

UNIwersytet Gdański  
Wydział Matematyki, Fizyki i Informatyki

KAROLINA AGATA LADEMANN

---

**Metody aproksymacji numerycznej  
liniowego równania Kleina-Gordona  
z masą czasowo- i przestrzenno-zależną**

---



*Rozprawa doktorska złożona jako spełnienie wymagań na stopień  
Doktora Nauk Ścisłych i Przyrodniczych*

*napisana pod kierunkiem  
dr. hab., prof. IM PAN Karoliny Kropielnickiej*

27 maja 2023

Załącznik nr 2  
do Zasad przeprowadzania przewodów doktorskich  
na Wydziale Matematyki, Fizyki i Informatyki  
Uniwersytetu Gdańskiego

OŚWIADCZENIE

Ja, niżej podpisana oświadczam, iż przedłożona praca doktorska została wykonana przez mnie samodzielnie, nie narusza praw autorskich, interesów prawnych i materialnych innych osób.

.....  
data

.....  
własnoręczny podpis

## Streszczenie

Tematem niniejszej rozprawy jest konstrukcja i analiza metod aproksymacji numerycznej dla równań liniowych oscylujących z różnymi częstotliwościami, na przykładzie równania Kleina-Gordona z masą czasowo- i przestrzennie-zależną, postaci

$$\partial_t^2 \psi(\mathbf{x}, t) = \Delta \psi(\mathbf{x}, t) + f(\mathbf{x}, t) \psi(\mathbf{x}, t). \quad (1)$$

Szczególnie ciekawym elementem rozważanego równania jest postać funkcji masy  $f(\mathbf{x}, t)$ , która może być charakteryzowana oscylacjami o różnych częstotliwościach, od małych do skrajnie wysokich. Przyjmujemy, że może być ona wyrażona lub przybliżona w następujący sposób

$$f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + \sum_{|n| \leq N} a_n(\mathbf{x}, t) e^{i\omega_n t}, \quad (2)$$

gdzie  $n \in \mathbb{N}$ ,  $\omega_n \in \mathbb{R}$ ,  $\omega_{\min} = \min_{n \leq N} |\omega_n| \geq 1$  oraz  $\omega_{\max} = \max_{n \leq N} |\omega_n| < \infty$ . Funkcje  $\alpha(\mathbf{x}, t)$  oraz  $a_n(\mathbf{x}, t)$ ,  $n \leq N$ , nie oscylują w zależności od  $\omega_n$  (tj. ich pochodne czasowe są ograniczone niezależnie od  $\omega_n$ ).

Funkcja  $f(\mathbf{x}, t)$  może przyjmować postać funkcji nieoscylującej (np.  $\forall_n a_n(\mathbf{x}, t) \equiv 0$ ), funkcji wysoko-oscylującej (np.  $f(\mathbf{x}, t) = a_2(\mathbf{x}, t) e^{i10^6 t}$ ) lub funkcji mieszanej (np.  $f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + a_1(\mathbf{x}, t) e^{it} + a_2(\mathbf{x}, t) e^{i10^6 t}$ ). Klasyczne metody numeryczne, oparte na twierdzeniu Taylora oraz kwadraturach, są bardzo efektywne w przypadku zagadnień nieoscylujących. Jednakże, w obliczu oscylacji wymagają one wprowadzenia zależności dla wielkości kroku czasowego  $h < 1/\omega_{\max}$ , co wpływa negatywnie na ich efektywność i koszt obliczeniowy, szczególnie gdy oscylacje są bardzo duże.

Liniowe równanie Kleina-Gordona (1) jest reprezentatywnym modelem, na przykładzie którego proponujemy i omówimy różne podejścia analizy obliczeniowej problemu oscylacji w komponencie zewnętrznym  $f(\mathbf{x}, t)$ , w tym rozwinięcie asymptotyczne, formułę Duhamela, rozwinięcie Magnusa oraz dekompozycję pola wektorowego.

Motywacją naszych badań nad równaniem (1) była praca [Bader et al., 2019], w której po raz pierwszy zaprezentowano metody obliczeniowe dla takiego równania. Proponowane tam schematy numeryczne sprawdzają się tylko w przypadkach nieoscylującej funkcji  $f(\mathbf{x}, t)$ . Zaciekawieni problemami wynikającymi z wysokich oscylacji, zaproponowaliśmy trzy inne podejścia numeryczne w obecności różnych oscylacji w równaniu Kleina-Gordona (1).

W pierwszej z prezentowanych metod zakładamy, że rozwiązanie równania Kleina-Gordona (1) przyjmuje formę nieskończonego zmodyfikowanego szeregu Fouriera. W wyniku otrzymujemy numeryczno-asymptotyczną metodę rozwiązywania równania Kleina-Gordona, której dokładność rośnie z częstotliwością oscylacji  $\omega_n$ . Metoda skonstruowana została z myślą wyłącznie o wysokich częstotliwościach funkcji  $f(\mathbf{x}, t)$

Druga z prezentowanych w rozprawie metod działa jednakowo dobrze, z trzecim rzędem zbieżności, w obliczu każdego rodzaju oscylacji. Ponadto, stała błędów tej metody nie rośnie wraz ze wzrostem  $\omega_n$ . Cel ten udało się osiągnąć dzięki przestawieniu równania Kleina-Gordona (1) w postaci niejednorodnego układu równań różniczkowych zwyczajnych oraz zastosowaniu do niego dobrze znanej metody niezmienniania stałej, znanej również jako reguła Duhamela. Ponadto, zastosowanie metody Filona do aproksymacji wysoko-oscylujących całek, które wchodzi w skład schematu numerycznego, pozwala na uniezależnienie kroku czasowego  $h$  od wielkości oscylacji  $\omega_n$ . Oprócz zaprezentowania metody, przedstawimy szczegółowy dowód zbieżności i jej globalnego rzędu dokładności oraz omówimy strukturę błędów metody i wyjaśnimy, dlaczego nie mają na nią wpływu możliwie ekstremalnie wysokie oscylacje.

W trzeciej prezentowanej metodzie numerycznej wykorzystujemy rozwinięcie Magnusa oraz różne techniki dekompozycji pola wektorowego, które są dobrze znanymi metodami aproksymacji rozwiązań równań ewolucyjnych. W przeciwieństwie do metody opartej na regule Duhamela, nie udowodnimy globalnego rzędu zbieżności, ale opierając się na zbieżności wykorzystywanych narzędzi, skoncentrujemy się na wyznaczeniu dominującego czynnika błędów, który odpowiada lokalnemu błędowi przybliżenia. Pokażemy, że

lokalna dokładność metody opisana jest funkcjami

$$\mathcal{O}(h^5 + \min\{h^3, \frac{h^2}{\omega_{\min}}, h^5\omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \neq 0$$

oraz

$$\mathcal{O}(\min\{h^3, \frac{h^2}{\omega_{\min}}, h^5\omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \equiv 0,$$

które zależą od relacji kroku czasowego  $h$ ,  $\omega_{\min}$  i  $\omega_{\max}$ . Metoda ta działa efektywnie dla wszystkich wielkości częstotliwości oscylacji w funkcji  $f(\mathbf{x}, t)$  i cechuje się najwyższą dokładnością z pośród zaproponowanych dotychczas metod dla liniowego równania Kleina-Gordona z masą czasowo- i przestrzenno-zależną.

Metody opisane są w oddzielnych rozdziałach, a każdy rozdział kończy się zaprezentowaniem wyników symulacji numerycznych i porównaniem do dotychczas opisanych metod.

# Abstract

In this doctoral thesis we are concerned with the construction and analysis of computational methods for linear equations, which oscillate with various frequencies. We present our ideas on the example of the linear Klein-Gordon equation with time- and space-dependent mass, of the form

$$\partial_t^2 \psi(\mathbf{x}, t) = \Delta \psi(\mathbf{x}, t) + f(\mathbf{x}, t) \psi(\mathbf{x}, t). \quad (3)$$

The form of the mass function  $f(\mathbf{x}, t)$  is a particularly interesting issue of this equations, as it can be characterized by oscillations of various frequencies, from low to extremely high. We assume that it can be expressed or approximated in the following form

$$f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + \sum_{|n| \leq N} a_n(\mathbf{x}, t) e^{i\omega_n t}, \quad (4)$$

where  $n \in \mathbb{N}$ ,  $\omega_n \in \mathbb{R}$ ,  $\omega_{\min} = \min_{n \leq N} |\omega_n| \geq 1$  and  $\omega_{\max} = \max_{n \leq N} |\omega_n| < \infty$ . Functions  $\alpha(\mathbf{x}, t)$  and  $a_n(\mathbf{x}, t)$ ,  $n \leq N$ , do not oscillate in  $\omega_n$ s (i.e. their time derivatives are bounded independently of  $\omega_n$ ).

The function  $f(\mathbf{x}, t)$  can take the form of a non-oscillatory function (e.g.  $\forall_n a_n(\mathbf{x}, t) \equiv 0$ ), a highly-oscillatory function (e.g.  $f(\mathbf{x}, t) = a_2(\mathbf{x}, t) e^{i10^6 t}$ ) or a mixed function (e.g.  $f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + a_1(\mathbf{x}, t) e^{it} + a_2(\mathbf{x}, t) e^{i10^6 t}$ ). Classical numerical methods, based on Taylor's expansion and quadratures, are very effective for non-oscillatory problems. However, in the presence of oscillations, they require restrictions on the time step size, namely  $h < 1/\omega_{\max}$ , which has a negative effect on their efficiency and computational cost, especially when the oscillations are very large.

The linear Klein-Gordon equation (3) is a representative model, on which we will propose and discuss various approaches to the computational analysis of the problem of oscillatory input term  $f(\mathbf{x}, t)$ , including asymptotic expansion, variation of constants, Magnus expansion and decomposition of the vector field.

The motivation for our research on the equation (3) was the paper [Bader et al., 2019], in which the computational methods for such an equation were presented for the first time. The numerical schemes proposed there work only in case of the non-oscillatory function  $f(\mathbf{x}, t)$ . The problems of the classical methods and those presented in the paper [Bader et al., 2019], resulting from the presence of high oscillations in the function  $f$ , seemed very interesting to us, therefore we proposed three other numerical approaches in the presence of different oscillations in the Klein-Gordon equation (3).

In the first presented method, we assume that the solution of the Klein-Gordon equation (3) may take the form of an infinite modified Fourier series. As a result we get numerical-asymptotic method, which accuracy increases with the increasing magnitude of oscillation frequencies  $\omega_n$ . The method was designed for the case, when we have only high frequencies in function  $f(\mathbf{x}, t)$

The second method presented in the thesis is a third-order in time method which is equally efficient in the full range of frequencies, e.g.  $f(x, t) = a_0(x, t)$ ,  $mf(x, t) = a_2(x, t) e^{i10^6 t}$  and  $f(x, t) = a_0(x, t) + a_1(x, t) e^{it} + a_2(x, t) e^{i10^6 t}$ , where the error constant does not grow with the size of  $\omega_n$ s, and no ratio between time step  $h$  and  $\omega_{\max}$  needs to be imposed. This goal was achieved by rewriting the Klein-Gordon equation (3) in the form of a non-homogeneous system of ordinary differential equations and by application of the well-known variation of constants method, also known as Duhamel's formula. Moreover, the use of the Filon method for approximation of highly oscillatory integrals, which are part of the numerical scheme, makes the time step size  $h$  independent of the magnitude of oscillations  $\omega_n$ . We will present a rigorous proof of convergence, discuss the structure of the error of the method and explain why it is not affected by possibly extremely high oscillations.

In the third presented numerical method, we use the Magnus expansion and various decompositions of the vector field, which are well-known methods of approximation of the solution of evolution equations. Unlike the method based on Duhamel's formula, we will not prove the global order of convergence, but will focus on local accuracy by determining the leading error terms appearing during the derivation of the

method. We will show that the local accuracy of the method is described by the expressions

$$\mathcal{O}(h^5 + \min\{h^3, \frac{h^2}{\omega_{\min}}, h^5\omega_{\max}^2\}) \quad \text{for } \alpha(\mathbf{x}, t) \neq 0$$

and

$$\mathcal{O}(\min\{h^3, \frac{h^2}{\omega_{\min}}, h^5\omega_{\max}^2\}) \quad \text{for } \alpha(\mathbf{x}, t) \equiv 0,$$

which depend on the relationship between time step  $h$  and  $\omega_{\min}$ ,  $\omega_{\max}$ . This method works effectively for all magnitudes of oscillation in function  $f(\mathbf{x}, t)$  and shows up the highest accuracy among the methods proposed so far for the linear Klein-Gordon equation with time- and space-dependent mass.

The methods are described in separate chapters, and each chapter ends with a presentation of the results of numerical simulations and a comparison to the methods presented before.

*Nauka nie zajmuje się takimi własnościami bytu, do których należy śmieszność.  
Nauka objaśnia świat, ale pogodzić się z nim może jedynie sztuka.  
Stanisław Lem, "Bajki robotów"*

## *Podziękowania*

Dziś, na zakończenie ważnego etapu mojego rozwoju naukowego, chciałbym podziękować w pierwszej kolejności osobom, z którymi tę drogę rozpoczęłam. Pierwsze podziękowania kieruję do dr Moniki Wrzosek oraz ś.p. dra hab. Prof. UG Henryka Leszczyńskiego za pokazanie mi naukowej ścieżki rozwoju oraz umożliwienie mi stawiania na niej pierwszych kroków.

Pragnę podziękować moim bliskim, Rodzicom Krystynie i Wiesławowi, Teściom Irenie i Ludwikowi, a w szczególności Mężowi Karolowi oraz Synowi Gabrielowi za wyrozumiałość, cierpliwość, serce i nieustanne wspieranie mnie w moim życiowym wyborze. Podziękowania należą się także Alicji Urbańskiej za motywowanie mnie do dalszej pracy, także dydaktycznej oraz dr Marcie Leśniak za wsparcie praktyczne i wiarę we mnie podczas całego okresu studiów doktoranckich.

Największe podziękowania składam mojej Pani Promotor, dr hab. Prof. IM PAN Karolinie Kropielniczej. To dzięki niej mogłam kontynuować rozpoczęty doktorat. Otoczyła mnie opieką dydaktyczną, ale także przyjacielską. Zadbala o mój wszechstronny rozwój, umożliwiając mi uczestnictwo w wielu ważnych konferencjach krajowych i zagranicznych, kontakt z wybitnymi matematykami oraz udział w licznych projektach naukowych. Służyła mi radą i doświadczeniem w każdej wątpliwości, zachęcała do szukania nowych rozwiązań i samorozwoju, jednocześnie budując we mnie poczucie przynależności do społeczności naukowej. Dziękuję za serce, czas, zaangażowanie i wsparcie w trakcie całego okresu naszej współpracy, a w szczególności za cierpliwość i wyrozumiałość podczas spisywania rozprawy doktorskiej.

Dziękuję ponadto Prof. Katharinie Schratz, dr Marissie Condon oraz Rafałowi Perczyńskiemu za owocną współpracę, Prof. Arieh Iserles za cenne uwagi dotyczące publikacji oraz Prof. Tomaszowi Szarkowi za pomoc w umożliwieniu kontynuacji doktoratu, a także za rozwój mojej ścieżki dydaktycznej.





# Spis treści

<b>1</b>	<b>Wstęp</b>	<b>1</b>
<b>2</b>	<b>Metoda numeryczna oparta na rozwinięciu asymptotycznym</b>	<b>5</b>
2.1	Wyprowadzenie metody asymptotycznej	5
2.2	Przykłady numeryczne	8
<b>3</b>	<b>Metoda numeryczna oparta na zastosowaniu formuły Duhamela</b>	<b>13</b>
3.1	Wyprowadzenie metody	13
3.1.1	Metoda Filon	15
3.2	Zbieżność metody	16
3.3	Przykłady numeryczne	22
<b>4</b>	<b>Metoda numeryczna oparta na rozwinięciu Magnusa oraz dekompozycji pola wektorowego</b>	<b>25</b>
4.1	Wyprowadzenie metody	26
4.1.1	Rozwinięcie Magnusa	26
4.1.2	Splitting Stranga	27
4.1.3	Składniki zewnętrzne splittingu (4.10)	28
4.1.4	Składnik wewnętrzny splittingu (4.10)	29
	Schemat $\Gamma_1^{[4]}$	30
	Schemat $\Gamma_2^{[4]}$	30
4.1.5	Kompletne schematy numeryczne $\Gamma_1^{[4]}$ oraz $\Gamma_2^{[4]}$	30
4.2	Oszacowania dominujących czynników błędu	32
4.2.1	Dominujący czynnik błędu obciążenia rozwinięcia Magnusa	33
4.2.2	Dominujący czynnik błędu dla splittingu Stranga	36
4.2.3	Dominujące czynniki błędu dla splittingu 4-go rzędu (4.13)	37
4.3	Struktura błędu	38
4.4	Przykłady numeryczne	40
4.5	Dodatek	45
4.5.1	Obliczenia dotyczące elementów rozwinięcia Magnusa (4.4)–(4.7)	45
4.5.2	Obliczenia dotyczące dominujących czynników błędu wynikających z zastosowania splittingu Stranga w podrozdziale 4.2.2	47
4.5.3	Wyprowadzenie splittingu 4-go rzędu typu Chin-Chen	48
<b>5</b>	<b>Podsumowanie i dalsze perspektywy</b>	<b>51</b>



# Rozdział 1

## Wstęp

Tematem niniejszej rozprawy jest konstrukcja i analiza metod obliczeniowych dla równań liniowych wysoko oscylujących, na przykładzie równania Kleina-Gordona z masą czasowo- i przestrzenno- zależną. Szczególnie ciekawym problemem rozważanego równania jest postać funkcji masy, która może oscylować z różnymi częstotliwościami, od małych do skrajnie dużych.

Oscylacje są naturalnym zjawiskiem otaczającego nas środowiska, często występującym w postaci fal, np. elektromagnetycznych, akustycznych, radiowych, świetlnych. Naturalnym jest, że równania różniczkowe opisujące te zjawiska są trudnym przedmiotem analizy obliczeniowej. Występujące w nich oscylacje powodują, że klasyczna analiza numeryczna, opierająca się na koncepcji twierdzenia Taylora zawodzi, gdyż błąd metody skalujący jak  $h^n$  (gdzie  $h$  jest krokiem czasowym) obarczony jest stałą błędu zależną od  $n$ -tej pochodnej oscylującego komponentu równania. Wystarczy pomyśleć o twierdzeniu Taylora zastosowanym do funkcji  $\cos(\omega t)$  z  $\omega \gg 1$ , aby stwierdzić, że reszta jego rozwinięcia jest wielkości  $\frac{h^n}{n!} \omega^n$ , która jest nieakceptowalnie duża gdy  $h > \frac{1}{\omega}$ .

Rozwój technologiczny napędzający naukę w dużej mierze bazuje na modelach matematycznych, które są niczym innym jak równaniami różniczkowymi z oscylującymi komponentami. Te z kolei, muszą być rozwiązywane i symulowane numerycznie w celu zrozumienia obserwowanych zjawisk. W związku z tym, w ostatnich dekadach nauka obfituje w aktywne badania dotyczące konstrukcji narzędzi obliczeniowych do rozwiązywania problemów wysoko-oscyłujących, [Engquist et al., 2009, Abdulle et al., 2012, Deaño et al., 2018]. Zaniedbanym przypadkiem wydaje się ten, w którym oscylacje występują w szerokim zakresie skal, od niskich do bardzo wysokich.

Liniowe równanie Kleina-Gordona jest reprezentatywnym modelem, na przykładzie którego zaproponujemy i omówimy różne podejścia analizy obliczeniowej do wysokich oscylacji. Oparte one będą na koncepcji rozwinięcia asymptotycznego, formule Duhamela, rozwinięcia Magnusa oraz dekompozycji pola wektorowego.

Zakładając, że stała Plancka  $\hbar$  i prędkość światła  $c$  przyjmują wartość 1 (tzw. *atomic scaling*), liniowe równanie Kleina-Gordona z masą czasowo- i przestrzenno-zależną przyjmuje postać

$$\partial_t^2 \psi(\mathbf{x}, t) = \Delta \psi(\mathbf{x}, t) + f(\mathbf{x}, t) \psi(\mathbf{x}, t) \quad (1.1)$$

i analizowane jest z warunkami początkowymi  $\psi(\mathbf{x}, 0) = \psi_0(\mathbf{x})$ ,  $\partial_t \psi(\mathbf{x}, t_0) = \varphi_0(\mathbf{x})$  oraz okresowymi warunkami brzegowymi<sup>1</sup>, gdzie  $\mathbf{x} \in \mathbb{T}^d$ .

Zakładamy, że funkcja  $f(\mathbf{x}, t)$  może być wyrażona lub przybliżona w następujący sposób

$$f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + \sum_{|n| \leq N} a_n(\mathbf{x}, t) e^{i\omega_n t}, \quad (1.2)$$

gdzie  $n \in \mathbb{N}$ ,  $\omega_n \in \mathbb{R}$ ,  $\omega_{\min} = \min_{n \leq N} |\omega_n| \geq 1$  oraz  $\omega_{\max} = \max_{n \leq N} |\omega_n| < \infty$ . Funkcja ta nazywana jest często funkcją masy lub komponentem zewnętrznym (tzw. *input term* lub *forcing term*). Przyjmujemy, że funkcje  $\alpha(\mathbf{x}, t)$  oraz  $a_n(\mathbf{x}, t)$ ,  $|n| \leq N$  nie oscylują jak  $\omega_n$  (tj. ich pochodne czasowe są ograniczone niezależnie od  $\omega_n$ ).

<sup>1</sup>Założenie okresowych warunków brzegowych ma na celu wyłącznie uproszczenie prezentacji, a wszystkie proponowane metody można łatwo uogólnić na inne warunki brzegowe.

Dla potrzeb analizy matematycznej prezentowanej w rozprawie nie potrzebujemy założenia o nieujemności funkcji  $f(\mathbf{x}, t)$ . Wystarczy założyć jej gładkość, ale dokładne, mniej restrykcyjne założenia o regularności funkcji  $\alpha(\mathbf{x}, t)$  oraz  $a_n(\mathbf{x}, t)$ ,  $|n| \leq N$  zależą od metody numerycznej i przedstawiamy je w poszczególnych rozdziałach pracy.

Funkcja  $f(\mathbf{x}, t)$  może przyjąć trzy formy:

1. funkcji *nieoscyłującej*, np. gdy  $a_n(\mathbf{x}, t) \equiv 0$ ,  $\forall |n| \leq N$ ,  $f(\mathbf{x}, t) = \alpha(\mathbf{x}, t)$ ;
2. funkcji *wysoko-oscyłującej*, np. gdy  $\alpha(\mathbf{x}, t) \equiv 0$ ,  $\omega_{\min} \gg 1$ ,  $f(\mathbf{x}, t) = a_1(\mathbf{x}, t)e^{i10^9 t}$ ;
3. funkcji *mieszanej*, która zawiera części nieoscyłujące i oscyłujące o różnych częstotliwościach, np. gdy  $f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + a_1(\mathbf{x}, t)e^{i10^t} + a_2(\mathbf{x}, t)e^{i10^4 t} + a_3(\mathbf{x}, t)e^{i10^9 t}$ .

Motywacja do studiowania równania Kleina-Gordona (1.1) ma dwa źródła - zapotrzebowanie na wydajne metody obliczeniowe w przypadku wysokich częstotliwości oscylacji oraz jego faktyczne zastosowania w nauce, np. w mechanice kwantowej ciał niebieskich. Istotnie, liniowe równanie Kleina-Gordona postaci

$$\left(\partial_t^2 - \mathcal{B}\right) \psi(\mathbf{x}, t) = 0, \quad \mathcal{B} := \Delta + f, \quad (1.3)$$

jest rozważane w wielu publikacjach naukowych. Jednym z podstawowych problemów związanych z równaniem (1.3) jest ujemna gęstość prawdopodobieństwa rozwiązania. Ali Mostafazadeh rozwiązał ten problem w pracach [Mostafazadeh, 2003], [Mostafazadeh, 2004], wprowadzając zależność funkcji masy  $f$  od zmiennej przestrzennej ( $f = f(\mathbf{x})$  w równaniu (1.3)). Problem ujemnej gęstości prawdopodobieństwa został rozwiązany, ale pojawił się nowy związany z naruszeniem zasady symetrii Lorentza<sup>2</sup>. Ten problem, z kolei, został rozwiązany przez Miloslava Znojila w pracach [Znojil, 2017a], [Znojil, 2017b] poprzez zezwolenie na czasową i przestrzenną zależność funkcji masy (tj.  $f = f(\mathbf{x}, t)$ ). Dzięki temu zabiegowi równanie (1.3) znajduje zastosowanie, np. w mechanice kwantowej ciał niebieskich, [Znojil, 2016], i nie jest obciążone niefizycznymi osobliwościami.

Dodatkowe informacje na temat równania Kleina-Gordona i jego znaczenia w nauce i inżynierii, można znaleźć w [Greiner et al., 1996, Shankar, 2012, Schwartz, 2014].

Wielu matematyków i fizyków przedstawiło zaawansowane metody obliczeniowe dla liniowych i nieliniowych równań Kleina-Gordona, [Chen and Liu, 2008], [Bao and Dong, 2012], [Faou and Schratz, 2014], [Shakeri and Dehghan, 2008], [Yusufoğlu, 2008], jednakże, w żadnej z tych prac nie rozważano przypadku gdy funkcja  $f$  jest zależna od czasu i przestrzeni. Prawdopodobnie związane jest to z tym, że prace M. Znojila pojawiły w 2017 r.

Według naszej wiedzy, pierwsza propozycja numerycznego rozwiązywania równania Kleina-Gordona z masą zależną od czasu i przestrzeni (1.1) została przedstawiona w pracy [Bader et al., 2019], gdzie autorzy proponują metody numeryczne rzędu czwartego i szóstego, wyprowadzone w oparciu o bezkomutatorowe rozwinięcie typu Magnusa, [Magnus, 1954]. Metody te sprawdzają się bardzo dobrze w przypadkach nieoscyłacyjnych, jednak wprowadzają duże stałe błędów w przypadku wysokich częstotliwości  $\omega_{\max} \gg 1$ .

Wizualizacja rozwiązania równania (1.1) wymaga implementacji metod numerycznych, co w praktyce sprowadza się do ustalenia kroku czasowego  $h$  obliczeń. W przypadku, gdy  $h \leq \frac{1}{\omega_{\max}}$  mówimy o przykładzie nieoscyłującym. Zauważmy jednak, że dla  $\omega_{\max} \gg 1$  obliczenia numeryczne z udziałem bardzo małego kroku czasowego, np.  $h = 10^{-9}$ , prowadzą do błędów obliczeniowych oraz wysokiego kosztu czasowego i pamięciowego algorytmów, co z kolei może uniemożliwić wykonywanie symulacji nawet na superkomputerach, np. Tryton i Kraken w Gdańsku. Gdybyśmy jednak zdecydowali się na tradycyjne metody oparte na twierdzeniu Taylora oraz na większy krok czasowy, czyli  $h \geq \frac{1}{\omega_{\min}} > \frac{1}{\omega_{\max}}$ , ominięlibyśmy problem wysokich kosztów czasowych i pamięciowych oraz błędów obliczeniowych uzyskując  $n$ -ty rząd zbieżności metody (czyli lokalna dokładność  $\mathcal{O}(h^{n+1})$ ). Wszystko to przestałoby jednak odgrywać istotną rolę w obliczu dużej

<sup>2</sup>W fizyce relatywistycznej symetria Lorentza, nazwana na cześć Hendrika Lorentza, jest równoważnością obserwacji lub symetrii obserwacji ze względu na szczególną teorię względności, co oznacza, że prawa fizyki pozostają takie same dla wszystkich obserwatorów poruszających się względem siebie w układzie inercjalnym. Został również opisany jako cecha natury, która mówi, że wyniki eksperymentów są niezależne od orientacji lub prędkości przyspieszania laboratorium w przestrzeni".

stałej błędu, która może skalować nawet jak  $(\omega_{\max})^n$ . Tak właśnie dzieje się w przypadku metod zaprezentowanych w pracy [Bader et al., 2019].

Naturalnym stało się zadanie pytania, jak powinna być skonstruowana metoda, która będzie efektywna w przypadku wysokich częstotliwości funkcji  $f(\mathbf{x}, t)$ . Odpowiedź na to pytanie zostanie sformułowana w Rozdziale 2 niniejszego doktoratu, który oparty jest na publikacji *Asymptotic numerical solver for the linear Klein–Gordon equation with space- and time-dependent mass* [Condon et al., 2021] napisanej we współpracy z Marrisą Condon<sup>3</sup>, Karoliną Kropielniczą<sup>4</sup> oraz Rafałem Perczyńskim<sup>5</sup>. Pokażemy, że zaproponowana metoda jest wysoce wydajna w przypadku wyłącznie wysokich oscylacji, tj.  $\alpha(\mathbf{x}, t) = 0$  i  $\omega_{\min} \gg 1$ , niestety zawodzi w przypadku nieoscylującym lub mieszanym, tj. gdy  $\omega_{\max} \sim 1$ .

Zauważmy, że metody przedstawione w pracy [Bader et al., 2019] działają tylko w przypadku nieoscylującym, podczas gdy metoda asymptotyczna z [Condon et al., 2021] jest efektywna tylko w obliczu bardzo wysokich oscylacji funkcji masy  $f(\mathbf{x}, t)$ . Nasuwa się zatem kolejne pytanie, czy możliwa jest konstrukcja takiej metody numerycznej, która działałaby w przypadku nieoscylującym, oscylującym i mieszanym. W prezentowanej pracy doktorskiej przedstawimy dwie takie metody.

Pierwsza tego typu metoda opisana jest w Rozdziale 3, który jest oparty na napisanej we współpracy z Karoliną Kropielniczą pracy *Third-order exponential integrator for linear Klein–Gordon equations with time and space-dependant mass* [Kropielnicka and Lademann, 2022]. Podstawą zaprezentowanej metody jest przedstawienie równania Kleina-Gordona (1.1) w postaci niejednorodnego układu równań różniczkowych zwyczajnych, a następnie zastosowanie do niego metody uzmienniania stałej (tzw. reguły Duhamela). W ten sposób otrzymaliśmy metodę rzędu trzeciego, którego stała błędu nie rośnie wraz ze wzrostem  $\omega_n$ . W Rozdziale 3, oprócz zaproponowania metody, przedstawimy szczegółowy dowód zbieżności i jej globalnego (a nie lokalnego) rzędu dokładności oraz omówimy strukturę błędu metody i wyjaśnimy, dlaczego nie mają na nią wpływu możliwie ekstremalnie wysokie oscylacje. Ponadto przedstawimy symulacje numeryczne porównując uzyskaną metodę ze standardowymi schematami typu Runge-Kutta 2-go i 4-go rzędu oraz z metodami z prac [Bader et al., 2019] oraz [Condon et al., 2021].

Dokładność metody z Rozdziału 3 poprawiona jest w pracy *Effective highly accurate time integrators for linear Klein–Gordon equations across the scales* [Kropielnicka et al., 2021], napisanej we współpracy z Karoliną Kropielniczą oraz Kathariną Schratz<sup>6</sup> i opisaney w Rozdziale 4. Zaproponowany schemat wykorzystuje rozwinięcie Magnusa oraz różne rodzaje dekompozycji pola wektorowego (tzw. splittingi). W przeciwieństwie do pracy [Kropielnicka and Lademann, 2022] nie udowadniamy globalnego rzędu zbieżności, ale opieramy się na zbieżności wykorzystywanych narzędzi i koncentrujemy się na wyznaczeniu dominującego czynnika błędu, który odpowiada lokalnemu błędowi przybliżenia. Jak się okaże, w tym przypadku nie warto mówić o zbieżności metody ale o jej dokładności. Pokażemy, że lokalna dokładność metody opisana jest funkcjami

$$\mathcal{O}(h^5 + \min\{h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \neq 0$$

oraz

$$\mathcal{O}(\min\{h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \equiv 0,$$

które zależą od relacji kroku czasowego  $h$ ,  $\omega_{\min}$  i  $\omega_{\max}$ . Okazuje się, że pesymistyczna dokładność metody [Kropielnicka et al., 2021] odpowiada optymistycznej dokładności schematu z [Kropielnicka and Lademann, 2022]. Rozdział 4 zakończymy graficzną interpretacją otrzymanych wyników oraz licznymi symulacjami numerycznymi uwzględniającymi porównania wydajności i efektywności wszystkich przedstawionych metod.

W rozdziale 5 przedstawimy syntetyczne podsumowanie otrzymanych wyników z krótkim opisem metod i konkluzjami dotyczącymi ich wydajności. Opiszemy też możliwości rozszerzenia otrzymanych wyników do innych typów równania Kleina-Gordona.

<sup>3</sup>School of Electronic Engineering, Glasnevin, Dublin 9, Irlandia

<sup>4</sup>Instytut Matematyki, Polska Akademia Nauk, Sopot

<sup>5</sup>Instytut Matematyki, Fizyki i Informatyki, Uniwersytet Gdański, Gdańsk

<sup>6</sup>Laboratoire Jacques-Louis Lions, Sorbonne Université, 4 place Jussieu, 75252 Paris, France

W związku z tym, że oscylacje występują tylko w zmiennej czasowej, w rozprawie koncentrujemy się na całkowaniu numerycznym względem czasu, tzw. *time integration*. Semi-dyskretyzacja (dyskretyzacja przestrzeni) zależy od warunków brzegowych zagadnienia i preferencji Użytkownika. Proponowane przez nas metody mogą być stosowane przy każdej semi-dyskretyzacji. Dla potrzeb symulacji numerycznych prezentowanych w rozprawie przyjmujemy, że równanie Kleina-Gordona (1.1) określone jest na torusie i przyjmujemy okresowe warunki początkowo-brzegowe. Pozwala to na wykorzystanie spektralnej kolokacji Fouriera względem zmiennej przestrzennej, [Kopriva, 2009, Trefethen, 2000].

Wszystkie eksperymenty numeryczne zaprezentowane w pracy zostały wykonane samodzielnie przez mnie w ośrodku komputerowym Trójmiejskiej Akademickiej Sieci Komputerowej (TASK) w Gdańsku, na mocy Grantu Obliczeniowego uzyskanego dnia 30 marca 2020 roku.

## Rozdział 2

# Metoda numeryczna oparta na rozwinięciu asymptotycznym

Rozdział ten oparty jest na opublikowanej w Applied Mathematics Letters pracy *Asymptotic numerical solver for the linear Klein–Gordon equation with space- and time-dependent mass* [Condon et al., 2021] napisanej we współpracy z Marissą Condon<sup>1</sup>, Karoliną Kropielnicką<sup>2</sup> oraz Rafałem Perczyńskim<sup>3</sup>.

Zaprezentujemy konstrukcję metody numerycznej opartej na rozwinięciu asymptotycznym w postaci zmodyfikowanego szeregu Fouriera. Numeryczno-asymptotyczne metody rozwiązywania zagadnień z wysoko oscylującym komponentem zewnętrznym  $f(\mathbf{x}, t)$ , znalazły zastosowanie w zagadnieniach inżynierskich [Condon et al., 2009], dla zwykłych równań różniczkowych z opóźnieniem czasowym [Condon et al., 2012] oraz dla niektórych typów równań falowych [Condon et al., 2019], ale według naszej wiedzy nigdy nie były badane w kontekście równań mechaniki kwantowej, gdzie wysokie oscylacje są charakterystyczną cechą układów, w szczególności tych modelowanych przez równanie Kleina-Gordona.

W tym Rozdziale przedstawiamy funkcję  $f(\mathbf{x}, t)$  w postaci

$$f(\mathbf{x}, t) = \sum_{n=-\infty}^{\infty} a_n(\mathbf{x}, t) e^{in\omega t}, \quad (2.1)$$

gdzie  $\omega \gg 1$  jest częstotliwością oscylacji. Postać ta jest szczególnym przypadkiem (1.2), gdzie  $n\omega = \omega_n$  oraz  $\alpha(\mathbf{x}, t) = a_0(\mathbf{x}, t) e^{i0\omega t}$ . Przyjmujemy tu, że rozwiązanie równania Kleina-Gordona może być przedstawione w postaci nieskończonego szeregu, którego rozwinięcie zależy od odwrotności oscylacji  $\omega$ . Wynika z tego, że im większa jest rozważana częstotliwość  $\omega$  tym szybciej kolejne składniki szeregu będą dążyły do zera. Dokładne wyprowadzenie metody znajduje się w podrozdziale 2.1. Oczywiście jest, że w praktycznych zastosowaniach nie możemy rozważać szeregu nieskończonego - będziemy stosowali skończone rozwinięcie, czyli  $|n| < N$ ,  $N \in \mathbb{N}$ .

### 2.1 Wyprowadzenie metody asymptotycznej

Równanie (1.1) zapisujemy w formie układu równań różniczkowych pierwszego rzędu

$$\begin{bmatrix} \partial_t \psi(\mathbf{x}, t) \\ \partial_t \varphi(\mathbf{x}, t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \Delta - f(\mathbf{x}, t) & 0 \end{bmatrix} \begin{bmatrix} \psi(\mathbf{x}, t) \\ \varphi(\mathbf{x}, t) \end{bmatrix}, \quad (2.2)$$

gdzie  $\varphi(\mathbf{x}, t) = \partial_t \psi(\mathbf{x}, t)$  oraz  $\psi(\mathbf{x}, 0) = \psi_0$ ,  $\varphi(\mathbf{x}, 0) = \varphi_0$  pochodzą z warunków początkowych równania (1.1).

<sup>1</sup>School of Electronic Engineering, Glasnevin, Dublin 9, Irlandia

<sup>2</sup>Instytut Matematyki, Polska Akademia Nauk, Sopot

<sup>3</sup>Instytut Matematyki, Fizyki i Informatyki, Uniwersytet Gdański, Gdańsk

Zakładamy, że istnieją funkcje:  $p_{0,0}(\mathbf{x},t)$ ,  $p_{1,0}(\mathbf{x},t)$  oraz  $p_{r,n}(\mathbf{x},t)$   $r \geq 2$ ,  $n \in \mathbb{Z}$ ,  $q_{0,0}(\mathbf{x},t)$  oraz  $q_{r,n}(\mathbf{x},t)$ ,  $r \geq 1$ ,  $n \in \mathbb{Z}$  takie, że układ  $[\psi(\mathbf{x},t), \varphi(\mathbf{x},t)]^T$

$$\psi(\mathbf{x},t) = p_{0,0}(\mathbf{x},t) + \frac{1}{\omega} p_{1,0}(\mathbf{x},t) + \sum_{r=2}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} p_{r,n}(\mathbf{x},t) e^{in\omega t}, \quad (2.3)$$

$$\varphi(\mathbf{x},t) = q_{0,0}(\mathbf{x},t) + \sum_{r=1}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} q_{r,n}(\mathbf{x},t) e^{in\omega t} \quad (2.4)$$

rozwiązuje (2.2). Rozwiązanie równania (1.1) jest określone przez funkcję  $\psi(\mathbf{x},t)$ .

Zadanie polega teraz na skutecznym (czyli przy niskim koszcie obliczeniowym) poszukiwaniu współczynników powyższego rozwiązania. Aby to zrobić, podstawiamy równania (2.3 - 2.4) do zagadnienia (2.2) i uzyskujemy następujące relacje

$$\begin{aligned} \partial_t p_{0,0}(\mathbf{x},t) + \frac{1}{\omega} \left[ \partial_t p_{1,0}(\mathbf{x},t) + \sum_{n=-\infty}^{\infty} in p_{2,n}(\mathbf{x},t) e^{in\omega t} \right] + \sum_{r=2}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} \left[ \partial_t p_{r,n}(\mathbf{x},t) + in p_{r+1,n}(\mathbf{x},t) \right] e^{in\omega t} \\ = q_{0,0}(\mathbf{x},t) + \sum_{r=1}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} q_{r,n}(\mathbf{x},t) e^{in\omega t}, \end{aligned} \quad (2.5)$$

$$\begin{aligned} \partial_t q_{0,0}(\mathbf{x},t) + \sum_{n=-\infty}^{\infty} in q_{1,n}(\mathbf{x},t) e^{in\omega t} + \sum_{r=1}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} \left[ \partial_t q_{r,n}(\mathbf{x},t) + in q_{r+1,n}(\mathbf{x},t) \right] e^{in\omega t} \\ = \Delta p_{0,0}(\mathbf{x},t) + \frac{1}{\omega} \Delta p_{1,0}(\mathbf{x},t) + \sum_{r=2}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} \Delta p_{r,n}(\mathbf{x},t) e^{in\omega t} \\ - \sum_{n=-\infty}^{\infty} a_n(\mathbf{x},t) e^{in\omega t} p_{0,0}(\mathbf{x},t) - \frac{1}{\omega} \sum_{n=-\infty}^{\infty} a_n(\mathbf{x},t) e^{in\omega t} p_{1,0}(\mathbf{x},t) \\ - \sum_{r=2}^{\infty} \frac{1}{\omega^r} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_k(\mathbf{x},t) p_{r,n-k}(\mathbf{x},t) e^{in\omega t}. \end{aligned} \quad (2.6)$$

Przyglądając się uważnie równaniom (2.5) oraz (2.6) zauważamy, że relacje opisujące współczynniki  $p_{r,n}$  oraz  $q_{r,n}$  można wyprowadzić poprzez przyrównywanie do siebie wyrazów (2.5) oraz (2.6) z tymi samymi wielkościami ( $\omega^{-r}$ ) i częstotliwościami ( $e^{in\omega t}$ ). Porównania te prowadzą do równań różniczkowych i rekurencyjnych, które stanowią podstawę metody obliczeniowej. Przedstawimy je w Tabelicy 2.1.

$r, n$	$n = 0$	$n \neq 0$
$r = 0$	$\begin{cases} \partial_t p_{0,0}(\mathbf{x},t) = q_{0,0}(\mathbf{x},t), \\ \partial_t q_{0,0}(\mathbf{x},t) = \Delta p_{0,0}(\mathbf{x},t) - a_0(\mathbf{x},t) p_{0,0}(\mathbf{x},t), \end{cases}$ $p_{0,0}(\mathbf{x},0) = \psi_0(\mathbf{x}), q_{0,0}(\mathbf{x},0) = \varphi_0(\mathbf{x}), x \in L,$ $p_{0,0}(\mathbf{x},t) = \text{const}, q_{0,0}(\mathbf{x},t) = \text{const}, t \geq 0, x \in \partial L,$	$q_{1,n}(\mathbf{x},t) = \frac{i}{n} a_n(\mathbf{x},t) p_{0,0}(\mathbf{x},t).$
$r = 1$	$\begin{cases} \partial_t p_{1,0}(\mathbf{x},t) = q_{1,0}(\mathbf{x},t), \\ \partial_t q_{1,0}(\mathbf{x},t) = \Delta p_{1,0}(\mathbf{x},t) - a_0(\mathbf{x},t) p_{1,0}(\mathbf{x},t), \end{cases}$ $p_{1,0}(\mathbf{x},0) = 0, q_{1,0}(\mathbf{x},0) = -\sum_{n \neq 0} q_{1,n}(\mathbf{x},0), x \in L,$ $p_{1,0}(\mathbf{x},t) = \text{const}, q_{1,0}(\mathbf{x},t) = \text{const}, t \geq 0, x \in \partial L,$	$q_{2,n}(\mathbf{x},t) = \frac{i}{n} \left( a_n(\mathbf{x},t) p_{1,0}(\mathbf{x},t) + \partial_t q_{1,n}(\mathbf{x},t) \right),$ $p_{2,n}(\mathbf{x},t) = -\frac{i}{n} q_{1,n}(\mathbf{x},t).$
$r \geq 2$	$\begin{cases} \partial_t p_{r,0}(\mathbf{x},t) = q_{r,0}(\mathbf{x},t), \\ \partial_t q_{r,0}(\mathbf{x},t) = \Delta p_{r,0}(\mathbf{x},t) - a_0(\mathbf{x},t) p_{r,0}(\mathbf{x},t) \\ \quad - \sum_{k \neq 0} a_k(\mathbf{x},t) p_{r,-k}(\mathbf{x},t), \end{cases}$ $p_{r,0}(\mathbf{x},0) = -\sum_{n \neq 0} p_{r,n}(\mathbf{x},0), x \in L,$ $q_{r,0}(\mathbf{x},0) = -\sum_{n \neq 0} q_{r,n}(\mathbf{x},0), x \in L,$ $p_{r,0}(\mathbf{x},t) = \text{const}, q_{r,0}(\mathbf{x},t) = \text{const}, t \geq 0, x \in \partial L.$	$q_{r+1,n}(\mathbf{x},t) = \frac{i}{n} \left( \partial_t q_{r,n}(\mathbf{x},t) - \Delta p_{r,n}(\mathbf{x},t) \right. \\ \quad \left. + \sum_{k=-\infty}^{\infty} a_k(\mathbf{x},t) p_{r,n-k}(\mathbf{x},t) \right),$ $p_{r+1,n}(\mathbf{x},t) = \frac{i}{n} (\partial_t p_{r,n}(\mathbf{x},t) - q_{r,n}(\mathbf{x},t)).$

TABELICA 2.1: Równania różniczkowe cząstkowe i rekurencyjne na współczynniki  $p_{r,n}$  oraz  $q_{r,n}$  rozwinięcia (2.3 - 2.4).

Wyznaczenie współczynników  $p_{r,n}$  oraz  $q_{r,n}$  przy użyciu relacji z powyższej Tabelicy pozwoli nam wyznaczyć rozwiązania (2.3 - 2.4), dzięki czemu otrzymamy asymptotyczne przybliżenie dla rozwiązania równania Kleina-Gordona (1.1).



**Uwaga 1** Oczywistym jest, że w praktycznych zastosowaniach obliczamy rozwiązania (2.3 - 2.4) dla skończonej ilości wyrazów rozwinięcia, mianowicie

$$\begin{aligned}\psi(\mathbf{x}, t) &= p_{0,0}(\mathbf{x}, t) + \frac{1}{\omega} p_{1,0}(\mathbf{x}, t) + \sum_{r=2}^R \frac{1}{\omega^r} \sum_{n=-N}^N p_{r,n}(\mathbf{x}, t) e^{in\omega t}, \\ \varphi(\mathbf{x}, t) &= q_{0,0}(\mathbf{x}, t) + \sum_{r=1}^R \frac{1}{\omega^r} \sum_{n=-N}^N q_{r,n}(\mathbf{x}, t) e^{in\omega t},\end{aligned}\tag{2.7}$$

gdzie  $R \geq 2$  oraz  $N \geq 1$  są ustalone.

Warto zwrócić uwagę, że dokładność przybliżenia asymptotycznego jest odwrotnie proporcjonalna do wielkości oscylacji  $\omega$ . Oznacza to, że im wyżej oscylująca jest funkcja  $f(\mathbf{x}, t)$ , tym lepsze otrzymujemy przybliżenie rozwiązania (1.1) dla danego  $R$ . Jednocześnie, można łatwo zauważyć, że im większa jest częstotliwość  $\omega$ , tym mniejsza wartość  $R$  (czyli krótsze rozwinięcie) wystarczy, aby osiągnąć określoną dokładność przybliżenia.

**Uwaga 2** Funkcja  $f(\mathbf{x}, t)$  może być zadana od razu jako nieskończony szereg (2.1) lub w postaci dowolnej funkcji dwóch zmiennych  $\mathbf{x}$  i  $t$ .

W pierwszym przypadku oczywistym jest, że w praktycznych zastosowaniach będziemy stosowali skończone obcięcie takiego szeregu do pewnej liczby  $N$ . Zatem to od Użytkownika zależy, jak dużo wyrazów z nieskończonego rozwinięcia zostanie wykorzystanych w metodzie numerycznej. Wybór liczby  $N$  będzie miał wpływ na dokładność metody asymptotycznej.

W drugim przypadku, to regularność funkcji  $f(\mathbf{x}, t)$  będzie miała wpływ na liczbę  $N$  w równaniu (2.7). Im wyższa regularność funkcji  $f(\mathbf{x}, t)$  tym szybsza jest zbieżność do zera współczynników  $a_n(\mathbf{x}, t)$  w zmodyfikowanym szeregu Fouriera, [Kumar et al., 2014]. Oznacza to, że jeśli  $f(\mathbf{x}, t)$  jest wystarczająco regularna, można ją dobrze przybliżyć za pomocą skończonej sumy

$$f_N(\mathbf{x}, t) = \sum_{n=-N}^{n=N} a_n(\mathbf{x}, t) e^{in\omega t}$$

przy stosunkowo małej liczbie  $N$ .

**Uwaga 3** Równania różniczkowe cząstkowe dla  $p_{r,0}$  (lewa strona Tablicy 2.1), jeśli nie mogą być rozwiązane analitycznie, to mogą być dość łatwo obliczone numerycznie dowolną preferowaną metodą obliczeniową. Zauważmy, że w przeciwieństwie do (1.1), te równania różniczkowe cząstkowe nie zależą od częstotliwości  $\omega$ , zatem nie są wysoko oscylujące i dlatego do ich obliczenia nie są wymagane bardzo małe kroki czasowe.

**Uwaga 4** Biorąc pod uwagę skończone przybliżenie  $f_N(\mathbf{x}, t)$  funkcji  $f(\mathbf{x}, t)$  oraz układ (2.7) możemy zaobserwować, że jeżeli

$$a_n(\mathbf{x}, t) = a_{-n}(\mathbf{x}, t)$$

to otrzymujemy zależności

$$p_{r,n}(\mathbf{x}, t) = \begin{cases} p_{r,-n}(\mathbf{x}, t), & \text{jeżeli } r \text{ jest parzyste, } r \geq 2, \\ -p_{r,-n}(\mathbf{x}, t) & \text{jeżeli } r \text{ jest nieparzyste, } r \geq 2. \end{cases}$$

Analogiczne zależności zachodzą dla współczynników  $q_{r,n}$ . Oznacza to, że  $p_{2r+1,0}(\mathbf{x}, t) = 0$  oraz  $q_{2r+1,0}(\mathbf{x}, t) = 0$  dla  $r \geq 0$ . Zatem, wystarczy obliczyć tylko co drugie równanie różniczkowe cząstkowe dla  $p_{r,0}(\mathbf{x}, t)$  i  $q_{r,0}(\mathbf{x}, t)$ , a w relacji rekurencyjnej obliczyć jedynie współczynniki  $p_{r,n}(\mathbf{x}, t)$  i  $q_{r,n}(\mathbf{x}, t)$  dla  $n > 0$ . Wszystko to prowadzi do znacznego skrócenia czasu obliczeniowego.

Z powyższych uwag wynika, że na dokładność zaprezentowanej metody asymptotycznej składa się kilka czynników:

- wielkość oscylacji  $\omega$ , co wiąże się bezpośrednio z wielkością liczby  $R$  elementów szeregu w (2.7);
- liczby  $N$ , która zależy ilości funkcji  $a_n(\mathbf{x}, t) e^{in\omega t}$  w zmodyfikowanym szeregu Fouriera (2.1);
- rzędu metody numerycznej zastosowanej do przybliżenia współczynników  $p_{r,0}$  oraz  $q_{r,0}$ .

Wpływ każdego z powyższych czynników zaprezentujemy i omówimy na przykładach numerycznych.

## 2.2 Przykłady numeryczne

Aby zilustrować proponowane podejście, rozważymy równanie (1.1) dla dwóch różnych postaci funkcji  $f(\mathbf{x}, t)$ . Zaprezentujemy

- porównania proponowanej metody z innymi znanymi metodami (Rysunki 2.1 oraz 2.2),
- błąd metody asymptotycznej w zależności od rzędu metody zastosowanej do przybliżenia współczynników  $p_{r,0}$  oraz  $q_{r,0}$  (Rysunek 2.1),
- koszt obliczeniowy metody asymptotycznej przy różnych wyborach  $R$ , czyli długości zmodyfikowanego rozwinięcia (Rysunek 2.2),
- wpływ  $N$ , czyli długość zmodyfikowanego szeregu Fouriera przybliżającego  $f(\mathbf{x}, t)$  na błąd metody asymptotycznej (Rysunek 2.3).

**Przykład 1** Rozważamy funkcję Kleina- Gordona (1.1) z funkcją masy

$$f_1(\mathbf{x}, t) = x^2 + \varepsilon \cos(\omega t)x^2,$$

warunkami początkowymi

$$\psi(\mathbf{x}, 0) = e^{-\frac{x^2}{2}}, \quad \partial_t \psi(\mathbf{x}, 0) = 0$$

oraz zerowymi warunkami brzegowymi. Wówczas, rozwinięcie (2.1) funkcji  $f_1(\mathbf{x}, t)$  zawiera tylko trzy niezerowe współczynniki:

$$a_0(\mathbf{x}) = x^2, \quad \text{oraz} \quad a_1(\mathbf{x}) = a_{-1}(\mathbf{x}) = \frac{\varepsilon}{2}x^2.$$

Aby przybliżyć rozwiązanie takiego zagadnienia, odwołujemy się do Tablicy 2.1 i rozwiązujemy powstające tam rekurencyjne i cząstkowe równania różniczkowe. Podczas gdy w tym konkretnym przypadku równania różniczkowe cząstkowe dla  $p_{i,0}$ ,  $q_{i,0}$ ,  $i = 0, 1, 3$ , można rozwiązać analitycznie, to pozostałe równania należy rozwiązać numerycznie dowolną preferowaną metodą. Ważną obserwacją jest, że otrzymane równania różniczkowe nie mają natury oscylacyjnej, przez co nie mamy relacji między krokiem czasowym  $h$  i oscylacjami  $\omega$ .

W naszym przypadku równania, które nie mają analitycznego rozwiązania, można zapisać w postaci macierzowej

$$\partial_t \begin{bmatrix} p_{i,0}(\mathbf{x}, t) \\ q_{i,0}(\mathbf{x}, t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \Delta - a_0(\mathbf{x}) & 0 \end{bmatrix} \begin{bmatrix} p_{i,0}(\mathbf{x}, t) \\ q_{i,0}(\mathbf{x}, t) \end{bmatrix} + \begin{bmatrix} 0 \\ s_i(\mathbf{x}, t) \end{bmatrix}, \quad (2.8)$$

gdzie

$$s_i(\mathbf{x}, t) = \sum_{k \neq 0} a_k(\mathbf{x}, t) p_{i,-k}(\mathbf{x}, t).$$

Rozwiązanie układu (2.8) można przedstawić za pomocą formuły Duhamela

$$\begin{bmatrix} p_{i,0}(\mathbf{x}, t) \\ q_{i,0}(\mathbf{x}, t) \end{bmatrix} = e^{At} \begin{bmatrix} p_{i,0}(\mathbf{x}, 0) \\ q_{i,0}(\mathbf{x}, 0) \end{bmatrix} + \int_0^t e^{A(t-\tau)} \begin{bmatrix} 0 \\ s_i(\mathbf{x}, \tau) \end{bmatrix} d\tau, \quad (2.9)$$

gdzie  $A = \begin{bmatrix} 0 & 1 \\ \Delta - a_0(\mathbf{x}) & 0 \end{bmatrix}$ .

Załóżmy, że czas  $t \in [0, T]$ ,  $T/h = K$  i  $t_k = kh$ , gdzie  $h$  jest wielkością kroku czasowego. Iterując powyższy wzór otrzymujemy wzór na funkcje  $p_{i,0}$  oraz  $q_{i,0}$  w kroku czasowym  $t_{k+1} = t_k + h$ ,  $k = 0, \dots, K-1$ :

$$\begin{aligned} \begin{bmatrix} p_{i,0}(\mathbf{x}, t_{k+1}) \\ q_{i,0}(\mathbf{x}, t_{k+1}) \end{bmatrix} &= e^{A(t_{k+1}-t_k)} \begin{bmatrix} p_{i,0}(\mathbf{x}, t_k) \\ q_{i,0}(\mathbf{x}, t_k) \end{bmatrix} + \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau)} \begin{bmatrix} 0 \\ s_i(\mathbf{x}, \tau) \end{bmatrix} d\tau \\ &= e^{Ah} \begin{bmatrix} p_{i,0}(\mathbf{x}, t_k) \\ q_{i,0}(\mathbf{x}, t_k) \end{bmatrix} + \int_0^h e^{A(h-\tau)} \begin{bmatrix} 0 \\ s_i(\mathbf{x}, t_k + \tau) \end{bmatrix} d\tau. \end{aligned} \quad (2.10)$$

Jak napisaliśmy we wstępie, w naszych przykładach do semi-dyskretyzacji wykorzystujemy spektralną metodę Fouriera. Wówczas operator  $\Delta$  zostanie przybliżony macierzą toeplizowską, a funkcja  $a_0(\mathbf{x})$  będzie macierzą diagonalną. Warto zauważyć, że funkcja eksponencjalna macierzy toeplizowskiej obliczona jest Szybką Transformacją Fouriera (FFT), podczas gdy funkcja eksponencjalna macierzy diagonalnej otrzymywana jest natychmiastowo. Funkcje wykładnicze macierzy, które nie są ani toeplizowskie ani diagonalne przybliżane są numerycznie zazwyczaj za pomocą rozwinięcia Taylora dla macierzy, które wymaga kosztownych (czasowo i pamięciowo) obliczeń  $n$ -tych potęg dużych (w sensie spektralnym) macierzy.

Macierz  $Ah$  jest macierzą antydiagonalną postaci  $\begin{bmatrix} 0 & h \\ Qh & 0 \end{bmatrix}$ , gdzie  $Qh = h\Delta - ha_0(\mathbf{x})$ . Eksponent takiej macierzy możemy obliczyć korzystając z definicji eksponentu macierzowego, mianowicie

$$\exp\left(\begin{bmatrix} 0 & h \\ Qh & 0 \end{bmatrix}\right) = \begin{bmatrix} \cos(h\sqrt{Q}) & h\operatorname{sinc}(h\sqrt{Q}) \\ -\sqrt{Q}\sin(h\sqrt{Q}) & \cos(h\sqrt{Q}) \end{bmatrix},$$

Jednakże, sinus oraz cosinus z wyrażenia  $h\sqrt{Q} = h\sqrt{\Delta - a_0(\mathbf{x})}$  jest skomplikowany obliczeniowo, ponieważ nie jest ono ani macierzą diagonalną, ani cykliczną symetryczną.

Z tego powodu w prezentowanych przykładach do przybliżenia eksponentu macierzy  $Ah$

$$e^{Ah} = \exp\left(\underbrace{\begin{bmatrix} 0 & 1 \\ -a_0(\mathbf{x}) & 0 \end{bmatrix}}_B h + \underbrace{\begin{bmatrix} 0 & 0 \\ \Delta & 0 \end{bmatrix}}_C h\right) \quad (2.11)$$

korzystamy z trzech rodzajów splittingów:

- rzędu pierwszego, postaci  $e^{hA} \sim e^{hB}e^{hC}$ , tzw. *Lie-Trotter splitting*;
- rzędu drugiego, postaci  $e^{hA} \sim e^{h/2C}e^{hB}e^{h/2C}$ , tzw. *Strang splitting*;
- rzędu trzeciego, postaci  $e^{hA} \sim e^{hC}e^{-h/24B}e^{-2h/3C}e^{3h/4B}e^{2h/3B}e^{7h/24C}$ , tzw. *Ruth splitting*.

Szczegółowe wyprowadzenie oraz analizę powyższych splittingów można znaleźć w [Strang, 1968], [McLachlan and Quispel, 2002, Ruth, 1983]. Ze względu na swoją strukturę, eksponenty  $e^{hB}$  oraz  $e^{hC}$  oblicza się w prosty sposób, korzystając z definicji eksponentu macierzowego, mianowicie

$$e^{hB} = \begin{bmatrix} \cos(h\sqrt{a_0(\mathbf{x})}) & h\operatorname{sinc}(h\sqrt{a_0(\mathbf{x})}) \\ -\sqrt{a_0(\mathbf{x})}\sin(h\sqrt{a_0(\mathbf{x})}) & \cos(h\sqrt{a_0(\mathbf{x})}) \end{bmatrix} \quad \text{oraz} \quad e^{hC} = \begin{bmatrix} 1 & 0 \\ h\Delta & 1 \end{bmatrix}, \quad (2.12)$$

gdzie

$$\operatorname{sinc}(x) = \begin{cases} \frac{\sin(x)}{x} & x \neq 0 \\ 1 & x = 0. \end{cases}$$

Do aproksymacji całki w (2.9) wykorzystujemy kwadraturę Gaussa-Legendre'a szóstego rzędu, z trzema węzłami.

Zaproponowaną w pracy asymptotyczną metodę numeryczną porównujemy z następującymi znanymi schematami:

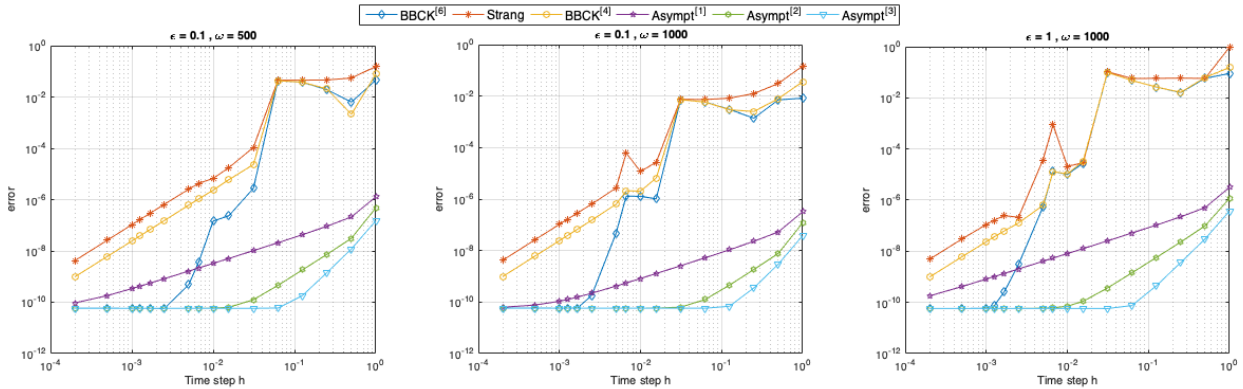
**BBCK** <sup>[4]</sup> - metoda czwartego rzędu  $\Sigma_{3c}^{[4]}$  z [Bader et al., 2019];

**BBCK** <sup>[6]</sup> - metoda szóstego rzędu  $\Sigma_{5c}^{[6]}$  z [Bader et al., 2019];

**Strang** - splitting Stranga [Strang, 1968] zastosowany bezpośrednio do zagadnienia (2.2).

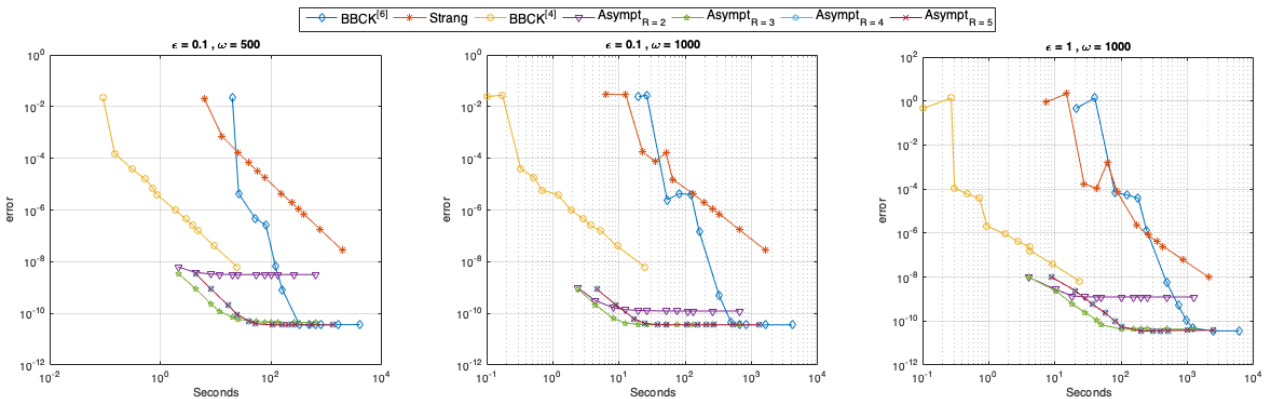
Rozwiązanie referencyjne obliczane jest przy użyciu metody 6-tego rzędu z prac [Blanes et al., 2009] z krokiem czasowym  $h = 10^{-6}$ , który jest znacznie mniejszy niż największa z rozważanych w przykładach oscylacja. Następnie stosujemy podane metody z różnymi krokami czasowymi i mierzymy błąd w końcowym czasie  $T = 1$  (Rysunek 2.1) lub  $T = 10$  (Rysunek 2.2 oraz 2.3) w normie  $\ell_2$ . Laplacian jest dyskretyzowany metodą Fouriera, [Kopriva, 2009, Trefethen, 2000] z  $M = 100$  modów.

Rysunek 2.1 przedstawia porównania dla trzech różnych par parametrów  $\omega$  i  $\epsilon$ . To porównanie przedstawia, jak zmienia się rząd metody asymptotycznej, gdy zmieniamy rząd splittingu macierzy (2.11). Wykresy *error vs time step* pokazują rząd metod, gdzie  $\text{Asympt}^{[1]}$ ,  $\text{Asympt}^{[2]}$  oraz  $\text{Asympt}^{[3]}$  opisują metody wykorzystujące odpowiednio splitting pierwszego, drugiego i trzeciego rzędu.



RYСУNEK 2.1: Porównanie rzędów metod w przedziale czasu  $[0,1]$ . Metody asymptotyczne  $\text{Asympt}^{[1]}$ ,  $\text{Asympt}^{[2]}$  and  $\text{Asympt}^{[3]}$  uzyskujemy wykorzystując  $N = 1$  oraz  $R = 5$  w (2.7), gdzie współczynniki  $p_{r,0}, q_{r,0}$  są obliczane zgodnie ze wzorem (2.9) używając splittingów 1-go, 2-go oraz 3-go rzędu dla (2.11), odpowiednio.

Rysunek 2.2 przedstawia niski koszt obliczeniowy metod asymptotycznych  $\text{Asympt}_{R=r}$ ,  $r = 2, 3, 4, 5$ , które otrzymywane są poprzez obcięcie rozwiązania (2.7) dla różnych wartości  $R$ . Łatwo zauważyć, że wszystkie metody asymptotyczne charakteryzują się bardzo dużą dokładnością przy znacznie niższym koszcie obliczeniowym, zwłaszcza w porównaniu z metodą szóstego rzędu. Jest to szczególnie ważne w kontekście dużych  $\omega$ . Zgodnie z Uwagą 1, im wyższe oscylacje funkcji  $f(x, t)$ , tym skuteczniejsze są metody asymptotyczne.



RYСУNEK 2.2: Porównanie kosztu obliczeniowego metod na przedziale czasu  $[0,10]$ . Metody asymptotyczne  $\text{Asympt}_{R=r}$ ,  $r = 2, 3, 4, 5$  uzyskujemy biorąc  $N = 1$  oraz  $R = r$  w (2.7). Do aproksymacji eksponentu (2.11) wykorzystujemy splitting drugiego rzędu.

**Przykład 2** Rozważamy funkcję Kliena-Gordona (1.1) z funkcją masy

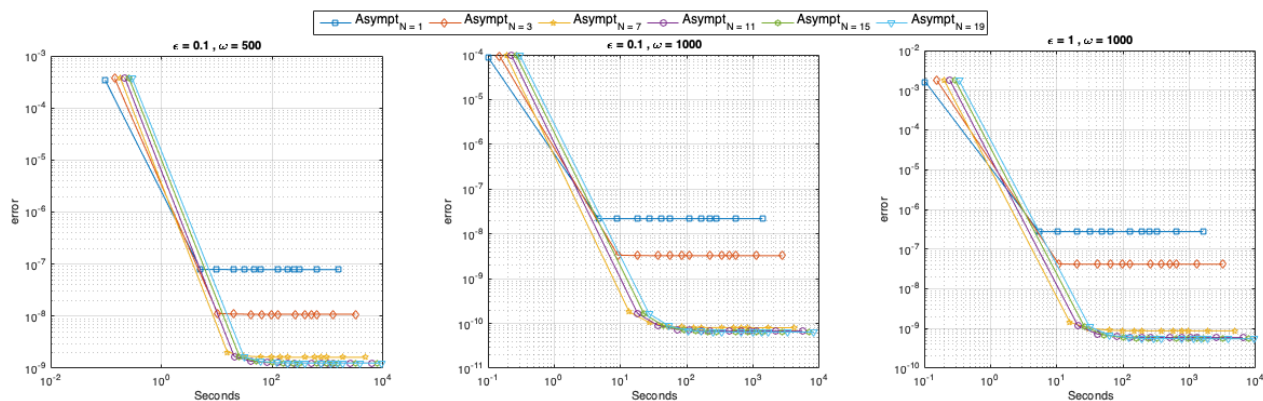
$$f_2(\mathbf{x}, t) = x^2 + \epsilon \cos^{19}(\omega t),$$

oraz warunkami początkowymi

$$\psi(\mathbf{x}, 0) = e^{-\frac{x^2}{2}}, \quad \partial_t \psi(\mathbf{x}, 0) = 0.$$

Przykład ten jest dobrany specjalnie w celu pokazania zależności metody asymptotycznej od obciążenia (2.7) względem  $N$  ponieważ rozwinięcie (2.1) funkcji  $f_2(\mathbf{x}, t)$  ma dwadzieścia jeden niezerowych współczynników, zatem  $N = 19$ .

Na Rysunku 2.3 przedstawiamy porównanie kosztu oraz dokładności metody dla wybranych  $N$ . Funkcje  $\text{Asymp}_{N=n}$ , gdzie  $n = 1, 3, 7, 11, 15, 19$  są obliczane z  $R = 3$  w (2.7) i przy użyciu splittingu drugiego rzędu dla współczynników  $p_{r,0}, q_{r,0}$ .



RYSUNEK 2.3: Porównanie kosztu oraz dokładności metod asymptotycznych w przedziale czasu  $[0,10]$ . Metody  $\text{Asymp}_{N=n}$ , gdzie  $n = 1, 3, 7, 11, 15, 19$  otrzymujemy biorąc  $N = n$  oraz  $R = 3$  w (2.7) i przy użyciu aproksymacji drugiego rzędu dla współczynników  $p_{r,0}, q_{r,0}$ .

Jak łatwo zauważyć metoda asymptotyczna wykazuje się bardzo dużą dokładności już przy zastosowaniu najmniejszej z możliwych ilości składników  $N$ , czyli dla  $N = 1$ . Ponadto, stosowanie rozwinięcia dłuższego niż  $N = 7$  nie poprawia dokładności metody.

Wszystkie powyższe symulacje numeryczne mają na celu pokazanie, że nie tylko w teorii ale także w praktyce, że zaproponowana metoda asymptotyczna jest efektywnym narzędziem do rozwiązywania równań typu (1.1) w przypadku, gdy funkcja  $f(x, t)$  jest wysoko oscylująca i gdy obliczenia mają być przeprowadzane w długim przedziale czasu i z dużymi krokami czasowymi  $h$ .



## Rozdział 3

# Metoda numeryczna oparta na zastosowaniu formuły Duhamela

Rozdział ten oparty jest na manuskrypcie *Third-order exponential integrator for linear Klein–Gordon equations with time and space-dependant mass* [Kropielnicka and Lademann, 2022] napisanym we współpracy z Karoliną Kropielnicką<sup>1</sup> i wstępnie przyjętym do Mathematical Modelling and Numerical Analysis.

Przypomnijmy, że zaprezentowane w pracy [Bader et al., 2019] metody są skutecznym narzędziem w przypadku nieoscylującej funkcji  $f(\mathbf{x}, t)$ . Z kolei metoda zaprezentowana w Rozdziale 2, jest bardzo efektywna w przypadku funkcji  $f(\mathbf{x}, t)$  o wyłącznie wysokich oscylacjach. Pozostaje jednak pewna luka, którą jest przypadek *mieszany* funkcji  $f(\mathbf{x}, t)$ , czyli gdy funkcja ta zależy od wielu różnych częstotliwości oscylacji, od małych do bardzo dużych, np.  $f(\mathbf{x}, t) = a_0(\mathbf{x}, t) + a_1(\mathbf{x}, t)e^{i10t} + a_2(\mathbf{x}, t)e^{i10^3t} + a_3(\mathbf{x}, t)e^{i10^9t}$ . Wówczas metody z pracy [Bader et al., 2019] zawodzą przez obecność dużych oscylacji, tj.  $\omega_{\max} = 10^9$ , a metoda asymptotyczna z Rozdziału 2 zawodzi przez obecność małych oscylacji, tj.  $\omega_{\min} = 0$ ,  $\omega_1 = 10$ .

W tym rozdziale przedstawimy metodę numeryczną rzędu trzeciego, która jest równie efektywna w stosunku do każdej z form funkcji  $f(\mathbf{x}, t)$ , tj. np. gdy  $f(\mathbf{x}, t) = a_0(\mathbf{x}, t) + a_1(\mathbf{x}, t)e^{i10t}$  lub  $f(\mathbf{x}, t) = a_3(\mathbf{x}, t)e^{i10^9t}$ , oraz w najtrudniejszym przypadku mieszanym, np. gdy  $f(\mathbf{x}, t) = a_0(\mathbf{x}, t) + a_1(\mathbf{x}, t)e^{i10t} + a_2(\mathbf{x}, t)e^{i10^3t} + a_3(\mathbf{x}, t)e^{i10^9t}$ .

Podstawą zaprezentowanej w tym Rozdziale metody jest przedstawienie równania Kleina - Gordona w postaci niejednorodnego układu równań różniczkowych zwyczajnych, a następnie zastosowaniu do tego układu metody niezmienniania stałej, znanej jako także reguła Duhamela. Pozwala to, między innymi, na wyodrębnienie całek z funkcji wysoko-oscylujących, które możemy policzyć analitycznie lub zastosować dedykowane im metody numeryczne, jak na przykład metodę Filona. Wówczas, w przeciwieństwie do metod z pracy [Bader et al., 2019] opartych na kwadraturach Gauss'a-Legendre, nie ma potrzeby wprowadzania zależności między wielkością kroku czasowego  $h$  oraz  $\omega_{\max}$ . Przedstawimy szczegółowy dowód zbieżności nowej metody, omówimy strukturę jej błędów i wyjaśnimy, dlaczego nie mają na nią wpływu możliwe bardzo wysokie oscylacje. Ostatnią część Rozdziału stanowią ilustracje symulacji numerycznych, na których przedstawimy dokładność nowej metody w zależności od zastosowanej metody aproksymacji całki z funkcji wysoko-oscylującej oraz porównania efektywności różnych istniejących metody z nowym podejściem i podkreślamy wyższość nowej metody w przypadku wysokich częstotliwości  $\omega_n$ .

### 3.1 Wyprowadzenie metody

Jak wspomnieliśmy we wstępie, zajmujemy się wyłącznie dyskretyzacją względem czasu. Z tego powodu w dalszej części tego Rozdziału pomijamy zależność od zmiennej przestrzennej  $\mathbf{x}$  i zapisujemy  $\psi(t), a_n(t), f(t)$  zamiast  $\psi(\mathbf{x}, t), a_n(\mathbf{x}, t), f(\mathbf{x}, t)$ . Wówczas możemy wyrazić (1.1) jako następujące abstrakcyjne równanie

$$\frac{d}{dt} \begin{bmatrix} \psi(t) \\ \psi'(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \Delta & 0 \end{bmatrix} \begin{bmatrix} \psi(t) \\ \psi'(t) \end{bmatrix} + \begin{bmatrix} 0 \\ f(t)\psi(t) \end{bmatrix} \quad (3.1)$$

<sup>1</sup>Institut Matematyki, Polska Akademia Nauk, Sopot

z rozwiązaniem danym za pomocą reguły Duhamela,

$$\begin{bmatrix} \psi(t) \\ \psi'(t) \end{bmatrix} = R(t - t_0) \begin{bmatrix} \psi_0 \\ \varphi_0 \end{bmatrix} + \int_{t_0}^t R(t - \tau) \begin{bmatrix} 0 \\ f(\tau)\psi(\tau) \end{bmatrix} d\tau, \quad (3.2)$$

gdzie

$$R(t) = \exp \begin{bmatrix} 0 & t \\ \Delta t & 0 \end{bmatrix} = \begin{bmatrix} \cos(tG) & G^{-1} \sin(tG) \\ -G \sin(tG) & \cos(tG) \end{bmatrix}, \quad \text{oraz} \quad G \cdot G = -\Delta.$$

Operator  $-\Delta$  jest dodatnio określony, zatem możemy przyjąć, że  $G$  jest operatorem dodatnio-określonym.

Założmy, że  $T/h = K$  i  $t_k = kh$ . Iterując (3.2) otrzymujemy wzór na rozwiązanie  $\psi$  oraz pochodną  $\psi'_t$  w kroku czasowym  $t_{k+1} = t_k + h$ ,  $k = 0, \dots, K - 1$ :

$$\begin{aligned} \begin{bmatrix} \psi(t_k + h) \\ \psi'(t_k + h) \end{bmatrix} &= R(t_k + h - t_k) \begin{bmatrix} \psi(t_k) \\ \psi'(t_k) \end{bmatrix} + \int_{t_k}^{t_k+h} R(t_k + h - \tau) \begin{bmatrix} 0 \\ f(\tau)\psi(\tau) \end{bmatrix} d\tau \\ &= R(h) \begin{bmatrix} \psi(t_k) \\ \psi'(t_k) \end{bmatrix} + \int_0^h R(h - \tau) \begin{bmatrix} 0 \\ f(t_k + \tau)\psi(t_k + \tau) \end{bmatrix} d\tau. \end{aligned} \quad (3.3)$$

Podobnie jak w przypadku metod typu Gautschiego [Hochbruck and Lubich, 1998], powyższa reprezentacja rozwiązania jest podstawą naszego schematu numerycznego.

**Uwaga 5** Przez  $\psi'(t_k)$  i  $\psi''(t_k)$  oznaczamy odpowiednio pierwszą i drugą pochodną funkcji  $\psi(\cdot, t)$  w punkcie  $t_k$ . W szczególności piszemy  $\psi'(t_0)$  zamiast  $\varphi_0$ .

Układ (3.3) możemy przedstawić równoważnie w postaci

$$\psi(t_k + h) = \cos(hG)\psi(t_k) + G^{-1} \sin(hG)\psi'(t_k) + \int_0^h G^{-1} \sin((h - \tau)G) f(t_k + \tau)\psi(t_k + \tau) d\tau, \quad (3.4)$$

$$\psi'(t_k + h) = -G \sin(hG)\psi(t_k) + \cos(hG)\psi'(t_k) + \int_0^h \cos((h - \tau)G) f(t_k + \tau)\psi(t_k + \tau) d\tau. \quad (3.5)$$

Główną trudnością układu (3.3), a co za tym idzie, także powyższego sformułowania jest to, że jest to niejawni układ równań, ponieważ nie znamy wartości  $\psi(t_k + \tau)$ . W celu przybliżenia wartości  $\psi(t_k + \tau)$  korzystamy z rozwinięcia Taylora

$$\psi(t_k + \tau) = \psi(t_k) + \tau\psi'(t_k) + \frac{\tau^2}{2!}\psi''(t_k) + E_k(\tau), \quad (3.6)$$

gdzie  $E_k(\tau) = \int_{t_k}^{t_k+\tau} \frac{(t_k + \tau - s)^2}{2!} \psi'''(s) ds$ .

**Uwaga 6** Wartości  $\psi''(t_k)$  można uzyskać bezpośrednio z warunku początkowego (dla  $k = 0$ ) i z zagadnienia (1.1) dla  $k \geq 1$ . To prowadziłoby jednak do dodatkowych kosztów obliczeniowych. Z tego powodu dla  $k \geq 1$  uciekamy się ponownie do wzoru Taylora

$$\psi''(t_k) = \frac{\psi'(t_k) - \psi'(t_{k-1})}{h} + \bar{E}_k, \quad (3.7)$$

gdzie  $\bar{E}_k = \frac{1}{h} \int_{t_{k-1}}^{t_k} (s - t_{k-1}) \psi'''(s) ds$ .

**Uwaga 7** Błędy  $E_k(\tau)$  i  $\bar{E}_k$  zawierają trzecią pochodną nieznaną funkcji  $\psi$ . Zauważmy, że

$$\begin{aligned} \partial_t^3 \psi(t) &= \partial_t \left( \partial_t^2 \psi(t) \right) = \partial_t \left( \Delta \psi(t) + f(t)\psi(t) \right) \\ &= \partial_t \Delta \psi(t) + (\partial_t f(t)) \psi(t) + f(t) \partial_t \psi(t), \end{aligned}$$

gdzie  $f(t)$  jest funkcją daną wzorem (1.2), zatem  $\partial_t f(t)$  skaluje jak  $\omega_{\max}$ . Oznacza to, że wraz ze wzrostem wartości  $\omega_{\max}$  będzie rosła wartość wyrażenia  $\partial_t^3 \psi(t)$ , a co za tym idzie także wartości błędów  $E_k(\tau)$  i  $\bar{E}_k$ , co wpłynęłoby negatywnie na ogólną stałą błędów aproksymacji.

Wykorzystamy jednak fakt, że  $\partial_t^3 \psi(t)$  we wzorach  $E_k(\tau)$  i  $\bar{E}_k$  występuje pod znakiem całki i w Lemacie 12 pokażemy, że ostateczny błąd metody nie rośnie wraz ze wzrostem oscylacji  $\omega_n$ .



Zastosowanie przybliżeń (3.6) i (3.7) w równaniach (3.4), (3.5) prowadzi bezpośrednio do przedstawionego w Tablicy 3.1 schematu numerycznego  $\Xi^{[3]}$ , gdzie  $\psi_k$  i  $\psi'_k$  są numerycznymi przybliżeniami odpowiednio  $\psi(t_k)$  i  $\psi'(t_k)$ .

Schemat numeryczny  $\Xi^{[3]}$

$$\begin{aligned} \psi_0 &= \psi(t_0), \quad \psi'_0 = \psi'(t_0), \quad \psi''_0 = \Delta\psi(t_0) - f(t_0)\psi(t_0), \\ \psi_1 &= \cos(hG)\psi(t_0) + G^{-1}\sin(hG)\psi'(t_0) + \int_0^h G^{-1}\sin((h-\tau)G)f(t_0+\tau) \left[ \psi_0 + \tau\psi'_0 + \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ \psi'_1 &= -G\sin(hG)\psi(t_0) + \cos(hG)\psi'(t_0) + \int_0^h \cos((h-\tau)G)f(t_0+\tau) \left[ \psi_0 + \tau\psi'_0 + \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ \text{for } k &= 1, N-1 \text{ do} \\ \psi_{k+1} &= \cos(hG)\psi_k + G^{-1}\sin(hG)\psi'_k + \int_0^h G_d^{-1}\sin((h-\tau)G)f(t_k+\tau) \left[ \psi_k + \tau\psi'_k + \frac{\tau^2}{2h}(\psi'_k - \psi'_{k-1}) \right] d\tau \\ \psi'_{k+1} &= -G\sin(hG)\psi_k + \cos(hG)\psi'_k + \int_0^h \cos((h-\tau)G)f(t_k+\tau) \left[ \psi_k + \tau\psi'_k + \frac{\tau^2}{2h}(\psi'_k - \psi'_{k-1}) \right] d\tau \\ \text{end do} \end{aligned}$$

TABLICA 3.1: Algorytm  $\Xi^{[3]}$  wyznaczenia przybliżonego rozwiązania  $\psi_k$  i  $\psi'_k$  w przedziale czasu  $[t_0, T]$  w punktach  $t_k = t_0 + hk$ ,  $h = (T - t_0)/N$ . Na tym etapie nie wprowadzamy jeszcze dyskretyzacji w przestrzeni.

### 3.1.1 Metoda Filon

Istotnym elementem wyprowadzonej w poprzednim podrozdziale metody numerycznej  $\Xi^{[3]}$  jest obecność dwóch całek

$$\int_0^h G_d^{-1}\sin((h-\tau)G)f(t_k+\tau) \left[ \psi_k + \tau\psi'_k + \frac{\tau^2}{2h}(\psi'_k - \psi'_{k-1}) \right] d\tau$$

oraz

$$\int_0^h \cos((h-\tau)G)f(t_k+\tau) \left[ \psi_k + \tau\psi'_k + \frac{\tau^2}{2h}(\psi'_k - \psi'_{k-1}) \right] d\tau.$$

Obie całki zawierają funkcję  $f(t_k + \tau)$ , która zgodnie z naszym założeniem, może być funkcją wysokooscylującą.

Naszym celem jest konstrukcja metody numerycznej w której nie występuje zależność między  $h$  a  $\omega_{\max}$ . Z tego powodu standardowe kwadratury, takie jak Gauss-Legendre, są nieodpowiednie (chyba że zastosujemy krok czasowy  $h < \omega_{\max}^{-1}$ ) i potrzebujemy bardziej specjalistycznego podejścia. Proponujemy zastosowanie metod typu Filon, które uzyskują żądany rząd w kroku czasowym  $h$  oraz ich stałe błędy nie zależą od (być może) dużej  $\omega_{\max}$  i gdzie nie ma potrzeby nałożenia warunków na zależność między  $h$  i  $\omega_{\max}$ . Porównanie efektywności schematu  $\Xi^{[3]}$  w zależności od metody użytej do aproksymacji całek w nim występujących przedstawiamy w podrozdziale 3.3, Rysunek 3.1.

Metody typu Filon można zastosować do aproksymacji całki z członem oscylującym postaci

$$I_\omega[f] = \int_a^b f(s)g_\omega(s)ds,$$

gdzie  $g_\omega(s)$  jest funkcją oscylującą z częstotliwością  $\omega$ . Podstawą metody Filon jest przybliżenie funkcji nieoscylacyjnej  $f(s)$  wielomianem  $p(s)$  stopnia  $r$ , takim, że

$$\forall i \in \{0, 1, \dots, r\} \quad p^{(i)}(a) = f^{(i)}(a), \quad p^{(i)}(b) = f^{(i)}(b).$$

W naszym przypadku wystarczy użyć funkcji kwadratowej, zatem

$$p(s) = \frac{f'(h) - f'(0)}{2h}s^2 + f'(0)s + f(0). \quad (3.8)$$

**Uwaga 8** W schemacie  $\Xi^{[3]}$  stosujemy metodę Filona do całek postaci  $\int_0^h f(s)e^{i\omega_n s} ds$ . Łatwo sprawdzić, że dla  $p(s)$  zdefiniowanego w (3.8) mamy

$$|f(s) - p(s)| \leq Ch^3 \max_{\xi \in [0, h]} |f'''(\xi)|,$$

gdzie stała  $C$  jest ograniczona niezależnie od  $\omega$ . To z kolei oznacza, że

$$\int_0^h f(s)e^{i\omega_n s} ds = \int_0^h \left[ \frac{f'(h) - f'(0)}{2h} s^2 + f'(0)s + f(0) \right] e^{i\omega_n s} ds + \mathcal{O}(h^4),$$

gdzie całkę po prawej stronie można obliczyć wprost przez całkowanie przez części.

Więcej szczegółów na temat metod Filona można znaleźć w [Gao and Iserles, 2017] i [Iserles and Nørsett, 2005], gdzie metody te zostały wprowadzone i wszechstronnie przeanalizowane.

## 3.2 Zbieżność metody

Rozpocniemy od pokazania, że naiwne liczenie błędu globalnego metody  $\Xi^{[3]}$ , czyli z wykorzystaniem formuł przedstawionych w Tabelicy 3.1, prowadziłyby do eksponencjalnej akumulacji błędów.

Faktycznie, zauważmy, że jeżeli  $\varepsilon_k = \|\psi(t_k) - \psi_k\|$  oraz  $\varepsilon'_k = \|\psi'(t_k) - \psi'_k\|$ , to  $\varepsilon_0 = \varepsilon'_0 = 0$  oraz

$$\begin{aligned} \psi(t_1) - \psi_1 &= G^{-1} \int_0^h \sin((h - \tau)G) f(t_0 + \tau) \left[ \frac{\tau^3}{6} \psi'''(\xi_0) \right] d\tau, \\ \psi'(t_1) - \psi'_1 &= \int_0^h \cos((h - \tau)G) f(t_0 + \tau) \left[ \frac{\tau^3}{6} \psi'''(\xi_0) \right] d\tau, \end{aligned}$$

dla pewnego  $\xi_0 \in [0, h]$ . Zatem

$$\varepsilon_1 \sim \mathcal{O}(h^5 \psi'''(\xi_0)), \quad \text{oraz} \quad \varepsilon'_1 \sim \mathcal{O}(h^4 \psi'''(\xi_0)).$$

Wprowadzamy oznaczenie  $\Psi := \max_{t \in [0, T]} \psi'''(t)$ . Wówczas dla  $k = 2, \dots, K$  mamy następujące oszacowania

$$\begin{aligned} \|\psi(t_k) - \psi_k\| &\leq \|\cos(hG)\| \|\psi(t_{k-1}) - \psi_{k-1}\| + \|G^{-1}\| \|\sin(hG)\| \|\psi'(t_{k-1}) - \psi'_{k-1}\| \\ &\quad + \|G^{-1}\| h \|\sin((h - \tau)G)\| |f(t_{k-1} + \tau)| \left( \varepsilon_{k-1} + h\varepsilon'_{k-1} + h\varepsilon_{k-1} + h\varepsilon_{k-2} + Ch^3\Psi \right), \\ \|\psi'(t_k) - \psi'_k\| &\leq \|G\| \|\sin(hG)\| \|\psi(t_{k-1}) - \psi_{k-1}\| + \|\cos(hG)\| \|\psi'(t_{k-1}) - \psi'_{k-1}\| \\ &\quad + h \|\cos((h - \tau)G)\| |f(t_{k-1} + \tau)| \left( \varepsilon_{k-1} + h\varepsilon'_{k-1} + h\varepsilon_{k-1} + h\varepsilon_{k-2} + Ch^3\Psi \right). \end{aligned}$$

gdzie  $C \in \mathbb{R}$  jest pewną ogólną stałą, niezależną od funkcji  $\psi(t)$ ,  $f(t)$  oraz oscylacji  $\omega_n$ .

Łatwo obliczyć, że powstające globalne błędy oszacowań akumulują się w następujący sposób:

$$\varepsilon_2 \sim 3 \mathcal{O}(h^5 \Psi), \quad \varepsilon_3 \sim 7 \mathcal{O}(h^5 \Psi), \quad \varepsilon_4 \sim 15 \mathcal{O}(h^5 \Psi), \quad \varepsilon_5 \sim 31 \mathcal{O}(h^5 \Psi) \dots$$

oraz

$$\varepsilon'_2 \sim 3 \mathcal{O}(h^4 \Psi), \quad \varepsilon'_3 \sim 7 \mathcal{O}(h^4 \Psi), \quad \varepsilon'_4 \sim 15 \mathcal{O}(h^4 \Psi), \quad \varepsilon'_5 \sim 31 \mathcal{O}(h^4 \Psi) \dots$$

Zatem dla  $k = 2, \dots, K$ , gdzie  $T/h = K$  i  $t_k = kh$ , otrzymujemy oszacowanie błędu globalnego postaci

$$\varepsilon_k \sim (2^k - 1) \mathcal{O}(h^5 \Psi) \quad \text{oraz} \quad \varepsilon'_k \sim (2^k - 1) \mathcal{O}(h^4 \Psi)$$

co oznacza, że dla  $k \rightarrow \infty$  błąd globalny metody  $\Xi^{[3]}$  dąży do nieskończoności.

Aby uniknąć tego problemu naszą analizę zbieżności i stabilności schematu  $\Xi^{[3]}$  zaczynamy od przeformułowania równań (3.4) i (3.5) oraz tych zaprezentowanych w Tabelicy 3.1.

**Twierdzenie 9** Formuły (3.4) i (3.5) mogą być przedstawione w równoważnej postaci

$$\psi(t_k) = \cos(khG)\psi(t_0) + G^{-1}\sin(khG)\psi'(t_0) + \sum_{l=0}^{k-1} G^{-1} \int_0^h \sin(((k-l)h - \tau)G) f(t_l + \tau) \psi(t_l + \tau) d\tau, \quad k = 1, \dots, K. \quad (3.9)$$

$$\psi'(t_k) = -G\sin(khG)\psi(t_0) + \cos(khG)\psi'(t_0) + \sum_{l=0}^{k-1} \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \psi(t_l + \tau) d\tau, \quad k = 1, \dots, K, \quad (3.10)$$

gdzie  $t \in [t_0, T]$ ,  $K \cdot h = T$ ,  $t_k = kh$ ,  $k = 0, 1, \dots, K$  i  $G \cdot G = -\Delta$ .

**Dowód:** Dowód poprowadzimy indukcyjnie. Można łatwo sprawdzić, że formuły dla  $\psi(t_1)$  i  $\psi'(t_1)$  otrzymane w (3.9) i (3.10) są równoważne do tych uzyskanych odpowiednio w (3.4) i (3.5). Załóżmy, że to samo dotyczy  $\psi(t_k)$  i  $\psi'(t_k)$  dla pewnego  $1 \leq k < K - 1$ . Wówczas

$$\begin{aligned} \psi(t_{k+1}) &= \cos(hG)\psi(t_k) + G^{-1}\sin(hG)\psi'(t_k) + \int_0^h G^{-1}\sin((h - \tau)G) f(t_k + \tau) \psi(t_k + \tau) d\tau \\ &= \cos(hG) \left[ \cos(khG)\psi(t_0) + G^{-1}\sin(khG)\psi'(t_0) \right] \\ &\quad + \cos(hG) \sum_{l=0}^{k-1} G^{-1} \int_0^h \sin(((k-l)h - \tau)G) f(t_l + \tau) \psi(t_l + \tau) d\tau \\ &\quad + G^{-1}\sin(hG) \left[ -G\sin(khG)\psi(t_0) + \cos(khG)\psi'(t_0) \right] \\ &\quad + G^{-1}\sin(hG) \sum_{l=0}^{k-1} \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \psi(t_l + \tau) d\tau \\ &\quad + \int_0^h G^{-1}\sin((h - \tau)G) f(t_k + \tau) \psi(t_k + \tau) d\tau \end{aligned}$$

Ponieważ  $G$  jest dodatnio określone, możemy zastosować tożsamości trygonometryczne

$$\begin{aligned} \cos(hG)\cos(khG) - [G^{-1}\sin(hG)][G\sin(khG)] &= \cos((k+1)hG) \\ \cos(hG)G^{-1}\sin(khG) + G^{-1}\sin(hG)\cos(khG) &= G^{-1}\sin((k+1)hG) \end{aligned}$$

otrzymując

$$\begin{aligned} \psi(t_{k+1}) &= \cos((k+1)hG)\psi(t_0) + G^{-1}\sin((k+1)hG)\psi'(t_0) \\ &\quad + \sum_{l=0}^k G^{-1} \int_0^h \sin(((k+1-l)h - \tau)G) f(t_l + \tau) \psi(t_l + \tau) d\tau. \end{aligned}$$

Dowodzi to równoważności wartości funkcji  $\psi$  uzyskanych przez (3.4) i (3.9). Analogiczny wynik można uzyskać dla  $\psi'$  przedstawionej za pomocą (3.5) i (3.10).  $\square$

**Uwaga 10** Równania ze schematu  $\Xi^{[3]}$  przedstawione w Tablicy 3.1 można przedstawić w podobnej postaci jak (3.9) i (3.10), mianowicie

$$\psi_k = \cos(khG)\psi_0 + G^{-1} \sin(khG)\psi'_0 \quad (3.11)$$

$$\begin{aligned} & + G^{-1} \int_0^h \sin((kh - \tau)G) f(t_0 + \tau) \left[ \psi_0 + \tau\psi'_0 + \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ & + \sum_{l=1}^{k-1} G^{-1} \int_0^h \sin(((k-l)h - \tau)G) f(t_l + \tau) \left[ \psi_l + \tau\psi'_l + \frac{\tau^2}{2h}(\psi'_l - \psi'_{l-1}) \right] d\tau, \\ \psi'_k & = -G \sin(khG)\psi_0 + \cos(khG)\psi'_0 \quad (3.12) \end{aligned}$$

$$\begin{aligned} & + \int_0^h \cos((kh - \tau)G) f(t_0 + \tau) \left[ \psi_0 + \tau\psi'_0 + \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ & + \sum_{l=1}^{k-1} \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \left[ \psi_l + \tau\psi'_l + \frac{\tau^2}{2h}(\psi'_l - \psi'_{l-1}) \right] d\tau, \end{aligned}$$

gdzie  $k = 2, \dots, K$ .

**Twierdzenie 11** Niech  $E_k(\tau) = \int_{t_k}^{t_k+\tau} \frac{(t_k+\tau-s)^2}{2!} \psi'''(s) ds$ ,  $k = 0, \dots, K$ ,  $\bar{E}_0 = 0$ , oraz  $\bar{E}_k = \frac{1}{h} \int_{t_{k-1}}^{t_k} (s - t_{k-1}) \psi'''(s) ds$ ,  $k = 1, \dots, K$ . Globalne błędy  $\varepsilon_k := \psi(t_k) - \psi_k$  oraz  $\varepsilon'_k := \psi'(t_k) - \psi'_k$ ,  $k = 1, \dots, K$ , metody  $\Xi^{[3]}$  można wyrazić za pomocą wzorów

$$\varepsilon_1 = \mathcal{R}_1,$$

$$\varepsilon_k = \varepsilon_{k-1}$$

$$\begin{aligned} & + 2G^{-1} \sin\left(\frac{h}{2}G\right) \sum_{l=1}^{k-2} \int_0^h \cos\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \varepsilon_l + \tau\left(1 + \frac{\tau}{2h}\right)\varepsilon'_l - \frac{\tau^2}{2h}\varepsilon'_{l-1} \right] d\tau \\ & + \int_0^h G^{-1} \sin((h - \tau)G) f(t_{k-1} + \tau) \left[ \varepsilon_{k-1} + \tau\left(1 + \frac{\tau}{2h}\right)\varepsilon'_{k-1} - \frac{\tau^2}{2h}\varepsilon'_{k-2} \right] d\tau + \mathcal{R}_k, \quad k \geq 2, \end{aligned}$$

$$\varepsilon'_1 = \mathcal{R}'_1,$$

$$\varepsilon'_k = \varepsilon'_{k-1}$$

$$\begin{aligned} & - 2 \sin\left(\frac{h}{2}G\right) \sum_{l=1}^{k-2} \int_0^h \sin\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \varepsilon_l + \tau\left(1 + \frac{\tau}{2h}\right)\varepsilon'_l - \frac{\tau^2}{2h}\varepsilon'_{l-1} \right] d\tau \\ & + \int_0^h \cos((h - \tau)G) f(t_k + \tau) \left[ \varepsilon_{k-1} + \tau\left(1 + \frac{\tau}{2h}\right)\varepsilon'_{k-1} - \frac{\tau^2}{2h}\varepsilon'_{k-2} \right] d\tau + \mathcal{R}'_k, \quad k \geq 2, \end{aligned}$$

gdzie

$$\begin{aligned} \mathcal{R}_k & = 2G^{-1} \sin\left(\frac{h}{2}G\right) \sum_{l=1}^{k-2} \int_0^h \cos\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \frac{\tau^2}{2}\bar{E}_l + E_l(\tau) \right] d\tau \\ & + \int_0^h G^{-1} \sin((h - \tau)G) f(t_{k-1} + \tau) \left[ \frac{\tau^2}{2}\bar{E}_{k-1} + E_{k-1}(\tau) \right] d\tau, \quad (3.13) \end{aligned}$$

$$\begin{aligned} \mathcal{R}'_k & = -2 \sin\left(\frac{h}{2}G\right) \sum_{l=1}^{k-2} \int_0^h \sin\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \frac{\tau^2}{2}\bar{E}_l + E_l(\tau) \right] d\tau \\ & + \int_0^h \cos((h - \tau)G) f(t_{k-1} + \tau) \left[ \frac{\tau^2}{2}\bar{E}_{k-1} + E_{k-1}(\tau) \right] d\tau \quad (3.14) \end{aligned}$$

**Dowód:** Na podstawie Twierdzenia 9 i Uwagi 10 globalne błędy  $\varepsilon_k$  i  $\varepsilon'_k$  można wyrazić za pomocą

$$\begin{aligned} \varepsilon_k &= G^{-1} \int_0^h \sin((kh - \tau)G) f(t_0 + \tau) \left[ \psi(t_0 + \tau) - \psi_0 - \tau\psi'_0 - \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ &\quad + \sum_{l=1}^{k-1} G^{-1} \int_0^h \sin(((k-l)h - \tau)G) f(t_l + \tau) \left[ \psi(t_l + \tau) - \psi_l - \tau\psi'_l - \frac{\tau^2}{2h}(\psi'_l - \psi'_{l-1}) \right] d\tau, \end{aligned} \quad (3.15)$$

$$\begin{aligned} \varepsilon'_k &= \int_0^h \cos((kh - \tau)G) f(t_0 + \tau) \left[ \psi(t_0 + \tau) - \psi_0 - \tau\psi'_0 - \frac{\tau^2}{2}\psi''_0 \right] d\tau \\ &\quad + \sum_{l=1}^{k-1} \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \left[ \psi(t_l + \tau) - \psi_l - \tau\psi'_l - \frac{\tau^2}{2h}(\psi'_l - \psi'_{l-1}) \right] d\tau. \end{aligned} \quad (3.16)$$

Korzystając z warunku początkowego  $\psi(t_0) = \psi_0$ ,  $\psi'(t_0) = \psi'_0$ , oraz faktu, że  $\psi''_0$  można wyznaczyć z równania (1.1) (pierwszy wiersz schematu  $\Xi^{[3]}$ ), otrzymujemy

$$\begin{aligned} \psi(t_0 + \tau) - \psi_0 - \tau\psi'_0 - \frac{\tau^2}{2}\psi''_0 &= E_0(\tau), \\ \psi(t_l + \tau) - \psi_l - \tau\psi'_l - \frac{\tau^2}{2h}(\psi'_l - \psi'_{l-1}) &= \varepsilon_l + \tau\varepsilon'_l + \frac{\tau^2}{2h}\varepsilon'_l - \frac{\tau^2}{2h}\varepsilon'_{l-1} + \frac{\tau^2}{2}\bar{E}_l + E_l(\tau), \quad l = 1, \dots, k-1. \end{aligned}$$

Następnie wiedząc, że  $\bar{E}_0 = 0$  i oznaczając

$$\begin{aligned} R_{k,l} &= \int_0^h G^{-1} \sin(((k-l)h - \tau)G) f(t_l + \tau) \left[ \frac{\tau^2}{2}\bar{E}_l + E_l(\tau) \right] d\tau, \\ R'_{k,l} &= \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \left[ \frac{\tau^2}{2}\bar{E}_l + E_l(\tau) \right] d\tau \end{aligned} \quad (3.17)$$

dla  $k = 0, \dots, K$  oraz  $l = 0, \dots, k-1$ , otrzymujemy

$$\begin{aligned} \varepsilon_k &= \sum_{l=1}^{k-1} G^{-1} \int_0^h \sin(((k-l)h - \tau)G) f(t_l + \tau) \left[ \varepsilon_l + \tau \left(1 + \frac{\tau}{2h}\right) \varepsilon'_l - \frac{\tau^2}{2h} \varepsilon'_{l-1} \right] d\tau + \sum_{l=0}^{k-1} R_{k,l}, \\ \varepsilon'_k &= \sum_{l=1}^{k-1} \int_0^h \cos(((k-l)h - \tau)G) f(t_l + \tau) \left[ \varepsilon_l + \tau \left(1 + \frac{\tau}{2h}\right) \varepsilon'_l - \frac{\tau^2}{2h} \varepsilon'_{l-1} \right] d\tau + \sum_{l=0}^{k-1} R'_{k,l}. \end{aligned}$$

Aby przedstawić różnicę następujących po sobie błędów globalnych  $\varepsilon_k - \varepsilon_{k-1}$ , stosujemy wzory trygonometryczne

$$\begin{aligned} \sin(((k-l)h - \tau)G) - \sin(((k-1-l)h - \tau)G) &= 2 \sin\left(\frac{h}{2}G\right) \cos\left((k-l)h - \frac{h}{2} - \tau\right)G, \\ \cos(((k-l)h - \tau)G) - \cos(((k-1-l)h - \tau)G) &= -2 \sin\left(\frac{h}{2}G\right) \sin\left((k-l)h - \frac{h}{2} - \tau\right)G. \end{aligned}$$

Wówczas

$$\begin{aligned} &\varepsilon_k - \varepsilon_{k-1} \\ &= \sum_{l=1}^{k-2} \int_0^h 2G^{-1} \sin\left(\frac{h}{2}G\right) \cos\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \varepsilon_l + \tau \left(1 + \frac{\tau}{2h}\right) \varepsilon'_l - \frac{\tau^2}{2h} \varepsilon'_{l-1} \right] d\tau \\ &\quad + \int_0^h G^{-1} \sin((h - \tau)G) f(t_{k-1} + \tau) \left[ \varepsilon_{k-1} + \tau \left(1 + \frac{\tau}{2h}\right) \varepsilon'_{k-1} - \frac{\tau^2}{2h} \varepsilon'_{k-2} \right] d\tau \\ &\quad + \sum_{l=0}^{k-2} [R_{k,l} - R_{k-1,l}] + R_{k,k-1} \end{aligned}$$

$$\begin{aligned}
& \varepsilon'_k - \varepsilon'_{k-1} \\
&= - \sum_{l=1}^{k-2} \int_0^h 2 \sin\left(\frac{h}{2}G\right) \sin\left((k-l)h - \frac{h}{2} - \tau\right)G f(t_l + \tau) \left[ \varepsilon_l + \tau\left(1 + \frac{\tau}{2h}\right) \varepsilon'_l - \frac{\tau^2}{2h} \varepsilon'_{l-1} \right] d\tau \\
&+ \int_0^h \cos((h-\tau)G) f(t_k + \tau) \left[ \varepsilon_{k-1} + \tau\left(1 + \frac{\tau}{2h}\right) \varepsilon'_{k-1} - \frac{\tau^2}{2h} \varepsilon'_{k-2} \right] d\tau \\
&+ \sum_{l=0}^{k-2} \left[ R'_{k,l} - R'_{k-1,l} \right] + R'_{k,k-1}
\end{aligned}$$

gdzie

$$\mathcal{R}_k := \sum_{l=0}^{k-2} [R_{k,l} - R_{k-1,l}] + R_{k,k-1} \quad \text{and} \quad \mathcal{R}'_k := \sum_{l=0}^{k-2} [R'_{k,l} - R'_{k-1,l}] + R'_{k,k-1},$$

co dowodzi tezy lematu. □

**Lemat 12** Wprowadzamy oznaczenia  $\mathcal{R} = \max_{k=1,\dots,K} \|\mathcal{R}_k\|$  oraz  $\mathcal{R}' = \max_{k=1,\dots,K} \|\mathcal{R}'_k\|$ , gdzie  $\mathcal{R}_k$  oraz  $\mathcal{R}'_k$  są zdefiniowane przez (3.13) oraz (3.14) w Twierdzeniu 11. Wówczas

$$\mathcal{R} \leq Ch^3 \min\{h\omega_{\max}, 1\} \quad \text{oraz} \quad \mathcal{R}' \leq C\|G\|h^3 \min\{h\omega_{\max}, 1\},$$

gdzie  $C$  zależy od  $N$ ,  $\max_{n \leq N, t \in [t_0, T]} |a_n(t)|$ ,  $\max_{n \leq N, t \in [t_0, T]} |a'_n(t)|$ ,  $\max_{t \in [t_0, T]} |\psi(t)|$ ,  $\max_{t \in [t_0, T]} |\psi'(t)|$ ,  $\max_{t \in [t_0, T]} \|\nabla \psi'(t)\|$  oraz  $T$ .

**Dowód:** Funkcje  $\mathcal{R}_k$  i  $\mathcal{R}'_k$ ,  $k = 0, \dots, K$  zależą od błędów  $E_k(\tau)$  i  $\bar{E}_k$ , które z kolei zależą od trzeciej pochodnej nieznannej funkcji

$$\psi'''(s) = \Delta \psi'(s) + i \sum_{n=1}^N \omega_n a_n(s) e^{i\omega_n s} \psi(s) + \sum_{n=1}^N a'_n(s) e^{i\omega_n s} \psi(s) + \sum_{n=1}^N a_n(s) e^{i\omega_n s} \psi'(s).$$

Drugi wyraz powyższego wyrażenia zawiera mnożenie przez  $\omega_n$ , które wymaga szczególnej uwagi. Z tego powodu zapisujemy błąd aproksymacji  $E_k(\tau)$  w postaci sumy  $E_k(\tau) = E_k^0(\tau) + E_k^\omega(\tau)$ ,  $k = 0, \dots, K$ , gdzie

$$\begin{aligned}
E_k^0(\tau) &= \int_{t_k}^{t_k+\tau} \frac{(t_k + \tau - s)^2}{2!} \left[ \Delta \psi'(s) + \sum_{n=1}^N a'_n(s) e^{i\omega_n s} \psi(s) + \sum_{n=1}^N a_n(s) e^{i\omega_n s} \psi'(s) \right] ds, \\
E_k^\omega(\tau) &= \int_{t_k}^{t_k+\tau} \frac{(t_k + \tau - s)^2}{2!} \sum_{n=1}^N i\omega_n a_n(s) e^{i\omega_n s} \psi(s) ds,
\end{aligned}$$

oraz  $\bar{E}_0 = 0$ ,  $\bar{E}_k = \bar{E}_k^0 + \bar{E}_k^\omega$ ,  $k = 1, \dots, K$ , gdzie

$$\begin{aligned}
\bar{E}_k^0(\tau) &= \frac{1}{h} \int_{t_{k-1}}^{t_k} (s - t_{k-1}) \left[ \Delta \psi'(s) + \sum_{n=1}^N a'_n(s) e^{i\omega_n s} \psi(s) + \sum_{n=1}^N a_n(s) e^{i\omega_n s} \psi'(s) \right] ds, \\
\bar{E}_k^\omega &= \frac{1}{h} \int_{t_{k-1}}^{t_k} (s - t_{k-1}) \sum_{n=1}^N i\omega_n a_n(s) e^{i\omega_n s} \psi(s) ds.
\end{aligned}$$

Łatwo zauważyć, że  $E_k^0(\tau) \sim \tau^3 C_k^0(\tau)$ ,  $k \geq 0$  i  $\bar{E}_k^0(\tau) \sim h \bar{C}_k^0$ , ( $k \geq 1$ ), gdzie  $C_k^0(\tau)$  i  $\bar{C}_k^0$  nie zależą od  $\omega_n$ . Ponadto,  $E_k^\omega(\tau) \sim \tau^3 \sum_{n=1}^N \omega_n C_k^\omega(\tau)$  i że  $\bar{E}_k^\omega(\tau) \sim h \sum_{n=1}^N \omega_n \bar{C}_k^\omega$ , gdzie  $C_k^\omega(\tau)$  i  $\bar{C}_k^\omega$  również nie zależą od częstotliwości

$\omega_n$ . Całkowanie przez części funkcji  $E_k^\omega(\tau)$

$$\begin{aligned} & \sum_n \frac{i\omega_n}{2!} \int_{t_k}^{t_k+\tau} (t_k + \tau - s)^2 a_n(s) e^{i\omega_n s} \psi(s) ds \\ &= \left| \begin{array}{l} (t_k + \tau - s)^2 a_n(s) \psi(s) \\ -2(t_k + \tau - s) a_n(s) \psi(s) + (t_k + \tau - s)^2 (a_n(s) \psi(s))'_s \end{array} \right|_{t_k}^{t_k+\tau} \frac{e^{i\omega_n s}}{i\omega_n} \\ &= \frac{1}{2!} \sum_n \left[ -\tau^2 a_n(t_k) \psi(t_k) e^{i\omega_n t_k} + \int_{t_k}^{t_k+\tau} 2(t_k + \tau - s) a_n(s) \psi(s) e^{i\omega_n s} ds \right] \\ & \quad - \frac{1}{2!} \sum_n \int_{t_k}^{t_k+\tau} (t_k + \tau - s)^2 (a_n(s) \psi(s))'_s e^{i\omega_n s} ds \end{aligned}$$

proceedzi to do wniosku, że  $E_k^\omega(\tau) \sim \tau^2 C_k^1(\tau)$  gdzie  $C_k^1(\tau)$ , nie zależy od częstotliwości  $\omega_n$ . Błąd  $\bar{E}_k^\omega$  reprezentujemy w podobny sposób, otrzymując że  $\bar{E}_k^0(\tau) \sim \bar{C}_k^1$  oraz  $\bar{C}_k^1$  nie zależy od częstotliwości  $\omega_n$ .

Podsumowując,

$$E_k(\tau) \leq C \min\{\tau^3 \omega_{\max}, \tau^2\} \quad \text{oraz} \quad \bar{E}_k \leq C \min\{h\omega_{\max}, h^0\}, \quad k \geq 1, \quad (3.18)$$

gdzie stała  $C$  jest niezależna od  $\omega_n$ , ale zależy od wielkości, takich jak  $N$ ,  $\max_{n \leq N, t \in [t_0, T]} |a_n(t)|$ ,  $\max_{n \leq N, t \in [t_0, T]} |a'_n(t)|$ ,  $\max_{t \in [t_0, T]} |\psi(t)|$ ,  $\max_{t \in [t_0, T]} |\psi'(t)|$ ,  $\max_{t \in [t_0, T]} \|\nabla \psi'(t)\|$ .

Przypomnijmy, że funkcje  $\mathcal{R}_k$  i  $\mathcal{R}'_k$  są postaci

$$\mathcal{R}_k := \sum_{l=0}^{k-2} [R_{k,l} - R_{k-1,l}] + R_{k,k-1} \quad \text{and} \quad \mathcal{R}'_k := \sum_{l=0}^{k-2} [R'_{k,l} - R'_{k-1,l}] + R'_{k,k-1}$$

gdzie wyrażenia  $R_{k,l}$  oraz  $R'_{k,l}$  zależą od wartości  $\frac{\tau^2}{2} \bar{E}_k + E_k$ . Na mocy (3.18) otrzymujemy następujące oszacowanie

$$\frac{\tau^2}{2} \bar{E}_k + E_k(\tau) \sim \min\{h^3 \sum_{n=1}^N \omega_n, h^2\} \bar{C}_k, \quad k \geq 0. \quad (3.19)$$

Ponadto, obserwujemy że  $\|2G^{-1} \sin\left(\frac{h}{2}G\right)\| \leq h$  i  $\|2\sin\left(\frac{h}{2}G\right)\| \equiv h\|G\|$ . Z połączenia tej obserwacji z oszacowaniem (3.19) z łatwością otrzymujemy tezę lematu.  $\square$

**Uwaga 13** Występujące w powyższym oszacowaniu wyrażenie  $\|G\|$  może budzić pewne obawy. Jak zdefiniowaliśmy na początku tego Rozdziału,  $G$  jest dodatnio określonym operatorem takim, że  $G \cdot G = -\Delta$ . Oznacza to jednak, że tak samo jak operator  $\Delta$ , operator  $G$  jest nieograniczony, zatem trudno jest rozważać jego normę  $\|G\|$ . Jednakże, wraz z wprowadzeniem do schematu numerycznego (symetrycznej) semi-dyskretyzacji w przestrzeni operator  $G$  będzie aproksymowany pewnym dodatnio określonym i ograniczonym operatorem.

Wszystkie zaprezentowane w tym podrozdziale Twierdzenia, Lematy oraz Uwagi pozwolą nam teraz na sformułowanie oraz dowód Twierdzenia o zbieżności metody  $\Xi^{[3]}$ :

**Twierdzenie 14** Metoda numeryczna  $\Xi^{[3]}$  jest zbieżna, jej błąd globalny  $\|\psi(T) - \psi_T\|$  nie rośnie wraz ze wzrostem  $\omega_{\max}$  i wykazuje trzeci (globalnie) rząd zbieżności względem czasu, rozumiejąc, że  $\|\psi(T) - \psi_T\| \leq Ch^3$  dla  $h \rightarrow 0$ .

**Dowód:** Niech  $M := \max_{t \in [t_0, T]} |f(t)|$ . Na podstawie Twierdzenia 11 i Lematu 12 otrzymujemy następujące nierówności dla  $k \geq 2$

$$\begin{aligned} \|\varepsilon_k\| &\leq \|\varepsilon_{k-1}\| + M \left( h \|\varepsilon_{k-1}\| + h^2 \|\varepsilon'_{k-1}\| + h^2 \|\varepsilon'_{k-2}\| \right) + \mathcal{R}, \\ \|\varepsilon'_k\| &\leq \|\varepsilon'_{k-1}\| + M \|G\| \left( h \|\varepsilon_{k-1}\| + h^2 \|\varepsilon'_{k-1}\| + h^2 \|\varepsilon'_{k-2}\| \right) + \mathcal{R}'. \end{aligned}$$

Niech  $\rho_k$  oraz  $\rho'_k$  będą górnymi ograniczeniami odpowiednio  $\|\varepsilon_k\|$  i  $\|\varepsilon'_k\|$ ,  $k = 0, \dots, K$ . Wówczas

$$\rho_0 = 0, \quad \rho'_0 = 0, \quad \rho_1 = \|\mathcal{R}_1\| \leq \mathcal{R}, \quad \rho'_1 = \|\mathcal{R}'_1\| \leq \mathcal{R}', \quad (3.20)$$

oraz dla  $k \geq 2$

$$\rho_k = \rho_{k-1} + M \left( h\rho_{k-1} + h^2\rho'_{k-1} + h^2\rho'_{k-2} \right) + \mathcal{R}, \quad (3.21)$$

$$\rho'_k = \rho'_{k-1} + M\|G\| \left( h\rho_{k-1} + h^2\rho'_{k-1} + h^2\rho'_{k-2} \right) + \mathcal{R}'. \quad (3.22)$$

Wnioskujemy, że  $\|\varepsilon_k\| \leq \rho_k$ ,  $\|\varepsilon'_k\| \leq \rho'_{k+1}$ ,  $\rho_k \leq \rho_{k+1}$ ,  $\rho'_k \leq \rho'_{k+1}$  i że spełnione są następujące nierówności:

$$\begin{aligned} \rho_k &\leq \rho_{k-1} + M \left( h\rho_{k-1} + 2h^2\rho'_{k-1} \right) + \mathcal{R}, \\ \rho'_k &\leq \rho'_{k-1} + M\|G\| \left( h\rho_{k-1} + 2h^2\rho'_{k-1} \right) + \mathcal{R}'. \end{aligned}$$

Przypomnijmy teraz, że  $\|G\| \geq 1$ . Oznacza to, że  $\mathcal{R} \leq \mathcal{R}'$ ,  $\rho_k \leq \rho'_k$  i że

$$\rho'_k \leq \left( 1 + M\|G\|(h + 2h^2) \right) \rho'_{k-1} + \mathcal{R}'.$$

Otrzymujemy zatem oszacowanie

$$\begin{aligned} \rho'_K &\leq \mathcal{R}' \sum_{j=0}^{K-1} \left( 1 + (h + 2h^2)M\|G\| \right)^j \leq \mathcal{R}' \sum_{j=0}^{K-1} (1 + M\|G\|3h)^j \\ &= \mathcal{R}' \frac{1 - (1 + M\|G\|3h)^{K-1}}{1 - (1 + M\|G\|3h)} = \mathcal{R}' \frac{-\sum_{j=1}^{K-1} \binom{K-1}{j} (M\|G\|3h)^j}{-M\|G\|3h} \\ &= \mathcal{R}' \frac{-\sum_{j=2}^{K-1} \binom{K-1}{j} (M\|G\|3h)^j - \binom{K-1}{1} (M\|G\|3h)}{-M\|G\|3h} \\ &= (K-1)\mathcal{R}' + \mathcal{R}' \frac{-\sum_{j=2}^{K-1} \binom{K-1}{j} (M\|G\|3h)^j}{-M\|G\|3h} = \mathcal{R}' \left( K-1 + 3M\|G\|\mathcal{O}(h^0) \right), \end{aligned}$$

oraz

$$\|\varepsilon_K\| \leq \|\varepsilon'_K\| \leq \rho'_K \leq TC\|G\| \min \left\{ h^3\omega, h^2 \right\}, \quad (3.23)$$

gdzie  $C$  zależy od  $N$ ,  $M$ ,  $\max_{n \leq N, s \in [t_0, T]} |a_n(s)|$ ,  $\max_{n \leq N, s \in [t_0, T]} |a'_n(s)|$ ,  $\max_{s \in [t_0, T]} |\psi(s)|$ ,  $\max_{s \in [t_0, T]} |\psi'(s)|$ ,  $\max_{s \in [t_0, T]} |\psi''(s)|$  oraz  $\max_{s \in [t_0, T]} \|\nabla \psi'(s)\|$ .

□

Z powyższego Twierdzenia wynika, że błąd globalny  $\|\psi(T) - \psi_T\|$  metody  $\Xi^{[3]}$ , czyli norma różnicy między dokładnym rozwiązaniem równania Kleina-Gordona (1.1) na końcu przedziału czasu  $\psi(T)$  oraz jego aproksymacji po  $K$  iteracjach  $\psi_T$ , jest ograniczony co najwyżej  $h^2$  i nie rośnie wraz ze wzrostem  $\omega_{\max}$ . Ponadto, dla  $h \rightarrow 0$  otrzymujemy w granicy globalny błąd trzeciego rzędu. Zgodnie z definicją oznacza to, że metoda  $\Xi^{[3]}$  zawsze jest rzędu 3-go globalnie względem czasu.

### 3.3 Przykłady numeryczne

W przykładach numerycznych skupimy się na dwóch równaniach. W Przykładzie 1 rozpatrzemy równanie falowe z wyrazem oscylującym w pojedynczej częstotliwości  $\omega$ . W pierwszej kolejności, na Rysunku 3.1, omówimy zachowanie metody  $\Xi^{[3]}$  w zależności od rodzaju metod użytych do aproksymacji całek występujących w schemacie. Na Rysunku 3.2 przedstawimy porównanie efektywności metody  $\Xi^{[3]}$  w stosunku do innych metod numerycznych znanych z literatury, w tym także do metody Asympt<sup>[3]</sup> opisaney w Rozdziale 2. W Przykładzie 2 rozważamy równanie Kleina-Gordona z wieloma częstotliwościami w funkcji masy  $f(\mathbf{x}, t)$ . Ten przykład posłuży nam do analizy efektywności metod w obliczu mieszanej postaci funkcji masy  $f(\mathbf{x}, t)$ , czyli gdy mamy duże spektrum częstotliwości  $\omega_n$ , od bardzo małych ( $\omega_{\min} = 1$ ) do bardzo dużych ( $\omega_{\max} = 10^5$ ).



Rozwiązanie referencyjne obliczane jest przy użyciu metody 6-tego rzędu [Blanes et al., 2009], z krokiem czasowym  $h = 10^{-6}$ . Następnie, stosujemy rozważane metody z różnymi krokami czasowymi i mierzymy błąd w końcowym czasie  $T$  w normie  $\ell_2$ . Laplacian jest dyskretyzowany metodą Fouriera, [Kopriva, 2009, Trefethen, 2000] z  $M = 200$  modami.

Na rysunkach 3.2 oraz 3.3 porównujemy metodę  $\Xi^{[3]}$  z innymi schematami z literatury

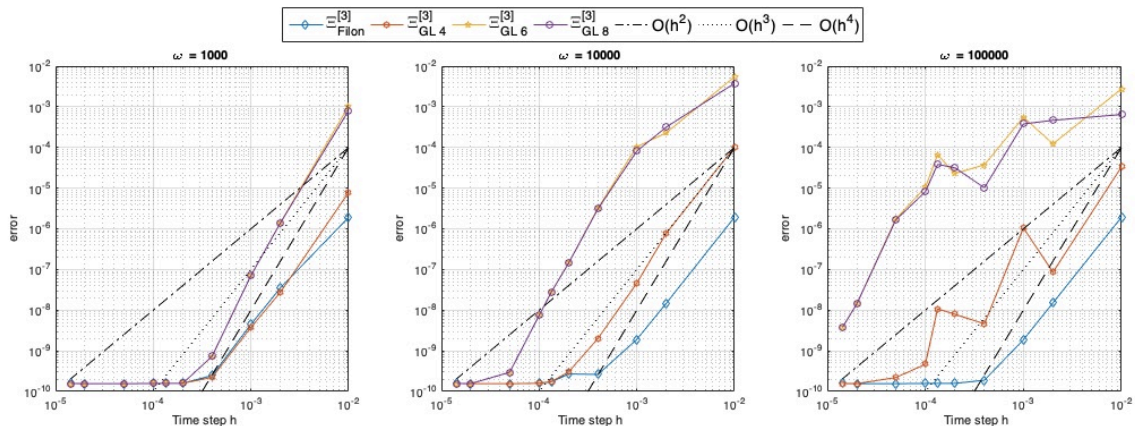
- RK<sup>[2]</sup> i RK<sup>[4]</sup> - metody Runge'a-Kutty drugiego i czwartego rzędu, odpowiednio;
- BBCK<sup>[4]</sup> - metoda czwartego rzędu  $\Sigma_{3c}^{[4]}$  z [Bader et al., 2019];
- Asympt<sup>[3]</sup> - metoda asymptotyczna rzędu trzeciego, opisana w Rozdziale 2.

*Przykład 1.*

Rozważmy następujące równanie falowe z funkcją masy zależną od czasu i przestrzeni

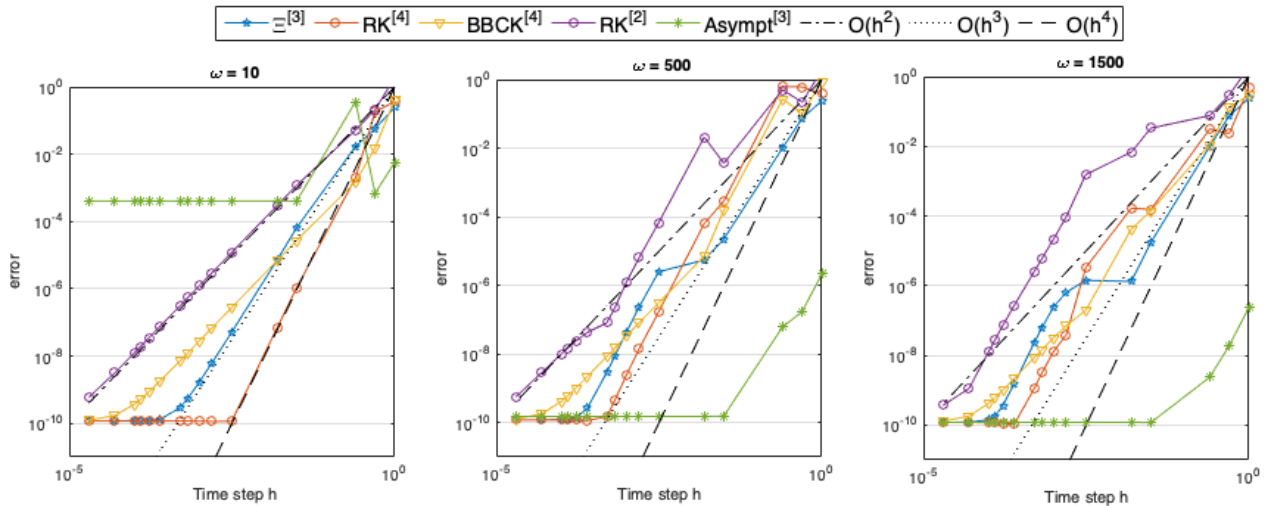
$$\begin{aligned} \partial_t^2 u &= \partial_x^2 u - [1 + \cos(\omega t)] x^2 u, \quad x \in [-10, 10], \quad t \in [0, 1], \\ u(x, 0) &= e^{-\frac{x^2}{2}}, \quad \partial_t u(x, 0) = 0, \\ u(-10, t) &= u(10, t). \end{aligned} \quad (3.24)$$

Na rysunku 3.1 parametr  $\omega$  przyjmuje wartości odpowiednio  $10^3$  (po lewej),  $10^4$  (w środku) i  $10^5$  (po prawej). Rozwiązanie równania (3.24) jest aproksymowane metodą  $\Xi^{[3]}$ , z całkami obliczonymi metodą Filona czwartego rzędu (przedstawionej w Uwadze 8) oraz z wykorzystaniem kwadratur Gaussa-Legendre'a czwartego, szóstego i ósmego rzędu. Jak widać na Rysunku 3.1 dokładność metod, w których zastosowano kwadratury Gaussa-Legendre'a znacznie odbiegają od dokładności metody w której zastosowano metody Filona, chyba że krok czasowy  $h$  jest znacznie mniejszy niż odwrotność częstotliwości  $\omega$ . Potwierdza to nasze twierdzenie, że metody typu Filon (lub inne metody dedykowane dla wysoko-oscyłujących całek) muszą być stosowane w przypadku zagadnień, w których pojawiają się duże oscylacje.



RYSUNEK 3.1: Dokładność numerycznego rozwiązania równania (3.24), gdzie wyniki uzyskano metodą  $\Xi^{[3]}$  z zastosowaniem kwadratur Filona i Gaussa-Legendre'a wysokich rzędów.

Rysunek 3.2 przedstawia porównania dokładności metod (*error*) w stosunku do wielkości kroku czasowego  $h$  i są analizowane dla trzech wielkości oscylacji:  $\omega = 10$ ,  $\omega = 500$  i  $\omega = 1500$ . Jak widać metoda Asympt<sup>[3]</sup> udowadnia swoją skuteczność przy średnich i dużych oscylacjach i zawodzi w przypadku  $\omega = 10$ . Z drugiej strony, metody RK<sup>[4]</sup> i BBCK<sup>[4]</sup> sprawdzają się bardzo dobrze w przypadku niskich i średnich oscylacji, ale zawodzą gdy  $\omega = 1500$ . Metoda drugiego rzędu RK<sup>[2]</sup> traci skuteczność już dla  $\omega = 500$ . Nowa metoda  $\Xi^{[3]}$  utrzymuje trzeci rząd zbieżności i udowadnia swoją skuteczność dla wszystkich wielkości oscylacji, niezależnie od wielkości kroku czasowego  $h$ .



RYSUNEK 3.2: Porównanie dokładności metod znanych z literatury, w których rozpatrywane są trzy wielkości oscylacji.

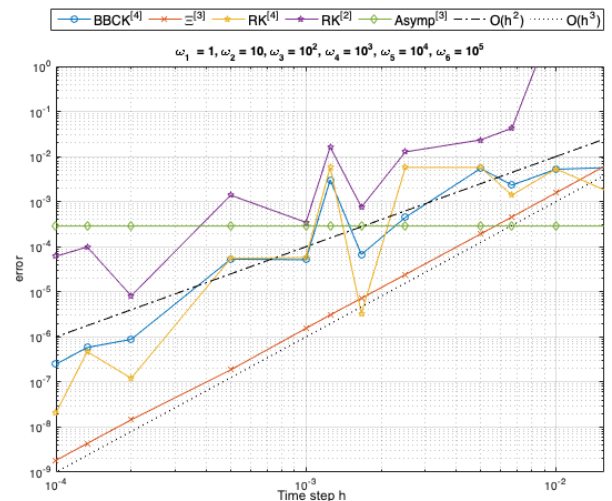
### Przykład 2.

W tym przykładzie rozważymy równanie, w którym funkcja  $f(x, t)$  charakteryzuje się dużą różnorodnością częstotliwości

$$\begin{aligned} \partial_t^2 u &= \partial_x^2 u - \sum_{n=0}^5 (1 + \cos(10^n t)) x^2 u, \quad x \in [-\pi, \pi], \quad t \in [0, 1], \\ u(x, 0) &= e^{-\frac{1}{2}(x-3)^2} + e^{-\frac{1}{2}(x+3)^2}, \quad \partial_t u(x, 0) = 0, \\ u(-\pi, t) &= u(\pi, t). \end{aligned} \quad (3.25)$$

Jak widać na Rysunku 3.3, zgodnie z teoretycznymi wynikami, błąd metody  $\Xi^{[3]}$  skaluje jak  $h^3$  niezależnie od wielkości  $\omega_n$ . Zauważmy, że na stałą błędnie nie ma wpływu bardzo wysoka częstotliwość  $\omega_{\max} = 10^5$ , i że nowa metoda zachowuje się dobrze dla wszystkich kroków czasowych. Zgodnie z oczekiwaniami, metoda asymptotyczna jest nieefektywna. Jest to spowodowane obecnością małych częstotliwości  $\omega_1 = 1$  oraz  $\omega_2 = 10$ . Z drugiej strony duże oscylacje, takie jak  $\omega_6 = 10^5$  sprawiają, że metody  $\text{RK}^{[2]}$ ,  $\text{RK}^{[4]}$  i  $\text{BBCK}^{[4]}$  stają się niewydajne.

Nasze dwa przykłady jasno potwierdzają efektywność proponowanej przez nas metody  $\Xi^{[3]}$  w porównaniu z metodami klasycznymi i asymptotycznymi.



RYSUNEK 3.3: Dokładność metod numerycznych zastosowanych do zagadnienia wieloczęstotliwościowego (3.25), gdzie funkcja  $f(x, t)$  obejmuje zakres częstotliwości od  $\omega_1 = 1$  to  $\omega_6 = 10^5$ .

## Rozdział 4

# Metoda numeryczna oparta na rozwinięciu Magnusa oraz dekompozycji pola wektorowego

Rozdział ten oparty jest na napisanej we współpracy z Karoliną Kropielnicą<sup>1</sup> oraz Kathariną Schratz<sup>2</sup> pracy *Effective highly accurate time integrators for linear Klein–Gordon equations across the scales* [Kropielnicka et al., 2021], która obecnie jest w recenzji.

Rozpocznijmy od przypomnienia dotychczas omówionych trzech sposobów numerycznego poszukiwania przybliżonego rozwiązania równania Kleina-Gordona (1.1). Pierwsze z nich, zaprezentowane w pracy [Bader et al., 2019], oparte jest na rozwinięciu Magnusa i kwadraturach typu Gauss-Legendre. Zaprezentowane w pracy metody 4-go i 6-go rzędu wykazują się bardzo dobrą dokładnością w przypadku małych oscylacji w funkcji masy  $f(\mathbf{x}, t)$ . Jednakże, zawodzą one w przypadku wysokich oscylacji  $\omega_{\max} \gg 1$ . W pracy [Condon et al., 2021], opisaney w Rozdziale 2, zaproponowaliśmy metodę specjalizującą się w czysto oscylacyjnych składnikach wejściowych, czyli  $\omega_{\min} \gg 1$ . Wadą tej metody jest to, że drastycznie zawodzi ona w przypadku nieoscylującym, czyli gdy  $\omega_{\max} \sim 1$ . Zauważmy, że żadna z tych metod nie radzi sobie skutecznie w przypadku gdy  $\omega_{\min} \sim 1$  i  $\omega_{\max} \gg 1$  jednocześnie. W poprzednim Rozdziale opisaliśmy wynik uzyskany w pracy [Kropielnicka and Lademann, 2022], gdzie zaproponowaliśmy metodę trzeciego rzędu opartą na regule Duhamela, której stała błędu nie rośnie wraz ze wzrostem oscylacji  $\omega_n$ .

W tym Rozdziale zaprezentujemy wynik z pracy [Kropielnicka et al., 2021], który przewyższa wszystkie zaprezentowane dotychczas metody pod względem dokładności i kosztu obliczeniowego. W przeciwieństwie do [Bader et al., 2019] oraz [Condon et al., 2021], proponowane metody są skuteczne dla wszystkich trzech postaciach funkcji masy  $f(\mathbf{x}, t)$ : nieoscylującej ( $\omega_{\max} \sim 1$ ), wysoce oscylującej ( $\omega_{\max} \gg 1$ ) oraz mieszanej ( $\omega_{\min} \sim 1$  oraz  $\omega_{\max} \gg 1$ ).

Proponowane podejście jest także ulepszeniem metody opisaney w [Kropielnicka and Lademann, 2022]. Po pierwsze, zbieżność trzeciego rzędu zostaje podniesiona do czwartego rzędu. Pojęcie rzędu zbieżności ma tu jednak charakter czysto formalny - duże stałe błędu nie mają wpływu na rząd zbieżności tylko gdy  $h < 1/\omega_{\max}$ . W zastosowaniach (zwłaszcza gdy  $h \geq 1/\omega_{\max}$ ) liczy się dokładność metody, a na tym polu zaproponowane w tym Rozdziale schematy numeryczne przewyższają wszystkie inne istniejące podejścia. Można to zaobserwować zarówno teoretycznie, jak i doświadczalnie. Zgodnie z Twierdzeniem 2 w tym Rozdziale, dominujący czynnik błędu zaproponowanych metod skaluje jak  $\mathcal{O}(h^5 + \min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  oraz jak  $\mathcal{O}(\min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  w szczególnym przypadku  $\alpha(\mathbf{x}, t) \equiv 0$ . Oznacza to, że najbardziej pesymistyczny przypadek tego oszacowania odpowiada najbardziej optymistycznemu oszacowaniu uzyskanemu w [Kropielnicka and Lademann, 2022].

Podrozdział 4.4 obfituje w liczne symulacje obliczeniowe oraz porównania pokazujące, że w przypadku zagadnień z dużymi oscylacjami w funkcji masy  $f(\mathbf{x}, t)$  kluczową rolę odgrywa dokładność metody, a nie jej rząd. W szczególności pokazuje to *Przykład 4*, Rysunek 4.10, gdzie w funkcji  $f(\mathbf{x}, t)$  mamy cały zakres oscylacji oraz gdzie dokładność oraz skuteczność proponowanych w tym Rozdziale metod okazują się lepsze niż we wszystkich innych podejściach.

<sup>1</sup>Institut Matematyki, Polska Akademia Nauk, Sopot

<sup>2</sup>Laboratoire Jacques-Louis Lions, Sorbonne Université, 4 place Jussieu, 75252 Paris, France

Nowe podejście opiera się na dobrze znanych narzędziach analizy numerycznej, takich jak rozwinięcie Magnusa i dekompozycja pola wektorowego. Motywację oraz wyprowadzenie metod przedstawiono w podrozdziale 4.1. Podrozdział 4.2 skupia się na wyznaczeniu oraz oszacowaniu dominujących czynników błędu, które zostały pominięte przy wyprowadzaniu schematów. W podrozdziale 4.3 przedstawiamy strukturę błędu aproksymacji oraz zilustrujemy ją graficznie. W podrozdziale 4.4 ilustrujemy nasze teoretyczne ustalenia za pomocą eksperymentów numerycznych, porównując różne istniejące metody z nowym podejściem oraz zwracamy uwagę na ich efektywność w obliczu wysokich częstotliwości  $\omega_n$ . Dodatki 4.5.1, 4.5.2 oraz 4.5.3 przedstawiają obliczenia, które stanowią uzupełnienie informacji z podrozdziału 4.2.

## 4.1 Wyprowadzenie metody

W tym podrozdziale przedstawiamy pełne wyprowadzenie metody numerycznej oraz wskażemy wyrażenia odpowiedzialne za błąd aproksymacji. Oszacowania wskazanych *dominujących czynników błędu* zostaną przedstawione w podrozdziale 4.2.

Poszukujemy funkcję  $\psi(t) := \psi(\cdot, t) \in C^2([0, T], H^{s+6}(\mathbb{T}^d))$  rozwiązującą równanie Kleina-Gordona (1.1) oraz zakładamy, że  $f(t) := f(\cdot, t) \in C^2([0, T], H^{s+4}(\mathbb{T}^d))$ ,  $s \geq 0$ .

Jak wspomnieliśmy we wstępie, zajmujemy się wyłącznie dyskretyzacją względem czasu. Z tego powodu w dalszej części tego Rozdziału pomijamy zależność od zmiennej przestrzennej  $x$  i zapisujemy  $\psi(t), a_n(t), \alpha(t), f(t)$  zamiast  $\psi(x, t), a_n(x, t), \alpha(x, t), f(x, t)$ . Wówczas możemy wyrazić (1.1) jako następujące abstrakcyjne równanie

$$\partial_t z(t) = A(t)z(t), \quad \text{gdzie} \quad z(t) = \begin{bmatrix} \psi(t) \\ \partial_t / \partial \psi(t) \end{bmatrix}, \quad (4.1)$$

gdzie

$$A(t) = \begin{bmatrix} 0 & 1 \\ N(t) & 0 \end{bmatrix} \quad \text{gdzie} \quad N(t) = \Delta + f(t). \quad (4.2)$$

Analityczne rozwiązanie układu (4.1) można przedstawić jako rozwinięcie Fera, Magnusa lub Dysona. Podobnie jak w pracy [Iserles et al., 2019], która dotyczyła równania Schrödingera, skupimy się na dwóch pierwszych wyrazach rozwinięcia Magnusa, do których zastosujemy dekompozycję pola wektorowego, znaną jako splitting Stranga, a następnie splitting zaproponowany pracy [Chin and Chen, 2002], aby otrzymać *metodę Ma-St-CC* (Magnus-Strang-Chin Chen) rzędu czwartego.

Zgodnie z literaturą [Iserles et al., 2000, Iserles et al., 2001, Blanes et al., 2009], pierwsze dwa wyrazy rozwinięcia Magnusa wystarczają do uzyskania aproksymacji czwartego rzędu. Co więcej, wyrazy te skalują odpowiednio jak  $\mathcal{O}(h)$  oraz  $\mathcal{O}(h^3)$ , więc splitting Stranga zastosowany do takiego obciążenia zachowuje dokładność czwartego rzędu, podobnie jak splitting typu [Chin and Chen, 2002] zastosowany do wewnętrznego składnika splittingu Stranga, rzędu  $\mathcal{O}(h)$ .

Podobnego rzędu zbieżności można by spodziewać się także w przypadku równania Kleina-Gordona. Jednakże, wysokie oscylacje w funkcji masy  $f(x, t)$  powodują ogromną stałą błędu, psując dokładność metody. Z tego powodu zamiast szukać zbieżności rzędu czwartego, skupimy się na poszukiwaniu dokładności metody, zależnej od relacji pomiędzy krokiem czasowym  $h$ ,  $\omega_{\min}$  oraz  $\omega_{\max}$ . W wyniku czego pokażemy, Twierdzenie 2, że lokalne oszacowania błędów są wielkości  $\mathcal{O}(h^5 + \min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  oraz  $\mathcal{O}(\min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  w szczególnym przypadku  $\alpha(x, t) \equiv 0$ .

### 4.1.1 Rozwinięcie Magnusa

Rozwinięcie Magnusa [Magnus, 1954] stosuje się w celu reprezentacji rozwiązania jednorodnego liniowego równania różniczkowego pierwszego rzędu w postaci nieskończonego szeregu

$$z(t+h) = e^{\Theta(t+h,t)} z(t), \quad (4.3)$$

gdzie  $\Theta(t+h, t) = \sum_{k=1}^{\infty} \Theta_k(t+h, t)$ , oraz każdy składnik  $\Theta_k(t+h, t)$  jest liniową kombinacją  $k$ -krotnie zagnieżdżonych całek z  $(k-1)$ -krotnie zagnieżdżonych komutatorów. Cztery pierwsze wyrazy rozwinięcia Magnusa dla (4.1) mają postać

$$\Theta_1(t+h, t) = \int_0^h A(t+t_1) dt_1 \quad (4.4)$$

$$\Theta_2(t+h, t) = -\frac{1}{2} \int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_2 dt_1 \quad (4.5)$$

$$\begin{aligned} \Theta_3(t+h, t) &= \frac{1}{6} \int_0^h \int_0^{t_1} \int_0^{t_2} [A(t+t_1), [A(t+t_2), A(t+t_3)]] dt_3 dt_2 dt_1 \\ &\quad + \frac{1}{6} \int_0^h \int_0^{t_1} \int_0^{t_2} [A(t+t_3), [A(t+t_2), A(t+t_1)]] dt_3 dt_2 dt_1 \end{aligned} \quad (4.6)$$

$$\begin{aligned} \Theta_4(t+h, t) &= \frac{1}{12} \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} [[[A(t+t_1), A(t+t_2)], A(t+t_3)], A(t+t_4)] dt_4 dt_3 dt_2 dt_1 \\ &\quad + \frac{1}{12} \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} [A(t+t_1), [[A(t+t_2), A(t+t_3)], A(t+t_4)]] dt_4 dt_3 dt_2 dt_1 \\ &\quad + \frac{1}{12} \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} [A(t+t_1), [A(t+t_2), [A(t+t_3), A(t+t_4)]]] dt_4 dt_3 dt_2 dt_1 \\ &\quad + \frac{1}{12} \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} [A(t+t_2), [A(t+t_3), [A(t+t_4), A(t+t_1)]]] dt_4 dt_3 dt_2 dt_1. \end{aligned} \quad (4.7)$$

Metoda Ma-St-CC oparta jest na dwóch pierwszych wyrazach powyższego rozwinięcia

$$z(t+h) \approx \exp\left(\int_0^h A(t+t_1) dt_1 - \frac{1}{2} \int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_2 dt_1\right) z(t). \quad (4.8)$$

Wyrażenia  $\Theta_3(t+h, t)$  oraz  $\Theta_4(t+h, t)$  stanowią zatem dominujące czynniki błędu metody Ma-St-CC. Jak wyjaśnimy na końcu podrozdziału 4.2.1 wielkości składników  $\Theta_k(t+h, t)$ ,  $k \geq 5$  są  $h^{k-4}$ -razy mniejsze niż  $\Theta_4(t+h, t)$ . Oszacowania  $\Theta_3(t+h, t)$  oraz  $\Theta_4(t+h, t)$  pokazujące, że skalują one jak  $\mathcal{O}(h^5 + \min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  oraz jak  $\mathcal{O}(\min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$  w przypadku, gdy  $\alpha(\mathbf{x}, t) \equiv 0$ , zostaną zaprezentowane w podrozdziale 4.2.1.

## 4.1.2 Splitting Stranga

Macierz  $A(t+t_1)$  jest anty-diagonalna, natomiast komutator  $[A(t+t_2), A(t+t_1)]$  redukuje się do macierzy diagonalnej. Oznacza to, że macierz  $\int_0^h A(t+t_1) dt_1 - \frac{1}{2} \int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_2 dt_1$  nie posiada struktury diagonalnej, anty-diagonalnej ani cyklicznie symetrycznej, i numeryczne liczenie jej eksponentu byłoby *kosztowne obliczeniowo* za pomocą np. metody podprzestrzeni Kryłowa lub rozwinięcia Taylora (analogicznie jak w Rozdziale 2). Z tego powodu podzielimy ten wykładnik za pomocą splittingu Stranga

$$\exp(X+Y) \approx \exp\left(\frac{1}{2}X\right) \exp(Y) \exp\left(\frac{1}{2}X\right), \quad (4.9)$$

którego zbieżność udowodniono np. w [Jahnke and Lubich, 2000], oraz którego dokładność zależy od komutatorów  $[Y, [Y, X]]$  oraz  $[X, [X, Y]]$ .

Dekompozycja (4.9) zastosowana do równania (4.8) prowadzi do przybliżenia

$$\begin{aligned} z(t+h) \approx & \exp\left(-\frac{1}{4} \int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_2 dt_1\right) \exp\left(\int_0^h A(t+t_1) dt_1\right) \\ & \exp\left(-\frac{1}{4} \int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_2 dt_1\right) z(t). \end{aligned} \quad (4.10)$$

Szczegółowe oszacowanie błędu występującego przy dekompozycji (4.10) znajduje się w podrozdziale 4.2.2.

W dalszej części tego rozdziału wyrażenia

$$\exp\left(-\frac{1}{4}\int_0^h\int_0^{t_1}[A(t+t_2),A(t+t_1)]dt_2dt_1\right)$$

nazywać będziemy *składnikami zewnętrznymi*, a wyrażenie

$$\exp\left(\int_0^h A(t+t_1)dt_1\right)$$

*składnikiem wewnętrznym* splittingu (4.10). Składniki te różnią się znacząco między sobą. Składnik wewnętrzny jest eksponentem z pojedynczej całki z macierzy anty-diagonalnej, a składniki zewnętrzne (jak pokażemy w następnym podrozdziale) można uprościć do eksponentu macierzy diagonalnej z pojedynczą całką. Ze względu na te różnice, wyrażenia te wymagają zastosowania innych narzędzi numerycznych.

**Uwaga 1** *Aproksymację podobnego typu*

$$\begin{aligned} z(t+h) \approx & \exp\left(\frac{1}{2}\int_0^h A(t+t_1)dt_1\right)\exp\left(-\frac{1}{2}\int_0^h\int_0^{t_1}[A(t+t_2),A(t+t_1)]dt_2dt_1\right) \\ & \exp\left(\frac{1}{2}\int_0^h A(t+t_1)dt_1\right)z(t), \end{aligned} \quad (4.11)$$

można również otrzymać bezpośrednio z symetrycznego rozwinięcia Fera, które zostało zaproponowane w [Zanna, 2001] oraz analizowane i ulepszone w [Blanes et al., 2002]. Różnica między (4.10) a (4.11) polega na tym, że (4.11) zawiera dwa składniki  $\exp\left(\frac{1}{2}\int_0^h A(t+t_1)dt_1\right)$ , co jak będzie oczywiste z następnymi dwoma podrozdziałami, sprawiłoby, że schemat numeryczny byłby bardziej skomplikowany i kosztowny obliczeniowo.

### 4.1.3 Składniki zewnętrzne splittingu (4.10)

Naszą pierwszą obserwacją dotyczącą składników zewnętrznych splittingu (4.10) jest, że występująca w nim całka podwójna z komutatora macierzy anty-diagonalnych redukuje się do macierzy diagonalnej z całek podwójnych

$$\begin{aligned} & \int_0^h\int_0^{t_1}[A(t+t_2),A(t+t_1)]dt_2dt_1 \\ &= \int_0^h\int_0^{t_1}\left[\begin{bmatrix} 0 & 1 \\ N(t+t_2) & 0 \end{bmatrix},\begin{bmatrix} 0 & 1 \\ N(t+t_1) & 0 \end{bmatrix}\right]dt_2dt_1 \\ &= \int_0^h\int_0^{t_1}\begin{bmatrix} N(t+t_1)-N(t+t_2) & 0 \\ 0 & N(t+t_2)-N(t+t_1) \end{bmatrix}dt_2dt_1 \\ &= \begin{bmatrix} \int_0^h\int_0^{t_1}(f(t+t_1)-f(t+t_2))dt_2dt_1 & 0 \\ 0 & \int_0^h\int_0^{t_1}(f(t+t_2)-f(t+t_1))dt_2dt_1 \end{bmatrix}. \end{aligned}$$

W związku z tym że numeryczne przybliżenie całek podwójnych jest kosztowne obliczeniowo (wymaga obliczeń w dwóch wymiarach) stosujemy całkowanie przez części

$$\begin{aligned} \int_0^h\int_0^s(f(t)-f(s))dtds &= \int_0^h\int_0^sf(t)dtds - \int_0^h\int_0^sf(s)dsdt = \begin{vmatrix} \int_0^sf(t)dt & 1 \\ f(s) & s \end{vmatrix} \\ &= \int_0^h\int_0^sf(t)dtds - \int_0^h sf(s)ds - \int_0^h\int_0^sf(s)dtds \\ &= h\int_0^hf(s)ds - \int_0^h sf(s)ds - \int_0^h sf(s)ds \\ &= \int_0^h(h-2s)f(s)ds = 2\int_0^h\left(\frac{h}{2}-s\right)f(s)ds. \end{aligned}$$

w wyniku czego otrzymujemy trywialny obliczeniowo eksponent macierzy diagonalnej z całek pojedynczych

$$\begin{aligned} & \exp \begin{bmatrix} -2 \int_0^h (\frac{h}{2} - t_1) f(\mathbf{x}, t_1) dt_1 & 0 \\ 0 & 2 \int_0^h (\frac{h}{2} - t_1) f(\mathbf{x}, t + t_1) dt_1 \end{bmatrix} \\ &= \begin{bmatrix} \exp \left( -2 \int_0^h (\frac{h}{2} - t_1) f(\mathbf{x}, t + t_1) dt_1 \right) & 0 \\ 0 & \exp \left( 2 \int_0^h (\frac{h}{2} - t_1) f(\mathbf{x}, t + t_1) dt_1 \right) \end{bmatrix}. \end{aligned}$$

**Uwaga 15** W przypadku wielowymiarowym, po semi-dyskretyzacji,

$$\mathbf{x} = (x_1^1, \dots, x_M^1) \otimes \dots \otimes (x_1^d, \dots, x_M^d)$$

całkę

$$\int_0^h \int_0^{t_1} [A(\mathbf{x}, t + t_2), A(\mathbf{x}, t + t_1)] dt_2 dt_1$$

można przybliżyć macierzą diagonalną, będącą iloczynem tensorowym macierzy diagonalnych

$$\begin{bmatrix} I(x_1^1, t) & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & I(x_M^1, t) & 0 & \dots & 0 \\ 0 & \dots & 0 & -I(x_1^1, t) & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & -I(x_M^1, t) \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} I(x_1^d, t) & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & I(x_M^d, t) & 0 & \dots & 0 \\ 0 & \dots & 0 & -I(x_1^d, t) & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & -I(x_M^d, t) \end{bmatrix},$$

gdzie

$$I(x_i^k, t) = 2 \int_0^h (t_1 - \frac{h}{2}) f(x_i^k, t + t_1) dt_1, \quad k = 1, \dots, d, i = 1, \dots, M.$$

#### 4.1.4 Składnik wewnętrzny splittingu (4.10)

W tym podrozdziale skupimy się na składniku wewnętrznym splittingu (4.10)

$$\exp \left( \int_0^h A(t + t_1) dt_1 \right) = \exp \left( \begin{bmatrix} 0 & h \\ D + F & 0 \end{bmatrix} \right), \quad (4.12)$$

gdzie dla uproszczenia notacji zapisujemy  $D := h\Delta$  oraz  $F := \int_0^h f(t + t_1) dt_1$ . Jak wspominaliśmy w Rozdziale 2, wykładnik macierzy anty-diagonalnej (4.12) można obliczyć za pomocą funkcji hiperbolicznych,

$$\exp \left( \begin{bmatrix} 0 & h \\ D + F & 0 \end{bmatrix} \right) = \begin{bmatrix} \cosh \left( \sqrt{h(D + F)} \right) & \sqrt{\frac{h}{D + F}} \sinh \left( \sqrt{h(D + F)} \right) \\ \sqrt{\frac{D + F}{h}} \sinh \left( \sqrt{h(D + F)} \right) & \cosh \left( \sqrt{h(D + F)} \right) \end{bmatrix}.$$

Obliczenie sinusa lub cosinusa hiperbolicznego z wyrażenia  $\sqrt{h(D + F)}$  może być skomplikowane i kosztowne, ponieważ nie jest ono ani macierzą diagonalną, ani cykliczną symetryczną. Z tego powodu posłużymy się dekompozycją macierzy  $\begin{bmatrix} 0 & h \\ D + F & 0 \end{bmatrix}$  taką, że eksponent każdego z uzyskanych składników będzie łatwy obliczeniowo.

W tym celu wykorzystamy splitting rzędu 4-go zaprezentowany w pracy [Chin and Chen, 2002], postaci

$$e^{hX+hY} \approx e^{\frac{1}{6}hX} e^{\frac{1}{2}hY} e^{\frac{2}{3}hX} e^{\frac{1}{72}h^3[X, Y, X]} e^{\frac{1}{2}hY} e^{\frac{1}{6}hX}, \quad (4.13)$$

którego rząd zbieżności zależy od czterokrotnie zagnieżdżonych komutatorów  $[Y, [Y, [Y, [Y, X]]], [X, [X, [X, [X, Y]]], [X, [Y, [Y, [Y, X]]], [Y, [X, [X, [X, Y]]], [Y, [X, [Y, [X, Y]]]$  oraz  $[X, [Y, [X, [Y, X]]]$ . Oszacowania tych komutatorów znajdują się w podrozdziale 4.2.3. Wyprowadzenie splittingu typu Chin-Chen znajduje się w Dodatku 4.5.3.

**Schemat  $\Gamma_1^{[4]}$** 

W pierwszej dekompozycji oddzielamy operator Laplace'a  $\begin{bmatrix} 0 & 0 \\ D & 0 \end{bmatrix}$  od funkcji masy  $\begin{bmatrix} 0 & h \\ F & 0 \end{bmatrix}$  a następnie stosujemy splitting (4.13) otrzymując następującą dekompozycję rzędu czwartego

$$\begin{aligned} \exp\left(\begin{bmatrix} 0 & h \\ D+F & 0 \end{bmatrix}\right) &= \exp\left(\begin{bmatrix} 0 & 0 \\ D & 0 \end{bmatrix} + \begin{bmatrix} 0 & h \\ F & 0 \end{bmatrix}\right) \\ &= \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{6}D & 0 \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & \frac{1}{2}h \\ \frac{1}{2}F & 0 \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{2}{3}D + \frac{0}{72}hD^2 & 0 \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & \frac{1}{2}h \\ \frac{1}{2}F & 0 \end{bmatrix}\right) \\ &\quad \times \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{6}D & 0 \end{bmatrix}\right) + \mathcal{O}(h^5). \end{aligned}$$

Zauważmy, że funkcje hiperboliczne  $\sqrt{hF}/2$  występujące w

$$\begin{aligned} \exp\left(\begin{bmatrix} 0 & \frac{1}{2}h \\ \frac{1}{2}F & 0 \end{bmatrix}\right) &= \begin{bmatrix} \cosh\left(\frac{\sqrt{hF}}{2}\right) & \sqrt{\frac{h}{F}} \sinh\left(\frac{\sqrt{hF}}{2}\right) \\ \sqrt{\frac{F}{h}} \sinh\left(\frac{\sqrt{hF}}{2}\right) & \cosh\left(\frac{\sqrt{hF}}{2}\right) \end{bmatrix} \\ &= \begin{bmatrix} \cosh\left(\frac{\sqrt{hF}}{2}\right) & \frac{h}{2} \operatorname{sinc}\left(\frac{\sqrt{hF}}{2}\right) \\ \sqrt{\frac{F}{h}} \sinh\left(\frac{\sqrt{hF}}{2}\right) & \cosh\left(\frac{\sqrt{hF}}{2}\right) \end{bmatrix} \end{aligned}$$

można łatwo obliczyć, ponieważ  $\frac{\sqrt{hF}}{2}$  staje się macierzą diagonalną po semi-dyskretyzacji. EkspONENTY z pozostałych dwóch macierzy można także policzyć w łatwy sposób z definicji eksponentu macierzowego, mianowicie

$$\exp\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{6}D & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ \frac{1}{6}D & 1 \end{bmatrix} \quad \text{oraz} \quad \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{2}{3}D + \frac{0}{72}hD^2 & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ \frac{2}{3}D + \frac{0}{72}hD^2 & 1 \end{bmatrix}.$$

**Schemat  $\Gamma_2^{[4]}$** 

Alternatywny splitting można uzyskać zachowując razem operator Laplace'a oraz funkcję masy, w którym suma  $D + F$  jest jedynym niezerowym elementem macierzy. Po zastosowaniu splittingu (4.13) pozostajemy z zadaniem wyliczenia eksponentów macierzy z tylko jednym niezerowym elementem

$$\begin{aligned} \exp\left(\begin{bmatrix} 0 & h \\ D+F & 0 \end{bmatrix}\right) &= \exp\left(\begin{bmatrix} 0 & 0 \\ D+F & 0 \end{bmatrix} + \begin{bmatrix} 0 & h \\ 0 & 0 \end{bmatrix}\right) \\ &= \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{6}(D+F) & 0 \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & \frac{1}{2}h \\ 0 & 0 \end{bmatrix}\right) \\ &\quad \times \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{2}{3}(D+F) + \frac{0}{72}h(D+F)^2 & 0 \end{bmatrix}\right) \exp\left(\begin{bmatrix} 0 & \frac{1}{2}h \\ 0 & 0 \end{bmatrix}\right) \\ &\quad \times \exp\left(\begin{bmatrix} 0 & 0 \\ \frac{1}{6}(D+F) & 0 \end{bmatrix}\right) + \mathcal{O}(h^5), \end{aligned}$$

co sprawia, że schemat jest niezwykle szybki obliczeniowo.

**4.1.5 Kompletnie schematy numeryczne  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$** 

W podrozdziale 4.1.3 przeprowadziliśmy analizę składników zewnętrznych splittingu (4.10) oraz zaprezentowaliśmy dwie dekompozycje rzędu 4-go pozwalające na efektywne obliczenie eksponentu składnika wewnętrznego splittingu (4.10), 4.1.4. W tym podrozdziale zaprezentujemy kompletne schematy numeryczne



$\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$ , pozwalające na wyznaczenie przybliżonego rozwiązania równania Kleina-Gordona (1.1).

Przyjmując wprowadzone wcześniej uproszczenia w notacji definiujemy

$$D := h\Delta, \quad F_k := \int_0^h f(t_k + t_1) dt_1 \quad \text{oraz} \quad \mathcal{F}_k := \frac{1}{2} \int_0^h (t_1 - \frac{h}{2}) f(t_k + t_1) dt_1.$$

Zakładając, że rozwiązanie  $z(t_k)$  jest znane w kroku czasowym  $t_k = kh$ , kompletny schemat numeryczny  $\Gamma_1^{[4]}$  wygląda następująco

$$\begin{aligned} z(t_k + h) \approx & \begin{bmatrix} \exp(-\mathcal{F}_k) & 0 \\ 0 & \exp(\mathcal{F}_k) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{6}D & 1 \end{bmatrix} \begin{bmatrix} \cosh\left(\frac{\sqrt{hF_k}}{2}\right) & \frac{h}{2}\text{sinc}\left(\frac{\sqrt{hF_k}}{2}\right) \\ \sqrt{\frac{F_k}{h}} \sinh\left(\frac{\sqrt{hF_k}}{2}\right) & \cosh\left(\frac{\sqrt{hF_k}}{2}\right) \end{bmatrix} \\ & \times \begin{bmatrix} 1 & 0 \\ \frac{2}{3}D + \frac{2}{72}hD^2 & 1 \end{bmatrix} \begin{bmatrix} \cosh\left(\frac{\sqrt{hF_k}}{2}\right) & \frac{h}{2}\text{sinc}\left(\frac{\sqrt{hF_k}}{2}\right) \\ \sqrt{\frac{F_k}{h}} \sinh\left(\frac{\sqrt{hF_k}}{2}\right) & \cosh\left(\frac{\sqrt{hF_k}}{2}\right) \end{bmatrix} \\ & \times \begin{bmatrix} 1 & 0 \\ \frac{1}{6}D & 1 \end{bmatrix} \begin{bmatrix} \exp(-\mathcal{F}_k) & 0 \\ 0 & \exp(\mathcal{F}_k) \end{bmatrix} z(t_k). \end{aligned}$$

Algorytm dla schematu  $\Gamma_1^{[4]}$  przedstawiono w Tabelicy 4.1.

T-krokowy algorytm dla schematu $\Gamma_1^{[4]}$
<p><b>do</b> <math>k = 0, T - 1</math></p> <p><math>q_0 = \exp(-\mathcal{F}_k)z_1; \quad p_0 = \exp(\mathcal{F}_k)z_2</math></p> <p><math>p_1 = \frac{1}{6}Dq_0 + p_0</math></p> <p><math>q_1 = \cosh\left(\frac{\sqrt{hF_k}}{2}\right)q_0 + \frac{h}{2}\text{sinhc}\left(\frac{\sqrt{hF_k}}{2}\right)p_1</math></p> <p><math>p_2 = \sqrt{\frac{F_k}{h}} \sinh\left(\frac{\sqrt{hF_k}}{2}\right)q_0 + \cosh\left(\frac{\sqrt{hF_k}}{2}\right)p_1</math></p> <p><math>p_3 = \left(\frac{2}{3}D + \frac{2}{72}hD^2\right)q_1 + p_2</math></p> <p><math>q_2 = \cosh\left(\frac{\sqrt{hF_k}}{2}\right)q_1 + \frac{h}{2}\text{sinhc}\left(\frac{\sqrt{hF_k}}{2}\right)p_3</math></p> <p><math>p_4 = \sqrt{\frac{F_k}{h}} \sinh\left(\frac{\sqrt{hF_k}}{2}\right)q_1 + \cosh\left(\frac{\sqrt{hF_k}}{2}\right)p_3</math></p> <p><math>p_5 = \frac{1}{6}Dq_2 + p_4</math></p> <p><math>q_3 = \exp(-\mathcal{F}_k)q_2; \quad p_6 = \exp(\mathcal{F}_k)p_5</math></p> <p><math>z_1 := q_3; \quad z_2 := p_6</math></p> <p><b>end do</b></p>

TABLICA 4.1: Algorytm  $\Gamma_1^{[4]}$  znajdowania przybliżonego rozwiązania równania (1.1) na przedziale czasu  $[t_0, t_T]$  z  $T$  krokami czasowymi  $h = (t_T - t_0)/T$ , gdzie  $z(t) = [z_1(t), z_2(t)]^\top$ . Dyskretyzacja w przestrzeni nie jest jeszcze zdefiniowana.

W podobny sposób prezentujemy końcowy schemat  $\Gamma_2^{[4]}$ , którego algorytm przedstawia Tablica 4.2.

$$\begin{aligned} z(t_k + h) \approx & \begin{bmatrix} \exp(-\mathcal{F}_k) & 0 \\ 0 & \exp(\mathcal{F}_k) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{6}(D + F_k) & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2}h \\ 0 & 1 \end{bmatrix} \\ & \times \begin{bmatrix} 1 & 0 \\ \frac{2}{3}(D + F_k) + \frac{2}{72}h(D + F_k)^2 & 1 \end{bmatrix} \\ & \times \begin{bmatrix} 1 & \frac{1}{2}h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1}{6}(D + F_k) & 1 \end{bmatrix} \begin{bmatrix} \exp(-\mathcal{F}_k) & 0 \\ 0 & \exp(\mathcal{F}_k) \end{bmatrix} z(t_k). \end{aligned}$$

T-krokowy algorytm dla schematu  $\Gamma_2^{[4]}$ 

```

do  $k = 0, T - 1$ 
   $q_0 = \exp(-\mathcal{F}_k)z_1; \quad p_0 = \exp(\mathcal{F}_k)z_2$ 
   $p_1 = \frac{1}{6}(D + F_k)q_0 + p_0$ 
   $q_1 = q_0 + \frac{1}{2}hp_1$ 
   $p_2 = \left(\frac{2}{3}(D + F_k) + \frac{2}{72}h(D + F_k)^2\right)q_1 + p_1;$ 
   $q_2 = q_1 + \frac{1}{2}hp_2$ 
   $p_3 = \frac{1}{6}(D + F_k)q_2 + p_2$ 
   $q_3 = \exp(-\mathcal{F}_k)q_2; \quad p_4 = \exp(\mathcal{F}_k)p_3$ 
   $z_1 := q_3; \quad z_2 := p_4$ 
end do

```

TABLICA 4.2: Algorytm  $\Gamma_2^{[4]}$  znajdowania przybliżonego rozwiązania (1.1) na przedziale czasu  $[t_0, t_T]$  z  $T$  krokami czasowymi  $h = (t_T - t_0)/T$ , gdzie  $z(t) = [z_1(t), z_2(t)]^\top$ . Dyskretyzacja w przestrzeni nie jest jeszcze zdefiniowana.

## 4.2 Oszacowania dominujących czynników błędu

W poprzednim podrozdziale przedstawiliśmy pełne wyprowadzenie schematów numerycznych. Wspomnieliśmy także, że błędy aproksymacji wynikające z obcięcia rozwinięcia Magnusa oraz zastosowania dekompozycji Stranga mogą zależeć od współczynników oscylacyjnych  $\omega_n$ . W tym podrozdziale przedstawiamy dokładną analizę i oszacowania tych składników. Poniższe twierdzenie będzie często wykorzystywane w naszej analizie.

**Twierdzenie 1** Niech  $a \in C^1[0, h]$  będzie funkcją rzeczywistą,  $h \leq 1$  oraz  $\omega \geq 1$ . Wówczas prawdziwe jest następujące oszacowanie:

$$\left| \int_0^h \int_0^{t_1} \int_0^{t_2} \dots \int_0^{t_{m-1}} a(t_k) e^{i\omega t_k} dt_m dt_{m-1} \dots dt_1 \right| \leq C \mathbb{A} \min \left\{ h^m, \frac{h^{m-1}}{\omega} \right\}, \quad 1 \leq k \leq m, \quad (4.14)$$

gdzie  $C$  jest stałą oraz  $\mathbb{A} = \max_{t \in [0, h]} \{|a(t)|, |a'(t)|\}$ .

**Dowód:** Przez  $C$  oznaczamy pewną ogólną stałą. Pierwszym oczywistym i natychmiastowym oszacowaniem jest

$$\left| \int_0^{t_{m-1}} t_m^r a(t_m) e^{i\omega t_m} dt_m \right| \leq C \mathbb{A} h^{r+1}. \quad (4.15)$$

Jednakże, przez proste całkowanie przez części otrzymujemy, że dla  $r = 0$

$$\int_0^{t_{m-1}} a(t_m) e^{i\omega t_m} dt_m = \frac{1}{i\omega} a(t_{m-1}) e^{i\omega t_{m-1}} - \frac{1}{i\omega} a(0) - \frac{1}{i\omega} \int_0^{t_{m-1}} a'(t_m) e^{i\omega t_m} dt_m,$$

oraz, że dla  $r = 1, 2, \dots$

$$\begin{aligned} \int_0^{t_{m-1}} t_m^r a(t_m) e^{i\omega t_m} dt_m &= \frac{1}{i\omega} t_{m-1}^r a(t_{m-1}) e^{i\omega t_{m-1}} - \frac{1}{i\omega} \int_0^{t_{m-1}} r t_m^{r-1} a(t_m) e^{i\omega t_m} dt_m \\ &\quad - \frac{1}{i\omega} \int_0^{t_{m-1}} t_m^r a'(t_m) e^{i\omega t_m} dt_m. \end{aligned} \quad (4.16)$$

Możemy zatem zaproponować bardziej subtelne oszacowanie postaci

$$\left| \int_0^{t_{m-1}} t_m^r a(t_m) e^{i\omega t_m} dt_m \right| \leq \frac{h^r}{\omega} C \mathbb{A} + r \frac{h^r}{\omega} C \mathbb{A} + \frac{h^{r+1}}{\omega} C \mathbb{A}, \quad r = 0, 1, 2, \dots \quad (4.17)$$

Z połączenia oszacowań (4.15) i (4.17) otrzymujemy nierówność

$$\left| \int_0^{t_{m-1}} t_m^r a(t_m) e^{i\omega t_m} dt_m \right| \leq (r+2)CA \min \left\{ h^{r+1}, \frac{h^r}{\omega} \right\}, \quad 0 \leq t_m < t_{m-1} \leq h. \quad (4.18)$$

Teza (4.14) Twierdzenia 1 wynika bezpośrednio z oszacowania (4.18).  $\square$

**Uwaga 2** Niech  $a_n \in C^1[0, h]$ ,  $n < |N|$ ,  $N \in \mathbb{N}$  będzie rodziną znanych funkcji o wartościach rzeczywistych oraz niech  $h \leq 1$  oraz  $\omega_{\min} \geq 1$ . Wówczas możemy pokazać wynik podobny do otrzymanego w Twierdzeniu 1, a mianowicie

$$\left| \int_0^h \int_0^{t_1} \int_0^{t_2} \dots \int_0^{t_{m-1}} \sum_{|n| \leq N} a_n(t_k) e^{i\omega_n t_k} dt_m dt_{m-1} \dots dt_1 \right| \leq CA \min \left\{ h^m, \frac{h^{m-1}}{\omega_{\min}} \right\}, \quad 1 \leq k \leq m, \quad (4.19)$$

gdzie  $C$  jest pewną ogólną stałą oraz  $A = \sum_{|n| \leq N} \max_{t \in [0, h]} \{|a_n(t)|, |a'_n(t)|; |n| < N\}$ .

**Uwaga 3** Aby ułatwić Czytelnikowi odbiór zaprezentowanych w tym podrozdziale oszacowań wprowadzamy uproszczenie notacji, pisząc

$$f(t_k) = \alpha(t_k) + \sum_{|n| \leq N} a_n(t_k) e^{i\omega_n t_k}$$

zamiast

$$f(t + t_k) = \alpha(t + t_k) + \sum_{|n| \leq N} a_n(t + t_k) e^{i\omega_n(t+t_k)}.$$

W dalszej części artykułu przez  $\|\cdot\|_s$  oznaczamy standardową normę Sobolewa  $H^s$  na torusie  $\mathbb{T}^d$  oraz zakładamy, że spełnione są następujące założenia

**Założenie 1** Poszukując rozwiązania  $\psi(t) := \psi(\cdot, t) \in C^2([0, T], H^{s+6}(\mathbb{T}^d))$  zakładamy, że

1.  $a_n(t) := a_n(\cdot, t) \in C^2([0, T], H^{s+4}(\mathbb{T}^d))$ ,
2.  $\alpha(t) := \alpha(\cdot, t) \in C^2([0, T], H^{s+4}(\mathbb{T}^d))$ ,
3.  $s \geq 0$ ,  $s + 6 > d/2$ .

**Definicja 1** We wszystkich obliczeniach tego podrozdziału oznaczamy przez  $C$  pewną ogólną stałą. Ponadto wprowadzamy następujące oznaczenia

$$\begin{aligned} \mathcal{E}_\omega^h &= \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\}, \\ \tilde{\mathcal{A}} &= \sum_{|n| \leq N} \max_{t \in [0, T]} \left\{ \|a_n(t)\|_{s+4}, \|a_n^2(t)\|_{s+2}, \|a_n^3(t)\|_s, \|\partial_t a_n(t)\|_{s+2}, \|\partial_t^2 a_n(t)\|_{s+2} \right\}, \\ \tilde{\mathcal{L}} &= \max_{t \in [0, T]} \left\{ \|\alpha(t)\|_{s+4}, \|\alpha^2(t)\|_{s+2}, \|\alpha^3(t)\|_s, \|\partial_t \alpha(t)\|_{s+2}, \|\partial_t^2 \alpha(t)\|_{s+2} \right\}. \end{aligned}$$

## 4.2.1 Dominujący czynnik błędu obcięcia rozwinięcia Magnusa

W zaprezentowanych poniżej obliczeniach analizujemy i szacujemy wyrażenia stanowiące dominujący czynnik błędu obcięcia rozwinięcia Magnusa (4.3). Dokładne obliczenia komutatorów występujących w wyrażeniach (4.6)–(4.7) rozwinięcia Magnusa obliczone są w Dodatku 4.5.1.

Wyrażenie  $\Theta_3(t + h, t)$  można przedstawić w postaci

$$\Theta_3(t + h, t) = \frac{1}{6} \int_0^h \int_0^{t_1} \int_0^{t_2} \begin{bmatrix} 0 & H_1 + H_3 \\ H_2 + H_4 & 0 \end{bmatrix} dt_3 dt_2 dt_1,$$

gdzie

$$\begin{aligned} H_1 + H_3 &= 2[-f(t_1) + 2f(t_2) - f(t_3)]; \\ H_2 + H_4 &= \Delta [f(t_3) - 2f(t_2) + f(t_1)] + [f(t_3) - 2f(t_2) + f(t_1)]\Delta \\ &\quad + 4f(t_1)f(t_3) - 2f(t_1)f(t_2) - 2f(t_3)f(t_2). \end{aligned} \quad (4.20)$$

W poniższych oszacowaniach oddzielamy części nieoscylujące (zawierające tylko funkcję  $\alpha$ ) od oscylujących otrzymując

$$\begin{aligned} \left| \int_0^h \int_0^{t_1} \int_0^{t_2} (H_1 + H_3) dt_3 dt_2 dt_1 \right| &\leq \left| \int_0^h \int_0^{t_1} \int_0^{t_2} 2[-\alpha(t_1) + 2\alpha(t_2) - \alpha(t_3)] dt_3 dt_2 dt_1 \right| \\ &\quad + \sum_{|n| \leq N} \left| \int_0^h \int_0^{t_1} \int_0^{t_2} 2[-a_n(t_1)e^{i\omega_n t_1} + 2a_n(t_2)e^{i\omega_n t_2} - a_n(t_3)e^{i\omega_n t_3}] dt_3 dt_2 dt_1 \right| \\ &\leq Ch^5 \max_{t \in [0, h]} |\alpha''(t)| + C \tilde{\mathcal{A}} \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\} \\ &\leq Ch^5 \tilde{\mathcal{L}} + C \tilde{\mathcal{A}} \mathcal{E}_\omega^h. \end{aligned} \quad (4.21)$$

Oszacowanie pierwszej sumy można uzyskać, rozwijając funkcje  $\alpha(t_i), i = 1, 2, 3$  w szereg Taylora w otoczeniu 0. Wówczas, istnieją takie  $\xi_1, \xi_2, \xi_3 \in [0, h]$ , że

$$\begin{aligned} \left| \int_0^h \int_0^{t_1} \int_0^{t_2} [\alpha(t_1) - 2\alpha(t_2) + \alpha(t_3)] dt_3 dt_2 dt_1 \right| \\ = \left| \int_0^h \int_0^{t_1} \int_0^{t_2} \alpha'(0)(t_1 - 2t_2 + t_3) dt_3 dt_2 dt_1 + \int_0^h \int_0^{t_1} \int_0^{t_2} t_1^2 \alpha''(\xi_1) - 2t_2^2 \alpha''(\xi_2) + t_3^2 \alpha''(\xi_3) dt_3 dt_2 dt_1 \right| \\ \leq 4h^5 \max_{t \in [0, T]} |\alpha''(t)|. \end{aligned}$$

Ostateczne oszacowanie mnożone jest przez  $h^5$  ponieważ pierwsza potrójna całka redukuje się do zera.

Więcej uwagi należy poświęcić przypadkowi, gdzie w naszym wyrażeniu występuje operator Laplace'a  $\Delta$ . Należy pamiętać, że będzie on działał nie tylko na funkcję  $f$ , ale także na rozwiązanie  $\psi$ , stąd też musimy podnieść nasze wymagania dotyczące regularności z  $\|\cdot\|_s$  do  $\|\cdot\|_{s+2}$ , aby mieć pewność, że istnieją drugie pochodne względem zmiennej przestrzennej z funkcji  $f$  i z rozwiązania  $\psi$ . Wówczas, dla pewnej odpowiednio gładkiej funkcji  $\varphi$  oszacowania w normie Sobolewa  $H^s$  przyjmują postać

$$\begin{aligned} \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} (H_2 + H_4) dt_3 dt_2 dt_1 \varphi \right\|_s &\leq \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} \Delta [f(t_1) - 2f(t_2) + f(t_3)] dt_3 dt_2 dt_1 \varphi \right\|_s \\ &\quad + \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} [f(t_1) - 2f(t_2) + f(t_3)] \Delta dt_3 dt_2 dt_1 \varphi \right\|_s \\ &\quad + \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} [4f(t_1)f(t_3) - 2f(t_1)f(t_2) - 2f(t_3)f(t_2)] dt_3 dt_2 dt_1 \varphi \right\|_s \\ &\leq Ch^5 \tilde{\mathcal{L}} \|\varphi\|_{s+2} + C \tilde{\mathcal{A}} \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\} \|\varphi\|_{s+2} + C \tilde{\mathcal{L}} \tilde{\mathcal{A}} \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\} \|\varphi\|_s \\ &\leq Ch^5 \tilde{\mathcal{L}} \|\varphi\|_{s+2} + C \tilde{\mathcal{A}} \mathcal{E}_\omega^h \|\varphi\|_{s+2} + C \tilde{\mathcal{L}} \tilde{\mathcal{A}} \mathcal{E}_\omega^h \|\varphi\|_s \end{aligned}$$

Aby wyjaśnić jak otrzymujemy powyższą nierówność, pokażemy dokładne oszacowanie jednego z komponentów sumy  $H_2 + H_4$

$$\begin{aligned} & \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} \Delta(f(t_1) - 2f(t_2) + f(t_3)) dt_3 dt_2 dt_1 \varphi \right\|_s \\ & \leq C \max_{t \in [0, T]} \left\{ h^5 \|\partial_t^2 \alpha(t) \Delta \varphi\|_s, h^5 \|\nabla \partial_t^2 \alpha(t) \nabla \varphi\|_s, h^5 \|\Delta \partial_t^2 \alpha(t) \varphi\|_s \right\} \\ & + C \sum_{|n| \leq N} \min \left\{ h^3 \max_{t \in [0, T]} \{ \|a_n(t) \Delta \varphi\|_s, \|\nabla a_n(t) \nabla \varphi\|_s, \|\Delta a_n(t) \varphi\|_s \}, \right. \\ & \quad \frac{h^2}{\omega_{\min}} \max_{t \in [0, T]} \{ \|\partial_t a_n(t) \Delta \varphi\|_s, \|\nabla \partial_t a_n(t) \nabla \varphi\|_s, \|\Delta \partial_t a_n(t) \varphi\|_s \}, \\ & \quad \left. h^5 \omega_{\max}^2 \max_{t \in [0, T]} \left\{ \|\partial_t^2 a_n(t) \Delta \varphi\|_s, \|\nabla \partial_t^2 a_n(t) \nabla \varphi\|_s, \|\Delta \partial_t^2 a_n(t) \varphi\|_s \right\} \right\}. \end{aligned}$$

Kolejną składową dominującego czynnika błędu obcięcia rozwinięcia Magnusa jest składnik  $\Theta_4$ , który można przedstawić w postaci

$$\Theta_4(t+h, t) = \frac{1}{12} \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \begin{bmatrix} \mathcal{H}_1 & 0 \\ 0 & \mathcal{H}_2 \end{bmatrix} dt_4 dt_3 dt_2 dt_1,$$

gdzie

$$\begin{aligned} \mathcal{H}_1 &= \Delta[2f(t_2) - 2f(t_3)] + 3[2f(t_2) - 2f(t_3)]\Delta + 4f(t_4)[f(t_2) - f(t_3)] + 4f(t_1)[f(t_2) - f(t_3)]; \\ \mathcal{H}_2 &= -3\Delta[2f(t_2) - 2f(t_3)] - [2f(t_2) - 2f(t_3)]\Delta - 4f(t_4)[f(t_2) - f(t_3)] - 4f(t_1)[f(t_2) - f(t_3)]. \end{aligned}$$

Szczegółowe obliczenia komutatorów występujących w składniku  $\Theta_4$  rozwinięcia Magnusa (4.3) prezentujemy w Dodatku 4.5.1.

Zwróćmy uwagę, że w każdym z wyrażeń sumy w  $\mathcal{H}_1$  występuje różnica funkcji  $f$  w dwóch różnych punktach czasu. Na mocy Twierdzenia Taylora otrzymujemy

$$f(t_i) - f(t_j) = h(f'(\xi_i) - f'(\xi_j)), \quad \text{for } 0 < t_i, t_j < h, \quad \text{oraz pewnych } 0 < \xi_i, \xi_j < h.$$

Dodatkowo, powołując się na Twierdzenie 1 wnioskujemy, że dla odpowiednio gładkiej funkcji  $\varphi$  otrzymujemy oszacowanie

$$\begin{aligned} \left\| \int_0^h \int_0^{t_1} \int_0^{t_2} \int_0^{t_3} \mathcal{H}_1 dt_4 dt_3 dt_2 dt_1 \varphi \right\|_s & \leq C h^5 \tilde{\mathcal{L}} \|\varphi\|_{s+2} + C \tilde{\mathcal{A}} \min \left\{ h^4, \frac{h^3}{\omega_{\min}}, h^5 \omega_{\max} \right\} \|\varphi\|_{s+2} \\ & + C \tilde{\mathcal{L}} \tilde{\mathcal{A}} \min \left\{ h^4, \frac{h^3}{\omega_{\min}}, h^5 \omega_{\max} \right\} \|\varphi\|_s. \end{aligned}$$

Oszacowanie całki poczwórnej  $\mathcal{H}_2$  wyprowadza się analogicznie.

Składniki  $\Theta_k(t+h, t)$ ,  $k \geq 5$  rozwinięcia Magnusa (4.3) podnoszą krotność zagnieżdżonych całek przynajmniej o  $(k-4)$  (z wyrażeń podobnych do  $\mathcal{H}_1$  i  $\mathcal{H}_2$ ), które powoduje zmniejszenie wielkości wyrażenia  $\Theta_k(t+h, t)$   $h^{k-4}$  razy. Z tego powodu  $\Theta_k(t+h, t)$ ,  $k \geq 5$  nie zaliczają się do dominujących czynników błędu obcięcia rozwinięcia Magnusa (4.3).

Podsumowując, dominujący czynnik błędu wynikający z obcięcia rozwinięcia Magnusa (4.3) równy sumie  $\Theta_3(t+h, t) + \Theta_4(t+h, t)$ , jest ograniczony przez wyrażenie

$$C \left[ \tilde{\mathcal{L}} \tilde{\mathcal{A}} \mathcal{E}_\omega^h \|\varphi\|_s + \left( \tilde{\mathcal{A}} \mathcal{E}_\omega^h + h^5 \tilde{\mathcal{L}} \right) \|\varphi\|_{s+2} \right].$$

## 4.2.2 Dominujący czynnik błędu dla splittingu Stranga

W zaprezentowanych poniżej obliczeniach analizujemy i szacujemy wyrażenia stanowiące dominujące czynniki błędu wynikające z zastosowania splittingu Stranga (4.10). Błąd tego splittingu zależy od dwóch komutatorów,  $[Y, [Y, X]]$  oraz  $[X, [X, Y]]$ . Dokładne obliczenia tych wyrażeń przedstawiono w Dodatku 4.5.2

Wprowadzając oznaczenia

$$X = \int_0^h \int_0^{t_1} [A(t_1), A(t_2)] dt_2 dt_1 = \int_0^h \int_0^{t_1} \begin{bmatrix} f(t_1) - f(t_2) & 0 \\ 0 & f(t_2) - f(t_1) \end{bmatrix} dt_2 dt_1,$$

$$Y = \int_0^h \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix} dt_1,$$

dwukrotnie zagnieżdżone komutatory  $[Y, [Y, X]]$  oraz  $[X, [X, Y]]$  przyjmują postać

$$[Y, [Y, X]] = \begin{bmatrix} -h^2 \Delta \mathcal{F} - 3h^2 \mathcal{F} \Delta - 4h \mathcal{F} \mathcal{F} & 0 \\ 0 & 3h^2 \Delta \mathcal{F} + h^2 \mathcal{F} \Delta + 4h \mathcal{F} \mathcal{F} \end{bmatrix};$$

$$[X, [X, Y]] = \begin{bmatrix} 0 & 4h \mathcal{F}^2 \\ 2h \mathcal{F} \Delta \mathcal{F} + h \mathcal{F}^2 \Delta + h \Delta \mathcal{F}^2 + 4 \mathcal{F}^2 \mathcal{F} & 0 \end{bmatrix},$$

gdzie

$$F = \int_0^h f(t + t_1) dt_1 \quad \text{oraz} \quad \mathcal{F} = \int_0^h \int_0^{t_1} [f(t + t_2) - f(t + t_1)] dt_2 dt_1.$$

Na mocy Twierdzenia 1 otrzymujemy następujące oszacowania

$$|F| \leq Ch \max_{t \in [0, T]} \{\|\alpha(t)\|_s\} + Ch \max_{t \in [0, T]} \{\|a_n(t)\|_s, \|\partial_t a_n(t)\|_s\} \leq Ch(\tilde{\mathcal{L}} + \tilde{\mathcal{A}}),$$

$$|\mathcal{F}| \leq Ch^2 \max_{t \in [0, T]} \{h \|\partial_t \alpha(t)\|_s\} + C \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \max_{t \in [0, T]} \{\|a_n(t)\|_s, \|\partial_t a_n(t)\|_s\}$$

$$\leq Ch^3 \tilde{\mathcal{L}} + C \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\},$$

$$\|\Delta \mathcal{F} \varphi\|_s \leq Ch^2 \max_{t \in [0, T]} \{h \|\partial_t \alpha(t) \Delta \varphi\|_s, h \|\nabla \partial_t \alpha(t) \nabla \varphi\|_s, h \|\Delta \partial_t \alpha(t) \varphi\|_s\}$$

$$+ C \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \|\varphi\|_{s+2}$$

$$\leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right) \|\varphi\|_{s+2},$$

$$\|\mathcal{F} \Delta \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right) \|\varphi\|_{s+2},$$

$$\|\mathcal{F} \mathcal{F} \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right) (h \tilde{\mathcal{L}} + \tilde{\mathcal{A}} h) \|\varphi\|_s,$$

$$\|\mathcal{F} \Delta \mathcal{F} \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right)^2 \|\varphi\|_{s+2},$$

$$\|\mathcal{F}^2 \Delta \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right)^2 \|\varphi\|_{s+2},$$

$$\|\Delta \mathcal{F}^2 \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right)^2 \|\varphi\|_{s+2},$$

$$\|\mathcal{F}^2 \mathcal{F} \varphi\|_s \leq C \left( h^3 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \min \left\{ h^2, \frac{h}{\omega_{\min}}, h^3 \omega_{\max} \right\} \right)^2 (h \tilde{\mathcal{L}} + \tilde{\mathcal{A}} h) \|\varphi\|_s.$$

Ostatecznie zauważamy, że dominujący czynnik błędu splittingu Stranga jest ograniczony przez

$$C \left[ \left( h^5 \tilde{\mathcal{L}}^2 + h^5 \tilde{\mathcal{L}} \tilde{\mathcal{A}} + \tilde{\mathcal{A}}^2 \mathcal{E}_\omega^h + \tilde{\mathcal{L}} \tilde{\mathcal{A}} \mathcal{E}_\omega^h \right) \|\varphi\|_s + \left( h^5 \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \mathcal{E}_\omega^h \right) \|\varphi\|_{s+2} \right].$$

Zgodnie z wynikami pracy [Jahnke and Lubich, 2000] błąd splittingu Stranga zależy tylko od podwójnie zagnieżdżonych komutatorów  $[Y, [Y, X]]$  oraz  $[X, [X, Y]]$ .

### 4.2.3 Dominujące czynniki błędu dla splittingu 4-go rzędu (4.13)

Zaproponowane w podrozdziale 4.1.5 metody numeryczne  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$  różnią się wyborem macierzy  $X$  oraz  $Y$  występujących w splittingu (4.13).

Przypomnijmy, że aby otrzymać schemat  $\Gamma_1^{[4]}$  przyjęliśmy, że  $X = \begin{bmatrix} 0 & 0 \\ D & 0 \end{bmatrix}$  oraz  $Y = \begin{bmatrix} 0 & h \\ F & 0 \end{bmatrix}$ , gdzie  $D := h\Delta$  oraz  $F := \int_0^h f(t + t_1) dt_1$ .

Nietrudno sprawdzić, że dominujący czynnik błędu przy takim doborze macierzy  $X$  i  $Y$  przyjmuje postać

$$\begin{aligned} [Y, [Y, [Y, [Y, X]]]] &= \begin{bmatrix} 0 & -4h^3FD - h^3DF \\ h^2F^2D + 6h^2FDF + h^2DF^2 & 0 \end{bmatrix} \\ [X, [X, [X, [X, Y]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [X, [Y, [Y, [Y, X]]]] &= \begin{bmatrix} 0 & 0 \\ 2h^2DFD + 3h^2D^2F + 3h^2FD^2 & 0 \end{bmatrix} \\ [Y, [X, [X, [X, Y]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [Y, [X, [Y, [X, Y]]]] &= \begin{bmatrix} 0 & -4h^3D^2 \\ 2h^2(FD^2 + D^2F) & 0 \end{bmatrix} \\ [X, [Y, [X, [Y, X]]]] &= \begin{bmatrix} 0 & 0 \\ 4h^2D^3 & 0 \end{bmatrix} \end{aligned}$$

W schemacie  $\Gamma_2^{[4]}$  przyjęliśmy natomiast, że  $X = \begin{bmatrix} 0 & 0 \\ D + F & 0 \end{bmatrix}$  oraz  $Y = \begin{bmatrix} 0 & h \\ 0 & 0 \end{bmatrix}$ , zatem dominujący czynnik błędu przyjmie postać

$$\begin{aligned} [Y, [Y, [Y, [Y, X]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [X, [X, [X, [X, Y]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [X, [Y, [Y, [Y, X]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [Y, [X, [X, [X, Y]]]] &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ [Y, [X, [Y, [X, Y]]]] &= \begin{bmatrix} 0 & 4h^3(D + F)^2 \\ 0 & 0 \end{bmatrix} \\ [X, [Y, [X, [Y, X]]]] &= \begin{bmatrix} 0 & 0 \\ -4h^2(D + F)^3 & 0 \end{bmatrix} \end{aligned}$$

Wnioskiem z powyższych obliczeń jest, że dominujące czynniki błędu splittingu (4.13) są ograniczone przez

$$Ch^5 \left[ (\tilde{\mathcal{L}} + \tilde{\mathcal{A}} + (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^2) \|\varphi\|_{s+2} + (1 + \tilde{\mathcal{L}} + \tilde{\mathcal{A}}) \|\varphi\|_{s+4} + \|\varphi\|_{s+6} \right]$$

w przypadku schematu  $\Gamma_1^{[4]}$  oraz przez

$$Ch^5 \left[ \left( (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^2 + (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^3 \right) \|\varphi\|_s + \left( \tilde{\mathcal{L}} + \tilde{\mathcal{A}} + (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^2 \right) \|\varphi\|_{s+2} + \left( 1 + \tilde{\mathcal{L}} + \tilde{\mathcal{A}} \right) \|\varphi\|_{s+4} + \|\varphi\|_{s+6} \right]$$

dla schematu  $\Gamma_2^{[4]}$ .

### 4.3 Struktura błędu

Podrozdział 4.2 dotyczył oszacowań elementów stanowiących dominujące czynniki błędu wynikające z obcięcia rozwinięcia Magnusa (4.3) do dwóch pierwszych wyrazów, zastosowania do nich splittingu Stranga (4.10) oraz wykorzystania splittingu (4.13) do obliczenia składnika wewnętrznego splittingu Stranga. W celu wyznaczenia całkowitego błędu metod numerycznych  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$  potrzebujemy oszacować różnicę między rozwiązaniem dokładnym a przybliżonym. Zaczniemy od następującej Uwagi, której dowód jest trywialny:

**Uwaga 16** Załóżmy, że macierz  $Z$  wymiaru  $2 \times 2$  można wyrazić w postaci  $Z = X + \xi \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ , gdzie  $\xi \in \mathbf{R}$  oznacza wielkość zaburzenia macierzy  $Z$ . Wówczas

$$e^Z = e^X + \mathcal{O}(\xi) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

W poniższym twierdzeniu prezentujemy lokalny błąd metod rozumiejąc, że w przypadku odpowiednio małych kroków jest on spójny z dominującym czynnikiem błędu. Nadużywając notacji, nie wprowadzamy rozróżnienia pomiędzy funkcjami skalarnymi i wektorowymi, i używamy normy Sobolewa  $H^{s+k}$ ,  $\|\cdot\|_{s+k}$ , w rozumieniu punktowym.

**Twierdzenie 2** Załóżmy, że spełnione są Założenia 1. Przez  $\phi^t$  oznaczmy dokładny potok, tj.  $z(t) = \phi^t(z_0)$  oraz oznaczamy przez  $\Phi^h$  potok numeryczny, tj.

$$z^{n+1} = \Phi^h(z^n)$$

gdzie  $\Phi^h$  odpowiada algorytmowi  $\Gamma_1^{[4]}$  lub  $\Gamma_2^{[4]}$ . Wówczas, dla  $\varphi \in H^{s+6}(\mathbb{T}^d)$  otrzymujemy

$$\begin{aligned} \|\phi^h(\varphi) - \Phi^h(\varphi)\|_s &\leq C \left( (\tilde{\mathcal{L}} \tilde{\mathcal{A}} + \tilde{\mathcal{A}}^2) \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\} + h^5 (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^2 + h^5 (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^3 \right) \|\varphi\|_s \\ &\quad + C \left( \tilde{\mathcal{A}} \min \left\{ h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2 \right\} + h^5 (\tilde{\mathcal{L}} + \tilde{\mathcal{A}} + (\tilde{\mathcal{L}} + \tilde{\mathcal{A}})^2) \right) \|\varphi\|_{s+2} \\ &\quad + C \left( h^5 (1 + \tilde{\mathcal{L}} + \tilde{\mathcal{A}}) \right) \|\varphi\|_{s+4} \\ &\quad + Ch^5 \|\varphi\|_{s+6} \end{aligned}$$

gdzie stałe  $\tilde{\mathcal{A}}, \tilde{\mathcal{L}}$  zdefiniowane są w Definicji 1.

Dowód twierdzenia wynika bezpośrednio z oszacowań w podrozdziale 4.2 oraz z Uwagi 16.

Z powyższego Twierdzenia, wynika że lokalny błąd aproksymacji numerycznej skaluje jak

$$\mathcal{O}(h^5 + \min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$$

oraz jak

$$\mathcal{O}(\min\{h^3, h^2/\omega_{\min}, h^5\omega_{\max}^2\})$$

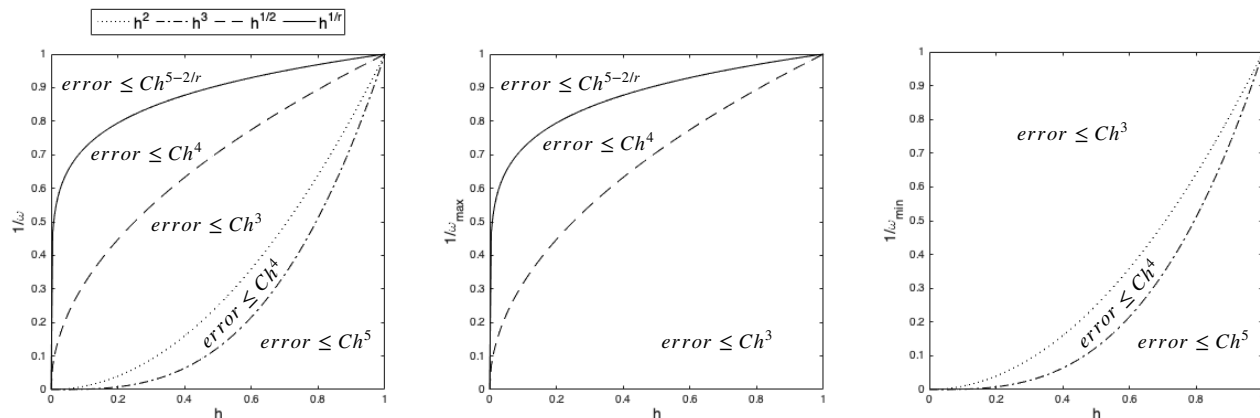
w szczególnym przypadku  $\alpha(\mathbf{x}, t) \equiv 0$ .

Zauważmy, że jeśli funkcja masy  $f(\mathbf{x}, t)$  nie zawiera składnika oscylującego, czyli  $\forall_n a_n(\mathbf{x}, t) \equiv 0$ , to metody  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$  są rzędu 5 (lokalnie). Jeśli jednak niektóre współczynniki  $a_n$  są niezerowe, to wystarczy

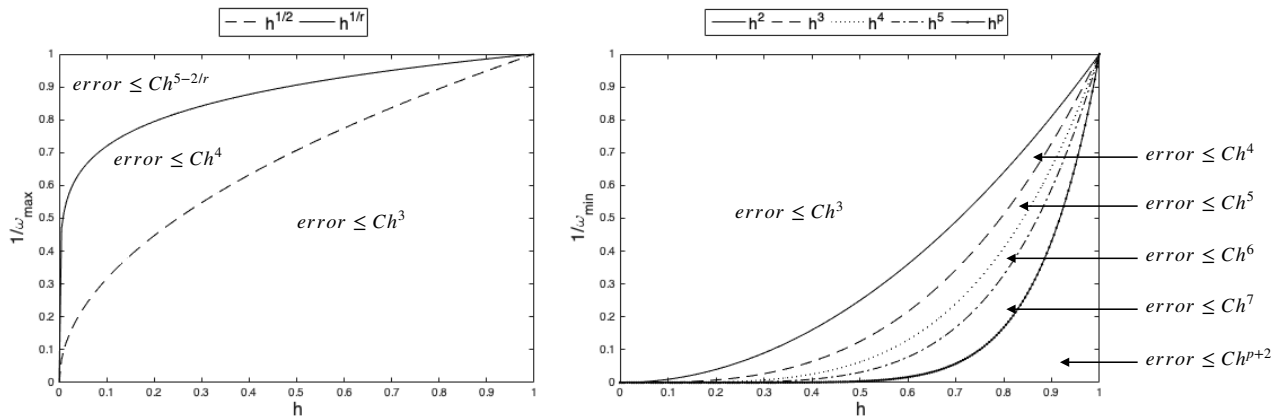


dobrac kroki czasowe  $h \gtrsim 1/\sqrt[3]{\min_n |\omega_n|}$ , aby obie metody były rzędu 5-go (lokalnie). Zjawisko to zaprezentujemy w przykładach numerycznych w podrozdziale 4.4. Kolejną ciekawą obserwacją jest to, że dla czysto oscylacyjnej funkcji  $f(t)$  (czyli gdy  $\alpha(t) \equiv 0$ ) możemy otrzymać metodę rzędu  $r$ , czyli błędu lokalnego  $\mathcal{O}(h^{r+1})$  dobierając krok czasowy  $h \gtrsim 1/\sqrt[r-1]{\min_n |\omega_n|}$ ,  $r$  dowolnie duże, Rysunek 4.2.

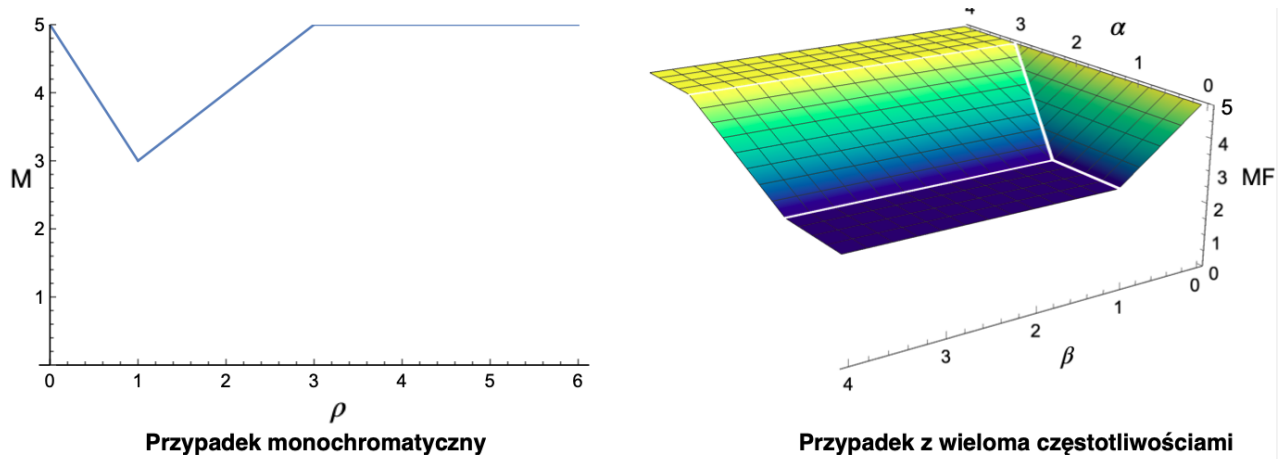
Zależność między wielkością oscylacji  $\omega_n$  a krokiem czasowym  $h$  ilustrują Rysunki 4.1–4.3.



RYSUNEK 4.1: Błąd lokalny proponowanych metod numerycznych jest uzależniony od stosunku wielkości oscylacji  $\omega_{\min}$  i  $\omega_{\max}$  i kroku czasowego  $h$ . Lewy rysunek ilustruje dokładność metod uzyskaną dla przypadku monochromatycznego, czyli gdy mamy tylko jedną częstotliwość  $\omega$  w funkcji masy  $f(\mathbf{x}, t)$  ( $\omega_{\min} = \omega_{\max}$ ). Środkowy rysunek pokazuje dokładność metod w odniesieniu do zależności pomiędzy największą częstotliwością  $\omega_{\max}$  a krokiem czasowym  $h$ , natomiast prawy rysunek ilustruje dokładność metod w odniesieniu do stosunku najmniejszej częstotliwości  $\omega_{\min}$  i kroku czasowego  $h$ . Ostateczna dokładność (przedstawiona na rysunku 4.3) zależy jednak od obu:  $\omega_{\min}$  oraz  $\omega_{\max}$ . Na powyższych wykresach zakładamy, że  $\alpha(\mathbf{x}, t) \neq 0$ .



RYSUNEK 4.2: W przeciwieństwie do Rysunku 4.1, zakładamy teraz, że  $\alpha(\mathbf{x}, t) = 0$ . Lewy rysunek ilustruje osiągalną dokładność metod w zależności od stosunku największej częstotliwości  $\omega_{\max}$  do kroku czasowego  $h$ , natomiast prawy rysunek ilustruje osiągalną dokładność metod w zależności od stosunku najmniejszej częstotliwości  $\omega_{\min}$  do czasu krok  $h$ .



RYSUNEK 4.3: Lewy rysunek ilustruje dokładność uzyskaną w przypadku monochromatycznym ( $\omega_{\min} = \omega_{\max}$ ) i należy ją rozumieć w następujący sposób: dla  $\omega_{\min} = h^{-\rho}$  błąd aproksymacji skaluje jak  $\mathcal{O}(h^M)$ , gdzie  $M = \max\{3, 2 + \rho, 5 - 2\rho\}$ . Rysunek po prawej stronie obejmuje przypadek wielu częstotliwości, gdzie dla  $\omega_{\min} = h^{-\alpha}$ ,  $\omega_{\max} = h^{-\beta}$  dokładność metody wynosi  $\mathcal{O}(h^{MF})$ , a  $MF = \max\{3, 2 + \alpha, 5 - 2\beta\}$ .

## 4.4 Przykłady numeryczne

W tym podrozdziale porównujemy nowo skonstruowane metody  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$  z metodami przedstawionymi w poprzednich Rozdziałach:

- **BBCK**<sup>[4]</sup>: metoda 4-go rzędu  $\Sigma_{3c}^{[4]}$  z pracy [Bader et al., 2019];
- **BBCK**<sup>[6]</sup>: metoda 6-go rzędu  $\Sigma_{5c}^{[6]}$  z pracy [Bader et al., 2019];
- **Asympt**<sup>[3]</sup>: metoda asymptotyczna 3-go rzędu z pracy [Condon et al., 2021], zaprezentowana w Rozdziale 2;
- $\Xi$ <sup>[3]</sup>: metoda 3-go rzędu z pracy [Kropielnicka and Lademann, 2022], zaprezentowana w Rozdziale 3.

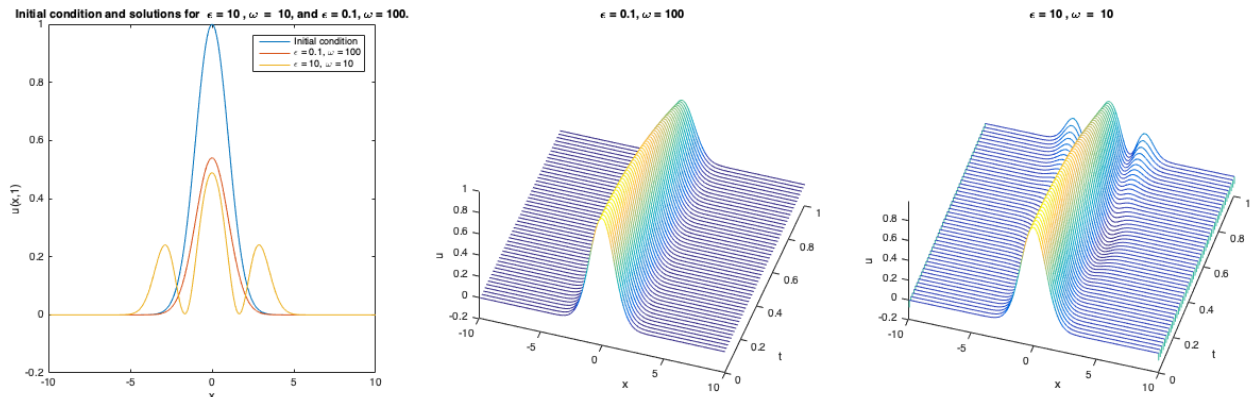
Rozwiązanie referencyjne obliczane jest przy użyciu metody 6-tego rzędu [Blanes et al., 2009], z krokiem czasowym  $h = 10^{-6}$ . Następnie, stosujemy podane metody z różnymi krokami czasowymi i mierzymy błąd w końcowym czasie  $T$  w normie  $\ell_2$ . Porównujemy zależność błędu poszczególnych metod od wielkości kroku czasowego  $h$  oraz w zależności do czasu obliczeniowego, liczonego w sekundach. Błąd ten jest wykreślany w podwójnej skali logarytmicznej (log-log). Laplacian jest dyskretyzowany metodą Fouriera, [Kopriva, 2009, Trefethen, 2000] z  $M = 200$  modami.

*Przykład 1.*

W pierwszym przykładzie rozważmy równanie Kleina-Gordona postaci

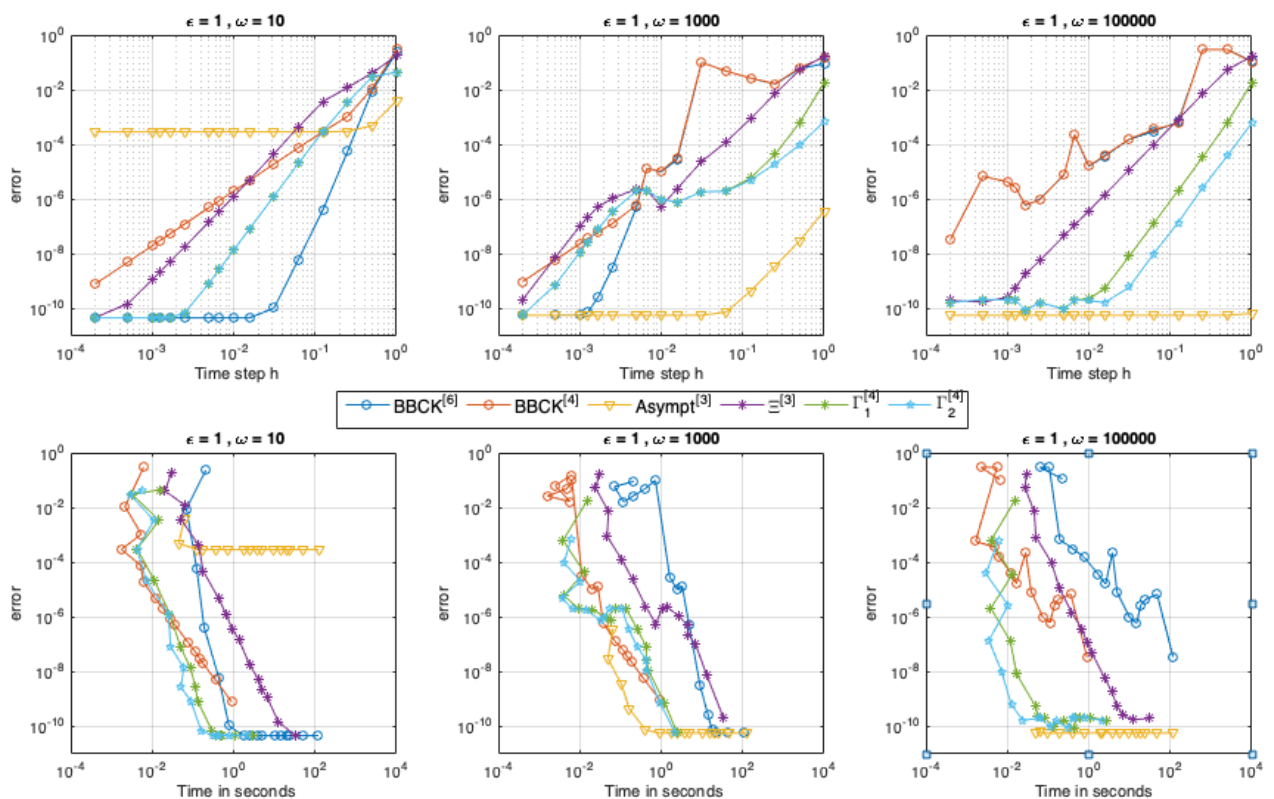
$$\begin{aligned}
 \partial_t^2 u &= \partial_x^2 u - (1 + \varepsilon \cos(\omega t)) x^2 u, & x \in [-10, 10], & t \in [0, 1], \\
 u(x, 0) &= e^{-\frac{x^2}{2}}, & \partial_t u(x, 0) &= 0, \\
 u(-10, t) &= u(10, t).
 \end{aligned}
 \tag{4.22}$$

Na pierwszym od lewej wykresie Rysunku 4.4 przedstawiamy warunek początkowy (niebieska linia), rozwiązanie w końcowym kroku czasowym dla  $\varepsilon = 10, \omega = 10$  (żółta linia) oraz rozwiązanie w końcowym kroku czasowym dla  $\varepsilon = 0, 1, \omega = 100$  (linia czerwona). Kolejne dwa wykresy przedstawiają ewolucję rozwiązań w czasie dla  $\varepsilon = 0.1, \omega = 100$  oraz dla  $\varepsilon = 10, \omega = 10$ .



RYSUNEK 4.4: Warunek początkowy, rozwiązania w końcowym kroku czasowym i ewolucja rozwiązań w czasie równania (4.22) dla dwóch par współczynników  $\epsilon$  i  $\omega$ .

Porównanie kosztu obliczeniowego i dokładności metod dla równania (4.22) przedstawiamy na Rysunkach 4.5 i 4.6. Przede wszystkim zauważmy, że metoda asymptotyczna  $\text{Asympt}^{[3]}$  przedstawiona w Rozdziale 2 jest niezrównana w przypadku bardzo wysoko-oscyłującej funkcji  $f(x,t)$ . Wynika to z tego, że metoda asymptotyczna została zaprojektowana specjalnie dla tego typu równań. Dla małych  $\omega$  metoda asymptotyczna  $\text{Asympt}^{[3]}$  jest nieefektywna.

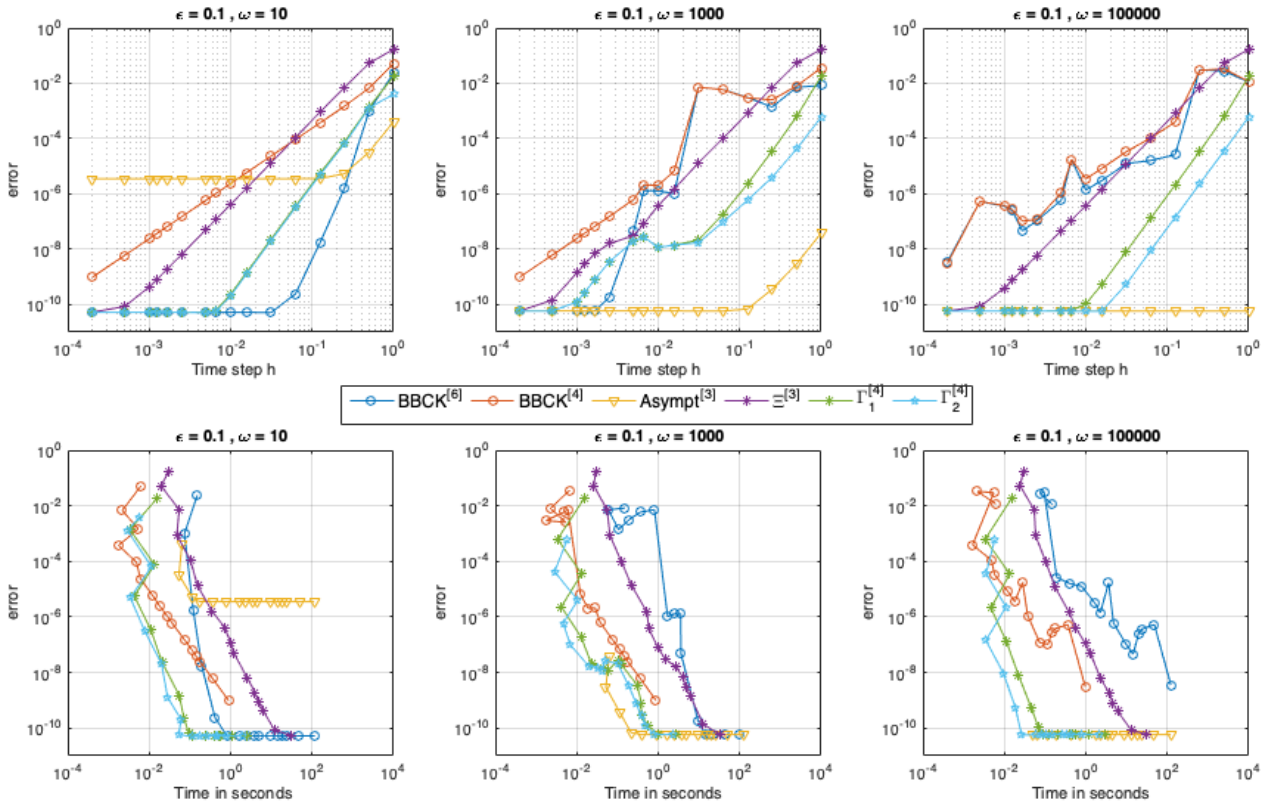


RYSUNEK 4.5: Porównanie rzędów (pierwszy wiersz) i kosztu obliczeniowego w sekundach (drugi wiersz) dla równania (4.22) dla  $\epsilon = 1$ .

W przypadku stosunkowo małych oscylacji, gdy  $\omega = 10$ , metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  zachowują się przewidywalnie - osiągają gorszy błąd niż metoda szóstego rzędu  $\text{BBCK}^{[6]}$ , ale znacznie mniejszy błąd niż metoda czwartego rzędu  $\text{BBCK}^{[4]}$ , metoda asymptotyczna  $\text{Asympt}^{[3]}$  oraz metoda  $\Xi^{[3]}$ . Wraz ze wzrostem oscylacji  $\omega = 1000$ , metoda asymptotyczna  $\text{Asympt}^{[3]}$  zaczyna bardzo szybko osiągać małe błędy aproksymacji, nawet dla dużych kroków czasowych  $h$ . Metody  $\text{BBCK}^{[4]}$  i  $\text{BBCK}^{[6]}$  wymagają natomiast małych kroków czasowych aby sprostać oscylacjom, podczas gdy nowe metody dają oczekiwane błędy dla wszystkich kroków czasowych

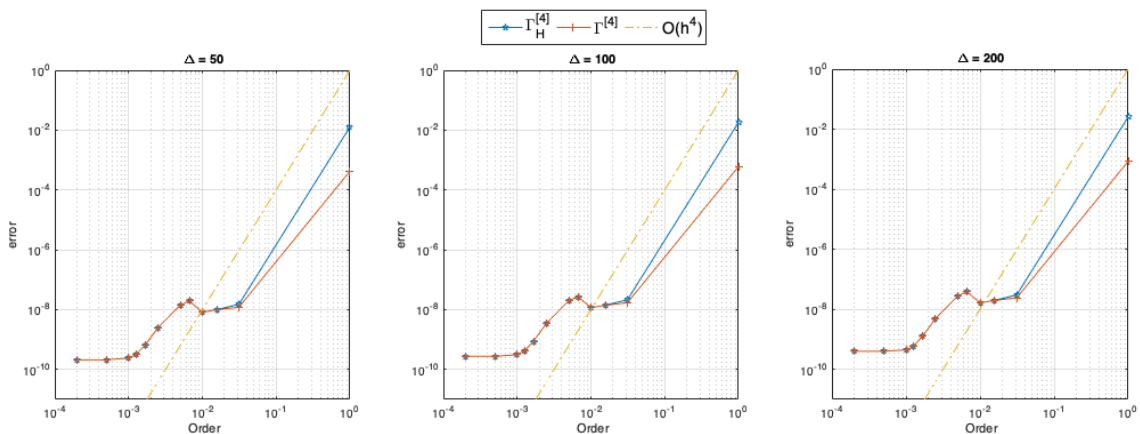
h. Metoda  $\Xi^{[3]}$  ma trzeci rząd zbieżności niezależnie od wielkości oscylacji, ale jest bardziej kosztowna od metod  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$

W przypadku dużych oscylacji w funkcji masy  $f(x, t)$ , metody  $\Xi^{[3]}$ ,  $\Gamma_1^{[4]}$  oraz  $\Gamma_2^{[4]}$  zachowują stabilny rząd zbieżności (trzeci i czwarty, odpowiednio), podczas gdy metody BBCK<sup>[4]</sup> i BBCK<sup>[6]</sup> wykazują ogromne wahania w rzędzie zbieżności, który zbliżony jest do rzędu drugiego.



RYSUNEK 4.6: Porównanie rzędów (pierwszy wiersz) i kosztu w sekundach (drugi wiersz) dla równania (4.22) dla  $\epsilon = 0.1$ .

Na Rysunku 4.7 przedstawiamy porównanie stałych błędów metod  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  w zależności od wyboru wielkości kroku przestrzennego  $\Delta_x$ . Jak widać na tym rysunku, wielkość kroku przestrzennego nie wpływa na dokładność przybliżenia.



RYSUNEK 4.7: Porównanie stałych błędów metod  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  dla równania (4.22) dla  $\epsilon = 0.1$  oraz  $\omega = 1000$ .

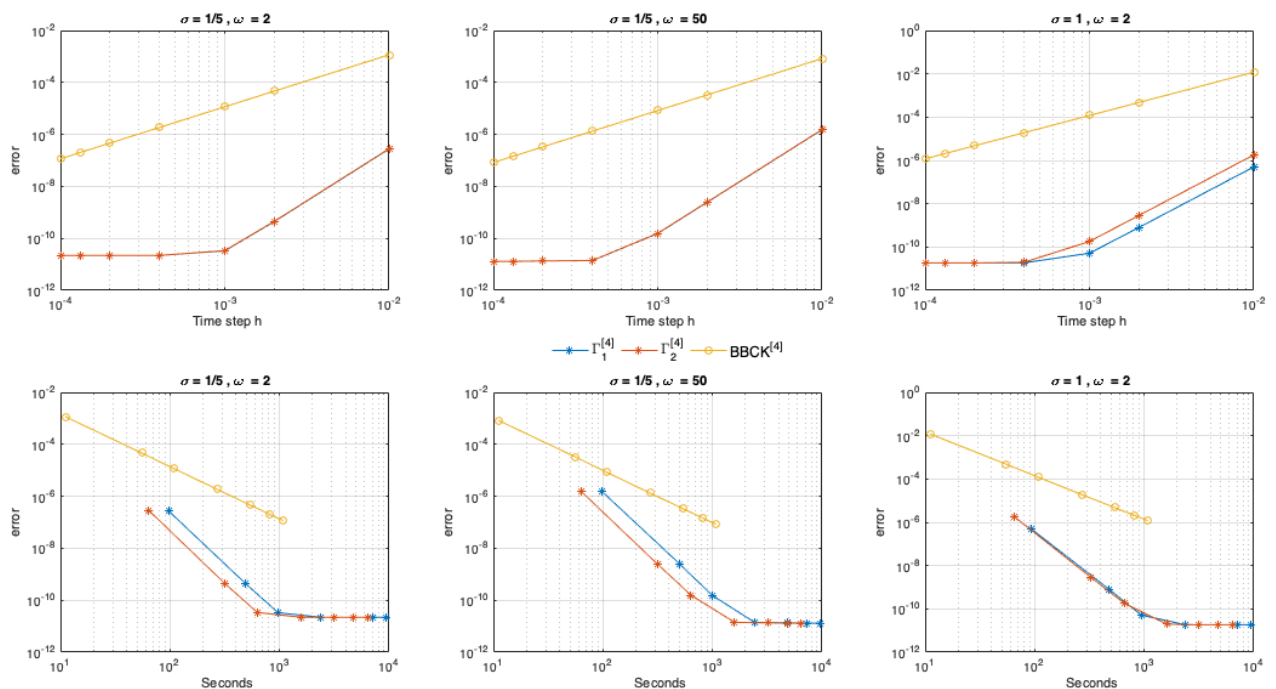
## Przykład 2.

W tym przykładzie rozważamy równanie falowe w dwóch wymiarach przestrzennych

$$\begin{aligned} \partial_t^2 u &= \Delta u - \sigma \left(1 + \frac{1}{5} \cos(\omega t)\right) x^2 y^2 u, \quad x, y \in [-\pi, \pi], \quad t \in [0, 1], \\ u(x, y, 0) &= e^{-\frac{1}{2}(x-3)^2} + e^{-\frac{1}{2}(x+3)^2} + e^{-\frac{1}{2}(y-3)^2} + e^{-\frac{1}{2}(y+3)^2}, \quad u'(x, y, 0) = 0; \\ u(-\pi, -\pi, t) &= u(\pi, \pi, t) = u(\pi, -\pi, t) = u(-\pi, \pi, t). \end{aligned} \quad (4.23)$$

Powyższy przykład służy pokazaniu, że metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  mogą być również używane do zagadnień w wielu wymiarach przestrzennych. Chociaż przedstawiamy przykład dwuwymiarowy, metody te można rozszerzyć na wyższe wymiary, ale wraz ze wzrostem liczby wymiarów przestrzennych  $d$ , wielkość macierzy wynikającej z semi-dyskretyzacji rośnie jak  $M^d$ , gdzie  $M$  rozmiarem siatki przestrzennej, zatem takie obliczenia wymagają komputerów o bardzo dużej mocy obliczeniowej.

Rysunek 4.8 przedstawia porównanie kosztu obliczeniowego i dokładności metod  $\Gamma_1^{[4]}$ ,  $\Gamma_2^{[4]}$  oraz BBCK<sup>[4]</sup> dla równania (4.23). Jak łatwo zauważyć, w przypadku dwuwymiarowym, metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  zachowują 4-ty rząd zbieżności. Ponadto, wykazują się stosunkowo niskim kosztem obliczeniowym.



RYSUNEK 4.8: Porównanie rzędów (pierwszy wiersz) i kosztu w sekundach (drugi wiersz) dla równania (4.23).

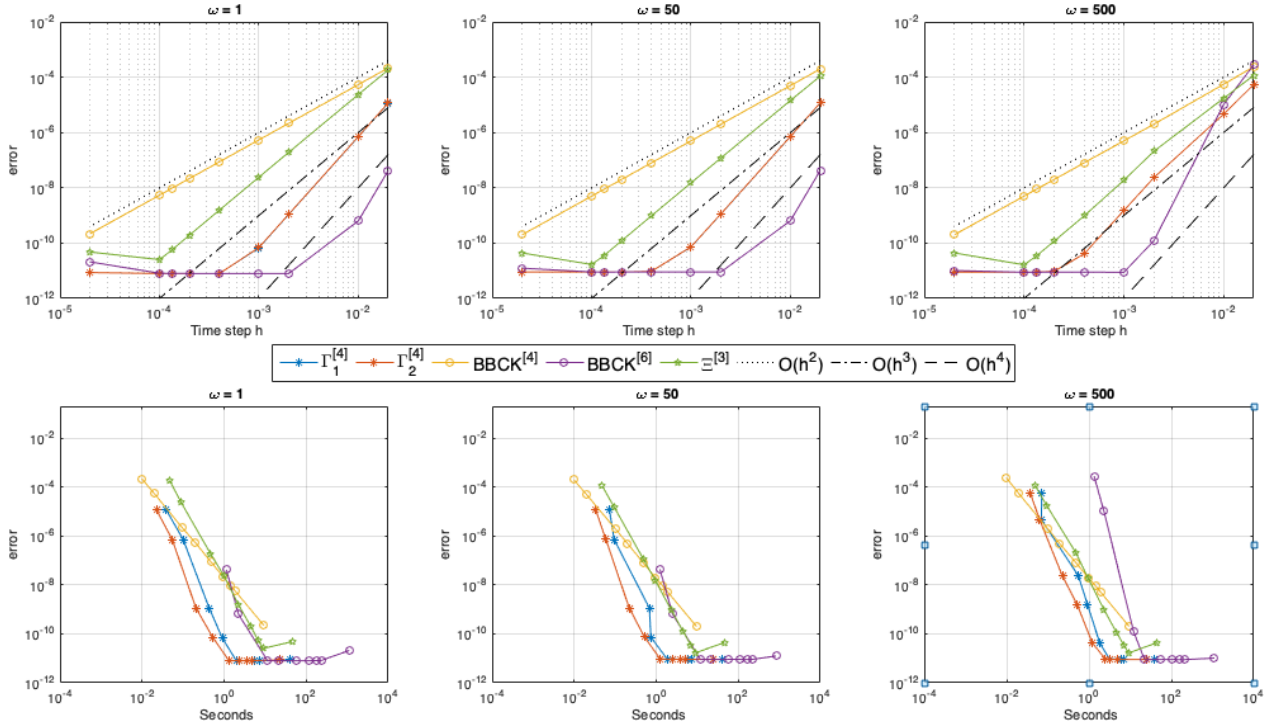
## Przykład 3.

W kolejnym przykładzie rozważamy równanie Kleina-Gordona postaci

$$\begin{aligned} \partial_t^2 u &= \partial_x^2 u - \frac{1}{1+t^2} \left(1 + \frac{1}{5} \cos(\omega t)\right) x^2 u, \quad x \in [-\pi, \pi], \quad t \in [0, 1]; \\ u(x, 0) &= e^{-\frac{1}{2}(x-3)^2} + e^{-\frac{1}{2}(x+3)^2}, \quad u'(x, 0) = 0; \\ u(-\pi, t) &= u(\pi, t). \end{aligned} \quad (4.24)$$

W przykładzie (4.24) rozważamy funkcję masy  $f(x, t)$ , w której część nieoscylicjąca  $\frac{-1}{1+t^2} x^2$  oraz część oscylująca  $\frac{-1}{5} \frac{1}{1+t^2} \cos(\omega t) x^2$  są zależne od czasu i przestrzeni. Porównania rzędu zbieżności oraz czasu obliczeniowego rozważanych metod zaprezentowane są na Rysunku 4.9. Metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  prezentują tę samą

dokładność i stałą błędu. Podobnie jak w przykładzie (4.22),  $\Gamma_1^{[4]}$  jest nieco bardziej kosztowna obliczeniowo niż  $\Gamma_2^{[4]}$ , ale obie metody są dokładniejsze niż metoda BBCK<sup>[4]</sup> i znacznie mniej kosztowne obliczeniowo niż metoda 6-go rzędu BBCK<sup>[6]</sup>. Metoda  $\Xi^{[3]}$  zachowuje stabilny 3-ci błąd zbieżności niezależnie od wielkości oscylacji  $\omega$ . Metoda ta jest bardziej kosztowna obliczeniowo niż metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$ .



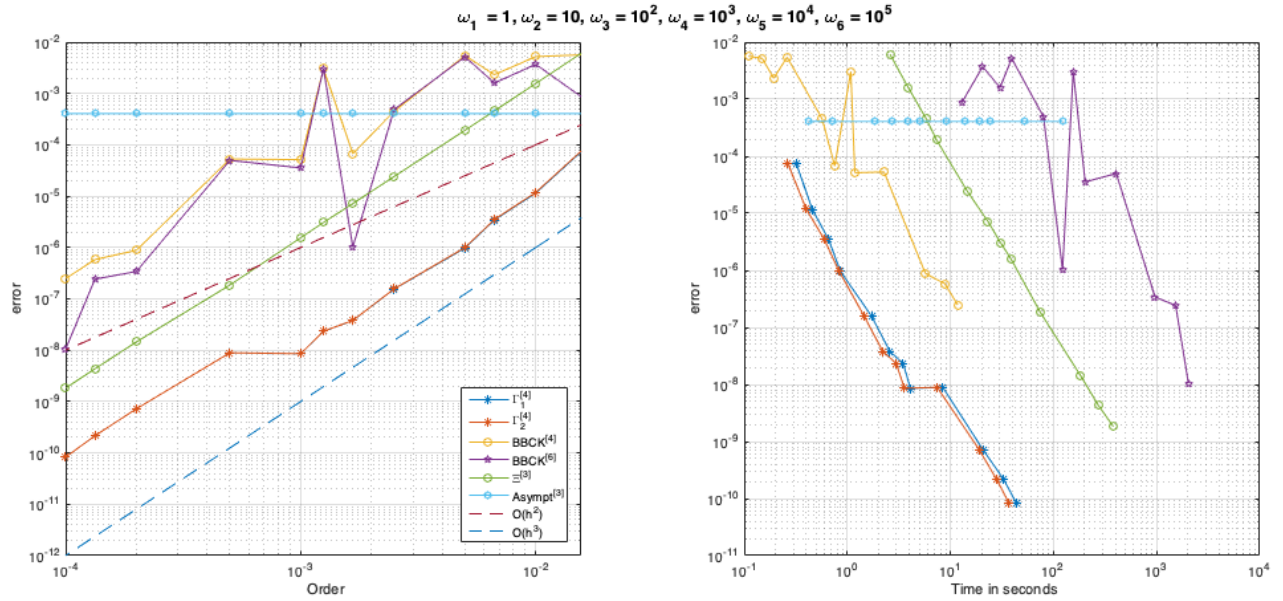
RYSUNEK 4.9: Porównanie rzędów (pierwszy wiersz) i kosztu w sekundach (drugi wiersz) dla równania (4.24).

*Przykład 4.*

Na koniec rozważmy równanie, w którym funkcja  $f(x, t)$  obejmuje szeroką gamę częstotliwości

$$\begin{aligned} \partial_t^2 u &= \partial_x^2 u - \sum_{k=0}^5 \left(1 + \cos(10^k t)\right) x^2 u, \quad x \in [-\pi, \pi], \quad t \in [0, 1], \\ u(x, 0) &= e^{-\frac{x^2}{2}}, \quad \partial_t u(x, 0) = 0, \\ u(-\pi, t) &= u(\pi, t). \end{aligned} \tag{4.25}$$

Zgodnie z oszacowaniami błędów (Rysunki 4.1–4.3) i Twierdzeniem 2 zaproponowane przez nas metody  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$  osiągają błąd wielkości co najwyżej  $\mathcal{O}(h^2)$  globalnie. Na rysunku 4.10 możemy zaobserwować, że nowe metody zachowują się stabilnie dla wszystkich przedziałów czasowych, osiągając dokładność  $\mathcal{O}(h^3)$ . Zgodnie z oczekiwaniami, metoda asymptotyczna osiąga stosunkowo słabą zbieżność i dużą stałą błędu, co jest to spowodowane niskimi oscylacjami  $\omega_1 = 1$  i  $\omega_2 = 10$ . Z drugiej strony duże oscylacje, takie jak  $\omega_6 = 10^5$ , sabotują metody BBCK<sup>[4]</sup> i BBCK<sup>[6]</sup>, gdyż kwadratury na których oparte są te metody zawodzą w aproksymacji silnie oscylacyjnych całek. Metoda  $\Xi^{[3]}$  osiąga stabilny 3-ci rząd zbieżności, ale jej dokładność jest mniejsza niż metod  $\Gamma_1^{[4]}$  i  $\Gamma_2^{[4]}$ . Ponadto, metoda  $\Xi^{[3]}$  jest dużo bardziej kosztowna obliczeniowo.



RYSUNEK 4.10: W tym przykładzie analizujemy bardzo ważne równanie (4.25), w którym funkcja  $f(x,t)$  zależy od wielu różnych oscylacji, od bardzo niskich jak  $\omega_1 = 1$  do wysokich, jak  $\omega_6 = 10^5$ .

## 4.5 Dodatek

### 4.5.1 Obliczenia dotyczące elementów rozwinięcia Magnusa (4.4)–(4.7)

Przypomnijmy, że  $f(t) = \alpha(t) + \sum_{|n| \leq N} a_n(t) e^{i\omega_n t}$  oraz

$$A(t_1) = \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix}, \quad [A(t_2), A(t_1)] = \begin{bmatrix} f(t_1) - f(t_2) & 0 \\ 0 & f(t_2) - f(t_1) \end{bmatrix}.$$

W pierwszej kolejności obliczamy dwukrotnie zagnieżdżone komutatory.

$$\begin{aligned} [A(t_1), [A(t_2), A(t_3)]] &= \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix} \begin{bmatrix} f(t_3) - f(t_2) & 0 \\ 0 & f(t_2) - f(t_3) \end{bmatrix} \\ &\quad - \begin{bmatrix} f(t_3) - f(t_2) & 0 \\ 0 & f(t_2) - f(t_3) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & f(t_2) - f(t_3) \\ [\Delta + f(t_1)][f(t_3) - f(t_2)] & 0 \end{bmatrix} \\ &\quad - \begin{bmatrix} 0 & f(t_3) - f(t_2) \\ [f(t_2) - f(t_3)][\Delta + f(t_1)] & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & H_1 \\ H_2 & 0 \end{bmatrix}, \end{aligned}$$

gdzie

$$\begin{aligned} H_1 &= 2[f(t_2) - f(t_3)] \\ H_2 &= [\Delta + f(t_1)][f(t_3) - f(t_2)] - [f(t_2) - f(t_3)][\Delta + f(t_1)] \\ &= \Delta f(t_3) + f(t_3)\Delta - \Delta f(t_2) - f(t_2)\Delta + 2f(t_1)f(t_3) - 2f(t_1)f(t_2). \end{aligned}$$

Analogicznie

$$[A(t_3), [A(t_2), A(t_1)]] = \begin{bmatrix} 0 & H_3 \\ H_4 & 0 \end{bmatrix},$$

gdzie

$$\begin{aligned} H_3 &= 2[f(t_2) - f(t_1)], \\ H_4 &= \Delta f(t_1) + f(t_1)\Delta - \Delta f(t_2) - f(t_2)\Delta + 2f(t_3)f(t_1) - 2f(t_3)f(t_2). \end{aligned}$$

Dla trzykrotnie zagnieżdżonych komutatorów mamy

$$\begin{aligned} [A(t_4), [[A(t_1), A(t_2)], A(t_3)]] &= \begin{bmatrix} 0 & 1 \\ \Delta + f(t_4) & 0 \end{bmatrix} \begin{bmatrix} 0 & H_5 \\ H_6 & 0 \end{bmatrix} \\ &\quad - \begin{bmatrix} 0 & H_5 \\ H_6 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \Delta + f(t_4) & 0 \end{bmatrix} \\ &= \begin{bmatrix} H_6 & 0 \\ 0 & [\Delta + f(t_4)]H_5 \end{bmatrix} - \begin{bmatrix} H_5[\Delta + f(t_4)] & 0 \\ 0 & H_6 \end{bmatrix} \\ &= \begin{bmatrix} H_6 - H_5[\Delta + f(t_4)] & 0 \\ 0 & [\Delta + f(t_4)]H_5 - H_6 \end{bmatrix}, \end{aligned}$$

gdzie

$$\begin{aligned} H_5 &= 2[f(t_2) - f(t_1)], \\ H_6 &= \Delta[f(t_1) - f(t_2)] + [f(t_1) - f(t_2)]\Delta + 2f(t_3)(f(t_1) - f(t_2)). \end{aligned}$$

Analogicznie,

$$[A(t_1), [[A(t_2), A(t_3)], A(t_4)]] = \begin{bmatrix} H_8 - H_7(\Delta + f(t_1)) & 0 \\ 0 & [\Delta + f(t_1)]H_7 - H_8 \end{bmatrix},$$

z

$$\begin{aligned} H_7 &= 2[f(t_3) - f(t_2)] \\ H_8 &= \Delta[f(t_2) - f(t_3)] + [f(t_2) - f(t_3)]\Delta + 2f(t_4)[f(t_2) - f(t_3)]. \end{aligned}$$

Podobnie,

$$\begin{aligned} [A(t_1), [A(t_2), [A(t_3), A(t_4)]]] &= \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix} \begin{bmatrix} 0 & H_9 \\ H_{10} & 0 \end{bmatrix} \\ &\quad - \begin{bmatrix} 0 & H_9 \\ H_{10} & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ \Delta + f(t_1) & 0 \end{bmatrix} \\ &= \begin{bmatrix} H_{10} & 0 \\ 0 & [\Delta + f(t_1)]H_9 \end{bmatrix} - \begin{bmatrix} H_9[\Delta + f(t_1)] & 0 \\ 0 & H_{10} \end{bmatrix} \\ &= \begin{bmatrix} H_{10} - H_9[\Delta + f(t_1)] & 0 \\ 0 & [\Delta + f(t_1)]H_9 - H_{10} \end{bmatrix}, \end{aligned}$$

gdzie

$$\begin{aligned} H_9 &= 2[f(t_3) - f(t_4)], \\ H_{10} &= \Delta[f(t_4) - f(t_3)] + [f(t_4) - f(t_3)]\Delta + 2f(t_2)[f(t_4) - f(t_3)] \end{aligned}$$

oraz

$$[A(t_2), [A(t_3), [A(t_4), A(t_1)]]] = \begin{bmatrix} H_{12} - H_{11}[\Delta + f(t_2)] & 0 \\ 0 & [\Delta + f(t_2)]H_{11} - H_{12} \end{bmatrix},$$

gdzie

$$\begin{aligned} H_{11} &= 2[f(t_4) - f(t_1)] \\ H_{12} &= \Delta[f(t_1) - f(t_4)] + [f(t_1) - f(t_4)]\Delta + 2f(t_3)[f(t_1) - f(t_4)]. \end{aligned}$$



Aby oszacować składnik  $\Theta_4$  musimy zsumować macierze pochodzące z poszczególnych czterokrotnie zagnieżdżonych komutatorów. W wyniku otrzymujemy macierz  $\begin{bmatrix} \mathcal{H}_1 & 0 \\ 0 & \mathcal{H}_2 \end{bmatrix}$ , gdzie

$$\begin{aligned} \mathcal{H}_1 &= -H_6 + H_5[\Delta + f(t_4)] + H_8 - H_7[\Delta + f(t_1)] + H_{10} - H_9[\Delta + f(t_4)] + H_{12} \\ &\quad - H_{11}[\Delta + f(t_2)] \\ &= \Delta[f(t_2) - f(t_1)] + [f(t_2) - f(t_1)]\Delta + 2f(t_3)[f(t_2) - f(t_1)] - 2[f(t_1) - f(t_2)][\Delta + f(t_4)] \\ &\quad + \Delta[f(t_2) - f(t_3)] + [f(t_2) - f(t_3)]\Delta + 2f(t_4)[f(t_2) - f(t_3)] - 2[f(t_3) - f(t_2)][\Delta + f(t_1)] \\ &\quad + \Delta[f(t_4) - f(t_3)] + [f(t_4) - f(t_3)]\Delta + 2f(t_2)[f(t_4) - f(t_3)] - 2[f(t_3) - f(t_4)][\Delta + f(t_1)] \\ &\quad + \Delta([f(t_1) - f(t_4)] + [f(t_1) - f(t_4)]\Delta + 2f(t_3)[f(t_1) - f(t_4)] - 2[f(t_4) - f(t_1)][\Delta + f(t_2)]) \\ &= \Delta[2f(t_2) - 2f(t_3)] + 3[2f(t_2) - 2f(t_3)]\Delta + 4f(t_4)([f(t_2) - f(t_3)] + 4f(t_1)[f(t_2) - f(t_3)]) \end{aligned}$$

oraz

$$\mathcal{H}_2 = -3\Delta[2f(t_2) - 2f(t_3)] - [2f(t_2) - 2f(t_3)]\Delta - 4f(t_4)[f(t_2) - f(t_3)] - 4f(t_1)[f(t_2) - f(t_3)],$$

gdzie

$$\begin{aligned} f(t_k)f(t_l) &= \left[ \alpha(t_k) + \sum_{|n| \leq N} a_n(t_k)e^{i\omega_n t_k} \right] \left[ \alpha(t_l) + \sum_{|n| \leq N} a_n(t_l)e^{i\omega_n t_l} \right] \\ &= \alpha(t_k)\alpha(t_l) + \alpha(t_k) \sum_{|n| \leq N} a_n(t_l)e^{i\omega_n t_l} + \alpha(t_l) \sum_{|n| \leq N} a_n(t_k)e^{i\omega_n t_k} \\ &\quad + \sum_{|n| \leq N} a_n(t_k)e^{i\omega_n t_k} \sum_{|m| \leq N} a_m(t_l)e^{i\omega_m t_l}. \end{aligned}$$

#### 4.5.2 Obliczenia dotyczące dominujących czynników błędu wynikających z zastosowania splittingu Stranga w podrozdziale 4.2.2

Oznaczając przez

$$\begin{aligned} X &= -\int_0^h \int_0^{t_1} [A(t+t_2), A(t+t_1)] dt_1 dt_2 = \begin{bmatrix} -\mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix}, \\ Y &= \int_0^h \begin{bmatrix} 0 & 1 \\ \Delta + f(t+t_1) & 0 \end{bmatrix} dt_1 = \begin{bmatrix} 0 & h \\ h\Delta + F & 0 \end{bmatrix}, \end{aligned}$$

otrzymujemy następujące komutatory

$$\begin{aligned} [Y, X] &= \begin{bmatrix} 0 & h \\ h\Delta + F & 0 \end{bmatrix} \begin{bmatrix} -\mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix} - \begin{bmatrix} -\mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix} \begin{bmatrix} 0 & h \\ h\Delta + F & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & h\mathcal{F} \\ -(h\Delta + F)\mathcal{F} & 0 \end{bmatrix} - \begin{bmatrix} 0 & -h\mathcal{F} \\ \mathcal{F}(h\Delta + F) & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 2h\mathcal{F} \\ -h\Delta\mathcal{F} - h\mathcal{F}\Delta - 2\mathcal{F}F & 0 \end{bmatrix}, \\ [Y, [Y, X]] &= \begin{bmatrix} 0 & h \\ h\Delta + F & 0 \end{bmatrix} \begin{bmatrix} 0 & 2h\mathcal{F} \\ -h\Delta\mathcal{F} - h\mathcal{F}\Delta - 2\mathcal{F}F & 0 \end{bmatrix} \\ &\quad - \begin{bmatrix} 0 & 2h\mathcal{F} \\ -h\Delta\mathcal{F} - h\mathcal{F}\Delta - 2\mathcal{F}F & 0 \end{bmatrix} \begin{bmatrix} 0 & h \\ h\Delta + F & 0 \end{bmatrix} \\ &= \begin{bmatrix} -h^2\Delta\mathcal{F} - 3h^2\mathcal{F}\Delta - 4h\mathcal{F}F & 0 \\ 0 & 3h^2\Delta\mathcal{F} + h^2\mathcal{F}\Delta + 4h\mathcal{F}F \end{bmatrix} \\ [X, Y] = -[Y, X] &= \begin{bmatrix} 0 & -2h\mathcal{F} \\ h\Delta\mathcal{F} + h\mathcal{F}\Delta + 2\mathcal{F}F & 0 \end{bmatrix} \end{aligned}$$

$$\begin{aligned}
[X, [X, Y]] &= \begin{bmatrix} -\mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix} \begin{bmatrix} 0 & -2h\mathcal{F} \\ h\Delta\mathcal{F} + h\mathcal{F}\Delta + 2\mathcal{F}\mathcal{F} & 0 \end{bmatrix} \\
&\quad - \begin{bmatrix} 0 & -2h\mathcal{F} \\ h\Delta\mathcal{F} + h\mathcal{F}\Delta + 2\mathcal{F}\mathcal{F} & 0 \end{bmatrix} \begin{bmatrix} -\mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix} \\
&= \begin{bmatrix} 0 & 4h\mathcal{F}^2 \\ 2h\mathcal{F}\Delta\mathcal{F} + h\mathcal{F}^2\Delta + h\Delta\mathcal{F}^2 + 4\mathcal{F}^2\mathcal{F} & 0 \end{bmatrix}.
\end{aligned}$$

### 4.5.3 Wyprowadzenie splittingu 4-go rzędu typu Chin-Chen

W podrozdziale 4.1.4 zastosowaliśmy splitting 4-go rzędu (4.13), zaprezentowany w pracy [Chin and Chen, 2002], gdzie został on wyprowadzony dla równania Schrödingera.

W ogólnej postaci dekompozycje takiego typu, czyli

$$e^{X+Y} = e^{c_2 X} e^{c_1 Y} e^{c_0 X + a[[X, Y], X] + b[[X, Y], Y]} e^{c_1 Y} e^{c_2 X} + \mathcal{O}(h^5)$$

można wyprowadzić przez dwukrotne zastosowanie formuły Baker–Campbell–Hausdorff postaci

$$\begin{aligned}
\exp[Y] \exp[X] \exp[Y] &= \exp[\text{sBCH}(Y, X)], \text{ gdzie} \\
\text{sBCH}(Y, X) &= 2Y + X - \frac{1}{6}[[X, Y], Y] - \frac{1}{6}[[X, Y], X] + \mathcal{O}(h^5)
\end{aligned}$$

oraz wyznaczenie stałych  $c_0, c_1$  oraz  $c_2$  tak, aby  $a = 0$  lub  $b = 0$ .

Poniżej przedstawiamy szczegółowe obliczenia prowadzące do używanej przez nas postaci splittingu typu Chin-Chen:

1.

$$\begin{aligned}
\underbrace{e^{c_1 Y} e^{c_0 X} e^{c_1 Y}}_{W_1} &= \exp[\text{sBCH}(c_1 Y, c_0 X)] \\
\text{sBCH}(c_1 Y, c_0 X) &= 2c_1 Y + c_0 X - \frac{1}{6}c_0 c_1^2 [[X, Y], Y] - \frac{1}{6}c_0^2 c_1 [[X, Y], X]
\end{aligned}$$

2.

$$\begin{aligned}
e^{c_2 X} \underbrace{e^{c_1 Y} e^{c_0 X} e^{c_1 Y}}_{W_1} e^{c_2 X} &= \exp[\text{sBCH}(c_2 X, W_1)] \\
\text{sBCH}(c_2 X, W_1) &= 2c_2 X + W_1 - \frac{1}{6}c_2^2 [[W_1, X], X] - \frac{1}{6}c_2 [[W_1, X], W_1] \\
&= 2c_2 X + 2c_1 Y + c_0 X - \frac{1}{6}c_0 c_1^2 [[X, Y], Y] - \frac{1}{6}c_0^2 c_1 [[X, Y], X] \\
&\quad - \frac{1}{3}c_1 c_2^2 [[Y, X], X] - \frac{2}{3}c_1^2 c_2 [[Y, X], Y] - \frac{1}{3}c_0 c_1 c_2 [[Y, X], X] \\
&= (c_0 + 2c_2)X + 2c_1 Y + \left(\frac{2}{3}c_1^2 c_2 - \frac{1}{6}c_0 c_1^2\right) [[X, Y], Y] \\
&\quad + \left(\frac{1}{3}c_1 c_2^2 - \frac{1}{6}c_0^2 c_1 + \frac{1}{3}c_0 c_1 c_2\right) [[X, Y], X]
\end{aligned}$$

gdzie

$$\begin{aligned}
-\frac{1}{6}c_2^2 [[W_1, X], X] &= -\frac{1}{6}c_2^2 [[2c_1 Y, X], X] - \frac{1}{6}c_2^2 [[c_0 X, X], X] = -\frac{1}{3}c_1 c_2^2 [[Y, X], X] \\
-\frac{1}{6}c_2 [[W_1, X], W_1] &= -\frac{1}{6}c_2 [[2c_1 Y + c_0 X, X], 2c_1 Y + c_0 X] = -\frac{2}{3}c_1^2 c_2 [[Y, X], Y] - \frac{1}{3}c_0 c_1 c_2 [[Y, X], X],
\end{aligned}$$

$$\text{zatem } a = \frac{1}{3}c_1c_2^2 - \frac{1}{6}c_0^2c_1 + \frac{1}{3}c_0c_1c_2 \text{ oraz } b = \frac{2}{3}c_1^2c_2 - \frac{1}{6}c_0c_1^2.$$

W zależności od tego, czy chcemy aby  $a = 0$  czy aby  $b = 0$ , otrzymujemy dwa układy równań dla współczynników  $c_0, c_1, c_2$ :

$b = 0$	$a = 0$
$\begin{cases} c_0 + 2c_2 = 1 \\ 2c_1 = 1 \\ a = \frac{1}{3}c_1c_2^2 - \frac{1}{6}c_0^2c_1 + \frac{1}{3}c_0c_1c_2 \\ b = \frac{2}{3}c_1^2c_2 - \frac{1}{6}c_0c_1^2 \\ b = 0 \end{cases}$	$\begin{cases} c_0 + 2c_2 = 1 \\ 2c_1 = 1 \\ a = \frac{1}{3}c_1c_2^2 - \frac{1}{6}c_0^2c_1 + \frac{1}{3}c_0c_1c_2 \\ b = \frac{2}{3}c_1^2c_2 - \frac{1}{6}c_0c_1^2 \\ a = 0 \end{cases}$
$c_0 = \frac{2}{3}, c_1 = \frac{1}{2}, c_2 = \frac{1}{6} \text{ and } a = -\frac{1}{72}$	$c_0 = \frac{1}{\sqrt{3}}, c_1 = \frac{1}{2}, c_2 = \frac{1}{2} - \frac{1}{2\sqrt{3}} \text{ and } b = \frac{1}{24}(2 - \sqrt{3})$

Zauważmy, że współczynniki w lewej kolumnie odpowiadają współczynnikom ze splittingu (4.13).

Wykorzystanie splittingu, w którym to współczynnik  $a = 0$  okazuje się być bardziej kosztowny obliczeniowo. Niezależnie od wyboru składników dekompozycji  $X$  i  $Y$  (gdzie  $X + Y = \begin{bmatrix} 0 & h \\ D + F & 0 \end{bmatrix}$ ) wewnętrzny składnik splittingu  $\exp\left(\frac{\sqrt{3}h}{3}X - \frac{(2-\sqrt{3})h^3}{24}[[X, Y], Y]\right)$  jest zawsze bardziej kosztowny obliczeniowo niż składnik  $\exp\left(\frac{2}{3}hX + \frac{1}{72}h^3[X, [Y, X]]\right)$ .



## Rozdział 5

# Podsumowanie i dalsze perspektywy

Podsumowując, w rozprawie doktorskiej skupiliśmy się na przedstawieniu trzech podejść analizy numerycznej do zagadnienia przybliżenia rozwiązań równań, w których występują oscylacje. Reprezentatywnym przykładem takiego równania jest równanie Kleina-Gordona

$$\partial_t^2 \psi(\mathbf{x}, t) = \Delta \psi(\mathbf{x}, t) + f(\mathbf{x}, t) \psi(\mathbf{x}, t). \quad (5.1)$$

W naszej analizie założyliśmy, że funkcja  $f(\mathbf{x}, t)$  może być przedstawiona lub przybliżona w postaci szeregu

$$f(\mathbf{x}, t) = \alpha(\mathbf{x}, t) + \sum_{|n| \leq N} a_n(\mathbf{x}, t) e^{i\omega_n t}, \quad (5.2)$$

gdzie  $n \in \mathbb{N}$ ,  $\omega_n \in \mathbb{R}$ ,  $\omega_{\min} = \min_{n \leq N} |\omega_n| \geq 1$  oraz  $\omega_{\max} = \max_{n \leq N} |\omega_n| < \infty$ . Funkcje  $\alpha(\mathbf{x}, t)$  oraz  $a_n(\mathbf{x}, t)$ ,  $n \leq N$ , nie oscylują w zależności od  $\omega_n$  (tj. ich pochodne czasowe są ograniczone niezależnie od  $\omega_n$ ).

W Rozdziale 2 przedstawiliśmy metodę numeryczno-asymptotyczną, która w przypadku wyłącznie wysokich oscylacji, tj. gdy  $\omega_{\min} \gg 1$ , wykazuje się nieporównywalnie lepszą dokładnością niż wszystkie inne znane metody numeryczne, jednakże zawodzi w obecności czynników nieoscylujących.

Drugą metodą, zaprezentowaną w Rozdziale 3, jest metoda rzędu trzeciego globalnie, niezależnie od wielkości oscylacji  $\omega_n$ . Metoda ta oparta jest na przedstawieniu równania (5.1) w postaci niejednorodnego układu równań różniczkowych zwyczajnych i zastosowaniu do niego reguły Duhamela. Ponadto, użycie metody Filona do aproksymacji wysoko-oscyłujących całek pozwala na uniezależnienie dokładności metody od wielkości kroku czasowego  $h$ . Przedstawiliśmy pełen dowód zbieżności metody oraz jej rzędu i pokazaliśmy, że stała błędów nie rośnie wraz ze wzrostem oscylacji.

Ostatnim, zaprezentowanym w Rozdziale 4 podejściu numerycznym, jest przedstawienie równania (5.1) w postaci układu równań różniczkowych zwyczajnych i zastosowanie do niego rozwinięcia Magnusa a następnie dwóch rodzajów dekompozycji pola wektorowego: Stranga i typu Chin-Chen. Pozwoliło to na konstrukcję dwóch metod numerycznych, których lokalną dokładność opisują wyrażenia

$$\mathcal{O}(h^5 + \min\{h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \neq 0$$

oraz

$$\mathcal{O}(\min\{h^3, \frac{h^2}{\omega_{\min}}, h^5 \omega_{\max}^2\}) \quad \text{dla } \alpha(\mathbf{x}, t) \equiv 0.$$

Zauważmy, że metody te wymagają bardziej rygorystycznych założeń dotyczących gładkości funkcji  $\psi(\mathbf{x}, t)$  oraz  $f(\mathbf{x}, t)$  w czasie,  $H^{s+6}$ , za to wykazują się one dużo lepszą dokładnością i niższym kosztem obliczeniowym niż wszystkie ze znanych dotychczas metod, niezależnie od wielkości oscylacji  $\omega_n$ . Co więcej, są one konkurencyjne w przypadku, gdy w funkcji  $f(\mathbf{x}, t)$  mamy szereg różnych oscylacji, od małych do bardzo dużych.

**Perspektywy dalszej pracy.** Doświadczenie zdobyte w matematyce obliczeniowej dla liniowego równania Kleina-Gordona planuję wykorzystać i poszerzyć pracując nad nieliniowymi równaniami falowymi. W pierwszej kolejności przyjrę się równaniom

- równanie Kleina–Gordona z potęgową nieliniowością

$$\begin{aligned}\frac{\partial^2}{\partial t^2} \psi(x, t) &= \Delta \psi(x, t) + a\psi(x, t) + b\psi(x, t)^n, \quad t \in [t_0, T], \quad x \in \mathbb{T}^d, \quad a, b \in \mathbb{R} \\ \psi(x, t_0) &= \psi_0(x), \quad \partial_t \psi(x, t_0) = \varphi_0(x);\end{aligned}$$

- równanie sinh-Gordon

$$\begin{aligned}\frac{\partial^2}{\partial t^2} \psi(x, t) &= \Delta \psi(x, t) + a \sinh(\lambda \psi(x, t)), \quad t \in [t_0, T], \quad x \in \mathbb{T}^d, \quad a, \lambda \in \mathbb{R} \\ \psi(x, t_0) &= \psi_0(x), \quad \partial_t \psi(x, t_0) = \varphi_0(x);\end{aligned}$$

- równanie sin-Gordon

$$\begin{aligned}\frac{\partial^2}{\partial t^2} \psi(x, t) &= \Delta \psi(x, t) + b \sin(\lambda \psi(x, t)), \quad t \in [t_0, T], \quad x \in \mathbb{T}^d, \quad b, \lambda \in \mathbb{R} \\ \psi(x, t_0) &= \psi_0(x), \quad \partial_t \psi(x, t_0) = \varphi_0(x);\end{aligned}$$

- równanie Kleina–Gordona opisujące zachowanie cząsteczki o masie  $m$  i ładunku  $q$  poruszającej się pod wpływem potencjałów elektrodynamicznych  $A(\mathbf{x}, t)$  oraz  $\phi(\mathbf{x}, t)$

$$\left[ \left( i\hbar \frac{\partial}{\partial t} - q\phi(\mathbf{x}, t) \right)^2 - c^2 \left( \frac{\hbar}{i} \nabla - qA(\mathbf{x}, t) \right)^2 - m^2 c^4 \right] \varphi(\mathbf{x}, t) = 0.$$

Zauważmy ponadto, że wszystkie zaproponowane przez nas metody dotyczą problemu występowania oscylacji, w szczególności tych bardzo wysokich, a nie samego równania Kleina-Gordona. Z tego powodu, w dalszej perspektywie zamierzam wykorzystać nasze techniki analizy numerycznej dla innych równań wysoko-oscyłujących, nie tylko tych falowych czy liniowych.

# Bibliografia

- [Abdulle et al., 2012] Abdulle, A., E. W., Engquist, B., and Vanden-Eijnden, E. (2012). The heterogeneous multiscale method. *Acta Numer.*, 21:1–87.
- [Bader et al., 2019] Bader, P., Blanes, S., Casas, F., and Kopylov, N. (2019). Novel symplectic integrators for the Klein-Gordon equation with space- and time-dependent mass. *J. Comput. Appl. Math.*, 350:130–138.
- [Bao and Dong, 2012] Bao, W. and Dong, X. (2012). Analysis and comparison of numerical methods for the Klein-Gordon equation in the nonrelativistic limit regime. *Numer. Math.*, 120(2):189–229.
- [Blanes et al., 2009] Blanes, S., Casas, F., Oteo, J. A., and Ros, J. (2009). The Magnus expansion and some of its applications. *Phys. Rep.*, 470(5-6):151–238.
- [Blanes et al., 2002] Blanes, S., Casas, F., and Ros, J. (2002). High order optimized geometric integrators for linear differential equations. *BIT Numerical Mathematics*, 42:262–284.
- [Chen and Liu, 2008] Chen, J.-B. and Liu, H. (2008). Multisymplectic pseudospectral discretizations for  $(3 + 1)$ -dimensional Klein-Gordon equation. *Commun. Theor. Phys. (Beijing)*, 50(5):1052–1054.
- [Chin and Chen, 2002] Chin, S. A. and Chen, C. (2002). Gradient symplectic algorithms for solving the schrödinger equation with time-dependent potentials. *The Journal of Chemical Physics*, 117(4):1409–1415.
- [Condon et al., 2012] Condon, M., Deaño, A., Iserles, A., and Kropielnicka, K. (2012). Efficient computation of delay differential equations with highly oscillatory terms. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(6):1407–1420.
- [Condon et al., 2009] Condon, M., Deano, A., Iserles, A., Maczyński, K., and Xu, T. (2009). On numerical methods for highly oscillatory problems in circuit simulation. *COMPEL-The international journal for computation and mathematics in electrical and electronic engineering*, 28(6):1607–1618.
- [Condon et al., 2019] Condon, M., Iserles, A., Kropielnicka, K., and Singh, P. (2019). Solving the wave equation with multifrequency oscillations. *Journal of Computational Dynamics*, 6(2):239–249.
- [Condon et al., 2021] Condon, M., Kropielnicka, K., Lademann, K., and Perczyński, R. (2021). Asymptotic numerical solver for the linear Klein-Gordon equation with space- and time-dependent mass. *Applied Mathematics Letters*, 115:106935.
- [Deaño et al., 2018] Deaño, A., Huybrechs, D., and Iserles, A. (2018). *Computing highly oscillatory integrals*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- [Engquist et al., 2009] Engquist, B., Fokas, A., Hairer, E., and Iserles, A. (2009). *Highly Oscillatory Problems*. Number 366 in London Maths Soc. Lecture Note Series. Cambridge University Press.
- [Faou and Schratz, 2014] Faou, E. and Schratz, K. (2014). Asymptotic preserving schemes for the Klein-Gordon equation in the non-relativistic limit regime. *Numer. Math.*, 126(3):441–469.
- [Gao and Iserles, 2017] Gao, J. and Iserles, A. (2017). Error analysis of the extended filon-type method for highly oscillatory integrals. *Research in the Mathematical Sciences*, 4.
- [Greiner et al., 1996] Greiner, W., Reinhardt, J., et al. (1996). *Field quantization*. Springer Science & Business Media.
- [Hochbruck and Lubich, 1998] Hochbruck, M. and Lubich, C. (1998). A gautschi-type method for oscillatory second-order differential equations. *Numerische Mathematik*, 83.

- [Iserles et al., 2019] Iserles, A., Kropielnicka, K., and Singh, P. (2019). Compact schemes for laser–matter interaction in schrödinger equation based on effective splittings of magnus expansion. *Computer Physics Communications*, 234:195–201.
- [Iserles et al., 2000] Iserles, A., Munthe-Kaas, H. Z., Nørsett, S. P., and Zanna, A. (2000). Lie-group methods. In *Acta numerica, 2000*, volume 9 of *Acta Numer.*, pages 215–365. Cambridge Univ. Press, Cambridge.
- [Iserles and Nørsett, 2005] Iserles, A. and Nørsett, S. P. (2005). Efficient quadrature of highly oscillatory integrals using derivatives. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461:1383 – 1399.
- [Iserles et al., 2001] Iserles, A., Nørsett, S. P., and Rasmussen, A. F. (2001). Time symmetry and high-order Magnus methods. *Appl. Numer. Math.*, 39(3-4):379–401. Special issue: Themes in geometric integration.
- [Jahnke and Lubich, 2000] Jahnke, T. and Lubich, C. (2000). Error bounds for exponential operator splittings. *BIT Numerical Mathematics*, 40:735–744.
- [Kopriva, 2009] Kopriva, D. A. (2009). *Implementing spectral methods for partial differential equations*. Scientific Computation. Springer, Berlin. Algorithms for scientists and engineers.
- [Kropielnicka and Lademann, 2022] Kropielnicka, K. and Lademann, K. (2022). Third order, uniform in low to high oscillatory coefficients, exponential integrators for Klein-Gordon equations. *arXiv preprint arXiv:2212.13762, wstępnie przyjęta w Mathematical Modelling and Numerical Analysis*.
- [Kropielnicka et al., 2021] Kropielnicka, K., Lademann, K., and Schratz, K. (2021). Effective highly accurate time integrators for linear Klein-Gordon equations across the scales. *arXiv preprint arXiv:2112.08908, w recenzji*.
- [Kumar et al., 2014] Kumar, D., Singh, J., Kumar, S., et al. (2014). Numerical computation of klein–gordon equations arising in quantum field theory by using homotopy analysis transform method. *Alexandria Engineering Journal*, 53(2):469–474.
- [Magnus, 1954] Magnus, W. (1954). On the exponential solution of differential equations for a linear operator. *Communications on pure and applied mathematics*, 7(4):649–673.
- [McLachlan and Quispel, 2002] McLachlan, R. I. and Quispel, G. R. W. (2002). Splitting methods. *Acta Numerica*, 11:341–434.
- [Mostafazadeh, 2003] Mostafazadeh, A. (2003). Hilbert space structures on the solution space of klein-gordon-type evolution equations. *Classical Quantum Gravity*, 20(1):155–171.
- [Mostafazadeh, 2004] Mostafazadeh, A. (2004). Quantum mechanics of Klein-Gordon-type fields and quantum cosmology. *Annals of Physics*, 309(1):1–48.
- [Ruth, 1983] Ruth, R. D. (1983). A canonical integration technique. *IEEE Trans. Nucl. Sci.*, 30(CERN-LEP-TH-83-14):2669–2671.
- [Schwartz, 2014] Schwartz, M. D. (2014). *Quantum field theory and the standard model*. Cambridge university press.
- [Shakeri and Dehghan, 2008] Shakeri, F. and Dehghan, M. (2008). Numerical solution of the Klein-Gordon equation via He’s variational iteration method. *Nonlinear Dynam.*, 51(1-2):89–97.
- [Shankar, 2012] Shankar, R. (2012). *Principles of quantum mechanics*. Springer Science & Business Media.
- [Strang, 1968] Strang, G. (1968). On the construction and comparison of difference schemes. *SIAM journal on numerical analysis*, 5(3):506–517.
- [Trefethen, 2000] Trefethen, L. N. (2000). *Spectral methods in MATLAB*. SIAM.
- [Yusufoğlu, 2008] Yusufoğlu, E. (2008). The variational iteration method for studying the Klein-Gordon equation. *Appl. Math. Lett.*, 21(7):669–674.
- [Zanna, 2001] Zanna, A. (2001). The Fer expansion and time-symmetry: A Strang-type approach. *Applied Numerical Mathematics*, 39(3-4):435–459.



- [Znojil, 2016] Znojil, M. (2016). Quantization of big bang in crypto-hermitian heisenberg picture. In *Non-Hermitian Hamiltonians in Quantum Physics: Selected Contributions from the 15th International Conference on Non-Hermitian Hamiltonians in Quantum Physics, Palermo, Italy, 18-23 May 2015*, pages 383–399. Springer.
- [Znojil, 2017a] Znojil, M. (2017a). Klein-Gordon equation with the time- and space-dependent mass:Unitary evolution picture. Technical report, Nuclear Physics Institute, Czech Academy of Sciences.
- [Znojil, 2017b] Znojil, M. (2017b). Non-Hermitian interaction representation and its use in relativistic quantum mechanics. *Ann. Physics*, 385:162–179.