ГОДИШНИК

НА

СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА

Том 93
1999

ANNUAIRE

DE

L'UNIVERSITE DE SOFIA "ST. KLIMENT OHRIDSKI"

FACULTE DE MATHEMATIQUES ET INFORMATIQUE

Tome 93
1999

СОФИЯ • 2000 • SOFIA
УНИВЕРСИТЕТСКО ИЗДАТЕЛСТВО „СВ. КЛИМЕНТ ОХРИДСКИ“
PRESSES UNIVERSITAIRES "ST. KLIMENT OHRIDSKI"

Annuaire de l' Université de Sofia "St. Kliment Ohridski"
Faculté de Mathématiques et Informatique

Годишник на Софийския университет „Св. Климент Охридски“
Факултет по математика и информатика

Address for correspondence:

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 Blvd J. Bourchier, P.O. Box 48
BG-1164 Sofia, Bulgaria

Fax xx(359 2) 687 180
electronic mail:
annuaire@fmi.uni-sofia.bg

Aims and Scope. The *Annuaire* is the oldest Bulgarian journal, founded in 1904, devoted to pure and applied mathematics, mechanics and computer sciences. It is reviewed by *Zentralblatt für Mathematik*, *Mathematical Reviews* and the Russian *Referativnii Jurnal*. The *Annuaire* publishes significant and original research papers of authors both from Bulgaria and abroad in some selected areas that comply with the traditional scientific interests of the Faculty of Mathematics and Informatics at the "St. Kliment Ohridski" University of Sofia, i.e., algebra, geometry and topology, analysis, mathematical logic, theory of approximations, numerical methods, computer sciences, classical, fluid and solid mechanics, and their fundamental applications.

ISSN 0205–0808

CONTENTS

IVAN GANCHEV. 110 Anniversary of the Faculty of Mathematics and Informatics at the Sofia University "St. Kl. Ohridski" (in Bulgarian) 5

OGNYAN CHRISTOV. Remark on the non-integrability of the perturbed motion of the particle in a central field in constant curvature spaces 25

ANGEL DITCHEV. Effective enumerations of families of recursive functions .. 31

VESSELA BALEVA. Regular enumerations for abstract structures 39

ANTON ZINOVIEV. Termal Equations and Finite Controllability 49

DONKA PASKOULEVA. On a class of meromorphic functions with negative coefficients 55

DIMITRINKA VLADEVA. Properties of the maximum likelihood estimation of constant drift and diffusion parameters in k-dimensional diffusion process observed at discrete random sampling 61

GEORGI CHOBANOV AND IVAN DIMOVSKI. Extension of the Duhamel principle for the heat equation with Dezin's initial condition................ 73

TOSHKO BOEV. A logarithmic class of semilinear wave equations 93

M. KOLEV AND K. MARKOV. A note on the shear modulus of a binary elastic mixture 109

IVAN MIHOVSKY. An explicit polynomial solution of the representative problem of the mechanics of fibrous composites 121

VALERY LEVIN AND THOMAS MICHELITSCH. Scattering of acoustoelectric waves on a cylindrical inhomogeneity in the transversely isotropic piezoelectric medium 137

J. PONTES, C. CHRISTOV AND M. VELARDE. Numerical approach to pattern selection in a model problem for Bénard convection in finite fluid layer 157

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

СТО И ДЕСЕТ ГОДИНИ ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА ПРИ СУ „СВ. КЛИМЕНТ ОХРИДСКИ“*

ИВАН ГАНЧЕВ

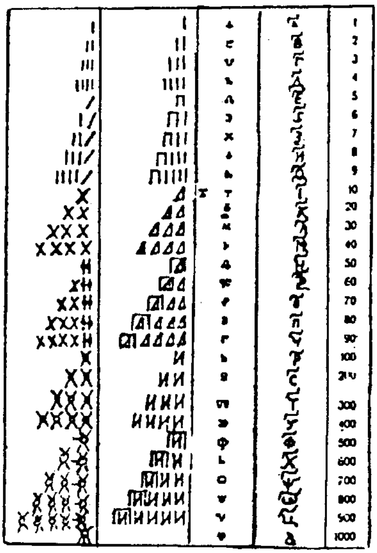
Математиката в България от абака до компютъра

Съдбата на всеки човек се определя от различни фактори. Основна роля всред тях играят обаче комплексът генетично наследени качества и средата, представляваща също комплекс от фактори, в която се развива дадена личност. С развитието на цивилизацията в рамките на втория основен фактор (средата) се появява един, който започва да играе важна роля в развитието на все повече хора. Това е училището. То като институция също има свое развитие, а с това развитие се изменя и влиянието му върху човека. Смея да твърдя, че все още малко е изследван въпросът как развитието на училището е влияло върху развитието на хората. Интуитивно обаче една значителна част от тях оценяват значимостта на това влияние и по различни поводи се обръщат към историята както на своя род, така и на своите училища. Вероятно имаме основание да кажем, че тази част от хората поне донякъде опровергават твърдението на известния немски философ Хегел „Историята учи, че никой не се учи от нея“. Нещо повече, те са достигнали до осъзнаване на ролята на знанията за миналото, за разбиране на настоящето, т.е. до мисълта, изказана от големия немски математик Г. В. Лайбниц още през 17 век, а именно: „Който иска да се ограничи с настоящето, той никога няма да го разбере.“

* Академично слово, прочетено на тържественото събрание във ФМИ, проведено на 24.11.1999 г. по случай 110 години от неговото създаване.

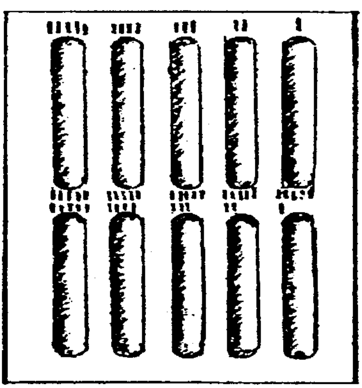
Най-сетне, с провеждането на самото настоящо честване на 110-годишнината на нашия факултет показваме, че според класацията на поета А. С. Пушкин за интелигентност и посредственост, ние попадаме в групата на интелигентните. Ще припомня, че някъде А. С. Пушкин беше писал: „Отношението към миналото отличава интелигентността от посредствеността.“ Обръщайки се към историята на нашия факултет, не може да не проявим интерес и към неговата предистория, към факторите, създали средата, в която се е почувствала необходимост от организирането му и възможност за неговото функциониране. От кога обаче започва предисторията на ФМИ?

Може би различни хора ще приемат за начало на тази предистория различни дати. В статия, публикувана в сп. „Обучението по математика“, академик Б. Петканчин започва тази предистория със създаването на Първата българска държава. Доверявайки се на усета към историческите факти и логиката на академик Б. Петканчин считам, че в един такъв тържествен момент, свързан със 110-годишния юбилей на Факултета, също е целесъобразно да представя на Вашето внимание съвсем накратко първо някои факти, които се отнасят до предисторията на факултета.



Фиг. 1

Първите сведения, свързани с математическите знания, използвани през този период, се отнасят до цифрите, с които прабългарите идват на Балканския полуостров (фиг. 1). Някои от тях са намерени при археологически изследвания на Плиска.



Фиг. 2

Съществен недостатък на старобългарските цифрови системи, а също и на славянобългарските е това, че те не са удобни за смятане. Същият недостатък имат старогръцката и α-бетичната цифрова система. Затова основното средство за смятане с по-големи числа е бил абакът. Поради ограниченост на времето тук само ще отбележа, че преди 20 години ми се удаде да възстановя български абак, използван през втората половина на 19 век в Ракитово (фиг. 2).

Турското робство откъсва българския народ от бурното развитие на математическите знания в Европа от 15 до 19 век. Известен проблясък в учебното дело в това отношение се появява само в Чипровското училище, за което за съжаление нямаме почти никакви конкретни сведения. Затова на този период няма да се спираме.

Интересен източник, който ни дава сведения за характера и равнището на математическите знания, разпространени между обикновените хора до 19 век, са задачите — народно творчество, някои от които са поместени при гатанки в статии още преди 1878 г. За отбелязване е обаче, че такива задачи почти няма в първите ни учебници по математика. Това показва, че някои математически знания са се предавали от едно поколение на друго, както са се предавали говоримият език и обикновените знания, свързани с бита.

До началото на 19 век единствените места, където организирано се предават съвсем елементарни математически знания, са манастирите, в които се подготвят свещеници. Обемът на знанията се е определял от границите, в които са се простирали нуждите за водене на прости финансови манастирски сметки и за пресмятанията, свързани с изработването на църковния календар. Едновременно с това от края на 18 век в създадените гръцки училища все по-често започват да постъпват и българчета, където те заедно с другите учебни предмети изучават и математика.

Първата книга на български език, в която има математически знания, изложени за усвояване, е „Рибният буквар“ на П. Берон, издаден през 1824 г. В нея на 9 страници са изложени таблици за четирите аритметични действия и някои примери на алгоритми за действия с по-големи числа. В таблиците числата са записани с арабски цифри, но не се използват знаците за аритметични действия и знакът „=“. Изложена е и кирилицата като цифрова система.

Първият български учебник по математика е аритметиката на Христаки Павлович, издадена през 1833 г., когато авторът е учител в създаденото от него през 1831 г. класно училище в Свищов. Този учебник съдържа знания, които в днешното наше училище се изучават в 4 и 5 клас. Освен обикновени дроби, в този учебник са изложени просто тройно правило, сложно тройно правило, лъжливото правило и известният ни алгоритъм на Евклид за НОД. С подобно съдържание е и вторият български учебник по аритметика, написан също в Свищов от Н. Бозвели и Ем. Васкидович като резултат от многогодишната им учителска дейност в създаденото от втория от авторите училище още през 1815 г. В това училище по-късно учи Ем. Иванов — един от създателите на СУ и на Физико-математическия факултет.

Третият учебник по аритметика на български език, издаден през 1843 г., е преведен от гръцки език и съдържа някои елементи, характерни за училищния курс по алгебра. В него се разглеждат например квадрат и куб на сбор и разлика на две числа. Преводът на учебника е осъществен от Сава Илиев Панагюреца.

Четвъртият български учебник по аритметика е издаден през 1845 г. от Сичан Николов, който е ученик на Хр. Павлович. По съдържание той е на равнището на преведения от Сава Илиев гръцки учебник, но в него е изложена и идеята за шифъра като средство за кодиране.

Първият български учебник по алгебра на Хр. Ваклидов излиза през 1859 г. и е авторизиран превод на френски учебник. Преди Освобождението през 1868 г. излиза на български език и втори учебник по алгебра, който също е преводен.

Според сведенията, с които разполагаме, геометрия като отделен учебен предмет се изучава в Даскалоливницата в Елена още през 40-те години на 19 век, а в ръкописния учебник по география на Хр. Павлович от около 1846/47 г. има изложени някои основни геометрични знания, които след това се използват при преподаването на астрономически и географски знания. Първият български самостоятелен учебник по геометрия обаче е издаден през 1867 г. от В. Груев. Вторият и третият учебник по геометрия на български език са преводни съответно от френски (през 1871 г. на Оливие) и от руски език (през 1873 г. на А. Ю. Давидов). Учебникът на Оливие е и първият учебник на български език, в който са изложени знания от тригонометрията и механиката.

Тук заслужава внимание следният факт: авторите на първите два български учебника по математика са учили в гръцки училища. Ученик на Хр. Павлович е авторът на четвъртия учебник Сичан Николов, а основната фигура в Даскалоливницата в Елена е Ив. Момчилов, който е ученик на ученика на Хр. Павлович, Андрей Робовски. Преводачът на третия учебник Сава Илиев също е учил в гръцко училище. Авторът на взаимоучителни таблици по математика и създателят на Габровското училище (през 1835 г.) Неофит Рилски също е учил в гръцко училище.

Въобще едно по-цялостно изследване на историята на обучението по математика у нас показва следното: Първите ни учители по математика в българските училища подготвят ученици, част от които стават направо учители. Те от своя страна вече подготвят ученици, които са били в състояние да продължат образованието си по математика в Русия или други европейски страни и да се върнат у нас като учители по математика преди или непосредствено след Освобождението, но вече с много по-задълбочена математическа подготовка. Такива учители или автори са споменатият вече Иван Момчилов, Добри Чинтулов, Тодор Икономов, Христо Ваклидов, Иван Гюзелев, Атанас Тинтеров, Емануил Иванов, Стефан Лафчиев и др.

Друг факт, който заслужава внимание, е следният: вторият учебник на български език по алгебра и вторият и третият учебник по геометрия на български език, както и някои от ръкописните авторизирани преводи на други чуждестранни учебници по математика, използвани от 1860 г. до 1877 г. (като този на Д. Чинтулов), са вече на равнището на учебниците, използвани в училищата на водещите европейски страни.

Трети факт, който заслужава внимание, е свързан с идването в България на чуждестранни учители непосредствено след Освобождението. По-конкретно, създаденият интерес в нашия народ към образованието още преди освобождението от турско робство и порасналите нужди от образовани кадри в младата българска държава довеждат до бързото разрастване на училищната мрежа. А от там и до увеличаване необходимостта от учителски кадри по математика. Това налага да бъдат поканени учители от чужбина. С особена отзивчивост поканата се приема от група чешки учители. Между тях са Карел Милде, Карел Шкорпил, Антон Шоурек, Франтишек Сплитек, Иван Немец, Владислав Шак и др. Тяхната задълбочена математическа и методическа подготовка изиграва важна роля за издигане математическата подготовка в укрепващите в Княжество България и в Източна Румелия училища, но и те не могат да задоволят все по-нарастващите нужди за учители по математика в гимназиите. Подобна положителна роля играе и дошлият през 1889 г. от Русия учител по математика Михаил Квартирников.

При това положение след средата на 80-те години на 19 век от една страна са налице чувствителни нужди от учители по математика с висше образование, от друга — има вече достатъчно подготвени младежи, завършили гимназия, които с успех могат да следват във ВУЗ, а от трета — има и подготвени в чужбина учители по математика с висше образование, които биха могли по подобие на учебните заведения, където са учили те самите, да подготвят учители. Аналогично е положението и при някои други научни области. Затова, особено след съединението на Княжество България с Източна Румелия през 1885 г., у нас се засилва движението за създаване на висше учебно заведение с първа задача, поне в началото, да подготвя учители за средните училища. Не случайно тук казвам „засилва движението“, защото самата идея за създаване на български университет води началото си още от 1880 г. Основна заслуга за нея, както изглежда, има чешкият учен Иречек. За първи път въпросът за откриване на висше училище или университет у нас се поставя в чл. 69 на проекта за „Основен закон за училищата в българското княжество“, минал само на първо четене в Народното събрание на 25.11.1880 г. Проектът е подготвен от Иван Гюзелев и Иречек. От тогава до 1888 г. въпросът за откриването на университета упорито се поставя от най-ревностните радетели за уредено българско образование, но среща и голяма съпротива. Ще отбележа само следните факти.

На 06.10.1882 г. Иречек пише в дневника си: „Днес ние трима Генчев, Вацов и аз пофантазирахме за български университет, пресметнахме разноските за юридическия факултет (7 души професори), физико-математически (6 души) и историко-филологически (5 души) всички с тригодишен курс, приблизително на 300 000 фр. годишно; сиреч, толкова, колкото струва държавният съвет. За сграда ще служи разбира се сегашната гимназия.“

На прощалната си вечеря на 12.09.1884 г. Иречек вдига наздравица за „бъдещия български университет“, макар че е „преситен с досади и интриги“ в България и че преди това изповядва „че е дошъл в България флегматичен и спокоен“, но че бил станал „отпосле твърде неврозен, при толкова неприятности и такъв труден живот“. След много перипетии тази идея се осъществява в твърде скромна форма през 1888 г. Министерството на народното просвещение решава да открие в София двугодишен Висш педагогически курс (ВПК). Една комисия към него изработва „Временни правила за уреждане на ВПК (Висш педагогически курс) в София“. За отбелязване е, че в нея влизат лица, които по-късно играят важна роля в историята на Софийския университет, а именно философът Иван Георгов, литературоведът Иван Шишманов, математикът Емануил Иванов, филологът Любомир Милетич. ВПК се открива на 01(13).10.1888 г., като работа започва само историко-филологическият отдел. Вторият — физико-математическият отдел на ВПК, наречен на 18.12.1888 г. „Висше училище“ (ВУ), се открива през месец октомври 1889 г. Една от специалностите на втория отдел е „математика и физика“. Първите преподаватели-математици в нея са чехът Теодор Монин, назначен като редовен преподавател от 01.09.1889 г., и Емануил Иванов, назначен като извънреден преподавател от 01.10.1889 г. С това, както отбелязва в цитираната по-горе статия академик Б. Петканчин, за пръв път влиза официално в нашата страна „висшата математика“.

Със закон от 24.12.1894 г. във ВУ се създават факултети, факултетни съвети и се въвеждат длъжността декан и званията професор и доцент, а със закон от 29.01.1904 г. ВУ се преименува в университет с три факултета, един от които е физико-математическият. Във втория от тези закони изрично се споменават вече оформилите се катедри във всички факултети. По математика те са: Основи на висшата математика, Висш анализ, Геометрия, Висша алгебра и Механика.

След Т. Монин и Ем. Иванов, до 1897 г. включително, във Физико-математическия факултет последователно постъпват математиците: Антон Шоурек — от 01.10.1890 г., Владислав Шак — от 01.10.1891 г., Михаил Момчилов — от 01.10.1891 г., Атанас Тинтеров — от 1893 г., Спиридон Ганев — от 01.10.1893 г. Това са математиците в нашия факултет от така нареченото първо поколение, които полагат основите на висшето математическо образование у нас.

От 1889 г. до 1900 г. престижът на Физико-математическия факултет постепенно нараства и докато в началото на този период се изказват опасения, че няма да се намерят достатъчно кандидати за записване в него, то в края на този период той почти напълно задоволява нуждите на страната ни от учители, а по-късно и за преподаватели в самия него.

Първият випуск „физико-математици“, завършили през 1892 г., се състои от „16 питомци от различни краища на страната“. От албум, за който се споменава в кн. 9 и 10, 1942 г., на „Списание на физико-математическото дружество“, се вижда, че тези първи питомци са: Кирил

Джанов от Карлово, Тодор Обрешков от Търново, Христо Мазнев от с. Радковци — Тревненско, Тодор Добринов от Свищов, Иван Куюмджиев от Ески Джумая, Иван П. Димчев от с. Странджа — Ямболско, Тодор Тодоров от Сливен, Стефан Радославов от Търново, Иван Гарванов от Ст. Загора, Коста Атанасов от Търново, Тодор Свещаров от Чирпан, Васил Русчев от Бяла черква — Търновско, Илия Т. Илиев от Стара Загора, Методий Божинов от с. Крива бара — Ломско, Тодор Радов от с. Али паша — Чирпанско и Антон Кражков от Казанлък.

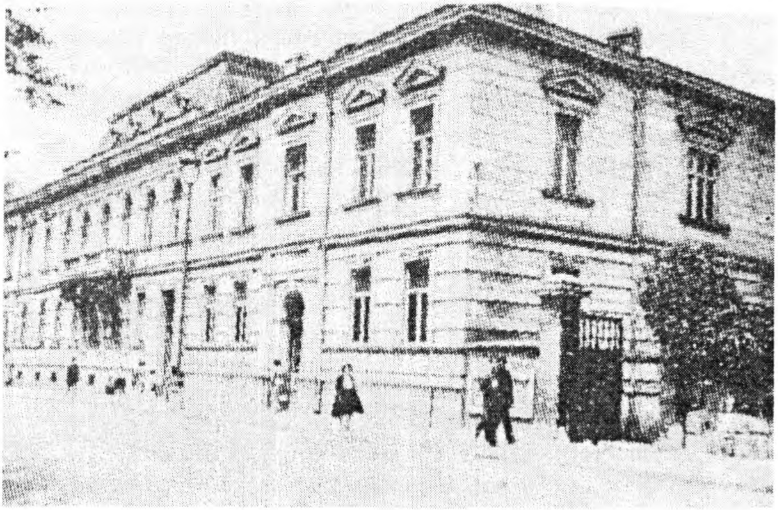
В цитирания материал четем следната бележка: „Редакцията търси портретите на II випуск — завършилите през 1893 г." Не знаем дали това желание на редакцията от 1942 г. се е осъществило, но може би проявеният от нея интерес тогава трябва да ни подсети и за някои дейности, свързани с историята на ФМИ, които и днес би било добре да вършим.

Първият доцент във факултета — негов възпитаник, е Георги Стоянов, назначен през 1909 г.

През 1907 г. Софийският университет, а заедно с него и Физико-математическият факултет прекарват тежко изпитание. Поради освиркването на Фердинанд от студентите, университетът е затворен, а професорите са уволнени. След няколко месеца университетът е отворен и са назначени други професори. След още няколко месеца обаче друг министър уволнява новоназначените професори и възстановява старите.

След две години ново нещастие сполетява математическия институт. Във връзка със скандал между проф. Сп. Ганев и студенти през 1910 г. професорите Ем. Иванов и Ат. Тинтеров демонстративно напускат уни-

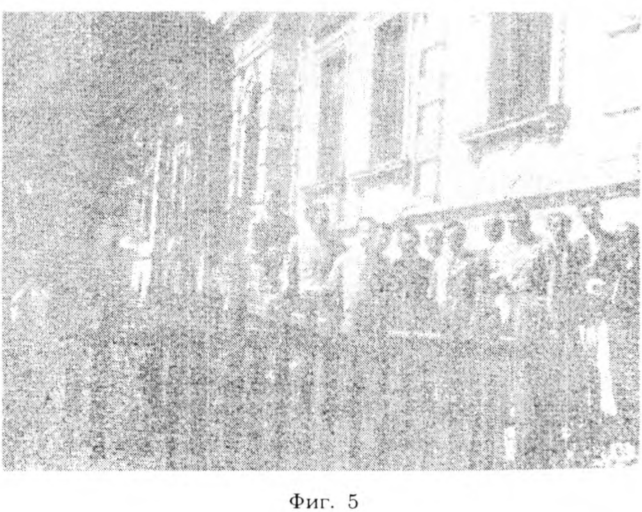
Фиг. 3



Фиг. 4

верситета, а през 1911 г. проф. Сп. Ганев е отстранен от института. През 1912 г., като участник в Балканската война, умира доц. Г. Стоянов. Така в математическия институт до 01.10.1914 г. остава само проф. А. Шоурек и младите асистенти, възпитаници на нашия университет, Кирил Попов, Иван Ценов и Любомир Чакалов.

Може би в този тържествен момент е интересно да си припомним и сградите, в които се е помещавала специалността математика. От 1889 г. до 1904 г. тя е била на ул. Московска 49 (фиг. 3), от 1904 г. до 1912 г. — на ул. Московска 13 (фиг. 4), от 1912 г. до 1931 г. — на ул. Сан Стефано 4 (фиг. 5 и фиг. 6 — сега там има голяма жилищна сграда), от 1931 г.



Фиг. 5

до 1940 г. отново е на ул. Московска 49, от 1941 до 1944 г. е на третия етаж на студентския дом на пл. Народно събрание 10 (фиг. 7). От края на 1944 г. до 1948 г. кабинетите на преподавателите са на ул. Раковска 108, четвъртия етаж, а занятията със студентите се водят в сградите на ул. Московска 13 и 49. От 1948 г. до 1951 г. кабинетите на преподавателите са на четвъртия етаж и тавана на ул. Стефан Караджа 4, а занятията продължават да се водят на ул. Московска 13 и 49. От 1951 г. до 1960 г. кабинетите на преподавателите и занятията по математика се преместват в северното крило на Ректората на Софийския университет на бул. Руски 15, от 1960 г. до 1971 г. те са на бул. Дж. Баучър 1, а от 1971 г. до днес са в сградата на бул. Дж. Баучър 5.

Фиг. 6

Ще си позволим да цитираме и бележката, направена от редакцията на Физико-математическото списание след материала в кн. 3 и 4 от 1940 г., в който са публикувани показаните 5 снимки: „Редакцията ще помести с удоволствие подробни бележки и описания на помещенията, в които са се обучавали наши физико-математици. Особено ценни биха били сведенията за първите години на Висшето училище: брой на аудиториите и

Фиг. 7

кабинетите, местоположението им, кой професор къде е чел и пр. и пр. И най-дребните факти представляват интерес.“ Няма да коментираме тази бележка, но се надяваме, че и тя ще накара всеки от нас да се позамисли над това доколко ние, увлечени в ежедневието, изпускаме някои важни неща, които нашите предшественици са разбирали и ценили по-добре от нас.

По нататък ще си позволя да дам последователно кратки сведения и за всеки от първите преподаватели по математика и механика в университета, включително и за Г. Стоянов.

1. **Теодор Монин** е роден в гр. Колин в Чехия през 1858 г., а е починал през 1893 г. пак в Чехия. Във ВУ работи от 01.09.1889 г. до 09.11.1891 г. Завършва Висшето техническо училище в Прага, след това е асистент там и като такъв е поканен в новооткритото у нас Висше училище. Тук чете курсове по Дескриптивна геометрия, Аналитична геометрия и Синтетична геометрия. През краткия си живот напечатва няколко научни статии и две книги, свързани с проективната и с елементарната геометрия.

2. **Емануил Иванов** е роден в Свищов през 1857 г., а е починал през 1925 г. в Кюстендил. Начално и средно образование завършва в Свищов, а от 1879 г. до 1883 г. завършва математика във Висшето техническо училище в Мюнхен. След завръщането си в България до юни 1885 г. работи като учител в Ломската гимназия, а от юни 1885 г. до 1890 г. е началник отдел и главен инспектор в Министерството на народното просвещение. През 1890/91 и 1893/94 г. е ректор на Висшето училище. Чете твърде разнообразни лекции в университета, където работи до 1923 г. (с прекъсване през 1907 г. и от 1910 г. до 1917 г.), а именно: Висш анализ, Диференциално смятане, Интегрално смятане, Диференциални уравнения, Въведение в теорията на комплексните функции, Теория на функциите, Елиптични интеграли и функции, Методически упътвания (елементарна математика). Интересно е да се отбележи, че проф. Ем. Иванов започва да чете лекции по елементарна математика точно тогава, когато за пръв път в света такива лекции започва да чете в Германия в курса „Елементарна математика от гледна точка на висшата математика“ известният немски математик Феликс Клайн — един от най-активните ръководители на международното реформаторско движение в обучението по математика в края на 19 век и началото на 20 век. В каталога на библиотеката на ФМИ фигурира картонче за циклостилни записки по „Висша алгебра“ от Емануил Иванов (360 стр.), но за съжаление те са изчезнали от

библиотеката. Научните му интереси са били главно в областта на теорията на числата. Написал е и немалък брой методически статии, които са допринесли съществено за развитието на методиката на обучението по математика у нас. Особено ценна е статията му „Принципите на днешното математическо обучение в средните училища“, чието съдържание и днес е актуално. Бил е член на Висшия учебен съвет в Министерството на просвещението.

3. **Антон Шоурек** е роден през 1857 г. в гр. Писек, Чехия, и е починал през 1926 г. в София. Учи математика в университета във Виена и във Висшето техническо училище в Прага. В университета чете много на брой курсове: Аналитична геометрия, Синтетична геометрия, Дескриптивна геометрия, Проективна геометрия, Кинетика на пространството, Диференциална геометрия, Висша геометрия, Теория на кривите линии и повърхнини, Алгебрически анализ, Висша алгебра, Методика на дескриптивната геометрия и геометричното чертане. Превел е от чешки на български език учебник по геометрия, учебник по алгебра и една логаритмична таблица. През 1914 г. издава голям университетски учебник по дескриптивна геометрия, а през 1926 г. — „Основи на проективната геометрия“. В библиотеката на ФМИ се съхраняват циклостилни записки на негови лекции по „Диференциална геометрия“ от 1914 г. и „Аналитична геометрия“ от 1913 – 1914 г. И той като Ем. Иванов е написал не малко методически статии, свързани с обучението по математика. Автор е и на 7 учебника за българските средни училища.

4. **Владислав Шак** е роден през 1869 г. в гр. Налжови, Чехия, а умира през 1941 г. в България. От 1891 г. до 1894 г. във Висшето училище чете курсовете Сферична тригонометрия, Алгебричен анализ, Аналитична геометрия и Алгебра. В България Владислав Шак развива богата дейност като учител, преводач, журналист и публицист.

5. **Михаил Момчилов** е роден през 1862 г. в гр. Горна Оряховица. Завършил е реална гимназия в гр. Николаев, Русия, а през 1888 г. завършва инженерство в политехниката в Дрезден. Във Висшето училище работи от 01.10.1891 г. до 01.03.1893 г., където чете лекции по Аналитична механика.

6. **Атанас Тинтеров** е роден през 1857 г. в Калофер, а е починал през 1927 г. в София. Учил е отначало в Калофер при бащата на Хр. Ботев, Ботьо Петков, а след това в Русия, където през 1881 г. завършва математика в Новоросийския университет в Одеса. В университета чете лекции по Алгебричен анализ, Основи на висшата математика, Диференциално и интегрално смятане, Теория на вероятностите и Геометрични приложения на интегралното смятане. В Университетската библиотека има запазени два тома литографски издадени записки на студенти по курсове на Ат. Тинтеров. Други учебници или статии от него за сега не са ми известни.

7. **Спиридон Ганев** е роден през 1869 г. в гр. Сливен, а е починал през 1961 г. в София. Завършва основно образование в гр. Сливен, средно в Пловдив (учител му е Антон Шоурек), а висше в Чехия и Белгия. През 1893 г. защитава докторска дисертация в Лиеж. След завръщането си в България отначало е назначен за учител в Софийската мъжка гимназия, а след това за доцент и по-късно професор във Висшето училище. Чете лекции по Висша алгебра, Аналитична механика, Механична теория на топлината, Теория на алгебричните уравнения, Понятие за функция в елементарната математика. До 1911 г. взема активно участие в работата по обновяване на обучението по математика в българските училища.

8. **Георги Стоянов** е роден през 1872 г. в гр. Нова Загора, а е починал, както вече отбелязахме, през 1912 г. Завършва математика и физика през 1900 г. От 17.08.1909 г. до смъртта си е редовен доцент в катедрата по основи на висшата математика. Чете лекции по Въведение във висшата математика, Въведение по висша алгебра, Основи на висшата математика, Диференциално смятане, Интегрално смятане и Диференциални уравнения. В алманаха са посочени като негови трудове 10 заглавия с математическо, методическо и историческо съдържание.

За съдържанието на курсовете, които е чел, до неотдавна нямахме никаква информация. На негово име в библиотеката на факултета до 1988 г. не фигурираше названието и на нито едни литопечатни записки. В склада на музея в гр. Елена обаче през 1988 г. случайно попаднах на четири тома литопечатни записки със следните заглавия: Въведение във

висшата алгебра, Лекции по интегрално смятане (1911 – 1912 г.), Лекции по висша алгебра (1914 г.), Диференциална геометрия. Върху първите два тома изрично е отбелязано, че са лекции на Георги Стоянов. За последните два тома за съжаление не е отбелязано по чии лекции са. Може би ще е интересно да се отдели известно внимание и да се попроучат тези записки. В един от томовете има „Упражнение по въведение във висшата математика“ и „Упражнение по Диференциално смятане“.

Със съдействието на ръководството на музея в гр. Елена през 1988 г. успях да осигуря за библиотеката на ФМИ ксерокс-копия на посочените по-горе четири тома литопечатни записки.

Позволих си да се спра по-подробно на намерените в гр. Елена материали, защото ми се струва, че по отношение на издирването и на този род информация е добре да се поучим от цитираната по-горе забележка на редакцията на Списанието на физико-математическото дружество, свързана със сградите на факултета. Може би още много подобни материали стоят неизползвани и постепенно се похабяват в различни складове, мазета или лични библиотеки на наследници на математици, които са твърде далеч от математиката.

Една от най-важните заслуги на нашите математици от първото поколение е тяхната грижа за подготовката на техни наследници. Независимо от финансовите трудности, които е имало в края на 19 и началото на 20 век у нас, математиците от първо поколение са успели да изпратят най-добрите свои ученици — вече като асистенти, на специализации във водещи университети в Германия или Франция.

Според академик Л. Илиев първият етап от развитието на математиката във Физико-математическия факултет на Софийския университет завършва през 1914 г., когато на 1 октомври същата година биват избрани за доценти младите и обещаващи научни работници Л. Чакалов, Ив. Ценов и К. Попов, разширили и задълбочили вече своите знания и в чужбина.

През 1920 г. и 1922 г. групата на младите доценти-математици във ФМФ се разширява съответно с Димитър Табаков и Н. Обрешков. Интересно е да се отбележи, че Н. Обрешков е избран за доцент в катедрата по Висша алгебра само две години след завършването на висшето си образование по математика във ФМФ на СУ.

Докато първото поколение математици във ФМФ създава организация и съдържание на обучението, близки до организацията и съдържанието на обучението във водещите световни университети, то второто поколение математици издига на световно равнище научната работа по математика у нас. Тъй като техните научни приноси са твърде много и не могат да бъдат обхванати в доклад за тържествено честване като настоящото, ще си позволя да не правя опит да ги засягам. Ще се спра накратко само на някои от основните моменти от техните биографии, като ще следвам годините на раждането им.

1. **Проф. Димитър Стефанов Табаков** е роден на 09.11.1879 г. в гр. Сливен. Завършил е математика и физика във Висшето училище в София през 1903 г. Специализирал е в Страсбург през 1903/1904 г. и в Нанси през 1904/1905 г. Доктор е на университета в Пиза от 1929 г. Бил е асистент по математика в Софийския университет от септември 1905 г. до 01.09.1907 г., редовен доцент по геометрия от 11.03.1920 г. до 06.03.1923 г., извънреден професор от 07.03.1923 г., а редовен професор и титуляр на катедрата по геометрия от 1941 г. до пенсионирането му на 01.10.1948 г. Чел е лекции по Аналитична геометрия, Проективна геометрия, Дескриптивна геометрия, Диференциална геометрия. Научните му интереси и приноси са в областта на геометрията. Заслужил деятел на науката е от 1966 г., а народен деятел — от 1969 г. Умира през 1973 г.

2. **Акад. Кирил Атанасов Попов** е роден на 03.05.1880 г. в гр. Шумен. От 1884 г. семейството му се премества във Варна, където Кирил Попов завършва средно образование. През 1897/98 г. е нередовен учител във варненското основно училище „Отец Паисий“ и публикува в сп. „Работник“ първия си популярен труд върху историята на астрономията. През 1898 г. постъпва във физико-математическия факултет на Висшето училище в София, в специалността Математика и физика, която завършва през 1902 г. През 1902/1903 и 1903/1904 учебна година е учител в класното училище в Свищов, а през 1904/1905 г. — в педагогическото училище в Казанлък. От 1905 г. е асистент по астрономия във Физико-математическия факултет в София, а от 1906 г. до 1910 г. специализира последователно в Мюнхен, Хайделберг, Ница и Париж. В Париж разработва докторска дисертация под ръководството на големия френски учен Анри Поанкаре. От 1914 г. е редовен доцент, от 1922 г. — извънреден професор, а от 1922 г. до 1952 г. е редовен професор и

ръководител на катедрата по Диференциално и интегрално смятане във ФМФ на Софийския университет. Чел е лекции по Небесна механика и балистика в Сорбоната в Париж, а отделни лекции — в почти всички известни европейски университети в Берлин, Рим, Виена, Белград, Гьотинген, Мюнхен, Хамбург, Торино и др. През 1947 г. е избран за академик. Удостоен е със званието „Народен деятел на науката". Умира през 1964 година.

3. **Акад. Иван Ценов Ангелов** е роден на 15.01.1883 г. във Враца. Баща му е един от активните участници във врачанския таен революционен комитет по време на турското робство. През 1899/1900 г. Иван Ценов завършва шестокласното училище и една година е начален учител в с. Нефела, Врачанско. След това постъпва в Софийската мъжка гимназия, която завършва през 1903 г. и отново става учител в едно основно училище във Враца. През 1904 г. постъпва студент във ФМФ в София, където учи пет семестъра. Шестия семестър учи в Белград, а седмия и осмия — в Загреб. След това се връща в София, завършва математика през 1909 г. и учителства в Софийската мъжка гимназия. Впоследствие става редовен асистент във ФМФ и е изпратен за две години на специализация в Париж. От 1914 г. е доцент, от 1919 г. — извънреден професор, а от 1922 г. до 1951 г. е редовен професор и ръководител на катедрата по Аналитична механика във ФМФ на Софийския университет. От 1925 г. Иван Ценов е член-кореспондент на БАН, а от 1928 г. — академик. През периода 1925–1930 г. е бил декан на ФМФ. Участвал е с научни доклади в различни научни конгреси в Стокхолм, Париж, Прага и др. Удостоен е с различни отличия и държавни награди. Умира през 1967 г.

4. **Акад. Любомир Николов Чакалов** е роден на 06.02.1886 г. в гр. Самоков. Завършил е математика и физика в Софийския университет през 1908 г. Специализирал е в Лайпциг и Гьотинген през 1910/1911 г. и 1911/1912 г., а в Париж и Неапол през 1924/1925 г. Доктор е на университета в Неапол от 1925 г. Асистент е по математика в Софийския университет от 10.09.1908 г. до 30.09.1914 г., редовен доцент в катедрата по Висш анализ — от 01.10.1914 г. до 13.10.1919 г., извънреден професор — от 14.10.1919 г. до 31.03.1922 г., а редовен професор, титуляр на катедрата по Висш анализ — от 01.04.1922 г. до пенсионирането му на 15.10.1952 г. Декан на Физико-математическия факултет е през 1923/1924 г., а ректор на Софийския университет — през 1943/1944 г. Чел е лекции по Теория на функциите, Диференциални уравнения, Вариационно смятане, Избрани въпроси от елементарната математика, Елиптични функции, Тригонометрични редове, Диференциално и интегрално смятане, Висша

алгебра, Аналитична геометрия, Дескриптивна геометрия. Научните му интереси и приноси са в областта на реалния и комплексен анализ, теория на числата, диференциални уравнения, конструктивна теория на функциите, алгебра, механични квадратури, елементарна математика, обучението по математика. Основава и ръководи в продължение на много години един от най-авторитетните научни кръжоци по математика в България (кръжокът по анализ), където получават първата си научна закалка много наши научни работници. Акад. Л. Чакалов поставя началото на националните ученически математически олимпиади в България през 1950 – 1951 г., превърнали се постепенно в основен стимул за разгръщане на една от най-богатите в света система за извънкласни дейности по математика. Член е на Българското физико-математическо дружество и на много чуждестранни научни дружества, член на Кралската чешка академия на науките, Географското дружество в Лима, Варшавската академия на науките, действителен член на Българската академия на науките от 1930 г., народен деятел на науката от 1963 г. Умира на 11.09.1963 г.

5. **Акад. Никола Димитров Обрешков** е роден на 06.03.1896 г. в гр. Варна. Завършил е математика и физика в Софийския университет през 1920 г. Доктор по математика на университета в Палермо е от 1932 г., а на Сорбоната в Париж — от 1933 г. Асистент е в катедрата по Диференциално и интегрално смятане в Софийския университет от 01.10.1920 г. до 31.03.1922 г., редовен доцент в катедрата по Висша алгебра — от 01.04.1922 г. до 31.03.1925 г., извънреден професор — от 01.04.1925 г. до 06.12.1928 г., редовен професор и завеждащ катедрата по Висша алгебра — от 07.12.1928 г. до края на живота си. Директор на Математическия институт на БАН е от 1951 г. до края на живота си. Чел е лекции по Висша алгебра, Теория на вероятностите, Сумиране на разходящите редове, Теория на числата. Чел е лекции по покана в университетите в Хамбург, Берлин, Женева, Рим, Палермо, Лайпциг, Дрезден и др. Научните му интереси и приноси са в областта на класическата алгебра, анализа, теория на вероятностите и математическата статистика, интегрална геометрия, топология, механика, математическа физика, теория на числата. Действителен член е на Българската академия на науките от 1945 г. Народен деятел на науката е

от 1963 г. Като израз на уважение и признателност към делото на Никола Обрешков след смъртта му бе учредена награда „Н. Обрешков“, която се присъжда за особени приноси в областта на физико-математическите науки. Умира на 11.08.1963 г.

Третото поколение професори-математици, работили във Физико-математическия факултет, повежда проф. Георги Брадистилов, а след него по годините на раждане се нареждат проф. Благовест Долапчиев, акад. Боян Петканчин, проф. Алипи Матеев, акад. Любомир Илиев, чл. кор. проф. Ярослав Тагамлицки. Към същото поколение трябва да добавим и доцентите Петко Иванов, Йордан Дуйчев и Димитричка Шопова. Това поколение продължи традицията на второто да се стреми да поддържа равнището на науката на ниво, близко до световното, и да подготвя кадри, които да бъдат достойни негови заместници. Особено продуктивен в това отношение беше научният кръжок по диференциално и интегрално смятане на проф. Я. Тагамлицки. Третото поколение, както и първото, отделя особено внимание на изпращането на специализанти в чужбина. Това поколение продължи традицията да се грижи и да се чувства отговорно за обучението по математика у нас на всички равнища и даде добър пример в това отношение на следващите поколения математици.

При провеждането на Първата международна математическа олимпиада, а и няколко години след това, научен ръководител на българския отбор е проф. А. Матеев. Завършващата подготовка на отбора от тогава и до днес се провежда от екипи, съставени предимно от математици от факултета и от МИ при БАН. След проф. А. Матеев научните ръководители на българските отбори също винаги са били от факултета или от МИ на БАН.

През 1961 г. по инициатива на младите тогава асистенти от факултета Йордан Кучинов и Иван Ганчев за пръв път не само в България, а и в света, се организира Национален радио-конкурс по математика. Той осигуряваше на всички ученици с повишен интерес по математика в продължение на 10 години еднакви възможности за развиване на умения за решаване на задачи със сложни решения. През този период основната част от участниците на международната математическа олимпиада бяха активни участници и в радио-конкурса. Много от тях днес са между водещите научни работници по математика в България.

Макар и с течение на времето все по-трудно да може да се отделят едно от друго новите поколения математици, смятам, че може да отделим четвъртото поколение математици във ФМФ. Това поколение може би трябва да започнем с доц. Александър Гьонов, първият математик-аспирант защитил кандидатска дисертация във факултета през 1953 г., и да продължим с проф. Ив. Чобанов, проф. Вл. Чакалов, доц. Боян Пенков, проф. Дойчин Дойчинов, доц. Васил Диамандиев, проф. Рачо Денчев, доц. Димитър Димитров, доц. Кирил Дочев, доц. Димитър Добрев, доц. Никола Мартинов, доц. Татяна Аргирова, проф. Тодор Генчев, акад. Благовест Сендов, проф. Запрян Запрянов, проф. Грозьо Станилов, доц.

Иванка Иванова, проф. Димитър Скордев, доц. Йордан Кучинов, доц. Иван Ганчев, доц. Костадин Петров, проф. Иван Проданов, доц. Анани Лангов, проф. Николай Хаджииванов, ст. н. с. Михаил Гаврилов, доц. Димитър Шишков.

Особена заслуга на третото поколение математици е, че главно по времето на неговата активна дейност се поставя началото на информатиката у нас със създаването през 1959/1960 г. със студенти по математика от III курс на профил „Изчислителна математика". Както за поставянето на това начало, така и за по-сетнешното развитие на изчислителната математика особено големи заслуги има акад. Л. Илиев.

Около 10 години преди това във ФМФ се поставя и началото на изучаването на Методика на обучението по математика (МОМ) в самия факултет, в рамките на обучението на студентите, готвещи се за учители. Това явление изиграва съществена положителна роля за изграждането на МОМ като съвременна система от научни знания и на кадри, които да развиват тази система на равнището, на което тя е в страните с развито училищно обучение по математика. Особено големи заслуги за успехите ни в това отношение имат доц. П. Иванов и проф. А. Матеев.

Ще си позволя да припомня съвсем накратко как беше поставено споменатото по-горе начало на развитието на информатиката (такъв термин, разбира се, тогава нямаше) преди 40 години във ФМФ на Софийския университет и в Математическия институт на БАН. По онова време тези две научни институции работеха така единно, особено при поставянето на началото на изчислителната математика, че да се търси някакъв приоритет на едната спрямо другата, според мен е безсмислено. За това време в доклада си на тържествената сесия по случай 100 години от създаването на факултета проф. П. Бърнев каза: „Цялото развитие на информатиката започна в Математическия институт при БАН и във ФМФ при Софийския университет. Това развитие беше закономерно и неизбежно, но върху него влияеха и субективни фактори. Пречка беше създалото се в СССР отрицателно отношение към кибернетиката, пренесено у нас от Т. Павлов. Благоприятен фактор беше енергичната и добре планирана дейност на Любомир Илиев. Тази дейност се осъществяваше в три основни направления:

1. Подготовка на кадри.
2. Създаване на организационни възможности.
3. Осигуряване на техника."

Казаното за дейността на акад. Л. Илиев като фактор в посочените три направления в началото на 60-те години на 20 век е една твърде скромна оценка. Макар и да бях тогава страничен наблюдател на тази дейност, смея да твърдя, че поне тогава той беше главният ѝ стратег не само в МИ на БАН и ФМФ на Софийския университет, но и в България. Негови първи помощници в това отношение бяха младите тогава наши колеги Благовест Сендов и Боян Пенков. Те, заедно с Апостол Обретенов, бяха и първите преподаватели в създадената през 1959/1960 г. първа специализация у нас по изчислителна математика. Първите специализанти в тази специализация бяха Димитър Шишков, Георги Пенчев, Веселин Спиридонов и Стефан Генчев.

През месец юни 1961 г. беше изпратена на дългосрочна работа в Дубна в бившия СССР за овладяване на работата с вече създадените там ЕСМ първата българска група от математици и инженери.

Такова беше началото, а за изключително бързото развитие на изчислителната математика след това във факултета и МИ, а от там и в България не се считам за компетентен да говоря. Само ще отбележа, че лично на мен ми е твърде неприятно, когато в по-ново време се говори за големите успехи на българските информатици, те да се приписват или на други институции, или да се разглеждат като появили се някак от само себе си.

През 1963 г. Математическият факултет се отдели като самостоятелно звено на Софийския университет с първи декан проф. А. Матеев. Тъй като това отделяне беше нежелано от колегите физици, дейността на Математическия факултет беше много затруднена. Той приличаше на изгонено младо семейство от недоволни възрастни родители, без да му се даде някаква покъщнина. Благодарение на общите усилия на цялата математическа колегия тогава под ръководството на първия, а след това и на втория декан на Математическия факултет, доц. Александър Гьонов, и с помощта на колегите от МИ на БАН работата във факултета според мен се развиваше много добре.

На 08.12.1970 г. студентите и преподавателите от Факултета по математика и механика преживяхме изключително събитие. Наш гост по време на тържествената вечеря по случай студентския празник беше конструкторът на първата в света електронна сметачна машина, американският учен от български произход Джон Атанасов. По инициатива на тогава младия професор Бл. Сендов, той беше поканен като гост на БАН и удостоен от Президиума на Народното събрание с орден „Кирил и Методий“ I степен.

В края на 60-те години интеграционните връзки на Факултета по математика и информатика с МИ на БАН се разшириха и задълбочиха толкова много, че съвсем естествено се достигна до създаването на ЕЦПКММ през 1970 г. (съкратено ЕЦММ), когато вече декан на МФ беше доц. Дойчин Дойчинов. Към момента на създаването на ЕЦ през 1970 г. в Математическия факултет имаше 8 катедри: Висша алгебра, Геометрия, Диференциално и интегрално смятане, Висш анализ, Механика, Обща и приложна математика, Изчислителна математика и Методика на обучението по математика. При създаването на ЕЦММ от катедрите на МФ и от секциите на МИ при БАН бяха създадени 13 интегрирани звена, наречени сектори. Създадена беше и нова структура на организацията на специалностите, съответстваща на едва сега утвърдената за цялата ни страна бакалавърско-магистърска система. Създадена беше също така и нова организация на учебния процес, която впоследствие беше почти изоставена.

По различен начин се възприе тогава създаването на ЕЦММ от различни колеги както от МФ, така и от МИ. Нещо повече, различните оценки на създаването и съществуването на ЕЦММ като явление и досега са факт. Тържественият характер на форум като настоящия обаче не ми позволява да се опитвам да оценявам кои са положителните и кои — отрицателните последствия. Може би единственият факт, който тогава се приемаше единодушно като положително последствие от интеграцията, беше даването на самостоятелна сграда на МФ и отпускането на значителни финансови средства за обзавеждането ѝ. Каква научна продукция даваха 13-те сектора, не се наемам да оценявам, но смятам, че имаше поне едно явление, свързано с научната дейност в тях, което повечето от колегите оценяват като положително. Това са ежегодните научно-отчетни конференции, които се провеждаха по сектори.

Докато беше звено, функционално интегрирано с МИ на БАН, Математическият факултет два пъти смени името си: отначало през 1970 г. той беше наречен Факултет по математика и механика, а по-късно — Факултет по математика и информатика. ЕЦММ беше разформирован в края на 1988 г. и от началото на 1989 г. Факултетът по математика и информатика съществува като самостоятелно звено само на СУ „Св. Кл. Охридски“. Тогава в него бяха сформирани 14 катедри и 3 лаборатории.

През последните 10–15 години изменение претърпя и номенклатурата на специалностите във факултета, която е в съответствие с новите потребности на обществото.

Преди да изкажа заключителните думи на своя доклад, смятам, че трябва да спомена имената и на колегите, които бяха декани на факултета след доц. Дойчин Дойчинов. Това са: акад. Бл. Сендов, доц. Димитър Димитров, проф. Рачо Денчев, проф. Генчо Скордев, проф. Боян Димитров, ст. н. с. Михаил Гаврилов, проф. Емил Хорозов. На всички тях и на техните помощници трябва да сме благодарни за усилията, които полагаха, за да може факултетът не само да съществува, но и независимо от различните трудности той да се развива, да се развиват неговите кадри, да се развива неговата материална база, да подготвя възпитаници, търсени и признавани не само у нас, но и в чужбина, да подпомага развитието на математическите кадри и в създадените след него други ВУЗ в България.

В заключение ще си позволя да предложа на почитаемото събрание да помислим дали не е добре, както през 1882 г. Иречек, Генчев и Вацов са пофантазирали за създаването на български университет и ние някога да пофантазираме да превърнем сградата на нашия факултет в нов вид средище; в средище, което и със своите коридори, фоайета и стени учи на математика, възпитава уважение към този най-висш продукт на човешката логика, към неговата история, към неговите творци, като заменим висящите по цели месеци по стените окъсани плакати и съобщения, понякога нямащи нищо общо с основната дейност на факултета, със скромна украса, представяща науката математика и нейните творци.

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

REMARK ON THE NON-INTEGRABILITY OF THE PERTURBED MOTION OF THE PARTICLE IN A CENTRAL FIELD IN CONSTANT CURVATURE SPACES

OGNYAN CHRISTOV

In the recent article [1] Kozlov and Harin generalize the motion of a particle in a central field to the case of constant curvature spaces. In this remark we show that the problem of the non-integrability of the perturbed motion in a central field on the sphere and on the Lobachevski's space is reduced to the flat case considered by Holmes and Marsden using Melnikov integrals.

Keywords: constant curvature spaces, central field motion, non-integrability, Melnikov integrals

MSC 2000: 37J30, 70F05, 70H07

1. INTRODUCTION

In [1] Kozlov and Harin generalize the motion of a particle in a central field to the case of constant curvature spaces. They study mainly the cases when all orbits are closed. It turns out that these cases are analogous to the gravitational potential and to the potential of an elastic string. Another important result is that the integrability of generalized two-center problem on a constant curvature surface is established and it is shown that the integrability remains even "elastic forces" are added.

It is natural to consider also the non-integrability of perturbed motion of the particle in a central field in constant curvature spaces. More precisely, we consider the interaction potentials which allow separatrices in the dynamics of the unperturbed problem and these separatrices split after small perturbation. As a tool we shall use the so-called Melnikov's integrals [2].

In this note we shall show that the problem of non-integrability of the perturbed motion in a central field on the sphere and on the Lobachevski's spaces can be reduced to the flat case, considered by Holmes and Marsden [3]. We shall first recall briefly their result, which is an example of a more general treatment of Melnikov's theory.

Consider the perturbed Hamiltonian

$$H^{\varepsilon} = H^0(r, p_r, p_\theta) + \varepsilon H^1(r, p_r, \theta, p_\theta)\,, \tag{1}$$

where $m = 1$, (r, θ) are the usual polar coordinates,

$$H^0 = \frac{1}{2}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) + V(r)\,. \tag{2}$$

Let $V(r)$ be a potential with a single maximum, so that for suitable values of $p_\theta \neq 0$ the effective potential has a minimum at r_- and a maximum at r_+ $(r_- < r_+)$ and

$$V(r) + p_\theta^2/(2r^2) \to \infty \quad \text{as} \quad r \to 0\,.$$

Thus H^0 has a homoclinic orbit

$$(\overline{r}(t), \overline{p_r}(t),\ \overline{\theta}(t) + \theta_0,\ \overline{p_\theta})\,,$$

where

$$\overline{r}(t) \to r_+, \quad \overline{p_r}(t) \to 0 \quad \text{as} \quad t \to \pm\infty\,,$$

$$\overline{p_r}(0) = 0, \quad \overline{p_\theta} \neq 0, \quad \overline{\theta}(t) = \int_0^t \Omega(t)\,\mathrm{d}t.$$

The derivative $\Omega(t) = \partial H^0/\partial p_\theta$ is evaluated on the homoclinic orbit.

Proposition 1 ([3]). *Let the Melnikov integral*

$$M(\theta_0) = \int_{-\infty}^{\infty} \left\{H^0, \frac{H^1}{\Omega}\right\}(t, \theta_0)\,\mathrm{d}t\,,$$

where $\{\cdot\}$ is the Poisson bracket, have simple zeros as a function of θ_0. Then for a sufficiently small ε, the system (1) has Smale horseshoes on the energy surface $H^{\varepsilon} = h$ and hence it is non-integrable.

The note is organized as follows. In section 2 we consider the motion of a particle in a central field on S^3, following Kozlov and Harin [1]. A simple construction reduces the problem to the flat case, already discussed in the foregoing. In section 3 we consider briefly the situation on Lobachevski's space. We conclude the note with several remarks.

In this section we consider the analogue of the classical problem in a central field on S^3. We follow closely Kozlov and Harin [1]. Consider the particle P with unit mass moving in a field of force with the potential V, depending only on the distance between the particle and some fixed point M (say the north pole) on the sphere S^3. Let θ be the length of the arc of the great circle connecting the points P and M. It is well-known that the potential of gravitational interaction satisfies the Laplace equation. Then V has to be a function of the angle θ only and the Laplace equation has to be replaced by the Laplace-Beltrami's one

$$\triangle V = \sin^{-2}\theta \frac{\partial}{\partial\theta}\left(\sin^2\theta \frac{\partial V}{\partial \theta}\right) = 0 .$$

Its solution is

$$V = -\frac{\gamma}{\tan\theta} + \alpha ,$$

where α, $\gamma > 0$ are constants. It is seen that in addition to the attracting center M, this field has a repulsive center at the antipodal point M'. It is proven also that when V is an arbitrary function of θ, the trajectories of P lie on two-dimensional sphere containing points M and M'.

Let (θ, φ) be the spherical coordinates on the above mentioned two-dimensional sphere. Then the Lagrangian is

$$L = \frac{1}{2}\left(\dot\theta^2 + \sin^2\theta\dot\varphi^2\right) - V(\theta) , \quad V(\theta) = U(\tan\theta) .$$

Introduce the polar coordinates by

$$\varphi = \varphi, \quad r = \operatorname{cotan}\frac{\theta}{2} ,$$

see Fig. 1. (A slightly different construction was used in Dubrovin *et al.* [4] or Kozlov *et al.* [1].) In these variables the Lagrangian becomes

$$L = \frac{1}{2}\left(\frac{4\dot r^2}{(1+r^2)^2} + \frac{4r^2\dot\varphi^2}{(1+r^2)^2}\right) - \tilde U(r) .$$

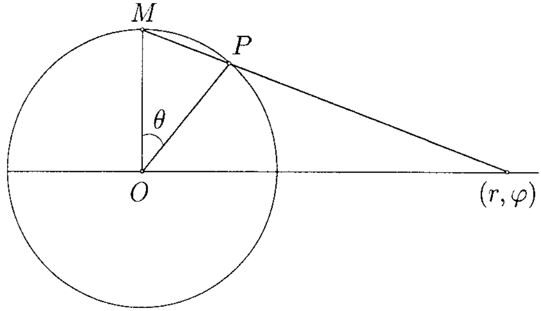


Fig. 1

Next, introduce the "new" time by

$$ds = (1 + r^2)dt/2\,.$$

The 'prime' denotes differentiation with respect to this new time variable, $' = d/ds$. Then

$$L = \frac{1}{2}\left(r'^2 + r^2(\varphi')^2\right) - \tilde{U}(r)\,.$$

Passing to the Hamiltonian, we get

$$H^0 = \frac{1}{2}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) + \tilde{U}(r)\,, \tag{3}$$

which has *exactly* the form (2) and hence it is integrable.

Consider a small perturbation of the Hamiltonian (4), namely

$$H^\varepsilon = H^0 + \varepsilon H^1(r, p_r, \varphi, p_\varphi). \tag{4}$$

The following proposition is true:

Proposition 2. *Given a potential $V(\theta)$ such that $\tilde{U}(r)$ has a single maximum and for suitable values of $p_\varphi \neq 0$, the effective potential has minimum at r_- and maximum at r_+ ($r_- < r_+$) and $\tilde{U}(r) + p_\varphi^2/(2r^2) \to \infty$ as $r \to 0$. Thus, $H^0 = h$ has a homoclinic orbit. Then, if*

$$M(\varphi_0) = \int_{-\infty}^{\infty} \left\{H^0, \frac{H^1}{\Omega}\right\}(t, \varphi_0)\, dt\,,$$

evaluated on the homoclinic orbit, has simple zeros, the system (4) *is non-integrable.*

Remark 1. For instance, the class of the potentials of the form

$$\tilde{U}(r) = ar^2 - br^4\,,$$

with positive constants a, b, satisfies the requirements we need.

3. MOTION IN A CENTRAL FIELD IN LOBACHEVSKI'S SPACE

This case is similar to that of the previous section and therefore it will be briefly discussed. Let (χ, φ) be the polar geodesic coordinates. Then the potential $V(\chi)$, analogous to the gravitational potential, has to satisfy the Laplace-Beltrami equation

$$\Delta V = \sinh^{-2}\chi \frac{\partial}{\partial \chi}\left(\sinh^2\chi \frac{\partial V}{\partial \chi}\right) = 0\,,$$

see [5]. Its obvious solution is

$$V = -\frac{\gamma}{\tanh \chi} + \alpha\,,$$

where α, $\gamma > 0$ are constants.

Consider now the Lagrangian of the point with unit mass with more general potential $V(\chi) = U(\tanh\chi)$:

$$L = \frac{1}{2}\left(\dot{\chi}^2 + \sinh^2\chi\dot{\varphi}^2\right) - U(\tanh\chi)\,.$$

Introduce the polar coordinates

$$\varphi = \varphi, \quad r = \tanh\frac{\chi}{2} \quad (r < 1)\,.$$

Then L reads

$$L = \frac{1}{2}\left(\frac{4\dot{r}^2}{(1-r^2)^2} + \frac{4r^2\dot{\varphi}^2}{(1-r^2)^2}\right) - \tilde{U}(r)\,.$$

Let us introduce the "new" time, whose definition in the case under study is

$$\mathrm{d}s = (1-r^2)dt/2\,.$$

Passing to the Hamiltonian, we get now

$$H^0 = \frac{1}{2}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) + \tilde{U}(r)\,, \tag{5}$$

which has *exactly* the form (2).

Consider, once again, the perturbed system with small Hamiltonian perturbation

$$H^\varepsilon = H^0 + \varepsilon H^1\left(r, p_r, \varphi, p_\varphi\right)\,. \tag{6}$$

The following proposition is true.

Proposition 3. *In the conditions of the Proposition* 2 (*note that here* $r < 1$), *if the Melnikov function* $M(\varphi_0)$ *has simple zeros, the system* (6) *is non-integrable.*

We shall conclude the note with several remarks.

Remark 2. Similarly to Holmes and Marsden [3], it is to be noted that for almost all choices of $\tilde{U}(r)$, the function $M(\varphi_0)$ has simple zeros.

Remark 3. The foregoing problem can be considered as well in higher dimensions. Then the Melnikov's vector can be used (see Wiggins [6]), but certain KAM conditions are needed.

Remark 4. It is seen that the analogue of the classical Kepler problem does not fall in our cases, since it does not possess a homoclinic orbit. We believe that the methods, used by Yoshida [7], can be applied to it for certain classes of perturbations, see also [8].

Acknowledgements. The partial support of the Bulgarian Ministry of Education and Science under Grant No MM 608/96 is acknowledged.

REFERENCES

1. Kozlov, V., A. Harin. Kepler's problem in constant curvature spaces. *Cel. Mech. and Dyn. Astronomy*, **54**, 1992, 393–399.
2. Melnikov, V. On the stability of center for time periodic perturbations. *Trans. Moscow Math. Soc.*, **12**, 1963, 1–57.
3. Holmes, P., J. Marsden. Horseshoes and Arnold diffusion for Hamiltonian systems on Lie groups. *Indiana Univ. Math. J.*, **32**, 1983, 273–309.
4. Dubrovin, B., S. Novikov, A. Fomenko. Contemporary Geometry. Vol. 1, Moscow, Nauka, 1979. (In Russian.)
5. Helgason, S. Differential Geometry and Symmetric spaces. Academic Press, New York, 1962.
6. Wiggins, S. Global Bifurcation and Chaos. Analytical Methods. Springer, New York, 1988.
7. Yoshida, H. Non-integrability of a class of perturbed Kepler problems in two dimension. *Phys. Lett. A*, **120**, 1987, 388–390.
8. Vigo-Aguiar, I. Non-integrability of some Hamiltonian systems in polar coordinates. *J. Phys. A, Math. Gen.*, **30**, 1997.

Received November 11, 1999

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 James Bourchier Blvd.
BG-1164 Sofia, Bulgaria
E-mail: `christov@fmi.uni-sofia.bg`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ"
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI"
FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

EFFECTIVE ENUMERATIONS OF FAMILIES OF RECURSIVE FUNCTIONS*

ANGEL V. DITCHEV

In this paper necessary and sufficient conditions for a family of finite functions and a family of totally recursive functions to have a universal partially recursive function are given.

Keywords: totally recursive function, universal function for a family of recursive functions, strong recursive sequence

2000 MSC: 03D20

The problem of finding necessary and sufficient conditions for a given family of partially recursive (p.r.) functions (recursively enumerable (r.e.) sets) to have a universal p.r. function (r.e. set) is one of the interesting problems in the Recursion theory. For example, if we want to find a recursive model for a given recursively enumerable theory in some cases, we have to know if a given family of recursive functions has a universal recursive function or not. It is well-known that the family of all p.r. functions (r.e. sets) has a universal p.r. function (r.e. set), while the family of all recursive functions (totally defined on some $\mathbb{N}^n$) has no universal recursive function. On the other hand, in the works [2, 3, 5] a related problem is considered. Some necessary and some sufficient conditions for the family of all recursive functions and some finite initial functions to have a universal r.e. set are obtained.

In [1, 4] Ishmuhametov and Selivanov have obtained sufficient conditions for a special class of families of r.e. sets. In [6] the author has characterized the families

* Research partially supported by the Ministry of Education, Contract No И-604, 1997.

of finite sets which admit a universal recursively enumerable set. But until now such a necessary and sufficient condition is not found for arbitrary families of r.e. sets (p.r. functions).

In this paper we notice that the characterization of the families of finite functions which admit a universal partially recursive function is the same (in some sense) as in the case of families of finite sets of natural numbers [6] and we give necessary and sufficient conditions for a given family of recursive functions to have a universal recursive function. We want to stress the analogy between both cases.

Here we use $\mathbb{N}$ to denote the set of all natural numbers $\{0,1,2,\ldots\}$ and N_n to denote the initial segment $\{k \mid k \in \mathbb{N} \,\&\, k < n\}$ of the set $\mathbb{N}$. We suppose that there is some fixed effective coding $\langle .,. \rangle$ of the pairs of natural numbers and $\lambda x.(x)_0$, $\lambda x.(x)_1$ are such recursive functions that $(\langle x_0, x_1 \rangle)_0 = x_0$ and $(\langle x_0, x_1 \rangle)_1 = x_1$. If f is a partial function, we use $Dom(f)$ to denote the domain, and $Ran(f)$ to denote the range of values of the function f. In the case when $Dom(f) \subseteq \mathbb{N}^k$ and $Ran(f) \subseteq \mathbb{N}$, we shall write $f : \mathbb{N}^k \multimap \mathbb{N}$, and if $Dom(f) = \mathbb{N}^k$ and $Ran(f) \subseteq \mathbb{N}$, we shall write $f : \mathbb{N}^k \to \mathbb{N}$. For the sake of simplicity, we use mainly unary functions defined on a subset of $\mathbb{N}$. If f and g are functions, we say that f is a subfunction of g (and write $f \subseteq g$) iff $\forall x(x \in Dom(f) \Rightarrow x \in Dom(g) \,\&\, f(x) = g(x))$. We say in such a case that g is an extension of the function f, as well. Denote by θ_v the finite function with cannonical index v. For example, if θ is such a function that $Dom(\theta)$ is finite and $Dom(\theta) = \{x_1, \ldots, x_k\}$, $x_1 < \ldots < x_k$, then we can consider $v = p_0^k . p_1^{2^{x_1}.3^{\theta(x_1)}} \ldots p_k^{2^{x_k}.3^{\theta(x_k)}}$. Here $p_0, p_1, \ldots$ is the increasing sequence of all prime numbers. If $Dom(\theta) = N_k$ for some natural k, we say that θ is defined on an initial segment. By φ_e we denote the e-th partially recursive function in the standard enumeration of the partially recursive functions.

Let $\Psi : \mathbb{N}^2 \multimap \mathbb{N}$ and $\mathfrak{F}$ be a family of partial functions defined on $\mathbb{N}$. The function Ψ is said to be *universal for the family* $\mathfrak{F}$ iff for any n the function $\lambda x.\Psi(n,x)$ is in the family $\mathfrak{F}$, and, conversely, for any function $f \in \mathfrak{F}$ there exists such n that $f = \lambda x.\Psi(n,x)$. If $\Psi : \mathbb{N}^2 \to \mathbb{N}$, then Ψ_n denotes the unary function $\lambda x.\Psi(n,x)$.

It is well-known [cf. 7, p. 38] that if $\Psi : \mathbb{N}^2 \to \mathbb{N}$ is a recursive function which is universal for the family $\mathfrak{F}$, then there exists a recursive function $f : \mathbb{N}^2 \to \mathbb{N}$ such that for all n $\theta_{f(n,0)} \subseteq \theta_{f(n,1)}, \ldots$ and $\lim\limits_{s} \theta_{f(n,s)}(x) = \Psi(n,x)$, i.e. $\forall n \forall x \exists s_0 \forall s \geq s_0 (\theta_{f(n,s)}(x) = \Psi(n,x))$.

The following theorems and their proofs are analogous to the ones in [6].

Theorem 1. *Let $\mathfrak{F}$ be a nonempty family of finite fuctions defined on $\mathbb{N}$. Then the family $\mathfrak{F}$ has a universal partial recursive function iff the following three conditions hold*:

(i) *The set $V = \{v \mid \exists f (f \in \mathfrak{F} \,\&\, \theta_v \subseteq f)\}$ is recursively enumerable*;

(ii) *The set $I = \{v \mid \theta_v \in \mathfrak{F}\}$ is Σ_2^0 (in the arithmetical hierarchy)*;

(iii) *There exists such a partial recursive function h that*

$$V \subseteq Dom(h) \quad and \quad \forall v (v \in V \Rightarrow \theta_v \subseteq \varphi_{h(v)} \in \mathfrak{F}).$$

Theorem 2. *Let $\mathfrak{F}$ be a nonempty family of finite functions defined on $\mathbb{N}$ such that for every $f \in \mathfrak{F}$ at most finitely many functions g exist in $\mathfrak{F}$ such that*

$f \subseteq g$. *Then the family has a universal partial recursive function iff the following two conditions hold*:

(i) *The set* $V = \{v \mid \exists f(f \in \mathfrak{F} \,\&\, \theta_v \subseteq f\}$ *is recursively enumerable*;

(ii) *The set* $I = \{v \mid \theta_v \in \mathfrak{F}\}$ *is in the class* Σ^0_2.

Theorem 3. *Let* $\mathfrak{F}$ *be a nonempty family of finite functions defined on* $\mathbb{N}$ *such that the set* $I = \{v \mid \theta_v \in \mathfrak{F}\}$ *is in the class* Π^0_1. *Then the family has a universal partial recursive function iff the following two conditions hold*:

(i) *The set* $V = \{v \mid \exists f(f \in \mathfrak{F} \,\&\, \theta_v \subseteq f)\}$ *is recursively enumerable*;

(ii) *There exists a partial recursive function* h *of two variables such that the following three conditions hold*:

(a) $\forall v(v \in V \Rightarrow \lambda n.h(v,n)$ *is totally defined*);

(b) $\forall v \in V \forall n_1 \forall n_2 (n_1 < n_2 \Rightarrow \theta_v \subseteq \theta_{h(v,n_1)} \subseteq \theta_{h(v,n_2)})$;

(c) $\forall v \in V \exists n(\theta_{h(v,n)} \in \mathfrak{F})$.

Now we shall consider the case of families with recursive functions.

Theorem 4. *Let* $\mathfrak{F}$ *be a nonempty family of unary recursive functions. Then the family* $\mathfrak{F}$ *has a universal recursive function iff the following two conditions hold*:

(i) *The set* $V = \{v \mid \exists f(f \in \mathfrak{F} \,\&\, \theta_v \subseteq f)\}$ *is recursively enumerable*;

(ii) *There exist a family* $\mathfrak{G}$ *such that* $\mathfrak{F} \subseteq \mathfrak{G}$ *and a recursive function* Ψ, *which is universal for the family* $\mathfrak{G}$, *such that the following two conditions are satisfied*:

a) *the set* $I = \{n \mid \Psi_n \in \mathfrak{F}\}$ *is* Σ^0_2 (*in the arithmetical hierarchy*);

b) *there exists such a partial recursive function* h *that*

$$V \subseteq Dom(h) \quad and \quad \forall v(v \in V \Rightarrow \theta_v \subseteq \Psi_{h(v)} \in \mathfrak{F}).$$

Proof. Suppose first that the family $\mathfrak{F}$ has a universal recursive function Ψ. Then the set

$$\begin{aligned} V &= \{v \mid \exists f(f \in \mathfrak{F} \,\&\, \theta_v \subseteq f)\} \\ &= \{v \mid \exists n \exists k \exists x_1 \ldots \exists x_k (v = p_0^k . p_1^{2^{x_1}.3^{\theta_v(x_1)}} \ldots p_k^{2^{x_k}.3^{\theta_v(x_k)}} \\ &\qquad \& \, \Psi(n,x_1) \cong \theta_v(x_1) \,\&\, \ldots \,\&\, \Psi(n,x_k) \cong \theta_v(x_k))\} \end{aligned}$$

is recursively enumerable.

Fix $\mathfrak{G}$=F. It is obvious that the set $I = \{n \mid \Psi_n \in \mathfrak{F}\} = \mathbb{N}$, so the condition a) from (ii) is satisfied.

Let us define the function h as follows:

$$\begin{aligned} h(v) \cong \mu n[\exists k \exists x_1 \ldots \exists x_k \exists y_1 \ldots \exists y_k (v = p_0^k . p_1^{2^{x_1}.3^{y_1}} \ldots p_k^{2^{x_k}.3^{y_k}} \\ \& \, \Psi(n,x_1) \cong y_1 \,\&\, \ldots \,\&\, \Psi(n,x_k) \cong y_k)]. \end{aligned}$$

It is clear that h is a p.r. function and satisfies b) from (ii).

Conversely, let the conditions (i) – (ii) hold and F be a unary recursive function such that $F(0) = 1$ and $Ran(F) = V$.

In addition, let exist a family $\mathfrak{G}$ such that $\mathfrak{F} \subseteq \mathfrak{G}$ and a recursive function Ψ, which is universal for the family $\mathfrak{G}$, and $\alpha : \mathbb{N}^2 \to \mathbb{N}$ be a unary recursive function such that for all n

$$\theta_{\alpha(n,0)} \subseteq \theta_{\alpha(n,1)} \subseteq \cdots \subseteq \theta_{\alpha(n,s)} \subseteq \cdots,$$

and for all n and x $\lim_s \theta_{\alpha(n,s)}(x) = \Psi(n,x)$. For the sake of simplicity, we suppose that $Dom(\theta_{\alpha(n,s)})$ is an initial segment for all n and s. At the end, let G be a ternary total recursive function such that

$$n \in I \iff \exists x \forall y [G(n,x,y) = 0]$$

and h be a partial recursive function such that $V \subseteq Dom(h)$ and $\forall v (v \in V \Rightarrow \theta_v \subseteq \Psi_{h(v)} \in \mathfrak{F})$.

We construct the so-called strong recursive sequence of finite functions by steps. On step s for any (n,x) we construct a finite function $\theta_{g(n,x,s)}$ for a recursive function g and at the end we will fix

$$\Theta(\langle n,x \rangle, z) = \lim_s \theta_{g(n,x,s)}(z).$$

Let us describe the construction of the recursive sequence of finite functions.

Step $s = 0$. Take $g(n,x,s) = 1$, i.e. $\theta_{g(n,x,s)} = \varnothing$.

Step $s > 0$. We consider two cases:

Case I. $\forall y [y \leq s \iff G(n,x,y) = 0]$.

Take $g(n,x,s) = F(t_{n,x,s})$, where $t_{n,x,s} \leq s$ is such that $Dom(\theta_{F(t_{n,x,s})})$ is a maximal initial segment in the set $\{Dom(\theta_{F(0)}), \ldots, Dom(\theta_{F(s)})\}$ such that $\theta_{F(t_{n,x,s})}$ is a subfunction of the function Ψ_n.

Case II. $\exists y [y \leq s \,\&\, G(n,x,y) \neq 0]$.

Take $\theta_{g(n,x,s)} = \theta_{\alpha(h(g(n,x,s_0-1)),s)}$, where $s_0 \cong \mu s[G(n,x,s) \neq 0]$.

Thus the construction is completed.

Obviously, the construction is effective, so the function g is recursive.

First of all, we shall see that for all fixed n, x and z the limit $\lim_s \theta_{g(n,x,s)}(z)$ exists and belongs to the family $\mathfrak{F}$. We consider two cases:

Case I. $\forall y [G(n,x,y) = 0]$. Then $\Psi_n \in \mathfrak{F}$ and for all s $g(n,x,s) = F(t_{n,x,s})$, where $t_{n,x,s}$ is such that $Dom(\theta_{F(t_{n,x,s})})$ is a maximal initial segment in the set $\{Dom(\theta_{F(0)}), \ldots, Dom(\theta_{F(s)})\}$ such that $\theta_{F(t_{n,x,s})}$ is a subfunction of the function Ψ_n. Thus, $\theta_{g(n,x,0)} \subseteq \theta_{g(n,x,1)} \subseteq \ldots$ and the limit exists and it is $\Psi_n(z)$, because for all k there exists s such that $\{0, \ldots, k\} \subseteq Dom(\theta_{g(n,x,s)})$.

Case II. $\exists y [G(n,x,y) \neq 0]$. Let $s_0 \cong \mu s[G(n,x,s) \neq 0]$. Then $g(n,x,s) = \alpha(h(g(n,x,s_0-1)),s)$ and for all $s \geq s_0$ $\theta_{g(n,x,s)} = \theta_{\alpha(h(g(n,x,s_0-1)),s)}$. Therefore the limit $\lim_s \theta_{g(n,x,s)}(z)$ exists and it is $\Psi_{h(g(n,x,s_0-1))}(z)$.

Now let $f \in \mathfrak{F}$. Then $f = \Psi_n$ for some $n \in I$. Therefore a natural x exists such that $\forall y [G(n,x,y) = 0]$. It is clear now that $f(z) = \Psi_n(z) = \lim_s \theta_{g(n,x,s)}(z)$.

Define the function Θ as follows:

$$\Theta(k,z) = \begin{cases} \lim_s \theta_{g(n,x,s)}(z), & \text{if } k = \langle n,x \rangle, \\ f_0(z), & \text{otherwise,} \end{cases}$$

where f_0 is a fixed element of the family $\mathfrak{F}$.

It is clear that Θ is total, recursive and universal for the family $\mathfrak{F}$.

The following examples show that none of the conditions (i) – (ii) can be skipped.

Example 1. Let A be a nonrecursively enumerable set. Define the family of recursive finctions by the following equality:

$$f_n(x) = \begin{cases} 0, & \text{if } x = n, \\ 1, & \text{otherwise.} \end{cases}$$

Then let $\mathfrak{F} = \{f_n \mid n \in A\}$. It is easy to see that the family $\mathfrak{F}$ does not have a universal recursive function but $\mathfrak{F}$ satisfies the condition (ii).

Example 2 a). Let $\mathfrak{F}$ be the family of all total recursive unary functions. It is well-known that the family $\mathfrak{F}$ has not a universal recursive function. On the other hand, it is obvious that the conditions (i) and a) from (ii) are fulfilled.

Example 2 b). Let A be the set of cannonical codes of finite sets such that A is not in Σ_2^0. Define the family $\mathfrak{G} = \{f_v\}_{v \in \mathbb{N}} \cup \{g_v\}_{v \in \mathbb{N}}$ of recursive functions by the following equalities:

$$f_v(x) = \begin{cases} 0, & \text{if } x \in E_v, \\ 1, & \text{otherwise,} \end{cases}$$

$$g_v(x) = \begin{cases} 1, & \text{if } x \in E_v, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathfrak{F} = \{f_v \mid v \in A\}$. It is easy to see that the family $\mathfrak{F}$ has not a universal recursive function, but $\mathfrak{F}$ satisfies the conditions (i) and b) from (ii).

Example 3. Let A be the set of cannonical codes of finite sets such that the family $\mathfrak{F}$ does not satisfy the condition (ii) from Theorem 1 (see the Example in [6]). We define the family of recursive finctions by the following equality:

$$f_v(x) = \begin{cases} 0, & \text{if } x \in E_v, \\ 1, & \text{otherwise.} \end{cases}$$

Then let $\mathfrak{F} = \{f_v \mid v \in A\}$. It is not difficult to see that the family $\mathfrak{F}$ has not a universal recursive function, but $\mathfrak{F}$ satisfies the conditions (i) and a) from (ii).

For the next theorem we need to recall a definition. The total function h is said to be a *boundary function* for the family $\mathfrak{F}$ iff for every finite subfunction θ of h there exists a function $f \in \mathfrak{F}$ such that θ is a subfunction of f. If $\mathfrak{F}$ is a family of unary recursive functions and a function h exists such that h is a boundary, we say that $\mathfrak{F}$ has a boundary function, otherwise we say that $\mathfrak{F}$ does not have a boundary function.

Theorem 5. *Let $\mathfrak{F}$ be a nonempty family of unary recursive functions which has not a boundary function. Then the family $\mathfrak{F}$ has a universal recursive function iff the following two conditions hold*:

(i) *The set $V = \{v \mid \exists f (f \in \mathfrak{F} \,\&\, \theta_v \subseteq f)\}$ is recursively enumerable*;

(ii) *There exist a family $\mathfrak{G}$ such that $\mathfrak{F} \subseteq \mathfrak{G}$ and a recursive function Ψ, which is a universal for the family $\mathfrak{G}$, and the set $I = \{n \mid \Psi_n \in \mathfrak{F}\}$ is Σ_2^0 (in the arithmetical hierarchy).*

Proof. The first part follows from Theorem 4. Let F be a unary recursive function such that $F(0) = 1$ and $Ran(F) = V$. Let in addition a family $\mathfrak{G}$ exist such that $\mathfrak{F} \subseteq \mathfrak{G}$ and a recursive function Ψ exists, which is a universal for the family $\mathfrak{G}$, and $\alpha : \mathbb{N}^2 \to \mathbb{N}$ be a unary recursive function such that for all n $\theta_{\alpha(n,0)} \subseteq \theta_{\alpha(n,1)} \subseteq \cdots \subseteq \theta_{\alpha(n,s)} \subseteq \cdots$ (again for the sake of simplicity we consider that $Dom(\theta_{\alpha(n,s)})$ is an initial segment for all n and s), and for all n and x $\lim_s \theta_{\alpha(n,s)}(x) = \Psi(n,x)$ and G be a ternary total recursive function such that

$$n \in I \iff \exists x \forall y [G(n,x,y) = 0].$$

We construct again a strong recursive sequence of finite functions by steps, analogously to Theorem 4. Let us describe the construction.

Step $s = 0$. Take $g(n,x,s) = 1$, i.e. $\theta_{g(n,x,s)} = \varnothing$.

Step $s > 0$. We consider two cases:

Case I. $\forall y [y \leq s \iff G(n,x,y) = 0]$.

Take $g(n,x,s) = F(t_{n,x,s})$, where $t_{n,x,s} \leq s$ is such that $Dom(\theta_{F(t_{n,x,s})})$ is the maximal initial segment in the set $\{Dom(\theta_{F(0)}), \ldots, Dom(\theta_{F(s)})\}$ such that $\theta_{F(t_{n,x,s})}$ is a subfunction of the function Ψ_n.

Case II. $\exists y [y \leq s \,\&\, G(n,x,y) \neq 0]$. Take $g(n,x,s) = F(t_{n,x,s})$, where $t_{n,x,s}$ is such that $Dom(\theta_{F(t_{n,x,s})})$ is an initial segment, $\theta_{g(n,x,s-1)} \subseteq \theta_{F(t_{n,x,s})}$ and $\theta_{F(t_{n,x,s})}$ is the maximal element of the set $\{\theta_{F(0)}, \ldots, \theta_{F(s)}\}$.

Thus the construction is completed. It is effective, so the function g is recursive.

Analogously, we shall see that for all fixed n, x and z the limit $\lim_s \theta_{g(n,x,s)}(z)$ exists and belongs to the family $\mathfrak{F}$. For fixed n, x and z we have to consider two cases, but the first is the same as in Theorem 4, so we shall consider only the second case.

Case II. $\exists y [G(n,x,y) \neq 0]$. From the construction it is clear that for all s $\theta_{g(n,x,s-1)} \subseteq \theta_{g(n,x,s)}$. Therefore, the limit $\lim_s \theta_{g(n,x,s)}(z)$ exists. Let us suppose that for some $z \in \mathbb{N}$ the limit $\lim_s \theta_{g(n,x,s)}(z)$ is undefined. Then there exists s_0 such that for all $s \geq s_0$ $\theta_{g(n,x,s)} = \theta_{g(n,x,s_0)}$. On the other hand, $\theta_{F(t_{n,x,s_0})} \subseteq f \in \mathfrak{F}$ for some f. Since for all natural k a finite functions $\theta \subseteq f$ exist such that $Dom(\theta) = N_{k+1}$ and $\theta \subseteq f$, then for all such θ there exists s such that $\theta_{F(s)} = \theta$. It is clear now that $\theta_{g(n,x,s)}(z)$ is defined for some s, which contradicts the supposition that the limit $\lim_s \theta_{g(n,x,s)}(z)$ is undefined. Therefore for all $z \in \mathbb{N}$ the limit $\lim_s \theta_{g(n,x,s)}(z)$ is defined.

Assume now that $\lim_s \theta_{g(n,x,s)}$ does not belong to the family $\mathfrak{F}$. Then according to the construction, natural numbers s_1, s_2, ... exist such that $\theta_{g(n,x,s_1)} \subset \theta_{g(n,x,s_2)} \subset \ldots$ and $\theta_{g(n,x,s_i)} \subseteq \Psi_{n_i}$ for all i. This means that $\lim_s \theta_{g(n,x,s)}$ is a boundary function for the family $\mathfrak{F}$.

The proof that if $f \in \mathfrak{F}$, then $f = \Psi_n$ for some $n \in I$, is the same as in Theorem 4.

At the end, let the function Θ be defined as in Theorem 4:

$$\Theta(k,z) = \begin{cases} \lim_s \theta_{g(n,x,s)}(z), & \text{if } k = \langle n,x \rangle, \\ f_0(z), & \text{otherwise}, \end{cases}$$

where f_0 is a fixed element of the family $\mathfrak{F}$. The theorem is proved.

Let us note that if a family of (total) recursive functions is finitely generated by some effective operations, then the family has a universal (total) recursive function.

At the end, we shall note the following

Proposition 6. *If the family $\mathfrak{F}$ of unary total recursive functions has a universal recursive function, then there exists a family $\mathfrak{G}$ such that $\mathfrak{F} \subseteq \mathfrak{G}$ and the family $\mathfrak{G}$ is finitely generated by some effective operations.*

Proof. Indeed, let Θ be universal for the family $\mathfrak{F}$. Fix the functions **I**, **S**, **O** and f defined as follows:

$\mathbf{I}(x) = x$, $\mathbf{S}(x) = x + 1$, $\mathbf{O}(x) = 0$ and

$$f(z) = \begin{cases} \Theta(n, x), & \text{if } z = \langle n, x \rangle, \\ 0, & \text{otherwise.} \end{cases}$$

Let us define the binary operation $(.,.)$ between the functions as follows:

$$(f_1, f_2)(x) = \langle f_1(x), f_2(x) \rangle.$$

It is easy to show that the family $\mathfrak{G}$, which is generated from the functions **I**, **S**, **O**, f, $\lambda x.(x)_0$, $\lambda x.(x)_1$ by the operations composition and $(.,.)$, contains the family $\mathfrak{F}$.

Open problem. Given a family $\mathfrak{F}$ of unary total recursive functions, which has a universal recursive function, is it true that the family $\mathfrak{F}$ is generated from finitely many functions belonging to $\mathfrak{F}$ by a finite number of effective operations?

REFERENCES

1. Ishmuhametov, S. T. On the families of recursively enumerable sets. In: *Probabilistic methods and cybernetics*, Publishing house of the University of Kazan, Kazan, **18**, 1982, 46–53 (in Russian).
2. Kumer, M. Negative results about the length problem. In: *Lecture Notes in Computer Science*, **270**, 1987, 237–248.
3. Menzel, W. & V. Sperschneider. Recursively enumerable extensions of $\mathbf{R}_1$ by finite functions. In: *Lecture Notes in Computer Science*, **171**, 1984, 62–76.
4. Selivanov, V. L. On the index sets of the computable classes of finite sets. In: *Algorithms and automata*, Publishing house of the University of Kazan, Kazan, 1978, 95–99 (in Russian).
5. Sperschneider, V. The length-problem. In: *Lecture Notes in Computer Science*, **171**, 1984, 88–102.
6. Ditchev, A. Effective enumerations of families of finite sets of natural numbers. *Zeit. f. math. Logik und Grundl. d. Math.*, **37**, 1991, 301–306.
7. Ershov, Yu. L. Enumeration theory. Nauka, Moscow, 1977 (in Russian).

Received December 20, 1999

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 James Bourchier Blvd.
BG-1164 Sofia, Bulgaria
E-mail: `ditchev@fmi.uni-sofia.bg`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

REGULAR ENUMERATIONS FOR ABSTRACT STRUCTURES*

VESSELA BALEVA

Using the method of regular enumerations in the context of definability, we obtain a normal form for the sets which are Σ_{n+1}-admissible in some partial structure.

Keywords: enumerations, enumeration reducibility, enumeration jump, forcing, abstract structures, admissible sets, definability

2000 MSC: 03D30

1. INTRODUCTION

In the present paper we are using the method of regular enumerations [9] in the context of definability on abstract structures.

For the sake of simplicity, we consider only unary sets. All the definitions and results can be easily generalized for sets of arbitrary finite arity.

Given two sets of natural numbers A and B, we say that A is enumeration-reducible to B ($A \leq_e B$) if $A = \Gamma_z(B)$ for some enumeration operator Γ_z [7, 1, 3, 5, 8]. In other words, if D_v denotes the finite set with a canonical code v and $W_0, \dots, W_z, \dots$ is the Gödel enumeration of the recursively enumerable (r. e.) sets, we have

$$A \leq_e B \iff \exists z \forall x (x \in A \Leftrightarrow \exists v (\langle v, x \rangle \in W_z \ \& \ D_v \subseteq B)).$$

Given a set A, denote by A^+ the set $A \oplus (\omega \setminus A)$. The set A is called *total* iff $A \equiv_e A^+$. Note that the graph G_f of each total function f is a total set.

* This work was partially supported by the Ministry of Education and Science, Contract И-604/96.

Given a set A, let $K_A^0 = \{\langle x, z\rangle \mid x \in \Gamma_z(A)\}$. We define the *e-jump* A'_e of A to be the set $(K_A^0)^+$.

Several properties of the e-jump are proved in [6, 9, 5]. Since we are going to consider only the e-jump here, we omit the subscript e in the notation of the e-jump. For each set B, $B^{(0)} = B$ and $B^{(n+1)}$ is the e-jump of $B^{(n)}$.

Let N be an infinite countable set and ω be the set of the natural numbers. We assume that we have an equality ($=_N$) and an inequality ($\neq_N$) in N. Consider $n+1$ sets $B_0, \ldots, B_n$ such that $B_i \subseteq N$ for each $i \in [0, n]$. The algebraic structure $\alpha = (N, \omega, =_N, \neq_N, G_f, D)$, where:

- $f : \omega \to N$ is a bijection,
- $D \subseteq \omega$ is a total set,

is called an *enumeration.* From now on we write $\alpha = (f, D)$ to denote the enumeration α and if $D = G_g$ for some total g, then we write $\alpha = (f, g)$.

The set $A \subseteq N$ is called *admissible* relatively $B_0, \ldots, B_n$ iff for each enumeration $\alpha = (f, D)$ it is true that

$$f^{-1}(B_0) \leq_e D \;\&\; \ldots \;\&\; f^{-1}(B_n) \leq_e D^{(n)} \Rightarrow f^{-1}(A) \leq_e D^{(n)}.$$

The aim of the present paper is to obtain a normal form of the admissible sets.

Consider a countable first-order language consisting of the binary predicate symbols $=$, $\neq$ (interpreted as $=_N$ and $\neq_N$) and unary predicate symbols T_i for each $i \in [0, n]$ (interpreted as B_i and taking only the value true (0), whenever defined).

An *elementary existential* formula is a formula in a prenex normal form with a finite number of quantifiers which are only existential, and a matrix which is a finite conjunction of atomic predicates of $=$, $\neq$ and T_0. These formulae are interpreted in the usual way and the quantifiers are over the set N. The elementary existential formulae can be effectively coded by natural numbers. If n is the code of a certain formula, the formula itself is denoted by $\lceil n \rceil$. We use the notation $\varphi(Z_1, \ldots, Z_a)$ for the formula φ with free variables among $Z_1, \ldots, Z_a$.

Bellow we define Σ_i^+-formulae and Π_i^+-formulae for each $i \leq n$.

A Σ_0^+-formula is a formula of the form $\bigvee_{\gamma(n)} \lceil \gamma(n) \rceil (Z_1, \ldots, Z_a)$, where γ is a recursive function and $\lceil \gamma(n) \rceil (Z_1, \ldots, Z_a)$ is an elementary existential formula. These formulae are interpreted in the usual way. The Π_0^+-formula $\Psi(Z_1, \ldots, Z_a)$ is a formula of the form $\neg\Phi(Z_1, \ldots, Z_a)$, where $\Phi(Z_1, \ldots, Z_a)$ is a Σ_0^+-formula. If $\overline{s} \in N^a$, then:

$$\Psi(\overline{Z}|\overline{s}) \simeq 0 \;\Leftrightarrow\; \Phi(\overline{Z}|\overline{s}) \not\simeq 0.$$

Proceeding by induction, suppose that $i < n$ and for each $j \in [0, i]$ we have defined Σ_j^+- and Π_j^+-formulae, which can be effectively coded by natural numbers. An elementary Σ_{i+1}^+-formula is a formula in a prenex normal form with a finite number of existential quantifiers and a matrix which is a finite conjunction of atomic predicates of T_{i+1}, $=$, $\neq$ and Σ_i^+- and Π_i^+-formulae. These formulae are interpreted in the usual way and they can be effectively coded by natural numbers.

A Σ_{i+1}^+-formula is a formula of the form $\bigvee_{\gamma(n)}\lceil\gamma(n)\rceil(Z_1,\ldots,Z_a)$, where γ is a recursive function and $\lceil\gamma(n)\rceil(Z_1,\ldots,Z_a)$ is an elementary Σ_{i+1}^+-formula. A Π_{i+1}^+-formula $\Psi(Z_1,\ldots,Z_a)$ is a formula of the form $\neg\Phi(Z_1,\ldots,Z_a)$, where $\Phi(Z_1,\ldots,Z_a)$ is a Σ_{i+1}^+-formula. These formulae are interpreted similarly to the Σ_0^+- and Π_0^+-formulae.

The set $A \subseteq N$ is called *definable* iff there exist a Σ_n^+-formula $\Phi(W_1,\ldots,W_r,Z)$ and $t_1,\ldots,t_r \in N$, such that for all $s \in N$,

$$s \in A \;\Leftrightarrow\; \Phi(\overline{W}|\bar{t}, Z|s) \simeq 0.$$

We are going to prove the following result, which gives a normal form for the admissible sets.

Theorem 1. *Let $A \subseteq N$. Then A is admissible iff A is definable.*

The "only if" part of the theorem is obvious, so we must prove only that if A is admissible, then A is definable.

2. REGULAR ENUMERATIONS

The method of regular enumerations is introduced and studied in [9]. In this paper we adapt it for abstract structures.

Let us fix $n \geq 0$ and subsets $B_0,\ldots,B_n$ of N. Since for every bijective mapping f of ω into N $f^{-1}(B_i) \equiv_e f^{-1}(B_i)\oplus\omega$, we may suppose that $f^{-1}(B_i)$ and hence B_i are not empty. We use the term *finite part* to denote an ordered pair $\tau = (f_\tau, g_\tau)$ such that:

- f_τ is a finite injective mapping of ω into N;
- g_τ is a finite mapping of ω into ω defined on a finite segment $[0, q-1]$ of ω.

The finite parts will be denoted by the letters τ, δ, ρ and Δ. If $\mathrm{dom}(g_\tau) = [0, q-1]$, then let $\mathrm{lh}(g_\tau) = q$. We assume that an effective coding of all sequences and all finite mappings of ω into ω, defined on a finite segment, is fixed. Let $\tau = (f_\tau, g_\tau)$ and $\rho = (f_\rho, g_\rho)$. If $f_\tau \subseteq f_\rho$ and $g_\tau \subseteq g_\rho$, we write $\tau \subseteq \rho$.

Bellow we define *i-regular* finite parts for each $i \leq n$.

A 0-regular finite part is a finite part $\tau = (f_\tau, g_\tau)$ such that $\mathrm{dom}(g_\tau) = [0, 2q+1]$ and for all odd $z \in \mathrm{dom}(g_\tau)$, $g_\tau(z) \in f_\tau^{-1}(B_0)$.

If $\mathrm{dom}(g_\tau) = [0, 2q+1]$, then the 0-rank, $|\tau|_0$, of τ is equal to $q+1$, the number of all odd elements of $\mathrm{dom}(g_\tau)$. For each 0-regular finite part τ, let $B_0^{g_\tau}$ be the set of the odd elements of $\mathrm{dom}(g_\tau)$.

Given a 0-regular finite part $\tau = (f_\tau, g_\tau)$, let

$$g_\tau \Vdash_0 F_e(x) \;\Leftrightarrow\; \exists v(\langle v, x\rangle \in W_e \;\&\; \forall u \in D_v(g_\tau((u)_0) \simeq (u)_1)),$$

$$g_\tau \Vdash_0 \neg F_e(x) \;\Leftrightarrow\; \forall(\text{0-regular } \rho = (f_\rho, g_\rho))(\tau \subseteq \rho \Rightarrow g_\rho \nVdash_0 F_e(x)).$$

Proceeding by induction, suppose that we have defined the i-regular finite parts for some $i < n$ and for each i-regular finite part $\tau = (f_\tau, g_\tau)$ we have defined its i-rank $|\tau|_i$, the set $B_i^{g_\tau}$ and the relations $g_\tau \Vdash_i F_e(x)$ and $g_\tau \nVdash_i F_e(x)$.

Let f_τ be a finite mapping of ω into N and g'_τ be a finite mapping of ω into ω such that $\mathrm{dom}(g'_\tau) = [0, q'-1]$ and $\tau' = (f_\tau, g'_\tau)$ is i-regular. Let

$$G = \{g_\rho \mid \rho = (f_\rho, g_\rho) \text{ is } i\text{-regular } \& \ \tau' \subseteq \rho \ \& \ g'_\tau \subset g_\rho \ \& \ g_\rho \Vdash_i F_e(x)\}.$$

We say that g''_τ is *appropriate* for f_τ, g'_τ, e and x (we denote this by $\mathrm{app}(g''_\tau, f_\tau, g'_\tau, e, x)$) iff one of the following is true:

- $G \neq \varnothing$, $g''_\tau \Vdash_i F_e(x)$, (f_τ, g''_τ) is i-regular, $g'_\tau \subset g''_\tau$ and $\mathrm{lh}(g''_\tau) = \min\{\mathrm{lh}(g) \mid g \in G\}$;
- $G = \varnothing$, (f_τ, g''_τ) is i-regular, $|(f_\tau, g''_\tau)|_i = |(f_\tau, g'_\tau)|_i + 1$ and $g'_\tau \subset g''_\tau$.

Let τ be a finite part, g_τ be defined on $[0, q-1]$, and $r \geq 0$. Then τ is $(i+1)$-regular with $(i+1)$-rank $r+1$ iff there exist natural numbers

$$0 < n_0 < b_0 < n_1 < b_1 \ldots < n_r < b_r < n_{r+1} = q$$

such that $(f_\tau, g_\tau \upharpoonright n_0)$ is an i-regular finite part with i-rank 1 and for all j, $0 \leq j \leq r$, it is true that:

1. $\mathrm{app}(g_\tau \upharpoonright b_j, f_\tau, g_\tau \upharpoonright (n_j+1), (j)_0, (j)_1)$;
2. $g_\tau(b_j) \in f_\tau^{-1}(B_{i+1})$;
3. $(f_\tau, g_\tau \upharpoonright n_{j+1})$ is an i-regular extension of $(f_\tau, g_\tau \upharpoonright (b_j+1))$ with i-rank $|(f_\tau, g_\tau \upharpoonright (b_j+1))|_i + 1$.

Let $B^{g_\tau}_{i+1} = \{b_0, \ldots, b_r\}$. The next lemma shows that the $(i+1)$-rank is well defined. Its proof follows easily from the definition of $(i+1)$-regular finite parts.

Lemma 1. *Let τ be an $(i+1)$-regular finite part. Then:*

(i) *Let $m_0, a_0, \ldots, m_p, a_p, m_{p+1}$ and $n_0, b_0, \ldots, n_r, b_r, n_{r+1}$ be two sequences of natural numbers satisfying 1–3. Then $r = p$, $n_{p+1} = m_{p+1}$ and for all $j \leq p$, $n_j = m_j$ and $b_j = a_j$;*
(ii) *If ρ is $(i+1)$-regular, $\tau \subseteq \rho$ and $|\tau|_{i+1} = |\rho|_{i+1}$, then $g_\rho = g_\tau$;*
(iii) *τ is i-regular and $|\tau|_i > |\tau|_{i+1}$.*

To complete the definition of the regular finite parts, let for each $(i+1)$-regular finite part τ

$$g_\tau \Vdash_{i+1} F_e(x) \ \Leftrightarrow \ \exists v(\langle v, x\rangle \in W_e \ \& \ \forall u \in D_v(u = \langle e_u, x_u, \varepsilon\rangle \ \& \ \varepsilon \in \{0,1\} \ \& \ g_\tau \Vdash_i (\neg)^\varepsilon F_{e_u}(x_u))),$$

$$g_\tau \Vdash_{i+1} \neg F_e(x) \ \Leftrightarrow \ \forall((i+1)\text{-regular } \rho = (f_\rho, g_\rho)) \ (\tau \subseteq \rho \Rightarrow g_\rho \nVdash_{i+1} F_e(x)).$$

Lemma 2. (i) *There exists an $(i+1)$-regular finite part with $(i+1)$-rank 1;*
(ii) *If τ is an $(i+1)$-regular finite part, then there exists an $(i+1)$-regular finite part ρ such that $\tau \subseteq \rho$ and $|\rho|_{i+1} = |\tau|_{i+1} + 1$.*

The proof of this lemma also follows immediately from the definitions.

The enumeration $\alpha = (f, g)$ is called *regular* iff the following two conditions hold:

- For each finite part $\delta \subseteq \alpha$ there exists an n-regular extension τ of δ such that $\tau \subseteq \alpha$;
- If $i \leq n$ and $z \in B_i$, then there exists an i-regular $\tau \subseteq \alpha$ such that $f_\tau^{-1}(z) \in B_i^{g_\tau}$.

Given a regular enumeration $\alpha = (f, g)$ and $i \leq k$, let

$$B_i^g = \{b \mid \exists(\tau = (f_\tau, g_\tau) \subseteq \alpha)(\tau \text{ is } i\text{-regular } \& \ b \in B_i^{g_\tau})\}.$$

Clearly, $f^{-1}(B_i) = g(B_i^g)$. Similarly to the analogous proposition 2.8, in [9], one can prove the following lemma:

Lemma 3. *Suppose that $\alpha = (f, g)$ is a regular enumeration. If $i \leq n$, then $f^{-1}(B_i) \leq_e g^{(i)}$.*

Let g be a total mapping of ω into ω. For each $i < n$, e and x we define the relation $g \vDash_i F_e(x)$ by induction on i:

$$g \vDash_0 F_e(x) \ \Leftrightarrow\ \exists v(\langle v, x\rangle \in W_e \ \&\ \forall u \in D_v(g((u)_0) \simeq (u)_1)),$$

$$g \vDash_{i+1} F_e(x) \ \Leftrightarrow\ \exists v(\langle v, x\rangle \in W_e \ \&\ \forall u \in D_v((u = \langle e_u, x_u, 0\rangle$$
$$\&\ g \vDash_i F_e(x)) \vee (u = \langle e_u, x_u, 1\rangle \ \&\ g \nvDash_i F_e(x)))).$$

Let for each $i \in [0, n]$

$$g \vDash_i \neg F_e(x) \ \Leftrightarrow\ g \nvDash_i F_e(x).$$

The following lemma can be proved by induction on i.

Lemma 4. *Let g be a total mapping on ω into ω, $A \subseteq \omega$ and $i \leq k$. Then $A \leq_e g^{(i)}$ iff there exists e such that for all x, $x \in A \Leftrightarrow g \vDash_i F_e(x)$.*

Lemma 5 (Truth lemma). *Let $\alpha = (f, g)$ be a regular enumeration. Then for all $i \leq n$,*

$$g \vDash_i F_e(x) \Leftrightarrow \exists \tau \subseteq \alpha(\tau \text{ is } i\text{-regular } \&\ g_\tau \Vdash_i F_e(x)).$$

Proof. We use an induction on i. The lemma is obviously true for $i = 0$. Suppose that $i < n$ and it is true for i. First, we are going to show that

$$g \vDash_i \neg F_e(x) \ \Leftrightarrow\ \exists \tau \subseteq \alpha(\tau \text{ is } i\text{-regular } \&\ g_\tau \Vdash_i \neg F_e(x)).$$

Suppose that $g \vDash_i \neg F_e(x)$ and for each i-regular $\tau \subseteq \alpha$, $g_\tau \nVdash \neg F_e(x)$. Then for each i-regular finite part τ of α there exists an i-regular ρ such that $\tau \subseteq \rho$ and $g_\rho \Vdash F_e(x)$. Let δ be an $(i+1)$-regular finite part of α such that $|\delta|_{i+1} > \langle e, x\rangle$. By the definition of the $(i+1)$-regular finite parts, there exists an i-regular $\rho' \subseteq \delta$ such that $g_{\rho'} \Vdash_i F_e(x)$. By induction $g \vDash_i F_e(x)$. A contradiction.

Suppose now that $\tau \subseteq \alpha$ is i-regular, $g_\tau \Vdash_i \neg F_e(x)$ and $g \vDash_i F_e(x)$. By induction, there exists an i-regular $\rho \subseteq \alpha$ such that $g_\rho \Vdash_i F_e(x)$. Using the monotonicity of $\Vdash_i$, we can assume that $\tau \subseteq \rho$ and get a contradiction. Now the lemma easily follows from the definitions and monotonicity.

Now we are ready to prove that if a set is admissible relatively $B_0, \dots, B_n$, then it is definable. First of all, we need to prove that each admissible set has a normal form based on forcing relation and regular finite parts [10]. After that we can "translate" this normal form into a Σ_n^+-formula.

We say that $A \subseteq N$ has a *forcing normal form* iff there exist a natural number e and an n-regular finite part δ such that for each $s \in N$ the following equivalence is true:

$$s \in A \iff \exists x \exists \tau \supseteq \delta(\tau \text{ is an } n\text{-regular finite part}$$
$$\& \ f_\tau(x) \simeq s \ \& \ g_\tau \Vdash_n F_e(x)).$$

Theorem 2 (Forcing normal form). *Let $A \subseteq N$. If A is admissible, then A has a forcing normal form.*

Proof. Suppose that A has not a forcing normal form. We are going to construct by steps a regular enumeration $\alpha = (f, g)$ such that for each $i \in [0, n]$ $f^{-1}(B_i) \leq_e g^{(i)}$, but $\neg(A \leq_e g^{(n)})$. At each step q we shall define an n-regular finite part δ_q such that $\delta_q \subseteq \delta_{q+1}$.

Let $s_0, s_1, \dots$ be an arbitrary enumeration of N and δ_0 be an arbitrary n-regular finite part with n-rank 1. Let $q > 0$ and let δ_r be defined for all $r < q$.

I. $(q)_0 = 3n$. Let s be the first element of the sequence $s_0, s_1, \dots$, which does not belong to the range$(f_{\delta_{q-1}})$, and z be the smallest natural number, which does not belong to dom$(f_{\delta_{q-1}})$. We define $f_{\delta_q}(z) \simeq s$ and $f_{\delta_q}(x) \simeq f_{\delta_{q-1}}(x)$ for $x \neq z$ and $g_{\delta_q} = g_{\delta_{q-1}}$.

II. $(q)_0 = 3n + 1$. Let δ_q be an arbitrary n-regular finite part such that $\delta_q \supseteq \delta_{q-1}$ and $|\delta_q|_n = |\delta_{q-1}|_n + 1$.

III. $(q)_0 = 3n + 2$ and $(q)_1 = e$. Since A has not a forcing normal form, for δ_{q-1} and e there exists $s \in N$ such that the following equivalence is not true:

$$s \in A \iff \exists x \exists \tau \supseteq \delta_{q-1}(\tau \text{ is an } n\text{-regular finite part}$$
$$\& \ f_\tau(x) \simeq s \ \& \ g_\tau \Vdash_n F_e(x)).$$

1. Let $s \in A$ and $\forall x \forall \tau \supseteq \delta_{q-1}(\tau$ is an n-regular finite part & $f_\tau(x) \simeq s \Rightarrow g_\tau \nVdash_n F_e(x))$. Let $\alpha = (f, g)$ be a regular enumeration such that $\alpha \supseteq \delta_{q-1}$. We shall prove that $f^{-1}(A) \neq \{x \mid g \models_n F_e(x)\}$. Let $x = f^{-1}(s)$. Suppose that $x \in \{y \mid g \models_n F_e(y)\}$. Using the Truth lemma and the monotonicity of the forcing, we obtain a finite part τ such that $f_\tau(x) = s$, $\delta_{q-1} \subseteq \tau$, and $g_\tau \Vdash F_e(x)$. A contradiction. In this case we define $\delta_q = \delta_{q-1}$.
2. Let $s \notin A$ and $\exists x \exists \tau \supseteq \delta_{q-1}(\tau$ is an n-regular finite part & $f_\tau(x) \simeq s$ & $g_\tau \Vdash_n F_e(x))$. Let us fix τ with the above properties and let $\alpha = (f, g)$ be a regular enumeration such that $\alpha \supseteq \tau$. Using the monotonicity of the forcing, we have that $g \models_n F_e(x)$ and $f(x) = s$, but $s \notin A$. Hence $f^{-1}(A) \neq \{x \mid g \models_n F_e(x)\}$. So in this case we define $\delta_q = \tau$.

Let $\alpha = (f, g)$ be a regular enumeration defined as follows: $f = \bigcup_{q \in \omega} f_{\delta_q}$ and $g = \bigcup_{q \in \omega} g_{\delta_q}$. Using Lemma 3 and Lemma 4, we obtain that A is not admissible, which proves the theorem.

Let us fix a variable Z. Denote by Var the set of all remaining variables. Let us fix a recursive bijective mapping var of the natural numbers onto Var.

We use the sign "$*$" to denote the concatenation operation on sequences of natural numbers, and "$\subseteq$" to denote the relation "is a subsequence of".

Bellow we define *i-patterns* of the i-regular finite parts for each $i \leq n$ and when an i-regular finite part τ is *coordinated* with the i-pattern σ.

Let $i = 0$. Then σ is a 0-pattern iff it is the code of a sequence of natural numbers of the form $\langle r_0, \ldots, r_{2q+1} \rangle$. The 0-rank of σ, $|\sigma|_0$, is $q + 1$. The 0-regular finite part τ is coordinated with the 0-pattern σ iff $\{r_1, r_3, \ldots, r_{2q+1}\} \subseteq \mathrm{dom}(f_\tau)$, $g_\tau(j) \simeq r_j$ for $j \in [0, 2q+1]$ and $\mathrm{lh}(g_\tau) = 2q + 2$. We denote $\overline{\sigma} = (\mathrm{var}(r_1), \mathrm{var}(r_3), \ldots, \mathrm{var}(r_{2q+1}))$ and $f_\tau(\overline{\sigma}) = (f_\tau(r_1), f_\tau(r_3), \ldots, f_\tau(r_{2q+1}))$.

Let $i > 0$. Then σ is an i-pattern iff it is the code of a sequence of natural numbers of the form

$$\langle \eta_0, n_0, \langle \xi_0, \varepsilon_0 \rangle, b_0, \eta_1, \ldots, \eta_r, n_r, \langle \xi_r, \varepsilon_r \rangle, b_r, \eta_{r+1} \rangle,$$

where η_0 is an $(i-1)$-pattern with $(i-1)$-rank 1 and for each $j \in [0, r]$ the following conditions are satisfied:

1. $\varepsilon_j \in \{0, 1\}$ and ξ_j is an $(i-1)$-pattern such that $\xi_j \supseteq \eta_j * \langle n_0 \rangle$ and if $\varepsilon = 1$, then $|\eta_j|_{i-1} = |\eta_j|_{i-1} + 1$;
2. η_{j+1} is an $(i-1)$-pattern such that $\eta_{j+1} \supseteq \xi_j * \langle b_0 \rangle$ and $|\eta_{j+1}|_{i-1} = |\xi_j|_{i-1} + 1$.

The i-rank of σ, $|\sigma|_i$, is $r + 1$. The i-regular finite part τ is coordinated with the i-pattern σ if the following conditions are satisfied:

- $\{b_0, \ldots, b_r\} \subseteq \mathrm{dom}(f_\tau)$;
- If $m_0, a_0, \ldots, m_r, a_r, m_{r+1}$ is a sequence of natural numbers satisfying 1–3 of the definition of i-regular finite part for τ, then $(f_\tau, g_\tau \restriction m_0)$ is coordinated with $(i-1)$-pattern η_0 and for each $j \in [0, r]$ we have:
  1. $g_\tau(m_j) = n_j$;
  2. $(f_\tau, g_\tau \restriction a_j)$ is coordinated with ξ_j;
  3. if $\varepsilon_j = 0$, then $g_\tau \Vdash_{i-1} F_{(j)_0}((j)_1)$, else $g_\tau \nVdash_{i-1} F_{(j)_0}((j)_1)$;
  4. $g_\tau(a_j) = b_j$;
  5. $(f_\tau, g_\tau \restriction m_{j+1})$ is coordinated with the $(i-1)$-pattern η_{j+1}.

Let $\overline{\sigma} = (\overline{\eta}_{r+1}, \mathrm{var}(b_0), \ldots, \mathrm{var}(b_r))$ and $f_\tau(\overline{\sigma}) = (f_\tau(\overline{\eta}_{r+1}), f_\tau(b_0), \ldots, f_\tau(b_r))$. Let for $i \in [0, n]$

$$\mathfrak{R}_i(\delta, x) = \{s \mid s \in N \ \& \ \exists \tau \supseteq \delta (f_\tau(x) \simeq s \ \& \ \tau \text{ is } i\text{-regular})\}.$$

Lemma 6. *There exists an uniform effective way, given g_δ, $y_1, \ldots, y_r$ such that $\delta = (f_\delta, g_\delta)$ is i-regular and $\mathrm{dom}(f_\delta) = \{y_1, \ldots, y_r\}$, and given natural numbers e and x, to define a Σ_i^+-formula $\Phi^{\delta,e,x}$ with free variables among* $\mathrm{var}(y_1) =$

$Y_1, \ldots, \mathrm{var}(y_r) = Y_r, Z$ *such that for all* $s \in \Re_i(\delta, x)$,

$$\Phi^{\delta,e,x}(\overline{Y}|\overline{f_\delta(y)}, Z|s) \simeq 0 \;\Leftrightarrow\; \exists\tau \supseteq \delta(\tau \textit{ is } i\textit{-regular} \;\&\; g_\tau \Vdash_i F_e(x)).$$

Proof. We prove the lemma by induction on i. For $i = 0$ it immediately follows from the definitions.

Let $i > 0$ and let assume that for each $j \in [0, i-1]$ the lemma is true. Using the inductive assumption, one can easily prove the next lemmas.

Lemma 7. *There exists an uniform effective way, given $(i-1)$-pattern σ, natural numbers e and x and a finite set $D = \{y_1, \ldots, y_r\}$, to define a Σ^+_{i-1}-formula $\Phi^{\sigma,e,x,D}$ with free variables among* $\mathrm{var}(y_1) = Y_1, \ldots, \mathrm{var}(y_r) = Y_r$, $\overline{\sigma}$ *such that for each $(i-1)$-regular finite part τ coordinated with σ and such that* $D \subseteq \mathrm{dom}(f_\tau)$ *it is true that*

$$\Phi^{\delta,e,x,D}(\overline{Y}|\overline{f_\delta(y)}, \overline{\sigma}|f_\tau(\overline{\sigma})) \simeq 0$$

$$\Leftrightarrow \exists\Delta \supseteq \tau(\Delta \textit{ is } (i-1)\textit{-regular} \;\&\; g_\Delta \Vdash_{i-1} F_e(x)).$$

Lemma 8. *There exists an uniform effective way, given an $(i-1)$-pattern σ, natural numbers e and x and a finite set $D = \{y_1, \ldots, y_r\}$, to define a Σ^+_{i-1}-formula $\Phi^{\sigma,e,x,D}$ with free variables among* $\mathrm{var}(y_1) = Y_1, \ldots, \mathrm{var}(y_r) = Y_r$, $\overline{\sigma}$ *such that for each $(i-1)$-regular finite part τ coordinated with σ and such that* $D \subseteq \mathrm{dom}(f_\tau)$ *it is true that*

$$\Phi^{\delta,e,x,D}(\overline{Y}|\overline{f_\delta(y)}, \overline{\sigma}|f_\tau(\overline{\sigma})) \simeq 0 \;\Leftrightarrow\; g_\tau \Vdash_{i-1} F_e(x).$$

Lemma 9. *There exists an uniform effective way, given an $(i-1)$-pattern σ, natural numbers l, e and x and a finite set $D = \{y_1, \ldots, y_r\}$, to define a Σ^+_{i-1}-formula $\Phi^{\sigma,l,e,x,D}$ with free variables among* $\mathrm{var}(y_1) = Y_1, \ldots, \mathrm{var}(y_r) = Y_r$, $\overline{\sigma}$ *such that for each $(i-1)$-regular finite part τ coordinated with σ and such that* $D \subseteq \mathrm{dom}(f_\tau)$ *it is true that*

$$\Phi^{\delta,l,e,x,D}(\overline{Y}|\overline{f_\delta(y)}, \overline{\sigma}|f_\tau(\overline{\sigma})) \simeq 0$$

$$\Leftrightarrow \exists\Delta \supseteq \tau(\Delta \textit{ is } (i-1)\textit{-regular} \;\&\; \mathrm{lh}(g_\Delta) < l \;\&\; g_\Delta \Vdash_{i-1} F_e(x)).$$

Lemma 10. *There exists an uniform effective way, given an $(i-1)$-pattern σ and a finite set $D = \{y_1, \ldots, y_r\}$, to define a Σ^+_{i-1}-formula $\Phi^{\sigma,D}$ with free variables $U_1, \ldots, U_k$ among* $\mathrm{var}(y_1) = Y_1, \ldots, \mathrm{var}(y_r) = Y_r$, $\overline{\sigma}$ *such that for all $\overline{s} \in N^k$ it is true that*

$$\Phi^{\delta,D}(\overline{U}|\overline{s}) \simeq 0 \;\Leftrightarrow\; \exists\tau(\tau \text{ is an } (i-1)\textit{-regular finite part coordinated with } \sigma$$
$$\& \{y_1, \ldots, y_r\} \subseteq \mathrm{dom}(f_\tau)$$
$$\& f_\tau(\mathrm{var}^{-1}(U_1)) = s_1, \;\& \ldots \& f_\tau(\mathrm{var}^{-1}(U_k)) = s_k).$$

Let us fix g_δ, $y_1, \ldots, y_r$, e and x. Let D be a finite set of natural numbers. We say that D is compatible with g_δ iff the following conditions are true:

- Each $u \in D$ is of the form $\langle e_u, x_u, \varepsilon_u \rangle$, where $\varepsilon_u \in \{0,1\}$;
- There are not elements u and w of D such that $u = \langle e, x, 0\rangle$ and $w = \langle e, x, 1\rangle$;
- If $\langle e, x, \varepsilon\rangle \in D$ and $\langle e, y\rangle \leq |\delta|_i$, and if $\varepsilon = 0$, then $g_\delta \Vdash_{i-1} F_e(y)$, else $g_\delta \nVdash_{i-1} F_e(y)$.

Let σ be an i-pattern. We say that D is compatible with $\sigma = \langle \eta_0, n_0, \langle \xi_0, \varepsilon_0\rangle, b_0, \eta_1, \dots, \eta_r, n_r, \langle \xi_r, \varepsilon_r\rangle, b_r, \eta_{r+1}\rangle$ iff $\langle e, y, \varepsilon\rangle \in D$ implies $\varepsilon_{\langle e,y\rangle} = \varepsilon$. We call σ compatible with g_δ iff for each τ coordinated with σ, $g_\delta \subseteq g_\tau$. We define $\mathrm{lh}(\sigma)$ as $\mathrm{lh}(g_\tau)$, where τ is $(i-1)$-regular and coordinated with σ.

Consider the r. e. set

$$W = \{\langle \sigma, v\rangle | \langle v, x\rangle \in W_e \ \&\ \sigma \text{ is an } i\text{-pattern} \ \&\ D_v \text{ is compatible with } \sigma \text{ and } g_\delta \ \&\ \sigma \text{ is compatible with } g_\tau\}.$$

Let $\langle \sigma, v\rangle \in W$, $\sigma = \langle \eta_0, n_0, \langle \xi_0, \varepsilon_0\rangle, b_0, \eta_1, \dots, \eta_r, n_r, \langle \xi_r, \varepsilon_r\rangle, b_r, \eta_{r+1}\rangle$, $j > |\delta|_i$ and $D = \{y_1, \dots, y_r, x\}$.

I. $\varepsilon_j = 0$. We define

$$\Phi^j = \Phi^{\xi_j, D} \wedge \Phi^{\eta_{j+1}, D} \wedge \Phi^{\xi_j, (j)_0, (j)_1, D} \wedge \Phi_1 \wedge T_i(\mathrm{var}(b_j)),$$

where $\Phi^{\xi_j,D}$ and $\Phi^{\eta_{j+1},D}$ are the formulae from Lemma 10 and $\Phi^{\xi_j,(j)_0,(j)_1,D}$ is the formula from Lemma 8, and if $|\xi_j|_{i-1} > |\eta_j|_{i-1} + 1$, then $\Phi_1 = \neg\Phi^{\eta_j * \langle n_0\rangle, \mathrm{lh}(\xi_j), (j)_0, (j)_1, D}$, where $\Phi^{\eta_j * \langle n_0\rangle, \mathrm{lh}(\xi_j), (j)_0, (j)_1, D}$ is the formula from Lemma 9, else $\Phi_1 = (Z = Z)$.

II. $\varepsilon_j = 1$. We define

$$\Phi^j = \Phi^{\xi_j, D} \wedge \Phi^{\eta_{j+1}, D} \wedge \neg\Phi^{\xi_j, (j)_0, (j)_1, D} \wedge T_i(\mathrm{var}(b_j)),$$

where the first two formulae are the same as above and $\Phi^{\xi_j,(j)_0,(j)_1,D}$ is the formula from Lemma 7.

We denote by E the set of all variables in $\overline{\sigma}$ and $\{Y_1, \dots, Y_r\}$. Let $\{W_1, \dots, W_p\}$ be the set $E \setminus \{Y_1, \dots, Y_r\}$. If $\mathrm{var}(x) \in E$, we define

$$\varphi^\sigma = \bigwedge_{U,W \in E, U \neq W} U \neq W \wedge \mathrm{var}(x) = Z,$$

else

$$\varphi^\sigma = \bigwedge_{U,W \in E, U \neq W} U \neq W.$$

Let $\Phi^{\langle \sigma, v\rangle}$ be the formula

$$\exists W_1, \dots, \exists W_p (\bigwedge_{j > |\delta|_i} \Phi^j \wedge \varphi^\sigma).$$

Note that the above is an elementary Σ_i^+-formula. Now we are ready to define our Σ_i^+-formula:

$$\Phi^{\delta, e, x} = \bigvee_{\langle \sigma, v\rangle \in W} \Phi^{\langle \sigma, v\rangle}.$$

This completes the proof of Lemma 6.

Using the previous lemma and Theorem 2, one can easily obtain our main result, Theorem 1.

4. CONCLUSIONS

In the papers [2] and [4], a normal form of the Σ_n-admissible sets in total structures is obtained. In the particular case, when $B_1 = \ldots = B_n = N$, we find a normal form for the sets which are Σ_{n+1}-admissible in some partial structure. It would be interesting to extend the method of regular enumerations for the constructive ordinals and to prove a similar theorem.

Acknowledgements. The author would like to thank a lot Ivan Soskov for his help and patience.

REFERENCES

1. Ash, C. J. Generalizations of enumeration reducibility using recursive infinitary propositional sentences. *Ann. Pure Appl. Logic*, **58**, 1992, 173–184.
2. Ash, C. J., J. F. Knight, M. Manasse, and T. Slaman. Generic copies of countable structures. *Ann. Pure Appl. Logic*, **42**, 1989, 195–205.
3. Case, J. Maximal arithmetical reducibilities. *Z. Math. Logik Grundlag. Math.*, **20**, 1974, 261–270.
4. Chisholm, J. Effective model theory vs. recursive model theory. *J. Symbolic Logic*, **55**, 1990, 1168–1191.
5. Cooper, B. S. Enumeration reducibility, nondeterministic computations and relative computability of partial functions. In: *Lecture Notes in Mathematics*, **1432**, Recursion theory week, Overwolfach, 1989 (G. E. Sacks, K. Ambos-Spies, G. Muler, eds.), Springer-Verlag, 1990, 57–110.
6. McEvoy, K. Jumps of quasi-minimal enumeration degrees. *J. Symbolic Logic*, **50**, 1985, 839–848.
7. Rogers Jr., H. Theory of recursive functions and effective computability. McGraw-Hill Book Company, New York, 1967.
8. Selman, A. L. Arithmetical reducibilities. *Z. Math. Logik Grundlag. Math.*, **17**, 1971, 335–350.
9. Soskov, I. N. A jump inversion theorem for the enumeration jump. *Arch. Math. Logic* (to appear).
10. Soskov, I. N. Intrinsically Π^1_1-relations. *Math. Logic Quarterly*, **42**, 1996, 109–126.

Received February 28, 2000

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 James Bourchier Blvd.
BG-1164 Sofia, Bulgaria
E-mail: `vbaleva@fmi.uni-sofia.bg`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ"
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI"
FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

TERMAL EQUATIONS AND FINITE CONTROLLABILITY

ANTON ZINOVIEV

The notions termal system and solution of termal system in a structure are given. It is shown that any unsolvable in some structure termal system is unsolvable in some finite structure as well. Then this result is applied to show the finite controllability of some classes of formulae.

Keywords: finite structure, finite controllability, termal system
MSC 2000: 03C13, 03C05

1. INTRODUCTION

The system of termal equations is a commonly used notion in the unification theory. This theory has many applications in the modern in recent years constraint logic programming. As it is known, the Robinson unification used in the traditional logic programming languages such as Prolog gives a complete set of solutions. In opposite to that, in the constraint solving it is usually sufficient just to decide satisfiability.

Many used in practice mathematical structures are finite. For example, it is possible to think of a database simply as a finite structure. An important part of mathematics is the study of finite structures as finite graphs or finite groups. This makes it interesting to analyse the connections between constraint solving and the finite-model theory.

In this work we will show that the solvability of the finite systems of termal equations is finite controllable problem. If such a system is unsolvable in some structure, then it is unsolvable in some finite structure, too. Then this result is applied to show the finite controllability of some simple classes of formulae.

In this paper the set of all variables will be denoted by Var. The universe of the structure $\mathfrak{M}$ will be denoted by $|\mathfrak{M}|$. $t^{\mathfrak{M}}[I]$ is the value of the term t in the structure $\mathfrak{M}$ with the interpretation $I\colon \mathrm{Var} \to |\mathfrak{M}|$. A sentence is finite satisfiable iff it is satisfiable in a finite structure. A sentence is finite controllable if it is either unsatisfiable or finite satisfiable.

2. TERMAL SYSTEMS

Definition 2.1. (i) *Termal equation* is an expression having the form $t \sim s$, where t and s are terms.

(ii) *Termal system* is a finite set of termal equations.

Definition 2.2. Given a structure $\mathfrak{M}$, we say the interpretation $I\colon \mathrm{Var} \to |\mathfrak{M}|$ is *a solution of the equation* $t \sim s$ iff $t^{\mathfrak{M}}[I] = s^{\mathfrak{M}}[I]$. We say the interpretation I is a solution of a termal system iff it is a solution of all equations in the system.

Lemma 2.1. *Any termal equation in the form $x \sim t$ or $t \sim x$, where $t \neq x$ and x is a variable occurring in the term t, is unsolvable in some finite structure.*

Proof. The term t is not variable, so t has the form $f(t_1, \dots, t_n)$. For some N, $1 \le N \le n$, the variable x occurs in the term t_N. Let T be the set of all subterms of t that contain the variable x. Define $|\mathfrak{M}| \stackrel{\mathrm{def}}{=} 2^T$ and let for any functional symbol g, $g^{\mathfrak{M}}(T_1, \dots, T_m)$ be the set

$$\{g(s_1, \dots, s_m) \in T\colon \forall i \in \{1, \dots, m\}(s_i \in T \Rightarrow s_i \in T_i)\} \cup M, \tag{2.1}$$

where $M = \{x\}$ iff $f = g$ and $t_N \in T_N$, and $M = \emptyset$ iff $f \neq g$ or $t_N \notin T_N$.

Let $I\colon \mathrm{Var} \to |\mathfrak{M}|$ be an arbitrary interpretation.

By induction on complexity of the term s we will show that $s \in T$ implies

$$s \in s^{\mathfrak{M}}[I] \Leftrightarrow x \in I(x). \tag{2.2}$$

If s is a variable and $s \in T$, then $s = x$ and hence (2.2) is obvious. Otherwise, if s is $g(s_1, \dots, s_m)$ and $s \in T$, then from (2.1) it follows that

$$s \in s^{\mathfrak{M}}[I] \Leftrightarrow \forall i \in \{1, \dots, m\}(s_i \in T \Rightarrow s_i \in s_i^{\mathfrak{M}}[I]). \tag{2.3}$$

For all $s_N \in T$, $N \in \{1, \dots, m\}$, from the induction hypothesis it follows that $s_N \in s_N^{\mathfrak{M}}[I] \Leftrightarrow x \in I(x)$. Moreover, there exists N such that $s_N \in T$. Hence

$$x \in I(x) \Leftrightarrow \forall i \in \{1, \dots, m\}(s_i \in T \Rightarrow s_i \in s_i^{\mathfrak{M}}[I]). \tag{2.4}$$

From (2.3) and (2.4) follows (2.2).

From (2.1) it follows $x \in t^{\mathfrak{M}}[I] \Leftrightarrow t_N \notin t_N^{\mathfrak{M}}[I]$, and from (2.2) it follows $t_N \notin t_N^{\mathfrak{M}}[I] \Leftrightarrow x \notin I(x) \Leftrightarrow x \notin x^{\mathfrak{M}}[I]$. Hence $t^{\mathfrak{M}}[I] \neq x^{\mathfrak{M}}[I]$. The given termal equation ($t \sim x$ or $x \sim t$) is unsolvable in the structure $\mathfrak{M}$. □

Theorem 2.2. *Any termal system, unsolvable in some structure, is unsolvable in some finite structure.*

Proof. Denote by δ the number of the equations in the system, by α the number of variables in the system, by β the depth of the most complex term in the system if $\delta \neq 0$, and $\beta = 0$ if $\delta = 0$. Denote by γ the number of the equations containing some term with depth β. We are going to prove the theorem by induction on the ordinal $\alpha\omega^3 + \beta\omega^2 + \gamma\omega + \delta$.

Case 0. The system contains no equations (i.e. it is the empty set). In this case the theorem is obviously true, because in any structure any interpretation is solution of the empty system.

Case 1. Among the equations containing a term with complexity β some equation has the form $x \sim x$, where x is a variable. Make new system containing the other equations and use the induction hypothesis for the new system.

Case 2. Among the equations containing a term with complexity β some equation has the form $x \sim t$ or $t \sim x$, where x is a variable occurring in the term t and $t \neq x$. According to Lemma 2.1 there exists a finite structure, where the equation $x \sim t$ is unsolvable, and hence in that structure the whole system is unsolvable too.

Case 3. Among the equations containing a term with complexity β some equation has the form $x \sim t$ or $t \sim x$ and the variable x does not occur in the term t. Make new system containing the other equations replacing everywhere in them x by t. It is obvious that the former system is unsolvable in the structures where the new system is unsolvable. Moreover, if I is a solution of the new system in some structure $\mathfrak{M}$, then I' is a solution of the former system in $\mathfrak{M}$ defining

$$I'(y) \overset{\text{def}}{=} \begin{cases} I(y) & \text{if } y \neq x, \\ t^{\mathfrak{M}}[I] & \text{if } y = x. \end{cases}$$

Hence the new system is unsolvable in the structures where the former one is unsolvable and the theorem follows from the induction hypothesis for the new system.

Case 4. Among the equations containing a term with complexity β some equation has the form

$$f(t_1, \ldots, t_n) \sim g(s_1, \ldots, s_m) \tag{2.5}$$

and $f \neq g$. Define $|\mathfrak{M}| \overset{\text{def}}{=} \{0, 1\}$, $f^{\mathfrak{M}}(\mu_1, \ldots, \mu_n) \overset{\text{def}}{=} 0$ and $g^{\mathfrak{M}}(\mu_1, \ldots, \mu_m) \overset{\text{def}}{=} 1$. Thus the equation (2.5) is unsolvable in $\mathfrak{M}$, so the system is unsolvable in $\mathfrak{M}$ too.

Case 5. Among the equations containing a term with complexity β some equation has the form

$$f(t_1, \ldots, t_n) \sim f(s_1, \ldots, s_n) \ . \tag{2.6}$$

Make new system replacing (2.6) by the equations $t_1 \sim s_1, \ldots, t_n \sim s_n$. In any structure the solutions of the new system are solutions of the former one and hence if the former system is unsolvable in some structure, the new one is unsolvable there, too. Moreover, given a finite structure $\mathfrak{M}$, define $|\mathfrak{N}| \overset{\text{def}}{=} |\mathfrak{M}| \times |\mathfrak{M}|^n$ and

$$g^{\mathfrak{N}}(\langle \alpha_1, \beta_1 \rangle, \ldots, \langle \alpha_m, \beta_m \rangle) \overset{\text{def}}{=} \begin{cases} \langle f^{\mathfrak{M}}(\alpha_1, \ldots, \alpha_m), \langle \alpha_1, \ldots, \alpha_m \rangle \rangle & \text{if } g = f, \\ \langle g^{\mathfrak{M}}(\alpha_1, \ldots, \alpha_m), \gamma \rangle & \text{if } g \neq f, \end{cases}$$

where γ is an arbitrary element of $|\mathfrak{M}|^n$. Let π be the projection of $|\mathfrak{N}|$ on $|\mathfrak{M}|$. For any interpretation $I\colon \mathrm{Var} \to |\mathfrak{N}|$ and any term t

$$t^{\mathfrak{M}}[\pi \circ I] = \pi \circ t^{\mathfrak{N}}[I].$$

Hence if I is a solution of the former system in $\mathfrak{N}$, then $\pi \circ I$ is a solution of the new system in $\mathfrak{M}$ and therefore if the new system is unsolvable in some finite structure, then the former one is unsolvable in some finite structure, too, and the theorem follows from the induction hypothesis for the new system. □

Corollary 2.3. *Let n termal systems be given and they have no solution in some n structures. Then there exists a finite structure where none of these systems has any solution.*

Proof. According to Theorem 2.2 there exist finite structures $\mathfrak{M}_1, \ldots, \mathfrak{M}_n$ where the n systems have no solution correspondingly. Let $\mathfrak{M}$ be the cartesian product of the structures $\mathfrak{M}_1, \mathfrak{M}_2, \ldots, \mathfrak{M}_n$, i.e. $|\mathfrak{M}| = |\mathfrak{M}|_1 \times \ldots \times |\mathfrak{M}|_n$, and for any functional symbol f the following equation is valid (where m is the arity of f):

$$\begin{aligned} f^{\mathfrak{M}}(\langle \mu_{11}, \ldots, \mu_{1n}\rangle, \ldots, \langle \mu_{m1}, \ldots, \mu_{mn}\rangle) \\ = \langle f^{\mathfrak{M}_1}(\mu_{11}, \ldots, \mu_{m1}), \ldots, f^{\mathfrak{M}_n}(\mu_{1n}, \ldots, \mu_{mn})\rangle. \end{aligned} \tag{2.7}$$

Suppose that one of the n systems has a solution in $\mathfrak{M}$, say the system

$$\{t_i \sim s_i \colon i \in \{1, 2, \ldots, k\}\} \tag{2.8}$$

has a solution $I\colon \mathrm{Var} \to |\mathfrak{M}|$. Therefore for $i = 1, 2, \ldots, k$ we have

$$t_i^{\mathfrak{M}}[I] = s_i^{\mathfrak{M}}[I]. \tag{2.9}$$

Let $I_j\colon \mathrm{Var} \to |\mathfrak{M}|_i$, $j = 1, 2, \ldots, n$, be the unique functions such that for any variable x we have $I(x) = \langle I_1(x), \ldots, I_n(x)\rangle$. By induction on term complexity using (2.7) it may be shown that for any term t

$$t^{\mathfrak{M}}[I] = \langle t^{\mathfrak{M}_1}[I_1], \ldots, t^{\mathfrak{M}_n}[I_n]\rangle. \tag{2.10}$$

From (2.9) and (2.10) it follows that for all $i = 1, \ldots, k$ and $j = 1, \ldots, n$

$$t_i^{\mathfrak{M}_j}[I_j] = s_i^{\mathfrak{M}_j}[I_j],$$

but this is a contradiction, because the system (2.8) has no solution in at least one of the structures $\mathfrak{M}_1, \mathfrak{M}_2, \ldots, \mathfrak{M}_n$. □

3. FINITE CONTROLLABILITY OF SOME CLASSES OF FORMULAE

Theorem 3.1. *Every satisfiable finite set A whose elements are closed formulae in the form $\forall x_1 \ldots \forall x_n \varphi$ or $\forall x_1 \ldots \forall x_n \neg\varphi$, where φ is an atomic formula, is satisfiable in some finite structure.*

Proof. Without a loss of generality we may assume that no variable occurs in two different formulae in A. Denote by D^+ the set of all atomic formulae $p(t_1, \dots, t_m)$ such that there is a formula $\forall x_1 \dots \forall x_n p(t_1, \dots, t_m)$ belonging to A. Similarly, denote by D^- the set of all atomic formulae $P(t_1, \dots, t_m)$ such that there is a formula $\forall x_1 \dots \forall x_n \neg P(t_1, \dots, t_m)$ belonging to A. Obviously, D^+ and D^- are finite sets. From Corollary 2.3 it follows that there is a finite structure $\mathfrak{M}$ where for any two atomic formulae $p(t_1, \dots, t_n) \in D^+$ and $p(s_1, \dots, s_n) \in D^-$ the termal system $\{t_1 \sim s_1, t_2 \sim s_2, \dots, t_n \sim s_n\}$ has no solution. Define new structure $\mathfrak{N}$ such that $|\mathfrak{N}| = |\mathfrak{M}|$ and for all functional symbols $f^{\mathfrak{N}} = f^{\mathfrak{M}}$. For predicate symbols p let $p^{\mathfrak{N}}$ be the set of all tuples $\langle \alpha_1, \dots, \alpha_n \rangle$ such that for some atomic formula $p(t_1, \dots, t_n) \in D^+$ and interpretation I and for all $i = 1, 2, \dots, n$ $\alpha_i = t_i^{\mathfrak{N}}[I]$. Clearly, all formulae $\forall x_1 \dots \forall x_n p(t_1, \dots, t_m)$ from the set A are true in $\mathfrak{N}$. Suppose a formula $\forall x_1 \dots \forall x_n \neg p(t_1, \dots, t_m) \in A$ is false in $\mathfrak{N}$. Then $\langle t_1^{\mathfrak{N}}[I], \dots, t_n^{\mathfrak{N}}[I] \rangle \in p^{\mathfrak{N}}$ for some interpretation I. By the definition of $p^{\mathfrak{N}}$ this means that there exist a formula $p(s_1, \dots, s_n) \in D^+$ and interpretation J such that $t_i^{\mathfrak{N}}[I] = s_i^{\mathfrak{N}}[J]$, $i = 1, 2, \dots, n$. According to the assumption in the beginning of the proof the terms t_i and s_i have no common variables and so we may assume that $I = J$. This is a contradiction, because we obtain a solution of the termal system $\{t_1 \sim s_1, \dots, t_n \sim s_n\}$ in $\mathfrak{N}$. □

Theorem 3.2. *Every satisfiable finite set A of closed formulae is satisfiable in a finite structure provided each of the formulae in A is built from atomic formulae and their negations by conjunction and quantifiers.*

Proof. Let φ be the conjunction of the formulae in A. Then φ is satisfiable just in the structures where A is satisfiable. Let φ' is equivalent to φ and in the form $Q_1 x_1 \dots Q_n x_n (L_1 \wedge \dots \wedge L_m)$, where all Q_i are either $\forall$ or $\exists$ and $L_1, \dots, L_m$ are atomic formulae and negations of atomic formulae. The formula φ' is true just in the structures where φ is true. By skolemization we obtain a formula φ'' in the form $\forall x_1 \dots \forall x_n (L'_1 \wedge \dots \vee L'_m)$ such that φ'' is satisfiable iff φ' is satisfiable and φ' is true in the structures where φ'' is true. Denote by B the set $\{\forall x_1 \dots \forall x_n L'_1, \dots, \forall x_1 \dots \forall x_n L'_n\}$. The formulae in B are simultaneously true just in the structures where φ'' is true. By Theorem 3.1 if B is satisfiable, then B is satisfiable in some finite structure. □

4. CONCLUSION

It may be expected that the result in §2 will have many other applications. The execution of a Prolog program can be thought as a constructing of a solvable termal system. More generally, it is possible to think of the searching of proof in a formal deductive system as searching of solvable termal system that has some additional property. This topic can be theme of a future publication.

Acknowledgements. The author would like to thank prof. Dimiter Skordev, who was interested in the truth problem of Theorem 3.1 and thus inspired this work.

REFERENCES

1. Chang, C. C., and H. J. Keisler. Model Theory. North-Holland, 1973.
2. Colmerauer, A. An Introduction to PROLOG III. *C. ACM*, **33**, 1990.
3. Barwise, J. An Introduction to First-Order Logic. In: *Handbook of Mathematical Logic* (J. Barwise, ed.). North-Holland, Amsterdam – New York – Oxford, 1977.
4. Fagin, R. Probabilities on finite models. *J. Symb. Logic*, **41**, 1, March 1976, 50–58.

Received March 16, 2000

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 James Bourchier Blvd.
BG-1164 Sofia, Bulgaria
E-mail: anton@lml.bas.bg

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

ON A CLASS OF MEROMORPHIC FUNCTIONS WITH NEGATIVE COEFFICIENTS

DONKA Z. PASKOULEVA

In this paper we obtain coefficient inequalities and distortion theorems for the class $T^*(\alpha, \beta, A)$ *of meromorphic functions with negative coefficients.*

Keywords: meromorphic functions, convex functions
MSC 2000: main 30C45, secondary 30C50

1. INTRODUCTION

Let Σ denote the class of functions of the form

$$f(z) = \frac{1}{z} + \sum_{n=1}^{\infty} a_n z^n \tag{1.1}$$

which are analytic in the punctured unit disk

$$U^* = \{z : z \in \mathbb{C} \text{ and } 0 < |z| < 1\}$$

with a simple pole at the origin and with residue 1 there. Let Σ^* denote the subclass of Σ consisting of functions $f(z)$, which are convex with respect to the origin, that is, satisfying the condition

$$\Re\left\{-\left(1 + \frac{zf''(z)}{f'(z)}\right)\right\} > 0, \quad z \in U^* . \tag{1.2}$$

Let $\Sigma^*(\alpha)$ denote the subclass of Σ consisting of functions $f(z)$ which are convex of order α, that is, satisfying the condition

$$\Re\left\{-\left(1+\frac{zf''(z)}{f'(z)}\right)\right\} > \alpha, \quad z \in U^*, \ 0 \leq \alpha < 1 . \tag{1.3}$$

Let $\Sigma^*(\alpha, A)$ denote the class of functions $f(z) \in \Sigma$ such that

$$1+\frac{zf''(z)}{f'(z)} = -\frac{1+(A-\alpha A+\alpha)w(z)}{1+w(z)} . \tag{1.4}$$

Here $w(z)$ is analytic in $U = \{z : |z| < 1\}$ and satisfies the conditions

$$w(0) = 0 \text{ and } |w(z)| < 1, \quad z \in U.$$

The condition (1.4) is equivalent to

$$\left|\frac{\frac{zf''(z)}{f'(z)}+2}{1+\frac{zf''(z)}{f'(z)}+(A-\alpha A+\alpha)}\right| < 1, \quad z \in U^*. \tag{1.5}$$

We note also that

$$\Sigma^*(\alpha, -1) = \Sigma^*(\alpha).$$

Let T denote the subclass of Σ consisting of functions of the form

$$f(z) = \frac{1}{z} - \sum_{n=1}^{\infty} |a_n| z^n. \tag{1.6}$$

A function $f(z) \in \Sigma$ is in the class $\Sigma^*(\alpha, \beta, A)$ if it satisfies the condition

$$\left|\frac{\frac{zf''(z)}{f'(z)}+2}{1+\frac{zf''(z)}{f'(z)}+(A-\alpha A+\alpha)}\right| < \beta, \tag{1.7}$$

$z \in U^*$, $0 \leq \alpha < 1$, $0 < \beta \leq 1$, $-1 \leq A < 1$.

Let us write

$$T^*(\alpha, \beta, A) = \Sigma^*(\alpha, \beta, A) \cap T .$$

We note that

$$T^*(\alpha, 1, -1) = T^*(\alpha)$$

is the class of meromorphically convex functions of order α with negative coefficients, which was studied by Uralegaddi and Ganigi [1].

In this paper we obtain coefficient inequalities and distortion theorems for the class $T^*(\alpha, \beta, A)$. We employ techniques similar to those used earlier by Silverman [2].

Theorem 2.1. *Let the function $f(z)$ defined by (1.1) be analytic in U^*. If*

$$\sum_{n=1}^{\infty}\{(n+1)+\beta[n+(1-A)\alpha+A]\}n|a_n| \leq (1-A)\beta(1-\alpha)\ , \tag{2.1}$$

$0 \leq \alpha < 1$, $0 < \beta \leq 1$, $-1 \leq A < 1$, *then* $f(z) \in \Sigma^*(\alpha, \beta, A)$.

Proof. Let (2.1) hold true for all admissible values of α, β, A. Let us consider the expression

$$F(f, f') = |zf''(z) + 2f'(z)| - \beta|f'(z) + zf''(z) + (A - A\alpha + \alpha)f'(z)|. \tag{2.2}$$

Replacing f and f' by their series expansions, for $0 < |z| = r < 1$ we have

$$F(f, f') = \left|\sum_{n=1}^{\infty}(n+1)na_n z^{n-1}\right| - \beta\left|\frac{(1-A)(1-\alpha)}{z^2} + \sum_{n=1}^{\infty}[n+(1-A)\alpha+A]na_n z^{n-1}\right|.$$

Now

$$\begin{aligned} r^2F(f, f') &\leq \sum_{n=1}^{\infty}(n+1)n|a_n|r^{n+1} \\ &\quad - \beta\left\{(1-A)(1-\alpha) - \sum_{n=1}^{\infty}[n+(1-A)\alpha+A]n|a_n|r^{n+1}\right\} \\ &= \sum_{n=1}^{\infty}\{(n+1)+\beta[n+(1-A)\alpha+A]n|a_n|r^{n+1}\} - (1-A)\beta(1-\alpha). \end{aligned}$$

Since the above inequality holds true for all r $(0 < r < 1)$, letting $r \to 1-$, we have

$$F(f, f') \leq \sum_{n=1}^{\infty}\{(n+1)+\beta[n+(1-A)\alpha+A]\}\, n|a_n| - (1-A)\beta(1-\alpha) \leq 0$$

by (2.1).

Hence it follows that

$$\left|\frac{zf''(z)}{f'(z)} + 2\right| < \beta\left|1 + \frac{zf''(z)}{f'(z)} + A - A\alpha + \alpha\right|,$$

so that $f(z) \in \Sigma^*(\alpha, \beta, A)$. □

Theorem 2.2. *Let the function $f(z)$ defined by (1.6) be abalytic in U^*. Then $f(z) \in T^*(\alpha, \beta, A)$ iff (2.1) is satisfied.*

Proof. In view of Theorem 2.1, let us assume that the function $f(z)$ defined by (1.6) is in the class $T^*(\alpha, \beta, A)$. Then

$$\left| \frac{\dfrac{zf''(z)}{f'(z)} + 2}{1 + \dfrac{zf''(z)}{f'(z)} + (A - A\alpha + \beta)} \right|$$

$$= \left| \frac{-\sum_{n=1}^{\infty}(n+1)n|a_n|z^{n-1}}{\dfrac{(1-A)(1-\alpha)}{z^2} - \sum_{n=1}^{\infty}[n + (1-A)\alpha + A]n|a_n|z^{n-1}} \right| < \beta, \quad z \in U^*.$$

But $\Re(z) \leq |z|$ for all z. Thus we have

$$\Re\left\{ \frac{\sum_{n=1}^{\infty}(n+1)n|a_n|z^{n-1}}{\dfrac{(1-A)(1-\alpha)}{z^2} - \sum_{n=1}^{\infty}[n + (1-A)\alpha + A]n|a_n|z^{n-1}} \right\} < \beta, \quad z \in U^*. \tag{2.3}$$

Now we choose the values of z on the real axis, so that $1 + \dfrac{zf''(z)}{f'(z)}$ is real. Upon clearing the denominator in (2.3) and letting $z \to 1-$ through real values, we obtain

$$\sum_{n=1}^{\infty}(n+1)n|a_n| \leq \beta \left\{ (1-A)(1-\alpha) - \sum_{n=1}^{\infty}[n + (1-A)\alpha + A]n|a_n| \right\}$$

or

$$\sum_{n=1}^{\infty} \{(n+1) + \beta[n + (1-A)\alpha + A]\}\, n|a_n| \leq (1-A)\beta(1-\alpha), \tag{2.4}$$

which proves the theorem. □

3. A DISTORTION THEOREM

Theorem 3.1. *Let the function $f(z)$ defined by (1.6) be in the class $T^*(\alpha, \beta, A)$. Then for $0 < |z| = r < 1$*

$$\frac{1}{r} - \frac{(1-A)\beta(1-\alpha)}{2 + \beta[1 + A + (1-A)\alpha]}r \leq |f(z)| \leq \frac{1}{r} + \frac{(1-A)\beta(1-\alpha)}{2 + \beta[1 + A + (1-A)\alpha]}r, \tag{3.1}$$

$$\frac{1}{r^2}-\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}\leq |f'(z)|$$
$$\leq \frac{1}{r^2}+\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}. \tag{3.2}$$

The result is sharp. The equality holds true for the function $f(z)$ given by

$$f(z)=\frac{1}{z}-\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}z \tag{3.3}$$

at $z=r$.

Proof. In view of Theorem 2.2, we have

$$\sum_{n=1}^{\infty}|a_n|\leq\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}.$$

Thus for $0<|z|=r<1$

$$|f(z)|\leq\frac{1}{r}+\sum_{n=1}^{\infty}|a_n|r^n\leq\frac{1}{r}+r\sum_{n=1}^{\infty}|a_n|\leq\frac{1}{r}+\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}r$$

and

$$|f(z)|\geq\frac{1}{r}-\sum_{n=1}^{\infty}|a_n|r^n\geq\frac{1}{r}-r\sum_{n=1}^{\infty}|a_n|\geq\frac{1}{r}-\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}r.$$

Furthermore, it follows from Theorem 2.2 that

$$\sum_{n=1}^{\infty}n|a_n|\leq\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}.$$

Hence

$$|f'(z)|\leq\frac{1}{r^2}+\sum_{n=1}^{\infty}n|a_n|r^{n-1}\leq\frac{1}{r^2}+\sum_{n=1}^{\infty}n|a_n|\leq\frac{1}{r^2}+\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}$$

and

$$|f'(z)|\geq\frac{1}{r^2}-\sum_{n=1}^{\infty}n|a_n|r^{n-1}\geq\frac{1}{r^2}-\sum_{n=1}^{\infty}n|a_n|\geq\frac{1}{r^2}-\frac{(1-A)\beta(1-\alpha)}{2+\beta[1+A+(1-A)\alpha]}.$$

It can easily be seen that the function $f(z)$ defined by (3.3) is extremal for Theorem 3. □

REFERENCES

1. Uralegaddi, B. A., M. D. Ganigi. Meromorphic convex functions with negative coefficients. *J. Math. Res. Exposition*, **1**, 1987, 21–26.
2. Silverman, H. Univalent functions with negative coefficients. *Proc. Amer. Math. Soc.*, **51**, 1975, 109–116.

Received February 28, 2000

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 8,
BG-1113 Sofia, Bulgaria

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ"

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI"

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATION OF CONSTANT DRIFT AND DIFFUSION PARAMETERS IN k-DIMENSIONAL DIFFUSION PROCESS OBSERVED AT DISCRETE RANDOM SAMPLING

DIMITRINKA IVANOVA VLADEVA

This article is concerned with the problem of a parameter estimation of the constant drift and diffusion coefficients: unknown vector A and unknown positive definite matrix B, respectively, of a k-dimensional diffusion type process, when the observations at the moment of random point process are given. We compute the means and variances of the maximum likelihood estimators and establish their asymptotic properties. The unbiasedness, the strong consistency and the asymptotical efficiency of the estimation for A are proved. The estimator of B is unbiased and consistent and the variance of this estimator does not depend on the distribution of the random moments of observations.

Keywords: diffusion process, k-dimensional Wiener process, discrete random sampling, maximum likelihood estimation, unbiasedness, consistency, efficiency, Fisher information

MSC 2000: 62H12, 62M05

1. INTRODUCTION

In this article we consider the diffusion process $X_t = (X_t^1, X_t^2, \dots, X_t^k)^T$, $t \geq 0$, defined by the stochastic differential equation

$$dX_t = A\,dt + B^{1/2}\,dW_t, \quad t \geq 0, \quad X_0 = 0, \tag{1}$$

where $A = (a^1, a^2, \dots, a^k)^T$ and

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ \dots & & & \\ b_{k1} & b_{k2} & \dots & b_{kk} \end{pmatrix}$$

are unknown constant vector and positive definite symmetric matrix, respectively, and $W_t = (W_t^1, W_t^2, ..., W_t^k)^T$ is a standard Wiener process with a mean 0 and a variance I_k.

The solution of the differential equation (1) exists in a strong sense, it is unique and is represented by the process

$$X_t = At + B_1 W_t, \quad t \geq 0. \tag{2}$$

More information about this problem can be found in [1].

The maximum likelihood estimation problem for the model (1), when we observe the process X_t, $t \geq 0$, continuously in the interval $[0, T]$, is solved in [2]. In the case when we have at disposal the discrete observations at equidistant points, many close to this one problems can be found in the monography [3]. At first, the random sampling scheme has been used by J. Beutler in [4]. Recently, many authors (see [5] and [6]) consider continuous diffusion processes when the observations are provided in discrete moments belonging to the interval $[0, T]$. In [7] A. Le Breton has solved the estimation problem for the model (1) when the points of observations are determinant.

Usually, the maximum of the distance between the points of observations tends to zero, while their number tends to infinity.

Our conditions are more natural. Let us denote the observations $X_{t_1}, \ldots, X_{t_N}$, $X_{t_i} = (X_{t_i}^1, X_{t_i}^2, \ldots, X_{t_i}^k)^T$. The moments $t_1, \ldots, t_N$ are the first N points of an arbitrary point process with independent identically distributed increments. The process $\{t_i\}$, $i = 1, \ldots, N$, is independent of the process X_t, $t \geq 0$, and we compute $E(F(X_t)) = E_t(E_X(F(X_t)))$. The problem is to find the maximum likelihood estimators of the unknown constant vector A and the matrix B and to establish their properties. In the one-dimensional case this problem was solved in [8].

Let us denote $X_i = X_{t_i}$, $\Delta X_i = X_i - X_{i-1}$, $\Delta_i = t_i - t_{i-1}$, $i = 1, \ldots, N$. For simplicity we denote $\overline{B} = B^{1/2}$.

2. MAXIMUM LIKELIHOOD ESTIMATION

Using the maximum likelihood method, we can prove the following natural results.

Theorem 1. *If $N \geq 2$, the statistic*

$$\hat{A}_N = \frac{X_N}{t_N} \tag{3}$$

is a maximum likelihood estimator for an unknown vector A.

We prove this theorem using the standart maximum likelihood procedure.

Theorem 2. *If $N \geq k$, the statistic*

$$\widetilde{B_N} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{\Delta X_i \Delta X_i^T}{\Delta_i} - \frac{X_N X_N^T}{t_N} \right\} \tag{4}$$

is a maximum likelihood estimator for an unknown matrix B, when $A \neq 0$.

For $A = 0$ the maximum likelihood estimator is

$$\widetilde{B_N} = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta X_i \Delta X_i^T}{\Delta_i}.$$

Our approach is different from the ones used in the proofs of similar propositions. (For example, see [9, p. 75].) Our proof is based on the following lemma.

Lemma 1. *Let $y_i = (y_i^1, \dots, y_i^k)^T$, $i = 1, \dots, N$, be k-dimensional vectors such that $B = \sum_{i=1}^{N} y_i y_i^T$ is a non-singular matrix. Then:*

a) *the matrix B is symmetric and positive definite;*

b) $C = \sum_{i=1}^{N} y_i^T \left(\sum_{i=1}^{N} y_i y_i^T \right)^{-1} y_i = k.$

Let us note that for the estimation of one of the parameters (A or B) it is not necessary to know the other one.

The estimator $\widetilde{B_N}$ is not unbiased. Therefore we prefer to use the estimator

$$\hat{B}_N = \frac{1}{N-1} \sum_{i=1}^{N} \left\{ \frac{\Delta X_i \Delta X_i^T}{\Delta_i} - \frac{X_N X_N^T}{t_N} \right\}. \tag{5}$$

The proofs of these theorems can be found in [12].

Our purpose here is to establish the properties of the maximum likelihood estimators, which are given in Theorems 1 and 2.

3. PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATORS

By arbitrary distribution of the process $t_1, \dots, t_N$ and natural conditions for its moments we can compute the means and variances of the considered estimators and to obtain their asymptotic properties.

Let $\max_{1 \le i \le N} \Delta_i$ do not tend to 0 when $N \to \infty$, and $t_N \to \infty$ when $N \to \infty$.

Theorem 3. *The estimator given by (5) is unbiased, strongly consistent and asymptotically effective for the unknown vector A if the condition*

$$\sum_{N=2}^{\infty} E(t_N^{-3}) < \infty \tag{6}$$

is satisfied.

Proof. We compute

$$E\hat{A}_N = E\left(\frac{X_N}{t_N}\right) = E_t E_X \left(\frac{X_N}{t_N}\right) = E_t \frac{A t_N}{t_N} = A.$$

Hence, $\hat{A}_N$ is an unbiased estimator.

The variance of this estimation is

$$E\hat{A}_N\hat{A}_N^T = E\left(\frac{X_N X_N^T}{t_N^2}\right) = \frac{E(At_N + \overline{B}W_N)(At_N + \overline{B}W_N)^T}{t_N^2}$$

$$= AA^T + B^2 E\frac{W_N W_N^T}{t_N^2} = AA^T + BE\frac{1}{t_N}\cdot$$

As $E\left(\frac{1}{t_N}\right) \to 0$ when $N \to \infty$, the considered statistic $\hat{A}_N$ is consistent.

A sufficient condition for the strong consistency of $\hat{A}_N$ is

$$\sum_{N=2}^{\infty} P\left(|\hat{A}_N - A| > \varepsilon\right) < \infty.$$

Let $\|A\|_4 = \left(E\sum_{i=1}^{k} a_i^4\right)^{1/4}$ be the norm of the random vector A and $\lambda = \|B\|$ be the maximum eigenvalue and the norm of the positive definite matrix B. Then using Markov's inequality, we obtain

$$\sum_{N=2}^{\infty} P\left(\left|\hat{A}_N - A\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^4}\sum_{N=2}^{\infty} E\left|\hat{A}_N - A\right|^4 \leq \frac{1}{\varepsilon^4}\sum_{N=2}^{\infty} E\left(\|\hat{A}_N - A\|_4\right)^4$$

$$= \frac{1}{\varepsilon^4}\sum_{N=2}^{\infty}\left\|\frac{\overline{B}W_N}{t_N}\right\|^4 \leq \frac{\lambda^2}{\varepsilon^4}\sum_{N=2}^{\infty} E\sum_{i=1}^{k}\left(\frac{W_N^i}{t_N}\right)^4 = \frac{3k\lambda^2}{\varepsilon^4}\sum_{N=2}^{\infty} E\frac{1}{t_N{}^3}\cdot$$

From condition (6) it follows that the estimator (3) is strongly consistent.

The Fisher information matrix for the estimator $\hat{A}_N$ is

$$I_{\hat{A}_N}(l,m) = E_A\left(\left\{\frac{\partial}{\partial a^l}\ln\frac{dP_{A,B}}{d\lambda}(x)\right\}\left\{\frac{\partial}{\partial a^m}\ln\frac{dP_{A,B}}{d\lambda}(x)\right\}\right),$$

where λ is the Lebesque measure in R^k.

Let c_{jl} be the (j,l)-th element of the matrix B^{-1} and $b_{i,j}$ be the (i,j)-th element of the matrix $\overline{B}$. We compute the first factor of $I_{\hat{A}_N}(l,m)$:

$$\frac{\partial}{\partial a^l}\ln\frac{dP_{A,B}}{d\lambda}(x) = \frac{\partial}{\partial a^l}l(A,B) = \sum_{i=1}^{N}\sum_{j=1}^{k} c_{lj}\left(a^j\Delta_i - \Delta X_i^j\right).$$

We use that

$$\Delta X_i^j = a^j\Delta_i + \sum_{n=1}^{k} b_{jn}\Delta W_i^n$$

and

$$\frac{\partial}{\partial a^l}l(A,B) = -\sum_{i=1}^{N}\sum_{j=1}^{k}\sum_{n=1}^{k} b_{jn}c_{lj}\Delta W_i^n.$$

Then

$$I_{\hat{A}_N}(l,m) = E\sum_{i=1}^{N}\sum_{j=1}^{k}\sum_{n=1}^{k}\sum_{p=1}^{k}\sum_{q=1}^{k} b_{jn}c_{lj}\Delta W_i^n b_{pq}c_{mp}\Delta W_i^q.$$

For $n \neq q$ the mathematical expectation is equal to zero and

$$E\sum_{i=1}^{N}(\Delta W_i^n)^2 = \sum_{i=1}^{N} E\Delta_i = Et_N,$$

so the information matrix is

$$I_{\hat{A}_N}(l,m) = E\sum_{i=1}^{N}\sum_{j=1}^{k}\sum_{p=1}^{k}\sum_{n=1}^{k} b_{jn}c_{lj}c_{mp}b_{pn}(\Delta W_i^n)^2$$

$$= Et_N\sum_{j=1}^{k}\sum_{p=1}^{k}\sum_{n=1}^{k} b_{jn}c_{lj}c_{mp}b_{pn} = Et_N\sum_{p=1}^{k}\sum_{j=1}^{k} B_{jp}c_{lj}c_{mp}$$

$$= Et_N\sum_{p=1}^{k} D_{lp}c_{mp} = Et_N c_{lm}.$$

Here B_{jp} denotes the (jp)-th element of the matrix B and D_{lp} denotes the (lp)-th element of the diagonal matrix $B^{-1}B = I_k$.

In this way we obtained that the information matrix is $B^{-1}Et_N$. Hence

$$\text{eff}\,\hat{A}_N \sim \left(BE\,\frac{1}{t_N}B^{-1}Et_N\right)^{-1} \to 1$$

when $N \to \infty$ and the estimator $\hat{A}_N$ is asymptotically effective.

The distribution of $\hat{A}_N$ (for a fixed t_N) is normal with parameters A and $\dfrac{B}{t_N}$, i.e. the vector

$$\left(\hat{A}_N - A\right)\overline{B}^{-1}\sqrt{t_N}$$

has a k-dimensional standard normal distribution.

Note. The sufficient conditions for strong consistency of the estimator can be written in terms of statistical moments of Δ_i.

For the estimator $\hat{B}_N$ we can establish the following properties:

Theorem 4. *The statistic* (5) *from Theorem* 2 *is unbiased and consistent estimation for the unknown matrix B and the variance of this estimation is*

$$E(\hat{B}_N - B)^2 = \frac{k+1}{N-1}B^2. \tag{7}$$

To prove these properties, we calculate the moments of the estimator. We need the next lemmas.

Lemma 2. *For every integer $k \geq 2$ the identity*

$$\sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} \min(i_1, \ldots, i_k) = \sum_{i=1}^{N} i^k$$

holds, where i_1, i_2, ..., i_k are natural numbers between 1 and N.

Proof. We consider the sum

$$\sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} \min(i_1, \ldots, i_k),$$

where $k \geq 2$. The minimum can accept all integer values from 1 to N. The number of all possible values of the k-dimensional variable $(i_1, i_2, \ldots, i_k)$ is N^k. The number 1 does not appear in $(N-1)^k$ cases. For the rest $N^k - (N-1)^k$ cases, $\min(i_1, \ldots, i_k) = 1$. Thus $\min(i_1, \ldots, i_k) = 2$ exactly $(N-1)^k - (N-2)^k$ times, ..., $\min(i_1, \ldots, i_k) = N$ only one time.

Therefore

$$\sum_{i_1,\ldots,i_k=1}^{N} \min(i_1, \ldots, i_k) = \sum_{i=1}^{N} (N - i + 1)[i^k - (i-1)^k]$$

$$= N.[1^k - 0^k] + \cdots + 1.[N^k - (N-1)^k] = \sum_{i=1}^{N} i^k.$$

If $k = 2$, we obtain

$$\sum_{i,j=1}^{N} \min(i, j) = \sum_{i=1}^{N} i^2 = \frac{N(N+1)}{2}.$$

Some characteristics of the considered processes will be usefull for our proofs. Let us denote

$$\Delta W_i = W_{t_i} - W_{t_{i-1}} = \left(W_{t_i}^1 - W_{t_{i-1}}^1, W_{t_i}^2 - W_{t_{i-1}}^2, \ldots, W_{t_i}^k - W_{t_{i-1}}^k\right)^T,$$

$\Delta_i = t_i - t_{i-1}$ and I_k is the identity matrix of dimension k. The arbitrary reneval point process t_i, $i = 1, \ldots, N$, is independent of the process X_t.

Lemma 3. *For the moments of the Wiener process the following equalities are satisfied:*

$$E\Delta W_i \Delta W_i^T = E\Delta_i I_k, \quad E\left(\Delta W_i \Delta W_i^T\right)^2 = (k+2)\, E\Delta_i^2 I_k,$$

$$E\Delta W_i \Delta W_j^T = 0,\ i \neq j, \quad E\left(\Delta W_i \Delta W_j^T\right)^2 = E\Delta_i^2 I_k,\ i < j,$$

$$E\Delta W_i \Delta W_i^T \Delta W_i = 0, \quad E\Delta W_i W_N^T = E\Delta_i I_k,$$

$$EW_N W_N^T = Et_N I_k, \quad E\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{\Delta_j}{t_N} = N - 1.$$

Proof. We will prove only the first two equalities.

We use the facts that B and $\hat{B}_N$ are symmetric matrices, the increments of the Wiener process are independent and their odd moments are equal to zero.

$$E\Delta W_i^l \Delta W_i^m = 0,\ l \neq m, \quad E(\Delta W_i^l)^2 = E\Delta_i,\ l,m = 1,\dots,k.$$

Then it holds:

$$E\Delta W_i \Delta W_i^T = \begin{pmatrix} E(\Delta W_i^1)^2 & E(\Delta W_i^1 \Delta W_i^2) & \dots & E(\Delta W_i^1 \Delta W_i^k) \\ & \dots & & \\ E(\Delta W_i^k \Delta W_i^1) & E(\Delta W_i^k \Delta W_i^2) & \cdots & E(\Delta W_i^k)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \Delta_i & 0 & \dots & 0 \\ \cdots & & & \\ 0 & 0 & \dots & \Delta_i \end{pmatrix} = E\Delta_i\, I_k.$$

Let d_{lm}, where $l,m = 1,2,\dots,k$, be the (l,m)-th element of the matrix $E(\Delta W_i \Delta W_i^T)^2$. Then

$$d_{lm} = E \sum_{\substack{n=1 \\ n\neq l,m}}^{k} (\Delta W_i^n)^2 \Delta W_i^l \Delta W_i^m + E\,\Delta W_i^l (\Delta W_i^m)^3$$
$$+ E\,(\Delta W_i^l)^3 \Delta W_i^m + E\,(\Delta W_i^l)^4.$$

So $d_{lm} = 0$ if $l \neq m,\ \ l,m = 1,\dots,k$, and

$$d_{ll} = E \sum_{\substack{n=1 \\ n\neq l}}^{k} (\Delta W_i^n)^2 (\Delta W_i^l)^2 + E\,(\Delta W_i^l)^4 = E \sum_{\substack{n=1 \\ n\neq l}}^{k} (\Delta W_i^n)^2 (\Delta W_i^l)^2 + E(\Delta W_i^l)^4$$

$$= (k-1)E\Delta_i^2 + 3E\Delta_i^2 = (k+2)E\Delta_i^2.$$

Hence

$$E(\Delta W_i \Delta W_i^T)^2 = (k+2)E\Delta_i^2\, I_k, \quad \forall i = 1,\dots,N,$$
$$E\Delta W_i W_N^T = E\Delta W_i^2 I_k = E\Delta_i I_k,$$
$$E W_N W_N^T = \sum_{i=1}^{N} E\Delta W_i \Delta W_i^T I_k = \sum_{i=1}^{N} E\Delta_i I_k = Et_N I_k.$$

Proof of Theorem 4. It is easy to establish that the estimator $\hat{B}_N$ is unbiased:

$$E\hat{B}_N = \frac{1}{N-1}\left(\sum_{i=1}^{N}\left(AA^T E\Delta_i + BE\frac{\Delta_i}{\Delta_i}\right) - AA^T t_N - BE\frac{t_N}{t_N}\right) = B.$$

After substituting $\Delta X_i = A\Delta_i + \overline{B}\Delta W_i$, $X_N = At_N + \overline{B}W_N$, we find:

$$E\left(\hat{B}_N^2 - B^2\right) = E\left(N-1\right)^{-2}\left(\left(\sum_{i=1}^{N}\Delta X_i \Delta X_i^T \Delta_i{}^{-1}\right)^2 - 2\sum_{i=1}^{N}\Delta X_i \Delta X_i^T X_N X_N^T \Delta_i t_N{}^{-1} + \left(X_N X_N^T\right)^2 t_N^{-2}\right) - B^2$$

$$= (N-1)^{-2}\sum_{i,j=1}^{N} E\left(\left(AA^T\right)^2\Delta_i\Delta_j + \overline{B}\Delta W_i \Delta W_i^T \overline{B}^T AA^T \Delta_j \Delta_i{}^{-1}\right.$$
$$+ AA^T\overline{B}\Delta W_j \Delta W_j^T \overline{B}^T \Delta_i \Delta_j{}^{-1} + \overline{B}\Delta W_i A\overline{B}\Delta W_j A$$
$$+ A\Delta W_i^T \overline{B}^T \overline{B}\Delta W_j A^T + \overline{B}\Delta W_i A^T A\Delta W_j^T \overline{B}^T + A\Delta W_i^T \overline{B}^T A\Delta W_j^T \overline{B}^T$$
$$+ \overline{B}\Delta W_i \Delta W_i^T \overline{B}^T \overline{B}\Delta W_j \Delta W_j^T \overline{B}^T \Delta_i^{-1}\Delta_j^{-1} - 2\left(AA^T\overline{B}W_N W_N^T \overline{B}^T \Delta_i t_N{}^{-1}\right.$$
$$+ \left(AA^T\right)^2 \Delta_i t_N + \overline{B}\Delta W_i \Delta W_i^T \overline{B}^T AA^T t_N \Delta_i{}^{-1} + A\Delta W_i A^T A W_N^T \overline{B}^T$$
$$+ A\Delta W_i^T B^T A W_N^T B^T + A\Delta W_i^T B^T B W_N A^T + \overline{B}\Delta W_i A^T \overline{B} W_N A^T$$
$$\left.+\overline{B}\Delta W_i \Delta W_i^T \overline{B}^T \overline{B} W_N W_N^T \overline{B}^T \Delta_i^{-1} t_N^{-1}\right) + \overline{B}W_N W_N^T \overline{B}^T \overline{B} W_N W_N^T \overline{B}^T t_N{}^{-2}$$
$$+ \left(AA^T\right)^2 t_N^2 + \overline{B}W_N W_N^T \overline{B}^T AA^T + AA^T \overline{B} W_N W_N^T \overline{B}^T + A W_N^T \overline{B}^T A W_N^T \overline{B}^T$$
$$\left.+ A W_N^T \overline{B}^T B W_N^T A t_N + \overline{B} W_N AA^T W_N^T \overline{B}^T + \overline{B} W_N A^T \overline{B} W_N A^T\right).$$

Using the formulas from Lemma 3, we calculate that a part of similar terms are equal to zero. For example:

$$E\left(\sum_{i,j=1}^{N}\overline{B}\overline{B}^T AA^T \Delta_j - 2\sum_{i=1}^{N}\overline{B}\overline{B}^T AA^T t_N + \overline{B}\overline{B}^T AA^T t_N\right.$$
$$\left.+ \sum_{i,j=1}^{N} AA^T B\Delta_i - 2\sum_{i=1}^{N} AA^T \overline{B} W_N W_N^T \overline{B}^T \Delta_i t_N^{-1} + AA^T \overline{B}\overline{B}^T t_N\right)$$
$$= BAA^T E\left(Nt_N - 2Nt_N + t_N + Nt_N - 2t_N + t_N\right) = 0.$$

Deleting these terms, we get:

$$E\left(\hat{B}_N^2 - B^2\right)$$
$$= (N-1)^{-2} E\left(\sum_{i=1}^{N}\overline{B}^T \Delta W_i \Delta W_i^T \overline{B}^T \Delta_i{}^{-1}\sum_{j=1}^{N}\overline{B}^T \Delta W_j \Delta W_j^T \overline{B}^T \Delta_j{}^{-1}\right.$$

$$+\sum_{i=1}^{N} B^2 \Delta W_i \Delta W_i^T \Delta W_i \Delta W_i^T \Delta_i^{-2}$$

$$-\sum_{i=1}^{N} \overline{B} \Delta W_i \Delta W_i^T B \sum_{j=1}^{N} \Delta W_j \sum_{l=1}^{N} \Delta W_l \overline{B}^T \Delta_i^{-1} t_N^{-1}$$

$$\left. + \frac{B^2 (k+1) t_N^2}{t_N^2} - (N-1)^2 B^2 \right)$$

$$= (N-1)^{-2} \left(\sum_{i=1}^{N} \overline{B}^T E \Delta W_i \Delta W_i^T \overline{B}^T \Delta_i{}^{-1} \sum_{j=1}^{N} \overline{B}^T E \Delta W_j \Delta W_j^T \overline{B}^T \Delta_j{}^{-1} \right.$$

$$+\sum_{i=1}^{N} B^2 E \Delta W_i \Delta W_i^T \Delta W_i \Delta W_i^T \Delta_i^{-2}$$

$$-E\left(\sum_{i=1}^{N} \overline{B} \Delta W_i \Delta W_i^T B \sum_{j=1}^{N} \Delta W_j \sum_{l=1}^{N} \Delta W_l \overline{B}^T \Delta_i^{-1} t_N^{-1}\right)$$

$$\left. + B^2 (k+1) - (N-1)^2 B^2 \right)$$

$$= (N-1)^{-2} \left(B^2 \left(N^2 - N\right) + N(k+2)B^2 - 2(k+2)B^2 \sum_{i=1}^{N} \Delta_i t_N{}^{-1} \right.$$

$$\left. - 2B^2 \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \Delta_j t_N{}^{-1} + B^2 - (N-1)^2 B^2 \right).$$

The means of some terms we calculate by the following reasoning. Let us denote

$$F(i,j,l) = E\left(\sum_{i=1}^{N} \overline{B} \Delta W_i \Delta W_i^T B \sum_{j=1}^{N} \Delta W_j \sum_{l=1}^{N} \Delta W_l \overline{B}^T \Delta_i t_N{}^{-1}\right).$$

Then $F(i,j,l) = 0$ in all cases when $i \neq j \neq l \neq i$.

If $j = l$, we obtain

$$F(i,j,j) = B^2 \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} E\left(\frac{\Delta W_i \Delta W_i^T}{\Delta_i}\right) E\left(\frac{\Delta W_j \Delta W_j^T}{t_N}\right) + F(i,i,i)$$

$$= B^2 \sum_{i=1}^{N} I_k \left(\frac{\Delta_i}{\Delta_i}\right) I_k \sum_{j=1, j \neq i}^{N} E\left(\frac{\Delta_i}{t_N}\right) = B^2(N-1) + F(i,i,i),$$

where

$$F(i,i,i) = \sum_{i=1}^{N} B^2 E\left(\frac{\Delta W_i \Delta W_i^T \Delta W_i \Delta W_i^T}{\Delta_i t_N}\right) = B^2(k+2)\sum_{i=1}^{N} E\left(\frac{\Delta_i^2}{t_N \Delta_i}\right)$$

$$= B^2(k+2).$$

Hence

$$E\left(\sum_{i=1}^{N} \overline{B}\Delta W_i \Delta W_i^T B \sum_{j=1}^{N} \Delta W_j \sum_{l=1}^{N} \Delta W_l \overline{B}^T \Delta_i t_N{}^{-1}\right) = B^2(N+k+1).$$

In the same way we calculate the means of all terms.

Finally, we obtain

$$E\left(\hat{B}_N^2 - B^2\right) = \frac{k+1}{N-1} B^2.$$

Hence $\hat{B}_N$ is the consistent estimator for the unknown matrix B.

The estimator $\hat{B}_N$ can be represented as follows:

$$\hat{B}_N = (N-1)^{-1}\left(\sum_{i=1}^{N} (\Delta X_i - A\Delta_i)(\Delta X_i - A\Delta_i)^T \Delta_i^{-1}\right.$$

$$\left. - (X_N - At_N)(X_N - At_N)^T t_N^{-1}\right) = (N-1)^{-1} \sum_{i,j=1}^{N} \alpha_{ij} Y_i Y_j^T,$$

where

$$Y_i = (X_i - A\Delta_i)\Delta_i^{-1/2} \sim N(0, B),$$

$$\alpha_{ii} = 1 - \Delta_i \left(\sum_{j=1}^{N} \Delta_j\right)^{-1}, \qquad \alpha_{ij} = \sqrt{\Delta_i \Delta_j}\left(\sum_{j=1}^{N} \Delta_j\right)^{-1}.$$

The random variables Y_i and Y_j are independent and

$$\sum_{i=1}^{N} \alpha_{ii} = N-1, \qquad \sum_{i,j=1}^{N} \alpha_{ij}^2 = N-1.$$

There exists an orthogonal transformation $Y = CZ$ such that

$$\hat{B}_N = (N-1)^{-1} \sum_{i=1}^{N} Z_i Z_i^T,$$

which is k-dimensional Wishard distribution with $N-1$ degrees of freedom.

4. COMMENTS

It is interesting to underline the next facts.

At first, the estimator $\hat{A}_N$ depends only on the last observation, the same as the continuous time sampling, and in the case when determined moments of observation are used. It is interesting to compare the estimators given by different sampling schemes. For example, the point process t_i, $i = 1, \dots, N$, can be Poisson's, geometric, uniform. Results of this kind can be found in [10] and [11].

The second fact is that the variance of $\hat{B}_N$ is independent of the distribution of the random point process $t_1, t_2, \dots, t_N, \dots$ and tends to zero as $O(N^{-1})$ by $N \to \infty$. The proved formula (7) for $k = 1$ is given in [8], i.e. the obtained results generalize the one-dimensional case.

The third fact is that the used sampling scheme is natural. We add the $(N+1)$-th observation to the first N observations and do not need new $N+1$ observations. We established good properties of the estimations without the condition $\max\limits_{1 \le i \le N} \Delta_i \to 0$ when $N \to \infty$, as in the other sampling schemes.

REFERENCES

1. Gichman, I., A. Skorochod. Stochastic Differential Equations. Naukova Dumka, Kiev, 1982.
2. Liptzer, R., A. Shiryayev. Statistics of Random Processes. Nauka, Moskow, 1974.
3. Basawa, B., B. L. M. Prakassa Rao. Statistical Inference for Stochastic Processes. Academic Press, London, 1980.
4. Beutler, P. Alias Free Randomly Time Sampling of Stochastic Processes. IEEE Trans. Inform. Theory, IT – 16, 1970, 147–154.
5. Genon-Catalot, J. Jacod. Estimation of the diffusion Coefficient for Diffusion Processes: Random Sampling. *Scand. J. of Stat.*, **21**, 2, 1994, 193–221.
6. Kessler, M. Estimation of an Ergodic Diffusion from Discrete Observations. *Scand. J. of Stat.*, **24**, 2, 1997, 211–230.
7. Le Breton, A. Estimation d'un Systemme Bilineare Autonome. *Rapport Rech.*, **166**, Univ. Grenoble, 1979.
8. Stoyanov, J., D. Vladeva. Estimation of Unknown Parameters of Continuous Time Stochastic Process by Observations at Random Moments. *Compt. rend. Bulg. Acad. Sci.*, **35**, 2, 1982, 153–156.
9. Giri, N. Multivariable Statistical Inference. Acad. Press, **4**, 1977.
10. Vladeva, D. Estimation of Unknown Parameters in a Diffusion — Geometric Model. In: *Math. and Educ. in Math.*, Proc. of 13-th Spring Conf., Sunny Beach, 1984, 265–271.
11. Vladeva, D. Parameter Estimation of k-Dimensional Wiener Process by Random Observations. In: *Conf. of Appl. Math.*, Tech. Univ. of Sofia, Sozopol, 1999, 154–156.
12. Vladeva, D. Maximum Likelihood Estimation of k-Dimensional Diffusion-Discrete Model. *Mathematica Balkanica*, **14**, 2000 (to appear).

Received February 28, 2000
Revised May 5, 2000

E-mail: `vladeva@mail.vvtu.bg`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

EXTENSION OF THE DUHAMEL PRINCIPLE FOR THE HEAT EQUATION WITH DEZIN'S INITIAL CONDITION

GEORGI I. CHOBANOV and IVAN H. DIMOVSKI

The classical Duhamel principle for the heat equation is extended to the case when the initial condition $u(x,0) = f(x)$ is replaced by the nonlocal A. Dezin's condition $\mu u(0) - u(T) = f(x)$, $\mu \neq 1$. To this end three types of operational calculi are developed: 1) operational calculus for $\frac{d}{dt}$ with the Dezin's functional, 2) operational calculus for $\frac{d^2}{dx^2}$ in a segment $[0,a]$ with boundary conditions $u(0) = 0$ and $u(a) = 0$, and 3) a combined operational calculus for functions $u(x,t)$ in $C(\Delta)$, $\Delta = [0,a] \times [0,T]$.

Keywords: convolution algebra, multiplier, operational calculus, divisor of zero, commutant, Duhamel principle

MSC 2000: 44A40

1. INTRODUCTION

In [3] a general operational calculus for $\frac{d}{dt}$ with arbitrary boundary value functional Φ is developed. Following the pattern of Mikusinski's operational calculus [6] in the space $C[0,1]$, the convolution

$$(f * g)(t) = \Phi_\tau \left\{ \int_\tau^t f(t + \tau - \sigma) g(\sigma)\, d\sigma \right\} \tag{1}$$

instead of Duhamel's convolution

$$(f * g)(t) = \int_0^t f(t - \tau) g(\tau)\, d\tau \tag{2}$$

is used. For the details connected with the convolution (1) found in 1974 ([1]) by one of the authors, see [2] and [3].

In order Mikusinski's scheme to work, only the restriction $\Phi\{1\} \neq 0$ is needed. Without a loss of generality then we may assume

$$\Phi\{1\} = 1.$$

In the case $\Phi\{1\} = 0$ the Mikusinski's scheme is also applicable, but with some modifications.

In [4] and [5] A. A. Dezin considered non-local boundary value problems for the differentiation operator with boundary value condition of the form

$$\mu y(0) - y(T) = 0 \tag{3}$$

with $\mu \neq 1$. For unessential technical simplifications, in the sequel we assume that μ is real. The case of a complex μ can be treated in almost the same way. Instead of the functional $\mu f(0) - f(T)$ we may take the normed functional

$$\Phi\{f\} = \frac{1}{\mu - 1}\,[\mu f(0) - f(T)]$$

and then (1) takes the form

$$(f * g)(t) = \frac{1}{\mu - 1}\left[\mu \int_0^t f(t-\tau)g(\tau)\,d\tau + \int_t^T f(T+t-\tau)g(\tau)\,d\tau\right]. \tag{4}$$

According to [2] operation (4) is a convolution of the right inverse operator

$$lf(t) = \int_0^t f(\tau)\,d\tau + \frac{1}{\mu - 1}\int_0^T f(\tau)\,d\tau \tag{5}$$

of $\dfrac{d}{dt}$ in $C[0,T]$. This means that $f * g$ is a bilinear, commutative and associative operation, such that

$$l(f * g) = (lf) * g.$$

Since $lf(t)$ is determined as the solution of the boundary value problem

$$y' = f, \qquad \mu y(0) - y(T) = 0,$$

then from the representation

$$lf(t) = \{1\} * f,$$

where $\{1\}$ is the constant function 1, it follows that (4) is a convolution of l in $C[0,T]$.

2. CHARACTERIZATION OF THE MULTIPLIERS OF THE CONVOLUTION ALGEBRA $(C[0,T],*)$

According to Larsen [8], an operator $M : C[0,T] \to C[0,T]$ is a multiplier of the convolution algebra $(C[0,T],*)$ iff the relation

$$M(f * g) = (Mf) * g$$

holds for all $f, g \in C[0,T]$. In [8] it is shown that each multiplier is a continuous linear operator. From a general result of [2], p. 32, it follows that each multiplier operator M of the convolution algebra $(C[0,T],*)$ has the convolution representation

$$Mf = \frac{d}{dt}(m * f), \tag{6}$$

where $m(t) \stackrel{\text{def}}{=} M\{1\}$.

In order to characterize the multipliers it remains to specify the representation functions m in (6).

Theorem 1. *A linear operator $M : C[0,T] \to C[0,T]$ is a multiplier of the convolution algebra $(C[0,T],*)$ iff it has the representation*

$$Mf(t) = \frac{d}{dt}\left\{\frac{\mu}{\mu-1}\int_0^t m(t-\tau)f(\tau)\,d\tau + \frac{1}{\mu-1}\int_t^T m(T+t-\tau)f(\tau)\,d\tau\right\}, \tag{7}$$

where m is a continuous function with bounded variation in $[0,T]$.

Remark 1. If $m \in C \cap BV$, then (7) can be written in the form

$$Mf(t) = \frac{d}{dt}\frac{\mu}{\mu-1}\int_0^t m(t-\tau)\,dm(\tau) + \frac{1}{\mu-1}\int_t^T m(T+t-\tau)\,dm(\tau). \tag{8}$$

Proof. From a more general result in [2], p. 32, it follows that each multiplier of $(C[0,T],*)$ has the form (7), where $m(t) \stackrel{\text{def}}{=} M\{1\}$ is a continuous function. It remains only to prove that m is a function with bounded variation in $[0,T]$.

To this end let us fix t $(0 < t \le T)$ and consider $(Mf)(t)$ as a linear functional on $C[0,T]$. According to the F. Riesz representation theorem $(Mf)(t)$ has the form

$$(Mf)(t) = \int_0^T f(\tau)\,d\alpha_t(\tau), \tag{9}$$

where $\alpha_t(\tau)$ is a function with bounded variation in $[0,T]$ depending on t as a parameter. For the sake of uniqueness we may assume that $\alpha_t(\tau)$ is continuous from the left. It would be possible to accomplish the differentiation in (6) termwise provided $m \in C \cap BV$, but we can assume only $m \in C$, which is not enough to ensure the differentiability. Therefore we apply the operator l to (7) and obtain

$$lMf = m * f = \frac{\mu}{\mu-1}\int_0^t m(t-\tau)f(\tau)\,d\tau + \frac{1}{\mu-1}\int_t^T m(T+t-\tau)f(\tau)\,d\tau \tag{10}$$

since for $g \in C^1[0,T]$ we have

$$lg' = g - \frac{\mu g(0) - g(T)}{\mu - 1}$$

and the function $g = m * f$ satisfies the boundary value condition $\mu g(0) - g(T) = 0$.

The operators M and l commute since they both are elements of the multipliers algebra and hence

$$lMf = Mlf = m * f. \tag{11}$$

From (9) we get

$$(Mlf)(t) = \int_0^T (lf)(\tau)\,d\alpha_t(\tau) = (lf)(T)\alpha_t(T) - (lf)(0)\alpha_t(0) - \int_0^T f(\tau)\alpha_t(\tau)\,d\tau. \tag{12}$$

Comparing (10), (11) and (12) we obtain the identity

$$\frac{\mu}{\mu-1}\int_0^t m(t-\tau)f(\tau)\,d\tau + \frac{1}{\mu-1}\int_t^T m(T+t-\tau)f(\tau)\,d\tau$$
$$= (lf)(T)\alpha_t(T) - (lf)(0)\alpha_t(0) - \int_0^T f(\tau)\alpha_t(\tau)\,d\tau.$$

Since

$$(lf)(T) = \mu(lf)(0) = -\frac{\mu}{\mu-1}\int_0^T f(\tau)\,d\tau,$$

then the right-hand side takes the form

$$-\int_0^T f(\tau)\left[\alpha_t(\tau) + \frac{\mu\alpha_t(T) - \alpha_t(0)}{\mu-1}\right]d\tau.$$

If τ is a point of continuity of α_t, then the following two functional identities should hold:

$$-\alpha_t(\tau) - \frac{\mu\alpha_t(T) - \alpha_t(0)}{\mu-1} = \frac{\mu}{\mu-1}m(t-\tau), \qquad 0 \le \tau \le t, \tag{13}$$

and

$$-\alpha_t(\tau) - \frac{\mu\alpha_t(T) - \alpha_t(0)}{\mu-1} = \frac{1}{\mu-1}m(T+t-\tau), \qquad t \le \tau \le T. \tag{14}$$

From (13) and (14) it follows that $m \in BV$. Moreover, from (13) and (14) it follows that m satisfies the boundary value condition $\mu m(0) - m(T) = 0$. Indeed, if we take $\tau = t$ in (13) and (14), we get $\mu m(0) = m(T)$.

If $m \in C \cap BV$, then the derivative $\frac{d}{dt}(m * f)$ exists as a function from $C[0,T]$ and hence the linear operator

$$Mf = \frac{d}{dt}(m * f)$$

is well defined in $C[0,T]$. Obviously, it is a multiplier of the convolution algebra $(C[0,T],*)$.

Since the operator l has a cyclic element — the constant function $\{1\}$, then the multipliers ring of convolution (4) coincides with the commutant of l (see [2], p. 33).

Thus we obtained a complete characterization of the linear operators $M : C[0,T] \to C[0,T]$, which commute with the integration operator l. This explicit characterization can be considered as the "solution" of the non-local spectral problem considered.

In abstract setting the aim of any operational calculus for the operator l reduces to characterizing the class of operators commuting with l. But since the commutant of l coincides with the multipliers ring of the convolution algebra $\big(C[0,T],*\big)$, in this abstract setting the spectral problem obtains its solution by means of Theorem 1.

Usually, in the general spectral theory (see [9], pp. 287–296) only analytic functions of a given operator are considered.

Here we prefer to develop a direct algebraic operational calculus for the operator l, following the multiplier quotients scheme instead of Mikusinski's approach.

3. OPERATIONAL CALCULUS BY MULTIPLIERS QUOTIENTS

Here the basic elements of an operational calculus for the integration operator (5) will be developed. One can follow either the Mikusinski's scheme or the multipliers quotients scheme proposed in [2].

The basic multiplier is the Dezin's integration operator $lf = \{1\} * f$. This multiplier is the convolution operator $l = \{1\}*$. Let $\mathcal{M}$ be the ring of the multipliers of the convolution algebra $\big(C[0,T],*\big)$ and $\mathcal{N}$ be the multiplicative set of non-zero non-devisors of 0 of this algebra.

Let us denote by $\mathcal{R}$ the quotient ring of $\mathcal{M}$ with respect to $\mathcal{N}$, i.e. $\mathcal{R} = \mathcal{N}^{-1}\mathcal{M}$ (see [7], Ch. 2, Sec. 3). The elements of the ring $\mathcal{R}$ are quotients of the form

$$m = \frac{P}{Q}, \qquad \text{where } P \in \mathcal{M} \text{ and } Q \in \mathcal{N}.$$

We should always bear in mind the equivalence

$$\frac{P}{Q} = \frac{R}{S} \quad \Longleftrightarrow \quad PS = RQ.$$

If $c \in \mathbb{C}$, we will use the same letter for the numerical multiplier $c\{f(t)\} = \{cf(t)\}$. By 1 we denote the unit of $\mathcal{R}$, which is different from the convolution multiplier $\{1\}* = l$. Then the algebraic inverse element of l will be denoted by

$$s = \frac{1}{l}.$$

Theorem 2. *If $f \in C^1[0,T]$, then the relation*

$$y' = sy - \frac{\mu y(0) - y(T)}{\mu - 1} \tag{15}$$

holds, where the term $\frac{\mu y(0) - y(T)}{\mu - 1}$ *is viewed not as a constant function, but as a numerical multiplier, and* y *and* y' *are the convolution multipliers* $\{y\}*$ *and* $\{y'\}*$, *respectively.*

Proof. It is easy to verify the identity

$$l\{y'(t)\} = y(t) - y(0) + \frac{1}{\mu - 1}\{y(T) - y(0)\} = y(t) - \frac{\mu y(0) - y(T)}{\mu - 1}\,.$$

If we express this equality as an identity of multipliers, it takes the form

$$ly' = y - \frac{\mu y(0) - y(T)}{\mu - 1} l\,. \tag{16}$$

Multiplying both sides by s, we obtain (15).

Remark 2. We will refer to (15) as the *basic formula of the Dezin's operational calculus.*

Theorem 3. *If* $\lambda \in \mathbb{C}$ *and* $\lambda \notin \frac{1}{T}(\ln|\mu| + 2m\pi i)$, $m \in \mathbb{Z}$, *then*

$$\frac{1}{s - \lambda} = \left\{\frac{e^{\lambda t}(\mu - 1)}{\mu - e^{\lambda T}}\right\}. \tag{17}$$

Proof. Using (15), we obtain

$$(s - \lambda)\{e^{\lambda t}\} = s\{e^{\lambda t}\} - \lambda\{e^{\lambda t}\} = \lambda\{e^{\lambda t}\} + \frac{\mu - e^{\lambda T}}{\mu - 1} - \lambda\{e^{\lambda t}\}$$

and (17) is obvious.

4. OPERATIONAL CALCULUS BY TRANSFORM APPROACH

An alternative approach is based on a finite integral transform associated with the Dezin's condition. (For a transform approach for a more general boundary value condition, see Dimovski [3].)

This finite integral transform can be defined, using the resolvent operator

$$\mathcal{R}_\lambda f = \left\{\frac{e^{\lambda t}(\mu - 1)}{\mu - e^{\lambda T}}\right\} * f. \tag{18}$$

Then the F. Riesz projector is defined as

$$\mathcal{P}_m\{f\} = -\frac{1}{2\pi i}\int\limits_{\Gamma_m} \mathcal{R}_\lambda f \, d\lambda, \tag{19}$$

where Γ_m is a small contour around the zero $\lambda_m = \dfrac{1}{T}(\ln|\mu| + 2m\pi i)$, $m \in \mathbb{Z}$, of

$$E(\lambda) = \frac{1}{\mu - 1}(\mu - e^{\lambda T}).$$

From (18) it follows that

$$\mathcal{P}_m\{f\} = \{e_m(t)\} * f,$$

where

$$e_m(t) = -\frac{1}{2\pi i}\int\limits_{\Gamma_m} \frac{e^{\lambda t}(\mu - 1)}{\mu - e^{\lambda T}}\, d\lambda = \frac{\mu - 1}{\mu T} e^{\lambda_m t}.$$

It is easy to verify the idempotency property

$$e_m^{*^2} = e_m * e_m = e_m.$$

It corresponds to the fact that $\mathcal{P}_m$ is a projector operator.

Using (4), it is easy to find that

$$\mathcal{P}_m\{f\} = \left(\frac{\mu}{\mu - 1}\int_0^T e^{-\lambda_m \tau} f(\tau)\, d\tau\right) e_m(t).$$

The coefficient of $e_m(t)$ is the corresponding finite Fourier transform

$$\mathcal{F}_m\{f\} = \frac{\mu}{\mu - 1}\int_0^T e^{-\lambda_m \tau} f(\tau)\, d\tau, \qquad m \in \mathbb{Z}. \tag{20}$$

This transform could be used as an alternative approach to the operational calculus we considered by a direct approach. In the following theorem we summarize the basic operational properties of the finite integral transform (20) for arbitrary $m \in \mathbb{Z}$.

Theorem 4. *For arbitrary $m \in \mathbb{Z}$ the following equalities hold:*

(i)
$$\mathcal{F}_m\{1\} = \frac{1}{\lambda_m},$$

(ii)
$$\mathcal{F}_m\{lf\} = \frac{1}{\lambda_m}\mathcal{F}_m\{f\},$$

(iii)
$$\mathcal{F}_m\{f'\} = \lambda_m \mathcal{F}_m\{f\} - \frac{\mu f(0) - f(T)}{\mu - 1},$$

(iv) $$\mathcal{F}_m\{f * g\} = \mathcal{F}_m\{f\}\mathcal{F}_m\{g\},$$

and the inversion formula

(v) $$f(t) = \sum_{-\infty}^{\infty} \mathcal{F}_m\{f\}\frac{\mu - 1}{\mu T}e^{\lambda_m t} = \frac{1}{T}\sum_{-\infty}^{\infty}\int_0^T e^{-\lambda_m \tau}f(\tau)\,d\tau e^{\lambda_m t},$$

when the series in the right-hand side converges uniformly.

Proof. (i) and (ii) are obvious.

(iii). From

$$lf' = f - \frac{\mu f(0) - f(T)}{\mu - 1}$$

and (i) it follows

$$\frac{1}{\lambda_m}\mathcal{F}_m\{lf'\} = \mathcal{F}_m\{f\} - \frac{\mu f(0) - f(T)}{\mu - 1}\frac{1}{\lambda_m}.$$

(iv) follows immediately from the representation

$$\mathcal{P}_m\{f\} = \mathcal{F}_m\{f\}e_m(t)$$

of the m-th Riesz projector (see Section 4).

(v) follows from the uniqueness theorem proven in [11], pp. 255–271.

Remark 3. Formula (iii) corresponds to the basic formula (15) and can be used in almost the same way.

5. OPERATIONAL CALCULUS FOR $\frac{D^2}{DX^2}$ WITH BOUNDARY VALUE CONDITIONS $U(0) = 0$, $U(A) = 0$

Following the multipliers quotients approach in [2], a survey of the operational calculus for the simplest boundary value problem for the second order differential operator $D = \frac{d^2}{dx^2}$ in $C[0, a]$ will be made. For more details one may consult our recent paper [10].

The starting point of this operational calculus is the operator $L_{-\lambda^2}$. For $f \in C[0, a]$ the function $y = L_{-\lambda^2}f$ is defined as the solution of the boundary value problem

$$\begin{aligned} &y'' + \lambda^2 y = f(x),\\ &y(0) = y(a) = 0. \end{aligned} \tag{21}$$

It is easy to obtain the explicit expression

$$L_{-\lambda^2}f(x) = \frac{1}{\lambda}\int_0^x \sin\lambda(x - \xi)f(\xi)\,d\xi - \frac{\sin\lambda x}{\lambda\sin\lambda a}\int_0^x \sin\lambda(a - \xi)f(\xi)\,d\xi. \tag{22}$$

The operation

$$(f * g)(x) = -\frac{1}{2a}\int_0^a \left[\int_x^\xi f(\xi + x - \eta)g(\eta)\,d\eta \right. \qquad (23)$$

$$\left. - \int_{-x}^{\xi} f(|\xi - x - \eta|)g(|\eta|)\mathrm{sgn}\left((\xi - x - \eta)\eta\right)d\eta\right] d\xi$$

is a convolution of $L_{-\lambda^2}$ such that

$$L_{-\lambda^2} f(x) = \left\{\frac{a\sin\lambda x}{\sin\lambda a}\right\} * f(x). \qquad (24)$$

The special case $\lambda = 0$ is used as the basic operator of the corresponding operational calculus. Denoting $L = L_0$, we have

$$Lf(x) = \{x\} * f(x), \qquad (25)$$

i.e. L may be considered as the convolution operator $\{x\}*$. For simplicity we will write $L = \{x\}$. Also, if $f \in C[0,a]$, then by $\{f\}$ we will denote the convolution multiplier operator $\{f\}*$.

Let $\mathcal{M}_x$ be the ring of the multipliers of the convolution algebra $\left(C[0,T], *\right)$ and let by $\mathcal{N}_x$ we denote the multiplicative set of the non-divisors of 0 in $\mathcal{M}_x\backslash\{0\}$.

Further we consider the ring of the multipliers quotients $\mathcal{R}_x = \mathcal{N}_x^{-1}\mathcal{M}_x$ of the form P/Q with $P \in \mathcal{M}_x$, $Q \in \mathcal{N}_x$.

Basic is the role of the multipliers quotient

$$S = \frac{1}{L}, \qquad (26)$$

where by 1 the identity operator in $\mathcal{M}_x$ is denoted.

The basic formula of the operational calculus under development can be obtained from the identity

$$Lf'' = f - \left(1 - \frac{x}{a}\right) f(0) + \frac{x}{a} f(a).$$

Writing it as an identity of multipliers operators, it takes the form

$$Lf'' = f - \left\{1 - \frac{x}{a}\right\} f(0) + \frac{1}{a} f(a)L. \qquad (27)$$

Multiplying by S, we obtain

$$f'' = Sf - S\left\{1 - \frac{x}{a}\right\} f(0) - \frac{1}{a} f(a), \qquad (28)$$

where the numbers $f(0)$ and $\frac{1}{a}f(a)$ are to be considered as "numerical operators", i.e. as numerical multipliers in $\left(C[0,a], *\right)$. Using (28), we can find that

$$\frac{1}{S + \lambda^2} = \left\{\frac{a\sin\lambda x}{\sin\lambda a}\right\} \qquad (29)$$

for $\lambda \neq \frac{n\pi}{a}$, $n \in \mathbb{N}$.

Proof. It is easy to see that

$$\frac{1}{S+\lambda^2} = L_{-\lambda^2}.$$

Indeed, if $y = L_{-\lambda^2} f$, then from (28) it follows $Sy = y'' = f - \lambda^2 y$. Now the assertion follows from (24).

The direct approach can be duplicated by the finite sine-transform (see [12]).

We will use a variant of the sine-transform having (23) as its convolution. It is slightly different from the finite Fourier sine-transform introduced by Churchill in [13], p. 349.

Definition 1. The transform $\mathcal{F}_n^s : C[0,a] \to \mathbb{C}^{\mathbb{N}}$ is defined by

$$\mathcal{F}_n^s\{f\} = \frac{(-1)^n}{n\pi} \int_0^a f(\xi) \sin \frac{n\pi}{a} \xi \, d\xi, \tag{30}$$

$n = 1, 2, \ldots$

Theorem 5. *The basic properties of the sine-transform are:*

(i)
$$\mathcal{F}_n^s\{x\} = -\left(\frac{a}{n\pi}\right)^2;$$

(ii)
$$\mathcal{F}_n^s\{Lf\} = -\left(\frac{a}{n\pi}\right)^2 \mathcal{F}_n^s\{f\};$$

(iii)
$$\mathcal{F}_n^s\{f''\} = -\left(\frac{n\pi}{a}\right)^2 \mathcal{F}_n^s\{f\} - \frac{1}{a}[f(a) - (-1)^n f(0)];$$

(iv)
$$\mathcal{F}_n^s\{f * g\} = \mathcal{F}_n^s\{f\}\mathcal{F}_n^s\{g\};$$

and the inversion formula

(v)
$$f(x) = \frac{2\pi}{a} \sum_{n=1}^{\infty} (-1)^n n \mathcal{F}_n^s\{f\} \sin \frac{n\pi}{a} x$$

holds when the right-hand side series converges uniformly.

Proof. (i) and (ii) can easily be obtained directly. (iii) follows from (27) using (i) and (ii):

$$\mathcal{F}_n^s\{Lf''\} = \mathcal{F}_n^s\{f\} - f(0)\mathcal{F}_n^s\{1 - \frac{x}{a}\} + \frac{1}{a} f(a)\mathcal{F}_n^s\{x\}$$

or

$$-\left(\frac{a}{n\pi}\right)^2 \mathcal{F}_n^s\{f''\} = \mathcal{F}_n^s\{f\} - f(0)\frac{(-1)^n a}{(n\pi)^2} + f(a)\frac{a}{(n\pi)^2}$$

since

$$\mathcal{F}_n^s\{1-\frac{x}{a}\} = \frac{(-1)^n a}{(n\pi)^2}\cdot$$

(iv) It can easily be verified that

$$\left(\sin\frac{n\pi}{a}x\right) * \left(\sin\frac{n\pi}{a}x\right) = \frac{(-1)^n a}{2n\pi}\sin\frac{n\pi}{a}x.$$

Since

$$\mathcal{F}_n^s\{f\}\sin\frac{n\pi}{a}x = \left(\sin\frac{n\pi}{a}x\right) * f,$$

then

$$\mathcal{F}_n^s\{f*g\}\sin\frac{n\pi}{a}x = \left(\sin\frac{n\pi}{a}x\right) * (f*g) = \mathcal{F}_n^s\{f\}\left(\sin\frac{n\pi}{a}x * g\right)$$
$$= \mathcal{F}_n^s\{f\}\mathcal{F}_n^s\{g\}\sin\frac{n\pi}{a}x.$$

(v) See [12], Ch. 3.

6. A TWO-VARIATE OPERATIONAL CALCULUS FOR THE HEAT EQUATION WITH DEZIN'S BOUNDARY VALUE CONDITION

Our first aim is to develop an operational approach to the following non-local boundary value problem for the heat equation in the rectangle $\Delta = [0,a]\times[0,T]$:

$$\begin{aligned} &u_t = u_{xx} + F(x,t), \quad (x,t)\in\Delta,\\ &\mu u(x,0) - u(x,T) = f(x), \quad 0\le x\le a,\\ &u(0,t) = \varphi(t),\ u(a,t) = \psi(t), \quad 0\le t\le T,\end{aligned} \tag{31}$$

where $\mu \neq 1$ is a real parameter.

To this end we need a two-variate operational calculus for functions $u(x,t)$ of a space variable and a time variable. We will follow the pattern from [10].

We have to find an inner operation for functions $f\in C(\Delta)$, which is a convolution both for the operator

$$l\{u(x,t)\} = \int_0^t u(x,\tau)\,d\tau + \frac{1}{\mu-1}\int_0^T u(x,\tau)\,d\tau \tag{32}$$

and the operator

$$L\{u(x,t)\} = \int_0^x (x-\xi)u(\xi,t)\,d\xi - \frac{x}{a}\int_0^a (a-\xi)u(\xi,t)\,d\xi.$$

According to (23) the operation

$$\begin{aligned}(f \overset{x}{*} g)(x) \;=\; &-\frac{1}{2a}\int_0^a \left[\int_x^\xi f(\xi+x-\eta)g(\eta)\,d\eta\right.\\ &\left.-\int_{-x}^\xi f(|\xi-x-\eta|)g(|\eta|)\operatorname{sgn}\left((\xi-x-\eta)\eta\right)d\eta\right]d\xi\end{aligned} \tag{33}$$

is a convolution for L in $C([0,a])$.

Now we are to combine (33) with (4) in order to obtain a two-variate convolution, such that l and L to be multipliers of the corresponding convolution algebra.

Theorem 6. *Let $u, v \in C(\Delta)$. Then the operation*

$$(f \overset{(x,t)}{*} g)(x,t) = -\frac{1}{2a(\mu-1)} \int_0^a \left[\int_0^t \mu \int_x^\xi f(x+\xi-\eta, t-\tau) g(\eta,\tau)\, d\eta d\tau \right.$$
$$+ \int_t^T \int_x^\xi f(x+\xi-\eta, T+t-\tau) g(\eta,\tau)\, d\eta d\tau \tag{34}$$
$$- \int_0^t \mu \int_{-x}^\xi f(|\xi - x - \eta|, t-\tau) g(|\eta|,\tau) \operatorname{sgn}[(\xi - x - \eta)\eta]\, d\eta d\tau$$
$$\left. - \int_t^T \int_{-x}^\xi f(|\xi - x - \eta|, T+t-\tau) g(|\eta|,\tau) \operatorname{sgn}[(\xi - x - \eta)\eta]\, d\eta d\tau \right] d\xi$$

is a bilinear, commutative and associative operation in $C(\Delta)$ such that the operators

$$l\{f\} = \{1\} \overset{t}{*} f$$

and

$$L\{f\} = \{x\} \overset{x}{*} f$$

are multipliers of the convolution algebra $\left(C(\Delta), \overset{(x,t)}{*}\right)$.

For a proof one can follow the lines of the corresponding proof in [10]. Since the bilinearity and the commutativity are almost obvious, only the associativity needs to be proved. We verify it for product functions $F(x,t) = f(x)\varphi(t)$ and $G(x,t) = g(x)\psi(t)$ using the identity

$$(F \overset{(x,t)}{*} G)(x,t) = (f \overset{x}{*} g)(x)(\varphi \overset{t}{*} \psi)(t),$$

followed by an approximation argument.

Let us denote by $\mathcal{R}$ the multipliers quotient ring of the convolution algebra $\left(C(\Delta), \overset{(x,t)}{*}\right)$. In $\mathcal{R}$ the one-variate identity

$$\{\varphi'(t)\} = s\{\varphi(t)\} - \frac{\mu\varphi(0) - \varphi(T)}{\mu - 1}$$

takes the form

$$\left\{ \frac{\partial u(x,t)}{\partial t} \right\} = s\{u(x,t)\} - \left[\frac{\mu u(x,0) - u(x,T)}{\mu - 1} \right]_t, \tag{35}$$

where by $[\cdot]_t$ it is denoted that the expression in the brackets is a numerical operator with respect to t, i.e. $[\cdot]_t = \{\cdot\} \overset{x}{*}$.

The corresponding identity with respect to x has the same form as in [10]:

$$\frac{\partial^2 u}{\partial x^2} = Su - S\left\{1 - \frac{x}{a}\right\}[u(0,t)]_x - \left[\frac{1}{a}u(a,t)\right]_x, \tag{36}$$

where $S = \frac{1}{L}$ and $[\cdot]_x$ is to be considered as a numerical multiplier with respect to x, i.e. $[\cdot]_x = \{\cdot\} \overset{t}{*}$.

For the proof of the uniqueness theorem in the next section we need to characterize the divisors of 0 of $\left(C(\Delta), \overset{(x,t)}{*}\right)$. This is done by means of the following two-variate finite integral transformation.

Definition 2. For $u \in C(\Delta)$ let

$$\mathcal{F}_{m,n}\{u\} = \frac{(-1)^n \mu}{(\mu - 1)n\pi} \int_0^a \int_0^T u(\xi, \tau) e^{-\lambda_m \tau} \sin \frac{n\pi}{a}\xi \, d\xi d\tau$$

be a two-variate finite integral transform corresponding to the two-variate convolution (34).

Theorem 7. *The following properties of $\mathcal{F}_{m,n}$ are satisfied:*

(i) $$\mathcal{F}_{m,n}\{Lu\} = -\left(\frac{a}{n\pi}\right)^2 \mathcal{F}_{m,n}\{u\},$$

(ii) $$\mathcal{F}_{m,n}\{lu\} = \lambda_m \mathcal{F}_{m,n}\{u\},$$

(iii) $$\mathcal{F}_{m,n}\{u\} e^{\lambda_m t} \sin \frac{n\pi}{a} x = \left\{e^{\lambda_m t} \sin \frac{n\pi}{a} x\right\} \overset{(x,t)}{*} u,$$

(iv) $$\mathcal{F}_{m,n}\{u \overset{(x,t)}{*} v\} = \mathcal{F}_{m,n}\{u\}\mathcal{F}_{m,n}\{v\},$$

(v) $$\mathcal{F}_{m,n}\left\{\frac{\partial^2 u}{\partial x^2}\right\} = -\left(\frac{n\pi}{a}\right)^2 \mathcal{F}_{m,n}\{u\} - \frac{1}{a}\mathcal{F}_m\{u(a,t) - (-1)^n u(0,t)\},$$

(vi) $$\mathcal{F}_{m,n}\left\{\frac{\partial u}{\partial t}\right\} = \lambda_m \mathcal{F}_{m,n}\{u\} - \mathcal{F}_n^s\left\{\frac{\mu u(x,0) - u(x,T)}{\mu - 1}\right\}.$$

Proof. Follows immediately from Theorems 4 and 5.

Lemma 1. *A function $u \in C(\Delta)$ is a divisor of 0 iff for some $m \in \mathbb{Z}$, $n \in \mathbb{N}$ we have $\mathcal{F}_{m,n}\{u\} = 0$.*

Proof. Indeed, let for some $m \in \mathbb{Z}$, $n \in \mathbb{N}$ we have $\mathcal{F}_{m,n}\{u\} = 0$. Then

$$\{e^{\lambda_m t} \sin \frac{n\pi}{a} x\} \overset{(x,t)}{*} \{u(x,t)\}$$
$$= \frac{(-1)^n \mu}{(\mu - 1)n\pi} \left[\int_0^a \int_0^T u(\xi, \tau) e^{-\lambda_m \tau} \sin \frac{n\pi}{a} \xi \, d\xi d\tau \right] e^{\lambda_m t} \sin \frac{n\pi}{a} x = 0.$$

Hence u is a divisor of 0.

Let us now conversely assume that u is a divisor of 0 of $\overset{(x,t)}{*}$, i.e. that $u \overset{(x,t)}{*} v = 0$ for some function $v \in C(\Delta)$, $v \neq 0$. If we assume that $\mathcal{F}_{m,n}\{u\} \neq 0$ for all $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then

$$(u \overset{(x,t)}{*} v) \overset{(x,t)}{*} \left\{ e^{\lambda_m t} \sin \frac{n\pi}{a} x \right\} = \frac{(-1)^n \mu}{(\mu - 1)n\pi} \mathcal{F}_{m,n}\{u\} \mathcal{F}_{m,n}\{v\} = 0,$$

whence $\mathcal{F}_{m,n}\{v\} = 0 \ \forall m \in \mathbb{Z}, \forall n \in \mathbb{N}$. This is equivalent to

$$\left(\sin \frac{n\pi}{a} x \right) \overset{x}{*} \left[e^{\lambda_m t} \overset{t}{*} v(x,t) \right] = 0.$$

If we denote the functions in the brackets with $F_m(x,t)$, the last equality implies that

$$F_m(x,t) = 0, \qquad \forall m \in \mathbb{Z}.$$

Fixing now $x \in [0, a]$, we obtain

$$e^{\lambda_m t} \overset{t}{*} v(x,t) = 0,$$

whence $v(x,t) = 0 \ \forall t \in [0, T]$, contrary to the assumptions.

Theorem 8. *$L - l$ is a divisor of* 0 *of the multipliers ring of the convolution algebra* $\left(C(\Delta), \overset{(x,t)}{*}\right)$ *iff* $\frac{a}{\pi} \sqrt{\frac{1}{T} \ln \frac{1}{\mu}} \in \mathbb{N}$.

Proof. If $(L - l)M = 0$ for some $M \neq 0$, then there exists a function $v \in C(\Delta)$, $v \neq 0$, such that $Mv = u \neq 0$ and $(L - l)u = 0$. The fact that $u \neq 0$ implies that for some $m \in \mathbb{Z}$ and $n \in \mathbb{N}$ we have

$$\mathcal{F}_{m,n}\{u\} \neq 0.$$

According to (i) and (ii) of Theorem 6 we have

$$\mathcal{F}_{m,n}\{Lu - lu\} = \left[-\left(\frac{a}{n\pi}\right)^2 - \lambda_m \right] \mathcal{F}_{m,n}\{u\} = 0.$$

This is possible only if

$$\left(\frac{a}{n\pi}\right)^2 + \lambda_m = 0.$$

Since $\lambda_m = \frac{1}{T}(\ln \mu + 2m\pi i)$, this is only possible for $m = 0$, i.e. when

$$\frac{1}{T} \ln \mu = - \left(\frac{a}{n\pi}\right)^2$$

for some $n \in \mathbb{N}$, whence the assertion follows. The converse is obvious.

7. GENERALIZED SOLUTIONS OF THE HEAT EQUATION WITH DEZIN'S INITIAL CONDITION

The common notion of a generalized solution in the sense of distribution theory is unpractical for boundary value problems. This is especially true for nonlocal boundary value problems in finite domains.

In the case of the boundary value problem (31) it is very useful to introduce the notion of a generalized solution in the framework of the algebraic analysis of D. Przeworska-Rolevicz [14].

Let $C^{2,1}(\Delta)$ be the space of functions that are twice continuously differentiable with respect to x and continuously differentiable with respect to t. Let us assume that (31) has a classical solution $u \in C^{2,1}(\Delta)$. Applying the operator Ll to the equation

$$u_t = u_{xx} + F(x,t),$$

we obtain

$$L(lu_t) = l(Lu_{xx}) + LlF(x,t).$$

From (16) we get

$$lu_t = u - \frac{1}{\mu - 1} f(x)$$

and from (27) we have

$$Lu_{xx} = u - \left(1 - \frac{x}{a}\right) \varphi(t) + \frac{x}{a} \psi(t).$$

Hence

$$L\left[u(x,t) - \frac{1}{\mu - 1} f(x)\right] = l\left[u(x,t) - \left(1 - \frac{x}{a}\right)\varphi(t) + \frac{x}{a}\psi(t)\right] + LlF(x,t)$$

or

$$(L - l)u = \frac{1}{\mu - 1} Lf(x) - \left(1 - \frac{x}{a}\right) l\varphi(t) + \frac{x}{a} l\psi(t) + LlF(x,t). \tag{37}$$

Conversely, if a function $u \in C^{2,1}(\Delta)$ satisfies (37), then it is a solution of (31). Indeed, if we apply the differential operator $\frac{\partial^3}{\partial x^2 \partial t}$ to (37), we obtain

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right) u = F(x,t)$$

and hence $u(x,t)$ is a solution of the heat equation.

In order to verify that the initial conditions are satisfied, let us apply $\frac{\partial^2}{\partial x^2}$ only. We obtain

$$u - lu_{xx} = \frac{1}{\mu - 1} f(x) + lF(x,t).$$

Applying to this equality the functional

$$\Psi\{g(t)\} = \mu g(0) - g(T), \tag{38}$$

we get

$$\Psi\{u\} = f(x),$$

i.e. the Dezin's initial conditions since $\Psi\{lg\} = 0$.

In a similar way applying $\frac{\partial}{\partial t}$ to (37) we get

$$Lu_t - u = -\left(1 - \frac{x}{a}\right)\varphi(t) + \frac{x}{a}\psi(t) + LF(x,t).$$

Since $(Lg)(x)$ satisfies the boundary conditions $(Lg)(0) = (Lg)(a) = 0$, then the above equality for $x = 0$ and for $x = a$ gives $u(0,t) = \varphi(t)$ and $u(a,t) = \psi(t)$.

The above considerations justify the next definition.

Definition 3. If $u \in C(\Delta)$ satisfies the integral relation (37), it is said to be a generalized solution of the boundary value problem (31).

Lemma 2. *If $u \in C(\Delta)$ satisfies (37), then $u(0,t) = \varphi(t)$, $u(a,t) = \psi(t)$ and $\mu u(x,0) - u(x,T) = f(x)$.*

Proof. From (37) for $x = 0$ we get

$$-lu(0,t) = -l\varphi(t),$$

whence $u(0,t) = \varphi(t)$. In a similar way, substituting $x = a$ in (37), we obtain $u(a,t) = \psi(t)$.

Applying the functional (38) to (37), we get

$$L\Psi_t\{u(x,t)\} = \frac{1}{\mu - 1} Lf(x)\Psi\{1\}$$

(since $L\Psi_t = \Psi_t L$) or

$$L\{\mu u(x,0) - u(x,T)\} = Lf(x).$$

Now Dezin's initial condition follows applying the operator $\frac{\partial^2}{\partial x^2}$ to both sides of the above equality.

The two-variate operational calculus developed above allows to algebraize completely the boundary value problem (31). We use formulas (35) and (36) to obtain

$$su - \frac{[f(x)]_t}{\mu - 1} = Su - S\left\{1 - \frac{x}{a}\right\}[\varphi(t)]_x - \frac{1}{a}[\psi(t)]_x + \{F(x,t)\}.$$

The solution u exists as an element of the ring $\mathcal{R}$, i.e. as a multipliers quotient and it can be represented in the algebraic form

$$u = \frac{\{F(x,t)\}}{s - S} + \frac{[f(x)]_t}{(\mu - 1)(s - S)} - \frac{S}{s - S}\left\{1 - \frac{x}{a}\right\}[\varphi(t)]_x - \frac{1}{s - S}\frac{1}{a}[\psi(t)]_x. \quad (39)$$

It is valid provided $s - S$ is a non-divisor of 0 in $\mathcal{R}$.

Theorem 9 (for uniqueness). *The element $s - S$ of $\mathcal{R}$ is a non-divisor of 0 provided $\frac{a}{\pi}\sqrt{\frac{1}{T}\ln\frac{1}{\mu}} \notin \mathbb{N}$.*

Proof. Let us assume the contrary. As in [10] it is easy to show that this assumption reduces to the existence of a non-zero function $u \in C(\Delta)$ such that

$$(L - l)u = 0.$$

Then the proof follows from Theorem 7.

Now the formal solution (39) exists provided $\frac{a}{\pi}\sqrt{\frac{1}{T}\ln\frac{1}{\mu}} \notin \mathbb{N}$. In order to interpret it, we introduce

$$\Omega = \frac{1}{sS(s - S)} = \frac{Ll}{s - S}.$$

Since $Ll = \{x\}$, then Ω can be interpreted as a solution to the boundary value problem (31) with $F(x,t) \equiv x$ and $f(x) = 0$, $\varphi(t) = 0$, $\psi(t) = 0$.

If such a solution $u(x,t)$ exists, we can find it by means of the finite sine transform. If $u(x,t)$ is the solution, then

$$\mathcal{F}_n^s\{u_t\} = \mathcal{F}_n^s\left\{\frac{\partial^2 u}{\partial x^2}\right\} + \mathcal{F}_n^s\{x\}.$$

Using (i) and (iii) of Theorem 5 and the boundary conditions for

$$u_n(t) = \mathcal{F}_n^s\{u(x,t)\},$$

we obtain the equations

$$\begin{aligned} &\frac{du_n}{dt} = -\left(\frac{n\pi}{a}\right)^2 u_n(t) - \left(\frac{a}{n\pi}\right)^2, \\ &\mu u_n(0) - u_n(T) = 0 \end{aligned} \quad (40)$$

for $n = 1, 2, \ldots$ For the functions $u_n(t)$ we find the explicit expressions

$$u_n(t) = \frac{(\mu-1)a^4}{(n\pi)^4(\mu - e^{-(\frac{n\pi}{a})^2 T})} e^{-(\frac{n\pi}{a})^2 t} - \frac{a^4}{(n\pi)^4}.$$

Taking into account the inversion formula (v) in Theorem 5, we now define

$$\begin{aligned} \Omega(x,t) &= \frac{2\pi}{a}\sum_{n=1}^{\infty}(-1)^n n u_n(t)\sin\frac{n\pi}{a}x = -2\left(\frac{a}{\pi}\right)^3\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3}\sin\frac{n\pi}{a}x \\ &+ 2\frac{(\mu-1)a^3}{\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3(\mu - e^{-(\frac{n\pi}{a})^2 T})}e^{-(\frac{n\pi}{a})^2 t}\sin\frac{n\pi}{a}x \qquad (41) \\ = -\frac{1}{6}(x^3 - a^2 x) &+ 2\frac{(\mu-1)a^3}{\pi^3}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^3(\mu - e^{-(\frac{n\pi}{a})^2 T})}e^{-(\frac{n\pi}{a})^2 t}\sin\frac{n\pi}{a}x. \end{aligned}$$

It is a matter of a simple verification that $\Omega(x,t)$ satisfies the integral relation

$$L\Omega - l\Omega = Ll\{x\}. \qquad (42)$$

Theorem 10. *Let $F(x,t) \in C^{2,1}(\Delta)$ and let it satisfy the zero initial-boundary conditions of (31). Then the problem (31) has a generalized solution $u(x,t)$, which has the Duhamel type representation*

$$u(x,t) = \Omega \overset{(x,t)}{*} \frac{\partial^3 F}{\partial x^2 \partial t}. \qquad (43)$$

Proof. In order to prove that u is a generalized solution of (31), we are to show that it satisfies the integral relation

$$(L-l)u = LlF.$$

From the assumptions made on f we have

$$Ll\frac{\partial^3 F}{\partial x^2 \partial t} = F.$$

Then

$$Ll(l-l)u = L^2 lu - Ll^2 u = L\left(\Omega \overset{(x,t)}{*} Ll\frac{\partial^3 F}{\partial x^2 \partial t}\right) - l\left(\Omega \overset{(x,t)}{*} Ll\frac{\partial^3 F}{\partial x^2 \partial t}\right)$$

$$= L(\Omega \overset{(x,t)}{*} F) - l(\Omega \overset{(x,t)}{*} F) = (L\Omega - l\Omega) \overset{(x,t)}{*} F.$$

But $L\Omega - l\Omega = Ll\{x\}$ (see (42)) and then

$$Ll(l-l)u = Ll\{x\} \overset{(x,t)}{*} F.$$

Since $\{x\} \overset{(x,t)}{*} F = LlF$, then

$$Ll(L-l)u = L^2 l^2 F.$$

Now "canceling" the term Ll, we get $Lu - lu = LlF$.

From (43) it follows that $u(x,t)$ is continuous on Δ and it satisfies the initial-boundary value conditions $u(0,t) = u(a,t) = 0$ and $\mu u(x,0) - u(x,T) = 0$ according to Lemma 2.

Theorem 11. *Let $f \in C^4[0,a]$ and let $f(0) = f''(0) = f(a) = f''(a) = 0$. Then the function $v \in C^{2,1}(\Delta)$ defined by*

$$v(x,t) = -\frac{1}{\mu-1}\left(\Omega \overset{x}{*} f^{(4)}(x)\right) - \frac{1}{\mu-1} f(x) \tag{44}$$

is a generalized solution of the equation $v_t = v_{xx}$ with $v(0,t) = v(a,t) = 0$, $\mu v(x,0) - v(x,T) = f(x)$.

Proof. It is not difficult to obtain this Duhamel-type representation from

$$v = -\frac{1}{\mu-1}\frac{[f(x)]_t}{s-S},$$

but it is easier to verify directly that (44) satisfies the equation $(L-l)v = Lf$.

Theorem 12. *Let $\psi \in C^2[0,T]$ satisfy $\mu\psi(0) - \psi(T) = 0$ and $\mu\psi'(0) - \psi'(T) = 0$. Then the function $w \in C^{2,1}(\Delta)$, given by*

$$w(x,t) = -\frac{1}{a}\Omega \overset{t}{*} \psi'' + \frac{x}{a}\psi(t), \tag{45}$$

is a generalized solution of the equation $w_t = w_{xx}$ with $w(0,t) = 0$, $w(a,t) = \psi(t)$ and $\mu w(x,0) - w(x,T) = 0$.

Proof. We have to show that w satisfies the integral relation

$$(L-l)w = -\frac{1}{a}x\psi(t).$$

This is a matter of simple calculations.

Remark 4. The case $F \equiv 0$, $\psi(t) \equiv 0$, $f(x) \equiv 0$ is not essentially different from the just considered case. Although the corresponding expression

$$\frac{-S\left\{1 - \frac{x}{a}\right\}[\varphi(t)]_x}{s-S}$$

looks more involved than the expression

$$\frac{-[\psi(t)]_x}{(\mu-1)(s-S)}$$

in the previous case, it can be simplified by introducing the new independent variable $z = a - x$.

Remark 5. In order the generalized solution of problem (31) to be a classical solution, only a slight increase of the smoothness assumptions on F, φ, ψ and f is necessary. It is sufficient to require the corresponding derivatives of the highest order to be not only continuous, but absolutely continuous.

REFERENCES

1. Dimovski, I. H. On an operational calculus for vector-valued functions. *Math. Balkanica*, **4**, 1974, 129–135.
2. Dimovski, I. H. Convolutional Calculus. Kluwer, Dordrecht—N. Y.—London, 1990.
3. Dimovski, I. H. Nonlocal operational calculi. *Proceedings of the Steklov Institute of Mathematics*, **3**, 1995, 53–65.
4. Dezin, A. A. Operators with a first derivative with respect to "time" and non-local boundary value conditions. *Izvestiya Acad. Nauk SSSR, ser. Math.*, **31**, 1967, 61–86.
5. Dezin, A.A. Partial Differential Equations. An Introduction to a General Theory of Linear Boundary Value Problems. Springer Verlag, Berlin—Heidelberg—N. Y., 1987.
6. Mikusinski, J. Operational Calculus. Oxford-Warszawa, 1959.
7. Lang, S. Algebra. Addison—Wesley, Reading (Mass.), 1965.
8. Larsen, R. An Intoduction to the Theory of Multipliers. Springer Verlag, Berlin—Heidelberg—N. Y., 1972.
9. Taylor, A. E. Introduction to Functional Analysis. J. Wiley, N.Y., 1958.
10. Chobanov, G. I., I. H. Dimovski. A two-variate operational calculus for boundary value problems. *Fractional Calculus and Applied Analysis*, **2**, 1999, 591–601.
11. Leont'ev, A. F. Series in Exponents. Nauka, Moscow, 1976. (In Russian.)
12. Sneddon, I. Fourier Transforms. N. Y.—Toronto—London, 1951.
13. Churchill, R. V. Operational Mathematics. 3rd. ed., McGraw-Hill, N. Y., 1972.
14. Przeworska-Rolevicz, D. Algebraic theory of right invertible operators. *Studia Math.*, **48**, 1973, 114–129.

Received March 24, 2000
Revised April 10, 2000

Georgi I. CHOBANOV
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev str., bl. 8
BG-1113 Sofia, BULGARIA
E-mail: chobanov@math.bas.bg

Ivan H. DIMOVSKI
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. G. Bonchev str., bl. 8
BG-1113 Sofia, BULGARIA

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ"

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI"

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

A LOGARITHMIC CLASS OF SEMILINEAR WAVE EQUATIONS

TOSHKO BOEV

We study the global existence, long-time behaviour and blow-up of classical solutions of the equation $\Box u = u \ln^q(1 + u^2)$ in $(3 + 1)$-space-time with arbitrary big initial data. Thus we have a case of a repellent potential energy term in the relevant energetic identity, contrary to the attractive energy case described by the well-known equation $\Box u = -u|u|^{p-1}$. The global existence result for $0 < q \leq 2$ is first established. Then special "counterdecay" (for $0 < q < 2$) and blow-up effects (for $q > 2$) are found, which show that $q = 2$ is a "critical" value. In this way it is answered, in particular, to a question that has arisen already in the pioneering works of Keller and Jörgens on the semilinear wave equation.

Keywords: global classical solutions, exponentially increasing solutions, blow-up solutions

MSC 2000: 35L70, 35B05

1. INTRODUCTION

Consider the semilinear wave equation

$$\Box u = f(u), \tag{1.1}$$

where $\Box \equiv \partial_t^2 - \Delta$, $\Delta = \Delta_x \equiv \sum_{j=1}^n \partial_{x_j}^2$, n is the spatial dimension. For the function $f(u)$ it is assumed that $f(u) = O(|u|^p)$, for $|u| \to +\infty$ or $|u| \to 0$, with a certain parameter $p > 0$.

As it is well-known, there are two critical numbers $p^*(n)$ and $p_0(n)$ that play a prominent role in the theory of Eq. (1.1). They are defined as follows:

$$p^*(n) = \frac{n+2}{n-2}, \quad p_0(n) = z^+, \tag{1.2}$$

where z^+ is the positive root of the equation

$$(n-1)z^2 - (n+1)z - 2 = 0.$$

The significance of the number $p^*(n)$ was revealed by Jörgens [6], by the results of many authors afterwards, see, e.g., [1, 9], and in the more recent papers [3, 10, 12] and the references therein. This number gives an answer to the question for the strongest nonlinearity of $f(u)$, as $|u| \to +\infty$, which admits global (in time) classical solutions to Eq. (1.1) without restrictions on the magnitude of the initial data. In the above cited papers global existence results have been obtained in the case $1 < p < p^*$, for an arbitrary magnitude of data, under the assumptions that the potential energy term

$$U_f[u] \equiv \int_0^u f(z)\,\mathrm{d}z$$

is attractive, i.e. $U_f[u] \leq \text{const}$, $\forall u$, within the main class of functions $f(u)$ of the form

$$f(u) = -u|u|^{p-1}. \tag{1.3}$$

The second critical number $p_0(n)$ shows up for u close to zero, in the quest for existence of global solutions with small enough data. The number $p_0(n)$ has been found by John [5] for $n = 3$, when $p_0 = 1 + \sqrt{2}$, and by Strauss [11] for an arbitrary n. It is noted that after the blow-up result in [5] for the equations $\Box u = \pm u^2$ (see also Lindblad [8]) with $p = 2$ — a number within the subcritical interval $1 < p < 1 + \sqrt{2}$, *and the global existence result in* [5] *for* $p \geq 1 + \sqrt{2}$, *many* authors paid a special attention to the cases of the critical value $p = p_0(n)$ and to the supercritical interval $p > p_0(n)$.

In the present paper we consider the classical solutions of the Cauchy problem

$$\Box u = u \ln^q(1 + u^2), \quad x \in R^3,\ t > 0,\ q > 0, \tag{1.4}$$

$$u(x, 0) = \varphi(x), \quad u_t(x, 0) = \psi(x), \quad x \in R^3, \tag{1.5}$$

where $\varphi \in C^3$, $\psi \in C^2$. (As usual, $C^k = C^k(R^3)$ is the space of the k-time smooth functions.) The main results below shed more light particularly on the small data problem for Eq. (1.1). (For more details see, e.g., [11, 13, 14].) When $|u|$ is small, we recover in Eq. (1.4) a particular case of Eq. (1.1) with $p = 1+2q$. However, in the subcritical interval $q < 1/\sqrt{2}$ $(1 < p < 1 + \sqrt{2})$ the problem (1.4), (1.5) possesses global solutions for arbitrary magnitudes of the data. From the view-point of the interest of many authors to the so-called supercritical interval for Eq. (1.1), let us note the following. For $q > 2$ we have in Eq. (1.4) a whole interval of supercritical powers $p = 1 + 2q$, provided $|u|$ is small; then, obviously, $p > 5 > 1 + \sqrt{2}$ (see [5], where the critical value $1 + \sqrt{2}$ is discussed in more details). Nevertheless, our blow-up results (established below in Section 4) show that classical solutions can exist for Eq. (1.1) (in the case (1.4)) with arbitrary small data, which blow up in the case of big enough supports of the data. Such an "anomaly" seems to be caused mainly by the repellent influence of the potential energy term U_f in

Eq. (1.4), when $U_f[u] \geq$ const, $\forall\, u$. Compared with (1.3), the nonlinearity of the source term is much weaker than that in the right-hand side of (1.4), as $|u| \to +\infty$, but the potential energy term $U_f[u]$ is repellent. Taking into account the Keller's pioneering result [7] (see also that of Glassey [4] and the counter-example in [9, 12] for (1.1) with $f = u|u|^{p-1}$), one can expect an absence of global solutions of (1.1) for repellent U_f, when the test life-span T_f^0,

$$T_f^0 \equiv \int_0^{+\infty} \left(1 + 2\int_0^z f(s)\,\mathrm{d}s\right)^{-1/2} \mathrm{d}z,$$

is finite. However, it remains an open question in general whether Eq. (1.1) possesses global classical solutions for arbitrary data (1.5) if $U_f \geq$ const, but $T_f^0 = +\infty$. For Eq. (1.4) we answer this question in Section 4 below. More precisely, we show that there exists a unique global solution of (1.4), (1.5) for $0 < q \leq 2$ (when $T_f^0 = +\infty$). In Section 3 the behaviour of the positive solutions of the Newton equation $\ddot{v} = v \ln^q(1 + v^2)$, associated with Eq. (1.4), is studied. In Section 4 it is proven that the classical solutions of Eq. (1.4) increase exponentially ("counterdecay") if $0 < q < 2$, and blow up if $q > 2$, for data, either positive or negative, which produce the so-called space-destinated waves. Thus it is shown, in particular, that $q = 2$ is a critical value for the global classical solutions of Eq. (1.4).

2. TWO BASIC PRINCIPLES

Consider the following Cauchy problem for the semilinear wave equation:

$$\Box u = f(u), \quad x \in R^3,\ t > 0, \tag{2.1}$$

$$u(x,0) = \varphi(x), \quad u_t(x,0) = \psi(x),\ x \in R^3, \tag{2.2}$$

where $f \in C^2(R^1), f(0) = 0$, and $\varphi \in C^3(R^3), \psi \in C^2(R^3)$. Recall that a solution u of Eq. (2.1) is called classical in a set $G \subset R^3 \times [0, +\infty)$ such that $G \cap \{t = 0\}$ contains some domain in R^3 if $u \in C^2(G) \cap C^1(\overline{G})$, $\overline{G}$ being the closure of G. For a given compact $K \subset R^3$ we denote by K_T^- the part of the backward light cone contained in the strip $0 \leq t \leq T$ and based on K at $t = 0$. Similarly, K_T^+ denotes the forward light cone issued from K for $0 \leq t \leq T$.

We suppose known the local existence theorem for (2.1), (2.2), and the theorems for uniqueness and the continuous dependence on the data.

Below we shall use the integral equation

$$u = u^0 + E * f(u^+). \tag{2.3}$$

Here u^0 is the solution of the free wave equation $\Box u = 0$, given by the classical Kirchhoff formula

$$u^0(x,t) = \partial_t \left(\frac{t}{4\pi} \int_{|\omega|=1} \varphi(x + t\omega)\,\mathrm{d}s_\omega\right) + \frac{t}{4\pi} \int_{|\omega|=1} \psi(x + t\omega)\,\mathrm{d}s_\omega, \quad \omega \in R^3, \tag{2.4}$$

where E is the fundamental solution of the wave operator, $u^+ = u$ $(t > 0)$, $u^+ = 0$ $(t < 0)$, and $E * f(u^+)$ is the convolution in R^4 in the usual distribution sense. Recall also that the following known formula is valid for $E * f(u^+)$, namely,

$$E * f(u^+) = \frac{1}{4\pi} \int_0^t \tau \int_{|\omega|=1} f\left(u(x + \tau\omega, t - \tau)\right) \mathrm{d}s_\omega \mathrm{d}\tau, \tag{2.5}$$

for $x \in R^3$ and $t > 0$.

2.1. COMPARISON PRINCIPLE

We realize here, in a new case, a classical comparison idea known, e.g., from the theory of the finite-dimensional dynamic systems. The essence is to compare the solution $u(x,t)$ of (2.1), (2.2) with the solutions of the Newton equation $\ddot{v} = f(v)$ from below or above. In the next theorem we impose the following requirements upon the function f:

$$f \in C^2(R^1), \quad f(0) = 0, \ f' \geq 0, \tag{2.6}$$

in R^1, and use backward light cones C_B^- in $\{0 \leq t\}$, based on closed balls $B \subset R^3$. (Obviously, a cone C_B^- can be represented as K_T^- with $K = B$ and big enough T.) Now the comparison principle reads:

Theorem 2.1. *Suppose the condition* (2.6) *holds for* $f(u)$. *Let the functions* $u(x,t)$ *and* $u^0(x,t)$ *be the solution of* (2.1), (2.2) *and the function from* (2.4), *respectively. Let* C_B^- *be a given backward light cone. Suppose also that* u^0 *satisfies the inequality*

$$u^0(x,t) \geq a_1 + b_1 t, \quad (x,t) \in C_B^-, \tag{2.7}$$

with certain constants a_1, $b_1 \geq 0$ *depending on* B. *Then the following estimate holds for the solution* u:

$$u(x,t) \geq v_1(t), \quad (x,t) \in C_B^- \cap G \cap \{0 \leq t < T_1\}, \tag{2.8}$$

where $v_1(t)$ *is the solution of the Cauchy problem*

$$\ddot{v} = f(v), \quad v(0) = a, \ \dot{v}(0) = b, \tag{2.9}$$

defined for $t \in [0, T_1)$, *for a positive* T_1, *with* $a = a_1$, $b = b_1$. *Similarly, we have the estimate*

$$|u(x,t)| \leq v_2(t), \quad (x,t) \in C_B^- \cap G \cap \{0 \leq t < T_2\}, \tag{2.10}$$

if $v_2(t)$ *is the solution of* Eq. (2.9), *defined for* $t \in [0, T_2)$, $T_2 > 0$, *with data* $a = a_2$, $b = b_2$, *where*

$$|u^0(x,t)| \leq a_2 + b_2 t, \quad (x,t) \in C_B^-, \tag{2.11}$$

and $a_2, b_2 \geq 0$ *are certain constants depending on* B.

Proof. Take an arbitrary backward light cone K_T^- such that

$$K_T^- \subset C_B^- \cap G \cap \{0 \le t \le T_1\}$$

and set

$$w_1(t,T) = \min_y u(y,t), \quad (y,t) \in K_T^-, \ t \in [0,T),$$

with fixed t. Note that the solution u is non-negative in K_T^-: this follows from the condition (2.7) and the positivity of the operator in (2.5). For a fixed $t \in [0,T)$, in virtue of Eq. (2.3) and the inequality (2.7), we have the estimate

$$w_1(t,T) \ge a_1 + b_1 t + \min_x E * f(u^+)(x,t), \quad (x,t) \in K_T^-. \tag{2.12}$$

Next, we can easily verify the inequality

$$E * f(u^+)(x,t) \ge E_0 * f(w_1^+)(t), \quad (x,t) \in K_T^-, \tag{2.13}$$

where E_0 is the fundamental solution of the operator d^2/dt^2, and

$$E_0 * f(w_1^+)(t) = \int_0^t s f\left(w_1(t-s,T)\right) ds. \tag{2.14}$$

Indeed, the formula (2.5) shows that the fundamental solution E represents a positive measure and because of the monotonicity of function f we have

$$E * f(u^+) \ge E * f(w_1^+) = E_0 * f(w_1^+), \quad (x,t) \in K_T^-.$$

Then from (2.12), (2.13) we obtain

$$w_1(t,T) \ge a_1 + b_1 t + E_0 * f(w_1^+)(t,T), \quad t \in [0,T). \tag{2.15}$$

Now it is natural to compare the function $w_1(t,T)$ with the solution $v_1(t)$, using that $v_1(t)$ solves the equation

$$v_1(t) = a_1 + b_1 t + E_0 * f(v_1^+)(t), \quad t \in [0,T). \tag{2.16}$$

Employing familiar arguments, it is not difficult to show that

$$w_1(t,T) \ge v_1(t), \quad \forall t \in [0,T),$$

which proves (2.8), because K_T^- is an arbitrary cone.

To prove the estimate (2.10), we utilize fully similar arguments. We now set

$$w_2(t,T) = \max_y |u(y,t)|, \quad (y,t) \in K_T^-, \ t \in [0,T),$$

with fixed t. Then

$$w_2(t,T) \le a_2 + b_2 t + E_0 * f(w_2^+)(t,T)$$

as a consequence of (2.3), (2.11) and the properties of convolutions $E * f(u^+)$, $E_0 * f(w^+)$. Next it remains to compare the function $w_2(t,T)$ with the solution $v_2(t)$, repeating the arguments from the comparison of w_1 and v_1. This completes the proof of Theorem 2.1. □

This principle gives an affirmative answer to the natural question, concerning time extension of the classical solutions of Eq. (2.1), dominated by certain "super-solutions."

Theorem 2.2. *Suppose* (2.6) *is valid for* $f(u)$ *and the following inequality holds for the function* u^0:

$$u^0 : |u^0(x,t)| \leq a + bt, \quad (x,t) \in K^-_{T_0},$$

for a given cone $K^-_{T_0}$, *with constants* $a, b \geq 0$ *depending on* T_0. *Let* $v(t)$ *be the solution of the problem* (2.9) *defined for* $t \in [0,T]$, $T \leq T_0$. *Then there exists a unique classical solution* u_T *of* (2.1), (2.2) *in* K^-_T, *which coincides with* u *in* $G \cap K^-_T$, *where*

$$K^-_T = K^-_{T_0} \cap \{0 \leq t \leq T\}.$$

The proof, being known, is omitted (see, e.g., [5] and the references cited therein). □

2.3. POSITIVE AND NEGATIVE SOLUTIONS

In this section we establish the existence of positive and negative solutions of Eq. (2.1). We introduce a class of solutions called space-destinated waves and note that they are positive or negative solutions, for $t > 0$, if the initial data are positive or negative, respectively. To this end we use the Lorentz pseudometric m_L,

$$m_L = dt^2 - |dx|^2, \quad |dx|^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2, \quad x = (x^1, x^2, x^3).$$

Following Friedlander [2], we recall the relations

$$m_L(\xi,\eta) = \xi^0\eta^0 - \sum_{j=1}^{3} \xi^j\eta^j, m_L(\xi,\xi) = (\xi^0)^2 - |\xi'|^2, \tag{2.17}$$

where $\xi' = (\xi^1, \xi^2, \xi^3)$, $|\xi'| = \sqrt{\sum_{j=1}^3 (\xi^j)^2}$, and $\xi = (\xi', \xi^0)$, $\eta = (\eta', \eta^0)$ are arbitrary vectors in R^{3+1}.

Definition 2.1. A classical solution $u(x,t)$ of (2.1) is called a *space-destinated wave* if the set

$$\Sigma_0 = \{(x,t) \in R^{3+1} : u(x,t) = u(x_0, 0)\}$$

is a non-degenerated smooth 3D-hypersurface in a neighbourhood of $(x_0, 0)$, $\forall x_0 \in R^3$, and the following inequality is fulfilled:

$$(-1)m_L(\xi,\xi) \geq 0, \quad \xi \in T_{(x_0,0)}(\Sigma_0), \ \xi \neq 0, \tag{2.18}$$

where $T_{(x_0,0)}(\Sigma_0)$ is the tangent hyperplane to Σ_0 at the point $(x_0, 0)$. A classical solution u is called a *strongly space-destinated wave* if the inequality (2.18) is strong.

Remark 2.1 (existence of space-destinated waves). It is not difficult to verify (see, e.g., [2]) that $(-1)m_L \geq 0$ on the tangent space from (2.18) if and only if $m_L\,(u'(x_0,0), u'(x_0,0)) \geq 0$, where $u'(x,t) \in R^{3+1}$ is the gradient of $u(x,t)$ at the point (x,t). Now, using the coordinate representation $u'(x,0) = (\nabla\varphi(x), \psi(x))$, where $u(x,0) = \varphi(x), u_t(x,0) = \psi(x)$, and $\nabla\varphi$ is the gradient of φ, we apply the formula (2.17) for $\xi = \eta = u'$ and reach the following conclusion. A classical solution $u(x,t)$ is a space-destinated wave if and only if the data $\varphi(x), \psi(x)$ have the properties

$$(\nabla\varphi(x), \psi(x)) \neq 0, \quad |\psi(x)| - |\nabla\varphi(x)| \geq 0,\ x \in R^3.$$

Proposition 2.1. *Suppose the function f satisfies the condition*

$$f(u) \in C^2(R^1), \quad uf(u) \geq 0,\ u \in R^1,$$

and the solution $u(x,t)$ of (2.1), (2.2), *defined on a set G, is a space-destinated wave. Then $u(x,t)$ is positive in G, for $t > 0$, if the initial data $u(x,0), u_t(x,0)$ are non-negative, for $x \in R^3$, and at least one of the following two conditions hold:*

(i) $$u(x,0) \neq 0,\ x \in R^3;$$

(ii) $u(x,t)$ *is a strongly space-destinated wave.*

Similarly, the solution $u(x,t)$ is negative in G, for $t > 0$, if the initial data are non-positive and at least one of the conditions (i), (ii) *holds.*

The proof is omitted, because the key ideas can be taken from the proof of Theorem 2.1. □

Since the nonlinear part $f(u)$ of Eq. (2.1) is odd, the following connection between positive and negative solutions of this equation is obvious.

Proposition 2.2. *Suppose the function $f(u)$ satisfies the condition*

$$f(u) \in C^2(R^1), \quad f(-u) = -f(u),\ \forall u.$$

Then the map $u(x,t) \to -u(x,t)$ is a bijection between the sets of the positive and the negative classical solutions of (2.1).

3. POSITIVE SOLUTIONS OF A NEWTON EQUATION

The so-called automodel solutions of the problem (1.4), (1.5), when the initial data are constants, satisfy the Cauchy problem for the Newton equation:

$$\ddot{u} = u \ln^q(1 + u^2), \quad t > 0, \tag{3.1}$$

$$u(0) = u_0, \quad \dot{u}(0) = u_1\,, \tag{3.2}$$

where u_0, u_1 are constants. We study here the behaviour of the positive solutions $u(t)$ of (3.1), (3.2), provided $u_0 \geq 0$, $u_1 \geq 0$ and $u_0 + u_1 > 0$. The positive solutions of (3.1), (3.2) play the role of sub- and super-solutions for the wave equation (1.4), estimating the solutions of (1.4), (1.5) under the assumptions of the comparison principle (Theorem 2.1). For the solution $u(t)$ of (3.1), (3.2) we shall use the following well-known formulae:

$$u(t) = u_0 + u_1 t + \int_0^t (t-s) f\left(u(s)\right)\, ds, \tag{3.3}$$

$$t = \int_{u_0}^{u(t)} \left(u_1^2 + 2\int_{u_0}^{z} f(s)\, ds\right)^{-1/2} dz, \tag{3.4}$$

where $f(u) = u \ln^q(1+u^2)$.

A basic property of the positive solutions under discussion is given by the next lemma.

Lemma 3.1. *If $u_0 \geq 0, u_1 \geq 0$, $u_0 + u_1 > 0$, then the solution $u(t)$ of the problem (3.1), (3.2) is positive and defined in a maximal interval $[0, T^0)$. Moreover, $u(t)$ is a monotonically increasing function in $[0, T^0)$ such that*

$$\lim u(t) = +\infty, \quad t \to T^0. \tag{3.5}$$

Proof. Let us write the relation (3.4) in the form $F(u) = t$, where

$$F(u) \equiv \int_{u_0}^{u} \left(u_1^2 + 2\int_{u_0}^{z} f(s)\, ds\right)^{-1/2} dz, \quad u \geq u_0.$$

Due to the monotonicity of the function F we have, for the solution $u(t)$ of the problem (3.1), (3.2), $u(t) = F^{-1}(t)$, where F^{-1} is the inverse function of F. The statements of the lemma directly follow from the classical theory of the Newton equation. □

In the next lemma we study a general estimate for the solution $u(t)$.

Lemma 3.2. *Let $u(t)$ be the solution from Lemma 3.1, defined for $t \in [0, T^0)$ and $\psi(z) = z(z-u_0)\ln^q(1+z^2)$. Then the solution $u(t)$ satisfies the estimate*

$$\int_{u_0}^{u} \left(u_1^2 + 2\psi(z)\right)^{-1/2} dz \leq t \leq \sqrt{1+q} \int_{u_0}^{u} \left((1+q)u_1^2 + \psi(z)\right)^{-1/2} dz \tag{3.6}$$

for all $t \in [0, T^0)$, $u = u(t)$.

Proof. We begin by studying the function $\varphi(u) = 2\int_{u_0}^{u} f(s)\, ds$:

$$\varphi(u) = 2\int_{u_0}^{u} s \ln^q(1+s^2)\, ds = \int_{1+u_0^2}^{1+u^2} \ln^q t\, dt.$$

Then

$$\varphi'(z) = 2z\ln^q(1+z^2),$$

$$\psi'(z) = (2z - u_0)\ln^q(1+z^2) + \frac{2qz^2}{1+z^2}(z-u_0)\ln^{q-1}(1+z^2).$$

Next we use the estimate for $\psi'(z)$:

$$\psi'(z) \le 2z\ln^q(1+z^2) + 2qz\ln^q(1+z^2), \quad z \ge u_0 .$$

It follows from the above formula for $\psi'(z)$, together with the inequality $z^2(1+z^2)^{-1} \le \ln(1+z^2)$. Estimating $\psi'(z)$ from below, we obtain

$$\psi'(z) \ge z\ln^q(1+z^2) = \frac{1}{2}\varphi'(z), \quad z \ge u_0.$$

Hence

$$\frac{1}{2}\varphi'(z) \le \psi'(z) \le (1+q)\varphi'(z), \quad z \ge u_0,$$

and because of the initial data $\varphi(u_0) = \psi(u_0) = 0$ we obtain the inequality

$$\frac{1}{2}\varphi(z) \le \psi(z) \le (1+q)\varphi(z), \quad z \ge u_0. \tag{3.7}$$

From (3.4) we see that

$$t = \int_{u_0}^{u} \left(u_1^2 + \varphi(z)\right)^{-1/2} dz, \quad u = u(t),$$

and applying (3.7) we establish the estimate

$$u_1^2 + \frac{\psi(z)}{1+q} \le u_1^2 + \varphi(z) \le u_1^2 + 2\psi(z),$$

which yields (3.6). □

The next lemma is a direct consequence of Theorems 2.1 and 2.2.

Lemma 3.3 (comparison and continuation principles). *Suppose $u(t)$, $v(t)$ are the solutions of* (3.1), (3.2) *with data* (u_0, u_1), (v_0, v_1), *defined in the maximal intervals* $[0, T_u)$, $[0, T_v)$, *respectively. Then:*

(i) $$u(t) \le v(t), \quad t \in [0,T_u) \cap [0,T_v),$$

if $0 \le u_j \le v_j$, $j = 0, 1$;

(ii) $$T_u \ge T_v,$$

if $u(t) \le v(t)$ *in* $[0,T_u) \cap [0,T_v)$.

Below we shall employ systematically the general estimate (3.6) in order to study the global solutions of the problem (3.1), (3.2) ($0 < q \le 2$), their large time behaviour and the blow-up phenomena ($2 < q$).

Lemma 3.4 (global solutions). *Let $0 < q \leq 2$ and $u_0 \geq 0$, $u_1 \geq 0$, $u_0+u_1 > 0$. Then the solution $u(t)$ of the problem* (3.1), (3.2) *is global, i.e. $T^0 = +\infty$, and for $0 < q < 2$ satisfies the estimate*

$$u(t) \leq \exp\left((4+u_1)t\right)^{2/(2-q)}, \quad t \geq 3 + u_0 + \frac{2}{d_{0,1}}, \quad 0 < q < 2, \tag{3.8}$$

where $d_{0,1} = \ln^{q/2}(1+u_0^2)$ if $u_1 = 0$, $d_{0,1} = u_1$ if $u_0 = 0$, and $d_{0,1} = \min\{\ln^{q/2}(1+u_0^2), u_1\}$ if $u_1 u_0 > 0$.

Proof. Suppose $0 < q < 2$. Consider firstly the case $u_0 \geq 0$, $u_1 > 0$. In virtue of (3.5) we can chose $t_1 \in (0, T^0)$ such that $u(t) \geq 2 + u_0$ for $t \geq t_1$. Then, due to (3.6), we have

$$t \geq \int_{2+u_0}^{u} \left(u_1^2 + 2(1+z)^2 2^q \ln^q(1+z)\right)^{-1/2} \mathrm{d}z \quad (t \geq t_1).$$

Therefore

$$(u_1^2 + 8)^{-1/2} \int_{2+u_0}^{u} (1+z)^{-1} \ln^{-q/2}(1+z)\, \mathrm{d}z \leq t, \quad t \geq t_1. \tag{3.9}$$

But

$$\int_{2+u_0}^{+\infty} (1+z)^{-1} \ln^{-q/2}(1+z)\, \mathrm{d}z = +\infty$$

and (3.9) shows that the solution $u(t)$ is global. The inequality $u(t) \geq u_0 + u_1 t \geq 2 + u_0$, valid for $t \geq 2/u_1$, and (3.9) yield

$$\ln^{1-q/2}(1+u) \leq (3+u_1)t + \ln^{1-q/2}(3+u_0) \leq (4+u_1)t \quad (t \geq 3 + u_0).$$

The estimate (3.8) obviously follows now with $d_{0,1} \leq u_1$. In the case $u_0 > 0$, $u_1 \geq 0$ we argue in a similar manner — noticing that $u(t)$ is global and employing the formula (3.3) to get

$$u(t) \geq u_0 + (t^2/2) f(u_0) \geq 2 + u_0 \quad \left(t \geq 2/\sqrt{f(u_0)},\ f(u_0) = u_0 \ln^q(1+u_0^2)\right).$$

Then (3.8) also follows but with $d_{0,1} \leq \sqrt{f(u_0)}$. Suppose now $q = 2$. In the case $u_0 \geq 0, u_1 > 0$ we reach again (3.9) and the solution $u(t)$ is thus global. □

Lemma 3.5 (lower estimates). *Let $0 < q < 2$ and $T \geq 1$ be a parameter. Suppose the initial data* (3.2) *depend on T, i.e. $u_0 = u_{0,T}$, $u_1 = u_{1,T}$, satisfying the assumptions*

$$0 \leq u_{0,T}, \quad \frac{c_1}{1+T} \leq u_{1,T}, \tag{3.10}$$

$c_1 > 0$ being a constant. Then the solution $u(t)$ of the problem (3.1), (3.2) *satisfies the estimate*

$$u(t) \geq (4/c_1) \exp\left(\frac{2-q}{16} h^* t\right)^{2/(2-q)}, \quad t \in [2T/3, T] \quad (0 < q < 2), \tag{3.11}$$

when $T \geq 24/(2-q)h^*$ *and* $h^* = c_1^2/(64+c_1^2)$.

The proof uses arguments fully similar to those from the proof of Lemma 3.4. That is why it is omitted. □

Lemma 3.6 (blow-up time). *Suppose* $q > 2$ *in* Eq. (3.1) *and* $u_0 u_1 \geq 0$, $u_0 + u_1 > 0$. *Then the life-span* T^0 *is finite for the solution* $u(t)$ *of* (3.1), (3.2), *which blows up at the moment* T^0. *Moreover, the blow-up time* T^0 *can be estimated as follows*:

$$\frac{1}{2^{q/2}(q-2)(1+u_1)(1+u_0)^{q/2-1}} \leq T^0 \leq \sqrt{1+q}\left(\frac{\sqrt{3}}{u_1} + \frac{2}{2^{q/2}(q-2)}\right) \quad (u_1 > 0); \tag{3.12}$$

$$\frac{1}{2^{q/2}(q-2)(1+u_1)(1+u_0)^{q/2-1}} \leq T^0 \leq \sqrt{1+q}\left(2\sqrt{\frac{3+u_0}{u_0 \ln^q(1+u_0^2)}} + \frac{4}{2^{q/2}(q-2)}\right) \quad (u_0 > 0). \tag{3.13}$$

Proof. To establish the left-hand sides of (3.12), (3.13), we shall use the left-hand side of (3.6). For the sum $u_1^2 + 2\psi(z)$ we have

$$u_1^2 + 2\psi(z) \leq u_1^2 + 2z^2 \ln^q(1+z^2) \leq (1+u_1)^2 (2z)^2 \ln^q (2z)^2,$$

when $z \geq 1 + u_0$. From (3.6) it follows

$$T^0 \geq \int_{1+u_0}^{+\infty} \frac{dz}{(1+u_1) 2^{q/2} 2z \ln^{q/2}(2z)} = \frac{\ln^{1-q/2}(2+2u_0)}{2^{q/2}(q-2)(1+u_1)}.$$

Afterwards we use the inequality $\ln 2(1+u_0) \leq 1+u_0$, which yields the upper estimates of (3.12), (3.13). In the case of $u_1 > 0$ we can consider $u_0 = 0$ (due to the comparison principle). Then from (3.6), taking the appropriate limit, it follows

$$T^0 \leq \sqrt{1+q} \int_0^{+\infty} \frac{dz}{\sqrt{3u_1^2 + \psi(z)}} \qquad \left(\psi(z) = z^2 \ln^q(1+z^2)\right).$$

But for the just defined function $\psi(z)$ we have $\psi(z) \geq 2^q z^2 \ln^q z$ for $z \geq 3$. Hence

$$T^0 \leq \sqrt{1+q} \int_0^3 \frac{dz}{\sqrt{3}u_1} + \frac{\sqrt{1+q}}{2^{q/2}} \int_3^{+\infty} \frac{dz}{z \ln^{q/2} z}$$

and the right-hand side of (3.12) immediately follows.

If $u_0 > 0$, we can similarly consider $u_1 = 0$ (again recalling the comparison principle). Next, employing the estimate

$$\psi(z) \geq (z-u_0)^2 \ln^q(1+z^2) \geq 2^{q-2} z^2 \ln^q z$$

for $z \geq 2u_0$, we obtain from (3.6)

$$T^0 \leq \sqrt{1+q} \int_{u_0}^{3+2u_0} \frac{\mathrm{d}z}{\sqrt{u_0 \ln^q(1+u_0^2)(z-u_0)}} + \sqrt{1+q} \int_{3+2u_0}^{+\infty} \frac{\mathrm{d}z}{2^{q/2-1} z \ln^{q/2} z},$$

which gives the right-hand side of (3.13). The blow-up property $\lim u(t) = +\infty$ for $t \to T^0$ has been already established in Lemma 3.1. □

4. GLOBALITY, COUNTER-DECAY AND EXPLOSION OF SOLUTIONS

In this section the main results of the paper are formulated and proven for the class (1.4), where the potential energy contains a repellent term. It becomes clear, in particular, that for Eq. (1.4) $q = 2$ is a critical value when the global existence problem for arbitrary big data is considered.

Theorem 4.1 (global solutions). *Let* $0 < q \leq 2$ *in* Eq. (1.4) *and* $\varphi(x) \in C^3$, $\psi(x) \in C^2$ *are arbitrary functions. Then the classical solution* $u(x,t)$ *of the problem* (1.4), (1.5) *is global.*

Proof. By the local theory the classical solution of (1.4), (1.5) exists in a set $G \subset R^3 \times [0, +\infty)$. Let $T > 0$ be arbitrary big. Choose an arbitrary ball $B \subset R^3$ with a radius $R \geq T$ and let K_T^- be the backward light cone (see Section 2) based on the compact $K = B$, $K_T^- \subset \{0 \leq t \leq T\}$. Let

$$a = \max |\varphi(x)|, \quad b = \max \left(|\nabla\varphi(x)| + |\psi(x)|\right), \ x \in B.$$

Then for the free wave $u^0(x,t)$ (see formula (2.4)) we have

$$|u^0(x,t)| \leq a + bt, \quad (x,t) \in C_B^-,$$

where C_B^- is the backward light cone based on B and $K_T^- = C_B^- \cap \{0 \leq t \leq T\}$ for $K = B$. Moreover, the solution $v(t)$ of the problem (2.9) is global in the case of $f(v) = v \ln^q(1+v^2)$, $q \in (0, 2]$, according to Lemma 3.4, and $v(t)$ is defined, in particular, for $t \in [0, T]$. Now the continuation principle (Theorem 2.2) assures that the solution $u(x,t)$ is indeed defined in K_T^-. This proves the theorem. □

Theorem 4.2 (exponential counter-decay). *Let* $0 < q < 2$ *and let the classical solution* $u(x,t)$ *of the problem* (1.4), (1.5) *be a strongly space-destinated wave, with either non-positive or non-negative initial data, satisfying the inequality*

$$\inf_x (1 + |x|) \left(|\psi(x)| - |\nabla\varphi(x)|\right) > 0, \quad x \in R^3. \tag{4.1}$$

Then for each compact $K \subset R^3$ *there exist positive constants* c', h', $t' = t'(K)$ *such that*

$$|u(x,t)| \geq (4/c') \exp\left(\frac{2-q}{16} h' t\right)^{2/(2-q)} \quad (0 < q < 2) \tag{4.2}$$

on $K_\infty^+ \cap \{t \geq t'\}$, *where* $K_\infty^+ = \cup K_T^+$, $T \in (0, +\infty)$, *is the forward light cone issued from* K. *If, moreover, we have*

$$\sup_x \left(|\varphi(x)| + |\nabla\varphi(x)| + |\psi(x)|\right) < +\infty, \ x \in R^3, \tag{4.3}$$

then the solution $u(x,t)$ *satisfies, besides* (4.2), *the estimate*

$$|u(x,t)| \leq \exp\left((4 + c'')t\right)^{2/(2-q)}, \forall t \geq t'', \tag{4.4}$$

as well, with certain positive constants c'' *and* $t'' = t''(c_0, c'')$.

Proof. For a given compact $K \subset R^3$ let us fix a ball $B_0 \subset R^3$ with a radius r_0 such that $K \subset B_0$. Translate next the origin in R^3 at the center of B_0. Obviously, it suffices to consider only the case of non-negative initial data, due to Proposition 2.2. Now (4.1) shows that the constant c_1, defined as

$$c_1 = \inf_x (1 + |x|)\left(\psi(x) - |\nabla\varphi(x)|\right), \tag{4.5}$$

is positive.

Let us take an arbitrary $T \geq 1 + r_0$ and an arbitrary point $(x_T, T) \in K_\infty^+$. Denote by $C^-(x_T)$ the backward light cone with the top at (x_T, T). From the Kirchhoff formula we obtain the following inequality for the free wave $u^0(x,t)$:

$$u^0(x,t) \geq \frac{1}{4\pi t}\int_{|x-y|=t} \left(\psi(y) - |\nabla\varphi(y)|\right) \mathrm{d}s_y. \tag{4.6}$$

When (x,t) varies in the cone $C^-(x_T)$, we have obviously

$$|y| \leq r_0 + 2T, \quad y \in R^3 : \ |x - y| = t,$$

and from (4.5) we find

$$\psi(y) - |\nabla\varphi(y)| \geq \frac{c_1}{1 + r_0 + 2T} \geq \frac{c_1}{1 + 3T} \geq \frac{c'}{1 + T}$$

with $c' = c_1/3$. Then (4.6) yields

$$u^0(x,t) \geq \frac{c'}{1+T}t, \quad (x,t) \in C^-(x_T).$$

Now we can apply the comparison principle (Theorem 2.1) to find

$$u(x,t) \geq v_T(t), \quad (x,t) \in C^-(x_T),$$

where $v_T(t)$ is the solution of the problem

$$\ddot{v} = v \ln^q(1 + v^2), \quad v(0) = 0, \ \dot{v}(0) = \frac{c'}{1+T}.$$

Afterwards it remains to apply Lemma 3.5 and to set $(x, t) = (x_T, T)$, then

$$u(x_T, T) \geq v_T(T) \geq (4/c') \exp(\frac{2-q}{16} h'T)^{2/(2-q)} \quad (0 < q < 2).$$

This proves the estimate (4.2). When (4.3) holds, we set

$$c_0 = \sup_x |\varphi(x)|, \quad c'' = \sup_x (|\nabla\varphi(x)| + |\psi(x)|),$$

and using the Kirchhoff formula we get

$$|u^0(x,t)| \leq c_0 + c''t, \quad (x,t) \in R^3 \times [0, +\infty).$$

Then, in virtue of Theorem 2.1, we conclude that

$$|u(x,t)| \leq U(t), \quad (x,t) \in R^3 \times [0, +\infty),$$

with $U(t)$ solving the equation

$$\ddot{U} = U \ln^q(1 + U^2), \quad U(0) = c_0, \; \dot{U}(0) = c''.$$

It is clear now that the estimate (4.4) follows from Lemma 3.4. □

In the next theorem we shall deal with the classical solutions $u(x,t)$ called space-destinated waves on a given compact $K \subset R^3$, which satisfy the requirements of Definition 2.1 for each $x \in K$. The notations

$$m_0 = \min_x |u(x,0)|, \; x \in K, \tag{4.7}$$

$$m_1 = \min_x (|u_t(x,0)| - |\nabla_x u(x,0)|), \; x \in K, \tag{4.8}$$

$$\tau_0 = \lim_{\varepsilon \to 0} \min \left(\frac{\sqrt{3}}{\varepsilon^2 + m_1}, 2\sqrt{\frac{3 + m_0}{\varepsilon^2 + m_0 \ln^q(1 + m_0^2)}} \right), \; m_0 + m_1 \neq 0, \tag{4.9}$$

$$M_0 = \max_x |u(x,0)|, x \in K; \tag{4.10}$$

$$M_1 = \max_x (|u_t(x,0)| + |\nabla_x u(x,0)|), \; x \in K, \tag{4.11}$$

shall be used for a given space-destinated wave $u(x,t)$ on K. By $T_q = T_q(u, K)$ we shall denote the supremum of all $T > 0$ such that a given classical solution u of (1.4) exists in the light cone K_T^-; T_q is usually called the life-span of u for the compact K.

Theorem 4.3 (blow-up of the solution). *Suppose $q > 2$ and $\varphi \in C^3$, $\psi \in C^2$ are arbitrary initial data, either non-negative or non-positive on a ball $B_r \subset R^3$, such that the solution $u(x,t)$ of* (1.4), (1.5) *is a space-destinated wave on B_r (r is the radius of the ball). If the numbers $m_j(r)$, $j = 0, 1$, and $\tau_0(r)$ satisfy the inequalities*

$$m_0(r) + m_1(r) > 0, \quad \sqrt{1+q}\left(\tau_0(r) + \frac{2^{2-q/2}}{q-2}\right) \leq r,$$

then the solution $u(x,t)$ blows up in a finite time and the life-span $T_q(r) = T_q(u, B_r)$ satisfies the estimates

$$\frac{2^{-q/2}\left(1+M_0(r)\right)^{1-q/2}}{(q-2)\left(1+M_1(r)\right)} \le T_q(r) \le \sqrt{1+q}\left(\tau_0(r)+\frac{2^{2-q/2}}{q-2}\right),$$

where $m_j(r)$, $M_j(r)$, $j = 0, 1$, and $\tau_0(r)$ are the constants from (4.7)–(4.11) with $K = B_r$.

Proof. By the Kirchhoff formula we obtain the estimates

$$m_0(r) + tm_1(r) \le |u^0(x,t)| \le M_0(r) + tM_1(r)$$

for $u^0(x,t)$ on the backward light cone $C^-(r) = K_T^-$, where $K = B_r, T = r$. In addition, $u^0(x,t)$ is either positive or negative on $C^-(r) \cap \{t > 0\}$. Then the comparison principle yields

$$v(t) \le |u(x,t)|, \quad (x,t) \in C^-(r) : 0 \le t < \min\left(r, T_v^0, T_q(r)\right), \tag{4.12}$$

$$|u(x,t)| \le U(t), \quad (x,t) \in C^-(r) : 0 \le t < \min\left(r, T_q(r), T_U^0\right), \tag{4.13}$$

where $v(t)$ and $U(t)$ satisfy the equation $\ddot{z} = z \ln^q(1+z^2)$ with the initial data (m_0, m_1) and (M_0, M_1) and the life-spans T_v^0, T_U^0, respectively. Next, from Lemma 3.6 we see that

$$T_U^0 \ge \frac{2^{-q/2}(1+M_0)^{1-q/2}}{(q-2)(1+M_1)}, \quad T_v^0 \le \sqrt{1+q}(\tau_0 + \frac{2^{2-q/2}}{q-2}). \tag{4.14}$$

Then the estimates (4.12)–(4.14), together with the inequalities

$$T_U^0 \le T_q(r) \le T_v^0,$$

prove the theorem. □

REFERENCES

1. Brenner, Ph., W. von Wahl. Global classical solutions of non-linear wave equations. *Math. Z.*, **176**, 1981, 87–121.
2. Friedlander, F. G. The wave equation on a curved space-time. Cambridge University Press, London, 1975.
3. Ginibre, J., G. Velo. Regularity of solutions of the critical and subcritical nonlinear wave equation. *Nonl. Anal. Theory Meth. Appl.*, **22**, 1994, 1–19.
4. Glassey, R. Blow-up theorems for nonlinear wave equations. *Math. Z.*, **132**, 1973, 183–203.
5. John, F. Blow-up of solutions of nonlinear wave equations in three space dimensions. *Manuscripta Math.*, **28**, 1979, 235–268.
6. Jörgens, K. Das Anfangswertproblem im Grossen für eine Klasse nicht-linearer Wellengleichungen. *Math. Z.*, **77**, 1961, 295–308.

7. Keller, J .B. On solutions of non-linear wave equations. *Comm. Pure Appl. Math.*, **10**, 1957, 523–530.
8. Lindblad, H. Blow-up for solutions of $\Box u = |u|^p$ with small initial data. *Comm. Part. Diff. Eq.*, **15**, 1990, 757–821.
9. Reed, M. Abstract nonlinear wave equations. In: *Springer Lecture Notes in Math.*, **507**, 1975, 1–128.
10. Shatah, J., M. Struwe. Regularity result for nonlinear wave equations. *Ann. Math.*, **138**, 1993, 503–518.
11. Strauss, W. Nonlinear wave equations, In: *C.B.M.S. Lecture Notes*, **73**, American Math. Soc., Providence, RI, 1989.
12. Struwe, M. Semilinear wave equations. *Bulletin (New series) of the A.M.S.*, **36**, 1992, 53–85.
13. Takamura, H. An elementary proof of the exponential blow-up for semilinear wave equations. *Math. Meth. Appl. Sci.*, **17**, 1994, 239–249.
14. Zhon, Y. Cauchy problem for semilinear wave equations in four space dimensions with small initial data. IDMF Preprint (9217), to appear in *J. Diff. Equations.*

Received March 15, 2000

Revised May 3, 2000

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
5 Blvd J. Bourchier, P.O. Box 48
BG-1164 Sofia, BULGARIA
E-mail: boev@fmi.uni-sofia.bg

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

A NOTE ON THE SHEAR MODULUS OF A BINARY ELASTIC MIXTURE

M. K. KOLEV and K. Z. MARKOV

The Hashin-Shtrikman and Walpole bounds on the effective shear modulus of a binary elastic mixture are revisited. A simple method of derivation is given as a generalization of the approach, recently proposed by one of the authors in the absorption and scalar conductivity problems for a two-phase medium.

Keywords: two-phase random media, effective shear modulus, variational estimates, Hashin-Shtrikman and Walpole bounds

MSC 2000: 74E30, 74E35, 60G60

The aim of this note is to present and discuss a simple derivation of the well-known two-point estimates on the effective shear modulus of a binary elastic mixture, due to Hashin and Shtrikman [1] and Walpole [2]. As a matter of fact, this is a continuation of the recent paper [3], where a similar analysis is performed for bounding the effective bulk modulus of the mixture. In the case of a shear modulus, however, a number of technical difficulties arise, which makes the analysis much more involving. The basic idea, as in [3], is a generalization of the approach, used by one of the authors in the absorption and scalar conductivity cases [4].

Let us recall first how the problem is posed, see, e.g. [5, 6]. Assume that the mixture is statistically homogeneous and isotropic. Let

$$\chi_i(x) = \begin{cases} 1, & \text{if } x \in \Omega_i, \\ 0, & \text{otherwise,} \end{cases}$$

be the characteristic function of the region Ω_i, occupied by one of the constituents, labelled 'i', $i = 1, 2$, so that $\chi_1(x) + \chi_2(x) = 1$. Hereafter, all quantities, pertaining to the region Ω_1 or Ω_2, are supplied with the subscript '1' or '2', respectively.

The statistical properties of the medium follow from the set of multipoint moments of one of the functions $\chi_i(x)$, say $\chi_2(x)$ for definiteness, or, which is the same, by the volume fraction $\eta_2 = \langle\chi_2(x)\rangle$ of the phase '2', and the multipoint moments

$$M_2(x) = \langle\chi_2'(0)\chi_2'(x)\rangle\,, \quad M_3(x,y) = \langle\chi_2'(0)\chi_2'(x)\chi_2'(y)\rangle\,, \dots\,, \tag{1}$$

with $\chi_2'(x) = \chi_2(x) - \eta_2$ being the fluctuating part of the field $\chi_2(x)$, see, e.g., [5, 6]. The angled brackets $\langle\cdot\rangle$ hereafter denote ensemble averaging. One point could be taken at the origin because of the assumed statistical homogeneity, as already done in (1).

Recall that for a statistically isotropic binary medium under study one has

$$M_2(0) = \left\langle\chi_2'^2(0)\right\rangle = \eta_1\eta_2, \quad M_3(0) = \left\langle\chi_2'^3(0)\right\rangle = \eta_1\eta_2(\eta_1 - \eta_2). \tag{2}$$

Assuming also the constituents isotropic, the fourth-rank tensor of elastic moduli of the medium, $\mathbf{L}(x)$, is a random field of the familiar form

$$\begin{aligned} \mathbf{L}(x) &= 3k(x)\mathbf{J}' + 2\mu(x)\mathbf{J}'', \\ k(x) &= k_1\chi_1(x) + k_2\chi_2(x) = \langle k\rangle + [k]\chi_2'(x), \\ \mu(x) &= \mu_1\chi_1(x) + \mu_2\chi_2(x) = \langle\mu\rangle + [\mu]\chi_2'(x), \end{aligned} \tag{3}$$

where k and μ stand, as usual, for the bulk and shear modulus, respectively. The square brackets denote the jumps of the appropriate quantities, say, $[k] = k_2 - k_1$, $[\mu] = \mu_2 - \mu_1$, etc. In Eq. (3), $\mathbf{J}'$ and $\mathbf{J}''$ are the basic isotropic fourth-rank tensors with the Cartesian components

$$J'_{ijkl} = \frac{1}{3}\delta_{ij}\delta_{kl}, \quad J''_{ijkl} = \frac{1}{2}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right). \tag{4}$$

The displacement field $u(x)$ in the medium, at the absence of body forces, is governed by the well-known equations

$$\begin{aligned} \nabla\cdot\boldsymbol{\sigma}(x) &= 0, \\ \boldsymbol{\sigma}(x) = \mathbf{L}(x) : \boldsymbol{\varepsilon}(x) &= k(x)\theta(x)\mathbf{I} + 2\mu(x)\mathbf{d}(x), \\ \boldsymbol{\varepsilon} = \frac{1}{2}(\nabla u + u\nabla), &\quad \mathbf{d}(x) = \boldsymbol{\varepsilon}(x) - \frac{1}{3}\theta(x)\mathbf{I}, \end{aligned} \tag{5}$$

where $\boldsymbol{\sigma}$ denotes the stress tensor, $\boldsymbol{\varepsilon}$ is the small strain tensor, generated by the displacement field $u(x)$, $\mathbf{d}$ is the strain deviator, and $\theta = \operatorname{tr}\boldsymbol{\varepsilon}$ is the volumetric strain. The colon designates contraction with respect to two pairs of indices and $\mathbf{I}$ is the unit second-rank tensor.

The system (5) is supplied with the condition

$$\langle\boldsymbol{\varepsilon}(x)\rangle = \mathbf{E}, \tag{6}$$

prescribing the macroscopic strain tensor $\mathbf{E}$, imposed upon the medium.

Recall [5] that the random problem (5), (6) is equivalent to the variational principle of classical type

$$W[\varepsilon(x)] = \langle \varepsilon(x) : \mathbf{L}(x) : \varepsilon(x) \rangle \to \min, \qquad \min W = \mathbf{E} : \mathbf{L}^* : \mathbf{E}. \tag{7}$$

The energy functional W is considered over the class of random fields $u(x)$ that generate strain fields $\varepsilon(x)$, complying with the condition (6). In Eq. (7), $\mathbf{L}^*$ is the tensor of effective elastic moduli for the medium which, in the isotropic case under study, has the form

$$\mathbf{L}^* = 3k^*\mathbf{J}' + 2\mu^*\mathbf{J}'', \tag{8}$$

where k^* and μ^* are the effective bulk and shear modulus of the mixture, respectively.

Consider, guided by [3] and [4], the class of trial fields for the variational principle (7):

$$\mathcal{K}^{(1)} = \Big\{ \tilde{u}(x) \;\Big|\; \tilde{u}(x) = \mathbf{E} \cdot x + \alpha \mathbf{E} : \int \Big(\nabla G(x-y) \otimes \mathbf{I} + \kappa \nabla\nabla\nabla F(x-y) \Big) \chi_2'(y)\, d^3y \Big\}, \tag{9}$$

having assumed now that $\mathbf{E}$ is deviatoric, $\operatorname{tr}\mathbf{E} = 0$. Since the solution, $u(x)$, of the problem (5), (6) linearly depends on $\mathbf{E}$, we can assume that $\operatorname{tr}\mathbf{E}\cdot\mathbf{E} = \mathbf{E} : \mathbf{E} = 1$. In the class of trial fields (9) α and κ are adjustable scalar parameters and the kernels there read

$$G(x) = \frac{1}{4\pi|x|}, \quad F(x) = \frac{|x|}{4\pi}. \tag{10}$$

Hereafter the integrals are over the whole $\mathbb{R}^3$ if the integration domain is not explicitly indicated.

It is noted that the class of trial fields (9) has been first employed by McCoy [7], when deriving the Beran's type bound [8] for the shear modulus. The only difference is that we have allowed the multiplier κ in (9) to be adjustable as well (an idea already used by Milton and Phan-Thien [9]). In the final stage of our procedure, the appropriate optimization will bring forth the "best" value $\kappa_{\text{opt}} = -1/\big(4(1-\nu_2)\big)$, see Eq. (23) below. This means that the integrand in the right-hand side of (9) would *exactly* coincide with the Green tensor of one of the constituents. Hence the original McCoy's class of trial fields [7] will show up eventually. (See also the discussion in [10].)

It is to be also noted that in [3], when studying in a similar way the effective bulk modulus, we have chosen $\mathbf{E}$ spherical. This assumption considerably simplified the analysis (in particular, there was no need to introduce the second term in the integrand of the right-hand side of (9), containing the triple gradient). In this case the result is the three-point bound on the bulk modulus, proposed by Beran and Molyneux [11].

The energy functional W, when restricted over $\mathcal{K}^{(1)}$, becomes a quadratic function of α and κ:

$$W[\widetilde{u}(x)] = A + 2B\alpha + C\alpha^2, \quad A = 2\langle\mu\rangle, \quad B = 2[\mu](\mathsf{Z}_1 + \kappa\mathsf{U}_1),$$

$$C = \langle\mu\rangle\left(\mathsf{Z}_2 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_2 + 4\kappa\mathsf{V}_2 + 2\kappa^2\mathsf{U}_2\right) \tag{11}$$

$$+ [\mu]\left(\mathsf{Z}_3 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_3 + 4\kappa\mathsf{V}_3 + 2\kappa^2\mathsf{U}_3\right) + (1 + 2\kappa)^2\left(\langle k\rangle\,\mathsf{T}_2 + [k]\mathsf{T}_3\right),$$

with the dimensionless statistical parameters for the medium, defined as follows:

$$\mathsf{Z}_1 = (\mathbf{E}\cdot\mathbf{E}) : \int \nabla\nabla G(y)\, M_2(y)\, d^3y,$$

$$\mathsf{Z}_2 = (\mathbf{E}\cdot\mathbf{E}) : \iint \nabla\nabla G(y_1)\cdot\nabla\nabla G(y_2) M_2(y_1 - y_2)\, d^3y_1 d^3y_2,$$

$$\mathsf{Z}_3 = (\mathbf{E}\cdot\mathbf{E}) : \iint \nabla\nabla G(y_1)\cdot\nabla\nabla G(y_2) M_3(y_1, y_2)\, d^3y_1 d^3y_2,$$

$$\mathsf{U}_1 = \mathbf{E} : \int \nabla\nabla\nabla\nabla F(y)\, M_2(y)\, d^3y : \mathbf{E},$$

$$\mathsf{U}_2 = \mathbf{E} : \iint \nabla\nabla\nabla\nabla F(y_1) : \nabla\nabla\nabla\nabla F(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 : \mathbf{E},$$

$$\mathsf{U}_3 = \mathbf{E} : \iint \nabla\nabla\nabla\nabla F(y_1) : \nabla\nabla\nabla\nabla F(y_2)\, M_3(y_1, y_2)\, d^3y_1 d^3y_2 : \mathbf{E}, \tag{12}$$

$$\mathsf{V}_2 = \mathbf{E} : \iint \nabla\nabla G(y_1)\cdot\nabla\nabla\nabla\nabla F(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 : \mathbf{E},$$

$$\mathsf{V}_3 = \mathbf{E} : \iint \nabla\nabla G(y_1)\cdot\nabla\nabla\nabla\nabla F(y_2)\, M_3(y_1, y_2)\, d^3y_1 d^3y_2 : \mathbf{E},$$

$$\mathsf{T}_2 = \mathbf{E} : \iint \nabla\nabla G(y_1)\otimes\nabla\nabla G(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 : \mathbf{E},$$

$$\mathsf{T}_3 = \mathbf{E} : \iint \nabla\nabla G(y_1)\otimes\nabla\nabla G(y_2)\, M_3(y_1, y_2)\, d^3y_1 d^3y_2 : \mathbf{E}.$$

Moreover, the parameters Z_1 and Z_2 can be easily evaluated in the statistically isotropic case under study. Indeed, $M_2(y) = M_2(|y|)$ then and the appropriate integrals in the definitions of Z_1 and Z_2 are isotropic second-rank tensors, thus proportional to $\mathbf{I}$:

$$\int \nabla\nabla G(y)\, M_2(y)\, d^3y = c_1\mathbf{I},$$

$$\iint \nabla\nabla G(y_1)\cdot\nabla\nabla G(y_2) M_2(y_1 - y_2)\, d^3y_1 d^3y_2 = c_2\mathbf{I},$$

with certain constants c_1 and c_2. To find the latter it suffices to make a contraction in the last two formulae, integrate by parts and recall that $G(x)$ is just the Green function for the Laplacian. This procedure yields

$$c_1 = -\frac{1}{3} M_2(0) = -\frac{1}{3} \eta_1 \eta_2, \quad c_2 = \frac{1}{3} M_2(0) = \frac{1}{3} \eta_1 \eta_2,$$

and therefore

$$\mathsf{Z}_1 = -\frac{1}{3} \eta_1 \eta_2, \quad \mathsf{Z}_2 = \frac{1}{3} \eta_1 \eta_2 \tag{13}$$

(recall that we have assumed $\mathbf{E} : \mathbf{E} = 1$).

Note that in the statistically isotropic case under study $c_2 = \frac{1}{3} P_2$, where

$$P_2 = \iint \nabla\nabla G(y_1) : \nabla\nabla G(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2$$

is the two-point statistical parameter that appeared in the appropriate bounds on the conductivity coefficient [4] and the bulk modulus [3] of the mixture. The above simple reasoning is just the evaluation of this parameter done, e.g. in [3] (see Eq. (12) there). Hence

$$\mathsf{Z}_1 = -\frac{1}{3} P_2, \quad \mathsf{Z}_2 = \frac{1}{3} P_2, \quad P_2 = \eta_1 \eta_2.$$

As we shall demonstrate below, the rest of the two-point statistical parameters in (12) are also proportional to P_2.

Due to the statistical isotropy of the medium the integrals

$$\begin{gathered}
\int \nabla\nabla\nabla\nabla F(y)\, M_2(y)\, d^3y = c_3 \mathbf{H}, \\
\iint \nabla\nabla G(y_1) \otimes \nabla\nabla G(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 = c_4 \mathbf{H}, \\
\iint \nabla\nabla G(y_1) \cdot \nabla\nabla\nabla\nabla F(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 = c_5 \mathbf{H}, \\
\iint \nabla\nabla\nabla\nabla F(y_1) : \nabla\nabla\nabla\nabla F(y_2)\, M_2(y_1 - y_2)\, d^3y_1 d^3y_2 = c_6 \mathbf{H}
\end{gathered} \tag{14}$$

are fourth-rank fully symmetric isotropic tensors, thus proportional to the tensor $\mathbf{H}$, whose Cartesian components read

$$H_{ijkl} = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk},$$

with certain constants c_3 to c_6. (The fact that the integrals in (14) represent fully symmetric tensors is easily seen if appropriate integration by parts is performed.) Making a full contraction in (14) and integrating by parts, we find the needed constants to be

$$c_3 = -\frac{2}{15} M_2(0), \quad c_4 = \frac{1}{15} M_2(0), \quad c_5 = \frac{2}{15} M_2(0), \quad c_6 = \frac{4}{15} M_2(0),$$

so that the parameters U_1, T_2, V_2 and U_2 thus become simply:

$$
\begin{aligned}
\mathsf{U}_1 &= -\frac{4}{15}\eta_1\eta_2, \quad \mathsf{T}_2 = \frac{2}{15}\eta_1\eta_2, \\
\mathsf{V}_2 &= \frac{4}{15}\eta_1\eta_2, \quad \mathsf{U}_2 = \frac{8}{15}\eta_1\eta_2,
\end{aligned}
\tag{15}
$$

taking into account Eq. (2) as well.

The variational principle (7), together with (11), implies

$$2\mu^* \leq W[\widetilde{u}(x)] = A + 2B\alpha + C\alpha^2, \quad \forall\alpha,\ \forall\kappa. \tag{16}$$

In particular, at $\alpha = 0$, one has

$$\mu^* \leq \langle\mu\rangle \tag{17}$$

which, obviously, is the elementary (Voigt) bound on μ^*.

Next, optimizing the right-hand side of (16) with respect to α, one gets another estimate on μ^*:

$$2\mu^* \leq A - \frac{B^2}{C},$$

i.e.

$$\mu^* \leq \langle\mu\rangle - \frac{2[\mu]^2(\mathsf{Z}_1 + \kappa\mathsf{U}_1)^2}{C}, \tag{18}$$

having taken into account the expressions for A and B, see (11).

In (18) we have fixed the constant κ. The next stage is to optimize it with respect to κ. The resulting bound will be then just the Milton–Phan-Thien's one [9] on the effective shear modulus μ^*. If κ has the special value $\kappa = -1/\big(4(1-\nu_2)\big)$, see (23) below (so that the integrand in (9) is just the appropriate Green tensor), then (18) is the McCoy's bound on μ^*. This is obviously a three-point estimate since for its evaluation three-point statistical information — the correlations $M_3(y_1, y_2)$ — is needed in the three-point parameters Z_3, T_3, V_3 and U_3, see (12).

The main problem in specifying the bound (18) are the three-point parameters Z_3, T_3, V_3 and U_3, whose evaluation for special and realistic random constitution is clearly a nontrivial problem. Note that the first of these parameters, Z_3, is

$$\mathsf{Z}_3 = \frac{1}{3}P_3, \quad P_3 = \iint \nabla\nabla G(y_1) : \nabla\nabla G(y_2) M_3(y_1, y_2)\, d^3y_1 d^3y_2,$$

where P_3 is the three-point parameter that showed up in conductivity and bulk modulus bounding procedures [3], [4]. This parameter is simply connected to the so-called ζ-parameter of Torquato and Milton [12–14], see also [3, Eqs. (18), (19)].

In the variational reasoning of [3], [4] we have excluded the parameter P_3, using the fact that the appropriate three-point bounds should be more restrictive than the elementary ones, whatever the properties of the constituents. This fact led us to an inequality between P_3 and P_2. Here we shall employ the same procedure; though a certain additional three-point parameter (a linear combination of Z_3, T_3, V_3 and U_3) will show up, we shall obtain two inequalities for the two such parameters as a consequence of the fact that now we can vary more material properties, namely, the bulk and shear moduli of the constituents.

Indeed, the bound (18) should be at least as good as the elementary bound (17) (since the energy functional is minimized over a broader class of trial fields). This implies that

$$C > 0, \quad AC - B^2 \geq 0, \tag{19}$$

because $\mu^* \geq 0$. Since $A = 2\langle\mu\rangle > 0$, one has

$$C \geq B^2/A > 0,$$

which means that the second inequality in (19) is the stronger one. Using the expressions for A, B and C from (11), we can write the latter in the form

$$\begin{aligned}
&\langle\mu\rangle\left(\mathsf{Z}_2 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_2 + 4\kappa\mathsf{V}_2 + 2\kappa^2\mathsf{U}_2\right)\\
&+ [\mu]\left(\mathsf{Z}_3 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_3 + 4\kappa\mathsf{V}_3 + 2\kappa^2\mathsf{U}_3\right)\\
&+ (1 + 2\kappa)^2\Big(\langle k\rangle\,\mathsf{T}_2 + [k]\mathsf{T}_3\Big) - \frac{2[\mu]^2(\mathsf{Z}_1 + \kappa\mathsf{U}_1)^2}{\langle\mu\rangle} \geq 0.
\end{aligned} \tag{20}$$

The inequality (20) should hold for every "realistic" choice of the elastic moduli of the constituents (i.e. for which the appropriate elastic energy is positive-definite). This implies

$$-\eta_2\mathsf{T}_2 \leq \mathsf{T}_3 \leq \eta_1\mathsf{T}_2, \tag{21a}$$

$$\begin{aligned}
&\frac{2(\mathsf{Z}_1 + \kappa\mathsf{U}_1)^2}{\eta_2} - \eta_2\left(\mathsf{Z}_2 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_2 + 4\kappa\mathsf{V}_2 + 2\kappa^2\mathsf{U}_2\right)\\
&\leq \mathsf{Z}_3 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_3 + 4\kappa\mathsf{V}_3 + 2\kappa^2\mathsf{U}_3\\
&\leq \eta_1\left(\mathsf{Z}_2 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_2 + 4\kappa\mathsf{V}_2 + 2\kappa^2\mathsf{U}_2\right) - \frac{2(\mathsf{Z}_1 + \kappa\mathsf{U}_1)^2}{\eta_1}.
\end{aligned} \tag{21b}$$

Hence we have indeed *two* sets of inequalities for the three-point parameters that enter the bound (18). (And this is a consequence, let us underline once again, of the fact that *two* material properties have been varied *independently* — the bulk and shear moduli of the constituents.) Following the idea of [4], we can exclude the "bad" three-point quantities

$$\mathsf{T}_3 \quad \text{and} \quad \mathsf{Z}_3 + \frac{1}{3}(1 - 8\kappa - 8\kappa^2)\mathsf{T}_3 + 4\kappa\mathsf{V}_3 + 2\kappa^2\mathsf{U}_3$$

from this bound, by means of (21), thus replacing them by the two-point quantities already evaluated. Depending on the signs of $[\mu] = \mu_2 - \mu_1$ and $[k] = k_2 - k_1$, we should use to this end the upper or lower bounds (21).

For example, in the case $[\mu] \geq 0$ (i.e. $\mu_2 \geq \mu_1$) and $[k] \geq 0$ (i.e. $k_2 \geq k_1$), the upper bounds (21) are to be used, which results in the estimate

$$\mu^* \leq \langle\mu\rangle - 2[\mu]^2\eta_1\eta_2 \frac{16\kappa^2 + 40\kappa + 25}{p\kappa^2 + q\kappa + r}, \tag{22}$$

$$p = 160\mu_2 + 120k_2 - 32[\mu]\eta_2,$$
$$q = 160\mu_2 + 120k_2 - 80[\mu]\eta_2,$$
$$r = 85\mu_2 + 30k_2 - 50[\mu]\eta_2.$$

Optimizing this bound with respect to κ, we find[1]

$$\kappa_{\text{opt}} = -\frac{\mu_2 + 3k_2}{2(4\mu_2 + 3k_2)} = -\frac{1}{4(1-\nu_2)}, \quad \nu_2 = \frac{3k_2 - 2\mu_2}{2\mu_2 + 6k_2}, \tag{23}$$

so that ν_2 is the Poisson ratio of the phase '2'. The best bound on μ^* thus becomes

$$\mu^* \leq \langle\mu\rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_2(8\mu_2 + 9k_2)}{6(2\mu_2 + k_2)}}, \quad \text{if } \mu_2 \geq \mu_1 \text{ and } k_2 \geq k_1. \tag{24a}$$

The calculations in the rest of the cases are fully similar, so that only the final results will be given:

$$\mu^* \leq \langle\mu\rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_1(8\mu_1 + 9k_1)}{6(2\mu_1 + k_1)}}, \quad \text{if } \mu_2 \leq \mu_1 \text{ and } k_2 \leq k_1,$$

$$\mu^* \leq \langle\mu\rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_1(8\mu_1 + 9k_2)}{6(2\mu_1 + k_2)}}, \quad \text{if } \mu_2 \leq \mu_1 \text{ and } k_2 \geq k_1, \tag{24b}$$

$$\mu^* \leq \langle\mu\rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_2(8\mu_2 + 9k_1)}{6(2\mu_2 + k_1)}}, \quad \text{if } \mu_2 \geq \mu_1 \text{ and } k_2 \leq k_1.$$

In the so-called "well-ordered" case, when $(k_2 - k_1)(\mu_2 - \mu_1) > 0$, (24a) and the first of the estimates (24b) coincide with the Hashin-Shtrikman bound on μ^*, see [1]. The general "non-ordered" case was considered by Walpole [2]. It is easily seen that our bounds (24) are just the Walpole bounds [2, 15].

The derivation of the lower bound, corresponding to (24), is fully similar. In this case we write the elastic energy (7) as a functional of the stress tensor field:

$$\begin{aligned} W[\boldsymbol{\sigma}(x)] &= \langle \boldsymbol{\sigma}(x) : \mathbf{L}^{-1}(x) : \boldsymbol{\sigma}(x) \rangle \to \min, \\ \min W &= \boldsymbol{\Sigma} : \mathbf{L}^{*-1} : \boldsymbol{\Sigma}. \end{aligned} \tag{25}$$

[1] The right-hand side of (22) has one more extremum point, $\kappa = -5/4$, but it corresponds to its maximum value and hence is of no interest for us.

The functional W is considered over the class of trial fields, such that

$$\nabla \cdot \boldsymbol{\sigma}(x) = 0, \quad \langle \boldsymbol{\sigma}(x) \rangle = \boldsymbol{\Sigma}, \tag{26}$$

with a prescribed macrostress tensor $\boldsymbol{\Sigma}$, imposed upon the mixture.

The natural counterpart of the class (9) of trial stress fields for the functional W in (25) now reads

$$\begin{aligned}\mathcal{N}^{(1)} = \Big\{ \tilde{\boldsymbol{\sigma}}(x) \mid \tilde{\boldsymbol{\sigma}}(x) = \boldsymbol{\Sigma} + \alpha \Big[\chi_2'(y)\, \boldsymbol{\Sigma} \\ - (1+2\kappa)\,\mathbf{I} \left(\boldsymbol{\Sigma} : \int \nabla\nabla G(x-y) \chi_2'(y)\, d^3y \right) \\ + \kappa \boldsymbol{\Sigma} : \int \nabla\nabla\nabla\nabla F(x-y) \chi_2'(y)\, d^3y \\ + 2\kappa \operatorname{def} \left(\boldsymbol{\Sigma} \cdot \int \nabla\nabla G(x-y) \chi_2'(y)\, d^3y \right) \Big] \Big\},\end{aligned} \tag{27}$$

with deviatoric $\boldsymbol{\Sigma}$, $\operatorname{tr} \boldsymbol{\Sigma} = 0$, and adjustable scalar parameters α and κ; $G(x)$ and $F(x)$ are the functions, defined in (10). In (27) 'def' denotes symmetrization of a second-rank tensor, i.e. $\operatorname{def} \mathbf{T} = \frac{1}{2}(\mathbf{T} + \mathbf{T}^*)$, see [10] for discussion.

The energy functional W, when restricted over $\mathcal{N}^{(1)}$, becomes a quadratic function of α:

$$W[\tilde{\boldsymbol{\sigma}}(x)] = A + 2B\alpha + C\alpha^2, \;\; A = \frac{1}{2} \langle \gamma \rangle \boldsymbol{\Sigma} : \boldsymbol{\Sigma},$$

$$B = \frac{1}{2}[\gamma]\Big(\eta_1\eta_2 \boldsymbol{\Sigma} : \boldsymbol{\Sigma} + 2\mathsf{Z}_1 + \kappa \mathsf{U}_1 \Big),$$

$$\begin{aligned} C = \frac{1}{2}\Bigg[\Big(\eta_1\eta_2(\eta_1\gamma_1 + \eta_2\gamma_2) + \left(\frac{4}{3} + \frac{8}{15}\kappa \right) [\gamma](\eta_2 - \eta_1)\eta_1^2 \Big) \boldsymbol{\Sigma} : \boldsymbol{\Sigma} \\ + 2(\eta_1\gamma_2 + \eta_2\gamma_1)(2\mathsf{Z}_1 + \kappa\mathsf{U}_1) + \langle\gamma\rangle \left(2\mathsf{Z}_2 + \frac{1}{3}(2 - 8\kappa - 4\kappa^2)\mathsf{T}_2 + 4\kappa\mathsf{V}_2 + \kappa^2\mathsf{U}_2 \right) \\ + [\gamma]\Big(2\mathsf{Z}_3 + \frac{1}{3}(2 - 8\kappa - 4\kappa^2)\mathsf{T}_3 + 4\kappa\mathsf{V}_3 + \kappa^2\mathsf{U}_3 \Big) \Bigg] + \frac{(1+4\kappa)^2}{9}\Big(\langle\beta\rangle\, \mathsf{T}_2 + [\beta]\mathsf{T}_3 \Big), \end{aligned} \tag{28}$$

where

$$\gamma(x) = \frac{1}{\mu(x)}, \quad \beta(x) = \frac{1}{k(x)}$$

are the respective compliances of the mixture, and $[\gamma] = \gamma_2 - \gamma_1$, $[\beta] = \beta_2 - \beta_1$ are their jumps. Note that the *same* two- and three-point statistical parameters (12), that already showed up in (11), enter (28) as well; the only difference is that the tensor $\mathbf{E}$ in their definitions (12) is to be replaced by the tensor $\boldsymbol{\Sigma}$.

The elementary (Reuss) lower bound on μ^* now follows from (28) at $\alpha = 0$:

$$\frac{1}{\mu^*} \leq \left\langle \frac{1}{\mu(x)} \right\rangle = \frac{\eta_1}{\mu_1} + \frac{\eta_2}{\mu_2} .$$

Almost literally, the arguments that have led us to the inequality (20) are to be repeated now — that is the estimate that results from (28) upon minimizing with respect to α should be always more restrictive than the elementary Reuss' one. The final result is another set of inequalities, similar to (21), namely,

$$-\eta_2 \mathsf{T}_2 \leq \mathsf{T}_3 \leq \eta_1 \mathsf{T}_2, \tag{29a}$$

$$\begin{aligned}
&\frac{(\eta_1\eta_2 \boldsymbol{\Sigma} : \boldsymbol{\Sigma} + 2\mathsf{Z}_1 + \kappa \mathsf{U}_1)^2}{\eta_2 \boldsymbol{\Sigma} : \boldsymbol{\Sigma}} - \eta_2^2 \left[\eta_1 + \left(\frac{4}{3} + \frac{8}{15}\kappa\right)(\eta_1 - \eta_2\right] \boldsymbol{\Sigma} : \boldsymbol{\Sigma} \\
&-2\eta_1(2\mathsf{Z}_1 + \kappa \mathsf{U}_1) - \eta_2 \left(2\mathsf{Z}_2 + \frac{1}{3}(2 - 8\kappa - 4\kappa^2)\mathsf{T}_2 + 4\kappa \mathsf{V}_2 + \kappa^2 \mathsf{U}_2\right) \\
&\leq 2\mathsf{Z}_3 + \frac{1}{3}(2 - 8\kappa - 4\kappa^2)\mathsf{T}_3 + 4\kappa \mathsf{V}_3 + \kappa^2 \mathsf{U}_3 \qquad (29b)\\
&\leq -\frac{(\eta_1\eta_2 \boldsymbol{\Sigma} : \boldsymbol{\Sigma} + 2\mathsf{Z}_1 + \kappa \mathsf{U}_1)^2}{\eta_1 \boldsymbol{\Sigma} : \boldsymbol{\Sigma}} + \eta_2 \left[\eta_1^2 - \left(\frac{4}{3} + \frac{8}{15}\kappa\right)(\eta_1 - \eta_2)\eta_2\right] \boldsymbol{\Sigma} : \boldsymbol{\Sigma} \\
&+2\eta_2\Big(2\mathsf{Z}_1 + \kappa \mathsf{U}_1\Big) + \eta_1 \left(2\mathsf{Z}_2 + \frac{1}{3}(2 - 8\kappa - 4\kappa^2)\mathsf{T}_2 + 4\kappa \mathsf{V}_2 + \kappa^2 \mathsf{U}_2\right).
\end{aligned}$$

Next we employ (29) in the estimates that follow from (28) in order to exclude the three-point parameters. The details are tedious and fully similar to those, already performed when deriving the bounds (24). The final result reads

$$\begin{aligned}
\mu^* &\geq \langle \mu \rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_2(8\mu_2 + 9k_2)}{6(2\mu_2 + k_2)}}, \quad \text{if } \mu_2 \leq \mu_1 \text{ and } k_2 \leq k_1, \\
\mu^* &\geq \langle \mu \rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_1(8\mu_1 + 9k_1)}{6(2\mu_1 + k_1)}}, \quad \text{if } \mu_2 \geq \mu_1 \text{ and } k_2 \geq k_1, \\
\mu^* &\geq \langle \mu \rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_2(8\mu_2 + 9k_1)}{6(2\mu_2 + k_1)}}, \quad \text{if } \mu_2 \leq \mu_1 \text{ and } k_2 \geq k_1, \\
\mu^* &\geq \langle \mu \rangle - \frac{\eta_1\eta_2[\mu]^2}{\eta_1\mu_2 + \eta_2\mu_1 + \dfrac{\mu_1(8\mu_1 + 9k_2)}{6(2\mu_1 + k_2)}}, \quad \text{if } \mu_2 \geq \mu_1 \text{ and } k_2 \leq k_1.
\end{aligned} \tag{30}$$

The inequalities (24), combined with (30), are just the Walpole bounds on the effective shear modulus of a binary mixture, see [2] and also [15], which are a direct generalization of the Hashin-Shtrikman result, with condition of "well-orderness" removed. Here we have shown how these classical estimates show up simply and naturally within the frame of the general method recently developed by one of the authors [4] in the absorption and scalar conductivity contexts.

Acknowledgements. The support of the Bulgarian Ministry of Education and Science under Grant No MM 805-98 is gratefully acknowledged.

REFERENCES

1. Hashin, Z., S. Shtrikman. A variational approach to the theory of the elastic behaviour of multiphase materials. *J. Mech. Phys. Solids*, **11**, 1963, 127–140.
2. Walpole, L. On the overall elastic moduli of composite materials. *J. Mech. Phys. Solids*, **17**, 1969, 235–251.
3. Markov, K. Z., M. K. Kolev. A note on the bulk modulus of a binary elastic mixture. *Annuaire de l'Univ. de Sofia, Fac. de Math. et Inf., Livre 2 – Mécanique*, **92**, 1998, 145–150.
4. Markov, K. Z. On the correlation functions of two-phase random media and related problems. *Proc. Roy. Soc. Lond. A*, **455**, 1999, 1069–1086.
5. Beran, M. Statistical continuum theories. John Wiley, New York, 1968.
6. Brown, W. F. Solid mixture permittivities. *J. Chem. Phys.*, **23**, 1955, 1514–1517.
7. McCoy, J. J. On the displacement field in an elastic medium with random variations in material properties. In: *Recent Advances in Engineering Sciences*, **5**, A. C. Eringen, ed., Gordon and Breach, New York, 1970, 235–254.
8. Beran, M. Use of a variational approach to determine bounds for the effective permittivity of a random medium. *Nuovo Cimento*, **38**, 1965, 771–782.
9. Milton, G. W., N. Phan-Thien. New bounds on effective elastic moduli of two-components materials. *Proc. Roy. Soc. Lond. A*, **380**, 1982, 305–331.
10. Markov, K. Z., K. D. Zvyatkov. Optimal third-order bounds on the effective properties of some composite media, and related problems. *Adv. in Mech. (Warsaw)*, **14**, No 4, 1991, 3–46.
11. Beran, M., J. Molyneux. Use of classical variational principles to determine bounds for the effective bulk modulus in heterogeneous medium. *Q. Appl. Math.*, **24**, 1966, 107–118.
12. Torquato, S. Microscopic approach to transport in two-phase random media. Thesis, State University of New York at Stony Brook, 1980.
13. Milton, G. W. Bounds on the electromagnetic, elastic and other properties of two-component composites. *Phys. Review Lett.*, **46**, 1981, 542–545.
14. Torquato, S. Random heterogeneous media: Microstructure and improved bounds on effective properties. *Appl. Mech. Rev.*, **44**, 1991, 37–76.

15. Walpole, L. Elastic behavior of composite materials: Theoretical foundations. In: *Adv. Appl. Mech.*, C. Yih, ed., vol. 21, Academic Press, New York, 1981, 169–242.

Received February 15, 2000

Mikhail K. KOLEV
Faculty of Mathematics and
Informatics, South-West University
Blagoevgrad, Bulgaria
E-mail: `mkkolev@aix.swu.bg`

Konstantin Z. MARKOV
Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
BG-1164 Sofia, Bulgaria
E-mail: `kmarkov@fmi.uni-sofia.bg`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

AN EXPLICIT POLYNOMIAL SOLUTION OF THE REPRESENTATIVE PROBLEM OF THE MECHANICS OF FIBROUS COMPOSITES

IVAN M. MIHOVSKY

The paper is focused on an important aspect of the central problem of the mechanics of the unidirectionally reinforced fibrous composites. This problem concerns the features of the matrix-fibre load transfer phenomenon which, actually, provides by itself the very reinforcing effect of the fibres. The paper illustrates that the successful analysis of this problem definitely needs, first, the exact solution of a certain typical or, say, representative axisymmetric boundary value problem of the elasticity theory and, second, a representation of this solution in a form, which is convenient enough for additional mathematical manipulations. Such an explicit with respect to the problem variables polynomial representation is derived in the paper.

Keywords: fibrous composites, elasticity, special functions

MSC 2000: 74E30, 74B05, 33C90

1. INTRODUCTION

A specific feature of the present contribution is that the purely mathematical result derived in it is, in practice, a useful and effective tool from the view point of the mechanics, i.e. of the study of the mechanical properties and the mechanical behaviour of an important for the engineering practice class of advanced structural materials, namely the unidirectionally reinforced fibrous composites. For this reason the analysis of the mathematical problem considered is preceded by a relatively extended preliminary section, the aim of which is to indicate at least the main problems associated with the mechano-mathematical modelling of the unidirectional fibrous composites and to illustrate, at the same time, the necessity and the practical value of the result obtained.

2. PRELIMINARIES

Composite materials consist, generally speaking, of a relatively weak and compliant continuous phase within which inclusions of different, most often stiff and strong materials, are discretely distributed and, as a rule, firmly bound to the surrounding continuous matrix phase. Depending on the material, the shape, the volume fraction and the geometry of the spatial distribution and mutual orientation of the inclusions, composite properties and, especially, their mechanical properties may differ considerably and even drastically from the respective properties of the matrix material. The strong effect that such inclusions produce on the properties of the matrix has been used since long in the fabrication of new structural materials with desired in advance unique combinations of mechanical properties. Basically, the goal of the practical use of this effect is to create materials with high strength and stiffness parameters, i.e. to reinforce the weak materials used as matrix phases. For this reason the effect is commonly referred to as a *reinforcing effect*.

This effect is especially pronounced in the so-called *fibre-reinforced composites*. The discrete reinforcing phase in these composites presents itself one or another network of rods, wires or whiskers, which are usually called *fibres*. An important for the practice class of such composites is that of the *unidirectionally reinforced fibrous composites*, i.e. of composites with straight parallelly aligned fibres. In most of the cases the fibres have the form of circular cylinders with a specific for each given composite radius-to-length ratio, which is often called fibre *slenderness ratio*.

The unidirectional composites are anisotropic or, speaking more precisely, transversely isotropic materials. They may have high strength and stiffness in the direction of the reinforcing fibres, but remain weak and compliant, as the matrix is, in the transverse direction. The practice proves that the degree of anisotropy or, which is the same, the strength of the very effect of the unidirectional fibrous reinforcement, depends on two basic structural parameters, namely the fibre volume fraction and the fibre slenderness ratio.

The strength and the stiffness of the unidirectional composites in their, say, strong direction, which is the direction of the reinforcing fibres, appear to be almost directly proportional to the fibre volume fraction. The simple "*rule of mixtures*"-type relations, used in the engineering practice, prove to be reliable quantitative approximations to the actual effect that the fibre volume fraction produces on the degree of strengthening and stiffenning of the matrix material.

The fibre slenderness ratio influences the strength and stiffness characteristics of the unidirectional composites in their strong direction in a more complicated way. In composites, which differ only in the lengths of the reinforcing fibres, but are otherwise identical in all remaining material and structural parameters, including fibre radii, those with longer fibres, i.e. with smaller slenderness ratios, are stronger and stiffer. A distinct feature of the practically observed dependencies of the strengthening effect on this ratio is that the decrease in the latter provides a respective increase in this effect only over certain restricted and specific for each composite structure ranges of increasing fibre lengths.

Briefly speaking, the unidirectional composites are characterized by specific slenderness ratios or, at fixed fibre radii, by specific fibre lengths. When the value of the specific for a given composite slenderness ratio is approached from above by means, for example, of a continuous increase of the fibre lengths at fixed fibre radii, the reinforcing effect continuously increases up to a specific for the composite maximum level and then, with further fibre lengths increase, this effect remains practically unchanged. The specific or, say, the *critical fibre length* upon which the reinforcing effect becomes insensitive to further fibre length increase, proves to be a distinct inherent characteristic of the unidirectional composites. Practical observations indicate that it is a complex and, as a matter of fact, still not quite well-known function of the material parameters of the composites. The existing experimental data certify that the critical fibre lengths for different fibre-matrix systems vary from a few to tens and even hundreds fibre radii. This fact is a clear indication that the theoretical approaches to the problem of determining the critical fibre lengths should not be based on rough initial physical models, neither should involve, from a mathematical view point, rough simplifications and approximations.

From its purely qualitative side the existence of such a critical fibre length is easily understandable. The point is that in composites loads are not directly applied to the fibres, but to the matrix into which they are embedded. The role of the matrix is, besides to serve as a binding medium, also to transfer loads from composite surfaces, where the loads are applied, to the fibres. This *matrix-fibre load transfer* is, in reality, the essence of the very mechanism of realization of the reinforcing effect. Due to this mechanism the strong and stiff fibres take or absorb and, respectively, carry most of the loads applied, as mentioned, to the matrix. This actually explains the high load bearing capacities of the fibrous composites, i.e. their potential to carry loads that considerably exceed the restricted load bearing capacities of the weak matrix materials.

Obviously enough, the matrix transfers loads to the fibres through their end and cylindrical surfaces. Due to a number of reasons the end fibre surfaces have little effect on the overall load transfer pattern. One of these reasons is that the area of the end fibre surfaces is much smaller than that of the cylindrical lateral surfaces. Therefore, the contribution of fibre ends to the load transfer should not be expected to be comparable to that of the cylindrical surfaces. The increase in the area of fibre end surfaces, which is achievable by preparing fibres with oval instead of flat ends, proves practically not to affect the load transfer features. This is due to the fact that, mainly for technological reasons, the fibre-matrix bonds at fibre ends are not of the same necessarily high quality as those over the cylindrical fibre surfaces. Thus, irrespectively of whether fibre ends are flat or oval, the stresses developing over the weak fibre ends-matrix interfaces prove to be simply small enough to play a more or less decisive role in the load transfer processes. That is why in most of the studies of the load transfer phenomena fibre ends are viewed as stress free.

It would be probably instructive to note, before considering the load transfer through the cylindrical fibre surfaces, that fibrous composites are designed mainly to carry loads in the reinforcement direction, i.e. in the axial fibre direction. The transfer of such loads is evidently due to the axial interfacial shear stresses developing as a result of the mechanical fibre-matrix interaction. To illustrate this, it would be sufficient just to consider the equilibrium in the axial direction of an arbitrary fibre portion and immediately to notice that these are exactly the axial interfacial shear stresses that balance the axial loads acting at the ends of such a fibre portion.This fact demonstrates that the study of the load transfer should be first of all focused on the problem of determining the longitudinal (i.e. along the fibre length) distribution of the axial interfacial shear stresses.

Coming closer to this problem, let us remind once more that the increase of fibre lengths up to certain specific critical values results in higher load bearing capacities or that, briefly speaking, longer fibres absorb and, respectively, carry higher axial loads. In part, this effect is certainly due to the fact that since the fibre length increase is also increase in the area of the cylindrical fibre surfaces, then the axial interfacial shear stresses, when acting over increased areas, should introduce into the fibres and, respectively, balance increased axial loads. At the same time one should not exclude as a reason for this effect the eventual changes and, in particular, the possible increase of the intensities of the same shear stresses. Moreover, the very distribution along the fibre length of these stresses may change and thus have also effect on the level of the axial loads transferred from the matrix to the fibre.

It is hoped that the foregoing considerations reveal to a certain extent the complex multiaspect nature of the load transfer problem. In fact, this central for the mechanics of the unidirectional composites problem has been since long a subject of extended research. The result of this research is the number of *load transfer models* that the current composite materials literature offers. Without a discussion on their positive and negative sides it will be mentioned here that all these models are practically similar in a few basic aspects.

First, they are based on the experimentally observed fact that the mechanical behaviour, mainly in the reinforcement direction, of a typical unidirectional composite is very close to that of a representative composite element, say, *unit composite cell*, which, being composed of a single fibre with a firmly bound coaxial matrix coating, is stressed axisymmetrically and, in addition, symmetrically with respect to its middle cross-section.

Second, the models reflect, but rather in a qualitative than in a more or less satisfactory quantitative manner, the concentration of stresses close to the ends of the fibres where, as a result and a manifestation of the *stress concentration*, local failure phenomena, mainly in the form of fibre-matrix debonding (i.e. of interfacial cracks), often take place. In this regard the similarity of the models lies in the fact that they consider as a major reason for these phenomena mainly the concentration of the interfacial axial shear stresses. Accordingly, they deal, basically, with the problem of deriving the patterns of the shear stress concentration fields close to the fibre ends.

Third, the shear stress concentration patterns that the models provide are derived from either simple strength of materials-type analyses or from too much simplified axisymmetric boundary value problems of the elasticity (or plasticity) theory. The simplifications concern both the mathematical treatments of the equations governing these problems and the mechano-mathematical models of the actual physical boundary conditions, especially the conditions for the stresses at the fibre ends. These conditions will be discussed below.

Finally, the models generally assume that no interfacial shear stresses develop along most of the fibre lengths and suggest one or another specific monotonous increase of the intensities of these stresses along the remaining relatively short end fibre portions. But, strictly speaking, such simple distributions of stresses are just inconsistent with the elasticity theory since the equations of the latter are, as it is known, of elliptic type. Accordingly, the distributions of fibre stresses, including the interfacial shear stresses, might be either uniform along the entire fibre length or strictly non-uniform along each non-vanishingly short fibre portion. Thus, from a quantitative view point, such simple axial stress distributions are acceptable only in the sense of the Saint Venant's principle and as approximations to certain particular exact solutions of the general axisymmetric elasticity problem. Since these solutions, i.e. the exact solutions of the particular boundary value problems associated with the models, are not known, one could hardly derive estimates of the errors that such approximations involve and decide, respectively, about the reliability of the models.

Irrespectively of their roughness and simplicity, these models are widely used in the practice of composites design and application since they provide quantitative estimates of the load transfer parameters and, thus, of the load bearing potentials of the fibrous composites, which, even if not realistic enough in the details, are quite acceptable as mean or integral estimates.

Another positive side of the models is that they have revealed in part the specific nature of the above mentioned *nontrivial dependence of the reinforcing* effect on the lengths of the fibres. The models definitely indicate that the separation of a fibre into a central uniformly stressed portion, which only carries axial load, but does not take load directly from the surrounding matrix, and neighbouring this portion with relatively short end portions, along which the matrix transfers load to the fibre, are quite realistic. This separation of the fibre into a load bearing, or, effective and load transfer, or, ineffective portions is commonly adopted today. In view of these models the main problem of the load transfer analysis is to determine the current load transfer fibre length as a function of both the fibre-matrix system and the current load, applied to the composite, and, in addition, to specify the characteristic for the composite critical maximum load transfer length, now as a function of the fibre-matrix system only. If this critical length is specified, then, in order to gain maximum load bearing effect from a given unidirectional fibre-matrix system, one should use fibres of length which is greater than or at least equal to twice the critical length. Such fibres are usually referred to as *long fibres*. Fibres of lengths below the critical length are viewed as *short fibres*. The respective composites are usually called composites with long, or continuous and short, or discontinuous fibres.

It is not surprising that the models suggest estimates of the critical load transfer lengths which differ considerably. These estimates depend on the different in their roughness approximations involved in each particular model. There is a number of reasons for which these estimates should not be viewed as sufficiently precise though, as was mentioned above, they are acceptable in integral sense. In part, this is due to the fact that, as was also mentioned, the very concept of specifying effective and ineffective fibre portions is a matter of approximate interpretation of an exact solution of the axisymmetric elasticity problem. This interpretation implicitly assumes that the exactly determined interfacial shear stresses, acting along the central (effective) fibre portion, are small, i.e. practically negligible, and that they monotonously increase towards fibre ends along the load transfer (ineffective) fibre portions. But such an interpretation, being, let us mention again, acceptable on the whole, necessarily requires, first of all, to specify the level below which the stresses in question could be viewed as negligible. The models do not suggest such levels. Moreover, they can not specify the latter since such specifications require by themselves as a reference basis the unknown exact stress distributions.

The basic limitation of the existing models is that they take practically no account of the actual boundary conditions at the fibre ends. Generally, they assume that high interfacial shear stresses develop close to fibre ends, but entirely ignore the corresponding considerable drop that these stresses should necessarily undergo along certain end fibre portions, adjacent to fibre ends, due to the fact that these ends are practically almost free of shear stresses. The models simply ignore these end portions as elements of the load transfer pattern.

Two remarks are due with respect to this fact. First, from the view point of the pure load transfer, the ignorance of these portions is not an important disadvantage since, from most general considerations, the latter should be expected to be short enough and thus not to cause considerable corrections in the estimates of the load transfer lengths. Second and more important is that the models ignore the entire complex interfacial stress field developing along these end fibre portions. This specific stress field has been constructed in a recent author's paper [1], where some of the results of an extended author's study in progress are briefly reported. As it is shown in the paper, this field develops along end fibre portions, which are really short. Their lengths do not exceed a few fibre radii, which proves that these portions really could not contribute considerably to the overall load transfer. But at the same time it is definitely proved in the paper that this stress concentration field is the factor that actually governs the onset of the failure phenomena close to the fibre ends. In fact, the paper illustrates, first of all, that these phenomena are not governed, as the considered models suggest, by the interfacial shear stress concentration, but by the much higher concentration of the interfacial radial stresses. Moreover, as it is shown in the paper, the radial interfacial stress changes its sign along the short end fibre portions. In other words, this high by itself stress is always positive, i.e. tensile, along certain parts of these portions. Obviously, this high tensile interfacial radial stress is the factor dominating the critical conditions of onset of failure phenomena close to the fibre ends.

The load transfer model proposed in [1] could be referred to as a *full fibre length model* since, in contrast to the models considered, it involves the very end fibre portions which these models ignore. Being actually derived by means of sewing of two exact axisymmetric solutions of the elasticity theory, the model is not only free from the shortcomings of the existing simple models, but, in view of the considerations of the present section, is a real reference basis on which one could estimate the reliability of these approximate models.

It is hoped that this section clearly indicates that the general axisymmetric problem of the theory of elasticity is of immediate relevance to the mechanics of the unidirectional fibrous composites. The realistic modelling and prediction of the properties and the behaviour of these composites definitely require not only the derivation of exact solutions of certain particular, but similar on the whole axisymmetric boundary value elasticity problems. In fact, it is equally important to be easily able to manipulate further such solutions in order, for example, to construct superpositions, to derive approximations with desired accuracy, to estimate the accuracy of certain existing approximations or, as in [1], to sew such exact solutions.

It would be very advantageous, for these reasons, to have at one's disposal a convenient, i.e. an easy for mathematical manipulations, form of the exact solution of a certain, say, *representative problem* of the mechanics of the unidirectional fibrous composites. In what follows, this problem is first formulated and its solution is then shown to be representable indeed in a really convenient explicit analytical polynomial form.

3. THE REPRESENTATIVE AXISYMMETRIC PROBLEM

The commonly adopted representative model problem for a typically loaded long reinforcing fibre assumes that the state of stress of the latter is, as was pointed out above, axisymmetric and, in addition, symmetric with respect to its middle cross section. Basing upon the superposition principle of the elasticity theory and on the standard assumption that fibre ends are flat and free from stresses, one may further specify the fibre stress state as resulting from uniformly distributed normal (tensile or compressive) stresses of intensity, say σ_0, applied to its ends (or acting over its middle cross section), and from interfacial stresses, developing along the lateral cylindrical fibre surface as a result of the mechanical fibre-matrix interactions, caused by the mismatch of the mechanical properties of the fibre and matrix materials.

The fibre, when referred to cylindrical coordinates $\{r, \theta, z\}$, is assumed to occupy the domain $\{0 \le r \le r_f,\ |z| \le L,\ 0 \le \theta \le 2\pi\}$, where, obviously, r_f and L are fibre radius and half length, respectively. Then the boundary conditions of the representative problem could be specified as

$$\tau_{rz}(r, 0) = \tau_{rz}(r, \pm L) = 0, \tag{1}$$

$$\sigma_z(r, \pm L) = \sigma_0, \quad (\text{or} \quad \sigma_z(r, 0) = \sigma_0), \tag{2}$$

where $\tau_{rz}(r,z)$ and $\sigma_z(r,z)$ are the shear and the axial fibre stresses, respectively. The notations $\sigma_r(r,z)$ and $w(r,z)$ will be used below for the radial fibre stress and the axial displacement in the fibre.

Obviously, the remaining boundary conditions, concerning the stresses $\tau_{rz}(r_f, z)$ and $\sigma_r(r_f, z)$, acting over the cylindrical fibre surface $r = r_f$, could not be specified in advance. In fact, to determine these unknown stresses is, as was highlighted above, the sense of the central for the mechanics of the unidirectional composites problem of the matrix-fibre load transfer.

The approach to this problem is based on the understanding, developed in the mentioned author's study in progress and briefly described in [1], according to which the role of the weak and compliant matrix is rather to conduct the applied loads to the surface of the stiff and strong fibre than to influence considerably the specific and in much independent manner in which the fibre absorbs these loads through its surfaces, transforms them into internal stresses, and creates its own stress distribution pattern of the above discussed type. Such understanding of the dominant role of the fibre, or, of its more or less independent behaviour, suggests that the actual fibre state could be interpreted as an optimum or, say, a *natural* one, or, in other words, as a state corresponding to a solution of a certain variational problem of the elasticity theory to which the interfacial stresses in question serve as natural boundary conditions.

This understanding is further combined with the concept of the stress function (or, stress potential) as a function which, once introduced as a solution of the *general variational problem* of the elasticity theory, provides full exact solutions of each particular boundary value problem when subject to the respective boundary conditions. Use is made of the known representations of the stresses and displacements by means of the stress function for the general axisymmetric elasticity problem. For the quantities of interest in the present study these representations read, cf. [2],

$$\sigma_z = \frac{\partial}{\partial z}\left[(2-\nu)\Delta\varphi - \frac{\partial^2\varphi}{\partial z^2}\right], \tag{3}$$

$$\tau_{rz} = \frac{\partial}{\partial r}\left[(1-\nu)\Delta\varphi - \frac{\partial^2\varphi}{\partial z^2}\right], \tag{4}$$

$$\sigma_r = \frac{\partial}{\partial z}\left(\nu\Delta\varphi - \frac{\partial^2\varphi}{\partial r^2}\right), \tag{5}$$

$$w = \frac{1}{2G}\left[2(1-\nu)\Delta\varphi - \frac{\partial^2\varphi}{\partial z^2}\right], \tag{6}$$

where $\varphi(r,z)$ is the stress function ($\Delta^2\varphi(r,z) = 0$), Δ is the Laplace operator, ν and G are, respectively, the Poisson's ratio and the shear modulus of the material considered or, in our case, of the fibre material.

Finally, use is also made of the productive Timoshenko's idea (cf.[2]) of representing the particular solutions for the stress function in terms of Legendre polynomials of the first kind $P_n(x)$, $n = 1, 2, 3, \ldots$, $x = z/R$, $R = \sqrt{r^2 + z^2}$. According to this representation each stress function of the form

$$\varphi_{2p+1}(r, z) = A_{2p+1} R^{2p-2} P_{2p-2}(x) + B_{2p+1} R^{2p-2} P_{2p-4}(x), \tag{7}$$

where $p = 1, 2, 3, \ldots$, A_{2p+1} and B_{2p+1} are constants to be determined from the boundary conditions, provides an axisymmetric stress state with stresses $\sigma_z(r, z; p)$ and $\tau_{rz}(r, z; p)$, which, as the representative problem requires, besides being axisymmetric, are, respectively, even and odd functions of z, i.e. symmetric and antisymmetric with respect to the middle cross-section of the fibre.

The practical realization of this approach to the representative model problem considered implies the following forms of the quantities of interest:

$$\sigma_z(r, z; p) = \sigma_0 \left(\frac{R}{L}\right)^{2p-2} x P_{2p-3}(x), \tag{8}$$

$$\begin{aligned} \tau_{rz}(r, z; p) = & -\sigma_0 \left(\frac{R}{L}\right)^{2p-2} \left\{ \frac{p(2p-3)}{(p-1)(2p-1)} \frac{x}{\sqrt{1-x^2}} \right. \\ & \left. \times \Big[P_{2p-4}(x) - x P_{2p-3}(x) \Big] + \frac{1}{2p-1} \sqrt{1-x^2} P_{2p-3}(x) \right\}, \end{aligned} \tag{9}$$

$$\begin{aligned} \sigma_r(r, z; p) = -\sigma_0 \left(\frac{R}{L}\right)^{2p-2} & \left\{ \frac{2(p+1-\nu)}{(2p-1)(2p-2)} \frac{1}{1-x^2} \Big(P_{2p}(x) - x P_{2p-1}(x) \Big) \right. \\ & \left. + \frac{1}{4p-3} \left[(2p-3) P_{2p-4}(x) + 2x P_{2p-3}(x) \right] + \frac{4p^2+4p-5}{(4p-3)(2p-2)} P_{2p-2}(x) \right\}, \end{aligned} \tag{10}$$

$$\begin{aligned} w(r, z; p) = & \frac{\sigma_0 L}{2G} \left(\frac{R}{L}\right)^{2p-1} \frac{1}{(2p-1)(2p-2)(4p+1)} \\ & \times \left[\left(4p^2 - 8p + 8p\nu + 2\nu + \frac{4(2p-1)^2}{4p-3} \right) P_{2p-1}(x) \right. \\ & \left. + \frac{(2p-1)(2p-2)(4p+1)}{4p-3} P_{2p-3}(x) \right]. \end{aligned} \tag{11}$$

In fact, the model proposed in [1] is a result of a procedure of sewing of two solutions of the type presented by Eqs. (8) — (11) with their own appropriately chosen p-indices. Each of these solutions satisfies a pair of boundary conditions of the type of Eqs. (1) and (2) and is thus a solution of the representative axisymmetric problem. One of the solutions concerns the central fibre portion, say $|z| < l$, $l < L$, while the other one governs the end fibre portions $l \leq |z| \leq L$, where, obviously, $L - l$ is the length of the latter fibre portions. These two solutions are sewed over the cross sections $|z| = l$ of the fibre.

The closed analytical form of the solution of the general representative problem, Eqs. (8) — (11), has its definite advantages, but, at the same time, this form is easily seen to be not convenient for further mathematical manipulations which one is necessarily forced to perform for one or another reason. Such manipulations are, for example, unavoidable part of the further elastic analysis of the complete problem of determining the stress-strain state of the entire unit composite cell (of which the fibre is only an element). The difficulties, arising when this form is eventually subject to further manipulations, result, on the one hand, from the fact that Eqs. (8) — (11) are implicit with respect of the problem variables r and z. On the other hand, the known analytical forms of the Legendre polynomials are by themselves complicated enough. To reduce to some extent these difficulties, most of which the author met during his work on [1], was the author's motive to try to derive a simpler representation of the latter equations, namely an explicit with respect to the variables r and z polynomial representation.

It should be recognized that the work on [1] and, in particular, the analysis of the axial fibre stress, Eq. (8), gave the author the hint for the form of the simple representation derived below. Unexpectedly, it appeared that to prove the generalization of this form was not a trivial combinatorics problem. As the reminder part of this contribution illustrates, the use of some familiar special functions provides a short and effective way to this generalization.

4. AXIAL FIBRE STRESS

The desired simplification of the axial fibre stress representation concerns, obviously, only the term $\sigma_z(r, z \, ; \, p)/\sigma_0$ in Eq. (8). Upon introducing dimensionless coordinates ρ, ζ this term takes, in accordance with one of the standard representations of the Legendre polynomials $P_n(x)$, cf. [3], the form

$$\frac{\sigma_z(\rho, \zeta; p)}{\sigma_0} = \zeta^{2p-2} \sum_{k=0}^{p-2} \frac{(-1)^k (4p - 2k - 6)!}{2^{2p-3} k!(2p - k - 3)!(2p - 2k - 3)!} \left(1 + \frac{\rho^2}{\zeta^2}\right)^k, \qquad (12)$$

where $(\ldots)!$ stays, as usual, for the factorials of the numbers in brackets.

The sum in Eq. (12) is a polynomial of ρ^2/ζ^2. Let this sum be denoted by Σ_z and written in the form

$$\Sigma_z = \sum_{m=0}^{p-2} \lambda_m \left(\frac{\rho^2}{\zeta^2}\right)^m. \qquad (13)$$

Provided a more or less convenient representation of the coefficients λ_m, $m = 1, 2, \ldots, p-2$, is found, the axial fibre stress will take the explicit polynomial form

$$\sigma_z(\rho, \zeta; p) = \sigma_0 \sum_{m=0}^{p-2} \lambda_m \rho^{2m} \zeta^{2p-2m-2}, \qquad (14)$$

which is definitely more informative and easier to deal with than the form suggested by Eq. (12) suggests.

It will be mentioned, before deriving the λ_m-presentation in question, that the solutions of the representative problem, which are of interest both from a mechanical view point and in the context of the present contribution, are those with relatively large p-values. Solutions with small p-values are physically unacceptable for the load bearing central fibre portion. As it is shown in [1], superpositions of such solutions govern, in actual fact, the state of stress of the very end fibre portions, so that they have their definite mechanical meaning. But, at the same time, as Eq. (12) proves, their explicit polynomial presentation is a matter of trivial transforms.

According to Eq. (8), the term $\sigma_z(\rho, \zeta; p)/\sigma_0$ here considered is

$$\frac{\sigma_z(\rho, \zeta; p)}{\sigma_0} = \zeta^{2p-2}\left(1+\frac{\rho^2}{\zeta^2}\right)^{\frac{2p-3}{2}} P_{2p-3}\left[\left(1+\frac{\rho^2}{\zeta^2}\right)\right]. \tag{15}$$

The general representation of the Legendre polynomials by means of the hypergeometric function $F(a, b; c; d)$ will be used below. According to this representation, cf. [3],

$$P_{2p-3}\left[\left(1+\frac{\rho^2}{\zeta^2}\right)^{-\frac{1}{2}}\right] = \frac{(4p-7)!!}{(2p-3)!!}\left(1+\frac{\rho^2}{\zeta^2}\right)^{-\frac{2p-3}{2}} F(a, b; c; d), \tag{16}$$

where, in our case, $a = -p + \frac{3}{2}$, $b = -p + 2$, $c = -2p + \frac{7}{2}$, $d = 1 + \frac{\rho^2}{\zeta^2}$, and $(\ldots)!!$ are double factorials.

Now, in view of Eqs. (12) — (16) and of the explicit form of $F(a, b; c; d;)$ with negative integer variable $b = -p + 2 < 0$, namely, cf. [4],

$$\begin{aligned} &F\left(-p+\frac{3}{2}, -p+2; -2p+\frac{7}{2}; 1+\frac{\rho^2}{\zeta^2}\right) \\ &= \sum_{k=0}^{p-2} \frac{\left(-p+\frac{3}{2}\right)_k (-p+2)_k}{\left(-2p+\frac{7}{2}\right)_k k!}\left(1+\frac{\rho^2}{\zeta^2}\right)^k, \end{aligned} \tag{17}$$

the coefficients λ_m, $m = 0, 1, \ldots, p-2$, take the form

$$\lambda_m = \frac{(4p-7)!!}{(2p-3)!!}\sum_{k=m}^{p-2} \frac{\left(-p+\frac{3}{2}\right)_k (-p+2)_k}{\left(-2p+\frac{7}{2}\right)_k k!}\binom{k}{m}, \tag{18}$$

where the Pochhammer symbol $(a)_i = a(a+1)(a+2)\ldots(a+i-1)$, $i = 1, 2, \ldots$, is used along with the binomial coefficients representation

$$\binom{k}{m} = \frac{k!}{m!(k-m)!} = \frac{(k-m+1)_m}{m!}. \tag{19}$$

Upon setting $s = k - m$ Eq. (18) takes, due to the obvious relation $(a)_{i+j} = (a)_i(a+i)_j$, the form

$$\lambda_m = \frac{(4p-7)!!}{(2p-3)!!} \frac{\left(-p+\frac{3}{2}\right)_m (-p+2)_m}{\left(-2p+\frac{7}{2}\right)_m m!} \times \sum_{s=0}^{p-m-2} \frac{\left(-p+m+\frac{3}{2}\right)_s (-p+m+2)_s}{\left(-2p+m+\frac{7}{2}\right)_s s!}. \tag{20}$$

The sum in Eq. (20) is easily seen to be, by definition (cf. Eq. (17)), equal to $F(-p+m+\frac{3}{2}, -p+m+2; -2p+m+\frac{7}{2}; 1)$ and to be thus, due to the relation, cf. [4], $F(a,b;c;d) = \Gamma(d)\Gamma(d-a-b)/\Gamma(d-a)\Gamma(d-b)$, representable as $\Gamma(-2p+m+7/2)\Gamma(-m)/\Gamma(-p+2)\Gamma(-p+3/2)$, where Γ denotes the well-known Gamma-function.

With the latter form and the known representations of the Γ-function Eq. (20) implies upon due transforms and manipulations the following compact expressions for the λ_m-coefficients, $m = 0, 1, 2, \ldots, p-2$:

$$\lambda_m = (-1)^m \frac{(2m-1)!!}{(2m)!!} \binom{2p-3}{2m}. \tag{21}$$

Other equivalent representations of the λ_m-coefficients become now available from Eq. (21) as, for example,

$$\lambda_m = (-1)^m \frac{1}{2^{2m}} \binom{2p-3}{2m} \binom{2p-m-3}{m}. \tag{22}$$

With Eq. (21) the final desired explicit in the $\{\rho, \zeta\}$-variables polynomial representation of the axial fibre stress becomes

$$\sigma_z(\rho, \zeta; p) = \sigma_0 \sum_{m=0}^{p-2} (-1)^m \frac{(2m-1)!!}{(2m)!!} \binom{2p-3}{2m} \rho^{2m} \zeta^{2p-2m-2}. \tag{23}$$

5. SHEAR FIBRE STRESS

The explicit polynomial representation of the shear fibre stress is, of course, derivable in a way, similar to the one just used for the axial stress. But it would be much simpler just to introduce Eq. (23) for the axial stress into the equilibrium equation

$$\frac{\partial \sigma_z}{\partial z} + \frac{1}{r}\frac{\partial}{\partial r}(r\tau_{rz}) = 0 \tag{24}$$

and to get almost immediately, upon satisfying the axial symmetry condition $\tau_{rz}(0, \zeta; p) = 0$, the form

$$\tau_{rz}(\rho, \zeta; p) = -\sigma_0 \sum_{m=0}^{p-2} \lambda_m \frac{p-m-1}{m+1} \rho^{2m+1} \zeta^{2p-2m-3}, \tag{25}$$

where λ_m are the same coefficients as in Eq. (21) (or Eq. (22)).

6. RADIAL FIBRE STRESS

As the structure of Eqs. (5), (6) suggests, the derivation of the explicit polynomial forms for the radial fibre stress σ_r and the axial displacement w is only a matter of further, mainly technical manipulations and transforms of the type already considered for the axial fibre stress. Omitting the details they involve, these manipulations reduce Eq. (10) to the form

$$\sigma_r(\rho, \zeta; p) = -\sigma_0 \sum_{m=0}^{p} \mu_m \rho^{2m-2} \zeta^{2p-2m}, \tag{26}$$

where the coefficients μ_m, $m = 1, 2, \ldots, p$, read

$$\begin{aligned} \mu_m = (-1)^m \frac{(2m-1)!!}{(2m)!!} \frac{p+1-\nu}{(2p-1)(p-1)} \frac{m}{p} \binom{2p}{2m} \\ - (-1)^m \frac{(2m-3)!!}{(2m-2)!!} \frac{2p-2m+3}{2p-2} \binom{2p-2}{2m-2}. \end{aligned} \tag{27}$$

Note that due to the standard convention $\binom{a}{b} = 0$, when $b < 0$, the term involving the multiplier $\binom{2p-2}{2m-2}$ in μ_0 is zero, so that the coefficient μ_0 is itself zero and the summation in Eq. (26) starts, practically, with $m = 1$.

7. AXIAL DISPLACEMENT

In a similar manner the axial displacement could be reduced to the form

$$w(\rho, \zeta; p) = \frac{\sigma_0 L}{2G(4p-3)} \sum_{m=0}^{p-1} \omega_m \rho^{2m} \zeta^{2p-2m-1}, \tag{28}$$

where

$$\omega_m = \frac{(2m-1)!!}{(2m)!!} \left[b \binom{2p-1}{2m} + \binom{2p-3}{2m} \right] - \frac{(2m-3)!!}{(2m-2)!!} \binom{2p-3}{2m-2} \tag{29}$$

and

$$b = \frac{16p^3 + 4(8\nu - 7)p^2 + 8(1-2\nu)p - 6\nu + 4}{(2p-1)(2p-2)(4p+1)}. \tag{30}$$

The terms $\binom{2p-3}{2m-2}$ in ω_0 and $\binom{2p-3}{2m}$ in ω_{p-1} are zeros due to the above mentioned convention for the binomial coefficients and the convention $\binom{a}{b} = 0$ for $b < a$, respectively.

8. CONCLUDING REMARKS

The σ_z; τ_{rz}, and σ_r representations, derived above, when introduced into the equilibrium equation

$$\frac{\partial \sigma_r}{\partial r} + \frac{\partial \tau_{rz}}{\partial z} + \frac{\sigma_r - \sigma_\theta}{r} = 0 \tag{31}$$

and the Hooke's law relation

$$u = \frac{r}{2(1+\nu)G} \left[\sigma_\theta - \nu(\sigma_r + \sigma_z) \right] \tag{32}$$

imply almost directly similar polynomial representations for the remaining basic quantities of the considered problem, namely the circumferential stress $\sigma_\theta(\rho, \zeta; p)$ and the radial displacement $u(\rho, \zeta; p)$. The coefficients in these representations are linear combinations of the coefficients λ_m and μ_m.

The following remark is due with respect to the forms of the coefficients λ_m, μ_m and ω_m. Obviously, Eqs. (21) (or (22)), (27) and (29) are only particular and certainly not the optimum forms of the otherwise large varieties of equivalent and maybe even simpler and more compact forms in which these coefficients are representable. Each of the particular forms derived above should be actually viewed as a basis for deriving other, eventually more convenient in one or another sense, equivalent forms of the same coefficients.

It should be probably mentioned in addition that the stresses and displacements in the fibre are not, as the first impression might be, independent of the mechanical properties of the matrix, the geometry of the unit composite cell (i.e. of the thickness of the matrix coating, or, which is the same, of the fibre volume fraction), and of the current loading parameter. In fact, these parameters enter the above derived expressions for the fibre stresses and displacements through the multiplier σ_0. The latter specifies the boundary conditions for the representative problem (cf. Eq. (2)) and presents itself the axial fibre stress in the trivial case of uniformly stressed fibre, i.e. the case which corresponds, formally, to the solution of the representative problem with $p = 1$. The coupling of this trivial fibre state with that of the surrounding matrix implies the so-called *plane cross sections-type* problems for the entire unit composite cell. The determination of the σ_0-stress is a basic element of the solution of these problems. References [5, 6] provide the σ_0-values for two particular but typical problems of thermal and mechanical loading of a unit composite cell, namely the problems of uniform cooling (heating) of the matrix phase and of longitudinal extension (compression) of a unit composite cell.

Acknowledgements. The support of the Bulgarian Ministry of Education and Science under Grant No MM 805-98 is gratefully acknowledged.

REFERENCES

1. Mihovsky, I. M. An end effect model for fibre-reinforced composites. In: *Continuum Models and Discrete Systems*, E. İnan and K. Markov, eds., World Scientific, 1998, 483–490.
2. Timoshenko, S. P., J. N. Goodier. Theory of elasticity. McGraw-Hill Kogakusha Ltd, Tokyo, 1970.
3. Gradshtein, I. S., E. M. Ryzhik. Tables of integrals, sums, series and products. Gos. Izd. Fiz.-mat. Lit., 4th ed., 1963.
4. Bateman, H., A. Erdely. Higher transcendental functions. Vol. 1, McGraw-Hill, 1953.
5. Herrmann, K., I. M. Mihovsky. Approximate analytical investigation of the elastic–plastic behaviour of fibrous composites. I. Thermal loading. *Annuaire de l'Univ. de Sofia, Fac. de Math. et Inf.*, Livre 2 — Mécanique, **83**, 1989, 5–20.
6. Herrmann, K., I. M. Mihovsky. Approximate analytical investigation of the elastic–plastic behaviour of fibrous composites. II. External loading. *Ibid.*, 21–35.

Received February 25, 2000

Faculty of Mathematics and Informatics
"St. Kliment Ohridski" University of Sofia
BG-1164 Sofia, Bulgaria
E-mail: mihovsky@fmi.uni-sofia.bg

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

SCATTERING OF ACOUSTOELECTRIC WAVES ON A CYLINDRICAL INHOMOGENEITY IN THE TRANSVERSELY ISOTROPIC PIEZOELECTRIC MEDIUM

VALERY M. LEVIN and THOMAS MICHELITSCH

Scattering of acoustoelectric waves on an inhomogeneity is studied. The scatterer is a circular continuous piezoelectric cylinder (fiber) embedded in the piezoelectric transversely isotropic medium. Expressions are found for the scattering amplitudes and total cross-sections of the three acoustoelectric waves propagating in the direction normal to the fiber axis. In the long-wave limit these expressions are obtained explicitly.

Keywords: piezoelectric medium, acoustoelectric waves, scattering, cylindrical inhomogeneity

MSC 2000: 74F15, 74J20

1. INTRODUCTION

The problem of the scattering of elastic waves on a single inhomogeneity in an elastic medium is of importance for several applications. First, such studies provide an information about the scatterer and therefore are relevant for the nondestructive evaluation of structural members. Another application is the investigation of the attenuation and velocity of elastic waves propagating through a medium, consisting of a set of noninteracting inhomogeneities. In recent years, significant progress has been achieved in solving this problem for ideally elastic materials [1–6].

In the present paper, we consider the scattering of acoustoelectric waves on a continuous cylindrical fiber embedded in a piezoelectric medium of hexagonal (transversely isotropic) symmetry. The expressions for scattering amplitudes of the acoustoelectric waves follow from the system of the integral equations for the electroelastic fields in the medium with inhomogeneity. This system is obtained in terms of Green's function of the coupled dynamic electroelastic problem (see Section 2). In Section 3, explicit expressions are obtained for the components of Green's function and scattering amplitudes for the quasiplane dynamic problem for the transversely isotropic piezoelectric medium. In Section 4, general formulae are derived for the total cross-section of acoustoelectric waves propagating in the direction normal to the fiber axis. Finally, explicit expressions are obtained for scattering amplitudes and total cross sections of three acoustoelectric waves in the long wave-length limit.

2. THE INTEGRAL EQUATIONS OF THE SCATTERING PROBLEM

We consider the piezoelectric medium obeying the following linear constitutive equations:

$$\begin{aligned} \sigma_{ij} &= C_{ijkl}\varepsilon_{kl} - e_{kij}E_k, \\ D_i &= e_{ikl}\varepsilon_{kl} + \eta_{ik}E_k, \end{aligned} \tag{2.1}$$

where σ and ε are the stress and strain tensors, E and D are the electric field intensity and electric displacement, respectively, $C = C^E$ is the tensor of elastic moduli at fixed E, $\eta = \eta^\varepsilon$ is the permittivity tensor at fixed strain ε, e is the piezoelectric constants tensor, and the superscript 'T' denotes the transposed tensor.

The substitution of Eqs. (2.1) into the equations of elastodynamics and Maxwell's equations leads to a coupled system of equations of linear electroelasticity. As usual, we disregard body sources of electrical nature. Hence, the equations of motion have the same form as in the theory of uncoupled elasticity

$$\partial_j\sigma_{ij} - \rho\ddot{u}_i = Q_i, \quad \partial_j = \partial/\partial x_j, \tag{2.2}$$

where u_i is the vector of elastic displacement, ρ is the material density, Q_i is the body force vector.

The solution of equation (2.2) together with Maxwell's equations describes the elastic-electromagnetic waves, i.e. elastic waves interacting with the electric field and the electromagnetic waves accompanying the deformation. If the characteristic velocity of the elastic waves is v, then the corresponding velocity of the electromagnetic waves has the order of $10^5\,v$. Therefore, we neglect the magnetic field generated by the elastic field propagating with the velocity v. It follows, then, that the magnetic effects can be neglected and the quasistatic approximation for the electric field can be used.

An additional field equation is the conservation of free electric charges:

$$\partial_i D_i = -q, \tag{2.3}$$

where q is the density of free electric charges and D_i is the dielectric displacement.

Since

$$E_i = -\partial_i\varphi, \quad \varepsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i), \tag{2.4}$$

where φ is the electric potential, the constitutive equations can be rewritten in the form

$$\sigma_{ij} = C_{ijkl}\partial_l u_k + e_{ijk}\partial_k\varphi,$$
$$D_i = e^T_{ikl}\partial_l u_k - \eta_{ik}\partial_k\varphi. \tag{2.5}$$

Substituting them into (2.2) and (2.3) yields a coupled system of linear differential equations of electroelasticity for the piezoelectric medium:

$$\partial_j C_{ijkl}\partial_l u_k + \partial_j e_{ijk}\partial_k\varphi - \rho\ddot{u}_i = -Q_i,$$
$$\partial_i e_{ikl}\partial_l u_k - \partial_i\eta_{ik}\partial_k\varphi = -q. \tag{2.6}$$

We consider now the harmonic oscillation of the medium with frequency ω. Since the dependence of quantities entering (2.6) on time is given by the multiplier $\exp(-i\omega t)$, the system (2.6) takes the form

$$\partial_j C_{ijkl}\partial_k u_l + \rho\omega^2 u_i + \partial_j e_{ijk}\partial_k\varphi = -Q_i,$$
$$\partial_i e^T_{ikl}\partial_k u_l - \partial_i\eta_{ik}\partial_k\varphi = -q. \tag{2.7}$$

Let the body forces Q_i and electric charges q be distributed in some domain V. The solution of the system (2.7) that vanishes at infinity can be represented as

$$u_i(x) = \int\limits_V G_{ik}(x-x')Q_k(x')\,\mathrm{d}x' + \int\limits_V \Gamma_i(x-x')q(x')\,\mathrm{d}x',$$
$$\varphi(x) = \int\limits_V \gamma_k(x-x')Q_k(x')\,\mathrm{d}x' + \int\limits_V g(x-x')q(x')\,\mathrm{d}x' \tag{2.8}$$

(the dependencies on frequency ω are omitted). The substitution of these expressions into the left-hand sides of (2.7) leads to a system of differential equations for the kernels $G_{ik}(x)$, $\Gamma_i(x)$, $\gamma_k(x)$ and $g(x)$ — the components of the electroelastic Green's function:

$$(\partial_j C_{ijkl}\partial_k + \rho\omega^2\delta_{il})G_{lm}(x) + \partial_j e_{ijk}\partial_k\gamma_m(x) = -\delta_{im}\delta(x),$$
$$(\partial_j C_{ijkl}\partial_k + \rho\omega^2\delta_{il})\Gamma_l(x) + \partial_j e_{ijk}\partial_k g(x) = 0,$$
$$\partial_i e^T_{ikl}\partial_k G_{lm}(x) - \partial_i\eta_{ik}\partial_k\gamma_m(x) = 0,$$
$$\partial_i e^T_{ikl}\partial_k\Gamma_l(x) - \partial_i\eta_{ik}\partial_k g(x) = -\delta(x), \tag{2.9}$$

where $\delta(x)$ is the Dirac function. Fourier transformation of these equations yields

$$\Lambda_{il}(k)G_{lj}(k) + h_i(k)\gamma_j(k) = \delta_{ij},$$
$$h^T_l(k)G_{lj}(k) - \lambda(k)\gamma_j(k) = 0,$$
$$\Lambda_{il}(k)\Gamma_l(x) + h_i(k)g(k) = 0,$$
$$h^T_l(k)\Gamma_l(k) - \lambda(k)g(k) = 1, \tag{2.10}$$

where

$$\Lambda_{il} = k_j C_{ijkl} k_k - \rho\omega^2 \delta_{il}, \quad h_i(k) = e_{ikl} k_k k_l,$$
$$h_l^T = e_{ikl}^T k_i k_k, \quad \lambda(k) = \eta_{ik} k_i k_k. \tag{2.11}$$

The solution of the system (2.10) can be written in the form

$$G_{ik} = \left(\Lambda_{ik} + \frac{1}{\lambda} h_i h_k^T\right)^{-1}, \quad g = -(\lambda + h_i^T \Lambda_{ij}^{-1} h_j)^{-1},$$
$$\gamma_i = \frac{1}{\lambda} h_k^T G_{ki}, \quad \Gamma_i = -\Lambda_{ik}^{-1} h_k g. \tag{2.12}$$

One can show that $\gamma_i = \Gamma_i$. Introducing the notation

$$\mathcal{G}(k,\omega) = \left\| \begin{array}{cc} G_{ik}(k,\omega) & \gamma_i(k,\omega) \\ \gamma_k^T(k,\omega) & g(k,\omega) \end{array} \right\|, \tag{2.13}$$

the x-presentation of Green's function can be obtained via the inverse Fourier transformation:

$$\mathcal{G}(x,\omega) = \frac{1}{(2\pi)^3} \int \mathcal{G}(k,\omega) \mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}}\, \mathrm{d}\mathbf{k}. \tag{2.14}$$

The equations of motion (2.7) can be written in the following symbolic form:

$$\mathcal{L}(\nabla) f(x) = 0, \quad \mathcal{L}(\nabla,\omega) = \mathcal{T}(\nabla) + \omega^2 \rho \mathcal{J}, \tag{2.15}$$

where

$$\mathcal{T}(\nabla) = \left\| \begin{array}{cc} T_{ik}(\nabla) & t_i(\nabla) \\ t_k^T(\nabla) & -\tau(\nabla) \end{array} \right\|, \quad \mathcal{J} = \left\| \begin{array}{cc} \delta_{ik} & 0 \\ 0 & 0 \end{array} \right\|, \quad f(x) = \left\| \begin{array}{c} u_k(x) \\ \varphi(x) \end{array} \right\|, \tag{2.16}$$

$$T_{ik}(\nabla) = \partial_j C_{ijkl} \partial_l, \quad t_i(\nabla) = \partial_j e_{ijk} \partial_k, \quad \tau(\nabla) = \partial_i \eta_{ik} \partial_k.$$

Consider now an unbounded medium with the electroelastic characteristics $\mathcal{L}^0$ and density ρ_0, where

$$\mathcal{L}^0 = \left\| \begin{array}{cc} C_{ijkl}^0 & e_{ijk}^0 \\ e_{ikl}^{0T} & -\eta_{ik}^0 \end{array} \right\|, \tag{2.17}$$

containing a region V with different electroelastic characteristics $\mathcal{L}$ and density ρ.

Let the harmonic vibrations of frequency ω propagate in the medium with the inhomogeneity. The electroelastic fields in such a medium satisfy equations (2.15) in which C, e, η and ρ are functions of coordinates. We represent these functions in the form

$$C(x) = C^0 + C^1 V(x), \quad e(x) = e^0 + e^1 V(x),$$
$$\eta(x) = \eta^0 + \eta^1 V(x), \quad \rho(x) = \rho_0 + \rho_1 V(x), \tag{2.18}$$

where $V(x)$ is the characteristic function of the region V, and the quantities with the superscript '1' denote the differences

$$C^1 = C - C^0, \quad e^1 = e - e^0, \quad \eta^1 = \eta - \eta^0, \quad \rho_1 = \rho - \rho_0. \tag{2.19}$$

The problem of electroelastic fields determination in the medium with an inclusion can be then reduced to the following system of integral equations:

$$f(x) = f^0(x) + \int_V S(x - x')\mathcal{L}^1 F(x')\,\mathrm{d}x' + \omega^2\rho_1 \int_V \mathcal{G}(x - x')\mathcal{J} f(x')\,\mathrm{d}x', \tag{2.20}$$

with $f^0(x)$ denoting the "incident" field. The latter satisfies the equation

$$\left[\mathcal{T}^0(\nabla) + \omega^2\rho_0\mathcal{J}\right] f^0(x) = 0 \tag{2.21}$$

with the notations

$$\mathcal{L}^1 = \left\| \begin{matrix} C^1 & e^1 \\ e^{1T} & -\eta^1 \end{matrix} \right\|, \quad S(x) = \left\| \begin{matrix} G_{ik,l}(x) & \gamma_{i,k}(x) \\ \gamma^T_{k,l}(x) & -g_{,k}(x) \end{matrix} \right\|,$$
$$F(x) = \left\| \begin{matrix} \varepsilon(x) \\ -E(x) \end{matrix} \right\|. \tag{2.22}$$

Here $G(x)$, $\gamma(x)$ and $g(x)$ are the respective x-representations of the functions entering (2.13).

When $x \in V$, Eq. (2.20) describes the electroelastic fields inside the inhomogeneity on which the fields outside of it can be constructed uniquely.

3. ELECTROELASTIC FIELDS IN THE TRANSVERSELY ISOTROPIC PIEZOELECTRIC MEDIUM CONTAINING A CONTINUOUS CYLINDRICAL FIBER

We consider an inhomogeneity having the shape of an infinite circular cylinder (continuous fiber) with the axis parallel to x_3-axis of the Cartesian coordinate system. Let the plane wave propagate in the direction normal to x_3-axis. Since $\mathcal{L}(x)$ and $\rho(x)$ are functions of x_1, x_2 only, the functions $f^0(x)$, $f(x)$, $F(x)$ are independent of x_3. Taking into account the relation

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-k_3 x'_3}\,\mathrm{d}x'_3 = \delta(k_3),$$

Eq. (2.20) transforms into the following one:

$$f(y) = f^0(y) + \int_S S(y - y')\mathcal{L}^1 F(y')\,\mathrm{d}y' + \omega^2\rho_1 \int_S \mathcal{G}(y - y')\mathcal{J} f(y')\,\mathrm{d}y', \tag{3.1}$$

where S is the cylindrical cross-section, $y = (x_1, x_2)$ and

$$\mathcal{G}(y - y') = \frac{1}{(2\pi)^2} \int_0^\infty \overline{k}\, d\overline{k} \int_0^{2\pi} \mathcal{G}(\overline{k}) \exp(-i\overline{\mathbf{k}} \cdot (\mathbf{y} - \mathbf{y}'))\, d\phi, \quad \overline{k} = (k_1, k_2). \tag{3.2}$$

The expression for $\mathcal{G}(\overline{k})$ has to be obtained from $\mathcal{G}(k, \omega)$, given by (2.13) by setting $k_3 = 0$.

Let x_3 be the axis of transverse isotropy. The material is characterized by five independent elastic moduli $\mathbf{C}^0 = \left\{C_{11}^0, C_{12}^0, C_{13}^0, C_{33}^0, C_{44}^0, C_{66}^0 = \frac{1}{2}(C_{11}^0 - C_{12}^0)\right\}$, three piezoelectric constants $\mathbf{e}^0 = \left\{e_{31}^0, e_{15}^0, e_{33}^0\right\}$ and two permeability coefficients $\eta^0 = \left\{\eta_{11}^0, \eta_{33}^0\right\}$. To simplify the needed in the sequel tensorial operations (inversion, contractions, etc), the tensors $\mathbf{C}^0, \mathbf{e}^0$ and η^0 are expressed in the form

$$\begin{aligned} \mathbf{C}^0 &= \frac{1}{2}(C_{11}^0 + C_{12}^0)\mathbf{T}^2 + 2C_{66}^0 \left(\mathbf{T}^1 - \frac{1}{2}\mathbf{T}^2\right) \\ &\quad + C_{13}^0(\mathbf{T}^3 + \mathbf{T}^4) + 4C_{44}^0 \mathbf{T}^5 + C_{33}^0 \mathbf{T}^6, \\ \mathbf{e}^0 &= e_{31}^0 \mathbf{U}^1 + e_{15}^0 \mathbf{U}^2 + e_{33}^0 \mathbf{U}^3, \quad \eta^0 = \eta_{11}^0 \mathbf{t}^1 + \eta_{33}^0 \mathbf{t}^2. \end{aligned} \tag{3.3}$$

The basic tensors $\mathbf{T}^1, \ldots, \mathbf{T}^6$, $\mathbf{U}^1, \mathbf{U}^2, \mathbf{U}^3$, $\mathbf{t}^1, \mathbf{t}^2$ are defined here by means of their components as follows:

$$\begin{aligned} &T_{ijkl}^1 = \theta_{i(k}\theta_{l)k}, \quad T_{ijkl}^2 = \theta_{ij}\theta_{kl}, \quad T_{ijkl}^3 = \theta_{ij} m_k m_l, \\ &T_{ijkl}^4 = m_i m_j \theta_{kl}, \quad T_{ijkl}^5 = \theta_{i)(k} m_{l)} m_{(j}, \quad T_{ijkl}^6 = m_i m_j m_k m_l, \\ &U_{ijk}^1 = \theta_{ij} m_k, U_{ijk}^2 = 2m_{(i}\theta_{j)k}, \quad U_{ijk}^3 = m_i m_j m_k, \\ &\qquad t_{ij}^1 = \theta_{ij}, \quad t_{ij}^2 = m_i m_j, \end{aligned} \tag{3.4}$$

in the Cartesian system whose x_3-axis is along the unit vector $\mathbf{m}$; the components of the tensor θ_{ij} are $\theta_{ij} = \delta_{ij} - m_i m_j$.

The appropriate formulae for the operations on these tensors are given in the Appendix.

The fiber material possesses the transverse isotropy aligned with the one of the matrix. The tensors of the elastic moduli, of the piezoelectric constants and of the dielectric coefficients of the fibers can be expressed in the same tensorial basis, similarly to (3.3) (without the superscript '0').

Using (3.3), one obtains

$$\begin{aligned} \Lambda_{ik}(\overline{k}) &= \Lambda_1 \overline{n}_i \overline{n}_k + \Lambda_2(\theta_{ik} - \overline{n}_i \overline{n}_k) + \Lambda_3 m_i m_k, \\ h_i(\overline{k}) &= h_i^T(\overline{k}) = \overline{k}^2 e_{15}^0 m_i, \quad \lambda(\overline{k}) = \overline{k}^2 \eta_{11}^0, \end{aligned} \tag{3.5}$$

where

$$\Lambda_1 = \overline{k}^2 C_{11}^0 - \rho_0\omega^2, \quad \Lambda_2 = \overline{k}^2 C_{66}^0 - \rho_0\omega^2, \quad \Lambda_3 = \overline{k}^2 C_{44}^0 - \rho_0\omega^2. \tag{3.6}$$

These expressions and (2.12) imply that

$$G_{ik}(\overline{k},\omega)=\frac{1}{\Lambda_1}\overline{n}_i\overline{n}_k+\frac{1}{\Lambda_2}(\theta_{ik}-\overline{n}_i\overline{n}_k)+\frac{1}{\Lambda_3'}m_im_k,$$

$$g(\overline{k},\omega)=-\frac{1}{\overline{k}^2\eta_{11}^0}\left[1-\frac{\overline{k}^2(e_{15}^0)^2}{\eta_{11}^0\Lambda_3'},\right],\quad \gamma_i=\frac{e_{15}^0}{\eta_{11}^0\Lambda_3'}m_i, \tag{3.7}$$

$$\Lambda_3'=\overline{k}^2C_{44}'-\rho_0\omega^2,\quad C_{44}'=C_{44}^0+\frac{(e_{15}^0)^2}{\eta_{11}^0}.$$

Introducing the quantities

$$\alpha^2=\frac{\rho_0\omega^2}{C_{11}^0},\quad \beta^2=\frac{\rho_0\omega^2}{C_{66}^0},\quad \beta_\perp^2=\frac{\rho_0\omega^2}{C_{44}'}, \tag{3.8}$$

the expressions (3.7) are recast as

$$\begin{aligned}G_{ik}(\overline{k},\omega)&=\frac{1}{\rho_0\omega^2}\left[\frac{\beta^2}{\overline{k}^2-\beta^2}\theta_{ik}\right.\\&\left.+\frac{\overline{k}_i\overline{k}_k}{\overline{k}^2}\left(\frac{\alpha^2}{\overline{k}^2-\alpha^2}-\frac{\beta^2}{\overline{k}^2-\beta^2}\right)+m_im_k\frac{\beta_\perp^2}{\overline{k}^2-\beta_\perp^2}\right],\\ g(\overline{k},\omega)&=\frac{1}{\eta_{11}^0}\frac{1}{\overline{k}^2}-\frac{1}{\rho_0\omega^2}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)^2\frac{\beta_\perp^2}{\overline{k}^2-\beta_\perp^2},\\ \gamma_i(\overline{k},\omega)&=\frac{1}{\rho_0\omega^2}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)\frac{\beta_\perp^2}{\overline{k}^2-\beta_\perp^2}m_i.\end{aligned} \tag{3.9}$$

To determine the x-representation of functions $G_{ik}(\overline{k},\omega)$, $\gamma_i(\overline{k},\omega)$ and $g(\overline{k},\omega)$, according to (2.14), we have to calculate the integral

$$I=\frac{1}{(2\pi)^2}\int_0^\infty\frac{\overline{k}\,\mathrm{d}\overline{k}}{\overline{k}^2-\beta^2}\int_0^{2\pi}\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{y}}\,\mathrm{d}\phi.$$

Since

$$\int_0^{2\pi}\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{y}}\,\mathrm{d}\phi=\int_0^{2\pi}\mathrm{e}^{-i\overline{k}y\cos\phi}\,\mathrm{d}\phi=2\int_0^{\pi}\cos(\overline{k}y\cos\phi)\,\mathrm{d}\phi=2\pi J_0(\overline{k}y),$$

where $J_0(z)$ is the Bessel function, we have

$$I=\frac{1}{2\pi}\int_0^\infty\frac{J_0(\overline{k}y)\overline{k}\,\mathrm{d}\overline{k}}{\overline{k}^2-\beta^2}=\frac{1}{2\pi}\lim_{\varepsilon\to0}\int_0^\infty\frac{J_0(\overline{k}y)\overline{k}\,\mathrm{d}\overline{k}}{\overline{k}^2+(\varepsilon+i\beta)^2}=\frac{i}{4}H_0^{(1)}(\beta y).$$

Here, $H_0^{(1)}(z)$ is the Hankel function of the first kind. Hence, the x-representation of these functions has the form

$$G_{ik}(r,\omega) = \frac{i}{4\rho_0\omega^2}\left\{\theta_{ik}\beta^2 H_0^{(1)}(\beta r) - \frac{\partial^2}{\partial y_i \partial y_k}\left[H_0^{(1)}(qr)\right]_\beta^\alpha + m_i m_k \beta_\perp^2 H_0^{(1)}(\beta_\perp r)\right\},$$

$$\gamma_i(r,\omega) = \frac{i}{4\rho_0\omega^2}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)\beta_\perp^2 H_0^{(1)}(\beta_\perp r) m_i, \tag{3.10}$$

$$g(r,\omega) = \frac{1}{2\pi\eta_{11}^0}\ln r - \frac{i}{4\rho_0\omega^2}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)^2 \beta_\perp^2 H_0^{(1)}(\beta_\perp r),$$

where

$$[f(qr)]_\beta^\alpha \equiv f(\alpha r) - f(\beta r), \quad r = |y|. \tag{3.11}$$

Eqs. (3.1), when written in detail, have the form

$$u_i(y) = u_i^0(y) + \int_S \Big[\Psi_{imn}(R)\varepsilon_{mn}(y') - \psi_{im}(R)E_m(y') + \rho_1\omega^2 G_{ik}(R)u_k(y')\Big]\,\mathrm{d}y', \tag{3.12}$$

$$\varphi(y) = \varphi^0(y) + \int_S \Big[\Phi_{mn}(R)\varepsilon_{mn}(y') - \phi_m(R)E_m(y') + \rho_1\omega^2 \gamma_k(R)u_k(y')\Big]\,\mathrm{d}y', \tag{3.13}$$

where

$$\begin{aligned}
\Psi_{imn}(R) &= G_{ik,l}(R)C_{klmn}^1 + \gamma_{i,k}(R)e_{kmn}^{1T},\\
\psi_{im}(R) &= G_{ik,l}(R)e_{klm}^1 - \gamma_{i,k}(R)\eta_{km}^1,\\
\Phi_{mn}(R) &= \gamma_{k,l}(R)C_{klmn}^1 - g_{,k}(R)e_{kmn}^{1T},\\
\phi_m(R) &= \gamma_{k,l}(R)e_{klm}^1 + g_{,k}(R)\eta_{km}^1, \quad R = |y - y'|.
\end{aligned} \tag{3.14}$$

Eqs. (3.12) and (3.13) allow one to find the far-field asymptotics of the electroelastic fields. Taking into account the asymptotic formulas at $r \to \infty$

$$R^{-1} \sim r^{-1}, \quad R \sim r - (\mathbf{n}\cdot\mathbf{y}'), \quad n_i = \frac{y_i}{y}, \quad y = |\mathbf{y}|,$$

$$\frac{\partial}{\partial y_{k_1}}\cdots\frac{\partial}{\partial y_{k_m}} H_0^{(1)}(qR) \sim (iq)^m n_{k_1}\dots n_{k_m}\sqrt{\frac{2}{\pi q y}}\,\mathrm{e}^{i(qy-\pi/4)}\mathrm{e}^{-iq(\mathbf{n}\cdot\mathbf{y}')},$$

one has

$$u_i(y) = u_i^0(y) + u_i^s(y), \quad \varphi(y) = \varphi^0(y) + \varphi^s(y), \tag{3.15}$$

and the "scattered" fields $u_i^s(y)$ and $\varphi^s(y)$ are determined by the expressions

$$u_i^s(y) = A_i(n)\frac{e^{i\alpha y}}{\sqrt{y}} + B_i(n)\frac{e^{i\beta y}}{\sqrt{y}} + C_i(n)\frac{e^{i\beta_\perp y}}{\sqrt{y}}\,,$$
$$\varphi^s(y) = c(n)\frac{e^{i\beta_\perp y}}{\sqrt{y}}\,. \tag{3.16}$$

Here $A_i(n)$, $B_i(n)$, $C_i(n)$ and $c(n)$ are the amplitudes of the three cylindrical waves that can be represented in the form

$$A_i(n) = n_i n_k f_k(\alpha n), \quad B_i(n) = (\theta_{ik} - n_i n_k) f_k(\beta n),$$
$$C_i(n) = m_i m_k f_k(\beta_\perp n) + m_i f(\beta_\perp n), \tag{3.17}$$
$$c(n) = \frac{e_{15}^0}{\eta_{11}^0}\Big[m_k f_k(\beta_\perp n) + f(\beta_\perp n)\Big],$$

with the notations

$$f_k(qn) = \frac{i}{2\rho_0\omega^2}\sqrt{\frac{q^3}{2\pi}}\,e^{-i\pi/4}\Bigg\{ iqn_l \int_S \Big[C_{klmn}^1 \varepsilon_{mn}(y')$$
$$-e_{klm}^1 E_m(y')\Big] e^{-iq(\mathbf{n}\cdot\mathbf{y}')}\,dy' + \rho_1\omega^2 \int_S u_k(R) e^{-iq(\mathbf{n}\cdot\mathbf{y}')}\,dy' \Bigg\}, \tag{3.18}$$
$$f(\beta_\perp n) = \frac{i}{2\rho_0\omega^2}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)\sqrt{\frac{\beta_\perp^3}{2\pi}}\,e^{-\frac{i\pi}{4}} i\beta_\perp n_k \int_S \Big[e_{kmn}^{1T}\varepsilon_{mn}(y')$$
$$+ \eta_{km}^1 E_m(y')\Big] e^{-i\beta_\perp(\mathbf{n}\cdot\mathbf{y}')}\,dy', \quad q = \alpha, \beta, \beta_\perp.$$

4. SCATTERING CROSS-SECTION IN A PIEZOELECTRIC MEDIUM

We define the intensity vector I_i, associated with a stress field σ_{ij}, the electric potential φ and the velocities $\dot{u}_i$ and $\dot{D}_i$ by the relation

$$I_i = \sigma_{ij}\dot{u}_j + \varphi\dot{D}_i. \tag{4.1}$$

Similarly, we denote by I_i^s the intensity vector associated with the scattered fields, and by I_i^0 the intensity vector associated with the incident fields. The term "intensity" refers to the rate of energy transfer per unit area in the direction normal to the one of propagation, that is

$$I = I_i n_i, \tag{4.2}$$

where n_i is the unit vector in the direction of propagation. The power flux (the rate of energy transfer across the surface S with unit normal n_i) is

$$Q = \int_S I_i n_i \, \mathrm{d}S = \int_S (\sigma_{ij} \dot{u}_j + \varphi \dot{D}_i) n_i \, \mathrm{d}S. \tag{4.3}$$

For a given angular frequency corresponding to period T, the total cross section $Q(\omega)$ is the ratio of the average power flux over all directions to the average intensity of the incident fields:

$$Q(\omega) = \frac{\langle Q^s \rangle_t}{\langle I^0 \rangle_t}, \tag{4.4}$$

where $\langle \cdot \rangle_t$ denotes the time averaging over the period T.

Having found the far-field asymptotics of the scattered electroelastic fields we can now compute the total cross-section according to relation (4.4). Since the power flux is a real number,

$$\langle Q \rangle_t = \frac{1}{4} \int_S \Big\langle (\sigma_{ij} + \sigma^*_{ij})(\dot{u}_j + \dot{u}^*_j) + (\varphi + \varphi^*)(\dot{D}_i + \dot{D}^*_i) \Big\rangle_t n_i \, \mathrm{d}S, \tag{4.5}$$

where '$*$' denotes the complex conjugate. Since we assume the vibrations to be harmonic,

$$\begin{aligned} \langle Q \rangle_t = \frac{i\omega}{4} \int_S \big\langle & -\sigma_{ij} u_i \mathrm{e}^{-2i\omega t} + \sigma^*_{ij} u^*_i \mathrm{e}^{2i\omega t} - \sigma^*_{ij} u_i + \sigma_{ij} u^*_i \\ & -\varphi D_j \mathrm{e}^{-2i\omega t} + \varphi^* D^*_j \mathrm{e}^{2i\omega t} - \varphi^* D_j + \varphi D^*_j \big\rangle_t n_j \, \mathrm{d}S. \end{aligned} \tag{4.6}$$

Computing the time average yields

$$\langle Q \rangle_t = -\frac{1}{2} \omega \, \mathrm{Im} \int_S (\sigma_{ij} u^*_i - D_j \varphi^*) n_j \, \mathrm{d}S. \tag{4.7}$$

Hence, the expression for the total cross-section takes the form

$$\begin{aligned} Q(\omega) &= -\frac{\omega}{2 \langle I^0 \rangle_t} \, \mathrm{Im} \int_S (\sigma^s_{ij} u^{*s}_i - D^s_j \varphi^{*s}) n_j \, \mathrm{d}S, \\ \langle I^0 \rangle_t &= -\frac{1}{2} \omega \, \mathrm{Im} \, (\sigma^0_{ij} u^{*0}_i - D^0_j \varphi^{*0}) n^0_j, \end{aligned} \tag{4.8}$$

where n^0_j is the normal to the front of the incident wave.

We apply now the general formula (4.8) to the scattering of the acoustoelectric waves on a continuous cylindrical surface of unit height and a radius r concentric with the fiber. Taking into account that the contribution to the energy flux through two cross-sections of this surface by the plane wave propagating normally to the fiber axis is zero, we have

$$Q(\omega) = -\frac{\omega}{2\langle I^0\rangle_t}\operatorname{Im}\int_0^{2\pi}(\sigma_{ij}^s u_i^{*s} - D_j^s\varphi^{*s})n_j r\,\mathrm{d}\phi\,, \tag{4.9}$$

where ϕ is the angle between the wave normal n^0 and an arbitrary normal to the fiber surface.

To compute $Q(\omega)$, we have to find (utilizing (3.16))

$$\sigma_{ij}^s = in_k\left[C_{ijkl}^0\left(A_l(n)\alpha\frac{\mathrm{e}^{i\alpha y}}{\sqrt{y}} + B_l(n)\beta\frac{\mathrm{e}^{i\beta y}}{\sqrt{y}}\right.\right.$$
$$\left.\left.+C_l(n)\beta_\perp\frac{\mathrm{e}^{i\beta_\perp y}}{\sqrt{y}}\right) + \mathrm{e}_{ijk}^0 c(n)\beta_\perp\frac{\mathrm{e}^{i\beta_\perp y}}{\sqrt{y}}\right].$$

Substituting the expressions for transversely isotropic tensors C_{ijkl}^0 and e_{ijk}^0, we obtain

$$\sigma_{ij}^s n_j = i\left\{\left[\frac{1}{2}(C_{11}^0 + C_{12}^0)n_i n_k A_k(n) + C_{66}^0 A_i(n)\right]\alpha\frac{\mathrm{e}^{i\alpha y}}{\sqrt{y}}\right.$$
$$\left.+C_{66}^0 B_i(n)\beta\frac{\mathrm{e}^{i\beta y}}{\sqrt{y}} + C_{44}' m_i m_k C_k(n)\beta_\perp\frac{\mathrm{e}^{i\beta_\perp y}}{\sqrt{y}}\right\}, \tag{4.10}$$

where the relation

$$C_{44}^0 m_k C_k(n) + e_{15}^0 c(n) = \left(C_{44}^0 + \frac{(e_{15}^0)^2}{\eta_{11}^0}\right)(m_k f_k(\beta_\perp n) + f(\beta_\perp n)) = C_{44}' m_k C_k(n)$$

is taken into account with C_{44}' determined by (3.7). Similarly,

$$D_i^s = i\left[e_{31}^0 m_i n_k A_k(n)\alpha\frac{\mathrm{e}^{i\alpha y}}{\sqrt{y}} + n_i(e_{15}^0 m_k C_k(n) - \eta_{11}^0 c(n))\beta_\perp\frac{\mathrm{e}^{i\beta_\perp y}}{\sqrt{y}}\right].$$

This implies that

$$D_i^s n_i = i(e_{15}^0 m_k C_k(n) - \eta_{11}^0 c(n))\beta_\perp\frac{\mathrm{e}^{i\beta_\perp y}}{\sqrt{y}}\,. \tag{4.11}$$

Since

$$e_{15}^0 m_k C_k(n) - \eta_{11}^0 c(n) = \Big(m_k f_k(\beta_\perp n) + f(\beta_\perp n)\Big)(e_{15}^0 - e_{15}^0) = 0,$$

the second term in (4.9) does not contribute to the total scattering cross-section.

The substitution of these expressions, alongside with the relation

$$u_i^{*s} = A_i^*(n)\frac{\mathrm{e}^{-i\alpha y}}{\sqrt{y}} + B_i^*(n)\frac{\mathrm{e}^{-i\beta y}}{\sqrt{y}} + C_i^*(n)\frac{\mathrm{e}^{-i\beta_\perp y}}{\sqrt{y}}\,,$$

into (4.9) yields eventually:

$$Q(\omega) = -\frac{\omega}{2\langle I^0\rangle}\int_0^{2\pi}\left[C_{11}^0\alpha|A_i|^2 + C_{66}^0\beta|B_i|^2 + C'_{44}\beta_\perp|C_i|^2\right]\,\mathrm{d}\phi, \tag{4.12}$$

where $|A_i|^2 = A_i A_i^*$.

We assume that the incident waves have the form

$$u_i^0(y,\omega) = U_i \mathrm{e}^{ik\mathbf{n}^0\cdot\mathbf{y}}, \quad \varphi^0(y,\omega) = \Phi\mathrm{e}^{ik\mathbf{n}^0\cdot\mathbf{y}}, \tag{4.13}$$

where k is the wave number, n_i^0 is the wave normal (perpendicular to the fiber axis), U_i is the polarization vector and Φ is the amplitude of the electric field. Since

$$\partial_j e_{jkl}^0 \partial_l u_k^0 - \partial_j \eta_{jk}^0 \partial_k \varphi^0 = 0,$$

it follows that

$$\Phi = \frac{e_{15}^0}{\eta_{11}^0} m_k U_k. \tag{4.14}$$

Hence, the expression for $\langle I^0\rangle_t$ can be represented in the form

$$\langle I^0\rangle_t = \frac{1}{2}\omega k\left[\frac{1}{2}(C_{11}^0 + C_{12}^0)(\mathbf{U}\cdot\mathbf{n}^0) + C_{66}^0(\mathbf{U}\cdot\mathbf{U}) + (C'_{44} - C_{66}^0)(\mathbf{U}\cdot\mathbf{m})\right]. \tag{4.15}$$

5. THE TOTAL SCATTERING CROSS-SECTION IN THE LONG-WAVE LIMIT

As it follows from expressions (3.17) for the amplitudes $A_i(n), B_i(n), C_i(n)$ and $c(n)$, the determination of vector $f_i(kn)$ and scalar $f(\beta_\perp n)$ plays a key role in the scattering problem. These quantities depend on the electroelastic fields u_i, φ (and the accompanying fields ε_{ij} and E_i), inside the region occupied by the scatterer. The mentioned fields have to be determined from the solution of the coupled electroelastic dynamic problem for the medium with the inhomogeneity. If these fields are found approximately, then the obtained formulae yield approximate expressions for the scattering cross-sections. Several approximations have been suggested (see the discussion of [4]), to mention only Born's approximation, quasistatic approximation and extended quasistatic approximation. We use the quasistatic (long-wave) approximation. The feature of this approximation is the replacement of the actual strain and electric fields inside of the inclusion by those of the static (infinite wavelength) problem. As it is well-known [7–11], if the external fields $F^0 = [\varepsilon^0, E^0]$ in the static limit ($\omega = 0$) are uniform in S, then the fields $F = [\varepsilon, E]$ inside this region are also uniform and have, after [11], the form

$$F = \mathcal{A}F^0, \quad \mathcal{A} = (\mathcal{I} + \mathcal{P}^0\mathcal{L}^1)^{-1}, \quad \mathcal{I} = \left\|\begin{array}{cc} I_{ijkl} & 0 \\ 0 & \delta_{ik}\end{array}\right\|, \quad I_{ijkl} = \delta_{i(k}\delta_{l)j}, \tag{5.1}$$

where $\mathcal{P}^0$ is an operator with constant components that can be represented in the tensor basis (3.4) as

$$\mathcal{P} = \left\| \begin{matrix} \mathbf{P}^0 & \mathbf{p}^0 \\ \mathbf{p}^0 & \pi^0 \end{matrix} \right\|,$$

$$\mathbf{P}^0 = -\left[\frac{1}{4C_{11}^0}\mathbf{T}^2 + \frac{1}{4}\left(\frac{1}{C_{11}^0} + \frac{1}{C_{66}^0}\right)\left(\mathbf{T}^1 - \frac{1}{2}\mathbf{T}^2\right) + \frac{\eta_{11}^0}{2\Delta_0}\mathbf{T}^5\right], \tag{5.2}$$

$$\mathbf{p}^0 = -\frac{e_{15}^0}{4\Delta_0}\mathbf{U}^2, \quad \pi^0 = \frac{C_{44}^0}{2\Delta_0}\mathbf{t}^1, \quad \Delta_0 = \eta_{11}^0 C_{44}^0 + (e_{15}^0)^2.$$

Note that only the product

$$\mathcal{L}^1 F = \mathcal{L}^1 \mathcal{A} F^0 = \mathcal{L}^{\mathcal{A}} F^0, \quad \mathcal{L}^{\mathcal{A}} = \mathcal{L}^1 \mathcal{A} \tag{5.3}$$

enters the right-hand sides of Eqs. (3.18). The components of the constant operator $\mathcal{L}^{\mathcal{A}}$ can be obtained by using tensorial operations in the basis (3.4), see the Appendix. Then

$$\mathcal{L}^{\mathcal{A}} = \left\| \begin{matrix} \mathbf{C}^{\mathcal{A}} & \mathbf{e}^{\mathcal{A}} \\ \mathbf{e}^{T\mathcal{A}} & -\eta^{\mathcal{A}} \end{matrix} \right\|,$$

$$\begin{aligned} \mathbf{C}^{\mathcal{A}} &= \frac{1}{2}(C_{11}^{\mathcal{A}} + C_{12}^{\mathcal{A}})\mathbf{T}^2 + 2C_{66}^{\mathcal{A}}\left(\mathbf{T}^1 - \frac{1}{2}\mathbf{T}^2\right) \\ &+ C_{13}^{\mathcal{A}}(\mathbf{T}^3 + \mathbf{T}^4) + 4C_{44}^{\mathcal{A}}\mathbf{T}^5 + C_{33}^{\mathcal{A}}\mathbf{T}^6, \end{aligned} \tag{5.4}$$

$$\mathbf{e}^{\mathcal{A}} = e_{31}^{\mathcal{A}}\mathbf{U}^1 + e_{15}^{\mathcal{A}}\mathbf{U}^2 + e_{33}^{\mathcal{A}}\mathbf{U}^3, \quad \eta^{\mathcal{A}} = \eta_{11}^{\mathcal{A}}\mathbf{t}^1 + \eta_{33}^{\mathcal{A}}\mathbf{t}^2,$$

with the notations

$$\frac{1}{2}\left(C_{11}^{\mathcal{A}} + C_{12}^{\mathcal{A}}\right) = \frac{1}{2}(C_{11}^1 + C_{12}^1)\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1},$$

$$C_{66}^{\mathcal{A}} = C_{66}^1\left[1 + \frac{C_{66}^1}{2}\left(\frac{1}{C_{11}^0} + \frac{1}{C_{66}^0}\right)\right]^{-1},$$

$$C_{13}^{\mathcal{A}} = C_{13}^1\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1}, \quad C_{44}^{\mathcal{A}} = \frac{1}{\Delta_f}\left[C_{44}^1 + \frac{C_{44}^0}{2\Delta_0}\left(C_{44}^1\eta_{11}^1 + (e_{15}^1)^2\right)\right],$$

$$C_{33}^{\mathcal{A}} = C_{33}^1 - \frac{(C_{13}^1)^2}{C_{11}^0}\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1}, \tag{5.5}$$

$$e_{31}^{\mathcal{A}} = e_{31}^1\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1}, \quad e_{15}^{\mathcal{A}} = \frac{1}{\Delta_f}\left[e_{15}^1 + \frac{e_{15}^0}{2\Delta_0}\left(C_{44}^1\eta_{11}^1 + (e_{15}^1)^2\right)\right],$$

$$e_{33}^{\mathcal{A}} = e_{33}^1 - \frac{C_{13}^1 e_{31}^1}{C_{11}^0}\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1},$$

$$\eta_{11}^{\mathcal{A}} = \frac{1}{\Delta_f}\left[\eta_{11}^1 + \frac{\eta_{11}^0}{2\Delta_0}\left(C_{44}^1\eta_{11}^1 + (e_{15}^1)^2\right)\right],$$

$$\eta_{33}^{\mathcal{A}} = \left[\eta_{33}^1 + \frac{(e_{31}^1)^2}{C_{11}^0}\left(1 + \frac{C_{11}^1 + C_{12}^1}{2C_{11}^0}\right)^{-1}\right],$$

$$\Delta_f = \left[1 + \frac{1}{2\Delta_0}\left(e_{15}^0 e_{15}^1 + C_{44}^0\eta_{11}^1\right)\right]\left[1 + \frac{1}{2\Delta_0}\left(e_{15}^0 e_{15}^1 + C_{44}^1\eta_{11}^0\right)\right]$$

$$-\frac{1}{4\Delta_0^2}\left(C_{44}^1 e_{15}^0 - C_{44}^0 e_{15}^1\right)\left(\eta_{11}^0 e_{15}^1 - \eta_{11}^1 e_{15}^0\right).$$

In the foregoing expressions the quantities with superscript '1' refer to the difference between the inclusion and the matrix electroelastic constants.

The determination of amplitudes $A_i(n)$, $B_i(n)$ and $C_i(n)$ utilizes the following relations that hold in the long-wave approximation:

$$\begin{aligned} &e^{-iq\mathbf{n}\cdot\mathbf{y}'} \simeq 1, \quad u_k^0 = U_k^0, \quad \varepsilon_{kl}^0 = ikU_{(k}^0 n_{l)}^0 \quad (k = \alpha, \beta, \beta_\perp), \\ &\varphi^0 = \frac{e_{15}^0}{\eta_{11}^0}(U_k^0 m_k), \quad E_k^0 = -i\beta_\perp n_k^0 \frac{e_{15}^0}{\eta_{11}^0}(U_k^0 m_k), \end{aligned} \tag{5.6}$$

where U_k^0 is the polarization vector, k is the wave number and n^0 is the normal to the wave front.

In accordance with (3.18), we have

$$\begin{aligned} f_k(qn) &= \frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{q^3}{2\pi}}e^{-i\pi/4}\Bigg[\rho_1\omega^2 U_k^0 \\ &\quad - qkC_{klpq}^{\mathcal{A}} n_l n_q^0 U_p^0 - q\beta_\perp \frac{e_{15}^0}{\eta_{11}^0}(U_k^0 m_k) e_{klp}^{\mathcal{A}} n_l n_p^0\Bigg], \\ f(\beta_\perp n) &= -\frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{q^3}{2\pi}}e^{-i\pi/4}\left(\frac{e_{15}^0}{\eta_{11}^0}\right)\beta_\perp \\ &\quad \times\left[ke_{kpq}^{\mathcal{A}T} n_k n_p^0 U_q^0 + \beta_\perp \frac{e_{15}^0}{\eta_{11}^0}(U_k^0 m_k)\eta_{kp}^{\mathcal{A}} n_k n_p^0\right], \end{aligned} \tag{5.7}$$

where $C^{\mathcal{A}}$, $e^{\mathcal{A}}$ and $\eta^{\mathcal{A}}$ are defined in (5.4) and (5.5), respectively, and a is the fiber radius.

We now consider several special cases.

5.1. LONGITUDINAL WAVES

In this case,

$$k = \alpha, \quad U_k^0 = n_k^0, \quad U_k^0 m_k = 0 \tag{5.8}$$

and the expressions for $f_k(\alpha n)$ and $f_k(\beta n)$ read

$$\begin{aligned} f_k(\alpha n) &= \frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{\alpha^3}{2\pi}}\mathrm{e}^{-i\pi/4}\Big\{\rho_1\omega^2 n_k^0 \\ &\quad - \alpha^2\left[\frac{1}{2}(C_{11}^A + C_{12}^A) + +C_{66}^A(2\cos\phi - 1)\right] n_k\Big\}, \\ f_k(\beta n) &= \frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{\beta^3}{2\pi}}\mathrm{e}^{-i\pi/4}\Big\{\rho_1\omega^2 n_k^0 \\ &\quad - \alpha\beta\left[\frac{1}{2}(C_{11}^A + C_{12}^A)n_k + C_{66}^A(2n_k^0\cos\phi - n_k)\right]\Big\}. \end{aligned} \tag{5.9}$$

According to (3.17), we find now

$$A_i(n) = \frac{i\pi a^2}{2}\sqrt{\frac{\alpha^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left\{\frac{\rho_1}{\rho_0}\cos\phi - \frac{1}{C_{11}^0}\left[\frac{1}{2}\left(C_{11}^A + C_{12}^A\right) + C_{66}^A\cos 2\phi\right]\right\} n_i,$$

$$B_i(\beta n) = \frac{i\pi a^2}{2}\sqrt{\frac{\beta^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left[\frac{\rho_1}{\rho_0} - 2\zeta\frac{C_{66}^A}{C_{66}^0}\cos\phi\right](n_i^0 - n_i\cos\phi), \tag{5.10}$$

where

$$\zeta = \alpha/\beta. \tag{5.11}$$

Obviously, $f(\beta_\perp n) = 0$ and the vector $f_k(\beta_\perp n)$ lies in the x_1x_2-plane. Therefore $f_k(\beta_\perp n)m_k = 0$ and $C_i(n) = c(n) = 0$.

Taking into account the relation

$$\langle I^0\rangle_t = -\frac{1}{2}\omega\alpha C_{11}^0 \tag{5.12}$$

and substituting (5.10) into the right-hand side of (4.12), we obtain, after integration with respect to φ,

$$\begin{aligned} Q_L(\omega) &= \frac{\pi^2}{8}a(\alpha a)^3\Big\{\frac{1}{(\rho_0 v_L^2)^2}\Big[\frac{1}{2}(C_{11}^A + C_{12}^A)^2 \\ &\quad + (C_{66}^A)^2\left(1 + \frac{1}{\zeta^4}\right)\Big] + \left(\frac{\rho_1}{\rho_0}\right)^2\left(1 + \frac{1}{\zeta^2}\right)\Big\}, \quad v_L^2 = C_{11}^0/\rho_0. \end{aligned} \tag{5.13}$$

5.2. SHEAR WAVES POLARIZED IN THE x_1x_2-PLANE

In this case,

$$\begin{aligned} U_k^0 &= e_k^0 \quad (e_k^0 n_k^0 = e_k^0 m_k^0 = 0), \quad k = \beta, \\ \langle I^0\rangle_t &= -\frac{1}{2}\omega\beta C_{66}^0. \end{aligned} \tag{5.14}$$

The scalar $f(\beta_\perp n)$ vanishes (as in the previous case) and

$$f_k(\alpha n) = \frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{\alpha^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left[\rho_1\omega^2 e_k^0 - \alpha\beta C_{66}^A(e_k^0\cos\phi + n_k^0\sin\phi)\right],$$
$$f_k(\beta n) = \frac{i\pi a^2}{2\rho_0\omega^2}\sqrt{\frac{\beta^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left[\rho_1\omega^2 e_k^0 - \beta^2 C_{66}^A(e_k^0\cos\phi + n_k^0\sin\phi)\right]. \tag{5.15}$$

The amplitudes $A_i(n)$ and $B_i(n)$ of the scattered waves take the form

$$A_i(n) = \frac{i\pi a^2}{2}\sqrt{\frac{\alpha^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left(\frac{\rho_1}{\rho_0}\sin\phi - \zeta\frac{C_{66}^A}{C_{66}^0}\sin 2\phi\right)n_i,$$
$$B_i(n) = \frac{i\pi a^2}{2}\sqrt{\frac{\beta^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left[\frac{\rho_1}{\rho_0}\left(e_i^0 - n_i\sin\phi\right)\right. \tag{5.16}$$
$$\left. -\beta^2\frac{C_{66}^A}{C_{66}^0}\left(e_i^0\cos\varphi + n_i^0\sin\phi - n_i\sin 2\phi\right)\right].$$

Hence, the total scattering cross-section of the waves, according to the general expression (4.12), is

$$Q_T(\omega) = \frac{\pi^2}{8}a(\beta a)^3\left[\frac{1}{(\rho_0 v_T^2)^2}(C_{66}^A)^2(1+\zeta^4)\right.$$
$$\left. + \left(\frac{\rho_1}{\rho_0}\right)^2(1+\zeta^2)\right], \quad v_T^2 = C_{66}^0/\rho_0. \tag{5.17}$$

As it follows from (5.13) and (5.17), the total scattering cross-sections of the longitudinal and shear waves polarized in the x_1x_2-plane do not contain any dielectric or piezoelectric constants. This was to be expected, since the x_1x_2-plane is the plane of isotropy, so that the piezoelectric behaviour does not manifest itself. The situation, however, is quite different when shear waves, polarized in the x_3-direction, are considered.

5.3. SHEAR WAVES POLARIZED IN x_3-DIRECTION

Indeed, we have in this case,

$$U_k^0 = m_k, \quad k = \beta_\perp, \quad \langle I^0\rangle_t = -\frac{1}{2}\omega\beta_\perp C'_{44}. \tag{5.18}$$

Obviously, $A_i(n) = B_i(n) = 0$ and the vector $f_k(\beta_\perp)$ and the scalar $f(\beta_\perp)$ become

$$f_k(\beta_\perp n) = \frac{i\pi a^2}{2}\sqrt{\frac{\beta_\perp^3}{2\pi}}\mathrm{e}^{-i\pi/4}\left[\frac{\rho_1}{\rho_0} - \frac{1}{C'_{44}}\left(C_{44}^A + \frac{e_{15}^A e_{15}^0}{\eta_{11}^0}\right)\right]\cos\phi,$$
$$f(\beta_\perp n) = -\frac{i\pi a^2}{2}\sqrt{\frac{\beta_\perp^3}{2\pi}}\mathrm{e}^{-i\pi/4}\frac{e_{15}^0}{\eta_{11}^0 C'_{44}}\left(e_{15}^A + \frac{e_{15}^0}{\eta_{11}^0}\eta_{11}^A\right)\cos\phi. \tag{5.19}$$

It follows, then, that

$$C_i(n) = \frac{i\pi a^2}{2}\sqrt{\frac{\beta_\perp^3}{2\pi}}\mathrm{e}^{-\frac{i\pi}{4}}\left\{\frac{\rho_1}{\rho_0} - \frac{1}{C'_{44}}\left[C_{44}^A + \left(\frac{e_{15}^0}{\eta_{11}^0}\right)^2 \eta_{11}^A\right]^2 \cos\phi\right\} m_i. \quad (5.20)$$

The total scattering cross-section for these waves is

$$Q_{T\perp}(\omega) = \frac{\pi^2}{8}(\beta_\perp a)^3 a \left\{\frac{1}{(\rho_0 v_{T\perp}^2)^2}\left[C_{44}^A + \left(\frac{e_{15}^0}{\eta_{11}^0}\right)^2 \eta_{11}^A\right]^2 + 2\left(\frac{\rho_1}{\rho_0}\right)^2\right\}, \quad (5.21)$$

$$v_{T\perp}^2 = C'_{44}/\rho_0.$$

For the purely elastic behaviour ($e_{15}^0 = 0$) this expression coincides with the one obtained in [12].

6. CONCLUSIONS

The obtained results for the scattering amplitudes and cross-sections of a circular scatterer in a piezoelectric medium of hexagonal (transversely isotropic) symmetry may be useful for many future applications, e.g., for the determination of the symmetry of the scatterer by measuring its scattering cross-section. The Green's function method, presented in Section 3, can be extended to scatterers of arbitrary symmetry. Here a similar amplitude equation as (3.16) occurs wherein the scattering amplitudes reflect the symmetry of the scatterer. Thus the presented method will hopefully stimulate further work in the treatment of the scattering of acoustoelectric waves at inhomogeneities.

REFERENCES

1. Gubernatis, J. E., E. Domany, J. A. Krumhansl. Formal aspects of the theory of the scattering of ultrasound by flows in elastic materials. *J. Applied Physics*, **48**, 1977, 2804-2811.
2. Gubernatis, J. E., E. Domany, J. A. Krumhansl, M. Huberman. The Born approximation in the theory of the scattering of elastic waves by flows. *J. Applied Physics*, **48**, 1977, 2812–2819.
3. Gubernatis, J. E., J. A. Krumhansl, R. M. Thomson. Interpretation of elastic wave scattering theory for analysis and design of flow characterization experiments: The long-wave limit. *J. Applied Physics*, **50**, 1979, 3338–3344.
4. Gubernatis, J. E. Long-wave approximations for the scattering of elastic waves from flows with applications to ellipsoidal voids and inclusions. *J. Applied Physics*, **50**, 1979, 4046–4058.
5. Kohn, W., J. R. Rice. Scattering of long-wavelength elastic waves from localized defects in solids. *J. Applied Physics*, **50**, 1979, 3346–3353.

6. Willis, J. R. A polarization approach to the scattering of elastic waves. I. Scattering by a single inclusion. *J. Mech. Phys. Solids*, **28**, 287–305.

7. Benveniste, Y. On the micromechanics of fibrous piezoelectric composites. *Mech. Mater.*, **18**, 1994, 183–193.

8. Huang, J. H., J. S. Yu. Electroelastic Eshelby's tensors for an ellipsoidal piezoelectric inclusion. *Compos. Eng.*, **4**, 1994, 1169-1182.

9. Dunn, M. L. Electroelastic Green's functions for transversely isotropic media and their application to the solution of inclusion and inhomogeneity problem. *Int. J. Eng. Sci.*, **32**, 1994, 119–131.

10. Chen, T. Micromechanical estimates of the overall thermoelectroelastic moduli of multiphase fibrous composites. *Int. J. Solids Structures.*, **31**, 1994, 3099–3111.

11. Levin, V. M. The effective properties of piezoactive matrix composite materials. *J. Appl. Math. Mech.*, **60**, 1996, 309–311.

12. Kanaun, S. K., V. M. Levin. The effective field method in mechanics of composite materials. Petrozavodsk, 1993, 600 pp. (In Russian.)

APPENDIX

The needed formulae, concerning the tensorial basis (3.4) used in the paper, are collected here. Their application, as already demonstrated, allows one to substantially simplify and standardize the appropriate tensorial operations in the problem under consideration.

If a certain tensor **A** is expressed in the **T**-basis as

$$\mathbf{A} = A_1\mathbf{T}^2 + A_2\left(\mathbf{T}^1 - \frac{1}{2}\mathbf{T}^2\right) + A_3\mathbf{T}^3 + A_4\mathbf{T}^4 + A_5\mathbf{T}^5 + A_6\mathbf{T}^6, \tag{A.1}$$

then the inverse tensor $\mathbf{A}^{-1}$ is given by the expression

$$\begin{aligned}\mathbf{A}^{-1} = &\frac{A_6}{2\Delta}\mathbf{T}^2 + \frac{1}{A_2}\left(\mathbf{T}^1 - \frac{1}{2}\mathbf{T}^2\right) - \frac{A_3}{\Delta}\mathbf{T}^3 - \frac{A_4}{\Delta}\mathbf{T}^4 \\ &+ \frac{4}{A_5}\mathbf{T}^5 + \frac{2A_1}{\Delta}\mathbf{T}^6, \quad \Delta = 2(A_1A_6 - A_3A_4).\end{aligned} \tag{A.2}$$

If two tensors **A** and **B** are given in the **T**-basis, the contraction of these tensors with respect to two pairs of indices reads

$$\begin{aligned}A_{ijkl}B_{klmn} = &(2A_1B_1 + A_3B_4)T^2_{ijmn} + A_2B_2\left(T^1_{ijmn} - \frac{1}{2}T^2_{ijmn}\right) \\ &+ (2A_1B_3 + A_3B_6)T^3_{ijmn} + (2A_4B_1 + A_6B_4)T^4_{ijmn} \\ &+ \frac{1}{2}A_5B_5T^5_{ijmn} + (A_6B_6 + 2A_4B_3)T^6_{ijmn}.\end{aligned} \tag{A.3}$$

Consider now two tensors **C** and **D**, presented in the **U**-basis,

$$C_{ijk} = \sum_{r=1}^{3} C_r U^r_{ijk}, \quad D_{ijk} = \sum_{s=1}^{3} D_s U^s_{ijk}. \tag{A.4}$$

The contraction of these tensors with respect to one index gives the tensor in **T**-basis:

$$C_{ijm} D^T_{mkl} = C_1 D_1 T^2_{ijkl} + C_1 D_3 T^3_{ijkl} + C_3 D_1 T^4_{ijkl} + 4C_2 D_2 T^5_{ijkl} + C_3 D_3 T^6_{ijkl}. \tag{A.5}$$

The contraction of the tensors **C** and **D** with respect to two pairs of indices gives a tensor, which is presented in the **t**-basis as

$$C^T_{ikl} D_{klj} = 2C_2 D_2 t^1_{ij} + (2C_1 D_1 + C_3 D_3) t^2_{ij}. \tag{A.6}$$

It can be shown that the **t**-basis is orthogonal in the sense that if

$$\alpha_{ij} = \alpha_1 t^1_{ij} + \alpha_2 t^2_{ij}, \quad \beta_{ij} = \beta_1 t^1_{ij} + \beta_2 t^2_{ij}, \tag{A.7}$$

then

$$\alpha_{ik}\beta_{kj} = \alpha_1\beta_1 t^1_{ij} + \alpha_2\beta_2 t^2_{ij} \tag{A.8}$$

and

$$\alpha^{-1}_{ij} = \frac{1}{\alpha_1} t^1_{ij} + \frac{1}{\alpha_2} t^2_{ij}. \tag{A.9}$$

The following formulae are also useful:

$$\begin{aligned}
A_{ijmn} C_{mnk} &= (2A_1 C_1 + A_3 C_3) U^1_{ijk} + \frac{1}{2} A_5 C_2 U^2_{ijk} + (2A_4 C_1 + A_6 C_3) U^3_{ijk}, \\
C^T_{imn} A_{mnkl} &= (2C_1 A_1 + C_3 A_4) U^{1T}_{ijk} + \frac{1}{2} C_2 A_5 U^{2T}_{ijk} + (2C_1 A_3 + C_3 A_6) U^{3T}_{ijk}, \\
\alpha_{im} C^T_{mkl} &= \alpha_2 C_1 U^{1T}_{ikl} + \alpha_1 C_2 U^{2T}_{ikl} + \alpha_2 C_3 U^{3T}_{ikl}, \\
C_{ijm}\alpha_{mk} &= C_1 \alpha_2 U^1_{ijk} + C_2 \alpha_1 U^2_{ijk} + C_3 \alpha_2 U^3_{ijk}.
\end{aligned} \tag{A.10}$$

Received February 29, 2000

Valery M. LEVIN
Division of Mechanics
Petrozavodsk State University
Lenin Ave. 33
Petrozavodsk 185640
Russia
E-mail: `levin@levin.pgu.karelia.su`

Thomas MICHELITSCH
Institute for Theoretical Physics I
University of Stuttgart
Pfaffenwaldring 57/4
D-70550 Stuttgart
Germany
E-mail: `michel@theo.physik.uni-stuttgart.de`

ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 93, 1999

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI“

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Tome 93, 1999

NUMERICAL APPROACH TO PATTERN SELECTION IN A MODEL PROBLEM FOR BÉNARD CONVECTION IN FINITE FLUID LAYER

J. PONTES, C. I. CHRISTOV and M. G. VELARDE

Long wave-length pattern formation is studied by means of numerical integration of a fourth-order in space nonlinear evolution equation subjected to Dirichlet lateral boundary conditions. Computationally efficient implicit difference scheme and algorithm are devised employing the method of operator splitting.
The case of Bénard convection in Boussinesq limit is considered. For different sets of the parameters different convective planforms are found: a pattern of hexagons (H^+) with upward flow in their centers, hexagons (H^-) with downward flow in the centers, coexisting hexagons and squares (S), and a case where the squares are selected. In the case when the critical wave-number vanishes (the wave-length diverges) the pattern selected is of a single cell which fills the whole domain under consideration.

Keywords: difference schemes, operator splitting, higher-order diffusion equations, pattern selection, nonlinearity, surface-tension driven convection

MSC 2000: 35K55, 65M06, 74S20, 65Z05

1. INTRODUCTION

Pattern formation in a thin layer of fluid heated from below occurs when the vertical temperature gradient exceeds certain threshold [12, 20, 10, 6]. The nature of the instability and the characteristics of the convective motion depend not only on the fluid parameters but also on the geometric and physical properties of the container. In sufficiently deep cells, or in cells in which the fluid is confined between rigid horizontal boundaries, the convective motion settles when the buoyancy force overcome the viscous forces (Rayleigh-Bénard problem). In sufficiently shallow layers with open surfaces, the inhomogeneities of the surface-tension distribution are responsible for the onset of the motion (Bénard-Marangoni problem) [3, 13]. Depending on whether the horizontal boundaries are good thermal conductors or not, the characteristic wave-length of the convective structure is either comparable to or is much larger than the depth of the cell.

In this work we deal with the problem of pattern selection and the long-term evolution of the planar field in horizontally limited systems, subject to rigid boundary conditions at the sidewalls: $u = \partial u/\partial n = 0$, where u is the temperature.

Different amplitude equations are derived in the literature as simplified models for the convective motion *in lieu* of the full system of the compressible Navier-Stokes (N-S) equations. Here belongs the Swift-Hohenberg (S-H) equation (see, for instance, [6]) and its generalizations [1], as well as the Knobloch equation [8, 9], to mention a few. The main difference between these two models is that the S-H equation has a Lyapunov potential, while the Knobloch equation has not, being thus physically much closer to the original model based on the full Navier-Stokes equations.

Knobloch's equation (1) is very similar to the 2D Navier-Stokes (N-S) system not only as a physical model but as a mathematical structure as well. It is simpler in the sense that it is a scalar equation containing no pressure, and hence no incompressibility constraint has to be satisfied. On the other hand, Eq. (1) is more complicated than the N-S one on the account of the numerous nonlinear terms which make it a multi-parametric model with a rich phenomenology.

Following [5], we employ the method of operator splitting to construct an efficient difference scheme and algorithm for solving the generalized diffusion equation (1) which contains fourth-order spatial derivatives. This is the numerical objective of the present paper. The proximity of the solution to its asymptotic state is assessed via monitoring the L_1-norm of the difference between two consecutive time steps; this norm is sensitive not only to the *changes* in the amplitude of the structure but also to the evolution of its *phase*.

The second objective of the present paper is to obtain physically relevant results. In this instance, the emphasis is placed on three aspects. First, the pattern selection is examined and the results are compared to the existing ones, particularly to those obtained in the framework of the Knobloch's equation. Second, we find numerically the possible non-stationary asymptotic states toward which the system evolves. Third, the long-time evolution of the patterns is tracked. In all these cases the wave-length and the spectral content of the patterns in terms of Fourier modes are thoroughly examined.

2. POSING THE PROBLEM

Chapman & Proctor [4] and Sivashinsky[15, 16] introduced an equation to describe the long wave-length pattern formation induced by buoyancy or by surface-tension instabilities in a thin layer of fluid. This equation was later generalized by Knobloch to the form

$$\frac{\partial u}{\partial t} = \alpha u - \mu \nabla^2 u - \nabla^4 u + \kappa \nabla \cdot |\nabla u|^2 \nabla u \\ + \beta \nabla \cdot \nabla^2 u \nabla u - \gamma \nabla \cdot u \nabla u + \delta \nabla^2 |\nabla u|^2. \tag{1}$$

Here, $u(x, y, t)$ is the horizontal planform of the temperature deviation from the conductive profile, x, y and t are *slow* variables, μ is the scaled bifurcation parameter, and α represents the effect of finite thermal conductivity of the horizontal boundaries, or finite Biot number. When the boundary conditions at the top and bottom of layer are not identical, then $\beta \neq 0$ and $\delta \neq 0$. Respectively, if non-Boussinesq effects are to be taken into account, one has $\gamma \neq 0$. In the present paper we do not deal with non-Boussinesq effects, hence, we set $\gamma = 0$. The coefficient κ can always be set to ± 1, see [8], except for solidification in binary alloys, where it vanishes. Specifically, in the case of Bénard's convection, $\kappa = +1$.

Knobloch considered modes forming square and hexagonal lattices, but he did not address the question of relative stability between squares and hexagons, and suggested that the problem should be studied numerically. He did not consider the case in which $\kappa = -1$, nor did he interrogate the non-stationary patterns that could emerge due to the non-potential character of Eq. (1). Shtilman & Sivashinsky [14] integrated Eq. (1) numerically in a square region of approximately 4×4 wavelengths, subjected to periodic boundary conditions. They set $\gamma = 0$, $\beta > \delta$ and obtained a structure of positive hexagons quite in accordance with experimental findings in Bénard–Marangoni convection and the analytical results of Knobloch for this range of parameters. Recently, Skeldon & Silber [17] extended the stability analysis performed by Knobloch and found some scenarios, where transition from hexagons to rectangles may occur. In some cases, these authors found that near onset some more exotic spatially periodic planforms are preferred to the usual rolls, squares and hexagons.

The existing results, concerning pattern selection in the framework of the dynamics represented by Eq. (1), can be summarized as follows:

- Square lattices:
  - Squares are stable if $\beta = \gamma = \delta = 0$;
  - Rolls are stable if $\beta = \delta \neq 0$ and $\gamma \neq 0$;
  - Rolls are stable for $\beta \neq 0$ and $\beta - \delta \neq 0$ if $\gamma = 0$.
- Hexagonal lattices:
  - Hexagons are stable if $\beta = \gamma = \delta = 0$;
  - If $|\beta + \gamma/q_c^2 - \delta| \ll |\beta|$, $|\delta|$, $|\gamma| = \mathcal{O}(1)$, the hexagons are stable at low amplitudes, having upward (H^+) or downward (H^-) flows in their centers depending on the sign of $(\beta + \gamma/q_c^2 - \delta)$, while the rolls are stable at larger amplitudes;
  - H^+ and H^- coexist at large amplitudes if $|\beta|, |\gamma|, |\delta| \ll 1$.

In this study we consider the Bénard convection in Boussinesq approximation when $\kappa = +1$ and $\gamma = 0$. Only in one of the calculations for the squared tessellations we use $\kappa = -1$ for the sake of comparison of the physical mechanisms. The simulations are conducted in the above selected parameter range for a square box with horizontal dimensions 150×150 (approximately 20 to 23 wave-lengths per side). Eq. (1) is integrated numerically in the points of a square grid of 402×402 points, which means roughly 17.5 points per wave-length. This is a significant improvement of the resolution in comparison with the works from the literature. The time-step used in our simulations is $\Delta t = 0.1$.

Section 4.1 presents the results concerning H^+ or H^- hexagons. Section 4.2 discusses squared tessellations S, obtained both with $\kappa = +1$ and $\kappa = -1$. Section 4.3 deals with the problem of the coexistence between squares and hexagons. In Section 4.4 the result of a simulation for $\alpha = 0$ is presented. This is a case when the critical wave-length diverges (zero critical wave-number).

3. NUMERICAL SCHEME

Following [5], we use here the idea of operator splitting to create efficient scheme which will allow extensive numerical experiment. In order to secure the desired properties of the operators to be inverted, we use a semi-implicit approximation of the nonlinear terms. First we recast the original PDE to a form stemming from a first order discrete representation of the time derivative:

$$\frac{u^{n+1} - u^n}{\Delta t} = \left(\Lambda_x^n + \Lambda_y^n\right) u^{n+1} + f^n, \tag{2}$$

or

$$\left[I - \left(\Lambda_x^n + \Lambda_y^n\right)\right] u^{n+1} = \Delta t \left(u^n + f^n\right), \tag{3}$$

where I is the identity operator, and u^n, u^{n+1} stand for the dependent variable on the respective time stage. Respectively, the operators Λ_x^n, Λ_y^n and the function f^n are defined as follows:

$$\Lambda_x^n = \frac{\alpha}{2} - \frac{\partial^4}{\partial x^4} + \kappa \frac{\partial}{\partial x}\left(|\nabla u^n|^2 \frac{\partial}{\partial x}\right),$$

$$\Lambda_y^n = \frac{\alpha}{2} - \frac{\partial^4}{\partial y^4} + \kappa \frac{\partial}{\partial y}\left(|\nabla u^n|^2 \frac{\partial}{\partial y}\right),$$

$$f^n = -2\frac{\partial^4 u^n}{\partial x^2 \partial y^2} - \mu \nabla^2 u^n + \beta \nabla \cdot \nabla^2 u^n \nabla u^n - \gamma \nabla \cdot u^n \nabla u^n + \delta \nabla^2 |\nabla u^n|^2.$$

Here it becomes clear why the scheme is called "semi-implicit". The nonlinear terms involving the third derivatives and proportional to the coefficient k are taken in divergent form in which the coefficients are from the "old" time-stage, while the derivatives itself are treated implicitly. Generally speaking, some more terms can be approximated implicitly, but such an approximation will destroy the negative definiteness of the operators to be inverted at each time step. For this reason we leave them on the "old" time stage.

Now the full inversion of the operator $[I - (\Lambda_x^n + \Lambda_y^n)]$ requires a considerable computational effort. We use here the operator-splitting scheme of *stabilizing correction*. We split Eq. (3) into the following two half-time steps [5]:

$$(I - \Delta t\, \Lambda_x^n)\, \tilde{u} = \left(I + \Delta t\, \Lambda_y^n\right) u^n + \Delta t\, f^n, \qquad (4)$$

$$\left(I - \Delta t\, \Lambda_y^n\right) u^{n+1} = \tilde{u} - \Delta t\, \Lambda_y^n\, u^n. \qquad (5)$$

Upon applying the operator $(I - \Delta t\, \Lambda_x^n)$ to Eq. (5) and adding the result to Eq. (4) one obtains

$$\left(I + (\Delta t)^2 \Lambda_x^n \Lambda_y^n\right) \frac{u^{n+1} - u^n}{\Delta t} = \left(\Lambda_x^n + \Lambda_y^n\right) u^{n+1} + f^n,$$

which means that within the second order approximation in time, the splitting scheme is equivalent to the original semi-explicit scheme (2). The advantage of the splitting is twofold. The operator in the left-hand side of Eqs. (4) and (5) displays a penta-diagonal structure when the space operators are approximated to second order on the difference level. The elements of these operators are just numbers and not matrices, as in the case of Eq. (3). Second, Eq. (4) can be solved line by line and Eq. (5) can be solved column by column, reducing considerably the storage requirements. The five-diagonal systems were then solved by Gaussian elimination with pivoting. The scheme (4), (5) generalizes for the fourth-order diffusion operators the classical operator splitting scheme of *stabilizing correction* [7, 21].

When a stationary pattern is investigated, the boundary conditions and coefficients of equation do not depend explicitly on the time t and the computed transient solution should converge to the steady solution unless a chaotic régime onsets. Of prime importance for computing the steady solutions is the selection of the criterion to judge whether the convergence is reached since the *amplitude* and the *phase dynamics* of the pattern evolve on different time scales. The phase evolves much slower than the amplitude does. Then the uniform norm of the difference between two consecutive iterations will follow the rate with which the maximal amplitude of the transient approaches the maximal amplitude of the steady solution. Using an uniform norm would send a false signal that solution converges long before the phase pattern reaches its stationary shape. Following [5], we track the time evolution of patterns by monitoring an L_1-type of norm which measures the rate of change of the *distance* between two successive states of the system. The L_1-norm is sensitive to the evolution of both the amplitude and the phase:

$$L_1 = \frac{1}{\Delta t} \frac{\sum_{i,j} |u_{i,j}^{n+1} - u_{i,j}^n|}{\sum_{i,j} |u^{n+1}|}, \qquad (6)$$

where the sums are made over all interior points of the grid.

4. NUMERICAL RESULTS

In this section we present the results of our numerical simulations of the Knobloch equation (1). All simulations begin at $t = 0$, from a random initial condition. As already mentioned in the foregoing discussion, we focus our attention on the case $k > 0$. We choose without loosing the generality $k = 1$. Only in one of the runs we use $\kappa = -1$ for the sake of comparison. More exhaustive treatment of the non-Bénard case $\kappa < 0$ is due elsewhere.

We consider only the cases when $\alpha \leq 0$, which means that there is no energy input proportional to the temperature u. For $\alpha < 0$ the said term accounts for a linear attenuation of the solution. Then for the dispersion relation for the Fourier modes of the linear part of Eq. (1) in unbounded region and for the interval of unstable wave-numbers we get

$$s = \alpha + \mu q^2 - q^4\,, \quad \frac{\mu}{2} - \sqrt{\frac{\mu^2}{4} - |\alpha|} < q < \frac{\mu}{2} + \sqrt{\frac{\mu^2}{4} - |\alpha|}\,,$$

where the negative sign of α is acknowledged.

One sees that there is a whole band of unstable modes. The fastest growing mode (the largest positive s) is $q_m = \sqrt{\mu/2}$ with exponent $s_m = \alpha + \frac{1}{2}\sqrt{\mu^2}$. In most of the cases treated in the present work q_m does not have much impact on the results and cannot effectively serve as a representative critical number. For this reason we propose a somewhat more elaborate definition of the critical wave number q_c. Namely, we find the value of μ for which $s_m = 0$, i.e. the value for which an unstable mode first appears. Clearly, this can happen only for $\mu = 2\sqrt{|\alpha|}$. Then we call "critical wave-number" the magnitude of q_m for this particular value of μ, $q_c = \sqrt{|\alpha|}$. As it will become clear in what follows, this critical wave-number is rather relevant to the wave motions under consideration.

The hexagon pattern with an upward flow in the centers of hexagons is denoted by H^+. Respectively, H^- stands for the pattern in which the flow in the centers of hexagons is downward. The grey shades of the plots are selected between white (regions with the most rapid upward flow) and black (the fastest downward flow). In order to gather more information about the motion, the Fourier transform of the pattern is shown in the figures. To this end, we include also a panel in the figures showing the sum of the amplitudes of all wave-numbers A in the interval between q and $q + \Delta q$, regardless to their orientation. Respectively, Δq is the grid spacing of the discrete Fourier transform. The respective result is depicted in the lowest panels of the figure as a function of (q/q_c). The position q_f of the peak identifies the fundamental mode of the structure which is, in general, different from the critical mode q_c.

4.1. FINITE WAVE-LENGTH HEXAGONS

The starting point of our simulations is the numerical work [14]. We adopt the same values used by those authors for the coefficients of the quadratic terms, namely $\beta = -0.125\sqrt{7}$, $\alpha = -0.8$ and $\mu = 2.7$.

For $\delta = -0.75\sqrt{7} < 0$, the system evolves into a pattern of positive hexagons which we designate as H^+. The structure obtained at the end of the simulation ($t = 19050$) is shown in the top-left panel of Fig. 1.

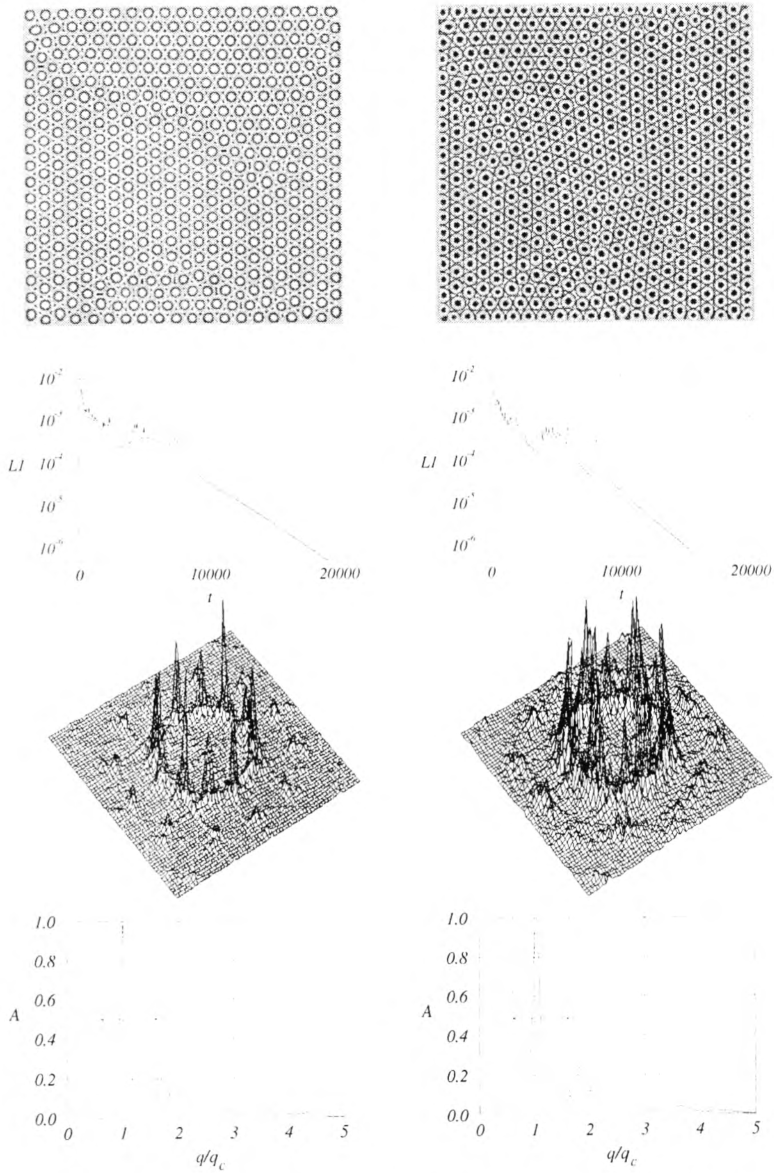


Fig. 1. Hexagons in square geometry for $\mu = 2.7$, $\alpha = -0.8$, $\beta = -0.125\sqrt{7}$. Left panels: H^+ at $t = 19050$ for $\delta = -0.75\sqrt{7}$. Right panels: H^- at $t = 15990$ for $\delta = 0.75\sqrt{7}$. From top to bottom: the pattern; the time evolution of L_1-norm; the Fourier transform; the wave-number content of the pattern

In the beginning, the small scales existing in the random initial condition are rapidly attenuated due to the filtering action of the dissipative part $\alpha - \nabla^4$ of the linear operator of Eq. (1). The most significant growth of the amplitude occurs in this phase and the decrease of the L_1-norm is very fast. The second phase is essential for the pattern evolution. It is characterized by an irregular behaviour of L_1 which on average decays much slower than in the initial stage. As the structure evolves, new cells are created or annihilated, dislocations move, and grains slide relative to each other. The number of defects usually diminishes as longer space correlations are created and the whole pattern moves. Tribelsky *et al.* show in [19] that just before and after a cell is created or annihilated, the evolution accelerates according to a power law. The peaks of the $L_1(t)$-curve reflect the time moments when *qualitative* changes occur in the pattern. Eventually, the system finds a configuration where further qualitative changes become extremely difficult and then the third stage of evolution begins characterized by a sustained exponential decay of the L_1-norm. We terminate the simulation after the L_1-norm decreases another order of magnitude since the moment of onset of the third phase of the evolution.

The pattern shown in the top-left panel of Fig. 1 contains three *grains*, the first one being adjacent to the upper boundary and having one of the axis of the lattice parallel to that wall. A second smaller grain of the same lattice orientation occupies a part of the lower wall. The third (the biggest) grain occupies the central part of the box and presents a lattice rotated by an angle of $\pi/6$ with respect to the other two. Several *hepta-penta* defects can be observed on the grain boundaries.

The third panel in the left column of Fig. 1 shows the core part of Fourier modes of the pattern at the end of the simulation $t = 19050$. Twelve peaks can be observed in the ring of fundamental modes, reflecting the existence of lattices with *two* orientations. In addition, this figure shows also the existence of secondary peaks (the bottom-left panel in Fig. 1) which can be explained using the arguments presented graphically in Fig. 2.

In order to identify the relative importance of the amplitude of the modes, irrespective to their orientation, we computed the sum (denoted by A) of the amplitudes found in circular rings of the Fourier transform centered in $q = 0$ and limited by

$$n\Delta q - \Delta q/2 \leq q < n\Delta q + \Delta q/2,$$

where $\Delta q = 2\pi/l$ and $n = 1, 2, \ldots$ The values of A, obtained as a function of q/q_c and normalized by $\max(A)$, are plotted in the bottom-left panel of Fig. 1. The curve displays an absolute maximum in $q/q_c = 0.98$, which we define as q_f/q_c, with q_f being the *fundamental* wave-number of the pattern. There are also several smaller peaks associated with modes $\sqrt{3}\,q_f$, $2\,q_f$, $\sqrt{7}\,q_f$ and $2\sqrt{3}\,q_f$. It is seen that the fundamental *wave-length* of the pattern is smaller than the critical one. We also outline by dashed vertical lines the limits on the band of linearly unstable modes. It can be seen that there are secondary peaks *out* of that band, i.e. active modes with negative eigenvalue. The next four relevant peaks are located in $q/q_c = \sqrt{3}, q_f/q_c$, $2q_f/q_c$, $2.65 \approx \sqrt{7}, q_f/q_c$ and $2\sqrt{3}\,q_f/q_c$. The amplitude of the peaks diminishes as q/q_c increases due to the fact that the eigenvalue of a mode becomes more and more negative, as the distance to the band of linearly unstable modes increases. This reveals an important signature of weakly-nonlinear systems: the interaction between modes promoted by the nonlinear dynamics merely introduces corrections to the fundamental modes, the latter being the most important in defining the essential features of the structure.

Thess & Bestehorn [18] found by direct integration of Navier-Stokes equations that an H^--structure appears in Bénard-Marangoni convection if the Prandtl number of the fluid is smaller than a critical value. We encounter negative hexagons H^- for positive $\delta = 0.75\sqrt{7}$. The size of the system and the values of the rest of the parameters are the same as in the preceding case H^+, except for the coefficient δ, for which we took the opposite sign.

The pertinent drawings are shown in the right panels of Fig. 1. In this case the end of the simulation appears at $t = 15990$. Now the system develops a more complex structure than in the H^+-case, which is also reflected by the Fourier transform of the pattern. Several *dislocations* and *hepta-penta defects* are observed. Yet, the curve $L_1(t)$ displays the same qualitative features found in the previous case and the time interval required to attain a sustained exponential decay of $L_1(t)$ is also comparable ($t = 15990$ here and $t = 19050$ in the previous configuration). The higher level of disorder of the H^--case is compatible with the higher forcing, applied to the system, which is reflected by a somewhat wider band of linearly unstable modes. The bottom-right panel of Fig. 1 shows that H^- contains the same structure of Fourier modes as the one found for H^+. The fundamental wave-number of the pattern is greater than the critical wave-number, namely $q_f = 1.03\, q_c$.

Fig. 2 focuses on an extended area around the core part of the Fourier transform of the H^+-pattern shown in the respective panels of Fig. 1. Fig. 2 is constructed using a nonlinear scale of gray shades, in order to enhance the weaker peaks for better observation. It is seen that the peaks fall in five concentric circumferences, whose diameters follow the same relations found for the abscissa of the peaks shown in the lowermost panels of Fig. 1. The first one (1) contains the fundamental modes of the pattern, q_f. Six directions (12 peaks) are observed on this circumference, due to the existence of grains with *two* orientations in the pattern under consideration. The second circumference (2) with radius $q = \sqrt{3}\, q_f$ contains wave-vectors generated by the interaction of fundamental modes of the hexagonal lattice (see also Fig. 2(b)). The third circumference (3) contains the second harmonic of the structure $2\, q_f$. The fourth circumference (4) with radius $q = \sqrt{7}\, q_f$ contains modes generated by the interaction of $\mathbf{q}_f$ and $2\,\mathbf{q}_f$, forming an angle of $\pi/3$. Two different orientations are possible in each $\pi/3$ sector leading to the existence of 12 *pairs* of peaks in this circumference, see Fig. 2(b) and Fig. 2(c). The fifth circumference contains modes with wave-number $q = 2\sqrt{3}\, q_f$. Fig. 2(d) displays a scheme of the pattern in the physical space as contained in each cell.

Twelve peaks can be observed in circumferences # 2, 3 and 5, but circumference # 4 contains twelve *pairs* of peaks. Fig. 2(b) and Fig. 2(c) clarify the origin of these peaks.

Fig. 2(b) shows a hexagon in the Fourier space composed of fundamental modes. The interaction of two aligned fundamental modes results in the second

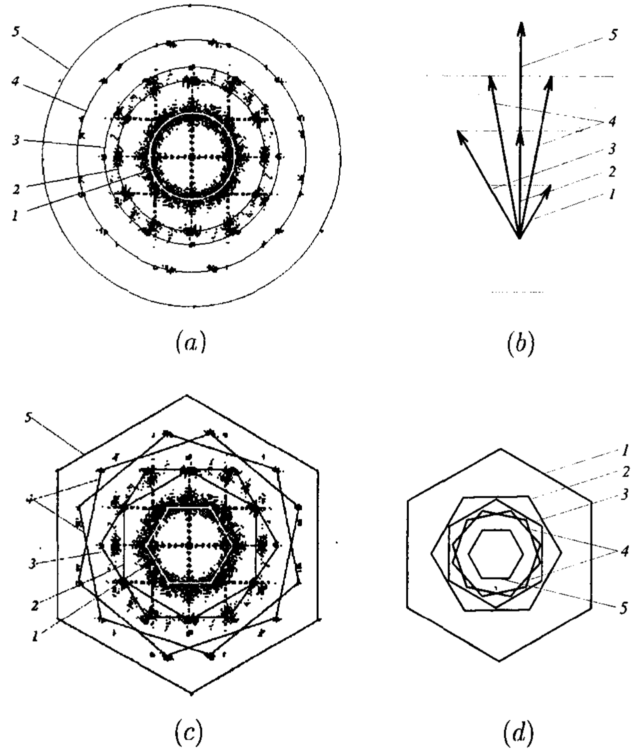


Fig. 2. H^+ in square geometry

harmonic, the interaction of two wave-vectors $\mathbf{q}_f$, forming an angle of $\pi/3$, gives a mode with $q = \sqrt{3}\, q_f$, the interaction of the latter one with itself originates the mode $2\sqrt{3}\, q_f$ and the interaction of a $\sqrt{3}\, q_f$-mode with a fundamental one creates a $\sqrt{7}\, q_f$-mode. In this case there are *two* possibilities in each $\pi/3$ sector. This is why twelve pairs of peaks appear in the circumference # 4.

Fig. 2(c) shows the hexagons in Fourier space, obtained by linking the peaks associated to one of the lattices of the pattern. Fig. 2(d) shows a scheme of a convective cell in the configurational space. Thus we compare the normalized fundamental mode q_f/q_c of the pattern, with the normalized average mode q_a/q_c, where the average mode is defined as

$$q_a = 2\pi \sqrt{\frac{N}{S_a}}\,, \tag{7}$$

where N is the number of cells in the pattern, and S_a is the area of the box.

Due to the existence of empty spaces at the grain boundaries, close to the sidewalls, the above estimate gives as a rule an average wave-number smaller than

q_f. In addition, the pattern contains some cells (like the heptagons) which are larger than the average cell.

4.2. FINITE WAVE-LENGTH SQUARES

We turn now to the question of identifying the conditions leading to square tessellations. Knobloch [9] shows that squares are stable when the system is a Boussinesq fluid with symmetric horizontal boundaries ($\beta = \gamma = \delta = 0$). Shtilman & Sivashinsky [14] obtain the same result by numerically integrating Eq. (1) in a 4×4 wave-length box assuming periodic boundary conditions. The squares can also occur when $\beta = \delta \neq 0$. We consider two configurations of parameters, one of them with $\kappa = +1$ and the other one with $\kappa = -1$.

The case $\kappa = +1$ evolves until $t = 41100$. The high forcing, applied to the system ($\mu - \mu_c = 0.602$), justifies the rather disordered structure obtained in this case (see the left panels in Fig. 3). This behavior is captured also by the Fourier transform. The most conspicuous feature of the pattern is the orientation of the dominant lattice parallel to the diagonals of the box (note that it is parallel to the sidewalls in the case $\kappa = -1$). The normalized fundamental mode is $q_f/q_c = 1$. The consecutive peaks are associated with the modes $\sqrt{2}\,q_f$, $2q_f$, $\sqrt{5}\,q_f$ and $3\,q_f$.

The pattern obtained for $\kappa = -1$ (right column of panels in Fig. 3) displays an almost perfect structure, possibly due to the lower forcing applied to the system. It consists of essentially one *grain* with several defects (*dislocations*). The almost perfect structure of the pattern is reflected by a very clear but rich Fourier spectra. There is no circle of fundamental modes. The inner part of the transform displays a *squared* geometry. Fundamental modes are found only in the directions parallel to the sidewalls and — as observed in the early stages of evolution — a second lattice, parallel to the diagonals of the box, coexists with the dominant lattice. The secondary lattice of modes clearly stems from the interaction of two orthogonal fundamental modes. The bottom-right panel of Fig. 3 shows that the normalized fundamental mode is $q_f/q_c = 1.01$, and several subsequent peaks, located at $\sqrt{2}\,q_f/q_c$, $2q_f/q_c$, $2.23q_f/q_c \approx \sqrt{5}\,q_f/q_c$ and $3q_f/q_c$. It is interesting to mention that the modes with $q = 2\sqrt{2}\,q_f$, which have a wave-vector shorter than $3q_f$, do not survive.

Among the cases, discussed in the present work, this is the most demanding in terms of computational effort, requiring 2×10^6 steps to attain the convergence to the steady solution, $t = 200000$ with time increment $\Delta t = 0.1$.

4.3. COEXISTENCE OF FINITE WAVE-LENGTH SQUARES AND HEXAGONS

The results from the previous sections show that hexagons are selected if β differs significantly from δ and that squares appear when $\beta = \delta$. Then the natural question is whether hexagons and squares can coexist if β differs slightly from δ. The second question is of whether or not H^+ and H^- can coexist if $|\beta|$, $|\delta| \ll 1$.

The answer to the first question is affirmative. The hexagons and squares do coexist when the value of β is in the vicinity of δ. Furthermore, the hexagons are

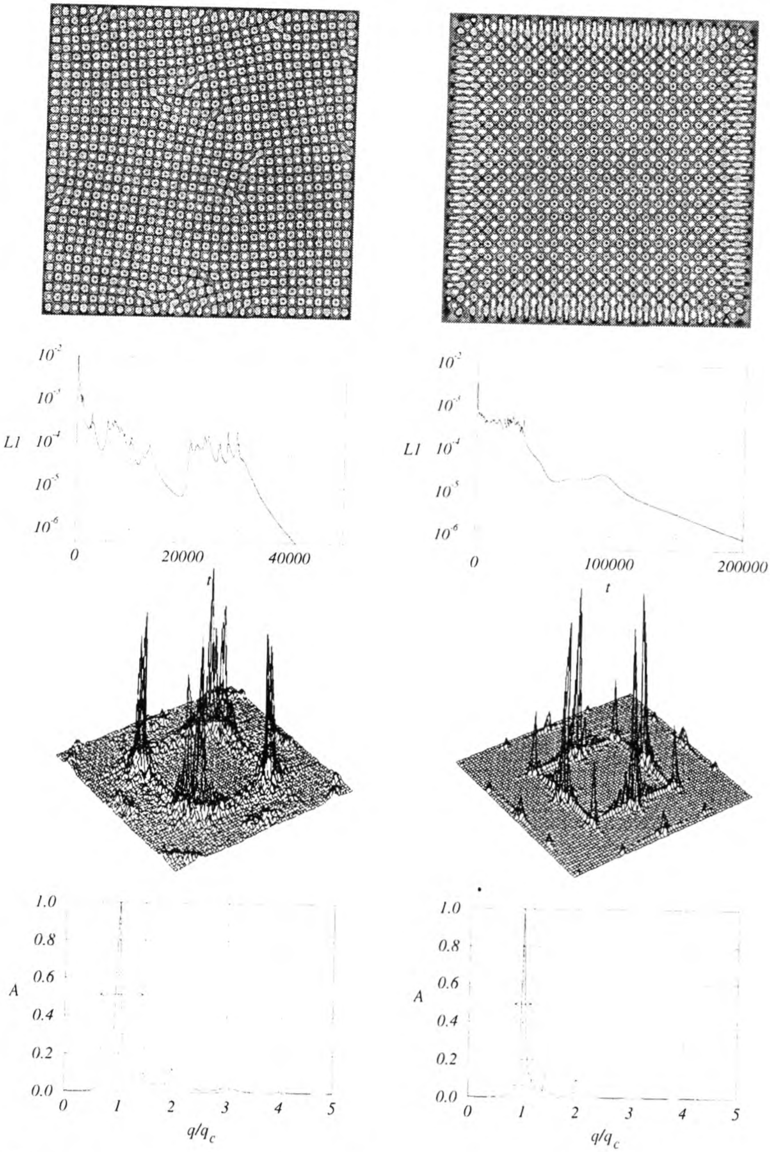


Fig. 3. Squared tessellations. Left panels: S-pattern for $\kappa = +1$ at $t = 41100$ and $\mu = 2.5$, $\alpha = -0.9$, $\beta = \delta = -0.25\sqrt{7}$. Right panel: S-pattern for $\kappa = -1$ at $t = 200000$ for $\mu = 2.0$, $\alpha = -0.9$, $\beta = \delta = -3\sqrt{7}$. From top to bottom: the pattern; the time evolution of L_1-norm; Fourier transform; wave-number content of the pattern shown

positive or negative, depending on the sign of $(\beta - \delta)$, exactly as in the case of the single-pattern tessellations, studied in Section 3.1. However, the conditions for the coexistence of both patterns appear to be much less robust than those that define a single pattern. We find that the coexistence of patterns depends not only on the coefficients of the quadratic terms of Eq. (1) but also on the magnitude of the forcing applied and on the system size, as well. A change in one of these parameters may lead to the collapse of one of the patterns. For instance, with the increase of the forcing, the dominant pattern changes from hexagons to squares. This result was also obtained in [2] by direct integration of the three-dimensional Navier-Stokes equations. In [11] the competition between hexagons and squares in a generalized Swift-Hohenberg equation was studied. It was found there that the front between competing patterns must be perpendicular to their corresponding modes in order to be stable.

The answer to the second question is negative. We do not find a coexistence of positive and negative hexagons. Squares emerge in the neighbourhood of the point where the transition $H^+ \longrightarrow H^-$ occurs, i.e. in the point where the sign of $(\beta - \delta)$ changes, and the transition occurs in the form $H^+ \longrightarrow S \longrightarrow H^-$, or *vice versa*. Besides, one of the hexagon types disappears *before* the onset of the other, and the squares are the only stable pattern in the case when $\beta = \delta$.

As in all of the previous figures the top panels of Fig. 4 show the state of the system at the end of the simulation. The drawings corresponding to the evolution of L_1-norm are placed on the next row. The third row consists of the Fourier transforms of the patterns. The last row shows the wave-number content of the pattern.

In the first numerical experiment (the left column of panels in Fig. 4) we set β to a value, slightly bigger than δ, which is supposed to be a case in the interface between the regions of H^+ and S. Indeed, our simulations show that the system eventually evolves to a state of coexistence of both patterns (designated by SH^+). Note that the system shows a tendency to develop squares close to the sidewalls. There are lines along which the distance between hexagons, and in some cases between squares, is larger than the average one. These lines are "fault lines" across which the neighbouring domains with different patterns slide during the evolution of the pattern towards a steady state. The orientations of the hexagonal and the square lattices are rather deformed and in some regions the directions of the dominant lattice keep rotating in time.

The Fourier transform indicates that the dominant modes fall in a circular ring of nearly critical modes. The modes are somewhat more evenly distributed along the ring than in the cases where the orientation of the lattice is less deformed. The fundamental wave-number here is $q_f = 1.01 q_c$. However, the curve of the wave content as a function of q/q_c (bottom-left panel of Fig. 4) does not suggest the existence of secondary modes, as it is the case in the above treated sets of parameters.

For the second example of coexistence we set β and δ so as to place the system in the limit between the regions of H^- and S. Our results show that it evolves to a coexistence of hexagons and squares, whose pattern is much more ordered than in the previous case. Hexagons appear close to the upper, left and right sidewalls, and also in the lower part of the box when the domain is larger. The squares occupy

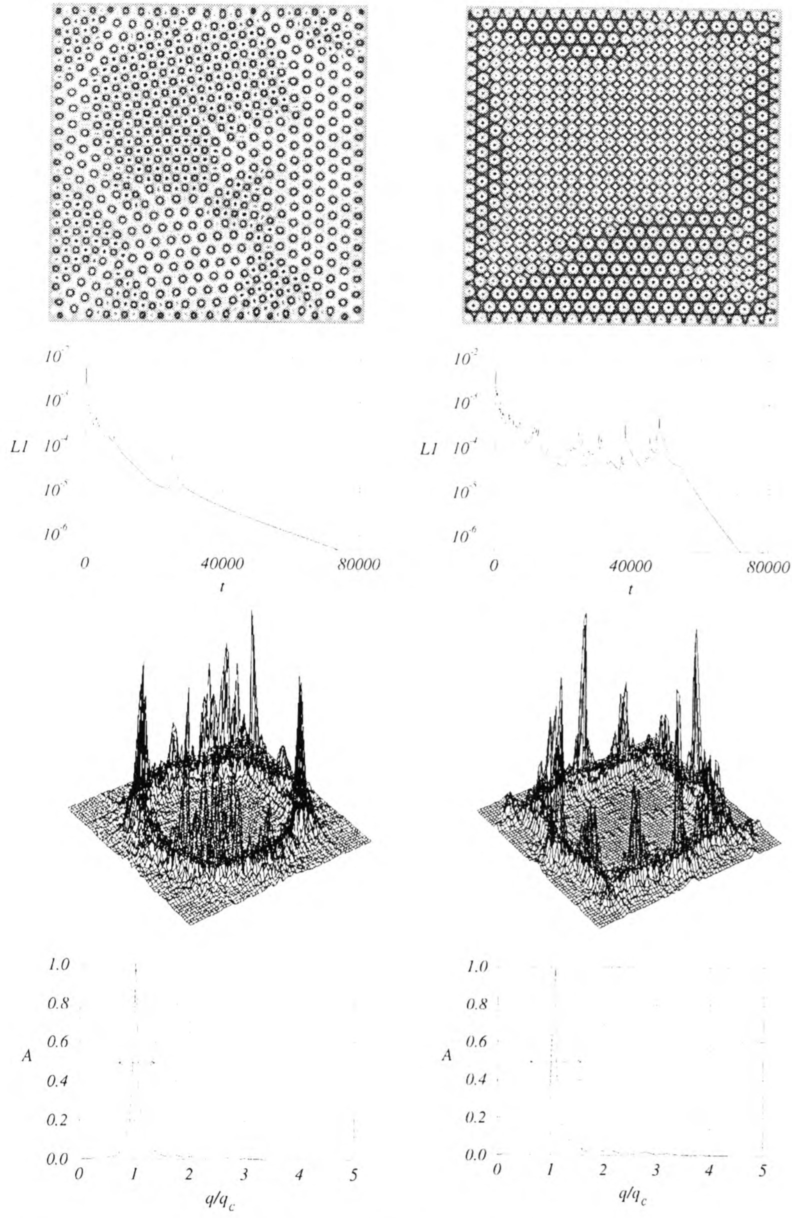


Fig. 4. Coexistence of squares S and hexagons H^+ or H^-. Left panels: SH^+ at $t = 73860$ for $\mu = 2.3$, $\alpha = -0.9$ and $\beta = -0.05\sqrt{7}$, $\delta = -0.07\sqrt{7}$. Right panels: SH^- at $t = 71950$ for $\mu = 2.8$, $\alpha = -0.9$ and $\beta = 0.1\sqrt{7}$, $\delta = 0.17\sqrt{7}$. From top to bottom: the pattern; the time evolution of the L_1-norm; Fourier transform; wave-number content of the pattern shown

mostly the inner part of the system.

The Fourier transform displays a squared structure, which is in a sharp contrast with the circular structure found in the previous case. Each corner of the square contains a small peak. In addition, there are three other peaks along each side of the square totaling twelve along the all four sides. The peak in the center of each side is slightly off the square and the twelve peaks actually fall in a circular ring of fundamental modes. These peaks define two hexagonal lattices rotated by an angle of $\pi/6$ with respect to each other. The hexagons, belonging to one of the lattices, appear close to the left and the right sidewalls of the system. The hexagons, belonging to the second lattice, appear closer to the upper sidewall. The two lattices do *not* coexist in the same parts of the configurational space. Both the image of the structure and the associated Fourier transform suggest that the amplitudes of the hexagons, belonging to each lattice, are of the same order, because they belong to the ring of fundamental modes.

The structure of squares results from the superposition of a lattice of fundamental modes with $q = q_f$ and a second one, rotated by an angle of $\pi/4$ relatively to the former and having $q = \sqrt{2}\, q_f$.

A different situation occurs with the other two lattices which are associated with the structure of squares. The first lattice is defined by two directions parallel to the sidewalls and by modes falling in the ring of fundamental modes. The four corresponding peaks in the Fourier transform coincide with those of the two hexagonal lattices which, consequently, are the highest peaks. The second lattice is rotated by an angle of $\pi/4$ with respect to the former, with modes $\sqrt{2}\, q_f$, generated by the nonlinear interaction of the fundamental modes. Here, the two square lattices do coexist in the same physical space.

In the last case the wave-number content of the pattern exhibits a maximum at $q_f = 1.07\, q_c$. The average wave-number, measured according to Eq. (7), is also larger than the critical one.

4.4. PATTERNS FOR VANISHING LINEAR ATTENUATION COEFFICIENT $\alpha = 0$

In this section we discuss the numerical findings for $\alpha = 0$ when, according to dispersion equation, the lower limit of the unstable wave-number is zero. The coefficients of the quadratic terms were selected to place the system in the H^--régime. Then all scales with $q < \sqrt{\mu}$ are supposed to be linearly unstable while the scales with $q > \sqrt{\mu}$ are damped.

The numerically obtained time evolution of the pattern is shown in Fig. 5. Smaller cells are indeed damped first. The more important finding, however, is that the cells increase with time and the system evolves towards a state where a single cell eventually occupies the whole box. This is somewhat counter-intuitive because one might expect in this case an onset of regular pattern with the most unstable wave-number $q_m = \sqrt{\mu/2} = 1$ for the selected value $\mu = 2$. Moreover, the box we have chosen is large enough to harbour more than 12 wave-lengths q_m, and hence this is not an effect connected with the distortion of the fastest-growing modes, due to the lack of space for their spatial extent.

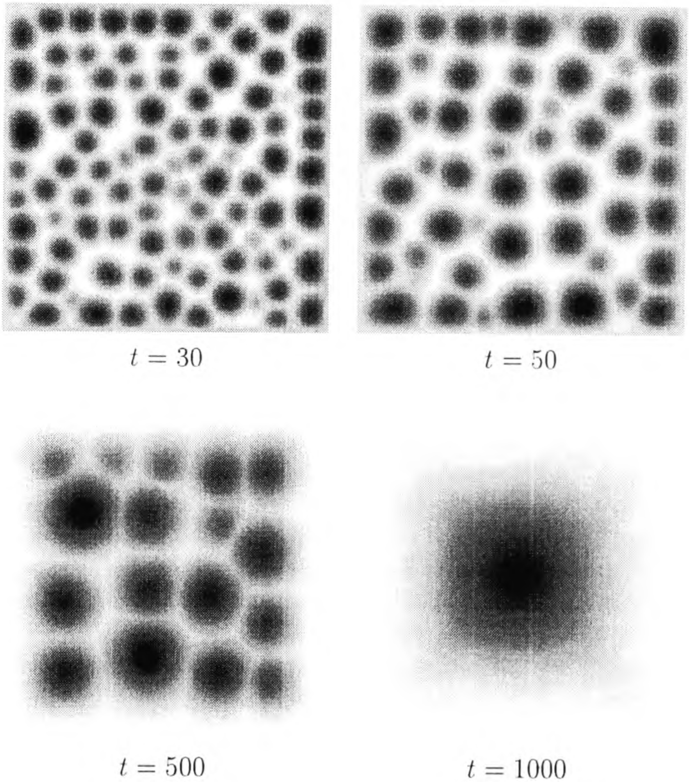


Fig. 5. Evolution of the H^{-}-pattern for $\alpha = 0$, $\mu = 2.0$, $\beta = -0.125\sqrt{7}$, and $\delta = 0.75\sqrt{7}$

Thus the interpretation of this simulation is that patterns with a characteristic size of the cell are obtained when larger than the box size wave-lengths are present among the linearly unstable wave-lengths of the system. The evolution eventually ends up in a stationary single-cell state, filling the whole bounding box. The influence of the boundaries on the motion makes the fastest growing (more unstable) modes to disappear eventually and only the largest convective cell, compatible with the box size, survives.

5. CONCLUSIONS

In the present work we investigate the properties of long wave-length patterns formed in systems which evolve according to the non-variational dynamics described by Eq. (1). The non-Boussinesq effects are neglected by setting $\gamma = 0$ in Eq. (1).

We found hexagonal tessellations H^+ (H^-) for $\beta > \delta$ ($\beta < \delta$) and tessellations of squares for $\beta = \delta$. By setting β to a value slightly larger than δ we identified states of coexistence between H^+ and S. Similarly, we found coexisting H^- and S patterns when choosing β to be slightly smaller than δ. These states of coexistence are less robust than those of single patterns. They depend on the forcing applied and on the horizontal dimensions of the system. A change in one of these parameters may lead to the disappearance of one of the patterns.

Typically, the speed of evolution of the patterns shows a rapid decay in the early stages. In the intermediate stages the speed is rather irregular because of the creation or annihilation of cells, transport of dislocations, deformation of grains. These are evolutions of the *phase* of the pattern in general. The last stage is a sustained monotone decay of the L_1-norm, suggesting that the system eventually attains a stationary state. The onset of the sustained decay serves as a criterion to terminate the computations.

The analysis of the spectral content of these patterns reveals that the fundamental wave-number of the structure is very close to the critical. The largest discrepancy occurs in the case of coexistence of hexagons H^- and squares S, where we find a fundamental mode which is 7% larger than the critical. The Fourier analysis also shows the existence of active modes *outside* the band of linearly unstable modes. In the case of single-pattern tessellations, these modes are not merely higher harmonics of the fundamental mode, but they can also originate from the interactions between modes with different orientations and/or wave-numbers. The ability of the system to generate and sustain secondary active modes leads in the case of square patterns to the coexistence, in the same subdomain in the configurational space, of a dominant lattice and a secondary one, the latter slanted by an angle $\pi/4$ with respect to the former.

For $\alpha = 0$ our calculations indicate that all scales shorter than the measures of the box are eliminated as the system evolves in time from a random initial condition. A single-cell stationary pattern is eventually reached, having the largest finite wave-length which is compatible with the dimensions of the box.

For $\kappa = +1$ we observe onset of patterns with larger density of defects when increasing the forcing, but no indication of the existence of possible unsteady asymptotic states. Our preliminary results for the regime $\kappa = -1$ (not shown in the present paper) suggest that at least in two cases there appear unsteady patterns of irregular polygons at high forcing. The numerical interrogation of this case will be published elsewhere.

Acknowledgements. The authors acknowledge fruitful discussions with G. Huang, H. Linde and A. A. Nepomnyashchy.

This work was supported by DGICYT 0.(Spain) under Grant PB 93–81, by the European Union under Grant Network ERBCHRXCT 93–107 and by "Fundación Ramón Areces". J.P. also acknowledges the support of a post-doctoral fellowship from the Spanish Ministry of Education and Science and fellowship No. 30167895-2 from the Brazilian National Council for the Research and Development (CNPQ).

The work of C.I.C. is supported in part by Grant LEQSF(1999-2002)-RD-A-49 from the Louisiana Board of Regents.

REFERENCES

1. Bestehorn, M. Phase instabilities for Bénard-Marangoni convection in fluid layers with large aspect ratio. *Phys. Rev. E*, **48**, 1993, 3622–3634.
2. Bestehorn, M. Square patterns in Bénard-Marangoni convection. *Phys. Rev. Lett.*, **76**, 1996, 46–49.
3. Block, M. J. Surface tension as the cause of Bénard cells and surface deformation in a liquid film. *Nature*, **178**, 1956, 650–651.
4. Chapman, C. J., M. R. E. Proctor. Non-linear Rayleigh-Bénard convection between poorly conducting boundaries. *J. Fluid Mech.*, **101**, 1980, 759–782.
5. Christov, C. I., J. Pontes, D. Walgraef, M. G. Velarde. Implicit time splitting for fourth-order parabolic equations. *Comp. Meth. Appl. Mech. & Eng.*, **148**, 1997, 209–224.
6. Cross, M. C., P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, **65**, 1993, 851–1112.
7. Douglass, J., H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, **82**, 1956, 421–439.
8. Knobloch, E. Pattern selection in long wavelength convection. *Physica D*, **41**, 1990, 450–479.
9. Knobloch, E. Onset of zero Prandtl number convection. *J. Phys. II France*, **2**, 1992, 995–999.
10. Koschmieder, E. L. Bénard cells and Taylor vortices. Cambridge Univ. Press, Cambridge, UK, 1993.
11. Kubstrup, C., H. Herrero, C. Pérez-Garcia. Fronts between hexagons and squares in a generalized Swift-Hohenberg equation. *Phys. Rev. E*, **54**, 1996, 1560–1569.
12. Normand, C., Y. Pomeau, M. G. Velarde. Convective instability: A physicist's approach. *Rev. Mod. Phys.*, **49**, 1977, 581–624.
13. Pearson, J. R. A. On convection cells induced by surface tension. *J. Fluid Mech.*, **4**, 1958, 89–500.
14. Shtilman, L., G. Sivashinsky. Hexagonal structure of large-scale Marangoni convection. *Physica D*, **52**, 1991, 477–488.
15. Sivashinsky, G. I. Large cells in nonlinear Marangoni convection. *Physica D*, **4**, 1982, 227–235.
16. Sivashinsky, G. I. On cellular instability in the solidification of a dilute binary alloy. *Physica D*, **8**, 1983, 243–248.
17. Skeldon, A. C., M. Silber. New stability results for patterns in a model of long wavelength convection. *Physica D*, **122**, 1998, 117–133.
18. Thess, A., M. Bestehorn. Planform selection in Bénard-Marangoni convection: *l* hexagons versus *g* hexagons. *Phys. Rev. E*, **52**, 1995, 6358–6367.
19. Tribelsky, M. I., S. Kai, H. Yamazaki, M. G. Velarde. Universal spatiotemporal scaling

in the dynamics of one-dimensional pattern selection. *Phys. Rev. E*, **51**, 1995, 5132–5135.

20. Velarde, M. G., C. Normand. Convection. *Sci. Am.*, **243**, 1980, 92–108.
21. Yanenko, N. N. Method of fractional steps. Gordon and Breach, NY, 1971.

Received March 24, 2000

J. PONTES
Dept. of Metallurgical
and Material Engineering
COPPE/UFRJ, PO Box 68505
21945-970 Rio de Janeiro, RJ
BRAZIL

C. I. CHRISTOV
Corresponding author
Department of Mathematics
University of Louisiana, Lafayette
LA 70504-1010, USA
E-mail: christov@louisiana.edu

M. G. VELARDE
Instituto Pluridisciplinar
Universidad Complutense de Madrid
Paseo Juan XXIII No. 1
E-28040 Madrid, SPAIN

Submission of manuscripts. The *Annuaire* is published once a year. No deadline exists. Once received by the editors, the manuscript will be subjected to rapid, but thorough review process. If accepted, it is immediately sheduled for the nearest forthcoming issue. No page charge is made. The author(s) will be provided with a total of 30 free of charge offprints of their paper.
The submission of a paper implies that it has not been published, or is not under consideration for publication elsewhere. In case it is accepted, it implies as well that the author(s) transfers the copyright to the Faculty of Mathematics and Informatics at the "St. Kliment Ohridski" University of Sofia, including the right to adapt the article for use in conjuction with computer systems and programs and also reproduction or publication in machine-readable form and incorporation in retrieval systems.

Instructions to Contributors. Preferences will be given to papers, not longer than 15 to 20 pages, written preferably in English and typeset by means of a TEX system. A simple specimen file, exposing in detail the instruction for preparing the manuscripts, is available upon request from the electronic address of the Editorial Board. Two copies of the manuscript should be submitted. Upon acceptance of the paper, the authors will be asked to send by electronic mail or on a diskette the text of the papers and the appropriate graphic files (in any format like `*.tif`, `*.pcx`, `*.bmp`, etc.).
The manuscripts should be prepared for publication in accordance with the instructions, given below.
The manuscripts must be *typed* on one side of the paper in double spacing with wide margins. On the *first* page the author should provide: a title, name(s) of the author(s), a short abstract, a list of keywords and the appropriate 2000 Mathematical Subject Classification codes (primary and secondary, if necessary). The affiliation(s), including the electronic address, is given at the end of the manuscripts.
Figures have to be inserted in the text near their first reference. If the author cannot supply and/or incorporate the graphic files, drawings (in black ink and on a good quality paper) should be enclosed separately. If photographs are to be used, only black and white ones are acceptable.
Tables should be inserted in the text as close to the point of reference as possible. Some space should be left above and below the table.
Footnotes, which should be kept to a minimum and should be brief, must be numbered consecutively.
References must be cited in the text in square brackets, like [3], or [5, 7], or [11, p. 123], or [16, Ch. 2.12]. They have to be numbered either in the order they appear in the text or alphabetically. Examples (please note order, style and punctuation):

For books: Obreshkoff, N. Higher algebra. Nauka i Izkustvo, 2nd edition, Sofia, 1963 (in Bulgarian).

For journal articles: Frisch, H. L. Statistics of random media. *Trans. Soc. Rheology*, **9**, 1965, 293–312.

For articles in edited volumes or proceedings: Friedman, H. Axiomatic recursive function theory. In: *Logic Colloquim 95*, eds. R. Gandy and F. Yates, North-Holland, 1971, 188–195.