\cdot) = f(\cdot)(1 + \varepsilon)$, gdy funkcja f wartościowana jest w punkcie $\zeta^0 = \zeta^l$ ($l \in \{1, 2, 3\}$). Na wstępie zauważmy, że $\mathbf{T}^l(\varepsilon) \subset \mathbf{T}$ ($l \in \{1, 2, 3\}$) dla $\varepsilon > 0$.

1. Niech $\zeta = (\zeta^1, \zeta^2, \zeta^3) \in \mathbf{T}$ będzie ustalone. Jeśli $f(\zeta_1)$, $f(\zeta_2)$ i $f(\zeta_3)$ są obliczone bezpośrednio tzn., bez zaburzenia, to $A(\zeta)$ jest niepusty na mocy dowodu z [25]. Rozpatrzmy przypadek, gdy jedna z wartości funkcji jest zaburzona błędem względnym $\varepsilon > 0$. Minimum paraboli skonstruowane w tym wypadku dane jest wyrażeniem $\tilde{\lambda}_l^*(\varepsilon; \zeta)$ gdzie $l \in \{1, 2, 3\}$ w zależności od tego, w którym punkcie wartość była aproksymowana. Rozpatrzmy wypadek, gdy $\tilde{\lambda}_l^*(\varepsilon, \zeta) \in [\zeta^1, \zeta^2]$ zakładając ponadto, że minimum $\lambda^*(\zeta)$ gdyby żadna z wartości nie była zaburzona również należy do $[\zeta^1, \zeta^2]$. Wówczas $A(\zeta)$ jest pusty wtedy i tylko wtedy, gdy zarówno $\tilde{u}_1(\varepsilon; \zeta)$ jak i $\tilde{u}_3(\varepsilon; \zeta)$ nie należą do $A(\zeta)$, tzn. wtedy i tylko wtedy, gdy

$$f(\tilde{\lambda}_1^*(\varepsilon; \zeta)) > \min\{f(\zeta^1)(1 + |\varepsilon|), f(\zeta^2)\} = f(\zeta^2) \quad (2.16)$$

oraz

$$f(\zeta^2) > \min\{f(\tilde{\lambda}_1^*(\varepsilon; \zeta)), f(\zeta^3)\} \geq \min\{f(\tilde{\lambda}_1^*(\varepsilon; \zeta)), f(\zeta^2)\}, \quad (2.17)$$

gdy wartość funkcji była zaburzona w punkcie ζ^1 , albo

$$f(\tilde{\lambda}_2^*(\varepsilon; \zeta)) > \min\{f(\zeta^1), f(\zeta^2)(1 + |\varepsilon|)\} = f(\zeta^2)(1 + |\varepsilon|) \quad (2.18)$$

oraz

$$f(\zeta^2)(1 + |\varepsilon|) > \min\{f(\tilde{\lambda}_2^*(\varepsilon; \zeta)), f(\zeta^3)\} \geq \min\{f(\tilde{\lambda}_2^*(\varepsilon; \zeta)), f(\zeta^2)(1 + |\varepsilon|)\}, \quad (2.19)$$

gdy wartość funkcji była zaburzona w punkcie ζ^2 , albo

$$f(\lambda_3^*(\varepsilon; \zeta)) > \min\{f(\zeta^1), f(\zeta^2)\} = f(\zeta^2) \quad (2.20)$$

oraz

$$f(\zeta^2) > \min\{f(\tilde{\lambda}_3^*(\varepsilon; \zeta)), f(\zeta^3)(1 + |\varepsilon|)\} \geq \min\{f(\tilde{\lambda}_3^*(\varepsilon; \zeta)), f(\zeta^2)\}, \quad (2.21)$$

gdy wartość funkcji była zaburzona w punkcie ζ^3 . Ponieważ nierówności (2.17) implikują, że $f(\zeta^2) \geq f(\tilde{\lambda}_1^*(\varepsilon; \zeta))$ otrzymujemy sprzeczność z nierównościami (2.16) co dowodzi tezy, gdy wartość funkcji była aproksymowana w punkcie ζ^1 . W podobny sposób otrzymujemy sprzeczności dla nierówności (2.19) i (2.18) dla aproksymacji w punkcie ζ^2 oraz (2.21) i (2.20) dla aproksymacji w punkcie ζ^3 . Zauważmy, że warunek, że odpowiadające $\lambda^*(\zeta)$ należy do $[\zeta_1, \zeta_2]$ zapewnia, że jeśli porównamy $\tilde{y} = (\tilde{y}^1, \tilde{y}^2, \tilde{y}^3) \in \tilde{A}^l(\zeta)$ ($l \in \{1, 2, 3\}$) z $y = (y^1, y^2, y^3) \in A^0(\zeta)$ to otrzymamy, że $\tilde{y}^1 \geq y^1$ oraz $\tilde{y}^3 \leq y^3$.

Przypadek kiedy $\lambda^*(\zeta)$ należy do $[\zeta_2, \zeta_3]$ jest symetryczny.

Powyższe warunki będą wyrażone względem ε w następnym punkcie dowodu i zostaną użyte w dowodzie głównego twierdzenia.

2. Załóżmy, że $\{\zeta_i\}_{i=0}^\infty$ oraz $\zeta_i \rightarrow \zeta_* \in \mathbf{T}$ i niech istnieje $\zeta_* \in \mathbf{T}$ (zwróćmy uwagę na fakt, że ζ_* i ζ^* oznaczają różne punkty) oraz nieskończony podciąg $K \subset \mathbb{N}$ taki, że $\zeta_{i+1} \in A(\zeta_i)$ dla każdego $i \in K$ takiego, że $\zeta_{i+1} \xrightarrow{K} \zeta_*$ przy $i \rightarrow \infty$. Wówczas musi istnieć $k \in \{1, 2, 3, 4\}$ oraz nieskończony podciąg $K' \subset K$ taki, że $\zeta_{i+1} = u_k(\zeta_i)$ lub $\zeta_{i+1} = \tilde{u}_k^{l_i}(\varepsilon; \zeta_i)$ (gdzie $l_i \in \{1, 2, 3\}$) dla każdego $i \in K'$. Jak zostało dowiedzione w [25] funkcje $u_k(\cdot)$ są ciągłe a więc jeśli ciąg $\{\zeta_i\}_{i=0}^\infty$ nie zawiera zaburzonego ζ_i to wówczas z ciągłości $u_k(\cdot)$ względem ζ i domkniętości zbioru $\mathbf{T}$ wynika, że $u_k(\zeta_i) \rightarrow u_k(\zeta_*) = \zeta_* \in A(\zeta_*)$. To dowodzi domkniętości przekształcenia A .

Wprowadzenie trójki ζ_{i+1} zawierającej punkt, w którym wartość była zaburzona, tzn. $\zeta_{i+1} \in \tilde{A}^l(\varepsilon; \zeta)$ wprowadza nieciągłość pierwszego rzędu do funkcji $u_k(\zeta)$ i powyższa argumentacja nie może zostać zastosowana jeśli nie zauważymy, że użycie algorytmu $A(\cdot)$ wprowadza jedynie skończoną liczbę izolowanych punktów nieciągłości.

Należy rozpatrzyć dwa przypadki w zależności od liczby wystąpień trójek zawierających elementy zaburzone w ciągu $\{\zeta_i\}_{i=0}^\infty$

- Kiedy liczba wystąpień $\tilde{u}_k^{l_i}$ w ciągu jest $\{\zeta_i\}_{i=0}^\infty$ jest skończona to dowód domkniętości przekształcenia $A(\cdot)$ podany w [25] może zostać zastosowany po usunięciu $\{\zeta_i\}_{i=0}^\infty$ wszystkich początkowych elementów zawierający podciąg wartości zaburzonych.
- Załóżmy teraz, że $\tilde{u}_k^{l_i}$ występuje nieskończoną liczbę razy w ciągu $\{\zeta_i\}_{i=0}^\infty$. Ponieważ zakładamy, że $\{\zeta_i\}_{i=1}^\infty$ jest zbieżny to dla dowolnego $\delta \in \mathbb{R}$ istnieje i_0 , takie że dla każdego $i > i_0$ mamy

$$\|\zeta_i - \zeta_{i+1}\|_2 = |(\zeta_i^1, \zeta_i^2, \zeta_i^3) - (\zeta_{i+1}^1, \zeta_{i+1}^2, \zeta_{i+1}^3)| < \delta \quad (2.22)$$

Wyberzmy więc i_1 , dla którego zachodzi nierówność (2.22). Ponieważ aproksymacja w kolejnych iteracjach występuje nieskończenie wiele razy to istnieje podciąg $\mathcal{K} = \{i : i > i_1\} \subset \mathbb{N}$ taki, że jeśli $i \in \mathcal{K}$ to

$$\zeta_{i+1} \in \tilde{A}^l(\varepsilon; \zeta_i)$$

dla pewnego $l \in \{1, 2, 3\}$. Bez straty ogólności możemy założyć, że $\tilde{\lambda}_l^* \in [\zeta_i^1, \zeta_i^2]$. Przypadek, gdy $\tilde{\lambda}_l^* \in [\zeta_i^2, \zeta_i^3]$ jest symetryczny. Udowodnimy następujący lemat

Lemat 5.

Niech $i_1 \in \mathbb{N}$ będzie takie, że zachodzi dla niego nierówność (2.22). Dla dostatecznie małego δ możliwy jest wybór takiego ε , że

$$\|\zeta_{i_1} - \zeta_{i_1+1}\|_2 > 2\delta. \quad (2.23)$$

Ponieważ założyliśmy, że $\tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}) \in [\zeta_{i_1}^1, \zeta_{i_1}^2]$, musimy rozpatrzeć dwa przypadki wyrażenia $\|\zeta_{i_1} - \zeta_{i_1+1}\|_2$

(a)

$$\|\zeta_{i_1} - \zeta_{i_1+1}\|_2 = \|(\zeta_{i_1}^1, \zeta_{i_1}^2, \zeta_{i_1}^3) - (\zeta_{i_1}^1, \tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}), \zeta_{i_1}^2)\|_2,$$

(b)

$$\|\zeta_{i_1} - \zeta_{i_1+1}\|_2 = \|(\zeta_{i_1}^1, \zeta_{i_1}^2, \zeta_{i_1}^3) - (\tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}), \zeta_{i_1}^2, \zeta_{i_1}^3)\|_2.$$

Rozważmy przypadek (a). Wyprowadzimy warunki na ε zapewniające, że zachodzą nierówności

$$\begin{aligned} & \|(\zeta_{i_1}^1, \zeta_{i_1}^2, \zeta_{i_1}^3) - (\zeta_{i_1}^1, \tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}), \zeta_{i_1}^2)\|_2^2 = \\ & (\zeta_{i_1}^2 - \tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}))^2 + (\zeta_{i_1}^3 - \zeta_{i_1}^2)^2 > (2\delta)^2 \quad (l \in \{1, 2, 3\}). \end{aligned} \quad (2.24)$$

Rozwiążemy powyższe nierówności względem ε używając transformacji $\hat{q}$ całej rodziny parabol z tym samym względnym położeniem minimum względem końców przedziału $\zeta_{i_1}^1$ i $\zeta_{i_1}^3$. Transformacja ta zdefiniowana jest w Dodatku A. Zastosowanie transformacji $\hat{q}$ pozwala uprościć powyższe nierówności. Przekształćmy więc parabolę skonstruowaną na punktach $(\zeta_{i_1}^1, f(\zeta_{i_1}^1))$, $(\zeta_{i_1}^2, f(\zeta_{i_1}^2))$ i $(\zeta_{i_1}^3, f(\zeta_{i_1}^3))$ używając transformacji $\hat{q}$. Transformacja $\hat{q}$ jest dobrze określona, gdy $f(\zeta_{i_1}^3) \neq f(\zeta_{i_1}^1)$ z zastrzeżeniem, że ponadto $a' \neq b'$ (patrz Dodatek A).

Rozpatrzmy przypadek kiedy $f(\zeta_{i_1}^3) > f(\zeta_{i_1}^1)$ ponieważ przypadek, gdy zachodzi $f(\zeta_{i_1}^1) > f(\zeta_{i_1}^3)$ jest symetryczny. Transformacja $\hat{q}$ przekształca nierówności (2.24) do

$$(\zeta_{(r)} - \tilde{\lambda}_l^*(\varepsilon_l, \tilde{\zeta}_{i_1}))^2 + (1 - \zeta_{(r)})^2 > \left(\frac{2\delta}{\zeta_{i_1}^3 - \zeta_{i_1}^1} \right)^2 \quad l \in \{1, 2, 3\}, \quad (2.25)$$

gdzie $\zeta_{(r)} = \frac{\zeta_{i_1}^2 - \zeta_{i_1}^1}{\zeta_{i_1}^3 - \zeta_{i_1}^1}$, $\tilde{\zeta}_{i_1} = (0, \zeta_{(r)}, 1)$ oraz $\tilde{\lambda}_l^*(\varepsilon_l, \tilde{\zeta}_i)$ dane są wzorami (6.8) gdzie

$$\begin{aligned} A &= -2(\zeta_{(r)} - 1), \\ B &= \hat{q}(\zeta_{(r)}), \\ C &= -\zeta_{(r)}, \end{aligned}$$

i są wyrażeniami na minimum parabol skonstruowanych na punktach $(0, -2)$, $(\zeta_{(r)}, \hat{q}(\zeta_{(r)}))$ oraz $(1, -1)$ zaburzonych w l -tym punkcie ($l \in \{1, 2, 3\}$) odpowiednio. Błąd względny aproksymacji ε w l -tym punkcie jest związany z przekształconym błędem względnym zależnością

$$\varepsilon_l = \begin{cases} \frac{2\varepsilon}{|f(\zeta_{i_1}^1)|}, & \text{dla zaburzenia w } \zeta_{i_1}^1, \\ \frac{\varepsilon \hat{q}(\zeta_{(r)})}{|f(\zeta_{i_1}^2)|}, & \text{dla zaburzenia w } \zeta_{i_1}^2, \\ \frac{\varepsilon}{|f(\zeta_{i_1}^3)|}, & \text{dla zaburzenia w } \zeta_{i_1}^3. \end{cases}$$

Rozwiązanie nierówności (2.25) musi być podzielone na dwa regiony.

- (a) Jeśli $\frac{4\delta^2}{(\zeta_{i_1}^3 - \zeta_{i_1}^1)^2} \leq (1 - \zeta_{(r)})^2$ to (2.25) są zawsze spełnione.
 (b) Jeśli $\frac{4\delta^2}{(\zeta_{i_1}^3 - \zeta_{i_1}^1)^2} > (1 - \zeta_{(r)})^2$ to (2.25) zapisują się jako

$$\begin{aligned} \pm|\zeta_{(r)} - \lambda_l^*(\varepsilon_l, \tilde{\zeta}_{i_1})| &> K_{i_1}, \\ K_{(a), i_1} &= \sqrt{\frac{4\delta^2}{(\zeta_{i_1}^3 - \zeta_{i_1}^1)^2} - (1 - \zeta_{(r)})^2}, \\ l &\in \{1, 2, 3\} \end{aligned}$$

a ponieważ transformacja $\hat{q}$ skaluje przedział liniowo i mamy $\tilde{\lambda}_l^*(\varepsilon_l, \tilde{\zeta}_i) < \zeta_{(r)}$, ponieważ $\tilde{\lambda}_l^*(\varepsilon_l, \tilde{\zeta}_i) \in (0, \zeta_{(r)})$, to otrzymujemy

$$\begin{aligned} \lambda_l^*(\varepsilon_l, \tilde{\zeta}_{i_1}) &< \zeta_{(r)} - K_{i_1}, \\ K_{(a), i_1} &= \sqrt{\frac{4\delta^2}{(\zeta_{i_1}^3 - \zeta_{i_1}^1)^2} - (1 - \zeta_{(r)})^2}, \\ l &\in \{1, 2, 3\}. \end{aligned} \tag{2.26}$$

Rozpatrzmy ostatnie nierówności dla $l = 1$ i $l = 2$ ponieważ dla $l = 3$ sytuacja jest symetryczna do $l = 1$. Stosując metodę rozwiązania powyższych nierówności przedstawioną w Dodatku B otrzymujemy, że jeśli

$$\varepsilon_1 > \frac{(B - C)(\zeta_{(r)} - K_a) - (B - \zeta_{(r)}C) - A(K_a - 1)}{A(K_a - 1)} \tag{2.27}$$

i ponadto

$$\begin{cases} A(1 + \varepsilon_1) + B - C > 0 \\ A(K_a - 1) < 0 \end{cases} \tag{2.28}$$

lub

$$\begin{cases} A(1 + \varepsilon_1) + B - C < 0 \\ A(K_a - 1) > 0 \end{cases} \quad (2.29)$$

zachodzą dla zaburzeń w ζ^1 spełniona jest nierówność (2.25) dla $l = 1$. Dla $l = 2$ nierówność (2.25) jest spełniona jeśli

$$\varepsilon_2 > \frac{A(1 + \zeta_{(r)}) + B - \zeta_{(r)}C - (A + B - C)(\zeta_{(r)} - K_a)}{B(\zeta_{(r)} - K_a - 1)} \quad (2.30)$$

i ponadto

$$\begin{cases} A + B(1 + \varepsilon_2) - C > 0 \\ B(\zeta_{(r)} - K_a - 1) > 0 \end{cases} \quad (2.31)$$

lub

$$\begin{cases} A + B(1 + \varepsilon_2) - C < 0 \\ B(\zeta_{(r)} - K_a - 1) < 0 \end{cases} \quad (2.32)$$

zachodzi dla zaburzenia w punkcie ζ^2 . Powyższe warunki implikują również, że spełnione są nierówności (2.24). Oznacza to, że jeśli zaburzenie wartości $f(\zeta^1)$ i $f(\zeta^2)$ spełniają (2.27) i (2.28) lub (2.29) i (2.30) i (2.31) lub (2.32) odpowiednia to dla dowolnego δ otrzymujemy nieskończenie wiele $i > i_1 > i_0$ takich, że zachodzi nierówność (2.23). W analogiczny sposób rozwiązujemy przypadki (b), tzn.

$$\begin{aligned} \|(\zeta_{i_1}^1, \zeta_{i_1}^2, \zeta_{i_1}^3) - (\tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}), \zeta_{i_1}^2, \zeta_{i_1}^3)\|_2^2 &= (\zeta_{i_1}^1 - \tilde{\lambda}_l^*(\varepsilon; \zeta_{i_1}))^2 \\ &> (2\delta)^2. \end{aligned} \quad (2.33)$$

Używając transformacji określonej w Dodatku A otrzymujemy

$$\lambda_l^*(\varepsilon_l; \tilde{\zeta}_i) > K_b \quad K_b = \frac{2\delta}{\zeta^3 - \zeta^1} \quad (i = \{1, 2, 3\}) \quad (2.34)$$

Stosując w analogiczny sposób metodę opisaną w Dodatku B dla $l = 1$ otrzymujemy

$$\varepsilon_1 > \frac{A[K_b - (1 + \zeta_{(r)})] + B(K_b - 1) - C(K_b - \zeta_{(r)})}{A[K_b - (1 + \zeta_{(r)})]} \quad (2.35)$$

i

$$\begin{cases} A(1 + \varepsilon_1) + B - C > 0 \\ A[(1 + \zeta_{(r)}) - K_b] > 0 \end{cases} \quad (2.36)$$

lub

$$\begin{cases} A(1 + \varepsilon_1) + B - C < 0 \\ A[(1 + \zeta_{(r)}) - K_b] < 0 \end{cases} \quad (2.37)$$

Dla $l = 2$ w ten sam sposób otrzymujemy

$$\varepsilon_2 > \frac{A[K_b - (1 + \zeta_{(r)})] + B(K_b - 1) - C(K_b - \zeta_{(r)})}{B(1 - K_b)} \quad (2.38)$$

$$\begin{cases} A + B(1 + \varepsilon_1) - C > 0, \\ B(1 - K_b) > 0, \end{cases} \quad (2.39)$$

$$\text{lub} \quad \begin{cases} A + B(1 + \varepsilon_1) - C < 0, \\ B(1 - K_b) < 0. \end{cases} \quad (2.40)$$

Oznacza to, że również i w tym przypadku dla dowolnego δ możemy wybrać takie ε , że zachodzi nierówność (2.23). Jest to sprzeczność z założeniem, że ciąg $\{\zeta_i\}_{i=0}^{\infty}$ jest ciągiem zbieżnym ponieważ możemy wybrać dwa podciągi, które są zbieżne do dwóch różnych punktów skupienia. Pierwszy podciąg jest złożony z ζ_{i_1} a drugi z ζ_{i_1+1} . Obydwa są ciągami nieskończonymi. Z tego wynika, że od pewnego i_0 ciąg $\{\zeta_i\}_{i=0}^{\infty}$ nie zawiera punktów z zaburzonymi wartościami funkcji tak więc począwszy od pewnego i_0 dowód z [25] pozostaje w mocy.

3. Załóżmy, że $\zeta \in \mathbf{T} \setminus \mathbf{S}$. Wówczas $\lambda^*(\zeta) \in (\zeta^1, \zeta^3)$. Ponieważ sytuacja jest symetryczna dla ζ^1 oraz ζ^3 rozważymy przypadek, w którym $\lambda^*(\zeta) \in (\zeta^1, \zeta^2]$ oraz $\tilde{\lambda}_l^*(\zeta) \in (\zeta^1, \zeta^2]$ ($l \in \{1, 2, 3\}$). Skorzystamy tu również z paraboli otrzymanej przez zastosowanie transformacji $\hat{q}$ z Dodatku A. Transformacja $\hat{q}$ zachowuje wszystkie własności konieczne w dowodzie. W rozważaniach skorzystamy z wyrażenia na $\Delta_l(\varepsilon; \zeta)$ będącego odległością pomiędzy zaburzonym i niezaburzonym minimum, tzn. $\Delta_l(\varepsilon; \zeta) = |\tilde{\lambda}_l^*(\zeta) - \lambda(\zeta)|$. Wzór na $\Delta_l(\varepsilon; \zeta)$ został wyprowadzony w Dodatku C. W przypadku tym mamy

- (a) $f(\zeta^2) \leq f(\zeta^3)$, jeśli zachodziłoby $f(\zeta^2) = f(\zeta^3)$, to $\tilde{\lambda}^*(\zeta) = 1/2(\zeta^2 + \zeta^3)$ ponieważ albo żadna z wartości nie jest zaburzona albo tylko wartość w punkcie ζ_1 jest zaburzona i zaburzenie to nie ma żadnego wpływu na położenie minimum. Obydwa przypadki są sprzeczne.
- (b) tylko $u_1(\zeta)$ i $\tilde{u}_1^l(\varepsilon; \zeta)$ ($l \in \{1, 2, 3\}$) oraz $u_3(\zeta)$ i $\tilde{u}_3^l(\zeta)$ ($l \in \{1, 2, 3\}$) mogą być w $A(\zeta)$. Dla niezaburzonych wartości funkcji otrzymujemy $\lambda^*(\zeta) < \zeta^2$ natomiast dla zaburzonych wartości funkcji mamy

$$\lambda_l^*(\varepsilon; \zeta) + \Delta_l(\varepsilon; \zeta) < \zeta^2 \quad (l \in \{1, 2, 3\}). \quad (2.41)$$

W Dodatku C rozwiązujemy (2.41) względem ε wyprowadzając warunki, dla których $\lambda_l^*(\varepsilon; \zeta) < \zeta^2$ co zapewnia, że dowód z [25] pozostaje w mocy. Otrzymujemy trzy nietrywialne przypadki

- i. $A(\zeta) = \{u_1(\zeta), \tilde{u}_1^1(\zeta), \tilde{u}_1^2(\zeta), \tilde{u}_1^3(\zeta)\}$. Ponieważ $u_3(\zeta) \notin A(\zeta)$ oraz $\tilde{u}_3^l(\varepsilon; \zeta) \notin A(\zeta)$ ($l \in \{1, 2, 3\}$) i $f(\lambda^*(\zeta)) < f(\zeta^2)$ oraz $f(\tilde{\lambda}_l^*(\varepsilon; \zeta)) < f(\zeta_2)$ ($l = 1, 3$)

$$f(\tilde{\lambda}_2^*(\varepsilon; \zeta)) < f(\zeta_2)(1 + |\varepsilon|) \quad (2.42)$$

musi zachodzić

$$\begin{aligned} c(u_1(\zeta)) &= f(\zeta^1) + f(\lambda^*(\zeta)) + f(\zeta^3) \\ &< f(\zeta^1) + f(\zeta^2) + f(\zeta^3) = c(\zeta), \end{aligned}$$

dla niezaburzonych wartości funkcji

$$\begin{aligned} c(\tilde{u}_1^l(\zeta)) &= \tilde{f}_l(\zeta^1) + f(\tilde{\lambda}_l^*(\zeta)) + \tilde{f}_l(\zeta^3) \\ &< \tilde{f}_l(\zeta^1) + \tilde{f}_l(\zeta^2) + \tilde{f}_l(\zeta^3) = c(\zeta), \end{aligned} \quad (2.43)$$

gdzie $\tilde{f}_l(\cdot) = f(\cdot)$ or $\tilde{f}_l(\cdot) = f(\cdot)(1+\varepsilon)$ w zależności od tego w którym punkcie wartość jest aproksymowana.

- ii. $A(\zeta) = \{u_3(\zeta), \tilde{u}_3^1(\zeta), \tilde{u}_3^2(\zeta), \tilde{u}_3^3(\zeta)\}$. Ponieważ $u_1(\zeta) \notin A(\zeta)$ oraz $\tilde{u}_1^l(\varepsilon; \zeta) \notin A(\zeta)$ ($l \in \{1, 2, 3\}$) musi zachodzić $f(\zeta^2) \leq f(\lambda^*(\zeta))$ oraz $f(\zeta^2) \leq f(\tilde{\lambda}_l^*(\zeta))$ ($l \in \{1, 2, 3\}$) zależnie od tego, w którym punkcie wartość funkcji była aproksymowana. Również zachodzi $f(\lambda^*(\zeta)) < f(\zeta_1)$ oraz $f_l(\tilde{\lambda}_l^*(\zeta)) < \tilde{f}_l(\zeta_1)$ ($l \in \{1, 2, 3\}$) ponieważ w przeciwnym razie istniałoby lokalne maksimum w przedziale $[\zeta^1, \zeta^2]$ co jest sprzeczne z unimodalnością. Tak więc w tym przypadku musi zachodzić

$$\begin{aligned} c(u_3(\zeta)) &= f(\lambda^*(\zeta)) + f(\zeta^2) + f(\zeta^3) \\ &< f(\zeta^1) + f(\zeta^2) + f(\zeta^3) = c(\zeta), \end{aligned}$$

dla niezaburzonych wartości funkcji oraz

$$\begin{aligned} c(\tilde{u}_3^l(\zeta)) &= f(\tilde{\lambda}_l^*(\zeta)) + \tilde{f}_2(\zeta^2) + \tilde{f}_3(\zeta^3) \\ &< \tilde{f}_l(\zeta^1) + \tilde{f}_l(\zeta^2) + \tilde{f}_l(\zeta^3) = c(\zeta), \end{aligned} \quad (2.44)$$

dla zaburzenia w punkcie ζ^l .

- iii. W końcu może zachodzić $A(\zeta) = \{u_1(\zeta), u_3(\zeta)\}$. W tym przypadku nie jesteśmy w stanie włączyć do $A(\zeta)$ żadnej trójki z punktem z wartością aproksymowaną $\tilde{u}_i(\varepsilon; \zeta)$ ($i = 1, 2, 3$). Wynika to z następujących własności

A. $f(\zeta_2) < f(\zeta_1)$ z założenia,

B. $\lambda^*(\zeta) \leq \zeta^2$,

C. $f(\lambda^*(\zeta)) = f(\zeta^2)$, co implikuje $\lambda^*(\zeta) = \zeta^2$.

Powyższe zależności zachodzą ponieważ w przeciwnym wypadku, mielibyśmy sprzeczność z unimodalnością funkcji $f(\cdot)$. Użycie aproksymacji wartości funkcji, w którymkolwiek z punktów oznaczałoby, że nie jesteśmy w stanie zagwarantować własności C. Wobec tego ponieważ $f(\zeta^2) < \min\{f(\zeta^1), f(\zeta^3)\}$ mamy $c(u_1(\zeta)) < c(\zeta)$ i $c(u_3(\zeta)) < c(\zeta)$ ¹. Z praktycznego punktu widzenia dla danego ζ_i , w którym jednej ze współrzędnych odpowiada wartość aproksymowana, możemy określić, czy możemy użyć wartości aproksymowanej sprawdzając warunek

$$|\tilde{\lambda}_l^*(\varepsilon; \zeta) - \zeta^2| > \Delta_l(\varepsilon; \zeta) \quad (l \in \{1, 2, 3\}). \quad (2.45)$$

¹Zauważmy, że w [25] własności B. i C. wyglądają nieco inaczej, mianowicie: B. $\lambda^*(\zeta) < \zeta^2$ i C. $f(\lambda^*(\zeta)) = f(\zeta^2)$. W szczególności własność B. w [25] zawiera błąd zawierając znak "<" zamiast " $\leq$ ".

To wyczerpuje wszystkie możliwości i kończy dowód trzeciego punktu. $\square$

Możemy teraz już sformułować algorytm Sekwencyjnej Interpolacji Parabolicznej z zaburzaniem (patrz [1] oraz [25] dla wersji bez zaburzeń).

Algorytm 2.

Algorytm Sekwencyjnej Interpolacji Parabolicznej z zaburzeniem wartości funkcji

Wejście : $\zeta_0 \in \mathbf{T}$ – punkt startowy,
 $\varepsilon > 0$ – błąd względny aproksymacji funkcji celu.

0. Ustaw $i = 0$.

1. Oblicz $\lambda^* = \lambda^*(\zeta_i)$ lub $\lambda^* = \tilde{\lambda}_l^*(\varepsilon; \zeta_i)$ w zależności od tego czy wartość funkcji jest obliczona bezpośrednio we wszystkich punktach $\zeta_i^1, \zeta_i^2, \zeta_i^3$ lub czy jest aproksymowana w punkcie ζ_i^l ($l \in \{1, 2, 3\}$) odpowiednio.

2. Jeśli $\lambda^* = \zeta_i^1$ or $\lambda^* = \zeta_i^3$ to koniec
W przeciwnym razie skonstruuj $A(\zeta_i)$.

(a) Jeśli aproksymacja wartości funkcji w żadnym punkcie trójki ζ_i nie jest dostępna to $A(\zeta_i) = A^0$ zgodnie ze wzorem (2.9).

(b) Jeśli aproksymacja wartości funkcji w punkcie ζ_i^l ($l \in \{1, 2, 3\}$) jest dostępna to

i. Oblicz transformację $\hat{q}$ określoną w Dodatku A.

ii. Oblicz $\Delta_l(\varepsilon; \zeta_i)$ zgodnie ze wzorem (6.9).

iii. Jeśli

$$|\lambda_l^*(\varepsilon; \zeta) - \zeta_i^2| < \Delta_l(\varepsilon; \zeta).$$

to $A(\zeta_i) = A^0$ to przejdź do kroku 3.

iv. Jeśli

$$\lambda_l^*(\varepsilon; \zeta) + \Delta_l(\varepsilon; \zeta) < \zeta_i^2 \quad \text{lub} \quad \lambda_l^*(\varepsilon; \zeta) - \Delta_l(\varepsilon; \zeta) > \zeta_i^2 \quad (2.46)$$

to $A = A^0 \cup \tilde{A}^l$

3. Oblicz

$$\zeta_{i+1} \in \arg \min \{c(\zeta) : \zeta \in A(\zeta_i)\}.$$

4. Podstaw $i := i + 1$ i przejdź do kroku 1.

Główny warunek określający czy można użyć dostępnej aproksymacji funkcji w jednym z punktów jest *własność separacji*, tzn. dla aproksymacji w l -tym punkcie $l \in \{1, 2, 3\}$ trójka ζ_i należy do $\mathbf{T}^l$ określonego przez (2.10), (2.11) lub (2.12) odpowiednio.

Rozważmy i -tą iterację, w której użyta jest aproksymacja funkcji w jednym z punktów ζ_i^1, ζ_i^2 or ζ_i^3 trójki ζ_i . Bez straty ogólności założmy, że $f(\zeta_2) < f(\zeta_1) < f(\zeta_3)$ oraz wartości funkcji w tych punktach są dostatecznie odseparowane tzn. w każdym z tych punktów może być użyta aproksymacja wartości funkcji zamiast jest dokładnej wartości i.e. $\tilde{\lambda}_l^*(\zeta) \in \mathbf{T}^l$ zależnie od tego, w którym punkcie wartość jest aproksymowana. Założmy ponadto, że aproksymacja jest użyta w jednym z punktów. Mamy trzy możliwości:

1. Jeśli aproksymacja wartości funkcji jest użyta dla ζ_i^3 to w iteracji $i+1$ ζ_i^3 nie wystąpi w trójce ζ_{i+1} i aproksymacja w $\tilde{\lambda}^*(\zeta_i)$ może zostać użyta.
2. Jeśli aproksymacja wartości funkcji jest użyta dla ζ_i^3 to w iteracji $i+1$ wartość $f(\zeta^1)$ będzie największa ponieważ ζ_i^3 wystąpi również w ζ_{i+1} zachowując również własność separacji. W ζ_{i+1} nie wystąpi natomiast ζ_i^3 .
3. Bardziej skomplikowaną sytuację mamy w wypadku aproksymacji w punkcie ζ_i^2 . W tej sytuacji nie mamy gwarancji, że własność separacji wartości funkcji w następnych dwóch iteracjach będzie zachowana. W tym wypadku $f(\lambda_l^*(\zeta_i))$ musi być obliczona bezpośrednio. Wówczas jeśli własność separacji wartości funkcji dla trójki ζ_{i+1} nie jest zachowana to nie możemy użyć aproksymacji w punkcie ζ_i^2 . Ponieważ $f(\lambda_l^*(\zeta_i))$ jest obliczane bezpośrednio nowa trójka ma postać

$$\zeta_i^{\text{new}} := (\zeta_i^1, \lambda_l^*(\zeta_i), \zeta_i^3). \quad (2.47)$$

Odpowiadające wartości funkcji są dokładne. Zauważmy, że w ten sposób zachowujemy własność 3. z Lematu 4. Poprawa wartości funkcji kosztu jest na poziomie $\varepsilon f(\zeta_i^2)$ a ponieważ w $c(\zeta_{i-1})$ własność separacji była zachowana lub nie użyto aproksymacji wartości funkcji to otrzymujemy

$$c(\zeta_{i-1}) \geq c(\zeta_i^{\text{new}}).$$

Założmy więc, że używamy aproksymacji wartości funkcji $f(\zeta_i^2)$ w ζ_i i ζ_{i+1} a następnie okazuje się, że dla ζ_{i+2} nie jest zachowana własność separacji. Wtedy nie bierzemy ani minimum z kroku i -tego ani $i+1$ -szego do skonstruowania ζ_i^{new} . W takiej sytuacji definiujemy

$$\begin{aligned} \zeta_i^{\text{new}} &:= (\zeta_i^1, \lambda_l^*(\zeta_i), \zeta_i^3) & \text{jeśli} & \quad f(\lambda_l^*(\zeta_i)) < f(\lambda_l^*(\zeta_{i+1})), \\ \zeta_i^{\text{new}} &:= (\zeta_i^1, \lambda_l^*(\zeta_{i+1}), \zeta_i^3) & \text{jeśli} & \quad f(\lambda_l^*(\zeta_i)) > f(\lambda_l^*(\zeta_{i+1})). \end{aligned} \quad (2.48)$$

Sterując użyciem aproksymacji funkcji zgodnie z powyższymi punktami zachodzą tezy Lematu 4 i ponadto $\zeta^* \in [\zeta_i^1, \zeta_i^3]$ oraz aproksymacja w jednym punkcie może być użyta do wygenerowania nawet trzech trójek.

Zbieżności powyższego algorytmu dowodzimy przez dowód twierdzenia (patrz [1]) analogicznego do twierdzenia dla algorytmu bez zaburzenia wartości funkcji (patrz [25] p. 155).

Twierdzenie 3.

Założmy, że $\{\zeta_i\}_{i=0}^{\infty}$ jest ciągiem skonstruowanym za pomocą Algorytmu 2 dla różniczkowalnej w sposób ciągły unimodalnej funkcji celu $f : \mathbb{R} \rightarrow \mathbb{R}$ ze sterowaniem użycia aproksymacji funkcji celu za pomocą strategii (2.47) i (2.48). Wówczas $\zeta_i \rightarrow \hat{\zeta}$ przy $i \rightarrow \infty$ gdzie $\hat{\zeta} \in S$.

Dowód jaki prezentujemy w tej pracy polega na pokazaniu przy jakich założeniach o błędzie aproksymacji funkcji celu oraz o miejscach jej użycia można zastosować dowód z [25].

Dowód

Podstawową trudnością w zastosowaniu metody dowodu z [25] jest fakt, że użycie aproksymacji w pewnych punktach dziedziny powoduje nieciągłości w funkcji celu c jak również w funkcjonale obliczania minimum paraboli skonstruowanej dla trójki ζ_i .

Zauważmy najpierw, że teza z punktu 3. w Lemacie 4 pokazuje, że dopuszczenie zaburzenia wartości funkcji zgodnie z (2.46) zapewnia, że $\{\zeta_i^1\}_{i=0}^{\infty}$ jest monotonicznie rosnący jak również $\{\zeta_i^3\}_{i=0}^{\infty}$ jest monotonicznie malejący. Ponieważ, obydwa te ciągi są ograniczone są więc zbieżne. Ponadto zapewniając, że $\Delta_l(\varepsilon; \zeta_i)$ jest takie, że powyższe własności byłyby zachowane gdybyśmy nie użyli aproksymacji, zapewnia nam, że $\tilde{\lambda}_l^*(\varepsilon, \zeta) \in [\zeta_i^1, \zeta_i^3]$ dla dowolnego $i \in \mathbb{N}$. Musimy rozpatrzyć dwa przypadki

1. Gdy $\{\zeta_i\}_{i=1}^{\infty} \rightarrow \hat{\zeta}$ i $\hat{\zeta} = (\hat{\zeta}^1, \hat{\zeta}^2, \hat{\zeta}^3)$ jest takim punktem skupienia, że $\hat{\zeta}^1 < \hat{\zeta}^2 < \hat{\zeta}^3$. Otrzymujemy tu sprzeczność używając ciągłości funkcji celu c jeśli użyty jest algorytm bez aproksymacji funkcji. Argument ciągłości może być użyty tutaj na podstawie dowodu Lematu 4, który pokazuje, że aproksymacja funkcji może być użyta tylko w skończonej liczbie iteracji. Po usunięciu pewnej liczby iteracji początkowych dowód z [25] pozostaje w mocy.
2. Nie zachodzi pierwszy przypadek, tzn., ciąg skonstruowany przez 2 może mieć dwa punkty skupienia. W [25] pokazano, że jest to ten sam punkt skupienia. Lemma 4 zapewnia, że argument zawarty w [25] jest w mocy jeśli algorytm kończy działanie w kryterium stopu w trójce bez punktów, w których wartości funkcji są aproksymowane. Ponieważ taki ciąg nie jest nieskończony musimy rozpatrzyć jeszcze jeden dodatkowy przypadek, tzn. gdy algorytm zatrzymuje się w trójce dla której w jednym z punktów wartość funkcji jest aproksymowana.

Dwoma punktami skupienia są $\zeta_* = (\hat{\zeta}^1, \hat{\zeta}^1, \hat{\zeta}^3)$ i $\zeta_{**} = (\hat{\zeta}^1, \hat{\zeta}^3, \hat{\zeta}^3)$. Aby zapewnić własność separacji wartości funkcji wartość funkcji może być aproksymowana tylko w punkcie $\hat{\zeta}^3$. W drugim przypadku wartość funkcji może być aproksymowana tylko w punkcie $\hat{\zeta}^1$. W powyższych ciągach mamy więc $\zeta_i^2 \rightarrow \zeta_i^1$ or $\zeta_i^2 \rightarrow \zeta_i^3$. Z drugiej strony, separacja między $\lambda^*(\varepsilon; \zeta_i)$ i ζ_i^2 musi być większa niż $\Delta_3(\varepsilon; \zeta_i)$ i $\Delta_1(\varepsilon; \zeta_i)$ dla odpowiednich przypadków. To daje sprzeczność ze zbieżnością.

□

2.5 Procedura ε -check

Po wyprowadzeniu warunków określających dopuszczalne zaburzenie wartości funkcji w transformacji minimalizującej funkcję wzdłuż kierunku zrealizowane za pomocą zaburzonej sekwencyjnej interpolacji kwadratowej możemy w formie algorytmicznej przedstawić procedurę ε -check z kroku 3.c) Algorytmu 1.

Algorytm 3.

Procedura ε -check – procedura sprawdzenia czy zaburzenie ε wartości funkcji zachowuje warunki zbieżności metody SPELROA

W wejście :	ε	– błąd względny aproksymacji,
	$\mathbf{x}_{k-2}, \mathbf{x}_{k-1}, \mathbf{x}_k$	– trzy ostatnie punkty wygenerowane przez metodę SPELROA połączoną z algorytmem optymalizacji bezpośredniej A,
	$\tilde{f}(\mathbf{x}_{k-2}), \tilde{f}(\mathbf{x}_{k-1}), \tilde{f}(\mathbf{x}_k)$	– odpowiadające wartości funkcji, aproksymacja funkcji może wystąpić conajwyżej w jednym z punktów.
W wyjście :	true	– jeśli warunki zbieżności są zachowane,
	false	– w przeciwnym razie.

0. Jeśli punkty nie są współliniowe i w procesie optymalizacyjnym nie są punktami iteracji wejściowymi do Algorytmu Sekwencyjnej Interpolacji Parabolicznej to zwróć **false**.

1. Oznaczmy $\zeta_k = (0, t_1, t_2)$ oraz $t_1 = \frac{\mathbf{x}_{k-2} - \mathbf{x}_{k-1}}{\mathbf{x}_{k-2} - \mathbf{x}_k}$, $t_2 = 1$.

2. Załóżmy, że $\tilde{f}(\mathbf{x}_{k-2}) < \tilde{f}(\mathbf{x}_k)$ (przypadek, gdy $\tilde{f}(\mathbf{x}_{k-2}) > \tilde{f}(\mathbf{x}_k)$ jest symetryczny względem $\zeta^0 = 1/2$ i wówczas rozpatrujemy $\zeta_k = (0, 1 - t_1, 1)$ z odpowiadającą trójką wartości $(\tilde{f}(\mathbf{x}_k), \tilde{f}(\mathbf{x}_{k-1}), \tilde{f}(\mathbf{x}_{k-2}))$).

Obliczmy transformację $\hat{q}$ określoną przez (6.1) aby otrzymać punkt $(\zeta(r), f(\zeta(r)))$. Przeskalujmy ε zależnie od tego, w którym punkcie jest użyta aproksymacja wg

$$\varepsilon := \begin{cases} \frac{2\varepsilon}{|f(\mathbf{x}_{k-2})|}, & \text{gdy aproksymacja jest użyta w } \mathbf{x}_{k-2}, \\ \frac{\varepsilon \hat{q}(t_1)}{|f(\mathbf{x}_{k-1})|}, & \text{gdy aproksymacja jest użyta w } \mathbf{x}_{k-1}, \\ \frac{\varepsilon}{|f(\mathbf{x}_k)|}, & \text{gdy aproksymacja jest użyta w } \mathbf{x}_k. \end{cases}$$

3. (a) Gdy aproksymacja wartości funkcji występuje w punkcie $\mathbf{x}_{k-2}$, to jeśli ε spełnia (6.11) to zwróć **true** w przeciwnym razie **false**.

(b) Gdy aproksymacja wartości funkcji występuje w punkcie $\mathbf{x}_{k-1}$, to jeśli ε spełnia (6.14) to zwróć **true** w przeciwnym razie **false**.

- (c) Gdy aproksymacja wartości funkcji występuje w punkcie $\mathbf{x}_k$, to jeśli ε spełnia (6.15) to zwróć **true** w przeciwnym razie **false**.

Zauważmy, że transformacja $\hat{q}$ jest tak skonstruowana, że $\hat{q}(x) < -1$ dla $x \in (0, 1)$. To gwarantuje, że zaburzenie względne wartości funkcji jest interpretowalne, ponieważ nigdy nie wystąpi osobliwość, którą byśmy mieli gdyby $\hat{q}(x) = 0$ dla pewnego $x \in (0, 1)$.

Rozdział 3

Przybliżanie funkcji ciągłych $f : \mathbb{R}^d \rightarrow \mathbb{R}$ sieciami RBF

W rozdziale tym zajmiemy się dwoma zagadnieniami – interpolacją oraz aproksymacją funkcji ciągłych d zmiennych. W obydwu zagadnieniach idzie o przybliżenie funkcji ciągłej $f : \mathbb{R}^d \rightarrow \mathbb{R}$, która zadana jest za pomocą zbioru dyskretnego $\mathbf{Z} = \{\mathbf{x}_i, y_i\}_{i=1}^N$.

3.1 Interpolacja oraz aproksymacja za pomocą regularyzacji Tikhonova

W zagadnieniu interpolacji zbiór $\mathbf{Z}$ jest taki, że

$$f(\mathbf{x}_i) = y_i, \quad i = 1, \dots, N. \quad (3.1)$$

Natomiast w wypadku aproksymacji dopuszczamy zaburzenie wartości funkcji f na poziomie ε , tzn.:

$$f(\mathbf{x}_i) = y_i + \delta_i, \quad \text{gdzie } |\delta_i| \leq \varepsilon \quad i = 1, \dots, N. \quad (3.2)$$

Rozwiązania powyższych zagadnień poszukujemy w zbiorze funkcji postaci

$$s_{f,Z}(\mathbf{x}) = \sum_{i=1}^N w_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) + \sum_{j=1}^Q b_j p_j(\mathbf{x}) \quad (3.3)$$

gdzie ϕ jest bazową funkcją radialną natomiast $\{p_j\}_{j=1}^Q$, $Q = \binom{m+d-1}{d}$ jest bazą przestrzeni wielomianów d zmiennych stopnia nie wyższego niż m .

Warunki istnienia oraz jednoznaczności rozwiązania zagadnienie interpolacji sprowadzają się do odpowiedzi na pytanie o odwracalność macierzy układu

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix} \quad (3.4)$$

gdzie $A = (\phi_j(\mathbf{x}_i))_{i,j=1,\dots,N}$ i $P = (p_j(\mathbf{x}_i))_{i=1,\dots,N,j=0,\dots,Q}$. Warunki jakie musi spełniać zbiór $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, funkcja ϕ oraz stopień m przestrzeni wielomianów $\pi_m(\mathbb{R}^d)$ po raz pierwszy zostały podane w [22].

Metodą rozwiązania zagadnienia aproksymacyjnego, jaką zajmujemy się w niniejszym rozdziale, jest metoda regularyzacji Tikhonova a jej korzenie sięgają prac nad problemami źle postawionymi (patrz [38] oraz referencje tamże). Zregularyzowane rozwiązanie zagadnienia aproksymacyjnego jest poszukiwane w przestrzeni Hilberta $\mathcal{H}$. Operator ograniczeń jest definiowany za pomocą innej przestrzeni unormowanej $\mathcal{G}$, dla której określony jest liniowy operator $L : \mathcal{H} \rightarrow \mathcal{G}$. Operator ograniczeń $J : \mathcal{H} \rightarrow \mathbb{R}$ określony jest wówczas jako $J(s) = \|Ls\|_{\mathcal{G}}^2$. Zagadnienie regularyzacji polega na znalezieniu dla ustalonej wartości parametru $\lambda > 0$, funkcji $s^* \in \mathcal{H}$ postaci (3.3), która jest rozwiązaniem zagadnienia minimalizacyjnego

$$\min_{s \in \mathcal{H}} \left\{ \sum_{j=1}^N [f(\mathbf{x}_j) - s(\mathbf{x}_j)]^2 + \lambda J(s) \right\}. \quad (3.5)$$

Jeśli przyjmiemy, że $J(s) = \|s\|_{\mathcal{G}}^2$ to zagadnienie to sprowadza się do rozwiązania układu

$$\begin{pmatrix} A + \lambda I & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix}, \quad (3.6)$$

dla ustalonej wartości parametru $\lambda > 0$. Warunki istnienia i jednoznaczności rozwiązania tego zagadnienia podała Wahba [40]. Oczywiście kluczowym zagadnieniem jest tu wybór wartości parametru λ . Dalej omówimy trzy metody wyboru wartości tego parametru – dwie znane z literatury oraz jedną wprowadzoną przez nas.

3.2 Radialne funkcje bazowe

Niech $\pi_m(\mathbb{R}^d)$ oznacza przestrzeń wielomianów d zmiennych stopnia nie większego niż m . Rozróżniamy radialne funkcje bazowe dodatnio określone oraz warunkowo dodatnio określone. Podajemy za [9] definicję obu klas.

Definicja 3.

Mówimy, że funkcja $\phi : \mathbb{R} \rightarrow \mathbb{R}$ jest całkowicie monotoniczna, jeśli

$$(-1)^r \phi^{(r)} \geq 0 \quad r = 1, 2, \dots \quad (3.7)$$

Twierdzenie 4. (patrz [9])

Przy założeniu, że punkty w zbiorze $\mathbf{X}$ są różne, macierz A jest dodatnio określona, jeśli funkcja ϕ z centrum w $\mathbf{c} \in \mathbb{R}^d$ jest taka, że dla $r = \|\mathbf{x} - \mathbf{c}\|$, funkcja $\phi(\sqrt{r})$ jest całkowicie monotoniczna i nie stała.

Definicja 4.

Mówimy, że funkcja $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ jest warunkowo dodatnio określona stopnia k , jeśli dla każdego zbioru punktów $\mathbf{X} \subset \mathbb{R}^d$ forma kwadratowa

$$\sum_{i=1}^N \sum_{k=1}^N w_i w_k \phi(\mathbf{x}_i - \mathbf{x}_k) \quad (3.8)$$

jest nieujemna dla wszystkich niezerowych $\{w_i\}_{i=1,\dots,N}$, które spełniają

$$\sum_{i=1}^N w_i p(\mathbf{x}_i) = 0, \quad (3.9)$$

gdzie $p(\mathbf{x})$ jest dowolnym wielomianem d zmiennych stopnia nie większego niż m . Mówimy, że funkcja ϕ jest ściśle warunkowo dodatnio określona stopnia m , jeśli powyższa forma kwadratowa jest zawsze dodatnia.

Twierdzenie 5. (patrz [9])

Dla funkcji ϕ , które są warunkowo dodatnio określone stopnia m mamy, że funkcja

$$(-1)^k \phi^{(k)}(\sqrt{t}) \quad (3.10)$$

jest całkowicie monotoniczna.

Najczęściej używanymi w praktyce funkcjami radialnymi w rozwiązaniu zagadnień interpolacji i aproksymacji funkcji są: spośród funkcji ściśle dodatnio określonych tzn. takie, że $m = 0$ (patrz [43], [32])

$$\begin{aligned} \phi(r) &= (r^2 + c^2)^\beta, & \beta < 0, \\ \phi(r) &= e^{-\alpha r^2}, & \alpha > 0, \\ \phi(r) &= (1 - r)_+^4 (4r + 1), \end{aligned}$$

oraz spośród warunkowo dodatnio określonych stopnia m funkcje

$$\begin{aligned} \phi(r) &= (-1)^{\beta/2+1} r^\beta \log r, & \beta \in 2\mathbb{N}, & \text{gdzie } m = \beta/2 + 1, \\ \phi(r) &= r^\beta & \beta \in \mathbb{R}_{>0}/2\mathbb{N}, & \text{gdzie } m = \lceil \beta/2 \rceil + 1, \\ \phi(r) &= (-1)^{\lceil \beta \rceil} (r^2 + c^2)^\beta, & \beta > 0 & \text{gdzie } m = \lceil \beta \rceil. \end{aligned}$$

Cechą charakterystyczną funkcji ściśle dodatnio określonych jest ich wykres w kształcie „dzwonu” (ang. *bell-shaped*) oraz $\lim_{r \rightarrow \infty} \phi(r) = 0$. Dla funkcji warunkowo dodatnio określonych czyli takich, że $m > 0$ mamy natomiast $\lim_{r \rightarrow \infty} \phi(r) \rightarrow \infty$.

Dla funkcji warunkowo dodatnio określonych aby macierz układu (3.6) była macierzą nieosobliwą wymagane jest, aby zbiór $\mathbf{X}$ spełniał warunek unisolwentności względem przestrzeni $\pi_m(\mathbb{R}^d)$.

Definicja 5. (unisolwentność zbioru $\mathbf{X}$)

Mówimy, że zbiór punktów $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \mathbb{R}^d$, gdzie $N > Q = \dim \pi_m(\mathbb{R}^d)$ jest unisolwentny względem przestrzeni $\pi_m(\mathbb{R}^d)$ jeśli wielomian zerowy jest jedynym wielomianem z przestrzeni $\pi_m(\mathbb{R}^d)$, który znika na wszystkich punktach $\mathbf{X}$ jednocześnie.

Warunek ten mówi, że zbiór punktów $\mathbf{X}$ na tyle dobrze próbkuje przestrzeń $\mathbb{R}^d$, w każdym wymiarze, że możliwe jest skonstruowanie wielomianu interpolacyjnego d zmiennych pełnego stopnia m . Przykładem zbioru, który nie jest unisolwentny względem $\pi_2(\mathbb{R}^d)$ jest zbiór 6 punktów położonych na okręgu. Na takim zbiorze nie można zbudować kwadratowego wielomianu interpolacyjnego.

3.3 Przestrzeń natywna

Przestrzeń funkcji, które można aproksymować za pomocą rozwinięcia (3.3) dla ustalonej (warunkowo) dodatnio określonej funkcji radialnej $\Phi(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|_2)$ jest przestrzenią Hilberta $\mathcal{H}$ z ustalonym iloczynem skalarnym $(\cdot, \cdot)_{\mathcal{H}}$ zawierającą samoreprodukującą funkcją jądrową. Ograniczenia błędu aproksymacji jakimi się zajmujemy w następnym paragrafie są postaci

$$|f - s_{f, \mathbf{Z}}| \leq CF(h(x)) \|f\|_{\mathcal{H}}$$

gdzie norma $\|\cdot\|_{\mathcal{H}}$ jest normą generowaną przez iloczyn skalarny $(\cdot, \cdot)_{\mathcal{H}}$, $C > 0$ jest stałą, $h(\mathbf{x})$ jest miarą gęstości danych a F jest funkcją ciągłą.

Niniejszy paragraf został przygotowany na podstawie [32] i [33].

3.3.1 Przestrzeń Hilberta z samoreprodukującą funkcją jądrową

Niech $\Omega \subseteq \mathbb{R}^d$ będzie dziedziną funkcji rzeczywistych, tworzących przestrzeń Hilberta $\mathcal{H}$ z iloczynem skalarnym $(\cdot, \cdot)_{\mathcal{H}}$. Załóżmy ponadto, że dla każdego $\mathbf{x} \in \Omega$ funkcyjonał wartościowania w punkcie $\delta_{\mathbf{x}} : f \rightarrow f(\mathbf{x})$ jest ciągły w $\mathcal{H}$ i.e.

$$\delta_{\mathbf{x}} \in \mathcal{H}^* \quad \text{dla każdego } \mathbf{x} \in \Omega,$$

gdzie $\mathcal{H}^*$ jest przestrzenią dualną do $\mathcal{H}$, tzn. przestrzenią operatorów liniowych z $\mathcal{H}$ w $\mathbb{R}$.

Twierdzenie 6. (patrz [32])

Jeśli przestrzeń Hilberta funkcji określonych na zbiorze Ω zawiera ciągły funkcyjonał wartościowania w punkcie to zawiera symetryczną funkcję nazywaną reprodukującą funkcją jądrową $\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$ o własnościach

$$\begin{aligned} \Phi(\mathbf{x}, \cdot) &\in \mathcal{H} \\ f(\mathbf{x}) &= (f, \Phi(\mathbf{x}, \cdot))_{\mathcal{H}} \\ \Phi(\mathbf{x}, \mathbf{y}) &= (\Phi(\mathbf{x}, \cdot), \Phi(\mathbf{y}, \cdot))_{\mathcal{H}} = \Phi(\mathbf{y}, \mathbf{x}) \\ \Phi(\mathbf{x}, \mathbf{y}) &= (\delta_{\mathbf{x}}, \delta_{\mathbf{y}})_{\mathcal{H}^*} = \Phi(\mathbf{y}, \mathbf{x}) \end{aligned}$$

dla wszystkich $\mathbf{x}, \mathbf{y} \in \Omega$, $f \in \mathcal{H}$.

3.3.2 Przestrzeń natywna

Definicja 6.

Jeśli symetryczna (warunkowo) dodatnio określona funkcja $\Phi : \Omega \times \Omega \rightarrow \mathbb{R}$ jest reprodukującą funkcją jądrową w przestrzeni Hilberta $\mathcal{H}$ funkcji o wartościach rzeczywistych określonych na zbiorze Ω , to przestrzeń $\mathcal{H}$ nazywamy przestrzenią natywną generowaną przez funkcję jądrową Φ i oznaczamy ją $\mathcal{N}_\Phi(\mathbb{R}^d)$.

Jednoznaczność przestrzeni natywnej generowanej przez funkcję Φ opisywana jest następującym twierdzeniem.

Twierdzenie 7. (patrz [32])

Przestrzeń natywna $\mathcal{H} = \mathcal{N}_\Phi(\mathbb{R}^d)$ generowana przez daną (warunkowo) dodatnio określoną funkcję Φ , o ile istnieje, jest określona jednoznacznie. Ponadto pokrywa się z domknięciem przestrzeni skończonych kombinacji liniowych funkcji $\Phi(\mathbf{x}, \cdot)$ względem iloczynu skalarnego

$$(\Phi(\mathbf{x}, \cdot), \Phi(\mathbf{y}, \cdot))_{\mathcal{H}} = \Phi(\mathbf{x}, \mathbf{y}) \quad \text{dla każdego } \mathbf{x}, \mathbf{y} \in \Omega.$$

Twierdzenie 7 mówi, że przestrzeń natywna generowana przez radialną funkcję bazową Φ jest domknięciem przestrzeni składającej się z funkcji postaci

$$f_{\mathbf{w}} := \sum_{j=1}^N w_j \Phi(\cdot - \mathbf{x}_j), \quad w_j \in \mathbb{R} \quad j = 1, \dots, N.$$

Iloczyn skalarny w tej przestrzeni zdefiniowany jest jako

$$(\cdot, \cdot)_{\mathcal{H}} = (f_{\mathbf{w}^{(1)}}, f_{\mathbf{w}^{(2)}})_{\Phi} := \sum_{j=1}^N \sum_{i=1}^M w_j^{(1)} w_i^{(2)} \Phi(\mathbf{y}_i - \mathbf{x}_j) \quad (3.11)$$

gdzie $f_{\mathbf{w}^{(1)}}$ jest funkcją interpolującą funkcję f na zbiorze $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ natomiast $f_{\mathbf{w}^{(2)}}$ jest funkcją interpolującą funkcję f na zbiorze $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_M\}$. Do udowodnienia ograniczeń błędów, które przytaczamy w następnym paragrafie potrzebne jest ponadto założenie o funkcji f , że jej transformacja Fouriera $\hat{f}$ jest zdominowana przez transformację Fouriera $\hat{\Phi}$ funkcji $\Phi(r) = \phi(\|r\|_2)$ w sensie

$$\int |\hat{f}|^2 \hat{\Phi}^{-1} dt < \infty. \quad (3.12)$$

Założenie to oznacza, że iloczyn skalarny (3.11) można zapisać w postaci całki Lebesgue'a

$$\begin{aligned} (\cdot, \cdot)_{\mathcal{H}} &= (f_{\mathbf{w}^{(1)}}, f_{\mathbf{w}^{(2)}})_{\Phi} := \sum_{j=1}^N \sum_{i=1}^M w_j^{(1)} w_i^{(2)} \Phi(\mathbf{y}_i - \mathbf{x}_j) \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{\Phi}(\omega) \left(\sum_{j=1}^N w_j^{(1)} e^{i\mathbf{x}_j^T \omega} \right) \left(\sum_{k=1}^M w_k^{(2)} e^{-i\mathbf{y}_k^T \omega} \right) d\omega. \end{aligned}$$

Mając daną dodatkowo określoną funkcję Φ interesuje nas więc pytanie, jakie funkcje należą do przestrzeni natywnej generowanej przez funkcję Φ . Odpowiedź w ogólności można znaleźć w [32]. My ograniczymy się do przestrzeni, które zawierają funkcje, które mogą być funkcjami celu w procesie optymalizacyjnym. Aby można było użyć algorytmu bezgradientowego funkcja celu musi być co najmniej ciągła. Algorytmy gradientowe lub drugiego rzędu wymagają ciągłości pochodnych kierunkowych oraz cząstkowych drugiego rzędu odpowiednio.

Dalej omówimy więc przestrzeń natywną, która zawiera funkcje całkowalne z pewną potęgą, przestrzenią taką, jest przestrzeń Sobolewa. Pokażemy jakie dodatkowo określone funkcje generują przestrzeń natywną izomorficzną z przestrzeniami Sobolewa odpowiedniego rzędu. Przestrzeń Sobolewa odpowiednich rzędów zawiera przestrzenie $C(\mathbb{R}^d)$, $C^1(\mathbb{R}^d)$ i $C^2(\mathbb{R}^d)$. Na koniec zajmiemy się przestrzenią natywną generowaną przez funkcje Gaussa. Jest to przestrzeń często używana w zastosowaniach praktycznych. Również w naszym zastosowaniu aproksymacja jest konstruowana przy użyciu funkcji Gaussa.

3.3.3 Przestrzeń Sobolewa

Przestrzeń Sobolewa definiujemy za pomocą operatora różniczkowania kierunkowego D^{α} ,

$$D^{\alpha} = \left[\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}}, \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}}, \dots, \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} \right]$$

gdzie $\alpha = \alpha_1 + \alpha_2 + \dots + \alpha_d$ i $\alpha_i \in \mathbb{N}$.

Definicja 7. (Przestrzeń Sobolewa całkowitego rzędu)

Przestrzeń Sobolewa $W_p^k(\Omega)$ określamy jako przestrzeń funkcji $u : \Omega \rightarrow \mathbb{R}$, dla których $D^{\alpha}u \in L_p(\Omega)$, $|\alpha| \leq k$. Jest to (semi-)przestrzeń Hilberta z (semi-)normą określoną odpowiednio

$$|u|_{W_p^k(\Omega)} := \left(\sum_{|\alpha|=k} \|D^{\alpha}u\|_{L_p(\Omega)}^p \right)^{1/p} \quad \text{oraz} \quad \|u\|_{W_p^k(\Omega)} := \left(\sum_{|\alpha| \leq k} \|D^{\alpha}u\|_{L_p(\Omega)}^p \right)^{1/p}.$$

Definicja 8. (Przestrzeń Sobolewa ułamkowego rzędu)

Przestrzeń Sobolewa $W_p^{k+s}(\Omega)$, $1 \leq p < \infty$, $k \in \mathbf{N}_0$, $0 < s < 1$ określamy jako przestrzeń funkcji $u : \Omega \rightarrow \mathbb{R}$, dla których poniższe normy są skończone

$$|u|_{W_p^{k+s}(\Omega)} := \left(\sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{|D^{\alpha}u(\mathbf{x}) - D^{\alpha}u(\mathbf{y})|^p}{\|\mathbf{x} - \mathbf{y}\|_2^{d+ps}} d\mathbf{x}d\mathbf{y} \right)^{1/p}, \quad (3.13)$$

$$\|u\|_{W_p^{k+s}(\Omega)} := \left(\|u\|_{W_p^k(\Omega)}^p + |u|_{W_p^{k+s}(\Omega)}^p \right)^{1/p}. \quad (3.14)$$

Założmy, że $\Phi \in L_1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ spełnia

$$c_1(1 + \|\omega\|_2^2)^{-\tau} \leq \hat{\Phi} \leq c_2(1 + \|\omega\|_2^2)^{-\tau}, \quad \omega \in \mathbb{R}^d \quad (3.15)$$

dla $\tau \in \mathbb{R}$ i $\tau > d/2$ oraz dwóch stałych dodatnich $c_1 < c_2$. Wówczas przestrzeń natywna $\mathcal{N}_{\Phi}(\mathbb{R}^d)$ odpowiadająca funkcji Φ pokrywa się z przestrzenią Sobolewa $W_2^{\tau}(\Omega)$, i norma w przestrzeni natywnej i norma w przestrzeni Sobolewa są równoważne.

Funkcjami Φ , które spełniają powyższy warunek są ściśle dodatnio określone funkcje Wendlanda (patrz [43]) oraz np. funkcje typu *thin plate splines* $\Phi(r) = (-1)^{\beta+1} \|r\|^{2\beta} \log \|r\|$, $\beta \in 2\mathbb{N}$, dla których

$$\hat{\Phi}(\omega) = 2^{d+2\beta-1} \pi^{d/2} \Gamma(d/2 + \beta) \beta! \|\omega\|^{-d-2\beta}$$

(patrz [33]).

3.3.4 Przestrzeń natywna dla funkcji Gaussa

Dla funkcji Gaussa $\Phi(r) = e^{-\alpha \|r\|^2}$ mamy

$$\hat{\Phi}(\omega) = \left(\frac{\pi}{\alpha} \right)^{d/2} e^{-\|\omega\|^2/(4\alpha)}.$$

Wydaje się, że taka postać transformacji Fouriera funkcji Φ może być argumentem przeciw aproksymacji za pomocą funkcji Gaussa, gdyż każda funkcja z przestrzeni natywnej $\mathcal{N}_{\Phi}(\mathbb{R}^d)$ funkcji musi być zdominowana przez transformację Fouriera $\hat{\Phi}$ w sensie (3.12). Warunek ten spełniają np. wszystkie funkcje o ograniczonym widmie, a przestrzeń tych funkcji odgrywa istotną rolę w teorii próbkowania, w szczególności w twierdzeniu Shannona o próbkowaniu (patrz [42]).

3.4 Regularyzacja Tikhonova

Zagadnienie treningu sieci z radialnymi funkcjami bazowymi jest zagadnieniem źle postawionym (ang. *ill-posed*) w sensie Hadamarda [17]. Zadanie jest zadaniem źle postawionym jeśli nie jest zadaniem dobrze postawionym, czyli gdy nie jest spełniony co najmniej jeden z warunków następującej definicji.

Definicja 9.

Zadanie nazywamy dobrze postawionym jeśli zachodzą następujące warunki:

1. *rozwiązanie zadania istnieje,*
2. *rozwiązanie zadania jest jednoznaczne,*
3. *rozwiązanie zadania zależy w sposób ciągły od danych.*

Zadanie treningu sieci z radialnymi funkcjami bazowymi przez rozwiązanie zagadnienia interpolacyjnego nie spełnia zazwyczaj warunku 3. Wynika to z faktu, że macierz interpolacyjna A jest macierzą bardzo źle uwarunkowaną. Złe uwarunkowanie jest większe dla funkcji dodatnio określonych i nieskończenie wiele razy różniczkowalnych tzn. np. dla funkcji Gaussa czy odwrotnej funkcji multikwadratowej.

Dla sieci trenowanych przez rozwiązanie zagadnienia aproksymacyjnego nie jest spełniony ponadto warunek 2.

Rozważmy metodę regularyzacji Tikhonova z operatorem ograniczeń $L \in \mathbb{R}^{n \times N}$, ($n < N$). Polega ona na rozwiązaniu zagadnienia minimalizacyjnego

$$\min_{\mathbf{w} \in \mathbb{R}^N} \{ \|A\mathbf{w} - \mathbf{y}\|^2 + \lambda^2 \|L\mathbf{w}\|_2^2 \}, \quad (3.16)$$

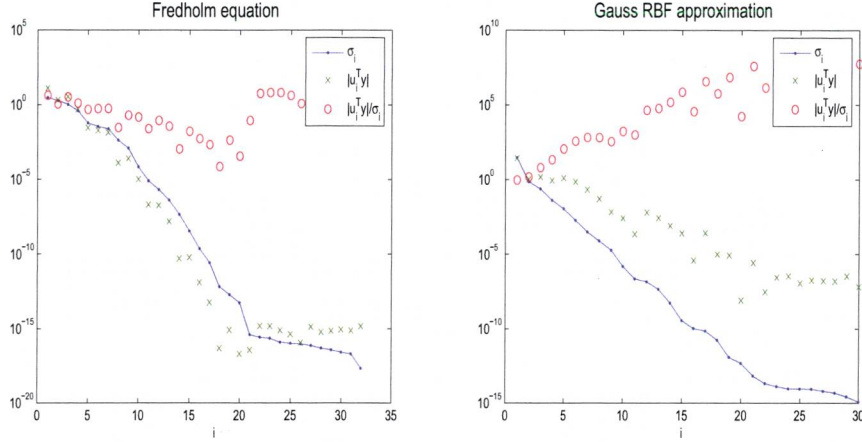
gdzie n jest numerycznym rzędem macierzy A a $\lambda \in \mathbb{R}$ jest nazywane parametrem regularyzacji. Stabilna metoda wyznaczania wektora $\mathbf{w}$ omówiona jest w paragrafie 3.5.

Metoda regularyzacji Tikhonova jest metodą rozwiązywania zadań źle postawionych i została zaproponowana jako metoda rozwiązywania zagadnień odwrotnych sformułowanych w postaci całki Fredholma

$$g(s) = \int_a^b K(s, t) f(t) dt.$$

Szukany rozwiązaniem zagadnienia odwrotnego jest funkcja $f(t)$ dla danej funkcji $g(s)$ i funkcji jądrowej $K(s, t)$. Dyskretyzacja równania Fredholma daje zawsze macierz źle uwarunkowaną.

Warunkiem wystarczającym na to, aby metoda regularyzacji Tikhonova dawała interpretowalne rozwiązanie dla zagadnienia odwrotnego zadanego całką Fredholma jest spełnienie dyskretnego warunku Picarda. Warunek ten określamy używając tu z wyprzedzeniem oznaczeń z wyprowadzenia rozwiązania zadania (3.16) podanego w paragrafie 3.5. Dyskretny warunek Picarda mówi, że ciąg wielkości $|\mathbf{u}_i^T \mathbf{y}|$, $i = 1, 2, \dots, r$ musi maleć szybciej niż ciąg σ_i , $i = 1, \dots, N$, gdzie $\mathbf{u}_i$ oraz σ_i są i -tym wektorem szczególnym i odpowiadającą wartością szczególną odpowiednio macierzy A . Jest to warunek, który łączy macierz równania z wektorem prawej strony. Dzięki temu, że $\mathbf{u}_i^T \mathbf{y} < \sigma_i$ dla wszystkich $i > p$ dla pewnego p to współczynniki filtrujące zmniejszają w rozwiązaniu wkład wektorów własnych odpowiadających małym wartościom własnym dla $i > p$. Dla zadania aproksymacji funkcjami



Rysunek 3.1: (a) Warunek Picarda dla równania całkowego Fredholma [17], (b) Warunek Picarda dla aproksymacji bazowymi funkcjami radialnymi funkcji Rosenbrocka dwóch zmiennych.

radialnymi warunek ten nie jest spełniony (patrz rysunek 3.1). Dla zagadnienia aproksymacyjnego rzeczywiste zmniejszenie wkładu wektorów własnych zachodzi dla $\sigma_p < \lambda < \sigma_{p-1}$ i wektorów własnych o indeksach $i > p + k$, gdzie k jest najmniejszą liczbą naturalną taką, że $u_{p+k}^T y > \lambda$.

Kluczowym elementem metody regularyzacji Tikhonova jest wybór wartości parametru λ . Celem jest taki wybór, który zapewni właściwą równowagę pomiędzy składnikiem aproksymacyjnym i regularyzacyjnym w wyrażeniu (3.16). Omówimy tutaj dwie metody wyboru wartości parametru λ z literatury oraz zaproponujemy nową metodę.

3.4.1 Metoda uogólnionej walidacji krzyżowej

Najpopularniejszą metodą wyboru parametru regularyzacji λ jest metoda uogólnionej walidacji krzyżowej (ang. *Generalized Cross Validation (GCV)*) (patrz [40]). Jest to metoda nie wymagająca znajomości *a priori* zaburzenia danych ani definicji dodatkowych parametrów.

Kryterium GCV jest wskaźnikiem określającym przybliżenie minimum predykcyjnego błędu średnio-kwadratowego (*Predictive Mean Square Error*) zdefiniowanego jako

$$T(\lambda) = \frac{1}{N} \sum_{i=1}^N (s_\lambda(\mathbf{x}_i) - f(\mathbf{x}_i))^2,$$

który zależy od estymowanej funkcji f .

Niech $A(\lambda)^\# = (A^T A + \lambda I)^{-1} A^T$ oznacza zregularyzowaną macierz odwrotną, gdzie A jest macierzą interpolacyjną. Wartość oczekiwaną błędu predykcyjnego $T(\lambda)$

można zapisać jako sumę obciążenia i wariancji

$$\begin{aligned} ET(\lambda) &= \frac{1}{N} E (\|A(\lambda)^\#(f + \epsilon) - f\|^2) \\ &= \frac{1}{N} \|(I - A(\lambda)^\#)f\|^2 + \frac{\sigma^2}{N} \text{Tr} [A(\lambda)^\#]^2 \\ &= b^2(\lambda) + \sigma^2 \mu_2(\lambda) \end{aligned}$$

gdzie $b(\lambda)$ jest obciążeniem modelu a σ^2 wariancją natomiast $\mu_2(\lambda)$ jest pewną funkcją wartości szczególnych macierzy A (patrz [40], [17]).

Dla zagadnienia (3.5) kryterium GCV zdefiniowane jest jako

$$V(\lambda) = \frac{1}{N} \sum_{k=1}^N (y_k - s_\lambda(\mathbf{x}_i))^2 / (1 - \mu_1(\lambda))$$

gdzie

$$\mu_1(\lambda) = \frac{1}{N} \sum_{i=1}^N a_{ii}(\lambda) = \frac{1}{N} \text{Tr} [A(\lambda)^\#],$$

a a_{ii} jest elementem diagonalnym macierzy $A(\lambda)^\#$.

W postaci macierzowej $V(\lambda)$ możemy zapisać

$$V(\lambda) = \frac{1}{N} \|(I - A(\lambda)^\#)y\|^2 / \left[\frac{1}{N} \text{Tr}(I - A(\lambda)^\#) \right]^2. \quad (3.17)$$

„Słabe twierdzenie” o zbieżności kryterium GCV (patrz [40]) mówi, że dla $N \rightarrow \infty$ istnieje ciąg $\tilde{\lambda}_N$, wartości λ minimalizujący wartość oczekiwaną kryterium GCV

$$EV(\lambda) = \frac{b^2(\lambda) + \sigma^2(1 - 2\mu_1(\lambda) + \mu_2(\lambda))}{(1 + \mu_1(\lambda))^2},$$

które przybliżają wartość λ dającą minimum predykcyjnego błędu średnio-kwadratowego

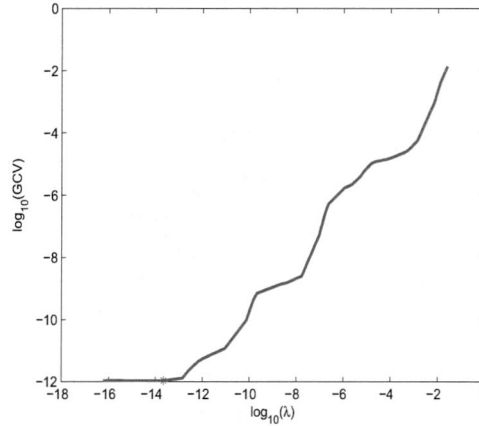
$$\lambda = \arg\min_{\lambda} ET(\lambda).$$

Dla danych nieregularnych i wolnych od błędów pomiaru metoda ta zawsze daje małe λ , które stabilizuje obliczanie odwrotności macierzy interpolacyjnej. λ mniejsze powodują powstanie błędu wynikającego z błędów zaokrągleń.

Rysunek 3.2 pokazuje przykładowy wykres wartości wskaźnika GCV względem λ .

3.4.2 Metoda L-krzywej

Drugą metodą wyboru parametru regularyzacji jest metoda L-krzywej zaproponowana przez Hansena [17] jako narzędzie w rozwiązywaniu zagadnień źle postawionych metodą regularyzacji Tikhonova.



Rysunek 3.2: Wykres wskaźnika $GCV(\lambda)$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.

Metoda ta polega na lokalizacji optymalnego parametru regularyzacji λ na wykresie normy $\|L\mathbf{w}_\lambda\|_2$ względem normy residuum $\|A\mathbf{w}_\lambda - y\|_2$. Krzywa ta zdefiniowana jest na skali logarytmicznej jako

$$(\zeta(\lambda), \eta(\lambda)) = (\log \|A\mathbf{w}_\lambda - y\|_2, \log \|L\mathbf{w}_\lambda\|_2).$$

i określa wkład tych wielkości w rozwiązaniu $\mathbf{w}_\lambda$ w zależności od λ .

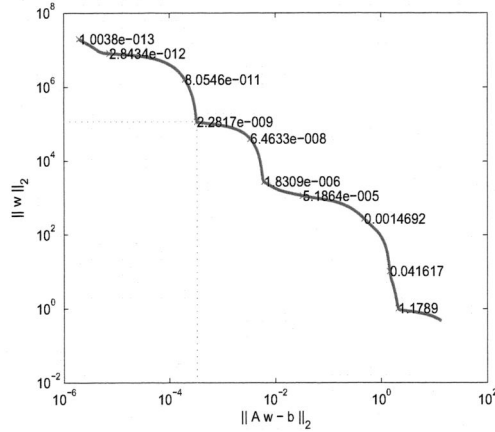
Rozwiązanie zagadnienia (3.16) dla dowolnego λ leży na tej krzywej. Rozwiązania dla małego λ znajdują się w górnej części wykresu ponad "narożnikiem" wykresu. Jeśli λ jest duże to rozwiązanie znajduje się w prawej dolnej części wykresu. Optymalny wybór parametru λ określa rozwiązanie znajdujące się w narożniku wykresu. Narożnik wykresu zdefiniowany jest jako punkt o maksymalnej krzywiznie

$$\kappa(\lambda) = \frac{\zeta''(\lambda)\eta''(\lambda) - \zeta'(\lambda)\eta'(\lambda)}{((\zeta'(\lambda))^2 + (\eta'(\lambda))^2)^{3/2}}.$$

Maksymalizacja funkcji $\kappa(\lambda)$ prowadzi do algorytmu wyboru wartości parametru regularyzacji λ , dla którego zachowany jest balans pomiędzy $\|A\mathbf{w}_\lambda - y\|_2$ i $\|L\mathbf{w}_\lambda\|_2$.

Obliczenie L-krzywej wymaga przeskanowania spektrum wartości szczególnych macierzy A co jest operacją czasochłonną. Algorytmy lokalizacji narożnika na wykresie opierają się na następującym schemacie.

1. Rozpocząć z kilkoma punktami (ζ_i, η_i) po obu stronach narożnika, tzn. dla dużego i dla małego λ .
2. Obliczyć trójparametryczną krzywą sklejaną trzeciego stopnia $\mathcal{S}$ dla punktów $(\zeta_i, \eta_i, \lambda_i)$.
3. Niech $\mathcal{S}_2$ oznacza pierwsze dwie współrzędne na krzywej $\mathcal{S}$ takie, że $\mathcal{S}_2$ przybliży L-krzywą. Obliczyć punkt na $\mathcal{S}_2$ o maksymalnej krzywiznie i odczytać odpowiadającą wartość λ_0 .



Rysunek 3.3: L-krzywa dla macierzy interpolacyjnej dla funkcji Gaussa dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.

4. Rozwiązać problem regularyzacyjny dla $\lambda = \lambda_0$ i dodać nowy punkt do $(\zeta(\lambda_0), \eta(\lambda_0), \lambda_0)$ od zbioru węzłów krzywej $\mathcal{S}$.
5. Przejść od punktu 2 jeśli nie uzyskaliśmy zbieżności.

Dla zagadnień źle postawionych w postaci całki Fredholma dla których spełniony jest warunek Picarda metoda L-krzywej daje wyniki zbliżone do metody GCV. Dla zagadnień interpolacyjnych oraz aproksymacyjnych dla małych zbiorów danych, które nie są zaburzone, takich jakie otrzymujemy z algorytmu optymalizacyjnego, metoda L-krzywej okazuje się dawać wyniki gorsze niż metoda GCV. Wynika to stąd, że L-krzywa może zawierać co najmniej kilka punktów o dużej krzywiznie. Pokazuje to rysunek 3.3.

3.4.3 Metoda ważonej wariancji gradientu

Wiadomo, że w wyrażeniu $\|Aw - y\|^2 + \lambda^2 \|Lw\|_2^2$ dla $w = w_\lambda$ w zależności od wartości parametru regularyzacji λ w różny sposób zachowują się składniki $\|Aw_\lambda - y\|^2$ (składowa rezydualna) i $\|Lw_\lambda\|_2^2$ (składowa regularyzacyjna). Gdy $\lambda \rightarrow 0$ to maleje część rezydualna natomiast część regularyzacyjna rośnie i odwrotnie wraz ze wzrostem wartości parametru λ , rośnie część residualna i maleje część regularyzacyjna.

Metoda dyskrepancyjna wyboru parametru λ polega na wybraniu największego λ , dla którego

$$\|Aw_\lambda - y\|^2 \leq \delta_e$$

gdzie δ_e jest ustalonym przez użytkownika poziomem odtwarzania całego zbioru danych. Oczywiście, gdy dane w zbiorze zawierają błąd to wartość δ_e powinna odzwierciedlać wiedzę o poziomie błędów danych. W metodzie tej pod uwagę brane są jednakowo wszystkie punkty danych zawarte w zbiorze $\mathbf{X}$.

Rozpatrzmy proces treningu sieci neuronowej $s_\lambda(\mathbf{x})$ dla kolejnych malejących wartości parametru λ za pomocą rozwiązania zagadnienia regularizacyjnego dla nieregularnego zbioru danych. Okazuje się, że w pewnych regionach otoczki wypukłej zbioru danych, gdy λ maleje, to odtwarzanie zbioru danych w otoczeniu punktu wartościowania maleje szybciej niż w innych regionach otoczki wypukłej zbioru danych.

W zastosowaniu sieci neuronowej do aproksymacji funkcji celu w procesach optymalizacyjnych praktycznie interesuje nas aproksymacja tylko w bliskim otoczeniu punktu wartościowania. Tak więc parametr λ możemy wybrać w zależności od lokalnego odtwarzania wartości funkcji na zbiorze danych $\mathbf{X}$.

W takim duchu zaproponowana została metoda WGV, która jest lokalną odmianą metody dyskrepancyjnej. W metodzie tej rozwiązuje się ciąg zagadnień regularizacyjnych dla malejącej wartości parametru λ . Wyboru parametru λ dokonujemy spośród tych wartości, dla których lokalnie jest uzyskane odtwarzanie zbioru danych na określonym przez użytkownika poziomie. Lokalne odtwarzanie zbioru danych $\mathbf{X}$ (*Normalized Local Mean Square Error*) definiujemy jako wskaźnik

$$NLMSE_{\lambda, \mathbf{z}}(\mathbf{x}) = \left(\sum_{k=1}^N \frac{[s_\lambda(\mathbf{x}_k) - y_k]^2}{y_k^2 r_k^2} \right) / \left(\sum_{k=1}^N \frac{1}{r_k^2} \right)^{\frac{1}{2}}, \quad (3.18)$$

gdzie $r_k = \|\mathbf{x}_k - \mathbf{x}\|$. Wybór λ ograniczamy do takiego zbioru $\Lambda(\mathbf{x})$ wartości parametru λ , dla których $NLMSE_{\lambda, \mathbf{z}} \leq \delta_e$. Fakt, że zbiór $\Lambda(\mathbf{x})$ rzeczywiście zależy od punktu wartościowania $\mathbf{x}$ pokazujemy na przykładzie zbioru danych otrzymanym podczas konstrukcji sieci aproksymującej w algorytmie optymalizacji bezgradientowej dwuwymiarowej funkcji Rosenbrocka – patrz rysunek 3.4. Rysunek 3.4a) przedstawia zbiór danych. Rysunek 3.4b) przedstawia wykresy wskaźnika NLMSE dla punktów z otoczenia punktu A oraz dla punktu B w zależności od wartości parametru λ . Jak widać odtwarzanie zbioru danych na poziomie błędu względnego 10^{-5} dla otoczenia punktu A otrzymujemy na zbiorze $\Lambda(A) \approx (0, 10^{-8})$. Dla punktu B mamy natomiast przedział $\Lambda(B) \approx (0, 10^{-11})$.

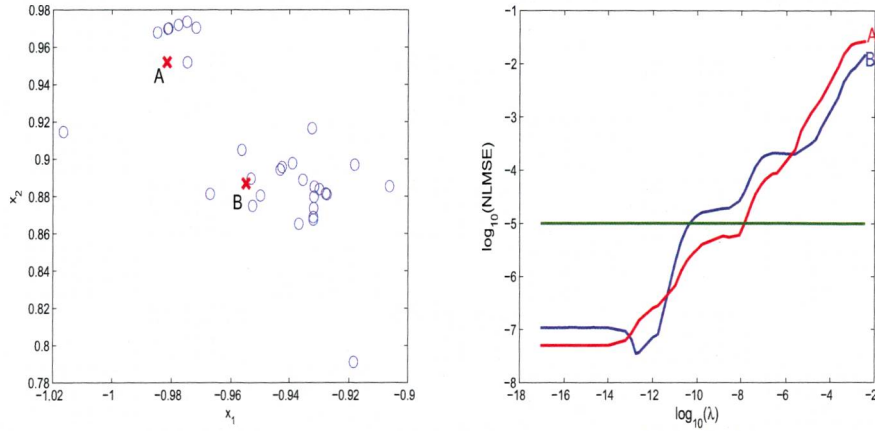
Ze zbioru $\Lambda(\mathbf{x})$ wybieramy takie λ , które minimalizuje funkcjonal wazonej wariancji gradientu zdefiniowanej jako

$$WGV_{\lambda, \mathbf{z}}(\mathbf{x}) = \sum_{k=1}^N \frac{\|\nabla s_\lambda(\mathbf{x}_k) - \vec{G}_\lambda(\mathbf{x})\|_2^2}{r_k^2} / \sum_{k=1}^N \frac{1}{r_k^2}, \quad (3.19)$$

gdzie $\vec{G}_\lambda(\mathbf{x})$ jest uśrednionym gradientem w punkcie $\mathbf{x}$, określonym jako

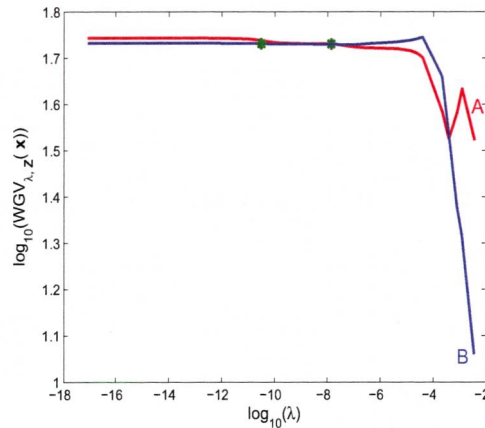
$$\vec{G}_\lambda(\mathbf{x}) = \sum_{k=1}^N \frac{\nabla s_\lambda(\mathbf{x}_k)}{r_k^2} / \sum_{k=1}^N \frac{1}{r_k^2} \quad (3.20)$$

oraz $r_k = \|\mathbf{x}_k - \mathbf{x}\|_2$. $\vec{G}_\lambda(\mathbf{x})$ oraz $WGV_{\lambda, \mathbf{z}}(\mathbf{x})$ zależą od $\mathbf{x}$ jedynie przez skalowanie, które wzmacnia wkład punktów $\mathbf{x}_k$ najbliższych położonych od punktu $\mathbf{x}$. Minimum funkcjonalu $WGV_{\lambda, \mathbf{z}}(\mathbf{x})$ względem parametru $\lambda \in \Lambda(\mathbf{x})$ określa ten model, dla



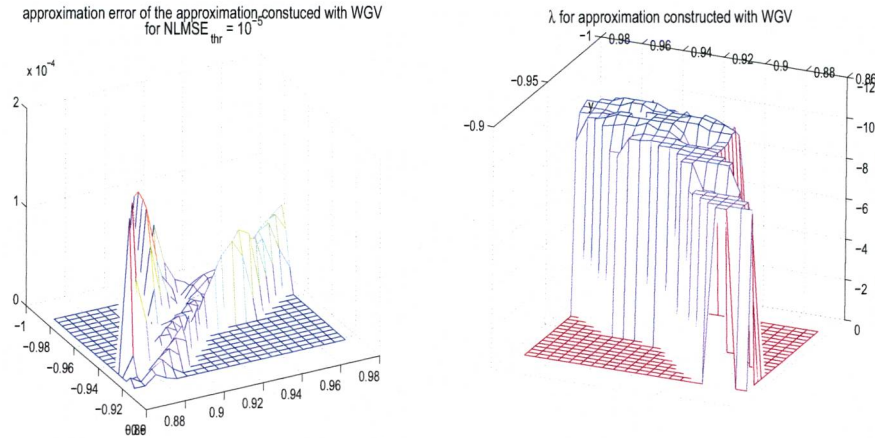
Rysunek 3.4: (a) Zbiór 30 punktów otrzymanych ze ścieżki optymalizacyjnej dla funkcji Rosenbrocka dwóch zmiennych, (b) Lokalne odtwarzanie zbioru treningowego w otoczeniu punktu A oraz punktu B.

którego $s_\lambda(\mathbf{x})$ ma najmniejsze oscylacje w otoczeniu punktu $\mathbf{x}$, gdyż odchylenie gradientu od uśrednionego gradientu $\tilde{G}_\lambda(\mathbf{x})$ na punktach $\mathbf{x}_k$ najbliższych położonych od punktu $\mathbf{x}$ jest najmniejsze. Na rysunku 3.5 pokazane są wykresy funkcjonalu $WGV_{\lambda, \mathbf{Z}}(\mathbf{x})$ odpowiadając wykresom $NLMSE_{\lambda, \mathbf{Z}}(\mathbf{x})$ na rysunku 3.4 b) dla punktów położonych w otoczeniu dwóch różnych punktów ze zbioru $\mathbf{Z}$.

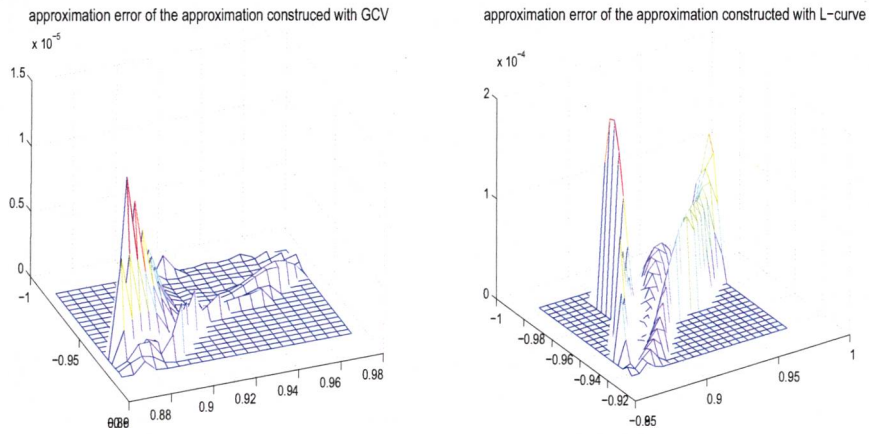


Rysunek 3.5: $WGV_{\lambda, \mathbf{Z}}(\mathbf{x})$ dla punktu A oraz punktu B z rysunku 3.4a). Symbolem * na wykresie oznaczona jest wartość λ wskazane przez metodę, dla $NLMSE_{\lambda, \mathbf{Z}}(\mathbf{x}) \leq 10^{-4}$.

Jak widać lokalność tych wskaźników jest uzyskiwana dzięki skalowaniu składnika sumy dla i -tego punktu przez kwadrat jego odległości od punktu wartościowania.



Rysunek 3.6: a) Błąd aproksymacji dla λ wybranego przez wskaźnik WGV dla $NLMSE = 5.0 \cdot 10^{-6}$, b) Wybrana wartość λ – jak widać w regionie uboższym w dane wskaźnik sugeruje większą wartość parametru λ .



Rysunek 3.7: a) Błąd aproksymacji dla $\lambda \approx 10^{-13}$ wybranego przez wskaźnik GCV, b) Błąd aproksymacji dla $\lambda \approx 10^{-8.5}$ wybranego za pomocą L-krzywej.

3.5 Stabilna metoda wyznaczania przybliżenia

Układ (3.4) rozwiązujemy przez rozwiązanie dwóch układów, mianowicie

$$\begin{cases} A\mathbf{w} + P\mathbf{b} = y, \\ P^T\mathbf{b} = 0. \end{cases} \quad (3.21)$$

Podobnie układ (3.6) rozwiązujemy przez rozwiązanie

$$\begin{cases} (A + L\lambda)\mathbf{w} + P\mathbf{b} = y, \\ P^T\mathbf{b} = 0. \end{cases} \quad (3.22)$$

Z równania $P^T\mathbf{b} = 0$ wyznaczamy składowe wektora $\mathbf{b}$ a następnie wstawiając je do równania pierwszego otrzymujemy układ o macierzy A lub $A + L\lambda$ odpowiednio.

Gdy macierz A nie jest zbyt duża, tak jak np. w aproksymacji funkcji celu, gdy ilość danych jest ograniczona, jako narzędzie stabilnego rozwiązywania powyższych

układów może zostać użyty rozkład macierzy A względem wartości szczególnych (SVD – *singular value decomposition*). Rozkład ten istnieje dla dowolnej nieosobliwej macierzy A i jest określony jako

$$A = \mathbf{U} \mathbf{S} \mathbf{V}^T$$

gdzie $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{N \times N}$ są macierzami ortogonalnymi a $\mathbf{S} = \text{diag}(\sigma_1, \dots, \sigma_N)$ oraz mamy porządek $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ macierzy szczególnych macierzy A . Rozkład SVD jest jednoznaczny do znaku kolumn macierzy $\mathbf{U}$ i $\mathbf{V}^T$. Gdy zdefiniujemy $y' := y - P\mathbf{b}$, to wektor $\mathbf{w}$ dla zagadnienia interpolacyjnego obliczamy za pomocą rozwinięcia względem wektorów szczególnych macierzy A

$$\mathbf{w} = \mathbf{V} \mathbf{S}^{-1} \mathbf{U}^T y' = \sum_{i=1}^N \frac{1}{\sigma_i} (\mathbf{u}_i^T y') \mathbf{v}_i$$

Gdy macierz A jest bardzo źle uwarunkowana, tak jak jest to np. dla funkcji Gaussa często mamy $\sigma_i < \epsilon$ dla pewnego $i < N$, gdzie ϵ jest dokładnością reprezentacji zmiennopozycyjnej. Wówczas zamiast całego rozwinięcia, którego końcowe wyrazy zawierają zaburzoną informację pochodzącą z błędów reprezentacji zmiennopozycyjnej wartości szczególnych. W takiej sytuacji ograniczamy się do obcięcia do r wyrazów, dla których $\sigma_i > \epsilon$ ($i = 1, \dots, r$).

Rozwiązanie zagadnienia aproksymacyjnego można natomiast zapisać za pomocą uogólnionego rozkładu SVD (GSVD) macierzy A i L . Rozkład GSVD dla macierzy A i L definiujemy (patrz [17]) jako

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & I_{N-r} \end{pmatrix} X^{-1}, \quad L = V X^{-1}.$$

Kolumny macierzy $U \in \mathbb{R}^{N \times N}$ oraz $V \in \mathbb{R}^{r \times r}$ są ortonormalne, natomiast kolumny macierzy $\mathbf{X} \in \mathbb{R}^{N \times N}$ są ortogonalne względem macierzy $A^T A$

$$X^T A^T A X = \begin{pmatrix} \Sigma^2 & 0 \\ 0 & I_{N-r} \end{pmatrix},$$

oraz spełniają

$$X^T L^T L X = \begin{pmatrix} M^2 & 0 \\ 0 & 0 \end{pmatrix},$$

gdzie macierze $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ oraz $M = \text{diag}(\mu_1, \dots, \mu_r)$ mają na przekątnej nieujemne elementy uporządkowane tak, że

$$0 \leq \sigma_1 \leq \dots \leq \sigma_r \leq 1, \quad 1 \geq \mu_1 \geq \dots \geq \mu_r \geq 0.$$

znormalizowane tak, że

$$\sigma_i^2 + \mu_i^2 = 1, \quad i = 1, \dots, r.$$

Wówczas

$$\gamma_i = \sigma_i / \mu_i \quad i = 1, \dots, r$$

nazywamy uogólnionymi wartościami szczególnymi. Jeśli macierz A jest pełnego rzędu to rozwiązanie $\mathbf{w}_\lambda$ powyższego problemu zapisać można za pomocą rozkładu SVD

$$\mathbf{w}_\lambda = XF \begin{pmatrix} \Sigma^\# & 0 \\ 0 & I_{N-r} \end{pmatrix} U^T y,$$

gdzie macierz $F = \text{diag}(f_i)$. Elementy diagonalne f_i nazywają się współczynnikami filtrującymi.

Rozwiązanie zregularyzowane $\mathbf{w}_\lambda$ oraz odpowiadający mu wektor residuów $y - A\mathbf{w}_\lambda$ można zapisać więc jako

$$\mathbf{w}_\lambda = \sum_{i=1}^r f_i \frac{u_i^T y}{\sigma_i} x_i + \sum_{k=r+1}^N u_k^T y x_k \quad (3.23)$$

i

$$L\mathbf{w}_\lambda = \sum_{i=1}^r f_i \frac{u_i^T y}{\gamma_i} v_i$$

oraz

$$y - A\mathbf{w}_\lambda = \sum_{i=1}^r (1 - f_i) u_i^T y u_i + (I_N - UU^T)y.$$

Współczynniki filtrujące zapisują się

$$f_i = \frac{\gamma_i^2}{\gamma_i^2 + \lambda^2} \quad i = 1, \dots, r.$$

Współczynniki filtrujące redukują w $\mathbf{w}_\lambda$ wpływ składowych odpowiadających wartościom szczególnym $\sigma_i \leq \sigma_p$ takim, że $\sigma_p < \lambda < \sigma_{p-1}$ ponieważ $f_i/\sigma_i \ll 1/\sigma_i$ dla każdego $p \leq i \leq N$. Tak więc składowe odpowiadające małym wartościom szczególnym, czyli te które wzmacniają zaburzenie danych, są odfiltrowywane. Zauważmy że i tutaj możemy zastosować obcięcie gdy $\gamma_i < \epsilon$ lub prosto $\mu_i < \epsilon$.

3.6 Ograniczenia błędu przybliżenia

W niniejszym paragrafie zajmiemy się dwoma rodzajami ograniczeń błędów przybliżeń sieciami o radialnych funkcjach aktywacji dostępnych w literaturze. Pierwsza grupa ograniczeń to ograniczenia wymagające założeń o gęstości zbioru danych $\mathbf{X}$ natomiast druga grupa to ograniczenia nie wymagające tego typu założeń.

Głównymi elementami wyprowadzeń ograniczeń z obydwu grup są:

1. własność odtwarzania wielomianów przez proces aproksymacyjny,
2. funkcjonal potęgowy,
3. funkcje Lebesque.

W kolejnych paragrafach omówimy więc najpierw po kolei te pojęcia.

Ograniczenia błędu omawiane w tym rozdziale zależą od gęstości zbioru $\mathbf{X}$ w dziedzinie Ω mierzonego za pomocą wskaźnika wypełnienia (ang. *fill distance*). Zanim więc przejdziemy do omówienia głównych elementów ograniczeń podamy definicję tego wskaźnika w formie lokalnej i globalnej oraz podamy jego własności. W rozdziale tym podamy również definicję wskaźnika wprowadzonego przez nas i mierzącego jakość otaczania punktu wartościowania przez punkty ze zbioru danych. Pokażemy również jak jakość aproksymacji radialnymi funkcjami bazowymi zależy od tego wskaźnika.

3.6.1 Wskaźniki gęstości danych

W rozdziale tym omówimy dwa wskaźniki mierzące lokalną jakość wypełnienia dziedziny Ω przez punkty zbioru danych $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$.

Norma siatki

Norma siatki (ang. *mesh norm*) jest wielkością określającą jak dobrze zbiór danych $\mathbf{X} \subseteq \Omega$ wypełnia dziedzinę Ω . Jest definiowana w dwóch wersjach

1. lokalnej [29] – dla parametru $\rho > 0$ dla każdego punktu $\mathbf{x} \in \Omega$

$$h_{\rho, \mathbf{X}, \Omega}(\mathbf{x}) := \max_{\mathbf{y} \in B(\mathbf{x}, \rho) \cap \Omega} \min_{\mathbf{x}_i \in \mathbf{X}} \|\mathbf{y} - \mathbf{x}_i\|_2, \quad (3.24)$$

gdzie $B(\mathbf{x}, \rho)$ oznacza kulę o środku $\mathbf{x}$ i promieniu ρ ,

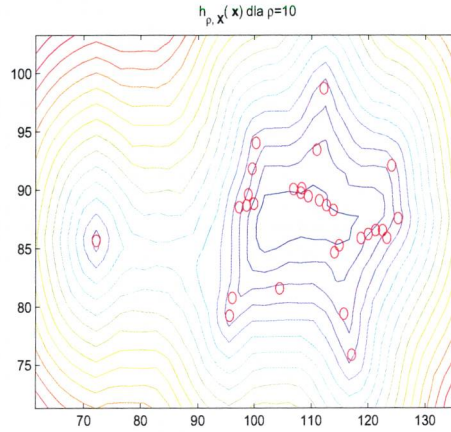
2. globalnej [43]

$$h_{\mathbf{X}, \Omega} := \max_{\mathbf{y} \in \Omega} \min_{\mathbf{x}_j \in \mathbf{X}} \|\mathbf{y} - \mathbf{x}_j\|_2. \quad (3.25)$$

W wersji lokalnej wskaźnik ten określa promień największej kuli nie zawierającej punktów danych ze zbioru $\mathbf{X}$ i zawartej w kuli o środku $\mathbf{x}$ i promieniu ρ . W wersji globalnej określa promień największej kuli nie zawierającej danych ze zbioru $\mathbf{X}$ zawartej w dziedzinie Ω .

Mając dany zbiór $\mathbf{X}$ oraz promień ρ od obliczenia lokalnej normy siatki można wykorzystując algorytm triangulacji Delaunaya [14]. Niestety konstrukcja triangulacji Delaunaya jest możliwa w przestrzeni $\mathbb{R}^d$ tylko, gdy zbiór danych złożony jest z punktów w pozycji ogólnej tzn., nie zawiera $d+1$ punktów współliniowych ani $d+2$ punktów położonych na okręgu. Ten warunek nie będzie zazwyczaj spełniony, gdy zbiór danych pochodzi ze ścieżki algorytmu optymalizacyjnego.

Rysunek 3.8 pokazuje przykładowy wykres wskaźnika $h_{\rho, \mathbf{X}, \Omega}$ dla 30 punktów przykładowego zbioru danych dwuwymiarowych.



Rysunek 3.8: $h_{\rho, \mathbf{X}, \Omega}(\mathbf{x})$ dla $\rho = 10$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.

Wskaźnik otoczenia $\gamma_{\mathbf{X}}(\mathbf{x})$

Drugim rozważanym przez nas wskaźnikiem mierzącym lokalną jakość rozkładu punktów zbioru $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ w otoczeniu punktu $\mathbf{x}$ jest wskaźnik mierzący długości odcinków $\bar{\mathbf{x}}\mathbf{x}_i$ ($i = 1, \dots, N$) i kąty $\angle \mathbf{x}_i \mathbf{x} \mathbf{x}_j$ ($i, j = 1, \dots, N$; $i < j$) między punktami zbioru $\mathbf{X}$. Lokalność takiego wskaźnika może być uzyskana w ten sposób, że nadaje się większą wagę krótszym odcinkom $\bar{\mathbf{x}}\mathbf{x}_i$ oraz większym kątom $\angle \mathbf{x}_i \mathbf{x} \mathbf{x}_j$. Wskaźnik o takich własnościach definiujemy następująco [3] :

$$\gamma_{\mathbf{X}}(\mathbf{x}) = \sum_{\substack{i,j \\ j < i}}^N a_{ij} W_{ij} / \sum_{\substack{i,j \\ j < i}}^N W_{ij}, \quad (3.26)$$

gdzie

$$a_{ij} = \frac{d_{ij}}{r_i + r_j}, \quad \text{dla} \quad d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2, \quad \text{a} \quad r_i = \|\mathbf{x} - \mathbf{x}_i\|_2,$$

oraz wagi W_{ij} są zdefiniowane jako

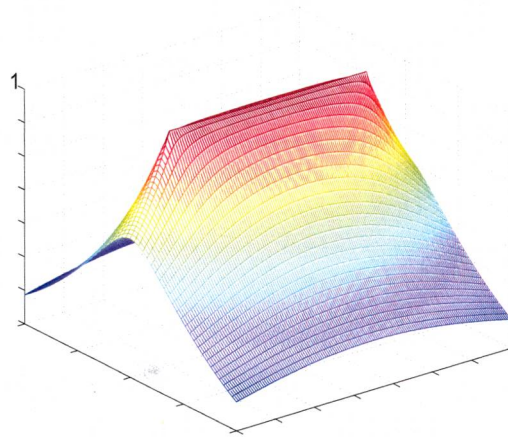
$$W_{ij} = \frac{1}{r_i + r_j}.$$

W przeciwieństwie do wskaźników zdefiniowanych w paragrafie 3.6.1, które zależą od gęstości wypełnienia kuli $B(\mathbf{x}, \rho)$ (w wypadku $h_{\rho, \mathbf{X}, \Omega}$) oraz gęstości wypełnienia całej dziedziny Ω przez punkty zbioru $\mathbf{X}$ (w wypadku $h_{\mathbf{X}, \Omega}$), wskaźnik $\gamma_{\mathbf{X}}(\mathbf{x})$ mierzy jak dobrze punkty danych otaczają punkt $\mathbf{x}$. Wartość $\gamma_{\mathbf{X}}(\mathbf{x})$ jest tym większa, im więcej punktów znajduje się w bliskim otoczeniu punktu $\mathbf{x}$ i są ułożone tak, że występuje dużo kątów rozwartych $\angle \mathbf{x}_i, \mathbf{x}, \mathbf{x}_j$ ($i, j = 1, \dots, N$; $i < j$) w bliskim otoczeniu punktu $\mathbf{x}$.

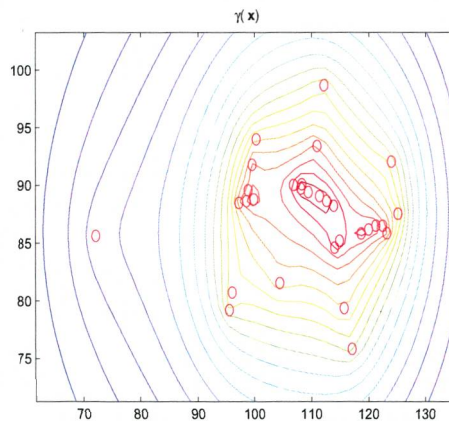
Własności takie wskaźnik $\gamma_{\mathbf{X}}(\mathbf{x})$ zawdzięcza własnościom funkcji a_{ij} i W_{ij} . Funkcje a_{ij} największą wartość równą 1 przyjmują dla punktów $\mathbf{x}$ położonych na

odcinku $\mathbf{x}_i\bar{\mathbf{x}}_j$. Maksimum jest więc osiągane, gdy kąt $\angle \mathbf{x}_i, \mathbf{x}, \mathbf{x}_j = \pi$. Wzmocnienie wpływu par punktów bliżej położonych punktu $\mathbf{x}$ jest uzyskiwane przez funkcję wagową W_{ij} , która maksymalną wartość równą $1/d_{ij}$ przyjmuje również na odcinku $\mathbf{x}_i\bar{\mathbf{x}}_j$. Im bliżej więc odcinka $\mathbf{x}_i\bar{\mathbf{x}}_j$ i im mniejsze d_{ij} tym większy wpływ pary punktów $\mathbf{x}_i, \mathbf{x}_j$ na wartość wskaźnika dla punktu $\mathbf{x}$.

Rysunek 3.9 pokazuje wykres funkcji a_{ij} oraz W_{ij} natomiast rysunek 3.10 pokazuje wykres wskaźnika $\gamma_{\mathbf{x}}(\mathbf{x})$ dla 30 punktów przykładowego zbioru danych z procesu optymalizacyjnego dwóch zmiennych.



Rysunek 3.9: Wykres funkcji a_{ij} . Maksymalna wartość przyjmowana jest na odcinku $\mathbf{x}_i, \mathbf{x}_j$. Wykres funkcji wagowej W_{ij} jest analogiczny. Maksymalna wartość funkcji wagowej również jest przyjmowana na odcinku $\mathbf{x}_i, \mathbf{x}_j$ i wynosi $1/d_{ij}$.



Rysunek 3.10: Wykres wskaźnika $\gamma_{\mathbf{x}}(\mathbf{x})$ dla przykładowego zbioru 30 punktów danych z dwuwymiarowego procesu optymalizacyjnego.

3.6.2 Odtwarzanie wielomianów

Podstawowym narzędziem używanym w dowodach ograniczeń błędów interpolacji oraz aproksymacji jest własność odtwarzania wielomianów. Odtwarzanie

wielomianów przez proces interpolacyjny/aproksymacyjny definiuje się następująco:

Definicja 10.

Proces, który określa dla każdego zbioru $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ rodzinę funkcji $u_j = u_j^{\mathbf{X}} : \Omega \rightarrow \mathbb{R}$, $1 \leq j \leq N$ zapewnia lokalne odtwarzanie wielomianów stopnia l na zbiorze Ω , jeśli istnieją stałe $h_0, C_1, C_2 > 0$ takie, że spełnione są warunki

1. $\sum_{j=1}^N p(\mathbf{x}_j)u_j = p$ dla każdego $p \in \pi_l(\mathbb{R}^d)|\Omega$,
2. $\sum_{j=1}^N |u_j| \leq C_1$ dla każdego $\mathbf{x} \in \Omega$,
3. $u_j(\mathbf{x}) = 0$, jeśli $\|\mathbf{x} - \mathbf{x}_j\|_2 \geq C_2 h_{\mathbf{X}, \Omega}$ i $\mathbf{x} \in \Omega$,

dla każdego $\mathbf{X}$ dla którego $h_{\mathbf{X}, \Omega} \leq h_0$.

Jeśli proces gwarantuje lokalne odtwarzanie wielomianów to ograniczenie błędu aproksymacji funkcji tym procesem opisuje następujące twierdzenie.

Twierdzenie 8. (patrz [43])

Niech $\Omega \subset \mathbb{R}^d$ jest ograniczony. Niech Ω^* będzie domknięciem zbioru $\bigcup_{\mathbf{x} \in \Omega} B(\mathbf{x}, C_2 h_0)$. Zdefiniujmy

$$s_{f, \mathbf{X}} = \sum_{j=1}^N f(\mathbf{x}_j) u_j,$$

gdzie $\{u_j\}_{j=1}^N$ jest lokalnym odtwarzaniem wielomianów stopnia m na zbiorze Ω . Wówczas jeśli $f \in C^{m+1}(\Omega^*)$ to istnieje stała $c > 0$ zależna tylko od stałych z definicji lokalnego odtwarzania wielomianów, taka że

$$|f(\mathbf{x}) - s_{f, \mathbf{X}}(\mathbf{x})| \leq C h_{\mathbf{X}, \Omega}^{m+1} |f|_{C^{m+1}(\Omega^*)} \quad (3.27)$$

dla każdego zbioru $\mathbf{X}$ o gęstości $h_{\mathbf{X}, \Omega} < h_0$. Seminorma po prawej stronie powyższej nierówności jest określona

$$|f|_{C^{m+1}(\Omega^*)} := \max_{|\alpha|=m+1} \|D^\alpha f\|_{L_\infty(\Omega^*)},$$

gdzie operator D jest zdefiniowany w paragrafie 3.3.3.

Na początku zadajmy sobie pytanie, jakie warunki muszą spełniać zbiory $\mathbf{X}$ i Ω , aby możliwe było lokalne odtwarzanie wielomianów stopnia nie większego niż m .

Warunkiem jaki musi spełniać zbiór $\mathbf{X}$ jest unisolwentność względem przestrzeni $\pi_m(\mathbb{R}^d)$. Natomiast dziedzina Ω musi spełniać warunek stożka wewnętrznego.

Definicja 11.

Mówimy, że zbiór $\Omega \subseteq \mathbb{R}^d$ spełnia warunek stożka, jeśli istnieje kąt $\theta \in (0, \pi/2)$ i promień $r > 0$, taki, że dla każdego $\mathbf{x} \in \Omega$ istnieje wektor jednostkowy $\xi(\mathbf{x})$ taki, że stożek

$$C(\mathbf{x}, \xi(\mathbf{x}), \theta, r) := \{\mathbf{x} + \eta \mathbf{y} : \mathbf{y} \in \mathbb{R}^d, \|\mathbf{y}\|_2 = 1, \mathbf{y}^T \xi(\mathbf{x}) \geq \cos \theta, \eta \in [0, r]\} \quad (3.28)$$

zawiera się w Ω .

Twierdzenie 9. (patrz [43])

Założmy, że $\Omega \subseteq \mathbb{R}^d$ jest zwarty i spełnia warunek stożka wewnętrznego z promieniem $r > 0$ i kątem $\theta \in (0, \pi/2)$. Niech $m \in \mathbb{N}$ będzie ustalone. Założmy, że $h > 0$ i zbiory $\mathbf{X}$ spełnia

1.

$$h \leq \frac{r \sin \theta}{4(1 + \sin \theta)m^2}, \quad (3.30)$$

2. dla każdej kuli $B(\mathbf{x}, h) \subseteq \Omega$ istnieje punkt $\mathbf{x}_j \in \mathbf{X} \cap B(\mathbf{x}, h)$,

to dla każdego $\mathbf{x} \in \Omega$ istnieją liczby rzeczywiste $u_j(\mathbf{x})$ takie, że

$$p(\mathbf{x}) = \sum_{j=1}^N u_j(\mathbf{x}) p(\mathbf{x}_j) \quad (3.31)$$

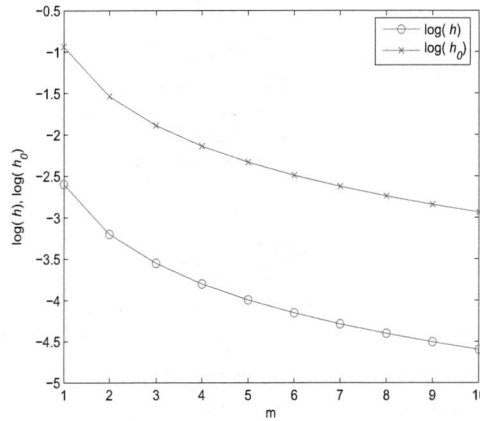
dla każdego $p \in \pi_m(\mathbb{R}^d)$. Ponadto

$$\sum_{j=1}^N |u_j(\mathbf{x})| \leq 2. \quad (3.32)$$

Dowód tego twierdzenia wykorzystujący pojęcie zbioru normującego oraz twierdzenie Hanha-Banacha o ograniczoności normy operatora liniowego podał Jetter *et al* w pracy [19]. Twierdzenie to określa stałe w definicji lokalnego odtwarzania wielomianów

$$\begin{aligned} C_1 &= 2, & C_2 &= \frac{16(1 + \sin \theta)^2 m^2}{3 \sin^2 \theta}, \\ h_0 &= \frac{r}{C_2}. \end{aligned} \quad (3.33)$$

Zależność wielkości h i h_0 od stopnia wymaganego odtwarzania wielomianów jest pokazana na rysunku 3.11.



Rysunek 3.11: Zależność wskaźnika h w (3.30) oraz h_0 w (3.33) od wymaganego stopnia odtwarzanych wielomianów w kuli jednostkowej.

3.6.3 Funkcjonał potęgowy

Dowody wszystkich ograniczeń błędów interpolacji i aproksymacji, które wykorzystują funkcjonal potęgowy wychodzą od ogólnej postaci błędu interpolacji

$$|f - s_{\mathbf{X},f}| \leq P_{\Phi,\mathbf{X}}(\mathbf{x}) \|f\|_{\mathcal{N}_{\Phi}(\mathbb{R}^d)},$$

gdzie $P_{\Phi,\mathbf{X}}$ jest funkcjonalem potęgowym, który zdefiniujemy w tym paragrafie. Funkcjonał potęgowy jest ściśle związany z reprezentacją Lagrange'a funkcji interpolującej dane ze zbioru $\mathbf{Z}$. Zaczniemy więc od twierdzenia o istnieniu reprezentacji Lagrange'a

Twierdzenie 10. (patrz [30])

Założmy, że Φ jest warunkowo dodatnio określona względem przestrzeni $\pi_m(\mathbb{R}^d)$ na zbiorze $\Omega \subseteq \mathbb{R}^d$. Założmy że $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ jest $\pi_m(\mathbb{R}^d)$ -unisolwentny. Wówczas istnieją funkcje $u_j^* \in V_{\mathbf{X}}$ takie, że $u_j^*(\mathbf{x}_k) = \delta_{jk}$, gdzie

$$V_{\mathbf{X}} := \pi_m(\mathbb{R}^d) + \left\{ \sum_{j=1}^N \alpha_j \Phi(\cdot, x_j) : \sum_{j=1}^N \alpha_j p(x_j) = 0, p \in \pi_m(\mathbb{R}^d) \right\}.$$

Ponadto istnieją funkcje v_j^* , $1 \leq j \leq Q$ takie, że wektory $u^*(x) = [u_1(\mathbf{x}), \dots, u_N(\mathbf{x})]^T \in \mathbb{R}^N$ i $v^*(x) = [v_1(\mathbf{x}), \dots, v_Q(\mathbf{x})]^T \in \mathbb{R}^Q$ tworzą rozwiązanie układu

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} u^*(\mathbf{x}) \\ v^*(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} R(\mathbf{x}) \\ S(\mathbf{x}) \end{pmatrix} \quad (3.34)$$

gdzie

$$\begin{aligned} A &= [\Phi(\mathbf{x}_i, \mathbf{x}_j)] \in \mathbb{R}^{N \times N}, \\ P &= [p_i(\mathbf{x}_j)] \in \mathbb{R}^{N \times Q}, \\ R(\mathbf{x}) &= [\Phi(\mathbf{x}, \mathbf{x}_1), \dots, \Phi(\mathbf{x}, \mathbf{x}_N)]^T \in \mathbb{R}^N, \\ S(\mathbf{x}) &= [p_1(\mathbf{x}), \dots, p_Q(\mathbf{x})]^T \in \mathbb{R}^Q. \end{aligned}$$

Na podstawie powyższego twierdzenia funkcję interpolującą możemy zapisać w postaci

$$s_{f,\mathbf{x}} = \sum_{j=1}^N f(\mathbf{x}_j) u_j^*(\mathbf{x}). \quad (3.35)$$

Definicja 12.

Założmy, że $\Omega \subseteq \mathbb{R}^d$ jest zbiorem otwartym natomiast $\Phi \in C^{2k}(\Omega \times \Omega)$ jest warunkowo dodatnio określoną funkcją jądrową na zbiorze Ω względem $\pi_m(\mathbb{R}^d)$. Jeśli $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ jest $\pi_m(\mathbb{R}^d)$ -unisolwentny, wówczas dla każdego $\mathbf{x} \in \Omega$ funkcjonal potęgowy jest zdefiniowany jako

$$[P_{\Phi,\mathbf{X}}(\mathbf{x})]^2 := \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_{j=1}^N u_j^*(\mathbf{x}) \Phi(\mathbf{x}, \mathbf{x}_j) + \sum_{i,j=1}^N u_i^*(\mathbf{x}) u_j^*(\mathbf{x}) \Phi(\mathbf{x}_i, \mathbf{x}_j)$$

Jeśli zdefiniujemy dla $\mathbf{x} \in \Omega$ formę kwadratową

$$\mathcal{Q}(u) := \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_{j=1}^N u_j \Phi(\mathbf{x}, \mathbf{x}_j) + \sum_{i,j=1}^N u_i u_j \Phi(\mathbf{x}_i, \mathbf{x}_j)$$

to

$$[P_{\Phi,\mathbf{X}}(\mathbf{x})]^2 = \mathcal{Q}(u^*(\mathbf{x})) = \left\| G(\cdot, \mathbf{x}) - \sum_{j=1}^N u_j^* G(\cdot, \mathbf{x}_j) \right\|_{\mathcal{N}_{\Phi}(\Omega)}, \quad (3.36)$$

gdzie $G(\cdot, \mathbf{x})$ jest ściśle dodatnio określoną funkcją zdefiniowaną następująco

$$G(\cdot, \mathbf{x}) := \Phi(\cdot, \mathbf{x}) - \sum_{k=1}^Q p_k(\mathbf{x}) \Phi(\cdot, \mathbf{x}_k).$$

Wyrażenie (3.36) jest więc normą funkcjonału wartościowania błędu interpolacji (3.3) w punkcie $\mathbf{x} \in \Omega$ w przestrzeni natywnej generowanej przez bazową funkcję radialną Φ .

3.6.4 Funkcja Lebesgue'a

Gdy mamy reprezentację Lagrange'a (3.35) lub odtwarzanie wielomianów zgodnie z twierdzeniem 9, to funkcja Lebesgue'a zdefiniowana jest jako

$$\sum_{j=1}^N |u_j(\mathbf{x})|. \quad (3.37)$$

Funkcja ta jest istotna w analizie błędu interpolacji i aproksymacji funkcjami radialnymi ponieważ za jej pomocą ogranicza się funkcjonal potęgowy. W wypadku

interpolacji funkcjami Gaussa Schaback [34] oraz Larsson *et al.* [20] niezależnie dowodzą, że gdy szerokość funkcji Gaussa zdąży do nieskończoności to funkcja Lebesgue'a (3.37) zdąży do wielomianu.

W twierdzeniu 9 o odtwarzaniu wielomianów z przestrzeni $\pi_m(\mathbb{R}^d)$, mamy że funkcja Lebesgue jest ograniczona

$$\sum_{j=1}^N |u_j(\mathbf{x})| < C_1 = 2, \quad (3.38)$$

przy czym stała C_1 zależy od gęstości zbioru $\mathbf{X}$ tzn. od h_0 . Natomiast h_0 zależy od m tak jak $o(1/m^2)$. Do dowodu tych zależności (patrz [19], [43]) jak i jednorodnego ograniczenia przez stałą 2, używa się nierówności Markowa o wielomianach i ich pochodnych, która zachodzi dla dowolnego wielomianu $p \in \pi_m(\mathbb{R}^d)$

$$|p'(t)| \leq m^2 \|p\|_{L_\infty[-1,1]}, \quad t \in [-1, 1]. \quad (3.39)$$

Z dowodu wiemy, jak dobrze zbiór $\mathbf{X}$ musi wypełniać dziedzinę Ω , aby zachowane było jednorodne ograniczenie $C_1 = 2$.

Jeśli chcemy, aby h_0 zależało od m tak jak $o(1/m)$ to niestety musimy poświęcić jednorodne ograniczenia funkcji Lebesgue'a stałą 2. Mówi o tym następujący lemat i twierdzenie.

Lemat 6. (spektralna wersja nierówności Markowa – patrz [43])

Niech $\gamma_1 = 2$ oraz $\gamma_d = 2d(1 + \gamma_{d-1})$ dla $d = 2, 3, \dots$. Niech m i q będą liczbami naturalnymi takimi, że $q > \gamma_d(m + 1)$. Niech Ω będzie sześciannem w $\mathbb{R}^d$. Niech dziedzina Ω będzie podzielona na q^d równych sześciannów rozłącznych. Jeśli $\mathbf{X} \subseteq \Omega$ jest zbiorem $N \geq q^d$ punktów takich, że każdy sześciann zawiera przynajmniej jeden z nich to dla każdego $p \in \pi_m(\mathbb{R}^d)$ mamy

$$\|p\|_{L_\infty(\Omega)} \leq e^{2d\gamma_d(m+1)} \|p\|_{L_\infty(\mathbf{X})}.$$

O odtwarzaniu wielomianów mówi następujące twierdzenie:

Twierdzenie 11. (patrz [43])

Niech $\Omega = W(\mathbf{x}_0, R)$ będzie sześciannem w $\mathbb{R}^d$. Istnieją stałe $c_0, c_2 > 0$ zależące tylko od Ω takie, że dla każdego $m \in \mathbb{N}$ i każdego $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ z $h_{\mathbf{X}, \Omega} \leq c_0/m$ możemy znaleźć funkcje $u_j : \Omega \rightarrow \mathbb{R}$ spełniające

1. $\sum_{j=1}^N u_j(\mathbf{x}) p(\mathbf{x}_j) = p(\mathbf{x})$ dla każdego $\mathbf{x} \in \Omega$ i $p \in \pi_m(\mathbb{R}^d)$,
2. $\sum_{j=1}^N |u_j(\mathbf{x})| \leq e^{2d\gamma_d(m+1)}$ dla każdego $\mathbf{x} \in \Omega$,
3. $u_j(\mathbf{x}) = 0$ jeśli $\|\mathbf{x} - \mathbf{x}_j\|_2 > c_2 m h_{\mathbf{X}, \Omega}$,

gdzie stała γ_d jest zdefiniowana jak w poprzednim lemacie.

Z dowodu tego twierdzenia mamy

$$h_0 = \frac{2R}{3q} = \frac{2R}{3\gamma_d(m+1)} =: \frac{c_0}{m}. \quad (3.40)$$

Jak widać z punktu 3. powyższego twierdzenia przeskalowaniu ulega również nośnik funkcji u_j w porównaniu z funkcjami u_j z twierdzenia 9. Jak widać z tego twierdzenia ograniczenie to jest bardzo „grube” tzn. $e^{2d\gamma_d(m+1)}$ jest wyrażeniem bardzo szybko rosnącym w zależności od m i od d .

Ograniczenia powyższe pokazują jakiego rzędu wielkościami są funkcje Lebesgue’a. Dalej zobaczymy, że znajomość wartości tej funkcji w punkcie $\mathbf{x}$ wystarcza do oszacowania błędu interpolacji bez konieczności założeń o gęstości danych w otoczeniu punktu x .

3.6.5 Ograniczenia błędu z założeniem o gęstości zbioru danych

W grupie tej wyróżnimy dwa typy ograniczeń:

1. wykorzystujące funkcjonal potęgowy,
2. ograniczenia błędu w przestrzeniach Sobolewa.

Ograniczenia błędu wykorzystujące funkcjonal potęgowy

Ograniczenia błędu interpolacji i aproksymacji za pomocą radialnych funkcji bazowych wykorzystujące funkcjonal potęgowy wychodzą od następującego ogólnego twierdzenia.

Twierdzenie 12. (patrz [43])

Założmy, że $\Omega \subseteq \mathbb{R}^d$ jest zbiorem otwartym natomiast $\Phi \in C^{2k}(\Omega \times \Omega)$ jest warunkowo dodatnio określoną funkcją jądrową na zbiorze Ω względem przestrzeni $\pi_m(\mathbb{R}^d)$. Zbiór $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ jest $\pi_m(\mathbb{R}^d)$ -unisolwentny. Niech $f \in \mathcal{N}_\Phi(\Omega)$ a jej interpolant oznaczmy przez $s_{f,\mathbf{X}}$. Wówczas dla każdego $\mathbf{x} \in \Omega$ błąd interpolacji funkcji f może być ograniczony przez

$$|f(\mathbf{x}) - s_{f,\mathbf{X}}(\mathbf{x})| \leq P_{\Phi,\mathbf{X}}(\mathbf{x})|f|_{\mathcal{N}_\Phi(\Omega)}. \quad (3.41)$$

W niniejszym paragrafie przytaczamy za [43] twierdzenie i pewne lematy dowodu najlepszego znanego, tzn. o najmniejszych znanych stałych, ograniczenia błędu interpolacji funkcjami Gaussa.

Twierdzenie 13. (patrz [43])

Funkcje kardynalne u_j^* obliczone dla $\mathbf{x} \in \Omega$ z układu (3.34) spełniają

$$P_{\Phi, \mathbf{x}}(\mathbf{x}) = \mathcal{Q}(u^*(\mathbf{x})) \leq \mathcal{Q}(u) \quad (3.42)$$

dla każdego $u \in \mathbb{R}^N$, gdzie mamy odtwarzanie wielomianów stopnia m , zgodnie z definicją 10, tzn.

$$\sum_{j=1}^N u_j p(x_j) = p(x) \quad \text{dla każdego } p \in \pi_m(\mathbb{R}^d). \quad (3.43)$$

Mając lokalne odtwarzanie wielomianów stopnia $l \geq m$ ograniczenie funkcjonału potęgowego wyznacza się zauważając, że dla dowolnego wielomianu $p \in \pi_l(\mathbb{R}^d)$

$$\begin{aligned} p(0) &= 2 \sum_{j=1}^N u_j(x) p(\mathbf{x} - \mathbf{x}_j) + \sum_{j=1}^N \sum_{k=1}^N u_j(\mathbf{x}) u_k(\mathbf{x}) p(\mathbf{x}_j - \mathbf{x}_k) \\ &= p(0) - 2p(\mathbf{x} - \mathbf{x}) + \sum_{j=1}^N u_j(\mathbf{x}) p(\mathbf{x}_j - \mathbf{x}) \\ &= p(0) - 2p(0) + p(0) = 0. \end{aligned}$$

Korzystając z własności zerowania się funkcji kardynalnych z definicji 10 łatwo pokazać (patrz [42]), że

$$P_{\Phi, \mathbf{x}}^2 \leq F(u(\mathbf{x})) \leq (1 + C_1)^2 \|\Phi - p\|_{L_\infty(B(0, 2C_2 h_{\mathbf{x}, \Omega}))}. \quad (3.44)$$

Dalej dla funkcji radialnej $\Phi = \phi(\|\cdot\|_2)$ użyjemy faktu, że każdy wielomian $p \in \pi_k(\mathbb{R})$ można zastąpić wielomianem $p(\|\cdot\|_2^2) \in \pi_{2k}(\mathbb{R}^d)$, uzyskując ograniczenie

$$P_{\Phi, \mathbf{x}}^2(\mathbf{x}) \leq \max_{0 \leq t \leq 4C_2^2 h^2} |\phi(\sqrt{t}) - p(t)|$$

jeśli $h = h_{\mathbf{x}, \Omega} \leq h_0(2k)$.

Dla zbioru Ω będącego kulą o promieniu R twierdzenie o odtwarzanie wielomianów redukuje się do

Lemat 7. (patrz [43])

Założmy, że $\Omega = B(\mathbf{x}_0, R)$ jest kulą o promieniu $R > 0$. Niech $l \in \mathbb{N}$. Dla ustalonego C takiego, że

$$C \geq \frac{8(2 + \sqrt{3})}{\sqrt{3}} l^2$$

Wówczas istnieje lokalne odtwarzanie wielomianów stopnia l ze stałymi $h_0 = R/C$, $C_1 = 2$, $C_2 = 2C$ zgodnie z definicją 10.

Ograniczenie błędu opisuje poniższe twierdzenie z wielkościami c_0 i c_2 takimi, że $h_0 = c_0/l^2$, $C_2 = c_2 l^2$ następujące twierdzenie

Twierdzenie 14. (patrz [43])

Założmy, że na zbiorze Ω mamy lokalne odtwarzanie wielomianów stopnia l dla każdego $l \in \mathbb{N}$ ze stałymi $h_0 = c_0/l^2$, C_1 niezależnego od l , $C_2 = c_2 l^2$. Niech $\Phi(r) = e^{-\alpha \|r\|_2^2}$ z $\alpha > 0$. Określmy $\tilde{c}_0 := \min\{c_0, (32\alpha c_2^2)^{-2/3}\}$. Jeśli $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ spełnia $h_{\mathbf{X}, \Omega} \leq \min\{1, \tilde{c}_0/4\}$ to funkcjonal potęgowy może być ograniczony przez

$$P_{\mathbf{X}, \Omega}^2(\mathbf{x}) \leq \frac{(1 + C_1)^2}{\sqrt{2\pi}} h^{\frac{\sqrt{\tilde{c}_0}}{4\sqrt{h}}}$$

dla każdego $\mathbf{x} \in \Omega$. Tak więc dla $f \in \mathcal{N}_{\Phi}(\Omega)$ błąd interpolacji może być ograniczony wyrażeniem

$$|f(\mathbf{x}) - s_{f, \mathbf{X}}(\mathbf{x})| \leq \frac{1 + C_1}{(2\pi)^{1/4}} h^{\frac{\sqrt{\tilde{c}_0}}{8\sqrt{h}}} \|f\|_{\Phi}.$$

Ograniczenia błędu dla funkcji z przestrzeni Sobolewa

Inne podejście do konstrukcji ograniczenia błędu interpolacji lub aproksymacji za pomocą bazowych funkcji radialnych można zastosować dla funkcji z przestrzeni natywnej $\mathcal{N}_{\Phi}(\Omega)$ pokrywającej się z przestrzenią Sobolewa (patrz [24]). Twierdzenie o ograniczeniu błędu interpolacji funkcji z przestrzeni $W_2^{\tau}(\Omega)$ cytujemy za [43]

Twierdzenie 15. (patrz [43])

Założmy, że $\Omega \subseteq \mathbb{R}^d$ jest ograniczony, i ma brzeg spełniający warunek Lipschitza, i spełnia warunek stożka wewnętrznego z promieniem r i kątem θ . Niech $\mathbf{X} \subseteq \Omega$ będzie zbiorem dyskretnym centrów a $s_{f, \mathbf{X}}$ będzie interpolantem. Założmy, że Φ jest funkcją jądrową przestrzeni Sobolewa, tzn. spełnia (3.15) z $\tau = k + s$ gdzie k jest dodatnią liczbą naturalną a $0 < s \leq 1$. Jeśli $m \in \mathbb{N}_0$ spełnia $k > m + d/2$, to błąd pomiędzy $f \in W_2^{\tau}(\Omega)$ a interpolantem $s_{f, \mathbf{X}}$ może być ograniczony

$$|f - s_{f, \mathbf{X}}|_{W_q^m} \leq C h_{\mathbf{X}, \Omega}^{\tau - m - d(1/2 - 1/q)_+} \|f\|_{W_2^{\tau}(\Omega)}$$

dla dostatecznie gęstego zbioru $\mathbf{X}$.

W dowodzie tego twierdzenia nie wykorzystuje się funkcjonału potęgowego. Korzysta się natomiast z

1. postaci normy, która zawiera operatory różniczkowania D^k $k \leq |\alpha|$,
2. twierdzenia o odtwarzaniu pochodnych wielomianów, tzn. z uogólnienia twierdzenia o odtwarzaniu wielomianów
3. dla funkcji

$$u(\mathbf{x}) = f(\mathbf{x}) - s_{f, \mathbf{X}}(\mathbf{x}) \tag{3.45}$$

z aproksymacji uśrednionym wielomianem Taylora

$$Q_k u(\mathbf{x}) := \sum_{|\alpha| < k} \frac{1}{\alpha!} \int_{B(0, \rho)} D^{\alpha} u(\mathbf{y}) (\mathbf{x} - \mathbf{y})^{\alpha} \phi_{\rho}(\mathbf{y}) d\mathbf{y} \tag{3.46}$$

gdzie $\phi_{\rho} \in C_0^{\infty}(\mathbb{R}^d)$ ma nośnik $B(0, \rho)$ i w sensie całkowym przybliża jedynekę.

Korzystając z powyższych własności dowodzi się że, dla $u \in W_p^{k+s}(\Omega)$, która znika na punktach ze zbioru $\mathbf{X}$ dla $k > d/p - m$, zachodzi kluczowe ograniczenie

$$|u|_{W_q^m(\Omega)} \leq ch^{k+s-|\alpha|-d(1/p-1/q)+} |u|_{W_p^{k+s}(\Omega)}, \quad (3.47)$$

gdzie h jest miarą gęstości siatki (3.25). Dowód polega na podzieleniu dziedziny Ω na zbiory gwieżdzisto-kształtne $\mathcal{D}$ o średnicy $\mathcal{O}(h_{X,\Omega})$. Na każdym obszarze $\mathcal{D}$ korzysta się z aproksymacji uśrednionym wielomianem Taylora.

Definicja 13.

Zbiór $\mathcal{D}$ nazywamy gwieżdzisto-kształtnym względem kuli $B(\mathbf{x}_c, \rho) := \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{x}_c\| \leq \rho\}$ jeśli dla każdego $\mathbf{x} \in \mathcal{D}$ domknięta otoczka wypukła $\{\mathbf{x}\} \cup B$ jest zawarta w $\mathcal{D}$. Jeśli $\mathcal{D}$ jest ograniczone to jego parametr kawałkowatości γ_c (ang. *chunkiness parameter*) jest zdefiniowany jako stosunek średnicy $d_{\mathcal{D}}$ do promienia $\rho_{\max}$ największej kuli względem, której zbiór $\mathcal{D}$ jest gwieżdzisto-kształtny.

Zbiór gwieżdzisto-kształtny spełnia warunek stożka o czym mówi lemat.

Lemat 8. (patrz [43])

Jeśli $\mathcal{D}$ jest ograniczony i jest gwieżdzisto-kształtny względem kuli $B(\mathbf{x}_c, \rho)$ i zawarty w kuli $B(\mathbf{x}_c, R)$ to spełnia warunek stożka wewnętrznego z promieniem ρ i kątem $\vartheta = 2\arcsin[\rho/(2R)]$.

Obszarami gwieżdzisto-kształtnymi pokrywamy dziedzinę Ω . Na każdym takim obszarze aproksymujemy funkcje u za pomocą wielomianu (3.46). Dowodzi się ograniczenia błędu takiej aproksymacji na każdym obszarze $\mathcal{D}$.

Lemat 9. (patrz [43])

Niech $0 < s \leq 1$ i $m \in \mathbb{N}$. Niech $1 < p < \infty$ i $k > m + d/p$ lub $p = 1$ i $k \geq m + d$. Dla $u \in W_p^{k+s}(\mathcal{D})$ mamy

$$\|u - Q_{k+1}u\|_{W_{\infty}^m(\mathcal{D})} \leq C(1 + \gamma_c)^{d(1+1/p)} d_{\mathcal{D}}^{k+s-m-d/p} |u|_{W_p^{k+s}(\mathcal{D})},$$

ze stałą $C > 0$ zależną tylko od k, d i p .

W dalszej części dowodu korzysta się z faktu, że funkcje u znikają na zbiorze $\mathbf{X}$. Korzysta się przy tym, z własności odtwarzania pochodnych wielomianu (uogólnienie twierdzenia 9). Mamy więc następujące lematy

Lemat 10. (patrz [43])

Niech k będzie stałą dodatnią, $1 \leq p < \infty, 0 < s \leq 1, 1 \leq q \leq \infty$, i niech $m \in \mathbb{N}_0$, spełniając $1 < p < \infty$ i $k > m + d/p$ lub $p = 1$ i $k \geq m + d$. Niech również $\mathbf{X} \subseteq \mathcal{D}$ będzie zbiorem dyskretnym spełniającym warunek 1. i 2. odtwarzania wielomianów z twierdzenia 9 z $h > 0$. Jeśli $u \in W_p^{k+s}(\mathcal{D})$ spełnia $u|_{\mathbf{X}} = 0$, to

$$|u|_{W_q^m(\mathcal{D})} \leq C d_{\mathcal{D}}^{k+s-m+d(1/q-1/p)} |u|_{W_p^{k+s}(\mathcal{D})}.$$

Przy czym stała C zależy tylko od k, d, p, m i kąta ϑ z warunku stożka dla $\mathcal{D}$.

Przejsście do jednorodnego ograniczenia na całej dziedzinie Ω wymaga określenia parametrów pokrycia zbioru Ω obszarami $\mathcal{D}$.

Lemat 11. (patrz [43])

Wprowadźmy wielkości

$$\begin{aligned}\vartheta &:= 2\arcsin\left(\frac{\sin\theta}{4(1+\sin\theta)}\right), \\ Q(k, \theta) &:= \frac{\sin\theta \sin\vartheta}{8k^2(1+\sin\theta)(1+\sin\vartheta)}, \\ R &:= Q(k, \theta)^{-1}, \\ \rho &:= \frac{\sin\theta}{2(1+\sin\theta)}R.\end{aligned}$$

Ponadto zdefiniujmy zbiory

$$T_\rho := \left\{ \mathbf{t} \in (2\rho/\sqrt{d})\mathbb{Z}^d : B(\mathbf{t}, \rho) \subseteq \Omega \right\},$$

oraz

$$\mathcal{D}_t = \{\mathbf{x} \in \Omega : \text{co}(\{\mathbf{x}\} \cup B(\mathbf{t}, \rho)) \in \Omega \cap B(\mathbf{t}, R)\}, \quad \text{dla } \mathbf{t} \in T_\rho$$

gdzie $\text{co}(A)$ jest otoczką wypukłą zbioru A .

Załóżmy, że $h = h_{\mathbf{X}, \Omega}$ spełnia $h \leq Q(k, \theta)r$. Następujące stwierdzenia są prawdziwe

1. każdy zbiór $\mathcal{D}_t$ jest gwiazdźdźisto-kształtny względem kuli $B(\mathbf{t}, \rho) \subseteq \mathcal{D}_t \subseteq \Omega \cap B(\mathbf{t}, R)$,
2. każdy zbiór $\mathcal{D}_t$ spełnia warunek stożka z kątem ϑ z promieniem ρ ,
3. $\Omega = \bigcup_{\mathbf{t} \in T_\rho} \mathcal{D}_t$ i $d_{\mathcal{D}_t} \leq 2R = 2h/Q(k, \theta)$,
4. $\sum_{\mathbf{t} \in T_\rho} \chi_{\mathcal{D}_t} \leq M_1$,
5. $|T_\rho| \leq M_2 \rho^{-d}$,

gdzie χ_B oznacza funkcję charakterystyczną zbioru B a stałe M_1, M_2 zależą tylko od k, θ i d .

Korzystając z tej geometrycznej konstrukcji pokrycia zbioru Ω dla $\mathbf{X} \subseteq \Omega$ z gęstością siatki h spełniającą $h < Q(k, \theta)r$ otrzymujemy w wyniku ograniczenie (3.47) na całym zbiorze Ω .

Korzystając z (3.47) dla $h < Q(k, \theta)r$ i $p = q = 2$ otrzymujemy wynik z twierdzenia 15.

Ograniczenia błędu dla zagadnienia aproksymacji

Analogiczny mechanizm dowodzenia jak dla interpolacji można zastosować do wyznaczenia ograniczenia błędu zregularyzowanej aproksymacji średniokwadratowej, tzn. rozwiązania zagadnienia (3.5) z $\Phi \in W_p^\tau$ i λ takiego, że $|f(\mathbf{x}_j) - s_{f,\mathbf{X}}(\mathbf{x}_j)| \leq \epsilon$ dla $1 \leq j \leq N$. Wówczas mamy ograniczenie

$$\|f - s_{f,\mathbf{X}}\|_{L_\infty(\Omega)} \leq Ch_{\mathbf{X},\Omega}^{\tau-d/p} |f|_{W_p^\tau(\Omega)} + 2\epsilon.$$

Wynika to z twierdzenia analogicznie dowodzonego co twierdzenie 15.

Twierdzenie 16. (patrz [41])

Założmy, że Ω jest ograniczony i spełnia warunek stożka. Niech k będzie stałą dodatnią, $0 < s \leq 1$, $1 \leq p < \infty$, $1 \leq q \leq \infty$ i niech $m \in \mathbb{N}_0$ spełnia $k > m + d/p$ dla $p > 1$ lub dla $p = 1$, $k \geq m + d$. Ponadto, niech $\mathbf{X} \subseteq \Omega$ będzie zbiorem dyskretnym z gęstością siatki h spełniającą $h < Q(k, \theta)r$. Jeśli $u \in W_p^{k+s}(\Omega)$ spełnia $|u|_{\mathbf{X}} \leq \epsilon$ to mamy ograniczenie

$$|u|_{W_q^m(\Omega)} = C \left(h^{k+s-m-d(1/p-1/q)+} |u|_{W_p^{k+s}(\Omega)} + h^{-m} \|u\|_{\mathbf{X}} \right), \quad (3.48)$$

gdzie $(x)_+ = x$ jeśli $x \geq 0$ lub 0 w przeciwnym wypadku. Stała C zależy tylko od k, d, p, q, m i θ .

Rozważając więc zagadnienie (3.5) możemy sformułować następujące twierdzenie

Twierdzenie 17. (patrz [41])

Założmy, że $\mathcal{H} \subseteq C(\Omega)$ jest unormowaną przestrzenią liniową funkcji ciągłych. Założmy ponadto, że $L : \mathcal{H} \rightarrow \mathcal{G}$ jest przekształceniem liniowym w przestrzeń unormowaną $\mathcal{G}$. Niech $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subseteq \Omega$ i $\{y_1, \dots, y_N\} \in \mathbb{R}$ definiują funkcjonaty

$$E(s) := \sum_{j=1}^N [y_j - s(\mathbf{x}_j)]^2, \quad J(s) := \|Ls\|_{\mathcal{G}}^2, \quad s \in \mathcal{H}.$$

Założmy, że $s_\lambda \in \mathcal{H}$ jest rozwiązaniem zagadnienia

$$\min_{s \in \mathcal{H}} E(s) + \lambda J(s)$$

dla ustalonego $\lambda > 0$. Założmy ponadto, że istnieje funkcja $s_0 \in \mathcal{H}$ z $E(s_0) = 0$ i $J(s_0) \leq J(f)$. Przy powyższych założeniach prawdziwe są ograniczenia

$$\begin{aligned} J(s_\lambda) &\leq J(f) \\ |f(x_j) - s_\lambda(x_j)| &\leq \sqrt{\lambda J(f)}, \quad 1 \leq j \leq N. \end{aligned}$$

3.6.6 Ograniczenie błędu bez założenia o gęstości zbioru danych

Ograniczenia błędu oparte na ogólnej postaci (3.41) bez założeń o gęstości zbioru danych rozważane są w [12]. Opierają się one na pojęciu funkcji wzrostu wielomianów zdefiniowanej

Definicja 14.

Dla zbioru $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \subseteq \Omega$ i $\mathbf{Y} \subseteq \mathbf{X}$ definiujemy funkcję wzrostu dla przestrzeni wielomianów $\pi_q(\mathbb{R}^d)$ na zbiorze $\mathbf{Y}$ w punkcie $\mathbf{x}$ jako

$$\rho_q(\mathbf{x}, \mathbf{Y}) = \max\{|p(\mathbf{x})| : p \in \pi_q(\mathbb{R}^d), \|p|_{\mathbf{Y}}\|_{\infty} \leq 1\}. \quad (3.49)$$

Funkcja wzrostu $\rho_q(\mathbf{x}, \mathbf{Y})$ ma skończoną wartość dla każdego $\mathbf{x}$, jeśli $\mathbf{Y}$ jest $\pi_q(\mathbb{R}^d)$ -unisolwentny. W przeciwnym wypadku $\rho_q(\mathbf{x}, \mathbf{Y}) = \infty$.

Twierdzenie 18.

Założmy, że dany mamy zbiór $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, taki, że $\mathbf{X} \subseteq \Omega$ oraz $y_i = f(\mathbf{x}_i)$, $i = 1, \dots, N$ dla funkcji $f \in \mathcal{N}_{\Phi}(\Omega)$. Dla rozwiązania zadania interpolacyjnego (3.3) i dowolnego niepustego podzbioru $\mathbf{Y} \subseteq \mathbf{X}$ oraz dowolnego $q \geq \max\{m, 0\}$ mamy

$$|f(\mathbf{x}) - s_{f, \mathbf{x}}| \leq (1 + \rho_q(\mathbf{x}, \mathbf{Y})) \sqrt{E(\Phi, \pi_q^d)_{C(B(\mathbf{0}, \text{diam}(\mathbf{Y} \cup \{\mathbf{x}\})))}} \|f\|_{\mathcal{N}_{\Phi}(\Omega)}, \quad \mathbf{x} \in \mathbb{R}^d \quad (3.50)$$

gdzie

1. $\rho_q(\mathbf{x}, \mathbf{Y})$ jest funkcją wzrostu wielomianów dla przestrzeni $\pi_q(\text{co}(\mathbf{Y}))$, gdzie $\text{co}(A)$ oznacza otoczkę wypukłą zbioru A ,
2. $B(\mathbf{0}, r)$ oznacza kulę w $\mathbb{R}^d$ o środku w punkcie $\mathbf{0}$ i promieniu r ,
3. $E(\Phi, \pi_q^d)_{C(B_{\mathbf{x}, \mathbf{Y}})}$ zdefiniowane jest jako

$$E(F, \mathcal{S})_{C(G)} := \inf_{g \in \mathcal{S}} \|F - g\|_{C(G)},$$

i oznacza błąd aproksymacji jednostajnej funkcji F za pomocą funkcji z $\mathcal{S}$ określonej na $G \subset \mathbb{R}^d$. W ograniczeniu (3.50) mamy $G = B(\mathbf{0}, \text{co}(\mathbf{Y} \cup \{\mathbf{x}\}))$, $\mathcal{S} = \pi_q^d(G)$ oraz $F = \Phi$.

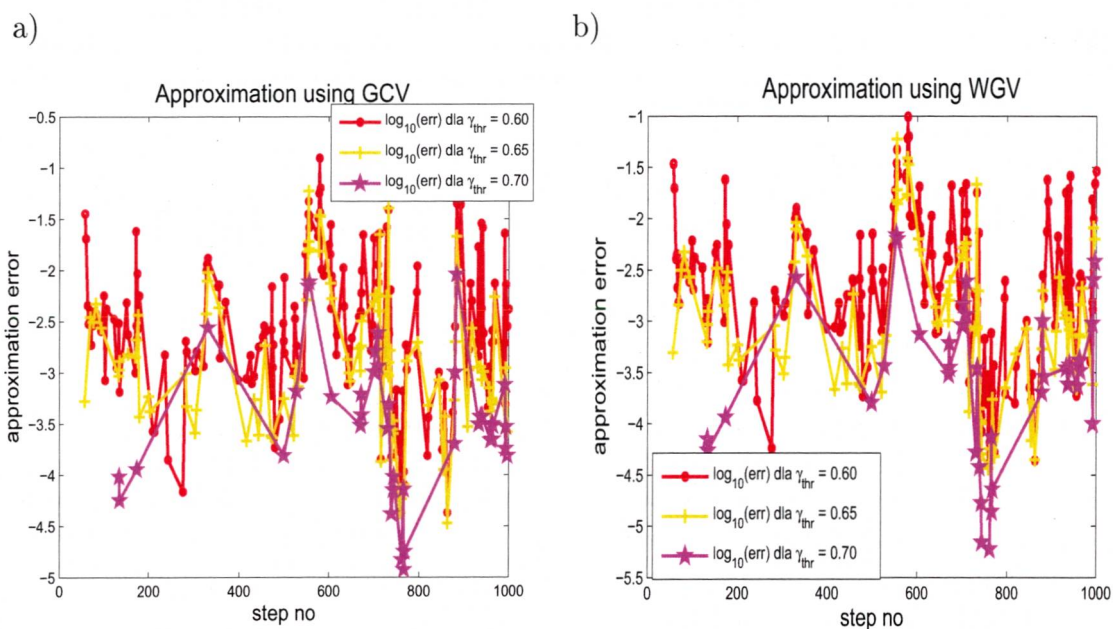
Zauważmy, że funkcja $\rho_q(\mathbf{x}, \mathbf{Y})$ jest równa funkcji Lebesgue'a (3.37), gdy $|\mathbf{Y}| = \dim(\pi_q(\mathbb{R}^d))$. Oszacowanie wartości $\rho_q(\mathbf{x}, \mathbf{X})$ dla dowolnego $\mathbf{x} \in \Omega$, gdy $\mathbf{X}$ znaleźć można w [13].

3.6.7 Ograniczenie błędu dla metody wariancji ważonej

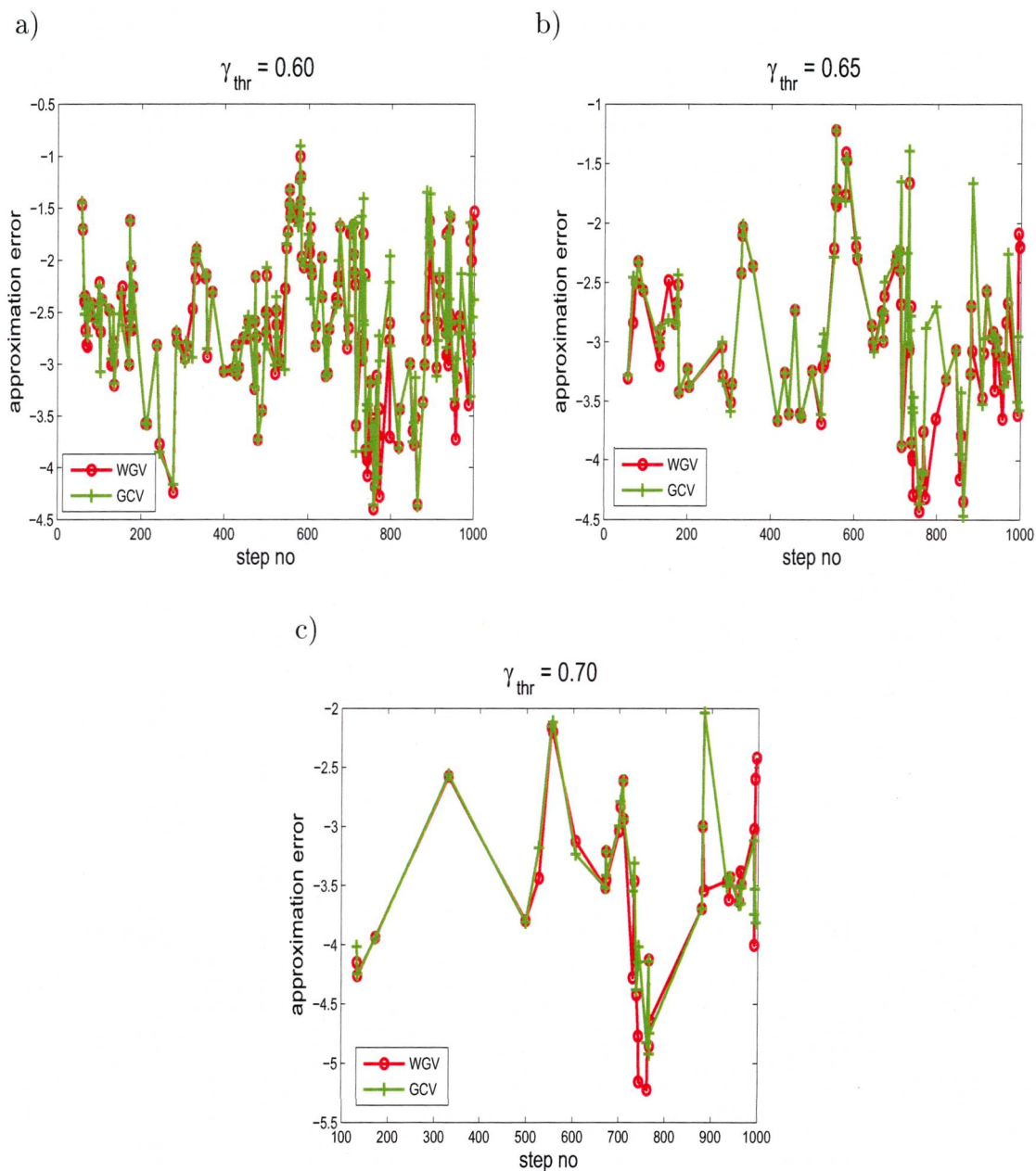
W niniejszym paragrafie pokazujemy na przykładzie trzech problemów optymalizacyjnych omawianych w [1] oraz w następnym rozdziale numeryczny argument na istnienie związku pomiędzy wartością wskaźnika $\gamma_{\mathbf{X}}(\mathbf{x})$ a błędem aproksymacji konstruowanej za pomocą metody GCV [40] oraz metody WGV. Dla każdej z trzech funkcji prezentujemy dwa rysunki. W obydwu z nich prezentujemy wykresy błędu aproksymacji dla tysiąca początkowych kroków algorytmu EXTREM w obszarach określanych przez wartość wskaźnika $\gamma_{\mathbf{X}}$ zdefiniowanego w paragrafie 3.6.1. Błąd aproksymacji w obszarze, dla którego wartość wskaźnika $\gamma_{\mathbf{X}}$ jest większa od wartości progowych $\gamma_{\mathbf{X}}(\mathbf{x}) > \gamma_{\text{thr}} = 0.6$, $\gamma_{\mathbf{X}}(\mathbf{x}) > \gamma_{\text{thr}} = 0.65$ oraz $\gamma_{\mathbf{X}}(\mathbf{x}) > \gamma_{\text{thr}} = 0.7$, obliczony został w 25 losowo wybranych punktach z otoczki wypukłej zbioru $\mathbf{X}$ z obszaru określonego przez wartość parametru γ_{thr} . Losowanie odbywało się z rozkładu normalnego względem składowych głównych zbioru danych $\mathbf{X}$.

Na rysunkach 3.12, 3.14 oraz 3.16 pokazana została zależność błędu aproksymacji od wartości progu γ_{thr} . Jak widać z wykresów na tych rysunkach błąd się zmniejsza zarówno dla metody GCV jak i dla metody WGV, gdy rośnie wartość progu γ_{thr} . Przy dużej wartości progu, takiej jak np. 0.7, punktów w których można użyć aproksymacji jest mało. Z doświadczenia wynika, i w miarę uniwersalną wartością parametru γ_{thr} jest 0.65. Dla takiej wartości wykonano obliczenia większości testów, których wyniki prezentujemy w następnym rozdziale. Na rysunkach 3.13, 3.15 oraz 3.17 zaprezentowane jest bezpośrednie porównanie jakości aproksymacji dla metod GCV oraz WGV dla ustalonych wartości progu γ_{thr} . Jak widać błąd w niektórych punktach jest mniejszy dla metody WGV. Generalnie obydwie metody w dla tych funkcji dały podobny błąd aproksymacji na punktów ze ścieżki algorytmu EXTREM. Niemniej jednak dla Algorytmu 1 lepszą zbieżność uzyskiwaliśmy metodą WGV, co prezentujemy w następnym rozdziale.

Określenie analitycznego związku pomiędzy wskaźnikiem $\gamma_{\mathbf{X}}(\mathbf{x})$ a błędem aproksymacji jedną z dwóch wymienionych metod wymaga jeszcze dalszej pracy badawczej. Rysunki 3.12, 3.14 oraz 3.16 sugerują, że związek taki istnieje. Związek taki jest innym związkiem niż zależności wymagające dużej gęstości danych. Istnienie takich zbiorów $\mathbf{X}$ oraz punktów w ich otoczce wypukłej, dla których mimo dużej wartości wskaźnika γ błąd względny aproksymacji jest duży (rzędu 10^{-1}) sugeruje, że związek taki zawierać powinien składową, która mierzy stopień unisolwentności zbioru $\mathbf{X}$. Funkcja wzrostu wielomianów (3.49) może być takim miernikiem.

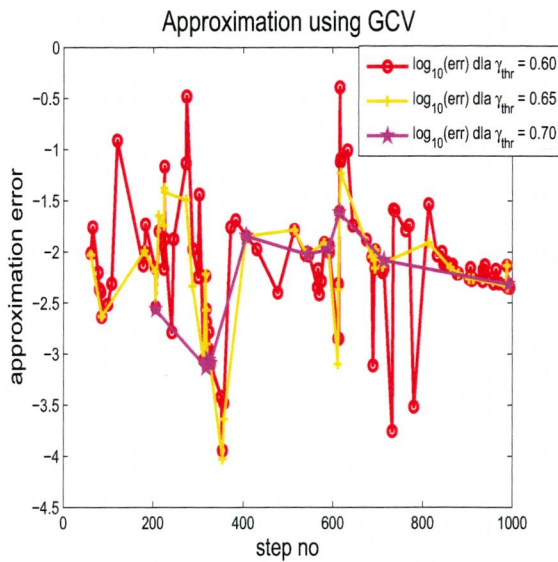


Rysunek 3.12: Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Rosenbrocka ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej (GCV), b) Weighted Gradient Variance z $NLMSE_{\text{thr}} = 10^{-5}$.

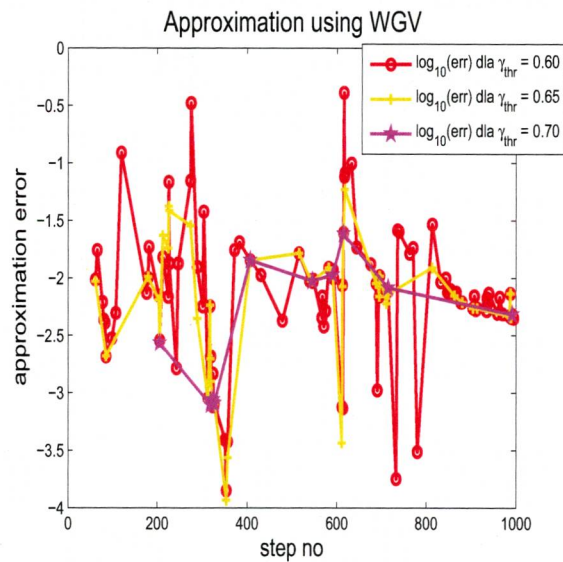


Rysunek 3.13: Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Rosenbrocka ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$.

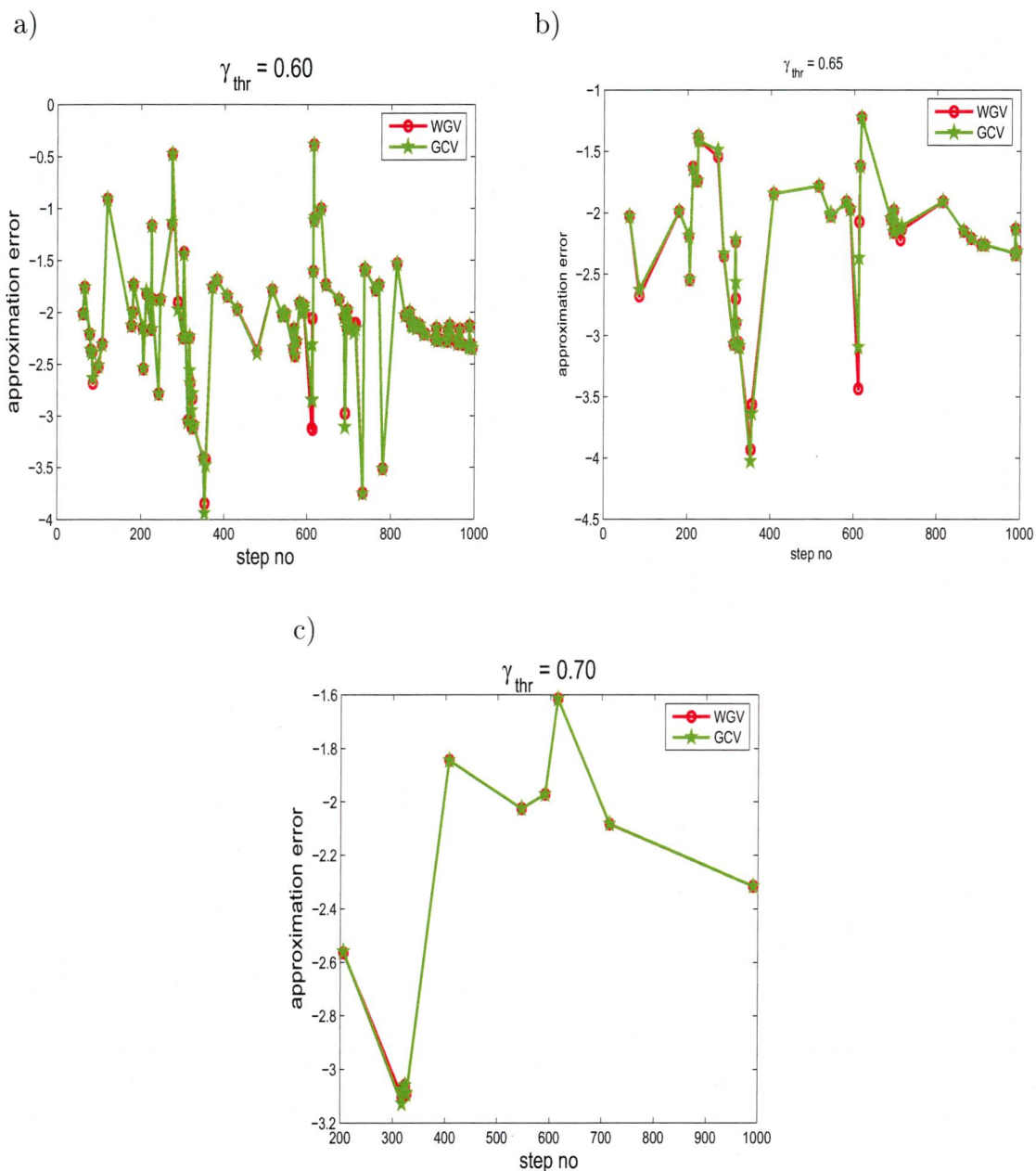
a)



b)

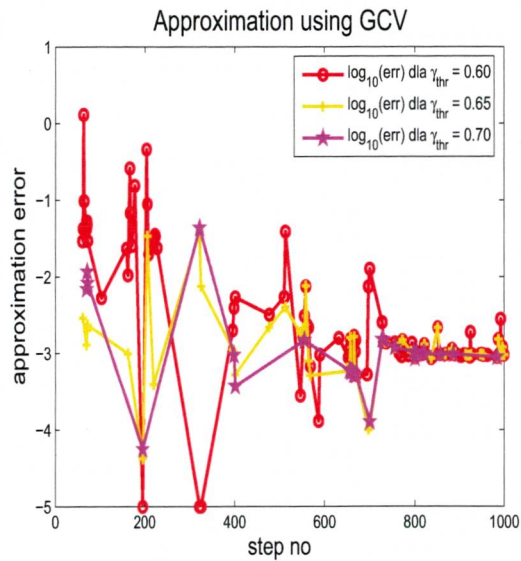


Rysunek 3.14: Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji „Chebquad” ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej, b) Weighted Gradient Variance z $NLMSE_{\text{thr}} = 10^{-5}$.

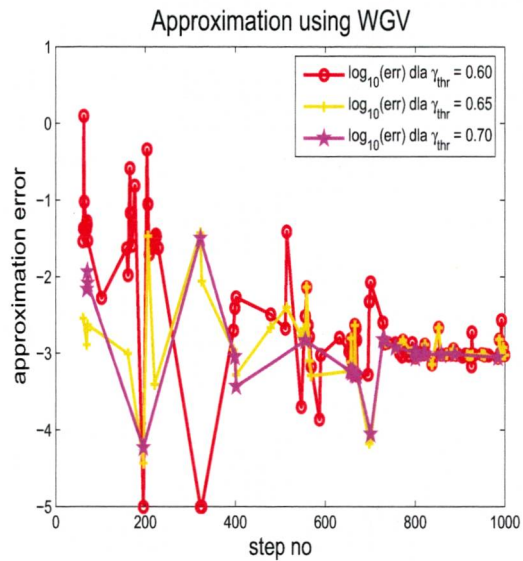


Rysunek 3.15: Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji „Chebquad” ośmiu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$.

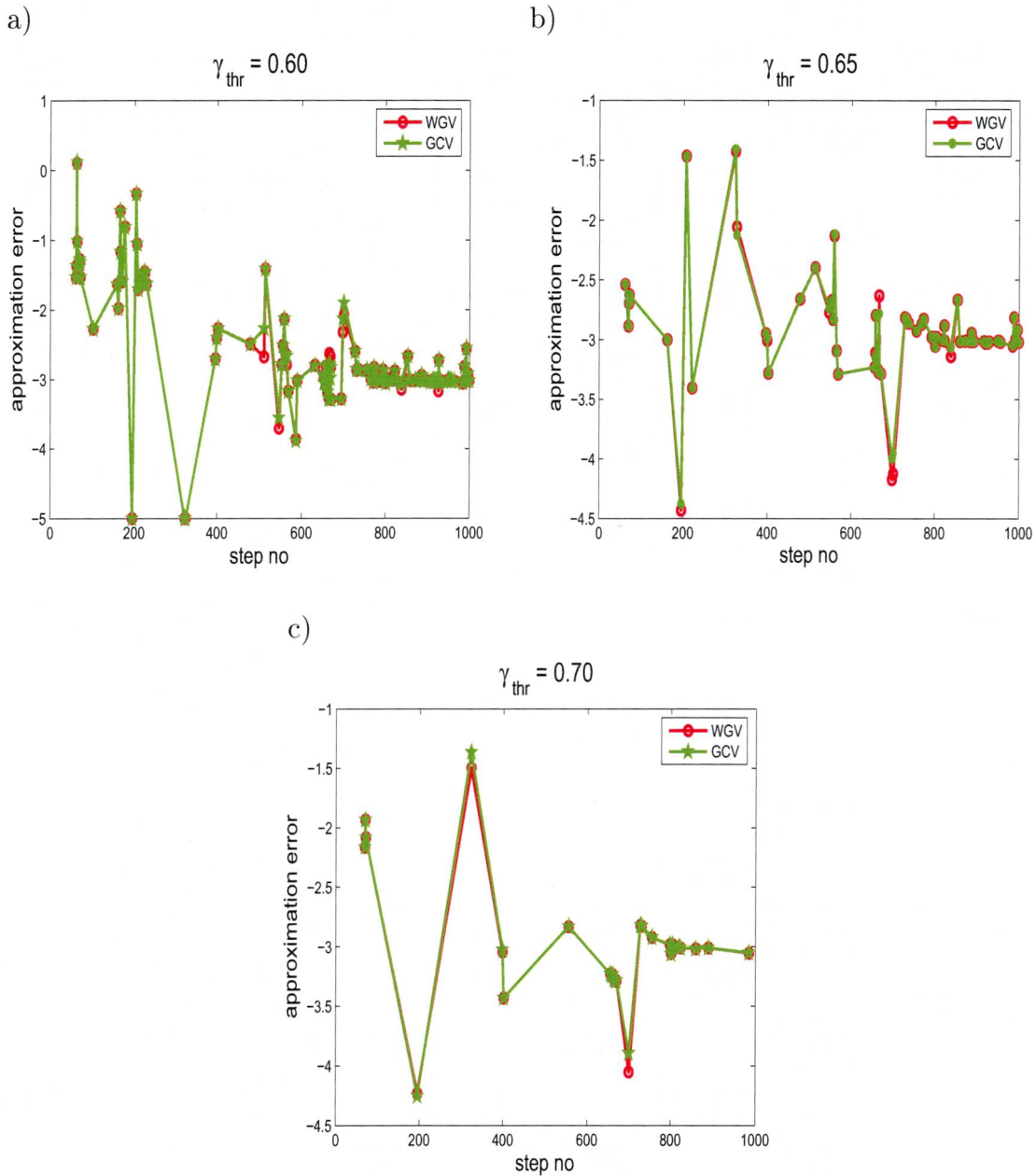
a)



b)



Rysunek 3.16: Błąd aproksymacji w punktach w początkowych tysiącu punktów ścieżki optymalizacyjnej funkcji Osborne 2 jedenastu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą a) metody uogólnionej walidacji krzyżowej, b) Weighted Gradient Variance z $NLMSE_{\text{thr}} = 10^{-5}$.



Rysunek 3.17: Błąd aproksymacji w początkowych tysiącu punktach ścieżki optymalizacyjnej funkcji Osborne 2 jedenastu zmiennych za pomocą sieci składającej się z 30 neuronów z radialnymi funkcjami aktywacji trenowanymi metodą regularyzacji z wyborem parametru λ za pomocą metody uogólnionej walidacji krzyżowej oraz Weighted Gradient Variance z $NLMSE_{thr} = 10^{-5}$ dla a) $\gamma_{thr} = 0.6$, b) $\gamma_{thr} = 0.65$, c) $\gamma_{thr} = 0.7$.

Rozdział 4

Zastosowania metody SPELROA

4.1 Optymalizacja funkcji testowych

Działania metody SPELROA prezentujemy na przykładzie ośmiu funkcji testowych ze zbioru problemów używanych do testowania algorytmów optymalizacyjnych zaprezentowanego w pracy [23]. Wszystkie funkcje w [23] są postaci

$$f(\mathbf{x}) = f([x_1, \dots, x_d]) = \sum_{i=1}^m f_i(\mathbf{x})^2,$$

gdzie m jest zależne problemu. Rozpatrzyliśmy następujące problemy o wzrastającej złożoności

1. Funkcja czterech zmiennych – funkcja kary II (problem 24 w [23])

Dla tej funkcji mamy $m = 2d$ oraz

$$\begin{aligned} f_1(\mathbf{x}) &= x_1 - 0.2 \\ f_i(\mathbf{x}) &= a^{1/2} (e^{0.1 \cdot x_i} + e^{0.1 \cdot x_{i-1}} - y_i), \quad 2 \leq i \leq n, \\ f_i(\mathbf{x}) &= a^{1/2} (e^{0.1 \cdot x_{i-n+1}} + e^{-0.1}), \quad n < i < 2n, \\ f_{2n}(\mathbf{x}) &= \left(\sum_{j=1}^n (n - j + 1) x_j^2 \right) - 1, \end{aligned}$$

gdzie $a = 10^{-5}$ oraz $y_i = e^{0.1 \cdot i} + e^{0.1 \cdot (i-1)}$. Punktem startowym jest $\mathbf{x}_0 = (0.5, \dots, 0.5)$ a minimum dla $d = 4$ wynosi $f^* = 9.37629... \cdot 10^{-4}$.

2. Funkcja sześciu i ośmiu zmiennych – rozszerzona funkcja Rosenbrocka (problem 21 w [23])

Dla tej funkcji $m = d$ i jest ona określona przez

$$\begin{aligned} f_{2i-1}(\mathbf{x}) &= 10(x_{2i} - x_{i-1}^2), \\ f_{2i}(\mathbf{x}) &= 1 - x_{2i-1}, \end{aligned}$$

Standardowym punktem startowym jest $\mathbf{x}_0 = (\xi_j)$, gdzie $\xi_{2j-1} = -1.2$ i $\xi_{2j} = 1$ a minimum wynosi $f^* = 0$ w punkcie $(1, \dots, 1)$.

3. **Funkcja ośmiu zmiennych II - funkcja „Chebquad”** (problem 35 w [23])
Jako drugą funkcję ośmiu zmiennych wybraliśmy tzw. funkcję „Chebquad”. Dla $m \geq d$ zdefiniowana jest ona przez f_i określone wzorem

$$f_i(\mathbf{x}) = \frac{1}{d} \sum_{j=1}^d T_i(x_j) - \int_0^1 T_i(x) dx,$$

gdzie T_i jest i -tym wielomianem Czebyszewa pierwszego rzędu przeskalowanym do przedziału $[0, 1]$. Standardowym punktem startowym jest $\mathbf{x}_0 = (\xi_i)$, gdzie $\xi_j = j/(d+1)$. Minimum dla $d = 8$ i $m = 8$ wynosi $f^* = 3.51687... \cdot 10^{-3}$.

4. **Funkcja dziesięciu zmiennych I – pasmowa funkcja Broydena** (problem 31 w [23])

Dla tej funkcji $m = d$ oraz

$$f_i(\mathbf{x}) = x_i(2 + 5x_i^2) + 1 - \sum_{j \in J_i} x_j(1 + x_j),$$

gdzie $J_i = \{j : j \neq i, \max(1, i-m_l) \leq j \leq \min(d, i+m_u)\}$ oraz $m_l = 5, m_u = 1$. Standardowym punktem startowym jest $\mathbf{x}_0 = (-1, \dots, -1)$. Minimum wynosi $f^* = 0$.

5. **Funkcja dziesięciu zmiennych II – dyskretna funkcja brzegowa** (problem 28 w [23])

Również dla tej funkcji $m = d$. Funkcje f_i dane są natomiast wzorami

$$f_i(\mathbf{x}) = 2x_i - x_{i-1} - x_{i+1} + h^2(x_i + t_i + 1)^3/2, \quad \text{gdzie} \\ h = 1/(d+1), t_i = ih, x_0 = x_{n+1} = 0.$$

Standardowym punktem startowym jest $\mathbf{x}_0 = (\xi_j)$, gdzie $\xi_j = t_j(t_j - 1)$. Minimum natomiast wynosi $f^* = 0$.

6. **Funkcja dziesięciu zmiennych III - funkcja kary I** (problem 23 w [23])

Dla tej funkcji mamy $m = d + 1$ oraz

$$f_i(\mathbf{x}) = a^{1/2}(x_i - 1), \quad \text{dla } 1 \leq i \leq d, \\ f_{d+1}(\mathbf{x}) = \left(\sum_{j=1}^d x_j^2 \right) - \frac{1}{4},$$

gdzie $a = 10^{-5}$. Punktem startowym jest $\mathbf{x}_0 = (\xi_j)$, gdzie $\xi_j = j$. Minimum dla $d = 10$ wynosi $f^* = 7.08765... \cdot 10^{-5}$.

7. **Funkcja jedenastu zmiennych – funkcja Osborne’a 2** (problem 19 w [23])

Jako funkcję jedenastu zmiennych wybraliśmy drugą funkcję Osborne’a. Dla tej funkcji mamy $m = 65$ a funkcje f_i dane są wzorami

$$f_i(\mathbf{x}) = y_i - (x_i e^{-t_i x_5} + x_2 e^{-(t_i - x_9)^2 \cdot x_6} + x_3 e^{-(t_i - x_{10})^2 \cdot x_7} + x_4 \cdot e^{-(t_i - x_{11})^2 \cdot x_8}),$$

gdzie $t_i = (i - 1)/10$ a y_i dla $i = 1, \dots, 65$ są stałymi, których wartości można znaleźć w [23]. Standardowym punktem startowym jest $\mathbf{x}_0 = [1.3, 0.65, 0.65, 0.7, 0.6, 3, 5, 7, 2, 4.5, 5.5]$ a minimum dla $d = 10$ wynosi $f^* = 4.01377... \cdot 10^{-2}$.

8. Funkcja piętnastu zmiennych – funkcja dyskretnego równania całkowego (problem 29 w [23])

Dla tej funkcji $m = d$ a funkcje f_i są określone wzorem

$$f_i(\mathbf{x}) = x_i + h \left[(1 - t_i) \sum_{j=1}^i t_j (x_j + t_j + 1)^3 + \right. \\ \left. t_i \sum_{j=i+1}^d (1 - t_j) (x_j + t_j + 1)^3 \right]^2, \quad \text{gdzie} \\ h = 1/(d + 1), t_i = ih, x_0 = x_{n+1} = 0.$$

Standardowym punktem startowym jest $\mathbf{x}_0 = (\xi_j)$, gdzie $\xi_j = t_j(t_j - 1)$. Minimum natomiast wynosi $f = 0$.

4.1.1 Wyniki numeryczne

Funkcja kary II czterech zmiennych

Tabela 4.1: Optymalizacja funkcji kary II czterech zmiennych za pomocą: a) algorytmu EXTREM, b) Algorytm 1 bez regularyzacji z $\gamma_{\text{thr}} = 0.65$. bez procedury ε -check, c) Algorytm 1 bez regularyzacji z $\gamma_{\text{thr}} = 0.65$ i z procedurą ε -check z $\varepsilon = 0.001$.

a)

	EXTREM	
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
50	0.0340864	0.002038
100	0.000009	0.000012
156	0.000000	0.000009

b)

	Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
50	0.0340864	0.002038	—
100	0.001679	0.000115	9
136	0.0020170	0.000010	13

c)

	Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
50	0.034086	0.002038	—
100	0.000029	0.000045	6
140	0.0000007	0.000010	15

Tabela 4.2: Optymalizacja funkcji kary II czterech zmiennych za pomocą Algorytmu 1 bez procedury ε -check z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 5 \cdot 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)

	Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
50	0.034086	0.002038	—
100	0.006357	0.002574	9
147	0.007686	0.000013	17

b)

	Alg. 1 z WGV, bez ε -check, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
50	0.034086	0.002038	—
100	0.002248	0.000102	7
130	0.001827	0.000012	11

Tabela 4.3: Optimalizacja funkcji kary II czterech zmiennych za pomocą Algorytmu 1 z procedurą ε -check z $\varepsilon = 10^{-3}$ oraz z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 5 \cdot 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)				b)			
Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
50	0.034086	0.002038	–	50	0.034086	0.002038	–
100	0.002158	0.004708	5	100	0.002141	0.004675	4
140	0.000007	0.000010	15	140	0.000007	0.000010	15

Tabela 4.3a: Przyspieszenie dla funkcji kary II czterech zmiennych uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	16.17%	15.00%	bez reg.	26.92%	23.71%
GCV	17.68%	14.28%	GCV	22.43%	23.07%
WGV	13.84%	13.57%	WGV	28.20%	22.43%

W przykładzie tym w tabel 4.2 i 4.3 widać, że zbieżność Algorytmu 1 do właściwego minimum jest znacznie gorsza dla bez użycia metody ε -check.

Rozszerzona funkcja Rosenbrocka sześciu zmiennych

Tabela 4.4: Optymalizacja rozszerzonej funkcji Rosenbrocka sześciu zmiennych za pomocą: a) algorytmu EXTREM, b) Algorytm 1 bez regularyzacji z $\gamma_{\text{thr}} = 0.65$ i bez procedury ε -check, c) Algorytm 1 bez regularyzacji z $\gamma_{\text{thr}} = 0.65$ i z procedurą ε -check z $\varepsilon = 0.001$.

a)

EXTREM		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	14.52000	72.60000
250	5.143157	2.885768
500	0.423567	0.109817
750	0.174880	0.031570
1000	0.060773	0.012666
1250	0.000899	0.001228
1523	0.000004	0.000001

b)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	14.52000	72.60000	—
250	6.047256	3.711824	22
500	1.738173	0.559350	45
750	0.455954	0.089372	35
933	0.017111	0.005606	10

c)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	14.52000	72.60000	—
250	6.034894	3.698337	5
500	1.733664	0.558246	15
750	0.433131	0.091551	40
1034	0.011905	0.002555	44

Tabela 4.5: Optymalizacja rozszerzonej funkcji Rosenbrocka sześciu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą GCV: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	14.52000	72.60000	—	1	14.52000	72.60000	—
250	5.328796	4.275823	9	251	5.636797	3.738582	16
500	1.242564	0.340772	24	500	1.198926	0.334819	28
750	0.332196	0.077669	48	750	0.606274	0.162569	48
908	0.038978	0.009351	30	1000	0.047958	0.009666	47
				1288	0.000652	0.000131	54

Tabela 4.6: Optymalizacja rozszerzonej funkcji Rosenbrocka sześciu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą WGV z $NLMSE = 5 \cdot 10^{-6}$: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

Alg. 1 WGV, bez ε -check, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	14.52000	72.60000	—	1	14.52000	72.60000	—
250	5.148554	2.890378	9	250	4.507976	2.973086	7
500	0.576409	0.145387	14	500	1.501108	0.556824	22
750	0.092298	0.017537	23	750	0.297285	0.448120	31
1000	0.004710	0.001764	21	1000	0.047480	0.010023	19
1298	0.000012	0.000002	26	1250	0.005590	0.001135	22
				1500	0.000046	0.002945	26
				1712	0.000005	0.000001	12

Tabela 4.6a: Przyspieszenie dla funkcji Rosenbrocka sześciu zmiennych uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	niezbędny	niezbędny	bez reg.	niezbędny	niezbędny
GCV	niezbędny	14.98%	GCV	niezbędny	28.10%
WGV	7.16%	—	WGV	20.87%	—

Jak widać z tabel 4.4, 4.5, 4.6 nie uzyskano zbieżności z użyciem aproksymacji RBF bez regularyzacji zarówno bez procedury ε -check jak i z tą procedurą oraz z regularyzacją z metodą GCV bez procedury ε -check. Dla aproksymacji z metodą WGV z $NLMSE = 5 \cdot 10^{-6}$ z procedurą ε -check z $\varepsilon = 10^{-3}$ możemy mówić tylko o przyspieszeniu względem ścieżki z algorytmu EXTREM. W tym wypadku algorytm z aproksymacją potrzebował więcej iteracji niż algorytm EXTREM. Możemy jednak mówić o przyspieszeniu dla tej samej liczby kroków. Wyniosło ono 8.33%.

Rozszerzona funkcja Rosenbrocka ośmiu zmiennych

Table 4.7: Optymalizacja rozszerzonej funkcji Rosenbrocka ośmiu zmiennych za pomocą: a) algorytmu EXTREM, b) Algorytm 1 bez regularyzacji i bez procedury ε -check, c) Algorytm 1 bez regularyzacji i z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{thr} = 0.65$.

a)

EXTREM		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	19.36000	96.80000
250	15.39404	10.438226
500	7.462217	5.606541
750	3.136710	1.032648
1000	1.009230	0.251387
1250	1.018545	0.050215
1500	0.075592	0.015386
1750	0.068872	0.012950
2000	0.043427	0.007465
3471	0.000000	0.000000

b)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{thr} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	19.36000	96.80000	—
250	11.59690	10.18228	13
500	5.753235	3.502973	15
750	2.594045	0.795321	30
1039	0.732554	0.189228	69

c)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{thr} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	19.36000	96.80000	—
250	9.181984	6.376938	22
500	4.639539	1.855113	23
750	1.904974	0.471618	30
1000	0.829651	0.215283	46
1250	0.130581	0.028169	43
1537	0.034711	0.006926	62

Tabela 4.8: Optymalizacja rozszerzonej funkcji Rosenbrocka ośmiu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą GCV: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	19.36000	96.80000	—	1	19.36000	96.80000	—
250	10.63705	9.590320	15	250	10.63705	9.590320	15
500	4.326629	1.770927	22	500	4.253577	1.718574	24
750	1.988725	0.564489	34	750	1.587166	5.075403	44
982	0.748296	0.184506	46	1000	0.573237	3.105517	56
				1225	0.556057	0.121451	58

Tabela 4.9: Optymalizacja rozszerzonej funkcji Rosenbrocka ośmiu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą WGV z $NLMSE = 5 \cdot 10^{-6}$: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

Alg. 1 WGV, bez ε -check, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	19.36000	96.80000	—	1	19.36000	96.80000	—
250	11.84563	10.45062	9	250	10.64878	9.379780	10
500	7.176414	5.804069	11	500	5.235753	2.394219	16
750	2.509801	0.766705	25	750	1.619362	0.435302	35
1097	1.128712	0.306324	67	1000	0.589236	0.143009	78
				1250	0.255143	1.383665	42
				1500	0.021409	0.004114	51
				1861	0.000839	0.000169	48

Tabela 4.9a: Przyspieszenie dla funkcji Rosenbrocka ośmiu zmiennych uzyskane Algorytmem 1 mierzone względem liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1.

	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	niebezpieczny	niebezpieczny
GCV	niebezpieczny	niebezpieczny
WGV	niebezpieczny	15.04%

Jak widać z tabel 4.7, 4.8, 4.9 zbieżność procesu optymalizacji uzyskano tylko dla aproksymacji RBF metodą WGV z $NLMSE = 5 \cdot 10^{-6}$. Ponieważ algorytm EXTREM wymagał znacznie więcej wartościowań funkcji niż algorytm z aproksymacją funkcji celu, o przyspieszeniu jest sens mówić tylko względem ścieżki z algorytmu z aproksymacją. Uzyskane przyspieszenie wyniosło 15%.

Funkcja „Chebquad” ośmiu zmiennych

Tabela 4.10: Optymalizacja funkcji „Chebquad” ośmiu zmiennych za pomocą: a) algorytmu EXTREM, b) Algorytm 1 bez regularyzacji i bez procedury ε -check, c) Algorytm 1 bez regularyzacji i z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)			
EXTREM			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	
1	0.026086	0.038617	
250	0.009547	0.006187	
500	0.003495	0.004681	
750	0.000271	0.003698	
1000	0.000000	0.003517	
1237	0.000000	0.003517	

b)			
Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.026086	0.038617	—
250	0.009686	0.006216	4
500	0.003531	0.004684	18
750	0.000098	0.003584	10
1000	0.000000	0.003517	48
1197	0.000000	0.003517	97

c)			
Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.026086	0.038617	—
250	0.007716	0.006214	19
500	0.000538	0.003873	42
805	0.000001	0.003518	80

Tabela 4.11: Optymalizacja funkcji „Chebquad” ośmiu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą GCV: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)			
Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.026086	0.038617	—
250	0.009686	0.006216	4
500	0.003537	0.004684	17
750	0.000076	0.003584	11
1095	0.000002	0.003518	188

b)			
Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.026086	0.038617	—
250	0.008672	0.006552	18
500	0.000459	0.003855	33
750	0.000000	0.003517	60
1000	0.000000	0.003517	73
1117	0.000000	0.003517	94

Tabela 4.12: Optymalizacja funkcji „Chebquad” ośmiu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą WGV z $NLMSE = 5 \cdot 10^{-6}$: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{thr} = 0.65$.

a)				b)			
Alg. 1 WGV, bez ε -check, $\gamma_{thr} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{thr} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.026086	0.038617	—	1	0.026086	0.038617	—
250	0.006810	0.005913	18	250	0.007802	0.006399	13
500	0.000084	0.003599	41	500	0.000976	0.004250	12
772	0.000000	0.003517	75	750	0.000001	0.003517	20
				808	0.000001	0.003517	6

Tabela 4.12a: Przyspieszenie dla funkcji „Chebquad” uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	14.79%	17.52%	bez reg.	17.34%	47.24%
GCV	20.09%	24.89%	GCV	29.26%	32.17%
WGV	17.36%	6.31%	WGV	48.42%	38.80%

Pasmowa funkcja Broydena dziesięciu zmiennych

Tabela 4.13: Optymalizacja pasmowej funkcji Broydena dziesięciu zmiennych za pomocą: a) algorytmu EXTREM, b),c) Algorytm 1 bez regularyzacji bez ε -check, d),e) Algorytm 1 bez regularyzacji z ε -check, $\varepsilon = 10^{-3}$.

a)			
EXTREM			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	
1	1.934202	360.0000	
200	0.713609	3.079440	
400	0.001211	0.060074	
600	0.000000	0.000000	

b)			
Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	0.728666	3.175358	27
400	0.000269	0.010127	37
600	0.000000	0.000002	59
880	0.000000	0.000001	55

c)			
Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	0.989317	3.300021	20
400	0.000404	0.017227	26
563	0.000016	0.000867	36

d)			
Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	0.899728	3.236088	36
400	0.003082	0.150238	38
631	0.000011	0.000035	61

e)			
Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	0.989317	3.300021	20
400	0.000369	0.005028	25
643	0.000000	0.000001	54

Tabela 4.14 Optymalizacja funkcji pasmowej Broydena dziesięciu zmiennych za pomocą Algorytmu 1 bez procedury ε -check z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.60$.

a)			
Alg. 1 z GCV			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	0.819230	3.192419	30
400	0.000562	0.026622	33
649	0.000002	0.000001	65

b)			
Alg. 1 z WGV			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—
200	1.026973	3.309025	21
400	0.018712	0.028031	12
584	0.000000	0.000000	42

Tabela 4.15: Optymalizacja funkcji pasmowej Broydena dziesięciu zmiennych za pomocą Algorytmu 1 z procedurą ε -check z $\varepsilon = 10^{-3}$ z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.60$.

a)				b)			
Alg. 1 z GCV				Alg. 1 z WGV			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	1.934202	360.0000	—	1	1.934202	360.0000	—
200	0.894353	3.241784	34	200	1.026973	3.309025	21
400	0.013946	0.445524	29	400	0.000871	0.038417	19
600	0.000844	0.004472	55	603	0.000001	0.000005	39
751	0.000003	0.000001	52				

W tabelach (4.14) oraz (4.15) przedstawiliśmy wyniki dla $\gamma_{\text{thr}} = 0.6$ ponieważ przy $\gamma_{\text{thr}} = 0.65$ liczba iteracji aproksymowanych była mniejsza niż 10% względem liczby wartościowań funkcji na ścieżce optymalizacyjnej algorytmu EXTREM. Takie ustawienie dla γ_{thr} może być wywnioskowane z tabel (4.13).

Tabela 4.15a: Przyspieszenie dla pasmowej funkcji Broydena dziesięciu zmiennych uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	—	17.33%	bez reg.	—	17.33%
GCV	13.17%	—	GCV	13.17%	—
WGV	12.83%	12.67%	WGV	15.17%	12.67%

Dyskretna funkcja brzegowa dziesięciu zmiennych

Tabela 4.16: Optymalizacja dyskretnej funkcji brzegowej dziesięciu zmiennych za pomocą: a) algorytmu EXTREM, b),c) Algorytmu 1 bez regularyzacji bez procedury ε -check, d),e) Algorytm 1 bez regularyzacji z ε -check, $\varepsilon = 10^{-3}$.

a)

EXTREM		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	0.041335	0.000789
500	0.013041	0.000223
1000	0.002631	0.000037
1500	0.000989	0.000012
2000	0.000062	0.000001
2500	0.000000	0.000000
2926	0.000000	0.000000

b)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—
500	0.012429	0.000148	77
929	0.009100	0.000102	61

c)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—
500	0.013145	0.000154	42
1000	0.008812	0.000096	36
1500	0.000367	0.000013	56
2057	0.000000	0.000000	79

d)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—
500	0.023221	0.000253	71
1000	0.006286	0.000072	92
1500	0.001413	0.000018	93
2000	0.000447	0.000007	88
2500	0.000172	0.000002	87
3000	0.000046	0.000001	89
3325	0.000033	0.000000	59

e)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—
500	0.013145	0.000154	42
1000	0.006799	0.000081	36
1259	0.000078	0.000003	44

Tabela 4.17: Optymalizacja dyskretnej funkcji brzegowej dziesięciu zmiennych za pomocą Algorytmu 1 bez procedury ε -check z regularyzacją z wyborem wartości parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{thr} = 10^{-6}$. W obydwu wypadkach $\gamma_{thr} = 0.60$.

a)				b)			
Alg. 1 z GCV, bez ε -check, $\gamma_{thr} = 0.60$				Alg. 1 z WGV, bez ε -check, $\gamma_{thr} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—	1	0.041335	0.000789	—
500	0.014178	0.000156	66	500	0.008977	0.000113	57
903	0.008044	0.000099	72	1000	0.001176	0.000021	68
				1168	0.000628	0.000008	32

Tabela 4.18: Optymalizacja dyskretnej funkcji brzegowej dziesięciu zmiennych za pomocą Algorytmu 1 z procedurą ε -check z $\varepsilon = 10^{-3}$ z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{thr} = 10^{-6}$. W obydwu wypadkach $\gamma_{thr} = 0.60$.

a)				b)			
Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{thr} = 0.60$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{thr} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.041335	0.000789	—	1	0.041335	0.000789	—
500	0.021260	0.000246	68	500	0.008452	0.000092	63
600	0.021060	0.000237	33	1000	0.004937	0.000054	75
				1500	0.000877	0.000010	69
				2000	0.000024	0.000000	92
				2456	0.000002	0.000000	51

Ponieważ algorytm EXTREM wykonuje dla tej funkcji bardzo dużo wartościowań funkcji w okolicy minimum natomiast Algorytm 1 potrzebuje tych wartościowań znacznie mniej, to w poniższej tabeli podajemy przyspieszenie procesu tylko względem ścieżki Algorytmu 1.

Tabela 4.18a: Przyspieszenie dla pasmowej funkcji brzegowej dziesięciu zmiennych uzyskane Algorytmem 1 mierzone względem liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1.

	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	14.85%	—
GCV	niezbędny	niezbędny
WGV	13.44%	14.25%

Funkcja kary I dziesięciu zmiennych

Tabela 4.19: Optimalizacja funkcji kary I dziesięciu zmiennych za pomocą: a) algorytmu EXTREM, b),c) Algorytmu 1 bez regularyzacji bez procedury ε -check, d),e) Algorytm 1 bez regularyzacji z ε -check, $\varepsilon = 10^{-3}$.

a)

EXTREM		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	369.48325	148033.565350
500	0.039733	0.000076
1000	0.030624	0.000074
1500	0.022666	0.000073
2000	0.014931	0.000073
2500	0.003847	0.000073
3135	0.000000	0.000072

b)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	–
274	3.786860	28.611277	40

c)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	–
557	0.223471	0.000083	47

d)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	–
337	0.171107	0.045111	43

e)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
0	369.4832	$1.48 \cdot 10^5$	–
429	0.111386	0.000346	36

Tabela 4.20: Optimalizacja dyskretnej funkcji brzegowej dziesięciu zmiennych za pomocą Algorytmu 1 bez procedury ε -check z regularyzacją z wyborem wartości parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)

Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	–
587	0.230362	0.000083	60

b)

Alg. 1 z WGV, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	–
454	0.052095	0.000077	27

Tabela 4.21: Optymalizacja dyskretnej funkcji brzegowej dziesięciu zmiennych za pomocą Algorytmu 1 z procedurą ε -check z $\varepsilon = 10^{-3}$ z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-6}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.65$.

a)				b)			
Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	369.4832	$1.48 \cdot 10^5$	—	1	369.4832	$1.48 \cdot 10^5$	—
465	0.136225	0.000084	55	500	0.044475	0.000076	19
				641	0.039146	0.000075	23

Tabela 4.21a: Przyspieszenie dla pasmowej funkcji Broydena dziesięciu zmiennych uzyskane Algorytmem 1 mierzone względem liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1.

	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	niezbieżny	niezbieżny
GCV	niezbieżny	niezbieżny
WGV	5.95%	6.55%

Dla tego przykładu tylko Algorytm 1 z procedurą ε -check z metodą WGV dał odtworzenie rzędu zbieżności algorytmu EXTREM

Funkcja Osborne’a II jedenastu zmiennych

Tabela 4.22: Optymalizacja funkcji Osborne II jedenastu zmiennych za pomocą:
a) algorytmu EXTREM, b) Algorytm 1 bez regularyzacji i bez procedury ε -check,
c) Algorytm 1 bez regularyzacji i z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu
wypadkach $\gamma_{\text{thr}} = 0.65$.

a)

	EXTREM	
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	22.61258	2.093420
250	1.008919	0.081672
500	0.009882	0.041034
750	0.005407	0.041772
1000	0.000060	0.040138
1250	0.000001	0.040138
1434	0.000000	0.040138

b)

	Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	22.61258	2.093420	—
250	2.816426	0.416705	32
346	2.783954	0.342864	48

c)

	Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	22.61258	2.093420	—
250	1.379902	0.977272	29
500	0.568322	0.059211	31
710	0.335485	0.041512	43

Tabela 4.23: Optymalizacja funkcji Osborne II jedenastu zmiennych za pomocą
Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą GCV: a) bez
procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach
 $\gamma_{\text{thr}} = 0.65$.

a)

	Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	22.61258	2.093420	—
250	2.798756	0.365735	31
500	0.571149	0.046544	35
750	0.093092	0.041709	35
1000	0.005360	0.040364	36
1228	0.004148	0.040266	34

b)

	Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	22.61258	2.093420	—
250	2.799374	0.365797	32
500	0.951048	0.045721	31
750	0.246694	0.041805	43
1000	0.004632	0.040462	35
1106	0.003884	0.040177	25

Tabela 4.24: Optymalizacja funkcji Osborne II jedenastu zmiennych za pomocą Algorytmu 1 z regularyzacją z wyborem parametru λ za pomocą WGV z $NLMSE = 5 \cdot 10^{-6}$: a) bez procedury ε -check, b) z procedurą ε -check z $\varepsilon = 10^{-3}$. W obydwu wypadkach $\gamma_{thr} = 0.65$.

a)				b)			
Alg. 1 WGV, bez ε -check, $\gamma_{thr} = 0.65$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{thr} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	22.61258	2.093420	—	1	22.61258	2.093420	—
250	1.009314	0.197551	11	250	0.909002	0.106864	14
500	0.029002	0.044174	24	500	0.111992	0.047136	38
604	0.026482	0.041259	40	750	0.001946	0.040209	49
				919	0.001441	0.040185	29

Tabela 4.24a: Przyspieszenie dla funkcji Osborne'a II jedenastu zmiennych uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	niezbieżny	niezbieżny	bez reg.	niezbieżny	niezbieżny
GCV	13.93%	15.01%	GCV	26.29%	34.45%
WGV	niezbieżny	14.15%	WGV	niezbieżny	43.53%

Jak widać z powyższych tabel najlepszą zbieżnością do właściwego minimum uzyskaliśmy również za pomocą Algorytmu 1 z procedurą ε -check oraz metodą WGV do konstrukcji aproksymacji RBF funkcji celu.

Funkcja dyskretnego równania całkowego piętnastu zmiennych

Tabela 4.25: Optymalizacja funkcji dyskretnego równania całkowego piętnastu zmiennych za pomocą: a) algorytmu EXTREM, b),c) Algorytmu 1 bez regularyzacji bez procedury ε -check, d),e) Algorytm 1 bez regularyzacji z ε -check, $\varepsilon = 10^{-3}$.

a)

EXTREM		
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f
1	0.059923	0.091463
200	0.000158	0.000210
421	0.000000	0.000000

b)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000182	0.000244	18
366	0.000115	0.000509	45

c)

Alg. 1 bez regularyzacji, bez ε -check, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000171	0.000230	14
436	0.000009	0.000010	40

d)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000157	0.000213	20
400	0.000000	0.000000	56
537	0.000000	0.000000	18

e)

Alg. 1 bez regularyzacji, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.65$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000158	0.000210	16
400	0.000000	0.000000	46
547	0.000000	0.000000	20

Tabela 4.26: Optymalizacja funkcji dyskretnego równania całkowego piętnastu zmiennych za pomocą Algorytmu 1 bez procedury ε -check z regularyzacją z wyborem wartości parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-4}$. W obydwu przypadkach $\gamma_{\text{thr}} = 0.60$.

a)

Alg. 1 z GCV, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000217	0.000298	22
366	0.000150	0.000183	63

b)

Alg. 1 z WGV, bez ε -check, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	—
200	0.000157	0.000209	18
427	0.000000	0.000000	47
495	0.000000	0.000000	2

Tabela 4.27: Optymalizacja funkcji dyskretnego równania całkowego piętnastu zmiennych za pomocą Algorytmu 1 z procedurą ε -check z $\varepsilon = 10^{-3}$ z regularyzacją z wyborem parametru λ za pomocą: a) metody GCV, b) metody WGV z $NLMSE_{\text{thr}} = 10^{-4}$. W obydwu wypadkach $\gamma_{\text{thr}} = 0.60$.

a)				b)			
Alg. 1 z GCV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$				Alg. 1 z WGV, ε -check z $\varepsilon = 10^{-3}$, $\gamma_{\text{thr}} = 0.60$			
step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.	step num.	$\ \mathbf{x} - \mathbf{x}^*\ ^2$	f	num. approx.
1	0.059923	0.091463	0	1	0.059923	0.091463	—
200	0.000738	0.000824	21	200	0.000867	0.000956	18
427	0.000000	0.000000	89	427	0.000000	0.000000	31
585	0.000000	0.000000	28	467	0.000000	0.000000	10

Tabela 4.27a: Przyspieszenie dla funkcji dyskretnego równania całkowego piętnastu zmiennych uzyskane Algorytmem 1 mierzone względem a) liczby bezpośrednich wartościowań funkcji celu w ścieżce optymalizacyjnej Algorytmu 1, b) liczby wartościowań funkcji celu w ścieżce optymalizacyjnej algorytmu EXTREM.

a)			b)		
	bez ε -check	z ε -check z $\varepsilon = 10^{-3}$		bez ε -check	z ε -check z $\varepsilon = 10^{-3}$
bez reg.	17.21%	—	bez reg.	28.03%	—
GCV	23.22%	23.59%	GCV	33.25%	—
WGV	13.54%	12.63%	WGV	—	4.45%

4.1.2 Synteza wyników numerycznych dla funkcji testowych

W paragrafie tym prezentujemy zestawienie wyników uzyskanych przyspieszeń dla optymalizacji algorytmem EXTREM oraz Algorytmem 1 połączonym z algorytmem EXTREM dla funkcji testowych z paragrafu 4.1.1. Tabele 4.28a i 4.28b zawierają więc wszystkie wyniki z tabel z paragrafu 4.1.1. W tabelach tych znak „—” oznacza, że nie uzyskano przyspieszenia, gdyż potrzebnych było więcej wartościowań bezpośrednich funkcji celu niż w algorytmie EXTREM, zachowana została natomiast zbieżność algorytmu.

Oznaczmy przez NUM_{BEZP} liczbę bezpośrednich obliczeń wartości funkcji celu wykonanych przez algorytm bezpośredniemu, przez $NUM_{SPELROA}$ liczbę bezpośrednich obliczeń wartości funkcji celu wykonanych przez algorytm SPELROA oraz przez $NUM_ALL_{SPELROA}$ liczbę obliczeń bezpośrednich funkcji celu przez algorytm SPELROA plus liczba aproksymacji funkcji celu wykonanych przez ten algorytmie. Przyspieszenie podane w tabelach 4.28a) i 4.28b) obliczone jest wg wzorów

$$100\% - \frac{NUM_{SPELROA}}{NUM_{BEZP}} \quad (4.1)$$

natomiast w tabeli 4.28b) wzorem

$$100\% - \frac{NUM_{SPELROA}}{NUM_ALL_{SPELROA}}. \quad (4.2)$$

Jak widać zbieżność dla wszystkich funkcji testowych uzyskaliśmy tylko dla metody SPELROA z aproksymacją z wyborem parametru regularyzacji za pomocą metody Weighted Gradient Variance połączonej z procedurą ε -check.

Tabela 4.28a: Przyspieszenie procesu optymalizacji uzyskane metodą SPELROA dla funkcji testowych. Przyspieszenie określone względem ilości bezpośrednich wartościowań funkcji celu w Algorytmie 1, tzn. obliczone ze wzoru (4.2).

			bez regularyzacji		GCV		WGV	
	Funkcja	liczba param.	bez ε -check	z ε -check	bez ε -check	z ε -check	bez ε -check	z ε -check
1	kary II	4	16.17%	15.00%	17.68%	14.28%	13.84%	13.57%
2	Rosenbrocka rozszerzona	6	nie-zbieżny	nie-zbieżny	nie-zbieżny	14.98%	7.16%	—
3	Rosenbrocka rozszerzona	8	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	15.04%
4	Chebquad	10	14.79%	17.52%	20.09%	24.89%	17.36%	6.31%
5	Pasmowa Broydena	10	—	17.35%	13.17%	—	12.83%	12.67%
6	Dyskretna brzegowa	10	14.85%	—	nie-zbieżny	nie-zbieżny	13.44%	14.25%
7	kary I	10	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	5.95%	6.55%
8	Osborne'a	11	nie-zbieżny	nie-zbieżny	13.93%	15.01%	nie-zbieżny	14.15%
9	Dyskretnego równ. całk.	15	17.21%	—	23.22%	23.59%	13.54%	12.63%

Tabela 4.28b: Przyspieszenie procesu optymalizacji uzyskane metodą SPELROA dla funkcji testowych. Przyspieszenie określone względem ilości bezpośrednich wartościowań funkcji celu w Algorytmie EXTREM, tzn. obliczone ze wzoru (4.1).

			bez regularyzacji		GCV		WGV	
	Funkcja	liczba param.	bez ε -check	z ε -check	bez ε -check	z ε -check	bez ε -check	z ε -check
1	kary II	4	26.92%	23.71%	22.43%	23.07%	28.20%	22.43%
2	Rosenbrocka rozszerzona	6	nie-zbieżny	nie-zbieżny	nie-zbieżny	28.10%	20.87%	—
3	Rosenbrocka rozszerzona	8	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	45.55%
4	Chebquad	10	17.34%	47.24%	29.26%	32.17%	48.42%	38.80%
5	Pasmowa Broydena	10	—	17.33%	13.17%	—	15.17%	12.67%
6	Dyskretna brzegowa	10	36.98%	—	nie-zbieżny	nie-zbieżny	65.45%	28.02%
7	kary I	10	nie-zbieżny	nie-zbieżny	nie-zbieżny	nie-zbieżny	86.41%	80.89%
8	Osborne'a	11	nie-zbieżny	nie-zbieżny	26.29%	34.45%	nie-zbieżny	43.53%
9	Dyskretnego równ. całk.	15	28.03%	—	33.25%	—	—	4.45%

4.2 Projektowanie magnesów nadprzewodzących

4.2.1 Zintegrowany proces projektowania magnesów dla zderzacza cząstek elementarnych LHC

4.2.2 Określenie funkcji celu

Celem optymalizacji magnesów nadprzewodzących w zderzaczu LHC było zapewnienie wymaganych własności generowanego przez nie pola magnetycznego – zakrzywiającego, skupiającego lub korygującego trajektorię wiązki przyspieszanych cząstek elementarnych. Wymagania te są określone przez geometrię tunelu oraz fizykę akceleratora. Ogólnie rzecz biorąc w procesie konstrukcji magnesu celem jest czyste pole wielobiegunowe o określonym natężeniu. Cel ten formułowany jest w kategoriach współczynników rozwinięcia Fouriera statycznego pola magnetycznego w aperturze magnesu. Statyczne pole magnetyczne w regionach niemagnetycznych i wolnych od prądu jest opisywane przez układ uproszczonych równań Maxwella

$$\begin{aligned}\nabla \times \vec{B} &= 0, \\ \nabla \cdot \vec{B} &= 0.\end{aligned}\tag{4.3}$$

Rozwinięcie pola magnetycznego w szereg Fouriera w aperturze magnesu

Dla zagadnień dwuwymiarowych układ (4.3) we współrzędnych kartezjańskich redukuje się do układu

$$\begin{aligned}\frac{\partial B_x}{\partial y} &= \frac{\partial B_y}{\partial x}, \\ \frac{\partial B_x}{\partial z} &= -\frac{\partial B_y}{\partial y},\end{aligned}\tag{4.4}$$

który jest równoważny równaniu Laplace'a

$$\nabla \cdot (\nabla \times A_z) = 0\tag{4.5}$$

spełnianego przez składową z potencjału wektorowego A zdefiniowanego standardowo jako

$$\vec{B} = \nabla \times \vec{A}$$

w przestrzeni trójwymiarowej.

Pole magnetyczne jest obliczane z zależności $B_x = \frac{\partial A_z}{\partial y}$ oraz $B_y = -\frac{\partial A_z}{\partial x}$. Przechodząc do współrzędnych biegunowych (r, φ) równanie (4.5) przyjmuje postać

$$\frac{\partial^2 A_z}{\partial r^2} + \frac{1}{r} \frac{\partial A_z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 A_z}{\partial \varphi^2} = 0.$$

Ustalając pole dla $r = 0$ otrzymujemy rozwiązanie

$$A_z(r, \varphi) = \sum_{n=1}^{\infty} r^n (G_n \cos(n\varphi) + H_n \sin(n\varphi)).$$

Radialna i kątowa składowa pola zapisują się wówczas jako szeregi

$$\begin{aligned} B_r(r, \varphi) &= \frac{1}{r} \frac{\partial}{\partial \varphi} A_z(r, \varphi) = \sum_{n=1}^{\infty} r^{n-1} (B_n \sin(n\varphi) + A_n \cos(n\varphi)), \\ B_\varphi(r, \varphi) &= -\frac{\partial}{\partial r} A_z(r, \varphi) = \sum_{n=1}^{\infty} r^{n-1} (B_n \cos(n\varphi) - A_n \sin(n\varphi)), \end{aligned} \quad (4.6)$$

gdzie $B_n = -nG_n$ i $A_n = nH_n$. Jakość pola w aperturze jest formułowana za pomocą współczynników rozwinięcia składowej radialnej (4.6) na zadanym promieniu referencyjnym $r = r_{\text{ref}}$. Dla magnesów zderzacza cząstek elementarnych LHC $r_{\text{ref}} = 17\text{mm}$.

Składowa pola generowana przez cewkę nadprzewodzącą

Pole magnetyczne w aperturze magnesu może zostać obliczone bez dyskretyzacji cewki nadprzewodzącej korzystając z prawa Biota-Sevarta oraz zredukowanego potencjału wektorowego $\vec{A}_r$, i.e.

$$\vec{B} = \vec{B}_s + \vec{B}_r = \mu_0 \vec{H}_s + \nabla \times \vec{A}_r, \quad (4.7)$$

gdzie $\vec{B}_s$ jest polem generowanym przez cewkę natomiast $\vec{B}_r$ jest polem generowane przez jarzmo. Pole magnetyczne $\vec{B}_s$ pobudzone przez cewkę o zadanym rozkładzie prądu może zostać obliczone korzystając z prawa Biota-Sevarta, które w trzech wymiarach zapisuje się

$$\vec{B}_s = \frac{\mu_0}{4\pi} \int_V \frac{\vec{J} \times \vec{r}}{r^3} dV,$$

gdzie $\vec{J}$ jest gęstością prądu, $\vec{r}$ jest wektorem z punktu całkowania prądu do punktu obliczania pola. Powyższa całka redukuje się na płaszczyźnie prostopadłej do kierunku prądu do

$$\vec{B}_s = \frac{\mu_0}{4\pi} \int_V \frac{J \vec{n}}{r} dA, \quad (4.8)$$

gdzie $\vec{n}$ jest wektorem jednostkowym prostopadłym do $\vec{r}$. Korzystając z (4.8) można analitycznie obliczyć współczynniki rozwinięcia (4.6) (c.f. [28]).

Składowa pola generowana przez jarzmo magnesu

Wstawiając równanie (4.7) do równania Maxwella

$$\nabla \times \vec{H} = \vec{J}$$

otrzymujemy

$$\nabla \times \frac{1}{\mu} \vec{B}_r = \vec{J} - \nabla \times \frac{1}{\mu} \vec{B}_s. \quad (4.9)$$

Równanie to rozwiązujemy w dwóch regionach materiałowych, tzn. w Ω_a , który jest częścią bez magnetyzacji ($\mu = \mu_0$) czyli w aperturze oraz w regionie Ω_i , która

stanowi jarzmo gdzie z kolei prąd jest równy zero. Otrzymujemy więc w każdej części materiałowej oddzielne równanie

$$\begin{cases} \nabla \times \frac{1}{\mu} \vec{B}_r = \vec{0} & \text{w } \Omega_a, \\ \nabla \times \frac{1}{\mu} \vec{B}_r = -\nabla \times \frac{1}{\mu} \vec{B}_s & \text{w } \Omega_i. \end{cases}$$

Dodatkowymi warunkami są warunki brzegowe Dirichleta

$$\vec{n} \cdot \vec{B}_r = -\vec{n} \cdot \vec{B}_s$$

lub Neumanna

$$\vec{n} \times \frac{1}{\mu} \vec{B}_r = -\vec{n} \times \vec{B}_s,$$

spełniane na elementach dyskretyzujących brzeg między regionami Ω_a i Ω_i , gdzie wektor $\vec{n}$ jest wektorem normalnym do brzegu. Ponadto konieczne są jeszcze tzw. *zewnętrzne warunki brzegowe* i.e. warunki w dostatecznej odległości od centrum apertury, gdzie pole jest zanedbywalnie małe. Równanie (4.9) wraz z warunkami brzegowymi na styku dwóch regionów materiałowych i z zewnętrznymi warunkami brzegowymi określa pełne zagadnienie brzegowe.

Używając dodatkowej funkcji wagowej $\vec{w}$, która spełnia homogeniczne warunki brzegowe Dirichleta, słabe sformułowanie problemu obliczenia pola magnetycznego w aperturze magnesu zapisuje się jako

$$\int_{\Omega} (\nabla \times \vec{w}) \cdot \frac{1}{\mu} (\nabla \times \vec{A}_r) d\Omega = \int_{\Omega} \left(1 - \frac{\mu_0}{\mu}\right) (\nabla \times \vec{w}) \cdot \vec{H}_s d\Omega.$$

Zagadnienie to rozwiązywane jest metodą elementów skończonych. Szczegóły powyższego sformułowania oraz metody rozwiązania znaleźć można w [28].

Optymalizacja wielocelowa – optymalność w sensie Pareto

W rzeczywistych problemach optymalizacyjnych istnieje zazwyczaj kilka celów, które są ze sobą w konflikcie, tzn. minima dla każdej funkcji celu różnią się od siebie. Zagadnienie optymalizacji K funkcji celu zapisuje się jako

$$„min” \vec{F}(\vec{X}) = „min”(f_1(\vec{X}), f_2(\vec{X}), \dots, f_K(\vec{X})),$$

gdzie $\vec{F} : \mathbb{R}^d \rightarrow \mathbb{R}^K$ a ograniczenia definiowane są przez funkcję $g_i, h_j : \mathbb{R}^d \rightarrow \mathbb{R}$ oraz stałe $c_i, d_j \in \mathbb{R}$

$$g_i(\vec{X}) \leq c_i \quad i = 1, 2, \dots, m \quad (4.10)$$

$$h_j(\vec{X}) = d_j \quad j = 1, 2, \dots, p \quad (4.11)$$

gdzie $\vec{X} = (x_1, \dots, x_d)$ jest takie, że

$$x_{l, \text{lower}} \leq x_l \leq x_{l, \text{upper}}, \quad l = 1, 2, \dots, d. \quad (4.12)$$

Z optymalizacją wielocelową związane jest pojęcie rozwiązania optymalnego w sensie Pareto [28].

Definicja 15. Zbiór $\vec{X}^*$ jest zbiorem optymalnym w sensie Pareto, gdy nie istnieje rozwiązanie $\vec{X}$ w dziedzinie $\{\vec{X} \in \mathbb{R}^d : g_i(\vec{X}) \leq 0; h_j(\vec{X}) = 0; x_{l,\text{lower}} \leq x_l \leq x_{l,\text{upper}}, \forall i = 1, \dots, m; j = 1, \dots, p; l = 1, 2, \dots, d\}$, dla którego

$$\begin{aligned} f_k(\vec{X}) &\leq f_k(\vec{X}^*) \quad \forall k \in \{1, \dots, K\} \\ f_k(\vec{X}) &< f_k(\vec{X}^*) \quad \text{dla przynajmniej jednego } k \in \{1, \dots, K\}. \end{aligned}$$

Innymi słowy rozwiązanie $\vec{X}$, dla którego poprawa jednego z celów powoduje degradację przynajmniej jednego z celów jest elementem zbioru rozwiązań optymalnych w sensie Pareto.

Zadanie poszukiwania rozwiązań optymalnych w sensie Pareto jest transformowane do zagadnienia optymalizacji bez ograniczeń funkcji skalarnej za pomocą :

1. ważenia celów oraz
2. funkcji odległości.

Ograniczenia (4.11), (4.11) oraz (4.12) są włączane do optymalizacji przez modyfikację funkcji celu za pomocą

1. funkcji kary dla ograniczeń zakresów poszukiwania (4.12) oraz
2. funkcji kary dla ograniczeń nieliniowych (4.11) i (4.11).

Metoda ważenia celów polega na połączeniu składowych funkcji wektorowej $F(\vec{X})$ w jedną skalarną funkcję celu

$$f(\vec{X}) = \sum_{k=1}^K w_k f_k(\vec{X}),$$

gdzie w_k są wagami określającymi ważność celu f_k w zagadnieniu optymalizacyjnym.

Metoda funkcji odległości polega połączeniu składowych funkcji wektorowej $F(\vec{X})$ w jedną skalarną funkcję celu dla ustalonych rezyduów $y_k \in \mathbb{R}; k = 1, \dots, K$

$$f(\vec{X}) = \left(\sum_{k=1}^K w_k |f_k(\vec{X}) - y_k|^r \right)^{1/r}.$$

Zazwyczaj ze względu na ciągłość pochodnych rozważa się $r = 2$. W praktyce wagi w_k są znane z rozważań analitycznych natomiast rezydual y_p są znane ze sformułowania problemu lub z założeń o szukanym rozwiązaniu.

Metoda funkcji kary dla ograniczeń zakresów poszukiwań polega na modyfikacji funkcji celu, gdy występuje przekroczenie zakresu

$$p(\vec{X}) = \begin{cases} f(\vec{X}) & \text{żadne ograniczenie nie jest przekroczone} \\ f(\vec{X}^*) + r(\vec{X}) & \text{jedno z ograniczeń jest przekroczone} \end{cases}$$

gdzie $\vec{X}^* = (x_1, x_2, \dots, x_l^*, \dots, x_n)$ oraz $x_l^* = x_{lu}$ jeśli $x_l > x_{lu}$ (tzn. górne ograniczenie jest przekroczone) oraz $x_l^* = x_{lu}$ jeśli $x_l < x_{lu}$ (tzn. dolne ograniczenie jest przekroczone). Składnik kary natomiast zapisuje się w postaci

$$r(\vec{X}) = \sum_l r_l \begin{cases} (x_l - x_{lu})^2 & \text{jeśli } x_l > x_{lu}, \\ (x_{lu} - x_l)^2 & \text{jeśli } x_l < x_{lu}, \\ 0 & \text{w przeciwnym razie,} \end{cases}$$

gdzie $r_l \in \mathbb{R}$ są odpowiednimi współczynnikami kary.

Metoda funkcji kary dla ograniczeń nieliniowych funkcji celu danych za pomocą (4.11), (4.11) polega modyfikacji funkcji celu przez dodanie składnika zależnego od tego jak bardzo ograniczenie funkcji celu jest przekroczone. Część funkcji celu odpowiadająca minimalizacji celu f_k modyfikowana jest przez dodanie składnika kary

$$p(\vec{X}) = f_i(\vec{X}) + \sum_i p_i \max^2(0, g_k(\vec{X}) - c_k) + \sum_j (h_j(\vec{X}) - d_j)^2$$

gdzie $p_i, q_j \in \mathbb{R}$ są odpowiednimi współczynnikami wagowymi.

Funkcja celu w optymalizacji magnesów zakrzywiających zderzacza cząstek elementarnych LHC

W następnym paragrafie omówione są wyniki optymalizacji nadprzewodzących zakrzywiających magnesów dipolowych zderzacza cząstek elementarnych LHC. Aby zdefiniować funkcję celu przekształćmy rozwinięcie (4.6) na okręgu o promieniu referencyjnym r_{ref} do postaci

$$B_r(r_{\text{ref}}, \varphi) = B_1(r_{\text{ref}}) \sum_{n=1}^{\infty} (b_n(r_{\text{ref}}) \sin n\varphi + a_n(r_{\text{ref}}) \cos n\varphi). \quad (4.13)$$

Ze względu na symetrię budowy magnesów zderzacza cząstek elementarnych LHC znikają z powyższego rozwinięcia składniki $a_1(r_{\text{ref}}), a_2(r_{\text{ref}}), \dots$. Celami optymalizacji magnesu dipolowego są

1. pole magnetyczne o zadanym natężeniu strumienia $B_1 = 9.3T$
2. czystość pola magnetycznego, tzn.

$$\begin{cases} \min b_2(r_{\text{ref}}), \\ \min b_3(r_{\text{ref}}), \\ \min b_4(r_{\text{ref}}), \\ \vdots \end{cases}$$

na nominalnym poziomie natężenia prądu pobudzającego,

3. czystość pola magnetycznego

$$\begin{cases} \min b_{2,\text{inj}}(r_{\text{ref}}), \\ \min b_{3,\text{inj}}(r_{\text{ref}}), \\ \min b_{4,\text{inj}}(r_{\text{ref}}), \\ \vdots \end{cases}$$

na poziomie prądu pobudzającego podczas wstrzyknięcia wiązki protonów do zderzacza cząstek elementarnych,

4. zmiana pola pomiędzy poziomem natężenia podczas wstrzyknięcia wiązki protonów do zderzacza cząstek elementarnych a poziomem nominalnym natężenia, tzn.

$$\begin{cases} \min \Delta b_2(r_{\text{ref}}) = \min(b_{2,\text{inj}}(r_{\text{ref}}) - b_2(r_{\text{ref}})), \\ \min \Delta b_3(r_{\text{ref}}) = \min(b_{3,\text{inj}}(r_{\text{ref}}) - b_3(r_{\text{ref}})), \\ \min \Delta b_4(r_{\text{ref}}) = \min(b_{4,\text{inj}}(r_{\text{ref}}) - b_4(r_{\text{ref}})), \\ \vdots \end{cases}$$

Optymalizowana funkcja celu, skonstruowana za pomocą metody ważenia celów, funkcji odległości, metod włączania ograniczeń opisanych w poprzednim paragrafie, powinna więc zawierać powyższe cele dla początkowych n . Zmiennymi, od których zależy optymalizowana funkcja celu są parametry techniczne modelu magnesu.

4.2.3 Wyniki numeryczne

W paragrafie tym prezentujemy wyniki działania metody SPELROA połączonej z algorytmem optymalizacji bezpośredniej EXTREM [18]. Algorytm ten ze względu na swoją stabilność i przyjazność dla użytkownika – nie wymaga od użytkownika podania żadnych parametrów – został użyty do optymalizacji nadprzewodzących magnesów dipolowych dla zderzacza LHC w Europejskim Laboratorium Fizyki Cząstek CERN [28]. W testach zastosowanie metody SPELROA wymagało podania 4 parametrów. Do konstrukcji aproksymatora (3.3) jako funkcji bazowej użyliśmy funkcji Gaussa $\phi(r) = e^{-\alpha\|r\|^2}$. Zebraliśmy je w tabeli 4.28. Dobór wartości tych parametrów oczywiście zależy od optymalizowanej funkcji celu oraz od liczby zmiennych w funkcji celu. Z naszych testów jednak można wywnioskować uniwersalne wartości tych parametrów. We wszystkich prezentowanych przykładach liczba użytych funkcji radialnych wynosiła $N = 30$ a liczba wykonanych kroków w fazie wstępnej wynosiła 40. Wartości progowe natomiast wynosiły $\gamma_{\text{thr}} = 0.65$.

LHC coil test facility, przykład optymalizacji funkcji dwóch zmiennych

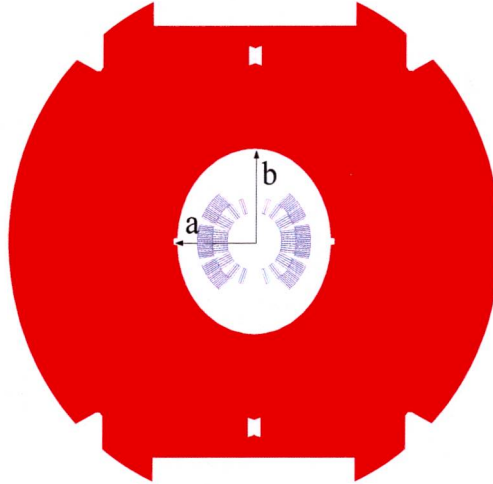
Głównym wysiłkiem podczas budowy zderzacza cząstek elementarnych LHC było zapewnienie spełnienia specyfikacji pola magnetycznego w głównych magnesach dipolowych.

Tabela 4.28: Parametry metody SPELROA definiowane przez użytkownika

I_s	–	liczba kroków w początkowej fazie optymalizacji, w których funkcja celu obliczana jest tylko w sposób bezpośredni
N	–	rozmiar zbioru treningowego a tym samym liczba funkcji bazowych użytych do budowy aproksymatora RBF (3.3). Zauważmy, że musi być ($N < I_s$).
γ_{thr}	–	wartość progowa dla wskaźnika (3.26) wyznaczenia regionu wiarygodności
$NLMSE_{thr}$	–	wartość progowa dla wskaźnika (3.18) do wyznaczenia parametru regularyzacji λ przez minimalizację (3.19)

Aby zapewnić możliwość efektywnego testowania rozwiązań technicznych wykorzystanych w głównym magnesie dipolowym LHC, skonstruowano krótki magnes z jedną aperturą, tzw. Coil Test Facility (CTF). Jest to jeden z trzech testowych modeli prototypowych magnesów dipolowych oprócz krótkiego i długiego modelu prototypowego końcowej specyfikacji.

CTF może służyć jako prosty przykład procesu optymalizacyjnego z dwoma parametrami: długościami osi (a, b) określającymi wewnętrzny kształt jarzma magnesu (patrz rys. 4.1).



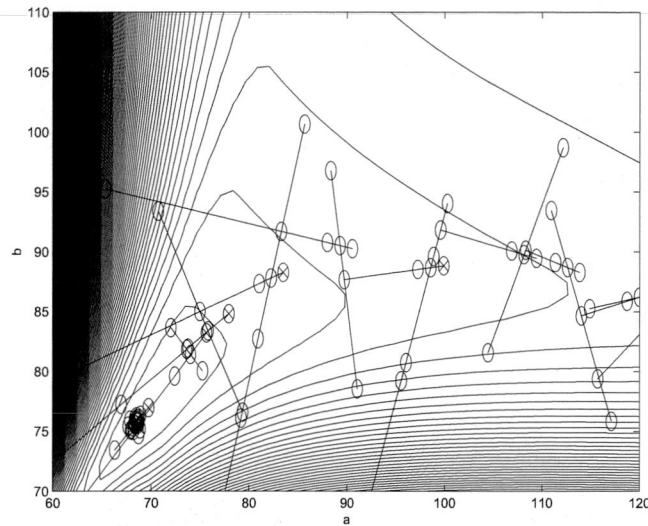
Rysunek 4.1: Model Coil Test Facility. Optymalizowane są rozmiary półosi elipsy określającej kształt kołnierza magnesu.

Funkcja celu jaką optymalizowaliśmy została skonstruowana za pomocą metody włączania ograniczeń nieliniowych i jest postaci

$$f_{obj}(a, b) = 50B_1 + 5 \max(0, \Delta b_3 - 1)^2 + \max(0, b_{3,inj} - 10)^2, \quad (4.14)$$

gdzie $b_{3,\text{inj}}$ jest wartością współczynnika sekstopolowego $b_3(r_{\text{ref}})$ w rozwinięciu (4.13) dla natężenia prądu pobudzającego magnes na poziomie natężenia prądu podczas wstrzyknięcia wiązki do zderzacza cząstek elementarnych, natomiast Δb_3 jest zmianą współczynnika $b_3(r_{\text{ref}})$ podczas ładowania magnesu od natężenia prądu pobudzenia podczas wstrzyknięcia wiązki do zderzacza do poziomu nominalnego.

Wykres funkcji celu $f_{\text{obj}}(a, b)$ wraz z wizualizacją punktów wygenerowanych w kolejnych krokach działania metody SPELROA z algorytmem EXTREM przedstawione są na rysunku 4.2. W tabeli 4.29 przedstawione są znalezione minima za pomocą algorytmu EXTREM oraz metody SPELROA połączonej z algorytmem EXTREM, liczba bezpośrednich obliczeń wartości funkcji celu oraz liczba aproksymacji funkcji celu. Uzyskane przyspieszenie obliczone jako stosunek liczby bezpośrednich wartościowań funkcji celu przez metodę SPELROA do liczby wartościowań funkcji celu przez algorytm EXTREM wyniosła 30%.



Rysunek 4.2: Ścieżka optymalizacyjna dla metody SPELROA połączonej z algorytmem EXTREM dla modelu CTF (c.f. [2]). Proces wystartowany został z punktu [130.0, 80.0]. Algorytm „zbiegł” do minimum funkcji znajdującego się w punkcie [68.55, 75.87] po 111 bezpośrednich wartościowaniach funkcji i konstrukcji aproksymacji RBF w 31 punktach (kółka z krzyżykiem). Algorytm EXTREM potrzebował 167 wartościowań funkcji. Uzyskano więc przyspieszenie rzędu 30%.

LHC Main Dipole, przykład optymalizacji funkcji trzech i pięciu zmiennych

Parametrami jakimi manipulujemy podczas optymalizacji są parametry geometryczne jarzma magnesu – patrz rysunek 4.3.

Tabela 4.29: Porównanie działania algorytmu EXTREM oraz metody SPELROA połączonej z algorytmem EXTREM do optymalizacji krótkiego modelu testowego coil-test facility dla zderzacza cząstek elementarnych LHC. Optymalizacja wystartowana została z punktu $[130.0, 80.0]$. Uzyskane przyspieszenie to około 30%.

	EXTREM	SPELROA
a	68.53	68.57
b	75.87	75.86
f_{obj}	-391.3003	-391.2974
Liczba obliczeń wart. f_{obj}	167	111
Liczba aproksymacji RBF f_{obj}	—	31

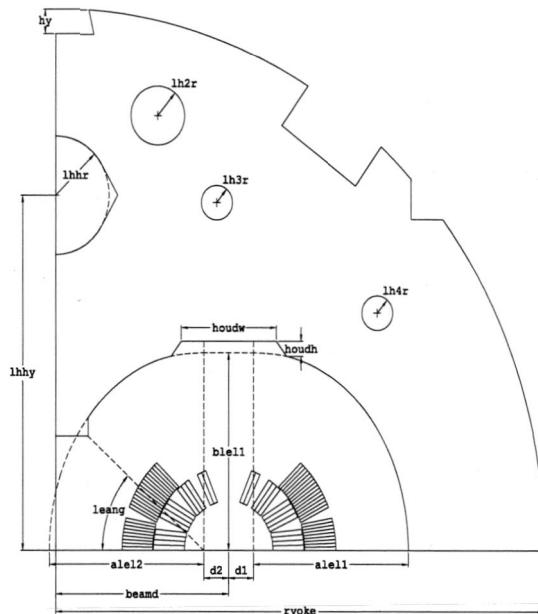
Funkcja celu jaką optymalizowaliśmy skonstruowana została za pomocą metody ważenia celów i jest postaci

$$f_{\text{obj}}(\vec{X}) = \sum_{n=2}^4 t_n \Delta b_n + \sum_{n=2}^4 b_{n,\text{inj}} \quad \text{dla} \quad (t_2 = t_3 = 5, t_4 = 50), \quad (4.15)$$

gdzie b_2, b_3 i b_4 są współczynnikami kwadrupolowymi, sekstopolowymi oraz oktopolowymi odpowiednio. Parametrami względem, których prowadzona jest optymalizacja trzech zmiennych, są $\vec{X} = (x_1, x_2, x_3)$, gdzie $x_1 = \text{alel1}$, $x_2 = \text{blel1}$ oraz $x_3 = \text{alel2}$ są parametrami kształtu kołnierza niemagnetycznego wokół cewki – patrz Rys. 4.3.

Do optymalizacji użyliśmy algorytmu EXTREM [18] połączonego z metodą SPELROA. Proces został wystartowany z punktu $[88.0, 105.0, 79.0]$. W tabeli 4.30 przedstawiamy porównanie znalezionych minimów funkcji celu oraz liczby wartościowań funkcji dla obydwu metod. Uzyskane za pomocą metody SPELROA przyspieszenie to około 30%.

W pracy [2] przedstawiliśmy również wyniki optymalizacji głównego magnesu dipolowego zderzacza cząstek elementarnych LHC dla funkcji (4.15) zależnej od pięciu parametrów projektu. Dodatkowo pod uwagę wzięto zmienne $x_4 = \text{ryoke}$ oraz $x_5 = \text{hy}$ – patrz Rys. 4.3. Metoda zastosowana tam różniła od prezentowanej w tej pracy jedynie sposobem określenia regionu wiarygodnego oraz sposobu wyboru parametru regularyzacji w konstrukcji aproksymatora funkcji celu. Różnica polegała na użyciu odległości Mahalanobisa zamiast wskaźnika $\gamma_{\mathbf{x}}$ oraz arbitralnego przedziału poszukiwania wartości parametru regularyzacji λ . Wynik przedstawiamy w tabeli 4.31



Rysunek 4.3: Parametryczna reprezentacja optymalizowanego jarzma magnesu LHC.

4.3 Strojenie regulatora PID – metoda SPELROA połączona z algorytmem quasi-Newtona

W niniejszej pracy formalnie dowiedliśmy, że metoda SPELROA jest zbieżna w połączeniu z algorytmami bezgradientowymi. Nic nie stoi jednak na przeszkodzie, żeby takie podejście zastosować w algorytmach gradientowych. Oczywiście dowód zbieżności przedstawiony w pracy nie przenosi się automatycznie na algorytmy gradientowe. Dowód zbieżności metody SPELROA dla algorytmów gradientowych wymagać będzie rozszerzenia, w którym rozważone będzie również zaburzenie wprowadzane do kierunku poszukiwań a nie jedynie do wartości funkcji w fazie minimalizacji wzdłuż kierunku.

W paragrafie tym prezentujemy zastosowanie metody SPELROA połączonej z algorytmem quasi-Newtona do procesu strojenia regulatora PID. Macierz aproksymująca hesjan była generowana za pomocą poprawki BFGS (patrz [15]).

Rozpatrzyliśmy regulator PID dla układu oscylującego danego impedancją

$$K(s) = \frac{1}{50s^3 + 43s^2 + 3s + 1}. \quad (4.16)$$

Do znalezienia parametrów K_p , K_d i K_i regulatora użyliśmy estymacji standardowych funkcji celu *Integral of time error squared* (ITES) oraz *Integral of time absolute error* (ITAE) danych wzorami

$$ITES = \int_0^\infty [te(t)]^2 dt \quad ITAE = \int_0^\infty t|e(t)|dt.$$

Tabela 4.30: Porównanie działania algorytmu EXTREM oraz metody SPELROA połączonej z algorytmem EXTREM do optymalizacji głównego magnesu dipolowego zderzacza cząstek elementarnych LHC. Optymalizacja wystartowana została z punktu $[88.0, 105.0, 79.0]$. Uzyskane przyspieszenie to około 30%.

	EXTREM	SPELROA
x_1	80.92	80.70
x_2	99.64	99.40
x_3	86.02	85.74
f_{obj}	-23.941	-24.179
Liczba obliczeń wart. f_{obj}	301	208
Liczba aproksymacji RBF f_{obj}	—	45

Minimum funkcji ITES daje rozwiązanie o przesterowaniu rzędu 4.5%. Natomiast dla funkcji ITAE przesterowanie znalezionej regulatora jest rzędu 3%. Wyniki optymalizacji przedstawiają tabele

Tabela 4.31: Porównanie działania algorytmu EXTREM oraz metody SPELROA połączonej z algorytmem EXTREM do optymalizacji głównego magnesu dipolowego zderzacza cząstek elementarnych LHC. Optymalizacja wystartowana została z punktu $[95.0, 96.0, 85.0, 275.0, 8.0]$. Uzyskane przyspieszenie to około 18.5%.

	EXTREM	SPELROA
x_1	87.10	87.44
x_2	105.39	104.69
x_3	92.90	93.55
x_4	280.97	280.99
x_5	3.91	4.10
f_{obj}	15.51	15.66
Liczba obliczeń wart. f_{obj}	343	281
Liczba aproksymacji RBF f_{obj}	—	52

Tabela 4.32: Metoda SPELROA w połączeniu z algorytmem BFGS quasi-Newton dla funkcji $f_{\text{obj}} = \log(ITES)$. Proces wystartowany z punktu $[0.2, 0.5, 5.5]$.

Algorytm	BFGS q-N	SPELROA
K_p	3.844896	3.826120
K_i	0.172128	0.172104
K_d	18.430017	18.351408
f_{obj}	1.910984	1.910988
Liczba obliczeń wart. f_{obj}	68	63
Liczba aproksymacji RBF f_{obj}	—	12

Tabela 4.33: Metoda SPELROA w połączeniu z algorytmem BFGS quasi-Newton dla funkcji $f_{\text{obj}} = \log(ITAE)$. Proces wystartowany z punktu $[0.2, 0.5, 6.5]$.

Algorytm	BFGS q-N	SPELROA
K_p	3.344571	3.291519
K_i	0.162876	0.162836
K_d	16.990756	16.787918
f_{obj}	1.489663	1.490390
Liczba obliczeń wart. f_{obj}	78	64
Liczba aproksymacji RBF f_{obj}	—	7

Rozdział 5

Zakończenie

5.1 Podsumowanie

Tematem niniejszej pracy jest optymalizacja funkcji celu, których obliczenie jest czasochłonne. Takie zadania optymalizacyjne pojawiają się w zagadnieniach optymalizacji i konstrukcji systemów sterowania jak i zagadnieniach projektowania różnego rodzaju urządzeń w licznych dziedzinach techniki. Obliczanie wartości funkcji w takim zadaniu wiąże się z zazwyczaj z przeprowadzeniem złożonej i czasochłonnej operacji w zależności od zmiennych projektowych. Do optymalizacji tego typu funkcji zazwyczaj stosuje się algorytmy bezgradientowe ze względu na to, że gradient nie jest dostępny a aproksymacja gradientu jest zbyt czasochłonna lub niemożliwa.

W niniejszej rozprawie autor zaprezentował nowe podejście do przyspieszania procesów optymalizacji funkcji ciągłych. Metoda, zaproponowana przez autora, polega na konstrukcji modelu aproksymującego w części punktów ścieżki optymalizacyjnej algorytmu bezgradientowego i na nieobliczaniu wartości funkcji w sposób bezpośredni lecz zastąpienia jej skonstruowanym modelem. W zaproponowanej metodzie modelem aproksymującym jest sieć z radialnymi funkcjami aktywacji trenowana za pomocą metody regularyzacji Tikhonova, stąd nazwana została Search Procedure Exploiting Locally Regularized Objective Approximation. Założeniem metody jest, że czas konstrukcji aproksymacji oraz obliczenia jej wartości w danym punkcie jest zanedbywalnie mały w porównaniu z czasem potrzebnym na obliczenie funkcji celu w sposób bezpośredni. Przy takim założeniu przyspieszenie algorytmu optymalizacji rozumiemy jako skrócenie czasu jego działania, tzn. zmniejszenie liczby obliczeń bezpośrednich wartości funkcji. W pracy autor udowodnił zbieżność metody dla algorytmów bezgradientowy. Zastosowanie zaprezentowanej metody w połączeniu z algorytmem Gaussa-Seidela z reortogonalizacją bazy kierunków poszukiwań, stosowanym do optymalizacji magnesów umożliwiło przyspieszenie procesu optymalizacji głównego magnesu zakrzywiającego zderzacza cząstek elementarnych LHC nawet do 30% dla funkcji celu, której obliczenie może trwać do 2h na komputerze DEC Alpha XP 1000. Dla trudnych funkcji testowych z literatury uzyskano od kilku do nawet kilkudziesięciu

procent przyspieszenia. Na przykładzie procesu strojenia regulatora PID autor pokazał, że istnieją również obiecujące perspektywy zastosowania nowej metody w połączeniu z algorytmami gradientowymi.

5.2 Proponowane dalsze kierunki badań

Możliwe dalsze kierunki badań to przede wszystkim :

- Poszukiwanie analitycznego związku pomiędzy wartością wskaźnika $\gamma_{\mathbf{x}}(\mathbf{x})$ określonego przez (3.26) mierzącego „jakość” otoczenia punktu $\mathbf{x}$ przez punkty ze zbioru danych $\mathbf{X}$ a jakością zregulowanej aproksymacji radialnymi funkcjami bazowymi. Gdyby taki związek udało się znaleźć możliwe byłoby określenie metody ustalenia wartości progowych γ_{thr} określających obszary otoczki wypukłej zbioru danych, w których błąd aproksymacji mógłby być precyzyjniej niż obecnie ograniczony.
- Zastosowanie aproksymacji radialnymi funkcjami bazowymi konstruowanej metodą WGV do aproksymacji gradientu funkcji celu i wykorzystanie tej aproksymacji w algorytmach gradientowych, dobre perspektywy rozwoju badań w tym kierunku potwierdza zaprezentowany przykład zastosowanie metody w połączeniu z efektywnym algorytmem optymalizacji gradientowej quasi-Newtona z poprawką aproksymacji hesjanu metodą BFGS zostało przedstawione na przykładzie optymalizacji układu regulacji PID.
- Porównanie metody SPELROA z metodami regionu wiarygodności.

Rozdział 6

Dodatki

6.1 Dodatek A

Minimum paraboli $q(\zeta)$ interpolujące punkty $(\zeta^1, f(\zeta^1))$, $(\zeta^2, f(\zeta^2))$ oraz $(\zeta^3, f(\zeta^3))$ gdzie $\{\zeta_1, \zeta_2, \zeta_3\} \subset \mathbb{R}$ dane jest wzorem

$$\lambda^* = \frac{1}{2} \frac{[(\zeta^3)^2 - (\zeta^2)^2]f(\zeta^1) + [(\zeta^3)^2 - (\zeta^1)^2]f(\zeta^2) + [(\zeta^2)^2 - (\zeta^1)^2]f(\zeta^3)}{(\zeta^3 - \zeta^2)f(\zeta^1) + (\zeta^3 - \zeta^1)f(\zeta^2) + (\zeta^2 - \zeta^1)f(\zeta^3)}.$$

Przekształćmy parabolę $q(x) = ax^2 + bx + c$; $x = \zeta^1 + t(\zeta^3 - \zeta^1)$; $t = (-\infty, \infty)$ do paraboli $\hat{q}(x) = \hat{a}x^2 + \hat{b}x + \hat{c}$ zakładając co następuje

1. Transformacja będzie postaci

$$\hat{q}(x) = E(p(x)) + D, \quad p(x) = L_2(x), \quad (6.1)$$

gdzie $L_2(x)$ jest wielomianem interpolującym punkty $(0, f(\zeta^1))$, $(\zeta_{(r)}, f(\zeta^2))$ oraz $(1, f(\zeta^3))$ gdzie $\zeta_{(r)} = \frac{\zeta^2 - \zeta^1}{\zeta^3 - \zeta^1}$.

2. Parabola przekształcona przyjmuje wartości

$$\begin{cases} \hat{q}(0) &= -1, \\ \hat{q}(1) &= -2, \end{cases} \quad \text{gdym } f(\zeta^1) > f(\zeta^3)$$

oraz

$$\begin{cases} \hat{q}(0) &= -2, \\ \hat{q}(1) &= -1, \end{cases} \quad \text{gdym } f(\zeta^1) < f(\zeta^3).$$

Z warunku 1. otrzymujemy

$$L_2(x) = a'x^2 + b'x + c'$$

gdzie

$$\begin{cases} a' &= \frac{f(\zeta^1)}{\zeta_{(r)}} + \frac{f(\zeta^2)}{\zeta_{(r)}(\zeta_{(r)}-1)} + \frac{f(\zeta^3)}{1-\zeta_{(r)}}, \\ b' &= -\left[\frac{f(\zeta^1)(\zeta_{(r)}+1)}{\zeta_{(r)}} + \frac{f(\zeta^2)}{\zeta_{(r)}(\zeta_{(r)}-1)} + \frac{f(\zeta^3)\zeta_{(r)}}{1-\zeta_{(r)}} \right], \\ c' &= f(\zeta^1). \end{cases}$$

Z warunków w punkcie 2. natomiast otrzymujemy

$$\begin{cases} E &= -\frac{1}{a'+b'}, \\ D &= -1 - Ec' = \frac{c'}{a'+b'} - 1. \end{cases} \quad \text{gdy } f(\zeta^1) > f(\zeta^3),$$

oraz

$$\begin{cases} E &= -\frac{3}{a'+b'}, \\ D &= -2 - Ec' = \frac{3c'}{a'+b'} - 2. \end{cases} \quad \text{gdy } f(\zeta^1) < f(\zeta^3).$$

Zapisując $\hat{q}(x)$ w formie kanonicznej $\hat{q}(x) = \hat{a}x^2 + \hat{b}x + \hat{c}$ mamy

$$\begin{cases} \hat{a} &= Ea', \\ \hat{b} &= Eb', \\ \hat{c} &= Ec' + D. \end{cases}$$

Interesujące dla nas własności przekształcenia $\hat{q}$ są następujące

1. $\hat{q}(\zeta_{(r)}) < -1$ dla $x \in [0, 1]$.
2. $p(\zeta_{(r)}) = f(\zeta^2)$.
3. $\hat{q}(\zeta_{(r)})$ nie zależy od $f(\zeta^2)$.
4. Dla minimum λ^* paraboli nieprzekształconej q takiej, że $\lambda^* \in [\zeta^1, \zeta^3]$ otrzymujemy minimum w punkcie $\hat{\lambda}^* = \frac{\lambda^* - \zeta^1}{\zeta^3 - \zeta^1}$ gdzie $\hat{\lambda}^*$ jest minimum paraboli $\hat{q}$.
5. Jedynym ograniczeniem jest osobliwość przekształcenia $\hat{q}$, gdy $a' + b' = 0$. Warunek ten nie jest spełniony dla parabol, których minimum znajduje się w punkcie $(\zeta^3 - \zeta^1)/2$.

Przekształcenie to redukuje liczbę stopni swobody z 6 do 3 a więc pozwala „narysować” część naszych wyników.

6.2 Dodatek B

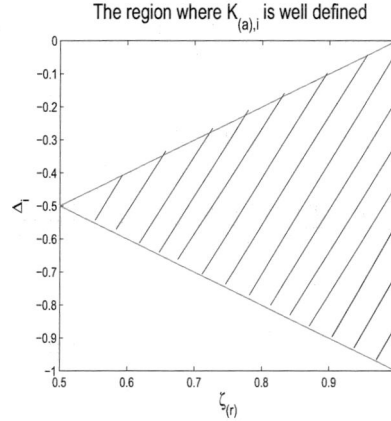
Aby rozwiązać nierówności (2.26), tzn.

$$\tilde{\lambda}_l^*(\varepsilon_l, \zeta_i) < \zeta_{(r)} - K_{(a),i}, \quad K_{(a),i} = \sqrt{\frac{4\delta^2}{(\zeta_i^3 - \zeta_i^1)^2} - (1 - \zeta_{(r)})^2} \quad (6.2)$$

$$l \in \{1, 2, 3\}$$

zauważmy, że $K_{(a),i}$ jest dobrze określone, gdy ζ_i i δ są takie, że

$$\left(\frac{2\delta}{\zeta_i^3 - \zeta_i^1} \right)^2 \geq (1 - \zeta_{(r)})^2 \quad (6.3)$$



Rysunek 6.1: Obszar w przestrzeni $(\Delta', \zeta_{(r)})$, w którym dobrze określona est wielkość $K_{(a),i}$ zdefiniowany w (6.2).

gdzie $\zeta_{(r)} = \frac{\zeta_i^2 - \zeta_i^1}{\zeta_i^3 - \zeta_i^1}$. W wypadku, gdy zachodzi nierówność

$$\left(\frac{2\delta}{\zeta_i^3 - \zeta_i^1} \right)^2 < (1 - \zeta_{(r)})^2$$

nierówności (2.25) są zawsze spełnione. Zdefiniujemy

$$\Delta'_i = \frac{2\delta}{\zeta_i^3 - \zeta_i^1}$$

Zastanówmy się w jakim obszarze przestrzeni $(\Delta', \zeta_{(r)}) \subset \mathbb{R}^2$ wielkość $K_{(a),i}$ jest dobrze określona. W obszarze tym spełnione jest

$$\Delta_i'^2 \geq (1 - \zeta_{(r)})^2$$

Zauważmy, że nierówność ta może być spełniona jedynie, gdy $\zeta_{(r)} > 1/2$. Mamy wtedy $1 - \zeta_{(r)} < 1/2$. Aby $\zeta_{(r)} - \Delta'_i$ było większe od zera musi zachodzić $\Delta'_i \in (\zeta_{(r)} - 1, -\zeta_{(r)})$. Powyższy obszar przedstawiony jest na rysunku 6.1. Zauważmy, że nie musimy rozpatrywać wyrażenia $\zeta_{(r)} + \Delta'_i$, gdyż jest ono większe od 1, gdy zachodzi (6.3).

Rozpatrzmy jedynie przypadek, gdy $f(\zeta_i^1) < f(\zeta_i^3)$. Dla przekształcenia $\hat{q}$ określonego wzorem (6.1) definiujemy wielkości

$$\begin{aligned} A &= -2(\zeta_{(r)} - 1), \\ B &= \hat{q}(\zeta_{(r)}), \\ C &= -\zeta_{(r)}. \end{aligned}$$

Wówczas (2.26) dla $l = 1$ zapisuje się

$$\frac{1}{2} \frac{A(\zeta_{(r)} + 1)(1 + \varepsilon_1) + B - \zeta_{(r)}C}{A(1 + \varepsilon_1) + B - C} < \zeta_{(r)} - K_{(a),i}.$$

Rozpatrzmy dwa przypadki w zależności od znaku mianownika lewej strony.

1. Jeśli $A(1 + \varepsilon_1) + B - C > 0$ to przez pomnożenie przez mianownik i po uproszczeniu mamy

$$A(2K_a + 1 - \zeta_{(r)})\varepsilon_1 < \zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B).$$

Ponieważ $A > 0$, to rozwiązanie zapisuje się

$$\begin{cases} \varepsilon_1 > \frac{C-B}{A} - 1, \\ \varepsilon_1 < \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{A(2K_a+1-\zeta_{(r)})}. \end{cases}$$

2. Jeśli $A(1 + \varepsilon_1) + B - C < 0$, to otrzymujemy

$$A(2K_a + 1 - \zeta_{(r)})\varepsilon_1 > \zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B),$$

a ponieważ $A > 0$ to rozwiązanie zapisuje się

$$\begin{cases} \varepsilon_1 < \frac{C-B}{A} - 1, \\ \varepsilon_1 > \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{A(2K_a+1-\zeta_{(r)})}. \end{cases}$$

Dla $l = 2$ nierówność (2.26) zapisuje się

$$\frac{1}{2} \frac{A(\zeta_{(r)} + 1) + B(1 + \varepsilon_2) - \zeta_{(r)}C}{A + B(1 + \varepsilon_2) - C} < \zeta_{(r)} - K_{(a),i}. \quad (6.4)$$

Zauważmy, że mamy $B < 0$. Musimy rozpatrzeć dwa przypadki w zależności od znaku mianownika

1. Jeśli $A + B(1 + \varepsilon_2) - C > 0$ to mamy

$$B(2K_{(a),i} + 1 - 2\zeta_{(r)})\varepsilon_2 < \zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B).$$

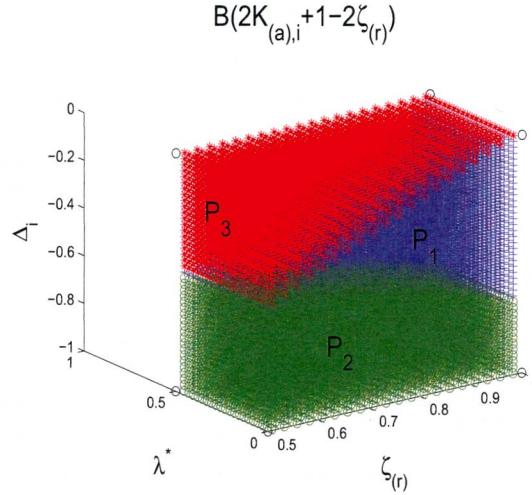
Ponieważ rozpatrujemy tylko $\zeta_{(r)}$, dla których dobrze określone jest $K_{(a),i}$, tzn. $\zeta_{(r)} \in (1/2, 1)$ to musimy również rozpatrzeć dwa przypadki zleżnie od znaku wyrażenia

$$B(2K_{(a),i} + 1 - 2\zeta_{(r)}). \quad (6.5)$$

Ponieważ $B = \hat{q}(\zeta_{(r)})$ zależy również od położenia minimum λ^* konstruowanej paraboli, to rozwiązania nierówności

$$B(2K_{(a),i} + 1 - 2\zeta_{(r)}) > 0$$

poszukujemy w przestrzeni $(\zeta_{(r)}, \lambda^*, \Delta_i) \subset \mathbb{R}^3$. Ponieważ rozpatrujemy tu parabole, dla których $f(\zeta_i^1) < f(\zeta_i^3)$ to musimy zauważyć, że z (6.13) wynika, że $\lambda^* \in (0, 1/2)$. Regiony, w których wyrażenie (6.5) jest dodatnie, ujemne albo nieokreślone przedstawiliśmy na rysunku 6.2. Mamy więc



Rysunek 6.2: W regionie P_1 wyrażenie (6.5) jest większe od zera. W regionie P_2 wyrażenie (6.5) jest mniejsze od zera. W regionie P_3 wyrażenie (6.5) nie jest określone.

(a) Jeśli $B(2K_{(a),i} + 1 - 2\zeta_{(r)}) > 0$, to rozwiązanie zapisuje się jako

$$\begin{cases} \varepsilon_2 < \frac{C-A}{B} - 1, \\ \varepsilon_2 < \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})}. \end{cases}$$

(b) Jeśli $B(2K_{(a),i} + 1 - 2\zeta_{(r)}) < 0$, to rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_2 < \frac{C-A}{B} - 1, \\ \varepsilon_2 > \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})}. \end{cases}$$

2. Jeśli $A + B(1 + \varepsilon_2) - C < 0$, to mamy rozwiązanie zapisuje się natomiast

(a) Jeśli $B(2K_{(a),i} + 1 - 2\zeta_{(r)}) > 0$, to rozwiązanie zapisuje się jako

$$\begin{cases} \varepsilon_2 > \frac{C-A}{B} - 1, \\ \varepsilon_2 > \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})}. \end{cases}$$

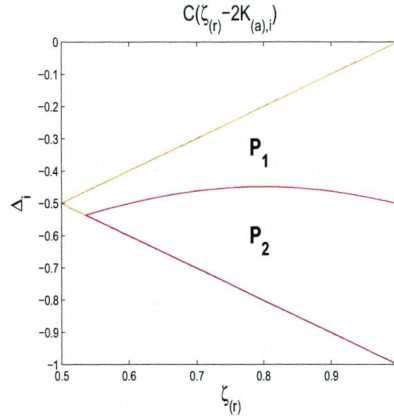
(b) Jeśli $B(2K_{(a),i} + 1 - 2\zeta_{(r)}) < 0$, to rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_2 > \frac{C-A}{B} - 1, \\ \varepsilon_2 < \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})}. \end{cases}$$

Dla $l = 3$ nierówność (2.26) przyjmuje postać

$$\frac{1}{2} \frac{A(\zeta_{(r)} + 1) + B - \zeta_{(r)}C(1 + \varepsilon_3)}{A + B - C(1 + \varepsilon_3)} < \zeta_{(r)} - K_{(a),i}.$$

Zauważmy, że $C < 0$. Rozwiązanie dzielimy na dwa przypadki



Rysunek 6.3: Wyrażenie (6.6) w obszarze $\mathbf{P}_1$ jest ujemne, natomiast w obszarze $\mathbf{P}_2$ jest dodatnie.

1. Jeśli $A + B - C(1 + \varepsilon_3) > 0$, to mamy

$$C(\zeta_{(r)} - 2K_{(a),i})\varepsilon_3 < \zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B).$$

Postać rozwiązania zależy od znaku wyrażenia

$$C(\zeta_{(r)} - 2K_{(a),i}). \quad (6.6)$$

Rysunek 6.3 przedstawia region, w którym wyrażenie (6.6) jest większe oraz region, w którym jest mniejsze od zera w dziedzinie określoności $K_{(a),i}$.

- (a) W regionie, w którym $C(\zeta_{(r)} - 2K_{(a),i}) > 0$ (patrz region $\mathbf{P}_2$ rysunek 6.3) rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_3 > \frac{A+B}{C} - 1, \\ \varepsilon_3 > \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{C(\zeta_{(r)} - 2K_{(a),i})}. \end{cases}$$

- (b) W regionie, w którym $C(\zeta_{(r)} - 2K_{(a),i}) < 0$ (patrz region $\mathbf{P}_1$ rysunek 6.3) rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_3 > \frac{A+B}{C} - 1, \\ \varepsilon_3 < \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{C(\zeta_{(r)} - 2K_{(a),i})}. \end{cases}$$

2. Jeśli $A + B - C(1 + \varepsilon_3) < 0$, to mamy

$$C(\zeta_{(r)} - 2K_{(a),i})\varepsilon_3 > \zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B).$$

- (a) W regionie, w którym $C(\zeta_{(r)} - 2K_{(a),i}) > 0$ (patrz region $\mathbf{P}_2$ rysunek 6.3) rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_3 < \frac{A+B}{C} - 1, \\ \varepsilon_3 < \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{C(\zeta_{(r)} - 2K_{(a),i})}. \end{cases}$$

- (b) W regionie, w którym $C(\zeta_{(r)} - 2K_{(a),i}) < 0$ (patrz region $\mathbf{P}_1$ rysunek 6.3) rozwiązanie przyjmuje postać

$$\begin{cases} \varepsilon_3 < \frac{A+B}{C} - 1, \\ \varepsilon_3 > \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{C(\zeta_{(r)} - 2K_{(a),i})}. \end{cases}$$

W jaki sposób pokazać, że dla $l \in \{1, 2, 3\}$ rozwiązanie powyższe nie jest sprzeczne zademonstrujemy na najbardziej skomplikowanym przypadku, tzn. dla $l = 2$. Pozostawiamy czytelnikowi przeprowadzenie analogicznego rozumowania dla $l = 1$ i $l = 3$.

Pokażemy, że dla każdego Δ_i można podać takie ε_0 , dla którego można wybrać ε_2 takie, że $|\varepsilon_2| > \varepsilon_0$ i dla którego spełnione będzie (6.4). Możliwość takiego wyboru oznacza, że ε_2 jest większe od 0 i dla dowolnego zaburzenia $\varepsilon > |\varepsilon_2|$ w $i + 1$ -szym iteracji algorytmu iteracji parabolicznej możemy wygenerować, trójkę $\zeta_{i+1} = (\zeta_{i+1}^1, \zeta_{i+1}^2, \zeta_{i+1}^1)$, która oddalona jest o 2δ od $\zeta_i = (\zeta_i^1, \zeta_i^2, \zeta_i^1)$ co dowodzi tezy twierdzenia. Musimy więc rozpatrzyć dwa przypadki

1. Gdy $B(2K_a + 1 - 2\zeta_{(r)}) > 0$, to rozwiązanie opisywane jest układami nierówności (6.6) i (6.6). Dla (6.6) mamy

$$\varepsilon < \min \left\{ \frac{C-A}{B} - 1, \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})} \right\}.$$

Ponieważ $\frac{C-A}{B} - 1 < 0$ to tak wybrane $\varepsilon < 0$. To rozwiązanie odrzucamy, gdyż musi być $\varepsilon > 0$. Dla (6.6) wybieramy natomiast

$$\varepsilon > \max \left\{ \left| \frac{C-A}{B} - 1 \right|, \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})} \right\}.$$

Taki wybór daje nam $\varepsilon > 0$.

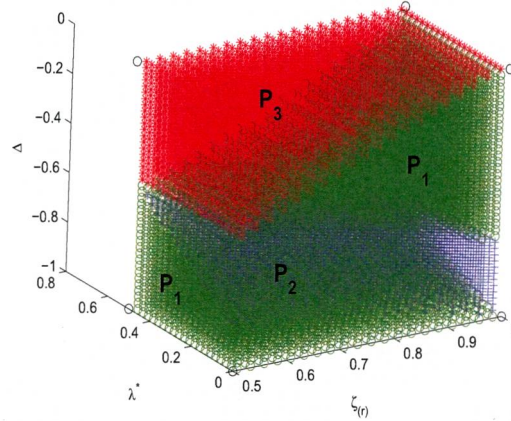
Ostatecznie więc ε_0 jest wartością bezwzględną z ε wybranych w powyższych dwóch sytuacjach. Praktycznie oznacza to tyle, że albo dla $\varepsilon_2 < \varepsilon_0 < 0$ lub $\varepsilon_2 > \varepsilon_0 > 0$, które można wyliczyć w każdym kroku, spełniona jest nierówność (2.23).

2. Gdy $B(2K_a + 1 - 2\zeta_{(r)}) < 0$, to rozwiązanie opisywane jest układami nierówności (6.6) i (6.6). Wprowadźmy oznaczenie

$$D = \frac{\zeta_{(r)}(A+2B-C) - 2K_{(a),i}(A+B-C) - (A+B)}{B(2K_a + 1 - 2\zeta_{(r)})} \quad (6.7)$$

Jeśli chodzi o (6.6) to układ ten jest sprzeczny, gdyż $\frac{C-A}{B} - 1 < 0$ natomiast $D > 0$, dlatego możemy go dorzucić z rozważań. Rozważmy więc (6.6), dla którego znak wyrażenia D przedstawiliśmy na rysunku 6.4.

Na rysunku 6.4 widać, że wyrażenie D jest mniejsze od 0 dla większe części regionu określoności. Tylko w tym regionie możemy określić przedział ε nie



Rysunek 6.4: W regionie $\mathbf{P}_1$ wyrażenie (6.7) jest mniejsze od zera. W regionie $\mathbf{P}_2$ wyrażenie (6.7) jest większe od zera. W regionie $\mathbf{P}_3$ wyrażenie (6.7) nie jest określone.

zawierający 0. Rysunek 6.4 pokazuje również, że Δ_i nie może być dowolne. Łatwo również pokazać, że w regionie gdzie $D < 0$ zachodzi

$$\frac{C - A}{B} - 1 < \frac{\zeta_{(r)}(A + 2B - C) - 2K_{(a),i}(A + B - C) - (A + B)}{B(2K_a + 1 - 2\zeta_{(r)})}.$$

Z tego wynika, że (6.6) nie jest sprzeczny, gdy Δ_i jest zbyt duże co do wartości bezwzględnej, tzn. np. gdy $\Delta_i < 1/2$ w regionie określoności $K_{(a),i}$. Nie jest to restrykcyjnym ograniczeniem ponieważ, ciąg ζ_i z założenia jest zbieżny. Warunek na Δ_i odpowiada założeniu z Lematu 5, że δ powinna być dostatecznie mała.

6.3 Dodatek C

6.3.1 Wyrażenie $\Delta_l(\varepsilon; \zeta)$

Minimum $\tilde{\lambda}_l^*(\varepsilon; \zeta)$ paraboli zaburzonej dla $l \in \{1, 2, 3\}$ możemy zapisać jako

$$\tilde{\lambda}_l^*(\varepsilon; \zeta) = \lambda^*(\zeta) \pm \Delta_l(\varepsilon; \zeta), \quad l \in \{1, 2, 3\},$$

gdzie $\lambda^*(\zeta)$ jest minimum paraboli niezaburzonej.

Założmy, że zarówno $\lambda^*(\zeta)$ jak i $\tilde{\lambda}_l^*(\varepsilon; \zeta)$ są minimami parabol otrzymanych za pomocą transformacji $\hat{q}(\cdot)$ opisanej w Dodatku A przy założeniu, że $f(\zeta^1) < f(\zeta^3)$, tzn. gdy zachodzi przeciwna nierówność to wykonujemy obrót paraboli względem $\zeta^0 = 1/2$ tak jak w punkcie 3. Algorytmu 3. Dla paraboli niezaburzonej mamy

$$\lambda^*(\zeta) = \frac{1}{2} \frac{A(\zeta_{(r)} + 1) + B - \zeta_{(r)}C}{A + B - C}$$

natomiast dla paraboli zaburzonej

$$\tilde{\lambda}_l^*(\varepsilon; \zeta) = \begin{cases} \frac{1}{2} \frac{A(\zeta_{(r)}+1)(1+\varepsilon)+B-\zeta_{(r)}C}{A(1+\varepsilon)+B-C} & \text{dla } l = 1, \\ \frac{1}{2} \frac{A(\zeta_{(r)}+1)+B(1+\varepsilon)-\zeta_{(r)}C}{A+B(1+\varepsilon)-C} & \text{dla } l = 2, \\ \frac{1}{2} \frac{A(\zeta_{(r)}+1)+B-\zeta_{(r)}C(1+\varepsilon)}{A+B-C(1+\varepsilon)} & \text{dla } l = 3. \end{cases} \quad (6.8)$$

dla zaburzeń w punktach 0, $\zeta_{(r)}$ oraz 1 odpowiednio, gdzie $\zeta_{(r)} = \frac{\zeta^2 - \zeta^1}{\zeta^3 - \zeta^1}$ a A, B, C są zdefiniowane w Dodatku B.

Upraszczając wyrażenie $|\lambda^* - \tilde{\lambda}_l^*(\varepsilon; \zeta)|$ otrzymujemy

$$\Delta_l(\varepsilon; \zeta) = \begin{cases} \frac{C-\zeta_{(r)}B}{2(A+B-C)} \cdot \frac{A\varepsilon}{A(\varepsilon+1)+B-C} & \text{dla } l = 1, \\ \frac{C+\zeta_{(r)}(A-C)}{2(A+B-C)} \cdot \frac{B\varepsilon}{A+B(1+\varepsilon)-C} & \text{dla } l = 2, \\ \frac{A+B(\zeta_{(r)}-1)}{2(A+B-C)} \cdot \frac{C\varepsilon}{A+B-C(1+\varepsilon)} & \text{dla } l = 3. \end{cases} \quad (6.9)$$

6.3.2 Główne warunki

Aby spełnione były nierówności (2.41) musimy zapewnić¹, że $\lambda^*(\zeta) < \zeta^2$. Jest tak, ponieważ jeśli $\tilde{\lambda}_l^*(\varepsilon; \zeta)$ jest przesunięty o $\Delta_l(\varepsilon; \zeta)$ w lewo to wówczas przesunięcie go ponownie w prawo nie spowoduje, że będzie ono większe niż ζ^2 . Jeśli $\tilde{\lambda}_l^*(\varepsilon; \zeta)$ jest przesunięte w prawo to mamy margines szerokości $2\Delta_l(\varepsilon; \zeta)$. Rozpatrzmy więc (2.41) dla zaburzeń kolejno dla $l = 1$, $l = 2$ oraz $l = 3$.

Dla $l = 1$ mamy

$$\frac{A(\zeta_{(r)}+1)(1+\varepsilon)+B-\zeta_{(r)}C}{A(1+\varepsilon)+B-C} < 2\zeta_{(r)} \quad (6.10)$$

Musimy rozpatrzeć dwa przypadki w zależności od znaku mianownika.

1. Jeśli $A(1+\varepsilon)+B-C > 0$, tzn.

$$\begin{aligned} \varepsilon &> \frac{C-B}{A} - 1 \quad \text{gdy } A > 0 \text{ lub} \\ \varepsilon &< \frac{C-B}{A} - 1 \quad \text{gdy } A < 0, \end{aligned}$$

otrzymujemy

$$(1+\varepsilon)A(1-\zeta_{(r)}) < \zeta_{(r)}(2B-C) - B$$

Również rozwiązanie tej nierówności dzielimy na dwa przypadki w zależności od znaku A

$$\begin{aligned} \varepsilon &< \frac{\zeta_{(r)}(2B-C) - B}{A(1-\zeta_{(r)})} - 1, \quad \text{gdy } A > 0 \text{ lub} \\ \varepsilon &> \frac{\zeta_{(r)}(2B-C) - B}{A(1-\zeta_{(r)})} - 1, \quad \text{gdy } A < 0. \end{aligned}$$

¹Zauważmy, że analogicznymi nierównościami są $\tilde{\lambda}_l^*(\varepsilon; \zeta) - \Delta_l(\varepsilon; \zeta) > \zeta^2$ dla $l \in \{1, 2, 3\}$. Aby były one spełnione zapewniamy $\lambda^*(\zeta) > \zeta^2$.

2. Jeśli $A(1 + \varepsilon) + B - C < 0$, tzn.

$$\begin{aligned}\varepsilon &< \frac{C-B}{A} - 1, \quad \text{gdy } A > 0 \text{ lub} \\ \varepsilon &> \frac{C-B}{A} - 1, \quad \text{gdy } A < 0,\end{aligned}$$

otrzymujemy

$$(1 + \varepsilon)A(1 - \zeta_{(r)}) > \zeta_{(r)}(2B - C) - B.$$

Analogicznie rozwiązanie tej nierówności dzielimy na dwa przypadki w zależności od znaku A .

$$\begin{aligned}\varepsilon &> \frac{\zeta_{(r)}(2B - C) - B}{A(1 - \zeta_{(r)})} - 1, \quad \text{gdy } A > 0 \text{ lub} \\ \varepsilon &< \frac{\zeta_{(r)}(2B - C) - B}{A(1 - \zeta_{(r)})} - 1, \quad \text{gdy } A < 0.\end{aligned}$$

Tak więc rozwiązanie dla $l = 1$ zapisujemy definiując b_1 i b_2

$$\begin{cases} r_1 = \frac{C-B}{A} - 1, \\ r_2 = \frac{A(\zeta_{(r)} - 1) + B(2\zeta_{(r)} - 1) - C(2\zeta_{(r)} - 1)}{A(1 - \zeta_{(r)})} \\ b_1 = \min(r_1, r_2), \quad b_2 = \max(r_1, r_2), \\ \varepsilon \in (b_1, b_2). \end{cases} \quad (6.11)$$

Rozwiązanie (6.11) nierówności (6.10) może być interpretowane jako ograniczenie błędu względnego tylko, gdy $b_1 < 0$ i $b_2 > 0$. Ostatecznie więc ograniczenie przyjmuje postać

$$|\varepsilon| < \min(|b_1|, |b_2|). \quad (6.12)$$

Zastosowanie transformacji $\hat{q}$ pozwala nam „narysować” region określony przez (6.12) w $\mathbb{R}^3$, opisaney zmiennymi $\lambda_1^*(\zeta; \varepsilon)$, ζ^2 i ε . Aby wyznaczyć region musimy rozpatrzeć wszystkie parabole postaci $y = ax^2 + bx + c$ przechodzące przez punkty $[0, K_1]$ i $[1, K_2]$ przy założeniu $K_2 > K_1$ z minimum w punkcie $0 < \lambda^* < \zeta^2$. Otrzymujemy wyrażenia na a, b oraz c postaci

$$\begin{cases} a = \frac{K_2 - K_1}{1 - 2\lambda^*}, \\ b = -2\lambda^* \frac{K_2 - K_1}{1 - 2\lambda^*}, \\ c = K_1. \end{cases} \quad (6.13)$$

Z tego otrzymujemy, że nie istnieje parabola z $\lambda^* > 1/2$ ponieważ wówczas byłoby $a < 0$ i skonstruowana parabola posiada maksimum w punkcie λ^* zamiast minimum. Tak więc rozpatrujemy jedynie $\lambda^* \in (0, 1/2)$. Po zastosowaniu transformacji $\hat{q}$ ta rodzina jest przekształcana do jednej paraboli z minimum w λ^* i przechodzi przez punkty $[0, -2]$, $[\zeta^2, \hat{q}(\zeta^2)]$ i $[1, -1]$. Jak można łatwo sprawdzić w interesującym nas regionie, tzn. dla $(\zeta_{(r)}, \lambda^*) \in (0, 1) \times (0, 1/2)$ mamy $b_1 < 0$ i $b_2 > 0$.

W ten sam sposób otrzymujemy ograniczenie na ε , gdy $l = 2$ oraz gdy $l = 3$. Rozwiązanie dla $l = 2$ możemy zapisać definiując r_1 i r_2 w postaci

$$\begin{cases} r_1 &= \frac{C-A}{B} - 1, \\ r_2 &= \frac{\zeta(r)(2A-C)-A(1+\zeta(r))}{B(1-2\zeta(r))} - 1, \\ |\varepsilon| &< \min(|r_1|, |r_2|). \end{cases} \quad (6.14)$$

Dla $l = 3$ natomiast mamy

$$\begin{cases} r_1 &= \frac{A+B}{C} - 1, \\ r_2 &= \frac{\zeta(r)(2A+2B-C)-(A(1+\zeta(r))+B)}{\zeta(r)C}, \\ |\varepsilon| &< \min(|r_1|, |r_2|). \end{cases} \quad (6.15)$$

Aby zapewnić, że wprowadzenie zaburzenia ε nie zaburzy własności wypukłości konstruowanej paraboli $\hat{q}(\cdot)$ potrzebujemy dodatkowych warunków. Zapisują się one

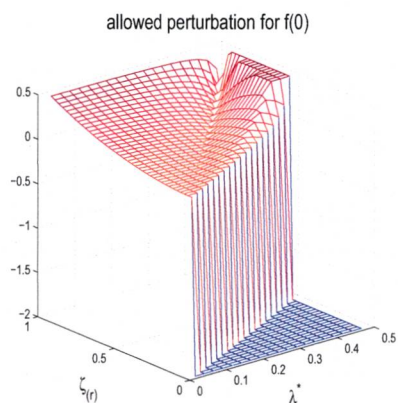
$$|\varepsilon| < \min\left(\left|-\frac{q(\zeta(r))}{2} - 1\right|, 0.5\right) \quad \text{dla } l = 1, \quad (6.16)$$

$$|\varepsilon| < \left|-\frac{q(\zeta(r))}{2} - 1\right| \quad \text{dla } l = 2, \quad (6.17)$$

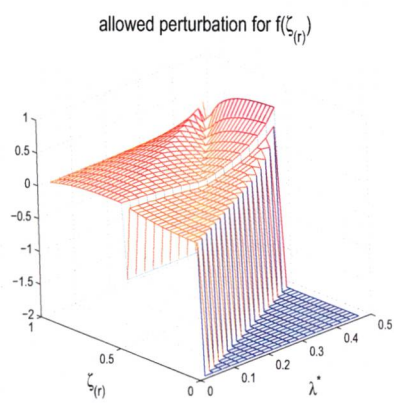
$$|\varepsilon| < 1 \quad \text{dla } l = 3, \quad (6.18)$$

Rysunek 6.5 pokazuje wykresy powyższych ograniczeń błędów.

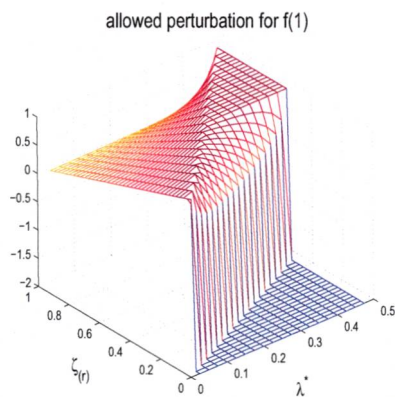
a)



b)



c)



Rysunek 6.5: Ograniczenie na ε dla $(\zeta_{(r)}, \lambda^*) \in (0, 1) \times (0, 0.5)$ (zauważmy, że $\zeta_{(r)} > \lambda^*$) a) dla $l = 1$ określone przez warunki (6.12) i (6.16) b) dla $l = 2$ określone przez warunki (6.14) i (6.17) (zauważmy, że dla $\zeta_{(r)} = 1/2$ mamy osobliwość) c) dla $l = 3$ określone przez warunki (6.15) i (6.18).

Bibliografia

- [1] M. Bazan, Search procedure exploiting locally regularized objective approximation. A convergence theorem for direct search algorithms. *Computational Intelligence in Optimization-Applications and Implementations*, Springer 2010.
- [2] M. Bazan, S. Russenschuck, Using neural networks to speed up optimization algorithms, *Eur. Phys. J.*, **AP 12**, pp. 109-115, 2000.
- [3] M. Bazan, M. Aleksa, S. Russenschuck, An improved method using radial basis function neural networks to speed up optimization algorithms, *IEEE Trans. on Magnetism*, **38**, p. 1081-1084, 2002.
- [4] M. Bazan, M. Aleksa, J. Lucas, S. Russenschuck, S. Ramberger, C. Völlinger, Integrated design of superconducting magnets with the CERN field computation program ROXIE, *Proc. 6-th International Computational Accelerator Physics Conference*, Darmstadt, Germany, September 2000.
- [5] M. Bazan, Using a regularized RBF approximation to speed-up optimization algorithms, *International Congress of Computational and Applied Mathematics*, Leuven, Belgium, July 2006.
- [6] M. Bazan, Ph. Bernard, C. Dalmas, W. Weingarten, A mono-cell thin niobium film cavity excited in the TE_{011} mode at 972MHz, *SL-Note-97-64-RF*, CERN, Geneva, 1997.
- [7] M. Bazan, H. Haebel, H. P. Kinderman, Numerical optimisation of the 1.2MW coupler for the SPS as LHC injector, *SL-Note-97-75-RF*, CERN, Geneva, 1997.
- [8] M. Bazan, Improved roots for semi-analytic eigenvalue processor, *SL-Note-97-76-RF*, CERN, Geneva, 1997,
Report 05/97, Inst. Informatyki, Uniwersytet Wrocławski, Wrocław, 1997.
- [9] M. D. Buhmann, Radial basis functions: Theory and Implementations, Cambridge University Press, 2005.
- [10] P. D. Christofides, A. Armaou, Y. Lou, A. Varshney, Control and Optimization of Multiscale Process Systems, Birkhäuser, 2008.
- [11] A. R. Conn, N. I. M. Gould, Ph. L. Toint, Trust region methods, SIAM, 2005.

- [12] O. Davydov, Error Bound for Radial Basis Interpolation in Terms of a Growth Function, Strathclyde Mathematics Research Report No. 24, 2006.
- [13] O. Davydov, F. Zeilfelder, Scattered data fitting by direct extension of local polynomials to bivariate splines, *Adv. Comp. Math.*, **21**, pp. 223-271, 2004.
- [14] J. E. Goodman, J. O'Rourke, Handbook of discrete and computational geometry, Chapman & Hall/CRC, 2004.
- [15] D. Goldfarb, A family of variable metric updates derived by variational means, *Mathematics of Computation*, **24**, pp. 23-26, 1970.
- [16] H. Górecki, Optymalizacja i sterowanie systemów dynamicznych, AGH, Kraków 2006.
- [17] P. Ch. Hansen, Rank-deficient and Discrete Ill-posed Problems", SIAM, Philadelphia 1998.
- [18] H. G. Jacob, Rechnergestützte Optimierung statischer und dynamischer Systeme, Springer, 1982.
- [19] K. Jetter, J. Stöckler, J. Ward, Error estimates for scattered data interpolation on Spheres, *Math. Comput.*, **68**, pp. 733-747, 1999.
- [20] E. Larsson, B. Fornberg, Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions, *Comp. and Math. with Applications*, **49**, pp. 103-130, 2005.
- [21] D. G. Luenberger, Introduction to linear and nonlinear programming, 2nd ed., Addison-Wesley, New York, 1984.
- [22] Ch. A. Micchelli, Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions, *Constructive Approximation*, Vol. 2, 1986, pp. 11-22.
- [23] J. J. Moré, B. S. Garbow, K. E. Hillstom, Testing unconstrained optimization software, *ACM Trans. Math. Software*, **7**, No. 1, pp. 17-41, 1981.
- [24] F. J. Norcowich, J. D. Ward, H. Wendland, Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting, *Math. Comput.*, **74**, pp. 643-763, 2005.
- [25] E. Polak, Optimization. Algorithms and Consistent Approximations, Applied Mathematical Sciences 124, Springer, 1997.
- [26] M. J. D. Powell, On calculation of orthogonal vectors, *The Computer Journal*, Vol. **11**(3), 1968, pp. 302-304.
- [27] H. H. Rosenbrock, An Automatic Method for finding the Greatest or Least Value of Function, *The Computer Journal*, Vol. **3**(3), 1960, pp. 175-184.

- [28] S. Russenschuck, ROXIE: Routine for the optimization of magnet X-section, inverse field calculation and coil end design, CERN Yellow Report 99-01, April 1999.
- [29] R. Schaback, Error estimates and condition number for radial basis function interpolation, *Adv. Comput. Math.*, **3**, pp. 251-264, 1995.
- [30] Z. M. Wu, R. Schaback, Local error estimates for radial basis function interpolation of scattered data, *I.M.A. J. Numer. Anal.* Vol. 13 (1993), p. 13-27.
- [31] R. Schaback, Error estimates and condition number for radial basis function interpolation, *Adv. Comput. Math.*, **3**, pp. 251-264, 1995.
- [32] R. Schaback, Native Hilbert Spaces for Radial Basis Functions I, The new development in Approximation Theory, Birkhäuser, Basel, 1999.
- [33] R. Schaback, Improved error bounds for scattered data interpolation by radial basis functions, *Mathematics of Computation*, **68**, pp. 201-216, 1999.
- [34] R. Schaback, Multivariate Interpolation by Polynomials and Radial Basis Functions, *Constr. Approx.*, **21**, pp. 293-317, 2005.
- [35] W. Y. Svrcek, D. P. Mahoney, B. R. Young, A real-time approach to process control. John Wiley & Sons, Ltd., 2006.
- [36] D. E. Seaborg, T. F. Edgar, D. A. Mallichamp, Process Dynamics and Control, Wiley & Sons, Inc., 2003.
- [37] P. Tatjewski, Sterowanie zaawansowanych obiektów przemysłowych. struktury i algorytmy, Akademicka Oficyna Wydawnicza EXIT, 2002.
- [38] A. N. Tikhonov, V. A. Arsenin, Solution of ill-posed problems, Winston & Sons, Washington, 1977.
- [39] R. Oeuvray, Trust region method based on radial basis functions with application on biomedical imaging, Ecole Polytechnique Federale de Lausanne, 2005.
- [40] G. Wahba, Spline models for observational data, SIAM, Philadelphia, 1990.
- [41] H. Wendland, C. Rieger, Approximate interpolation with applications to selecting smoothing parameters, *Numerische Mathematik*, Vol. 101 (2005), p. 643-662.
- [42] H. Wendland, Gaussian Interpolation Revisited, In K. Kopotun, T. Lyche, M. Neamtu, eds., *Trends in Approximation Theory*, pp. 427-436. Nashville, Vanderbilt University Press, 2001.
- [43] H. Wendland, Scattered Data Approximation, Cambridge University Press, 2005.

- [44] Yun-Wei Shang, Yu-Huang Qui, A note on the extended Rosenbrock function, *Evolutionary Computation*, **14**, No. 1, pp. 119-126, 2006.
- [45] S. Węgrzyn, Podstawy automatyki, PWN, 1972.
- [46] W. I. Zangwill, Nonlinear Programming; a Unified Approach, Prentice-Hall International Series, Englewood Cliffs, 1969.
- [47] K. Zhou, J. C. Doyle, K. Glover, Robust and optimal control, Prentice Hall, 1995.