ANNALES
UNIVERSITATIS SCIENTIARUM BUDAPESTINENSIS DE ROLANDO EÖTVÖS NOMINATAE

SECTIO MATHEMATICA

TOMUS XXXV.

1992

ANNALES
UNIVERSITATIS SCIENTIARUM BUDAPESTINENSIS DE ROLANDO EÖTVÖS NOMINATAE

SECTIO BIOLOGICA
incepit anno MCMLVII
SECTIO CHIMICA
incepit anno MCMLIX
SECTIO CLASSICA
incepit anno MCMXXIV
SECTIO COMPUTATORICA
incepit anno MCMLXXVIII
SECTIO GEOGRAPHICA
incepit anno MCMLXVI
SECTIO GEOLOGICA
incepit anno MCMLVII
SECTIO GEOPHYSICA ET METEOROLOGICA
incepit anno MCMLXXV
SECTIO HISTORICA
incepit anno MCMLVII
SECTIO IURIDICA
incepit anno MCMLIX
SECTIO LINGUISTICA
incepit anno MCMLXX
SECTIO MATHEMATICA
incepit anno MCMLVIII
SECTIO PAEDAGOGICA ET PSYCHOLOGICA
incepit anno MCMLXX
SECTIO PHILOLOGICA
incepit anno MCMLVII
SECTIO PHILOLOGICA HUNGARICA
incepit anno MCMLXX
SECTIO PHILOLOGICA MODERNA
incepit anno MCMLXX
SECTIO PHILOSOPHICA ET SOCIOLOGICA
incepit anno MCMLXII

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 3–6

АСИМПТОТИКА РЕШЕНИЙ ЛИНЕЙНЫХ СИСТЕМ ПО ПЕРВОМУ ПРИБЛИЖЕНИЮ

Й. ФОДОР

Кафедра прикладного анализа университета им. Л. Этвеша, Будапешт

(Поступило 13. 7. 1988)

Вместе с исходной системой

(1) $$\dot{x} = A(t)x$$

в $\mathbb{R}^n$ с ограничеными кусочно непрерывными при $t \geq 0$ коэффициентами рассмотрим и возмущенную систему

(2) $$\dot{y} = A(t)y + f(t,y)$$

где f — возмущения (определены на произведении $[0,+\infty) \times \mathbb{R}^n$ кусочно непрерывны по t) порядка $m > 1$ (кратко m-возмущения) т.е.

$$f(t,0) \equiv 0, \quad |f(t,x_1) - f(t,x_2)| \leq g(t) \max\{|x_1|^{m-1}, |x_2|^{m-1}\}|x_1 - x_2|$$
$$g(t) \geq 0 \quad \text{непрерывная функция. (см. [1] и [3])}$$

Пусть $e_1(t), e_2(t), \ldots, e_n(t)$ такой нормальный базис для (1), что $e_1(0)=[1,0,\ldots,0]^*$, $e_2(0)=[0,1,0,\ldots,0]^*$, $\ldots$, $e_n(0)=[0,0,\ldots,0,1]^*$. Далее обозначим

$$P_i = \operatorname{diag}[\overset{1}{\smile}\!\!\!\!0, 0, \ldots, 0, \overset{i}{\smile}\!\!\!\!1, 0, \ldots, \overset{n}{\smile}\!\!\!\!0],$$

$$P = \sum_{i=1}^{k} P_i, \quad Q = \sum_{i=k+1}^{n} P_i \quad \text{и} \quad I = P + Q.$$

Введем функцию Грина

$$G(t,s) := \begin{cases} X(t)PX^{-1}(s) & \text{если} \quad t \geq s \geq 0, \\ -X(t)QX^{-1}(s) & \text{если} \quad s > t \geq 0, \end{cases}$$

где X — операторная функция Коши уравнения (1), т.е.

$$X(t) := [e_1(t), e_2(t), \ldots, e_n(t)].$$

Рассмотрим интегральное уравнение

$$(3)\qquad y(t)=X(t)a+\int\limits_0^\infty G(t,s)f(s,y(s))\mathrm{d}s.$$

Легко увидеть, что всякая функция, удовлетворяющая этому уравнению является также решением (2) ($t\geq 0$, $a\in\mathbb{R}^n$) (см. [3] стр. 148).

Можно показать, что

$$(4)\qquad |X(t)P_iX^{-1}(s)|=\frac{|e_i(t)|}{|e_i(s)|\sin\alpha_i(s)}$$

$$\text{где}\quad \alpha_i(s)=\sphericalangle\{X(s)P_i\mathbb{R}^n,X(s)(I-P_i)\mathbb{R}^n\}$$

(см. [3] стр. 151–153).

Пусть $\beta(\cdot):[0,+\infty)\to(0,+\infty)$ непрерывная функция, что

$$|e_\ell(t)|\leq\beta(t),\qquad t\geq 0$$

для некоторого фиксированного $\ell\in\{1,\dots,n\}$.

Предложим далее, что

$$(5)\qquad \int\limits_0^t \frac{g(s)\beta^m(s)}{|e_i(s)|\sin\alpha_i(s)}\mathrm{d}s\leq c\frac{\beta(t)}{|e_i(t)|},\quad i\in\{1,\dots,k\},\quad t\geq 0$$

и

$$(6)\qquad \int\limits_t^{+\infty} \frac{g(s)\beta^m(s)}{|e_i(s)|\sin\alpha_i(s)}\mathrm{d}s\leq c\frac{\beta(t)}{|e_i(t)|},\quad i\in\{k+1,\dots,n\},\quad t\geq 0$$

где $0<c=\text{const}$.

Рассмотрим оператор T:

$$(Ty)(t):=\int\limits_0^{+\infty} G(t,s)f(s,y(s))\mathrm{d}s,\qquad t\geq 0,$$

в пространстве $\mathbb{B}(\beta)$ в котором вектор-функциями являются ($t\geq 0$), подчинённые условию

$$\sup_{t\geq 0}\frac{|y(t)|}{\beta(t)}=:\|y\|<+\infty.$$

Известно, что $\mathbb{B}(\beta)$ есть банахово пространство в норме $\|\cdot\|$ (см. [2] стр. 510).

Лемма 1. Если $y \in \mathbb{B}(\beta)$, то $Ty \in \mathbb{B}(\beta)$.

Доказательство.

$$\|Ty\| = \sup_{t\geq 0} \frac{1}{\beta(t)} \left| \int_0^{+\infty} G(t,s) f(s,y(s)) \mathrm{d}s \right| \leq$$

$$\leq \sup_{t\geq 0} \frac{1}{\beta(t)} \left[\sum_{i=1}^{k} \int_0^t |X(t) P_i X^{-1}(s)| g(s) |y(s)|^m \mathrm{d}s + \right.$$

$$\left. + \sum_{i=k+1}^{n} \int_t^{+\infty} |X(t) P_i X^{-1}(s)| g(s) |y(s)|^m \mathrm{d}s \right] \leq$$

$$\leq c\|y\|^m \sup_{t\geq 0} \left[\sum_{i=1}^{k} \frac{|e_i(t)|}{\beta(t)} \int_0^t \frac{g(s)\beta^m(s)}{|e_i(s)| \sin\alpha_i(s)} \mathrm{d}s + \right.$$

$$\left. + \sum_{i=k+1}^{n} \frac{|e_i(t)|}{\beta(t)} \int_t^{+\infty} \frac{g(s)\beta^m(s)}{|e_i(s)| \sin\alpha_i(s)} \mathrm{d}s \right] \leq n \cdot c\|y\|^m.$$

Аналогично предыдущей лемме получаем:

Лемма 2. Если y_1, $y_2 \in \mathbb{B}(\beta)$, то

$$\|Ty_1 - Ty_2\| \leq nc \max\{\|y_1\|^{m-1}, \|y_2\|^{m-1}\} \|y_1 - y_2\|.$$

Пусть $r > 0$ такой, что $q = ncr^{m-1} < 1$, $\delta = (1-q)r$ и

$$S_r := \{y : y \in \mathbb{B}(\beta),\ \|y\| < r\}.$$

Из лемм 1. и 2. следует

Теорема. *Если у системы (1) и функции выполняется (5) и (6), то при любом $0 < \varepsilon < \delta$ существует решение y системы (2) такое, что*

$$|y(t)| \leq \frac{\varepsilon}{1-q} \beta(t) \quad \text{и} \quad |y(t) - \varepsilon e_\ell(t)| \leq \frac{q\cdot\varepsilon}{1-q} \beta(t), \quad t \geq 0.$$

Доказательство. Пусть $a = [\overset{1}{\smile}{0}, 0, \ldots, 0, \overset{\ell}{\varepsilon}, 0, \ldots, \overset{n}{0}]$, то $X(t)a = \varepsilon e_\ell(t)$ поэтому $\|Xa\| \leq \varepsilon$.

Рассмотрим оператор Φ_a определённый равенством

$$\Phi_a y = \varepsilon e_\ell + Ty \qquad (y \in \mathbb{B}(\beta)).$$

Φ_a отображает шар S_r в себя и действует в этом шаре как сжимающий с коэффициентом сжатия q.

В самом деле

$$\|\Phi_a y_1 - \Phi_a y_2\| = \|Ty_1 - Ty_2\| \le ncr^{m-1}\|y_1 - y_2\| = q\|y_1 - y_2\|$$

и

$$\|\Phi_a y\| \le \|\varepsilon e_\ell\| + \|Ty\| \le \delta + qr = r.$$

Поэтому уравнение $\Phi_a y = y$ рассматриваемое лишь на шаре S_r при любом $0 < \varepsilon < \delta$ имеет единственное решение y для которого

а) $(1-q)\|y\| \le \varepsilon\|e_\ell\|$,

б) $\|Ty\| \le \dfrac{q}{1-q}\|y\|$,

в) $|y(t)| \le \beta(t)\|y\| \le \dfrac{\varepsilon}{1-q}\beta(t)$,

г) $|y(t) - \varepsilon e_\ell(t)| = |(Ty)(t)| \le \|Ty\|\beta(t) \le \dfrac{q\varepsilon}{1-q}\beta(t)$.

Следствие. При $\beta(t) = |e_\ell(t)|$ и $nc < 1$ выбираем $r < 1$ (это возможно), получим, что у системы (2) при любом $\varepsilon < (1 - ncr^{m-1})r$ существует такое решение y, что

$$(1-r)\varepsilon|e_\ell(t)| \le |y(t)| \le (1+r)\varepsilon|e_\ell(t)|, \qquad t \ge 0.$$

Литература

[1] В. М. Миллионщиков: ДАН СССР, **162** (1965), 266–268.

[2] Былов Б. Ф., Виноград Р. Э., Гробман Д. М., Немыцкий В. В.: *Теория показателей Ляпунова,* Наука, 1966г.

[3] Я. Фодор: О задаче Ляпунова, *Annales Univ. Sci. Budapest.* Sectio Mathematica, **29** (1986), 143–182.

ANNALES UNIV. SCI. BUDAPEST., 35 (1992), 7–14

FILTERFORMITIES AND CARTESIAN CLOSEDNESS

By

D. LESEBERG

Technical University of Braunschweig

(Received June 27, 1991)

This paper is dedicated in honour to Prof. H. HERRLICH (Bremen)

0. Introduction

The category *Top* of topological spaces and continuous maps fails to have some desirable properties, e.g. the product of two quotient maps need not be a quotient map and there is in general no natural function space topology, i.e. *Top* is not cartesian closed. To remedy this situation inconvenient for algebraic topologists, especially homotopists, several substitutes for *Top* have been suggested. M. KATĚTOV invented *filter-merotopic spaces* or simply filter spaces and several other authors examined various categories of *convergence structures* (KENT, NEL, FISCHER, KOWALSKY, CHOQUET, MACHADO, BINZ, KELLER, SCHWARZ, FRÖHLICHER, ANTOINE and others). Cartesian closed categories within the realm of *nearness spaces* were discovered and studied by ROBERTSON, BENTLEY and HERRLICH. *GRILL* is a cartesian closed coreflective subcategory of *SNEAR* and it is isomorphic to the category of filter spaces and merotopically continuous maps. It also contains suitably restricted convergence spaces as a coreflective subcategory, namely those that satisfy the following axiom:

(R_0) If a filter $\underline{F}$ converges to x and y belongs to every member of $\underline{F}$, then $\underline{F}$ converges to y.

In generalizing topological structures, CSÁSZÁR introduced the concept of the so called *syntopogenous spaces.* It contains in particular *Top* but doesn't solve the above mentioned problems. COOK and FISCHER defined *uniform convergence structures* which generalize the uniform spaces in the sense of WEIL. LEE showed that the corresponding category is cartesian closed but unfortunately it doesn't contain all topological spaces. A common generalization of the last two mentioned structures was introduced by the author under the name of *syntopoformic spaces, syntopogenous limit spaces* and *syntopogenous convergence spaces* respectively.

Now the *filterformic spaces* are introduced which fill the gap arising from the existence of all these structures. It is shown that the categories *FMER* of filter spaces and *FCONV* of filter convergence spaces can be considered as a bireflective subcategory of the topological category *FILFORM* of filterformic spaces and maps respectively. On the other hand we prove that *FILFORM* is cartesian closed. Thus, equalities such as $A^{B\times C}\cong\cong(A^B)^C$, $(\Pi A_i)^B\cong\Pi(A_i^B)$, $A^{\coprod B_i}\cong\Pi(A^{B_i})$ and $AX\coprod B_i\cong\coprod(AXB_i)$ are obtained.

1. Filterformities and related structures

1.1. DEFINITIONS: For a set X FIL(X) denotes the set of all filters on X. A function $F:\underline{P}X\rightarrow\underline{P}$ FIL(X) is called a *filterformity* on X, and the pair (X,F) is called a *filterformic space*, iff the following axioms are satisfied:

(Ff1) $A\in\underline{P}X$, $\underline{F}_1\in F(A)$ and $\underline{F}_1\subset\underline{F}_2\in$ FIL(X) imply $\underline{F}_2\in F(A)$,

(Ff2) A_1, $A_2\in\underline{P}X$ and $A_1\subset A_2$ imply $F(A_2)\subset F(A_1)$,

(Ff3) $x\in X$ implies $\dot{x}\in F(\{x\})$; where $\underline{P}X$ denotes the power set of X and $\dot{x}:=\{B\subset X\mid x\in B\}$.

$\underline{F}$ is called an *A-convergent filter* on X iff $\underline{F}\in F(A)$.

Now let (X,F_1), (Y,F_2) be filterformic spaces. A function $f:X\rightarrow Y$ is called a *filterformic map* from (X,F_1) to (Y,F_2) — shortly an *ff-map* — iff the following hold: $A\in\underline{P}X$ and $\underline{F}\in F_1(A)$ imply $f(\underline{F})\in F_2(f[A])$. $f(\underline{F})$ denotes the filter generated by $f\underline{F}$, where

$$f\underline{F}:=\{f[B]\mid B\in\underline{F}\}.$$

The (concrete) category of filterformic spaces and ff-maps is denoted by *FILFORM*.

1.2. EXAMPLES: Let X be a set.

(i) For a filter convergence structure τ on X we put for $A\in\underline{P}X$: $F_\tau(A):=\{\underline{F}\in$ FIL$(X)\mid\forall\,x\in A\ \exists\,\underline{F}'\in\tau(x)\underline{F}'\subset\underline{F}\}$; s. [22],

(ii) For a filter-merotopy Γ on X we put for $A\in\underline{P}X$: $F_\Gamma(A):=\Gamma$; s. [11],

(iii) For a syntopogenous convergence structure $\underline{K}$ on X we put for $A\in\underline{P}X$: s. [15], [16]
$F_{\underline{K}}(A):=\{\underline{F}\in$ FIL$(X)\mid\exists\,\Omega\in\underline{K}\ \forall\,x\in A\ \forall\,t\in\Omega t(\{x\})\subset\underline{F}\}$,

(iv) For a syntopogenous structure $\underline{S}$ on X we put for $A\in\underline{P}X$: $F_{\underline{S}}(A):=$ $:=\{\underline{F}\in$ FIL$(X)\mid\exists\,<\in\underline{S}\ \forall\,x\in A<(x)\subset\underline{F}\}$, where $<(x):=\{B\subset$ $\subset X\mid\{x\}<B\}$; s. [6],

(v) For a uniform limit structure $\underline{L}$ on X we put for $A \in \underline{P}X$: $F_{\underline{L}}(A) := \{\underline{F} \in \mathrm{FIL}(X) \mid \exists\, \underline{U} \in \underline{L}\ \forall\, x \in A\, \underline{U}(x) \subset \underline{F}\}$, where $\underline{U}(x) := \{R(x) \mid R \in \underline{U}\}$ with $R(x) := \{y \in X \mid (x,y) \in R\}$; s. [5],

(vi) For a uniformity $\underline{U}$ on X we put for $A \in \underline{P}X$: $F_{\underline{U}}(A) := \{\underline{F} \in \mathrm{FIL}(X) \mid \exists\, R \in \underline{U}\ \forall\, x \in A\, R(x) \in \underline{F}\}$,

(vii) For a proximal limit structure $\underline{P}$ on X we put for $A \in \underline{P}X$: $F_{\underline{P}}(A) := \{\underline{F} \in \mathrm{FIL}(X) \mid \exists\ \delta \in \underline{P}\ \forall\ x \in A\, \delta(x) \subset \underline{F}\}$, where $\delta(x) := \{B \subset X \mid (X \setminus B, \{x\}) \notin \delta\}$; s. [18].

To obtain nice embeddings of the important categories *FMER* and *FCONV* into *FILFORM* some additional axioms will be now exhibited.

1.3. DEFINITION: For a filterformic space $(X,F)F$ is called a *filtermeroformity* on X if it satisfies the following axiom:

(M) $F(\Phi) \subset F(X)$. Then the pair (X,F) is called a *filtermeroformic space.*

1.4. THEOREM. *The full subcategory MERFORM of FILFORM whose objects are the filtermeroformic spaces is*

(a) *bireflective in FILFORM;*

(b) *isomorphic to the category FMER.*

PROOF. To (a): For a filterformic space (X,F) we put for $A \in \underline{P}X$: $F^m(A) := F(\Phi)$. Then (X,F^m) is a filtermeroformic space and $1_X : (X,F) \to (X,F^m)$ is the *FILFORM* bireflection of (X,F).

To (b): For a filter-merotopy Γ on X we put for $A \in \underline{P}X$: $F_\Gamma(A) := \Gamma$ (see also Example 1.2. (ii)). Then F_Γ is a filtermeroformity on X. Therefore we obtain an isomorphism between the categories *FMER* and *MERFORM*.

1.5. DEFINITION: For a filterformic space (X,F), F is called a *filterconformity* on X if it satisfies the following axiom:

(C) $A \in \underline{P}X$ implies $\bigcap_{x \in A} F(\{x\}) \subset F(A)$.

(In this context, if $A = \Phi$ then $\bigcap_{x \in A} F(\{x\}) = \mathrm{FIL}(X)$.) Then the pair (X,F) is called a *filterconformic space.*

1.6. THEOREM. *The full subcategory CONFORM of FILFORM whose objects are the filterconformic spaces is*

(a) *bireflective in FILFORM;*

(b) *isomorphic to the category FCONV.*

PROOF. To (a): For a filterformic space (X,F) we put for $A \in \underline{P}X$: $F^q(A) := \bigcap_{x \in A} F(\{x\})$. Then (X,F^q) is a filterconformic space and $1_X : (X,F) \to (X,F^q)$ is the *FILFORM* bireflection of (X,F).

To (b): For a filter convergence structure τ on X we put for $A \in \underline{P}X$:

$$F_\tau(A) := \{\underline{F} \in \text{FIL}(X) \mid \forall\, x \in A\ \exists\, \underline{F}' \in \tau(x) \underline{F}' \subset \underline{F}\}$$

(see also Example 1.2. (i)). Then F_τ is a filterconformity on X. Therefore we obtain an isomorphism between the categories *FCONV* and *CONFORM*.

Note that for two filterformities F_1 and F_2 on a set X there exists a natural order by setting:

$$F_1 \leq F_2 :\Leftrightarrow \forall\, A \in \underline{P}X\ F_1(A) \subset F_2(A).$$

1.7. REMARK: From a categorical point of view there is no essential difference between the categories *MERFORM* and *FMER* respectively *CONFORM* and *FCONV*. Since the categories *MERFORM* respectively *CONFORM* are bireflective in *FILFORM*, limits in *MERFORM* (respectively *FMER*) respectively *CONFORM* (respectively *FCONV*) are formed in *FILFORM*. Moreover we observe that *MERFORM* is a full subcategory of *CONFORM*.

2. Categorical properties of *FILFORM*

First we note that for a set X the *FILFORM* – fiber of X, denoted by FILFORM(X), is a set, i.e. FILFORM(X) is the set of all filterformities on X.

In addition we observe that *FILFORM* has the terminal separator property which means that for any set X with cardinality one there exists precisely one element in FILFORM(X).

At last in proving *FILFORM* is topological we have to show the existence of initial structures.

2.1. THEOREM. *For any set X, any family $(X_i, F_i)_{i \in I}$ of filterformic spaces, and any family $(f_i : X \to X_i)_{i \in I}$ of functions there exists a unique filterformity $F_{\{f_i^{-1}\}}$ on X which is initial with respect to the given data $(X, f_i, (X_i, F_i), I)$, i.e. such that for any filterformic space (Y,F) a map $g : (Y,F) \to (X, F_{\{f_i^{-1}\}})$ is an ff-map iff for every $i \in I$ the composite map $f_i \circ g : (Y,F) \to (X_i, F_i)$ is an ff-map.*

PROOF. Let A be a subset of X. We define

$$F_{\{f_i^{-1}\}}(A) := \{\underline{F} \in \text{FIL}(X) \mid \forall\, i \in I\ f_i(\underline{F}) \in F_i(f_i[A])\}.$$

Now it is straightforward to verify the postulated properties. In order to show that *FILFORM* is cartesian closed we firstly prove that for given filterformic spaces (X,F_1), (Y,F_2) there is always a function space Y^X available, structured strongly enough to make the natural evaluation function $e : X \times Y^X \to Y((x,f) \mapsto f(x))$ an ff-map.

2.2. THEOREM. *For any pair $((X,F_1), (Y,F_2))$ of filterformic spaces the set $Y^X := \{f^* \mid f^* : (X,F_1) \to (Y,F_2)$ is an ff-map $\}$ can be supplied in a natural way with a filterformity such that the evaluation map preserves convergence.*

PROOF. Let (X,F_1), (Y,F_2) be filterformic spaces. We define a filterformity F^* on Y^X by setting for $A^* \subset Y^X$:

$$F^*(A^*) := \{\underline{F}^* \in \text{FIL}(Y^X) \mid \forall A \in \underline{P}X \ \forall \underline{F} \in F_1(A) \\ e(\underline{F} \times \underline{F}^*) \in F_2(A^*(A))\},$$

where $e(\underline{F} \times \underline{F}^*)$ denotes the filter generated by $\{e[B \times B^*] \mid B \in \underline{F}, B^* \in \underline{F}^*\}$, and $A^*(A)$ is defined by setting $A^*(A) := \{f^*(x) \mid f^* \in A^*, x \in A\}$. It is easy to verify that F^* fulfils axioms (Ff1) and (Ff2) in the definition of a filterformity (see Definition 1.1). Now we prove F^* also fulfills axiom (Ff3).

Let f^* be an ff-map. We show that $\dot{f}^* \in F^*(\{f^*\})$. Let A be a subset of X and let $\underline{F}$ be an A - convergent filter. We have to verify that $e(\underline{F} \times \dot{f}^*) \in F_2(f^*[A])$. By supposition it remains to prove the statement $f^*(\underline{F}) \subset e(\underline{F} \times \dot{f}^*)$ is valid. $C \in f^*(\underline{F})$ implies $C \supset f^*[B]$ for some $B \in \underline{F}$. Therefore $B \times \{f^*\} \in \underline{F} \times \dot{f}^*$. We show that the following inclusion holds: $e[B \times \{f^*\}] \subset f^*[B]$. $z \in e[B \times \{f^*\}]$ implies $z = e(x, f^*)$ for some $x \in B$, so that $z = f^*(x) \in f^*[B]$ holds which concludes the proof.

In the following step we show that the evaluation map is an ff-map. So let $R \subset X \times Y^X$ and let $\underline{F}$ be an element of $(F_1 \times F^*)(R)$, where $F_1 \times F^*$ denotes the *product filterformity* on $X \times Y^X$, i.e. the filterformity on $X \times Y^X$ that is initial with respect to the data $(X \times \times Y^X, P_X, P_{Y^X}, ((X,F_1),(Y^X,F^*)))$; here P_X denotes the projection from $X \times Y^X$ to X and P_{Y^X} denotes the projection from $X \times Y^X$ to Y^X respectively (see also Theorem 2.1.). The aim is to show that $e(\underline{F}) \in F_2(e[R])$. By supposition we obtain the two statements $P_X(\underline{F}) \in F_1(P_X[R])$ and $P_{Y^X}(\underline{F}) \in F^*(P_{Y^X}[R])$. By definition of F^* we have $e(P_X(\underline{F}) \times P_{Y^X}(\underline{F})) \in \in F_2(P_{Y^X}[R](P_X[R]))$. It remains to prove that the two statements

(i) $P_{Y^X}[R](P_X[R]) \supset e[R]$ and

(ii) $e(P_X(\underline{F}) \times P_{Y^X}(\underline{F})) \subset e(\underline{F})$

are valid, because then by paying attention to the axioms (Ff1) and (Ff2) the desired result follows.

To (i): $y \in e[R]$ implies $y = e(r)$ for some $r \in R$ which means $r = (x, f^*)$ for some $x \in X$, $f^* \in Y^X$, so that $y = f^*(x)$. The following two equations hold:

$$f^* = P_{Y^X}(x, f^*) = P_{Y^X}(r) \quad \text{and} \quad x = P_X(x, f^*) = P_X(r).$$

Therefore the statement $y \in P_{Y^X}[R](P_X[R])$ follows.

To (ii): $E \in e(P_X(\underline{F}) \times P_{Y^X}(\underline{F}))$ implies $E \supset e[B \times B^*]$ for some $B \in P_X(\underline{F})$, $B^* \in P_{Y^X}(\underline{F})$. We have $B \supset P_X[B_1]$ for some $B_1 \in \underline{F}$ and $B^* \supset P_{Y^X}[B_2]$ for some $B_2 \in \underline{F}$. Since $\underline{F}$ is a filter the statement $B_1 \cap B_2 \in \underline{F}$ follows. Now, our goal is to show that $e[B_1 \cap B_2] \subset E$.

We have $e[B_1 \cap B_2] \subset e[P_X[B_1 \cap B_2] \times P_{Y^X}[B_1 \cap B_2]]$, and the following two inclusions hold:

$$P_X[B_1] \supset P_X[B_1 \cap B_2] \tag{1}$$

and

$$P_{Y^X}[B_2] \supset P_{Y^X}[B_1 \cap B_2]. \tag{2}$$

Therefore the inclusion $P_X[B_1] \times P_{Y^X}[B_2] \supset P_X[B_1 \cap B_2] \times P_{Y^X}[B_1 \cap B_2]$ is valid, which shows that $e[P_X[B_1 \cap B_2] \times P_{Y^X}[B_1 \cap B_2]]$ is a subset of the set $e[B \times B^*]$. But now we have $e[B_1 \cap B_2] \subset E$, which concludes the proof.

Now, on the other hand we shall prove that for given filterformic spaces (X, F_1), (Y, F_2) and (Z, F_3), the filterformity F^* on Y^X is weak enough to ensure that for any ff-map f: $(X \times Z, F_1 \times F_3) \to (Y, F_2)$ the associated function $\overline{f} : (Z, F_3) \to (Y^X, F^*)$ is also an ff-map.

2.3. THEOREM. *For a triple $((X, F_1), (Y, F_2), (Z, F_3))$ of filterformic spaces let $f : (X \times Z, F_1 \times F_3) \to (Y, F_2)$ be an ff-map. Then the function $\overline{f} : (Z, F_3) \to (Y^X, F^*)$ defined by $\overline{f}(z)(x) := f(x, z)$ for each $z \in Z$ and for each $x \in X$ is an ff-map.*

PROOF. Let $\overline{A}$ be a subset of Z and let $\underline{F}'$ be an $\overline{A}$-convergent filter on Z, we have to show that $\overline{f}(\underline{F}')$ is an element of $F^*(\overline{f}[\overline{A}])$. In applying the definition of F^* let A be a subset of X and let $\underline{F}$ be an A-convergent filter on X, we must prove that $e(\underline{F} \times \overline{f}(\underline{F}')) \in F_2(\overline{f}[\overline{A}](A))$.

In showing this statement we verify that $\underline{F} \times \underline{F}' \in (F_1 \times F_3)(A \times \overline{A})$. Then by supposition we have $f(\underline{F} \times \underline{F}') \in F_2(f[A \times \overline{A}])$, since f is an ff-map. Because $F_2(f[A \times \overline{A}]) \subset F_2(\overline{f}[\overline{A}](A))$ — note that the inclusion $\overline{f}[\overline{A}](A) \subset f[A \times \overline{A}]$ is valid — the statement $f(\underline{F} \times \underline{F}') \in F_2(\overline{f}[\overline{A}](A))$ follows. Since $f(\underline{F} \times \underline{F}')$ is coarser than the filter $e(\underline{F} \times \overline{f}(\underline{F}'))$ the desired result will be proved.

To show the last inclusion let $\overline{B} \in f(\underline{F} \times \underline{F}')$, so $\overline{B} \supset f[B \times B']$ for some $B \in \underline{F}$ and for some $B' \in \underline{F}'$. We have $B \times \overline{f}[B'] \in \underline{F} \times \overline{f}(\underline{F}')$. It remains to show that $f[B \times B'] \supset ev[B \times \overline{f}[B']] \cdot y \in ev[B \times \overline{f}[B']]$ implies $y = ev(x, f^*)$ for some $x \in B$ and for some $f^* \in \overline{f}[B']$. Now $f^* = \overline{f}(z)$ for some $z \in B'$ and the equality $y = ev(x, \overline{f}(z)) = \overline{f}(z)(x) = f(x, z)$ holds with $(x, z) \in B \times B'$.

Now we prove the still open statement $\underline{F} \times \underline{F}' \in (F_1 \times F_3)(A \times \overline{A})$. Therefore we have to check that

(i) $P_X(\underline{F} \times \underline{F}') \in F_1(P_X[A \times \overline{A}])$,

(ii) $P_Z(\underline{F} \times \underline{F}') \in F_3(P_Z[A \times \overline{A}])$

are valid. In proving the above two statements it remains to show the following four inclusion hold:

(1) $P_X[A \times \overline{A}] \subset A$,

(2) $P_X(\underline{F} \times \underline{F}') \supset \underline{F}$,

(3) $P_Z[A \times \overline{A}] \subset \overline{A}$,

(4) $P_Z(\underline{F} \times \underline{F}') \supset \underline{F}'$;

because by supposition we firstly get $\underline{F} \in F_1(A) \subset F_1(P_X[A \times \overline{A}])$ and $\underline{F}' \in \in F_3(\overline{A}) \subset F_3(P_Z[A \times \overline{A}])$ and secondly we have $P_X(\underline{F} \times \underline{F}') \in F_1(P_X[A \times \overline{A}])$ and $P_Z(\underline{F} \times \underline{F}') \in F_3(P_Z[A \times \overline{A}])$, which concludes the proof.

It is a triviality to verify (1) and (3).

In proving statement (2) let B be an element of $\underline{F}$. Since $\underline{F}' \neq \Phi$ we can choose $B' \in \underline{F}'$, so that $B \times B' \in \underline{F} \times \underline{F}'$ follows. But now with respect to (1) we get $P_X[B \times B'] \subset B$, hence $B \in P_X(\underline{F} \times \underline{F}')$. Statement (4) can be proved in an analogous way.

3. *FILFORM* and exponential laws

To give a summary we note that *FILFORM* is a cartesian closed topological category in which the important cartesian closed topological categories *FMER* and *FCONV* can be embedded up to isomorphism, moreover the corresponding full subcategories are bireflective in *FILFORM*. By purely categorical arguments the following three exponential laws hold in *FILFORM*, s. [9]:

(1) *First exponential law*: $X^{Y \times Z}$ is isomorphic to $(X^Y)^Z$,

(2) *Second exponential law*: $(\prod_{i \in I} X_i)^Y$ is isomorphic to $\prod_{i \in I} (X_i^Y)$,

(3) *Third exponential law*: $X^{\coprod_{i \in I} Y_i}$ is isomorphic to $\prod_{i \in I} (X^{Y_i})$.

At last we mention that the interesting *distributive law* also holds in *FILFORM*, i.e. $X \times \coprod_{i \in I} Y_i$ is isomorphic to $\coprod_{i \in I} (X \times Y_i)$.

References

[1] P. Antoine, Étude élémentaire des catégories d'ensembles structurés, *Bull. Soc. Math. Belge*, **18** (1966), 142–164 and 387–414.

[2] L. H. Bentley, H. Herrlich, and W. A. Robertson, Convenient categories for topologist, *Comm. Math. Univ. Carolinae*, **17** (1976), 207–227.

[3] E. Binz, *Continuous Convergence on $C(X)$*, Springer Lecture Notes in Math, **469** (1975).

[4] G. Choquet, Convergence, *Ann. Univ. Grenoble*, **23** (1947/48), 57–112.

[5] C. H. Cook, and H. R. Fischer, Uniform convergence structures, *Math. Ann.*, **173** (1967), 290–306.

[6] Á. Császár, *Foundations of general topology*, Pergamon Press, Oxford-London-New York-Paris, (1963).

[7] H. R. Fischer, Limesräume, *Math. Ann.*, **137** (1959), 269–303.

[8] A. Frolicher, and W. Bucher, *Calculus in vector spaces without norm*, Springer Lecture Notes in Math., **30** (1966).

[9] H. Herrlich, Cartesian closed topological categories, *Math. Coll. Univ. Cape Town*, **9** (1974), 1–16.

[10] H. Herrlich, *Some topological theorems which fail to be true*, Springer Lecture Notes in Math., **540** (1976).

[11] M. Katětov, On continuity structures and space of mappings, *Comm. Math. Univ. Carolinae*, **6** (1965), 257–278.

[12] D. C. Kent, Convergence functions and their related topologies, *Fund. Math.*, **54** (1964), 125–133.

[13] H. J. Kowalsky, Limesräume und Komplettierung, *Math. Nachr.*, **12** (1954), 301–340.

[14] D. Leseberg, Syntopoformic spaces (Syntopogenous limit spaces), *Quaest. Math.*, **7** (1984), 39–63.

[15] D. Leseberg, Syntopogene und topogene Konvergenzräume, *Annales Univ. Sci. Budapest.*, Sectio Math., **30** (1987), 47–63.

[16] R. S. Lee, The category of uniform convergence spaces is cartesian closed, *Bull. Austr. Math. Soc.*, **15** (1976), 461–465.

[17] A. Machado, Espaces d'Antoine et pseudo-topologies, *Cahiers de Top. et Géom. Diff.*, **11** (1973), 309–327.

[18] Th. Marny, Proximal limit spaces, *Quaest. Math.*, **2** (4) (1978), 443–452.

[19] L. D. Nel, *Cartesian closed topological categories*, Springer Lecture Notes in Math., **540** (1976).

[20] G. Preuss, *Theory of topological structures*, D. Reidel, Dordrecht Holland (1988).

[21] W. A. Robertson, *Convergence as a nearness concept*, Thesis, Carleton Univ. (1975).

[22] F. Schwarz, Powers and exponential objects in initially structured categories and applications to categories of limit spaces, *Quaest. Math.*, **6** (1983), 227–254.

ANNALES UNIV. SCI. BUDAPEST., 35 (1992), 15-21

ON CARDINALITY OF SETS OF METRICS GENERATING METRIC SPACES OF PRESCRIBED PROPERTIES

By

T. ŠALÁT, J. TÓTH and L. ZSILINSZKY

Department of Algebra and Number Theory of Comenius University Bratislava and Department of Math, College of Education Nitra, Czechoslovakia

(Received July 1, 1991)

Introduction

The paper [4] deals with some problems formulated in [1]. It is shown in [4] for example that the cardinality of the set of all topologies induced by metrics on a given set X with the cardinality $\aleph_0$ is equal to $2^{\aleph_0}$. In connection with this result it seems to be of some interest to investigate the questions concerning the cardinality of the set of all metrics on a given set X and the cardinality of sets of metrics leading to metric spaces of prescribed kinds (e.g. separable, complete, compact metric spaces). These considerations will be made without any restriction about the cardinality of the set X.

Throughout this paper let X be a given set. We shall consider the sets

$$\mathcal{M}(X)=\{d;\ d \text{ is a metric on } X\}$$
$$\mathcal{U}(X)=\{d\in\mathcal{M}(X);\ (X,d) \text{ is a complete metric space }\}$$
$$\mathcal{S}(X)=\{d\in\mathcal{M}(X);\ (X,d) \text{ is a separable metric space }\}$$
$$\mathcal{K}(X)=\{d\in\mathcal{M}(X);\ (X,d) \text{ is a compact metric space }\}$$
$$\mathcal{C}(X)=\{d\in\mathcal{M}(X);\ (X,d) \text{ is a connected metric space }\}.$$

First of all we recall some basic definitions and notations. The symbol $\mathcal{P}(X)$ stands for the power-set of Y, and $|Y|$ denotes the cardinality of the set Y.

For a constant $v>0$, the function $d_v: X\times X\to\mathbb{R}$ is said to be a trivial metric on X if

$$d_v(x,y)=\begin{cases} v, & \text{for } x\neq y\\ 0, & \text{for } x=y\end{cases}$$

($\mathbb{R}$ denotes the set of all real numbers).

Let d, $d' \in \mathcal{M}$ and x, $x_n \in X$ $(n=1,2,\dots)$. The metrics d, d' are said to be equivalent if

$$\lim_{n\to\infty} d(x_n, x) = 0 \quad \text{if and only if} \quad \lim_{n\to\infty} d'(x_n, x) = 0.$$

In our further considerations the following lemmas will take an important place.

LEMMA 1. Suppose that a is an infinite cardinal, and $\aleph_0$ denotes the power of the set of all natural numbers. Then

i) $a+\aleph_0=a$

ii) $\aleph_0 \cdot a = a$

iii) $a+a=a$

iv) $a \cdot a = a$

holds.

PROOF. See [2] (pp. 25–26.) ∎

Let $Y \subset \mathbb{R}$ and $f: X \to Y$ be a bijection. Then the function $d: X \times X \to \mathbb{R}$,

(1) $$d(x,y) = |f(x) - f(y)|, \qquad \text{for } x,\ y \in X$$

is obviously a metric on X, and for the metric space (X,d) the following assertion holds (Y is considered as a metric space with the euclidean metric):

LEMMA 2 The mapping $f: X \to Y$ is a homeomorphism.

PROOF. The proof follows at once from the definition of f. ∎

Cardinality of sets $\mathcal{M}(X)$, $\mathcal{U}(X)$, $\mathcal{S}(X)$, $\mathcal{K}(X)$, $\mathcal{C}(X)$

If $X=\emptyset$ or X has only one element, then obviously $|\mathcal{M}(X)| = |\mathcal{U}(X)| = |\mathcal{S}(X)| = |\mathcal{K}(X)| = |\mathcal{C}(X)| = 1$, therefore it is sufficient to consider the case if $|X| \geq 2$.

Denote by c the power of the continuum.

THEOREM 1. *We have*

$$|\mathcal{M}(X)| = \begin{cases} c, & \text{if } 2 \leq |X| < \aleph_0 \\ 2^{|X|}, & \text{if } \aleph_0 \leq |X|. \end{cases}$$

PROOF. If $|X| = n$ is a finite cardinal (≥ 2), then $c \leq |\mathcal{M}(X)|$, since c is the cardinality of the set of all trivial metrics d_v $(v>0)$ on X.

On the other hand

$$\mathcal{M}(X)\subset {}^{X\times X}\mathbb{R}, \text{ so } |\mathcal{M}(X)|\leq c^{n\cdot n}=c, \text{ thus } |\mathcal{M}(X)|=c.$$

Suppose now, that X is an infinite set and put $|X|=a$. For every $A\subseteq X$ define the function $\varrho_A : X\times X\to\mathbb{R}$ as follows. If $x\neq y$, then

$$\varrho_A(x,y)=\varrho_A(y,x)=\begin{cases}1, & \text{for } x,\ y\in A\\ 2, & \text{for } x,\ y\in X\setminus A\\ \Theta, & \text{for } x\in A,\ y\in X\setminus A, \text{ where } 1<\Theta<2,\end{cases}$$

and naturally $\varrho_A(x,x)=0$ for all $x\in X$.

It is easy to verify that $\varrho_A\in\mathcal{M}(X)$, so

$$|\mathcal{M}(X)|\geq|\mathcal{P}(X)|=2^{|X|} \tag{2}$$

as $\varrho_A\neq\varrho_{A'}$ for $A\neq A'$, A, $A'\subseteq X$.

Conversely we have $\mathcal{M}(X)\subset {}^{X\times X}\mathbb{R}$ and according to Lemma 1 we get

$$|\mathcal{M}(X)|\leq c^{a\cdot a}=c^a=\left(2^{\aleph_0}\right)^a=2^{\aleph_0\cdot a}=2^a=2^{|X|} \tag{2'}$$

From (2), (2') we get by the Cantor–Bernstein theorem $|\mathcal{M}(X)|=2^{|X|}$. ∎

THEOREM 2. *We have*

$$|\mathcal{U}(X)|=\begin{cases}c, & \textit{for } 2\leq|X|<\aleph_0\\ 2^{|X|}, & \textit{for } \aleph_0\leq|X|.\end{cases}$$

PROOF. Evidently all trivial metrics d_v $(v>0)$ belong to $\mathcal{U}(X)$, if X is a finite set with at least two elements. Hence from Theorem 1 we obtain

$$c\leq|\mathcal{U}(X)|\leq|\mathcal{M}(X)|=c.$$

Let $\aleph_0\leq|X|=a$. It is not hard to see, that all metrics ϱ_A $(A\subseteq X)$ from the proof of Theorem 1 are elements of $\mathcal{U}(X)$, so

$$2^a=|\mathcal{P}(X)|\leq|\mathcal{U}(X)|\leq|\mathcal{M}(X)|=2^a,$$

the theorem follows. ∎

THEOREM 3. *We have*

$$|\mathcal{S}(X)|=\begin{cases}c, & \textit{for } 2\leq|X|\leq\aleph_0\\ 2^{|X|}, & \textit{for } \aleph_0<|X|\leq c\\ 0, & \textit{for } |X|>c.\end{cases}$$

PROOF. It is well-known that there is no separable metric space X with $|X|>c$.

Suppose, that $2 \leq |X| < \aleph_0$. Then X is a countable and dense subset of the metric space (X,d) for every $d \in \mathcal{M}(X)$, thus by Theorem 1 $|\mathcal{S}(X)| = |\mathcal{M}(X)| = c$ holds.

Let us finally examine the case if $\aleph_0 \leq |X| \leq c$. Put $|X| = a$, then there exists a set $D \subset \left[0, \frac{1}{6}\right)$ of cardinality a. Denote by $D_0 = D \cup \left(\mathbb{Q} \cap \left[0, \frac{1}{6}\right)\right)$, where $\mathbb{Q}$ is the set of all rational numbers.

Using Lemma 1 we obtain

$$a \leq |D_0| \leq |D| + \left|\mathbb{Q} \cap \left[0, \frac{1}{6}\right)\right| = a + \aleph_0 = a, \text{ so } |D_0| = a.$$

Put

$$D_k = \left\{ \frac{1}{2 \cdot 3^k} + \frac{v}{3^k};\ v \in D_0 \right\}$$
$$E_k = \left\{ 1 - \frac{1}{2 \cdot 3^k} + \frac{v}{3^k};\ v \in D_0 \right\}$$
$$F_k = \left\{ \frac{1}{3^{k+1}} + \frac{v}{3^k};\ v \in D_0 \right\} \qquad (k = 1, 2, \ldots)$$

and

$$H = \left[\bigcup_{k=1}^{\infty} (D_k \cup E_k \cup F_k) \right] \cup \{0, 1\}.$$

Using Lemma 1 we receive

$$a \leq |H| \leq \sum_{k=1}^{\infty} (|D_k| + |E_k| + |F_k|) + |\{0,1\}| = \aleph_0 \cdot (a + a + a) + 2 \leq$$
$$\leq \aleph_0 \cdot a + \aleph_0 = a + \aleph_0 = a, \quad \text{since} \quad |D_k| = |E_k| = |F_k| = |D_0| = a$$

$(k = 1, 2, \ldots)$.

The previous inequalities imply that $|H| = a$. Hence there exists a bijection $f : X \to H$.

The set $\mathbb{Q} \cap H$ is obviously a dense and countable subset of H, thus H is a separable metric space, and it follows from Lemma 2 that (X, d) with the metric (1) is a separable metric space as well.

Define the sequences of elements of X:

$$x_{2k-1}^{(v)} = f^{-1}\left(\frac{1}{2 \cdot 3^k} + \frac{v}{3^k}\right), \qquad x_{2k}^{(v)} = f^{-1}\left(1 - \frac{1}{2 \cdot 3^k} + \frac{v}{3^k}\right),$$
$$y_k^{(v)} = f^{-1}\left(\frac{1}{3^{k+1}} + \frac{v}{3^k}\right) \qquad \text{for} \quad v \in D_0 \text{ and } k = 1, 2, \ldots,$$

furthermore put $z=f^{-1}(0)$ and $u=f^{-1}(1)$.

These elements (included z and u) are mutually different for every $k=1,2,\dots$ and $v\in D_0$ as one can easily check on the basis of bijectivity of f.

Let us define new bijections $f_A: X\to H$ for sets $A\subseteq D_0$ as follows (the symbol $\mathbb{N}$ stands for the set of all positive integers):

$$f_A(x)=f(x), \text{ for } x\notin \bigcup_{\substack{v\in A\\ k\in\mathbb{N}}} \{x_k^{(v)}, y_k^{(v)}\}$$

$$f_A(x_k^{(v)})=f(y_k^{(v)}),\ f_A(y_k^{(v)})=f(x_k^{(v)}), \text{ for } v\in A,\ k=1,2,\dots .$$

By the help of the mapping f_A analogously to (1) it can be introduced a metric $d_A\in\mathscr{S}(X)$ for every $A\subseteq D_0$.

We shall prove, that for different subsets A, $B\subseteq D_0$, the metrics d_A and d_B are not equivalent and therefore they are different.

Let $A\neq B$, A, $B\subseteq D_0$. Then without loss of generality we may assume that there exists a $v_0\in A\setminus B$. Then

$$d_A(x_k^{(v_0)},z)=\left|f_A(x_k^{(v_0)})-f_A(z)\right|=\left|f(y_k^{(v_0)})-f(z)\right|=$$

$$(3)\qquad =\frac{1}{3^k}\left(v_0+\frac{1}{3}\right)\to 0, \quad \text{as} \quad k\to\infty$$

Furthermore

$$d_B(x_{2k-1}^{(v_0)},z)=|f_B(x_{2k-1}^{(v_0)})-f_B(z)|=|f(x_{2k-1}^{(v_0)})-f(z)|=$$

$$=\frac{1}{3^k}\left(v_0+\frac{1}{2}\right)\to 0, \quad \text{as} \quad k\to\infty,$$

and

$$d_B(x_{2k}^{(v_0)},u)=|f_B(x_{2k}^{(v_0)})-f_B(u)|=|f(x_{2k}^{(v_0)})-f(u)|=$$

$$=\frac{1}{3^k}\left|v_0-\frac{1}{2}\right|\to 0, \quad \text{as} \quad k\to\infty.$$

It means, that the sequence $\{x_k^{(v_0)}\}_{k=1}^{\infty}$ does not converge with respect to the metric d_B. From (3) we see that the metrics d_A and d_B are not equivalent. Then it follows from Theorem 1 that

$$2^a=|\mathscr{P}(D_0)|\le|\mathscr{S}(X)|\le|\mathscr{M}(X)|=2^a, \quad \text{thus} \quad |\mathscr{S}(X)|=2^a. \qquad \blacksquare$$

REMARK. We have proved more than the statement of the Theorem 3, is namely that the cardinality of non-equivalent metrics from $\mathscr{S}(X)$ is $2^{|X|}$ if $\aleph_0\le|X|\le c$.

THEOREM 4. *We have*

$$|\mathcal{K}(X)| = \begin{cases} c, & \text{for } 2 \le |X| \le \aleph_0 \\ 0, & \text{for } \aleph_0 < |X| < c \\ 2^c, & \text{for } |X| = c \\ 0, & \text{for } |X| > c. \end{cases}$$

PROOF. Since every compact metric space is also a separable metric space, therefore $|\mathcal{K}(X)|=0$ if $|X|>c$. Furthermore a compact metric space (X,d) is a complete and separable metric space as well, so if $|X|>\aleph_0$ then according to the result of [3] (see p. 351, IV.2.) we have $|X|=c$ which is a contradiction to our assumption.

If X is a finite set, then evidently $\mathcal{K}(X)=\mathcal{M}(X)$, so it follows from Theorem 1 that $|\mathcal{K}(X)|=|\mathcal{M}(X)|=c$.

Suppose now that $|X|=\aleph_0$ and denote by

$$G=\left\{0,1,\frac{1}{2},\ldots,\frac{1}{n},\ldots\right\}\subset \mathbb{R}.$$

Then there exists a bijection $f: X \to G$. For $p>0$ define a metric $d_p : X\times X \to \mathbb{R}$ as follows: $d_p(x,y)=p\cdot d(x,y)$, where d is the metric from (1) $(x,\ y\in X)$.

From the compactness of G and Lemma 2 one can derive, that the metric space (X,d_p) is compact for every $p>0$, thus using Theorem 1 we get

$$c=|(0,\infty)|\le|\mathcal{K}(X)|\le|\mathcal{M}(X)|=2^{\aleph_0}=c, \quad \text{so } |\mathcal{K}(X)|=c.$$

Finally assume, that $|X|=c$ and let $f: X\to[0,1]$ be a one-to-one correspondence. Then a metric d can be defined on X analogously to (1), and according to Lemma 2 (X,d) is a compact metric space (since the interval $[0,1]$ is compact, too).

In the rest of the proof we can use the idea of the proof of Theorem 3 putting $D_0=\left[0,\frac{1}{6}\right)$ and $H=[0,1]$. ∎

THEOREM 5. *We have*

$$|\mathcal{C}(X)| = \begin{cases} 0, & \text{for } 2 \le |X| < c \\ 2^c, & \text{for } |X| = c. \end{cases}$$

PROOF. Suppose that there is a $d\in\mathcal{C}(X)$ if $2\le|X|<c$. Since (X,d) is connected, the space $(X\times X,\varrho)$ $(\varrho=\sqrt{d^2+d^2})$ is connected, too (cf. [5] p.144, Theorem E). As the mapping $d: X\times X\to\mathbb{R}$ is continuous (cf. [5], p.144, Theorem C) the set $d(X\times X)$ is an interval $I\subset[0,\infty)$ and so $|I|=c$.

Therefore $|X \times X| = c$. But by Lemma 1 and our assumption we have $|X \times X| = |X| < c$ — a contradiction. Hence $|\mathscr{C}(X)| = 0$ if $2 \leq |X| < c$.

If $|X| = c$, then there exists a bijection $f : X \to [0,1]$. Since the interval $[0,1]$ is a connected set, then according to Lemma 2 the space (X, d) is connected as well, where d is the metric constructed in (1).

The rest of the proof is analogous to the proof of Theorem 3. ∎

REMARK It is an open problem to determine the cardinality of $\mathscr{C}(X)$ if $|X| > c$.

References

[1] E. ČECH, Topologické prostory, *Čas. pěst. mat. fyziky*, **66** (1937), 225–264.

[2] E. HEWITT, V. STROMBERG, *Real and Abstract Analysis*, Springer–Verlag, Berlin–Heidelberg–New York, 1969.

[3] K. KURATOWSKI, *Topologie I*. PWN, Warszawa, 1958.

[4] B. POSPÍŠIL, Sur le nombre des topologies d'un ensemble donné, *Čas. pěst. mat. fyziky*, **67** (1938), 100–102.

[5] G. F. SIMMONS, *Introduction to Topology and Modern Analysis*, McGraw-Hill Comp., New York–San Francisco–Toronto–London, 1963.

ANNALES UNIV. SCI. BUDAPEST., 35 (1992), 23–35

ON HERMITE FUNCTIONS

By

I. JOÓ

Department of Analysis, L. Eötvös University, Budapest

(Received July 5, 1992)

Denote $H_n(x)$ the n-th Hermite polynomial and $0<x_1<x_2<\ldots<x_\nu$ its positive zeros. Let $\varrho(m)=\varrho^*(m)/m$, where $0<\varrho^*(m)$ increases arbitrarily slowly, further let $\nu=O(m/\beta(m))$, where $\beta(m)$ is a sufficiently large constant. Then we have the

THEOREM 1. *We have the following estimates*

$$x_\nu=\frac{\nu\pi-\frac{\pi}{2}}{\sqrt{4m+1}}+O\left(\frac{\varrho(m)}{\sqrt{m}}\right)+O\left(\frac{\nu^3}{m^{5/2}}\right),\qquad \text{if } n=2m; \tag{1}$$

$$x_\nu=\frac{\nu\pi}{\sqrt{4m-1}}+O\left(\frac{\varrho(m)}{\sqrt{m}}\right)+O\left(\frac{\nu^3}{m^{5/2}}\right),\qquad \text{if } n=2m-1; \tag{2}$$

further if $0<\beta^*(m)$ *increases arbitrarily slowly, then*

$$x_{\nu+1}-x_\nu=\frac{\pi}{\sqrt{2n+1}}+O\left(\frac{\varrho(n)}{n}\right)+O(1)\frac{1}{\sqrt{n}}\sum_{k=2}^{\infty}\frac{k}{c_3^k}\cdot\frac{\nu^k}{n^k}\quad . \tag{3}$$

where $0<c_3<1$ *is an absolute constant,* $\nu=O(m/\beta^*(m))$.

The proof follows the method given on Szegő's book 8.9 on pp. 237–240, where the method is used for the zeros belonging to a fixed interval. We needed several modifications of this idea. In the second part of this paper we give a weaker result but for all zeros in the Theorem 2. We give applications of these results in subsequent papers. Before the proof we formulate also the Theorem 2.

THEOREM 2. *Now denote* $x_1>x_2>\ldots>x_n$ *the zeros of* $H_n(x)$. *Then we have*

$$x_{|\nu|}=\sqrt{2n+1}\cdot\cos\theta_0(\nu)+\frac{\varrho(v)\cdot O(1)}{n^{1/6}\cdot|\nu|^{1/3}},\qquad |\nu|=1,\ldots,\left[\frac{n}{2}\right], \tag{4}$$

where $\varrho(\nu)=\frac{a}{|\nu|}$, *a is an absolute constant and* $\theta_0(\nu)$ *is given by*

$$\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta_0(\nu)-2\theta_0(\nu))=\nu\pi-\frac{3\pi}{4}. \tag{5}$$

Here we can not write $o(1)$ in place of $O(1)$. In the special case of $|\nu|\le dn$, where $d>0$ is a small absolute constant, we get from (4) and (5)

$$x_{|\nu|}=\sqrt{2n+1}-\frac{O(1)|\nu|^{2/3}}{n^{1/6}}. \tag{6}$$

From (4) and (5) we will deduce the

$$x_{|\nu+1|}-x_{|\nu|}=\frac{O(1)}{n^{1/6}\cdot|\nu|^{1/3}},\qquad |\nu|=2,\dots,\left[\frac{n}{2}\right], \tag{7}$$

where we can not write $o(1)$ in place of $O(1)$.

PROOF OF THE THEOREM 1. We need the

LEMMA 1. We have

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H_n(\sqrt{2n+1}\cos\theta)= \tag{8}$$

$$=k_n(\theta)\cdot(-2n)\sqrt{2n+1}\cdot\sin\theta\cdot 2^{\frac{n}{2}-\frac{1}{4}}\cdot\sqrt{(n-1)!}\cdot(\pi(n-1))^{-\frac{1}{4}}\cdot$$
$$\cdot\left[e^{-\cos^2\theta}\left\{\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n}\right)\right\}+O(1)\cos\theta\right],$$

where $k_n(\theta)=e^{\frac{2n+1}{2}\cos^2\theta}\cdot(\sin\theta)^{-\frac{1}{2}}$, $0<\varepsilon\le\theta\le\pi-\varepsilon$, $\varepsilon\le\frac{\pi}{2}$ fixed, and the remainder terms are uniform in θ.

PROOF. We need ([1], 8.22.12)

$$e^{-\frac{x^2}{2}}H_n(x)=2^{\frac{n}{2}+\frac{1}{4}}\cdot\sqrt{n!}\cdot(\pi n)^{-\frac{1}{4}}\cdot(\sin\theta)^{-\frac{1}{2}}\cdot \tag{9}$$
$$\cdot\left\{\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n}\right)\right\},$$

where $x=\sqrt{2n+1}\cos\theta$, $0<\varepsilon\le\theta\le\pi-\varepsilon$, $\varepsilon\le\frac{\pi}{2}$ fixed. The remainder term is uniform in θ.

Since $H_n'(x)=2nH_{n-1}(x)$ ([1], 5.5.10), hence

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H_n(\sqrt{2n+1}\cos\theta)=-2n(2n+1)^{\frac{1}{2}}\cdot\sin\theta\cdot H_{n-1}(\sqrt{2n+1}\cos\theta). \tag{10}$$

By Lagrange's mean value theorem

$$H_{n-1}(\sqrt{2n+1}\cos\theta)-H_{n-1}(\sqrt{2n-1}\cos\theta)=O(1)\cdot\frac{\cos\theta}{\sqrt{n}}\cdot H_{n-1}'(\xi), \tag{11}$$

where

(12) $\sqrt{2n-1}\cos\theta \le \xi \le \sqrt{2n+1}\cos\theta, \quad H'_{n-1}(\xi) = O(1)nH_{n-2}(\xi);$

hence in order to apply (9) for $H_{n-2}(\xi)$, the following must be fulfilled: There exists a constant $\varepsilon^* > 0$ depending only on ε (ε^* independent on n) such that

$$\{\sqrt{2n+1}\cos\theta:\ 0<\varepsilon\le\theta\le\pi-\varepsilon\} \subseteq \{\sqrt{2n-3}\cos\theta^*:\ 0<\varepsilon^*\le\theta^*\le\pi-\varepsilon^*\}$$

for $n \ge n_0$, where n_0 is an absolute constant depending only on ε. This means the validity of the inequality

$$\sqrt{2n+1}\cos\varepsilon \le \sqrt{2n-3}\cos\varepsilon^*,$$

hence

$$\cos\varepsilon \le \sqrt{1-\frac{4}{2n+1}}\cos\varepsilon^*.$$

Substituting $\varepsilon^* = \varepsilon/2$ we get

$$\frac{2\cos^2\varepsilon}{1+\cos\varepsilon} \le 1 - \frac{4}{2n+1}.$$

Because for fixed $\varepsilon > 0$ the left hand side is a constant smaller than 1 and the right hand side is monotone increasing to 1 as $n \to +\infty$, hence for $n \ge n_0 = n_0(\varepsilon)$ the inequality is fulfilled. Hence we can apply (9) to estimate $H_{n-2}(\xi)$. Because $\xi = \sqrt{2n-3}\cos\theta^*$, hence from (12) $(1+O(1/n))\cos\theta = \cos\theta^*$, consequently $(1+O(1/n))\sin\theta = \sin\theta^*$, i.e.

$$H_{n-2}(\xi) = H_{n-2}(\sqrt{2n-3}\cos\theta^*) = e^{\frac{2n-3}{2}\cos^2\theta^*}\cdot 2^{\frac{n}{2}-\frac{3}{4}}\cdot\sqrt{(n-2)!}\cdot$$
$$\cdot(\pi(n-2))^{-\frac{1}{4}}(\sin\theta^*)^{-\frac{1}{2}}\left\{\sin\left[\left(\frac{n}{2}-\frac{3}{4}\right)(\sin 2\theta^* - 2\theta^*)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n}\right)\right\} =$$
$$= O(1)e^{n\cos^2\theta}\cdot 2^{\frac{n}{2}}\cdot\sqrt{(n-2)!}\cdot n^{-\frac{1}{4}}\cdot(\sin\theta)^{-\frac{1}{2}}.$$

That is

(13) $$H_{n-1}(\sqrt{2n+1}\cos\theta) - H_{n-1}(\sqrt{2n-1}\cos\theta) = O(1)\cdot\frac{\cos\theta}{\sqrt{n}}\cdot H'_{n-1}(\xi) =$$
$$= O(1)\cdot\sqrt{n}\cos\theta H_{n-2}(\xi) = O(1)e^{n\cos^2\theta}\cdot 2^{\frac{n}{2}}\cdot\sqrt{(n-1)!}\cdot n^{-\frac{1}{4}}\cos\theta(\sin\theta)^{-\frac{1}{2}}.$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H_n(\sqrt{2n+1}\cos\theta) = -2n(2n+1)^{\frac{1}{2}}\cdot\sin\theta\cdot$$
$$\cdot\left[H_{n-1}(\sqrt{2n+1}\cos\theta) - H_{n-1}(\sqrt{2n-1}\cos\theta) + H_{n-1}(\sqrt{2n-1}\cos\theta)\right],$$

hence we get by (13)

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H_n(\sqrt{2n+1}\cos\theta) =$$

$$= k_n(\theta)\cdot(-2n(2n+1)^{\frac{1}{2}}\sin\theta)\cdot\Big[e^{-\cos^2\theta}\cdot 2^{\frac{n}{2}-\frac{1}{4}}\cdot\sqrt{(n-1)!}\cdot(\pi(n-1))^{-\frac{1}{4}}\cdot$$

$$\cdot\left\{\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n}\right)\right\}+O(1)2^{\frac{n}{2}}\sqrt{(n-1)!}n^{-\frac{1}{4}}\cos\theta\Big]=$$

$$= k_n(\theta)\cdot(-2n)\sqrt{2n+1}\sin\theta\cdot 2^{\frac{n}{2}-\frac{1}{4}}\cdot\sqrt{(n-1)!}\cdot(\pi(n-1))^{-\frac{1}{4}}\cdot$$

$$\cdot\left[e^{-\cos^2\theta}\left\{\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n}\right)\right\}+O(1)\cos\theta\right].$$

Lemma 1 is proved.

In the following let $\theta=\frac{\pi}{2}-\theta^*$, where $0<\theta^*<\frac{1}{\beta(n)}\le\frac{1}{10}$ and $0<\beta(n)$ be a sufficiently large constant. We may suppose also that it is slightly increasing and in this case we know the estimate only in the interval $(0,n/\beta(n))$, but the remainder term will be better. Our proof in the following works for both cases.

We consider two cases: $n=2m$ and $n=2m+1$ separately.

(a) *Case* $n=2m$.

In this case the main term in (9) can be written as follows:

$$\sin\left[\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)-m\pi+\frac{\pi}{2}\right]=$$

$$=(-1)^m\sin\left[\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)+\frac{\pi}{2}\right].$$

Let

$$\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)=\nu\pi-\frac{\pi}{2}\pm\varrho(m), \tag{14}$$

where $\nu>0$ is integer, $0<\varrho(m)\le\frac{\pi}{4}$. Then for the main term of (9) we have

$$\sin\left[\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)-m\pi+\frac{\pi}{2}\right]=(-1)^m\cdot(-1)^\nu\cdot(\pm\sin\varrho(m)).$$

This is a main term in (9) indeed, if for $\varrho(m):=\frac{\varrho^*(m)}{m}$, $0<\varrho^*(m)\nearrow+\infty$ and $\varrho^*(m)\le\frac{m}{20}$ be fulfilled. Remark, that it follows from (14) that $\theta^*\asymp\frac{\nu}{m}$, which means according to $0<\theta^*<1/\beta(m)$, $\nu=O\left(\frac{m}{\beta(m)}\right)$. Now denote the θ^* corresponding to $+\varrho(m)$ in (14) by θ^*_+ and that of corresponding to $-\varrho(m)$ by θ^*_-. It follows from the above considerations that

$$\operatorname{sign}H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*_+\right)\right)=(-1)^m\cdot(-1)^\nu,$$

$$\operatorname{sign}H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*_-\right)\right)=(-1)^m\cdot(-1)^\nu\cdot(-1),$$

i.e. H_n changes sign in the interval

$$\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta_-^*\right),\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta_+^*\right)\right)$$

which means that H_n has at least one zeros in this interval. If H_n does not change sign in this interval, i.e. $\operatorname{sign} H_n$ is constant, then H_n has exactly one zero in this interval. Now examine, how we can ensure this. Let $\theta_-^* \le \theta^* \le \theta_+^*$. According to Lemma 1, for the main term of H_n we have

$$\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta - 2\theta)+\frac{3\pi}{4}\right]=$$
$$=\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta)+\frac{3\pi}{4}-\frac{\sin 2\theta-2\theta}{2}\right]=$$
$$=\sin\left[\left(m+\frac{1}{4}\right)(\sin 2\theta^* + 2\theta^*)-(m-1)\pi-\frac{\sin 2\theta^*+2\theta^*}{2}\right].$$

Because $f(\theta^*)=\sin 2\theta^*+2\theta^*$ strictly increasing in $[0,1/4]$, hence

$$|f(\theta^*)-f(\theta_+^*)|\le|f(\theta_-^*)-f(\theta_+^*)|\le\frac{2\varrho(m)}{m},$$

so

$$\sin 2\theta^*+2\theta^*=\sin 2\theta_+^*+2\theta_+^*+c_1\frac{\varrho(m)}{m},$$

where $|c_1|\le 2$. Applying (14) we obtain for the main term of H_n'

$$\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta - 2\theta)+\frac{3\pi}{4}\right]=$$
$$=\sin\left[\nu\pi-\frac{\pi}{2}+\varrho(m)+2c_1\varrho(m)-\frac{1}{2}\frac{\nu\pi-\frac{\pi}{2}+\varrho(m)}{m+\frac{1}{4}}+\frac{c_1}{2}\cdot\frac{\varrho(m)}{m}\right]\cdot(-1)^{m-1}.$$

According to the assumption made on $\varrho(m)$ and ν we have for $m\ge m_0$ (m_0 is an absolute constant):

$$0<c_2<\left|\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]\right|\le 1.$$

Because in (8) we have $O(1)\cos\theta = O(1)\sin\theta^* = O\left(\frac{1}{\beta(n)}\right)$, hence the main term (8) is indeed the main term thus the notation in (8) is justified. On the other hand we obtain from the above considerations that $\operatorname{sign} H_n' = =(-1)^m\cdot(-1)^{\nu-1}$ in the interval $\theta_-^*\le\theta^*\le\theta_+^*$. Hence, H_n has exactly one zeros in the interval

$$(\sqrt{2n+1}\sin\theta_-^*,\sqrt{2n+1}\sin\theta_+^*).$$

Because θ_-^* and θ_+^* depend on ν, next we use the notation $\theta_-^*(\nu)$, $\theta_+^*(\nu)$ for them. Now we prove that if $0<\theta^*\le\theta_-^*(1)$ or $\theta_+^*(\nu)\le\theta^*\le\theta_-^*(\nu+1)$ then $H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*\right)\right)\neq 0$, and hence Theorem 1 follows immediately.

First consider the case $0<\theta^*\le\theta_-^*(1)$. In this case

$$\frac{\pi}{2}<\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)+\frac{\pi}{2}\le\left(m+\frac{1}{4}\right)(\sin 2\theta_-^*(1)+2\theta_-^*(1))+\frac{\pi}{2}=$$
$$=\frac{\pi}{2}-\varrho(m)+\frac{\pi}{2}=\pi-\varrho(m).$$

Hence the order of the main term in (9) is $\ge c\varrho(m)$. According to the assumptions made on $\varrho(m)$, the main term is indeed main. The sign of the main term is $(-1)^m$ which is equal to $\operatorname{sign} H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta_-^*(1)\right)\right)$. This means that $H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*\right)\right)\neq 0$ if $0<\theta^*\le\theta_-^*(1)$. Now investigate the case $\theta_+^*(\nu)\le\theta^*\le\theta_-^*(\nu+1)$. Define θ_0^* as follows:

$$\left(m+\frac{1}{4}\right)(\sin 2\theta_0^*+2\theta_0^*)=\nu\pi.$$

Then $\theta_+^*(\nu)<\theta_0^*<\theta_-^*(\nu+1)$.

First consider the case $\theta_+^*(\nu)<\theta^*\le\theta_0^*$. Then

$$\nu\pi+\varrho(m)<\left(m+\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)+\frac{\pi}{2}\le\nu\pi+\frac{\pi}{2}.$$

Hence the main term of (9) has the order of magnitude $\ge c\varrho(m)$. According to the assumptions made on $\varrho(m)$ it follows that the main term in (9) is the main. The sign of the main term is $(-1)^m\cdot(-1)^\nu$ which is equal to $\operatorname{sign} H_n\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta_+^*(\nu)\right)\right)$, i.e. H_n has no zeros in the interval $\theta_+^*(\nu)<\theta^*\le\theta_0^*$. We obtain similarly, that H_n has no zeros in the interval $\theta_0^*\le\theta^*<\theta_-^*(\nu+1)$. Summarizing our considerations we get

$$\sqrt{2n+1}\sin\theta_-^*(\nu)<x_\nu<\sqrt{2n+1}\sin\theta_+^*(\nu).$$

Taking into account

$$|\sin\theta_+^*(\nu)-\sin\theta_-^*(\nu)|\le|f(\theta_+^*(\nu))-f(\theta_-^*(\nu))|\le\frac{2\varrho(m)}{m},$$

the length of this interval is $O\left(\frac{\varrho(m)}{\sqrt{m}}\right)$, hence

$$x_\nu=\sqrt{2n+1}\sin\theta_-^*(\nu)+O\left(\frac{\varrho(m)}{\sqrt{m}}\right).$$

Because $\sin 2y+2y=A$, $0<A<1/4$ implies $\sin y=A/4(1+O(y^2))$, (namely $\frac{\sin 2y+2y}{4\sin y}-1=\frac{\sin 2y+2y-4\sin y}{4\sin y}=\frac{O(y^3)}{\sin y}=O(y^2)$), hence

$$\sin\theta^*_-(\nu)=\frac{\nu\pi-\frac{\pi}{2}-\varrho(m)}{4m+1}\left(1+O\left(\frac{\nu^2}{m^2}\right)\right),$$

consequently

$$x_\nu=\frac{\nu\pi-\frac{\pi}{2}}{\sqrt{4m+1}}+O\left(\frac{\varrho(m)}{\sqrt{m}}\right)+O\left(\frac{\nu^3}{m^{5/2}}\right),$$

and this proves the Theorem 1 for $n=2m$.

(b) Case $n=2m-1$.

In this case the main term in (9) is

$$(-1)^{m-1}\sin\left[\left(m-\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)\right].$$

Define $\theta_\pm{}^*$ by

$$\left(m-\frac{1}{4}\right)(\sin 2\theta_\pm{}^*+2\theta_\pm{}^*)=\nu\pi\pm\varrho(m), \tag{15}$$

where $\nu>0$ is an integer and $0<\varrho(m)\le\frac{\pi}{4}$. If we impose on $\varrho(m)$ the same conditions as in (a), then we obtain similarly that H_n changes sign on the interval

$$\left(\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*_-\right),\sqrt{2n+1}\cos\left(\frac{\pi}{2}-\theta^*_+\right)\right)$$

where θ^*_+ resp. θ^*_- denotes the solution of (15) corresponding to $+\varrho(m)$ resp. $-\varrho(m)$.

Now investigate the sign of H_n' on this interval. Let $\theta^*_-\le\theta^*\le\theta^*_+$. According to Lemma 1 we obtain for the main term of H_n':

$$\sin\left[\left(m-\frac{1}{4}\right)(\sin 2\theta^*+2\theta^*)-(m-1)\pi+\frac{\pi}{2}-\frac{\sin 2\theta^*+2\theta^*}{2}\right]=$$
$$=(-1)^{m-1}\sin\left[\nu\pi+\varrho(m)+2c_1^*\varrho(m)+\frac{\pi}{2}+\frac{c_1^*}{2}\frac{\varrho(m)}{m}-\frac{1}{2}\frac{\nu\pi+\varrho(m)}{m-\frac{1}{4}}\right],$$

where $|c_1^*|\le 4$. Here we have used the same idea as in (a). Consequently, according to the assumptions made on $\varrho(m)$ and ν, the main term in the expression of H_n' is main part indeed, hence the sign of H_n' is $(-1)^m\cdot(-1)^\nu$ on the interval $\theta^*_-\le\theta^*\le\theta^*_+$. This means that H_n has exactly one zeros in the interval. Because θ^*_- and θ^*_+ depend on m and on ν, hence we write in the following $\theta^*_-(2m-1,\nu)$ and $\theta^*_+(2m-1,\nu)$ further for θ^*_- and θ^*_+ used in (a) we will write $\theta^*_-(2m,\nu)$ and $\theta^*_+(2m,\nu)$. We have to prove only that if $0<\theta^*\leq\theta^*_-(2m-1,1)$, or $\theta^*_+(2m-1,\nu)\leq\theta^*\leq\theta^*_-(2m-1,\nu+1)$ then $H_n\left(\sqrt{2n+2}\cos\left(\frac{\pi}{2}-\theta^*\right)\right)\neq 0$. We have seen in the case of (a), that the sign of $H'_{2m}=4mH_{2m-1}$ is $(-1)^m\cdot(-1)^{\nu-1}$ in the interval $\theta^*_-(2m,\nu)\leq$ $\leq\theta^*\leq\theta^*_+(2m,\nu)$ further $(-1)^m\cdot(-1)^\nu$ if $\theta^*_-(2m,\nu+1)\leq\theta^*\leq\theta^*_+(2m,\nu+1)$. This means that H_{2m-1} has zeros in the interval

$$\left(\sqrt{4m+1}\sin\theta^*_+(2m,\nu),\sqrt{4m+1}\sin\theta^*_-(2m,\nu+1)\right).$$

It is known that the zeros of H_n and H_{n-1} separate each other, hence H_{2m-1} has exactly one zeros in the interval

$$\left(\sqrt{4m+1}\sin\theta^*_+(2m,\nu),\sqrt{4m+1}\sin\theta^*_-(2m,\nu+1)\right).$$

Let

$$A_\nu:=\left(\sqrt{4m+1}\sin\theta^*_+(2m,\nu),\sqrt{4m+1}\sin\theta^*_-(2m,\nu+1)\right),$$
$$B_\nu:=\left(\sqrt{4m-1}\sin\theta^*_-(2m-1,\nu),\sqrt{4m-1}\sin\theta^*_+(2m-1,\nu)\right).$$

We obtain from (15) analogously as in (a):

$$\sin\theta_\pm{}^*(2m-1,\nu)=\frac{\nu\pi\pm\varrho(m)}{4m-1}\left(1+O\left(\frac{\nu^2}{m^2}\right)\right).$$

Using this a simple calculation shows that $B_\nu\subset A_\nu$. Hence taking into account the earlier considerations we get that

$$H_{2m-1}\left(\sqrt{4m-1}\cos\left(\frac{\pi}{2}-\theta^*\right)\right)\neq 0$$
$$\text{if}\quad\theta^*_+(2m-1,\nu)\leq\theta^*\leq\theta^*_-(2m-1,\nu+1).$$

In the case of $0<\theta^*\leq\theta^*_-(2m-1,1)$ use the fact that $H_{2m-1}(0)=0$. A similar calculation as above gives:

$$H_{2m-1}\left(\sqrt{4m-1}\cos\left(\frac{\pi}{2}-\theta^*\right)\right)\neq 0.$$

Thus we have proved that

$$\sqrt{4m-1}\sin\theta^*_-(2m-1,\nu)<x_\nu<\sqrt{4m-1}\sin\theta^*_+(2m-1,\nu),$$

and hence by a similar calculation as in (a) we get

$$x_\nu=\frac{\nu\pi}{\sqrt{4m-1}}+O\left(\frac{\varrho(m)}{\sqrt{m}}\right)+O\left(\frac{\nu^3}{m^{5/2}}\right),$$

and this proves (2).

Now we return to the proof of (3). To this let

$$g(y):=\frac{\sin 2y+2y-4\sin y}{4\sin y},\qquad\left(0<y<\frac{1}{4}\right).$$

As we have seen on p.28, $g(y)=O(y^2)$, where the implicit constant is absolute and effective, hence there exists an absolute constant $c_3>0$ such that $|g(y)|<1$ if $0<y\le c_3$. Let $\sin 2y+2y=A$, then $g(y)=\frac{A}{4\sin y}-1$ and hence

$$\sin y=\frac{A}{4}\cdot\frac{1}{1+g(y)}. \tag{16}$$

The Taylor expansion of $g(y)$ has the form

$$g(y)=\sum_{k=2}^{\infty}e_k y^k.$$

Because $|g(y)|<1$, hence any formal calculations are permitted in the Taylor series of $g(y)$, and taking into account the geometric series we get:

$$\frac{1}{1+g(y)}=1+\sum_{k=2}^{\infty}d_k y^k.$$

Next suppose $n=2m$ (the case $n=2m-1$ is similar). We have seen that

$$x_\nu=\sqrt{4m+1}\sin\theta_-^*(\nu)+O\left(\frac{\varrho(m)}{\sqrt{m}}\right),$$

where

$$\sin 2\theta_-^*(\nu)+2\theta_-^*(\nu)=\frac{4\left(\nu\pi-\frac{\pi}{2}-\varrho(m)\right)}{4m+1}.$$

Apply (16) and use the substitutions

$$y=\theta_-^*(\nu),\qquad A=\frac{4\left(\nu\pi-\frac{\pi}{2}-\varrho(m)\right)}{4m+1}.$$

We get

$$\sin\theta_-^*(\nu)=\frac{\nu\pi-\frac{\pi}{2}-\varrho(m)}{4m+1}\left(1+\sum_{k=2}^{\infty}d_k[\theta_-^*(\nu)]^k\right),$$

hence

$$x_{\nu+1}-x_\nu=\frac{(\nu+1)\pi-\frac{\pi}{2}-\varrho(m)}{\sqrt{4m+1}}\sum_{k=2}^{\infty}d_k[\theta_-^*(\nu+1)]^k-$$

$$-\frac{\nu\pi-\frac{\pi}{2}-\varrho(m)}{\sqrt{4m+1}}\sum_{k=2}^{\infty}d_k[\theta_-^*(\nu)]^k+\frac{\pi}{\sqrt{4m+1}}+O\left(\frac{\varrho(m)}{\sqrt{m}}\right).$$

From (14)
$$|\theta_-^*(\nu+1)-\theta_-^*(\nu)|=O\left(\frac{1}{m}\right).$$

Because

$$|a^k-b^k|\le|a-b|k\cdot\max(a^{k-1},b^{k-1}),\qquad a,\ b>0,$$

hence

$$[\theta^*_-(\nu+1)]^k - [\theta^*_-(\nu)]^k = O(1)k\cdot\frac{1}{m}\cdot\frac{\nu^{k-1}}{m^{k-1}},$$

so

$$(\nu+1)[\theta^*_-(\nu+1)]^k - \nu[\theta^*_-(\nu)]^k = [\theta^*_-(\nu+1)]^k + \nu\{[\theta^*_-(\nu+1)]^k - [\theta^*_-(\nu)]^k\} =$$
$$= O(1)k\frac{\nu^k}{m^k} + O(1)\frac{\nu^k}{m^k} = O(1)k\cdot\frac{\nu^k}{m^k},$$

and using this we get

$$x_{\nu+1} - x_\nu = \frac{\pi}{\sqrt{4m+1}} + O\left(\frac{\varrho(m)}{\sqrt{m}}\right) + O(1)\frac{1}{\sqrt{m}}\sum_{k=2}^{\infty}|d_k|k\frac{\nu^k}{m^k}.$$

We have seen that $\sum_{k=2}^{\infty} d_k y^k$ is an absolutely and uniformly convergent Taylor series in $0\le y\le c_3$, hence $|d_k|\le\frac{o(1)}{c_3^k}$, so, taking into account $\frac{\nu}{m} = O\left(\frac{1}{\beta(m)}\right)$ we get

$$|d_k|k\frac{\nu^k}{m^k} = o(1)\frac{k}{c_3^k}\cdot\frac{1}{\beta^k(m)}.$$

Let $\beta(m)\ge\frac{2}{c_3}$, then

$$\sum_{k=2}^{\infty}|d_k|k\frac{\nu^k}{m^k} = o(1)\sum_{k=2}^{\infty}\frac{k}{2^k} = o(1),$$

i.e. the remainder term is remainder indeed, and the Theorem 1 is proved.

PROOF OF THE THEOREM 2. We need

$$e^{-\frac{x^2}{2}}H_n(x) = (2N^n)^{\frac{1}{2}}\cdot\exp\left(-\frac{N}{4}\right)\cdot(\sin\theta)^{\frac{1}{2}}\cdot$$
$$\cdot\left\{\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n\theta^3}\right)\right\}, \tag{17}$$

where $x := \sqrt{2n+1}\cos\theta$, $0<\theta\le\pi/2$, $N = 2n+1$, [2], p.22–23. Here the remainder terms are uniform in θ. The main term of (17) is

$$\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta)+\frac{3\pi}{4}\right].$$

Let

$$\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta) = \nu\pi - \frac{3\pi}{4}\pm\varrho(n,\nu), \tag{18}$$

where $\nu < 0$ is an integer and $0 < \varrho(n,\nu) \le \frac{\pi}{8}$. Then for the main term of (17) we get

$$\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta) + \frac{3\pi}{4}\right] = \sin(\nu\pi \pm \varrho(n,\nu)) = (-1)^\nu \cdot (\pm \sin\varrho(n,\nu)).$$

Because by (18) we have $n\theta^3 \asymp |\nu|$, hence for $n \ge n_0$ the main term is the main term indeed in (17) if and only if

$$\varrho(n,\nu) = \frac{\varrho^*(n,\nu)}{|\nu|}, \quad \varrho^*(n,\nu) \ge a > 0, \quad \varrho(n,\nu) \le \frac{|\nu|}{20},$$

where a is a "large" absolute constant. In (18) denote θ_+ resp. θ_- the θ corresponding to $+\varrho(n,\nu)$ resp. $-\varrho(n,\nu)$. It follows from the above considerations that

$$\operatorname{sign} H_n(\sqrt{2n+1}\cos\theta_+) = (-1)^\nu, \quad \operatorname{sign} H_n(\sqrt{2n+1}\cos\theta_-) = (-1)^\nu \cdot (-1),$$

which means that H_n changes sign in the interval

$$(19) \qquad (\sqrt{2n+1}\cos\theta_-, \sqrt{2n+1}\cos\theta_+)$$

which means that H_n has at least one zeros in this interval. Let $\varrho(n,\nu) = \frac{a}{|\nu|}$. Then in the case of $|\nu| \ge \nu_0(a)$ the main term in (17) is the main term indeed (for $n \ge n_0(\nu_0)$). This means that in $n-c$ interval of type (19) contains zeros of H_n, where c is an absolute constant. In $(-\infty,\infty)$ H_n has n zeros, hence we are not able to say what is the position of "finitely many" zeros of H_n, but we shall see, that the asymptotic formula given below works also for them, further in applications (usually) finitely many zeros do not effect the final result.

The length of the interval is

$$\sqrt{2n+1}\cos\theta_+(\nu) - \sqrt{2n+1}\cos\theta_-(\nu) = \sqrt{2n+1}(\cos\theta_+(\nu) - \cos\theta_-(\nu))$$

and by the mean-value theorem we get

$$\cos\theta_+(\nu) - \cos\theta_-(\nu) = -\sin\xi\cdot(\theta_+(\nu) - \theta_-(\nu)) \asymp -\theta_+(\nu)\cdot(\theta_+(\nu) - \theta_-(\nu)) \asymp$$

$$(20) \qquad \asymp \frac{-|\nu|^{1/3}}{n^{1/3}} \cdot (\theta_+(\nu) - \theta_-(\nu)).$$

From (18) we get

$$\sin 2\theta_+(\nu) - 2\theta_+(\nu) - (\sin 2\theta_-(\nu) - 2\theta_-(\nu)) = \frac{2\varrho(n,\nu)}{\frac{n}{2}+\frac{1}{4}}.$$

Applying the mean value theorem for the function $f(\theta) = \sin 2\theta - 2\theta$ we obtain

$$\frac{2\varrho(n,\nu)}{\frac{n}{2}+\frac{1}{4}} = (\theta_+(\nu) - \theta_-(\nu))2(\cos 2\xi - 1) = -2(\theta_+(\nu) - \theta_-(\nu))\sin^2\xi,$$

$\theta_+(\nu)\leq\xi\leq\theta_-(\nu)$, i.e.

(21) $$\frac{\varrho(n,\nu)}{n}\asymp-\theta_+^2(\nu)\cdot(\theta_+(\nu)-\theta_-(\nu))\asymp-\frac{|\nu|^{\frac{2}{3}}}{n^{2/3}}(\theta_+(\nu)-\theta_-(\nu)).$$

From (20) and (21) we get

$$\cos\theta_+(\nu)-\cos\theta_-(\nu)\asymp\frac{\varrho(n,\nu)}{n^{2/3}\cdot|\nu|^{1/3}},$$

i.e. the length of the interval is

$$\sqrt{2n+1}\cos\theta_+(\nu)-\sqrt{2n+1}\cos\theta_-(\nu)\asymp\frac{\varrho(n,\nu)}{n^{1/6}\cdot|\nu|^{1/3}}.$$

Denote $0<\theta_0(\nu)\leq\frac{\pi}{2}$ the unique solution of

(22) $$\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta-2\theta)=\nu\pi-\frac{3\pi}{4},$$

then

(23) $$x_{|\nu|}=\sqrt{2n+1}\cos\theta_0(\nu)+\frac{O(1)\cdot\varrho(n,\nu)}{n^{1/6}\cdot|\nu|^{1/3}},$$

where it is not possible to write $o(1)$ in place of $O(1)$. It follows from (22) that $\theta_0^3(\nu)\asymp\frac{|\nu|}{n}$, hence

$$\cos\theta_0(\nu)=1-\frac{\theta_0^2(\nu)}{2}+O(1)\theta_0^4(\nu)=1-O(1)\frac{|\nu|^{2/3}}{n^{2/3}}+O(1)\frac{|\nu|^{4/3}}{n^{4/3}}.$$

Let $|\nu|\leq dn$, where $d>0$ is a small constant. Then

$$\cos\theta_0(\nu)=1-O(1)\frac{|\nu|^{2/3}}{n^{2/3}},\qquad O(1)>0,$$

where it is not possible to write $o(1)$ in place of $O(1)$. We get

(24) $$x_{|\nu|}=\sqrt{2n+1}\left(1-O(1)\frac{|\nu|^{2/3}}{n^{2/3}}\right)+\frac{O(1)\varrho(n,\nu)}{n^{1/6}|\nu|^{1/3}}=$$

$$=\sqrt{2n+1}-O(1)\frac{|\nu|^{2/3}}{n^{1/6}}+\frac{O(1)\varrho(n,\nu)}{n^{1/6}|\nu|^{1/3}}=\sqrt{2n+1}-O(1)\frac{|\nu|^{2/3}}{n^{1/6}},$$

where $\varrho(n,\nu)=\frac{a}{|\nu|}$, $d\cdot n\geq|\nu|\geq\nu_0$ with a suitable small constant d, $\nu_0=\nu_0(a)$ is a fixed constant; here we can not write $o(1)$ in place of $O(1)$.

Now estimate the distance of the zeros. Using (23) we get

$$x_{|\nu+1|}-x_{|\nu|}=\sqrt{2n+1}(\cos\theta_0(\nu+1)-\cos\theta_0(\nu))+\frac{O(1)\varrho(n,\nu)}{n^{1/6}|\nu|^{1/3}},$$

and similarly to the earlier estimates

$$\cos\theta_0(\nu+1)-\cos\theta_0(\nu)\asymp\frac{1}{n^{2/3}|\nu|^{1/3}},$$

hence

$$x_{|\nu+1|}-x_{|\nu|}=\frac{O(1)}{n^{1/6}|\nu|^{1/3}}+\frac{O(1)\varrho(n,\nu)}{n^{1/6}|\nu|^{1/3}}=\frac{O(1)}{n^{1/6}|\nu|^{1/3}},\tag{25}$$

because $\varrho(n,\nu)=\frac{a}{|\nu|}$. We cannot write $o(1)$ in place of $O(1)$.

Now investigate, what can we say on the excluded "finitely many" zeros. In the case of $|\nu|\le\nu_0$, the distribution of the zeros is well known, we can express them using the zeros of the Airy functions. This coincides with (23) and (24) extending them to $|\nu|\le\nu_0$. This means that (23) holds for every zeros and the proof of the Theorem 2 is complete.

We remark that the method of the proof of the Lemma 1 gives the following stronger result:

LEMMA 1*.

$$\frac{\mathrm{d}}{\mathrm{d}\theta}H_n(\sqrt{2n+1}\cos\theta)=\tag{26}$$

$$=k_n(\theta)\cdot(-2n)\sqrt{2n+1}\sin\theta\cdot(2(2n-1)^{n-1})^{\frac{1}{2}}\cdot\exp\left(-\frac{2n-1}{4}\right)\cdot$$

$$\cdot\left[e^{-\cos^2\theta}\left\{\sin\left[\left(\frac{n}{2}-\frac{1}{4}\right)(\sin 2\theta-2\theta)+\frac{3\pi}{4}\right]+O\left(\frac{1}{n\theta^3}\right)\right\}+O(1)\cos\theta\cdot\frac{1}{n\theta^3}\right],$$

where $0<\frac{c}{\sqrt{n}}\le\theta\le\frac{\pi}{2}$, further the remainder terms are uniformly in θ.

Recently S. SZABÓ extended the results of this paper for the case of Laguerre functions [3].

References

[1] G. SZEGŐ, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publ., vol 23., Amer. Math. Soc., Providence, Rhode Island, 1939, 4-th ed., 1975.

[2] A. ERDÉLYI, Asymptotic solutions of differential equations with transition points or singularities, *J. Mathematical Phys.,* 1 (1960), 16–26. MR 22#2773.

[3] S. SZABÓ, On the distribution of the zeros of Laguerre functions. (Personal communication.)

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 37–57

KLASSIFIKATION DER KEGELSCHNITTBÜSCHEL MIT VIER REELLEN GRUNDPUNKTEN UND EINEM PARALLELEN GRUNDPUNKTEPAAR IN DER ISOTROPEN EBENE

Von

KATHARINA MOLNÁR-SÁSKA

"Ybl Miklós" Technische Fachhochschule, Budapest

(Eingegangen am 16. Juli, 1991)

1. Einleitung

Eine reelle affine Ebene A_2, die durch eine Gerade f mit einem inzidenten Punkt $F \in f$ metrisiert wird, heißt eine isotrope Ebene I_2; in diesem Zusammenhang wird f als absolute Gerade und F als absoluter Punkt bezeichnet. Die interessante Geometrie dieser Ebene kann in [1] nachgelesen werden.

Ein Kegelschnittbüschel wird bekanntlich durch zwei Kegelschnitte eindeutig bestimmt. Diese beiden Kegelschnitte schneiden sich in der komplexen Erweiterung $I_2(C)$ im algebraischen Sinn in vier Punkten A, B, C und D. Sie sind die Grundpunkte des Kegelschnittbüschels; die Geraden durch die vier Grundpunkte heißen Grundgeraden. Ein beliebiger Punkt der Ebene bestimmt zusammen mit den vier Grundpunkten einen Kegelschnitt dieses Büschels eindeutig. Je nach der Realität bzw. der Vielfachheit dieser Grundpunkte läßt sich in I_2 eine große Zahl von Büscheltypen unterscheiden (vgl. [2]–[5]). In dieser Arbeit untersuchen wir nur die Büscheln vom Typ II d.h. wir befassen uns mit jenen Kegelschnittbüscheln, welche 4 reelle und verschiedene Grundpunkte besitzen, wobei aber stets 2 Grundpunkte parallel sind, d.h. auf einer isotropen Geraden liegen. Die Unterfälle vom Typ II der isotropen Ebene werden vollständig klassifiziert und durch vollständige Invariantensysteme beschrieben, wobei wir uns analytischer Methoden bedienen.

Wir verwenden projektive Koordinaten $(x_0 : x_1 : x_2)$ in P_2 und affine Koordinaten (x, y) in I_2. Bekanntlich gilt dann $x = \frac{x_1}{x_0}$, $y = \frac{x_2}{x_0}$ für $x_0 \neq 0$. Es ist üblich zu setzen: $f \ldots x_0 = 0$; $F(0:0:1)$.

Dann erhält man für die allgemeinen isotropen Ähnlichkeiten die Darstellung

$$(1.1) \qquad \begin{cases} \bar{x} = c_1 + c_2 x \\ \bar{y} = c_3 + c_4 x + c_5 y \end{cases} \qquad c_2 \cdot c_5 \neq 0 .$$

Alle möglichen Typen von $\mathbb{I}$ Kegelschnittbüscheln der isotropen Ebene werden analytisch bestimmt und unter geometrischen Gesichtspunkten beschrieben. Als Fundamentalgruppe legen wir hierbei die dreigliedrige isotrope Bewegungsgruppe B_3

$$(1.2) \qquad \begin{cases} \bar{x} = c_1 + x \\ \bar{y} = c_3 + c_4 x + y \end{cases}$$

zugrunde.

Bei der Klassifikation beschränken wir uns hierbei stets auf nicht ausgeartete Kegelschnittbüschel, d.h. wir betrachten keine Paare von Geradenbüscheln.

In dieser Arbeit wird die Klassifikation mit Hilfe der Mittelpunktskurve k_m und der isotropen Brennpunktskurve k_f vorgenommen. Wir unterscheiden weiters drei Fälle vom Typ $\mathbb{I}$ in der isotropen Ebene:

ERSTER FALL ($\mathbb{I}_0$): Es gibt zwei parallele Punktepaare unter den vier Grundpunkten.

ZWEITER FALL ($\mathbb{I}_A$): Es gibt ein paralleles Punktepaar unter den vier Grundpunkten und die anderen zwei Grundpunkte liegen auf derselben Seite der isotropen Grundgeraden.

DRITTER FALL ($\mathbb{I}_B$): Es gibt ein paralleles Punktepaar unter den vier Grundpunkten und die anderen zwei Grundpunkte liegen auf verschiedenen Seiten der isotropen Grundgeraden.

Dann gilt zusammenfassend als Hauptergebnis der

SATZ 1.1: *In der isotropen Ebene I_2 existieren bezüglich der isotropen Bewegungsgruppe B_3 genau 2 verschiedene Typen der Klasse $\mathbb{I}_0$; genau 15 Typen der Klasse $\mathbb{I}_A$ und genau 14 Typen der Klasse $\mathbb{I}_B$.*

Auflistung und Kurzbeschreibung der Büscheltypen

$\mathbb{I}_0 1$. Das Grundviereck $BDAC$ ist ein nichtzulässiges Parallelogramm in der isotropen Ebene. Der Hauptteil der Mittelpunktskurve is ein Punkt, genauer der Mittelpunkt des Grundparallelogramms. Der Hauptteil der Brennpunktskurve ist eine reelle Mittellinie des Grundparallelogramms.

$\mathbb{I}_0 2$. Das Grundviereck $BDAC$ ist ein nichtzulässiges Trapez (aber kein Parallelogramm) in der isotropen Ebene, wobei die parallelen Seiten die isotropen Grundgeraden sind. Der Hauptteil der Mittelpunktskurve und der Brennpunktskurve ist eine reelle Mittellinie des Grundtrapezes.

$\mathrm{I\!I}_A 1$. Das Grundviereck $BDAC$ ist ein nichtzulässiges Trapez (aber kein Parallelogramm) in der isotropen Ebene, wobei die parallelen Seiten die reellen Grundgeraden sind. Der Hauptteil der Mittelpunktskurve ist die reelle Mittellinie des Grundtrapezes. Der Hauptteil der Brennpunktskurve ist eine Hyperbel 1. Art.

$\mathrm{I\!I}_A 2$, 3, 4. Das Grundviereck $BDAC$ ist ein nichtzulässiges konvexes Viereck in der isotropen Ebene, wobei es kein reelles paralleles Geradenpaar unter den 6 durch die vier Grundpunkte legbaren Geraden gibt. Die Mittelpunktskurve ist eine Hyperbel 1. Art. Wenn man die Lage der Mittelpunktshyperbel zu einer Diagonale des Grundvierecks in betracht zieht, so ergeben sich 3 Typen von Kegelschnittbüscheln. Der Halbierungspunkt einer Diagonale des Grundvierecks ist:

$\mathrm{I\!I}_A 2$ der Mittelpunkt einer Hyperbel 1. Art;

$\mathrm{I\!I}_A 3$ der Mittelpunkt eines reellen, sich schneidenden Geradenpaares;

$\mathrm{I\!I}_A 4$ der Mittelpunkt einer Hyperbel 2. Art im Kegelschnittbüschel.

In allen drei Fällen ist der Hauptteil der Brennpunktskurve eine Hyperbel 1. Art.

$\mathrm{I\!I}_A 5$, 6, 7. Die Grundpunkte B, D, A, C bestimmen ein nichtzulässiges konvexes Viereck in der isotropen Ebene, wobei es kein reelles parelleles Geradenpaar unter den 6 durch die vier Grundpunkte legbaren Geraden gibt. Die Mittelpunktskurve ist eine Hyperbel 2. Art. Wenn man die Lage der Mittelpunktshyperbel zu einer Diagonale des Grundvierecks in Betracht zieht, so ergeben sich 3 Typen von Kegelschnittbüscheln. Der Halbierungspunkt einer Diagonale des Grundvierecks ist der Mittelpunkt:

$\mathrm{I\!I}_A 5$ einer Hyperbel 1. Art;

$\mathrm{I\!I}_A 6$ eines reellen, sich schneidenden Geradenpaares;

$\mathrm{I\!I}_A 7$ einer Hyperbel 2. Art im Kegelschnittbüschel.

In allen drei Fällen ist der Hauptteil der Brennpunktskurve eine Hyperbel 1. Art.

$\mathrm{I\!I}_A 8$, 9, 10, 11, 12, 13, 14, 15. Die konvexe Hülle der vier Grundpunkte B, D, A, C ist ein nichtzulässiges Dreieck der isotropen Ebene I_2. Die Mittelpunktskurve ist eine Ellipse. Wenn man die Lage der Mittelpunktsellipse zu den Seiten der konvexen Hülle in Betracht zieht, so ergeben sich 8 Typen von Kegelschnittbüscheln. Die isotrope Grundgerade schneidet die Mittelpunktsellipse in 2 Punkten und die Halbierungspunkte der 2 nicht isotropen Seiten der konvexen Hülle sind;

$\mathrm{I\!I}_A 8$ die Mittelpunkte einer Hyperbel 1. Art;

$\mathrm{I\!I}_A 9$ die Mittelpunkte einer Hyperbel 2. Art;

$\mathbb{I}_A 10$ einer ist der Mittelpunkt einer Hyperbel 1. Art, der andere der Mittelpunkt einer Hyperbel 2. Art;

$\mathbb{I}_A 11$ einer ist der Mittelpunkt eines reellen, sich schneidenden Geradenpaares, der andere der Mittelpunkt einer Hyperbel 1. Art;

$\mathbb{I}_A 12$ einer ist der Mittelpunkt eines reellen, schneidenden Geradenpaares, der andere der Mittelpunkt einer Hyperbel 2. Art im Kegelschnittbüschel.

Die isotrope Grundgerade berührt die Mittelpunktsellipse und die Halbierungspunkte der zwei nicht isotropen Seiten der konvexen Hülle sind die Mittelpunkte:

$\mathbb{I}_A 13$ einer Hyperbel 1. Art;

$\mathbb{I}_A 14$ einer Hyperbel 2. Art;

$\mathbb{I}_A 15$ eines reellen, sich schneidenden Geradenpaares.

In allen 8 Fällen ist der Hauptteil der Brennpunktskurve eine Ellipse.

$\mathbb{I}_B 1$. Das Grundviereck $DABC$ ist ein zulässiges Parallelogramm der isotropen Ebene, wobei die Diagonale DB die isotrope Grundgerade ist. Der Hauptteil der Mittelpunktskurve ist ein Punkt, genauer der Mittelpunkt des Grundparallelogramms. Der Hauptteil der Brennpunktskurve ist eine Hyperbel 1. Art.

$\mathbb{I}_B 2$. Das Grundviereck $DABC$ ist ein zulässiges Trapez (kein Parallelogramm) in der isotropen Ebene, wobei die Diagonale DB die isotrope Grundgerade ist. Der Hauptteil der Mittelpunktskurve ist die reelle Mittellinie des Grundtrapezes. Der Hauptteil der Brenpunktskurve ist eine Hyperbel 1. Art.

$\mathbb{I}_B 3$, 4. Das Grundviereck $DABC$ ist ein zulässiges konvexes Viereck der isotropen Ebene (kein Trapez), wobei die Diagonale DB die isotrope Grundgerade ist. Die Mittelpunktskurve ist eine Hyperbel 1. Art. Wenn man die Lage der Mittelpunktshyperbel zu einer Diagonale des Grundvierecks in Betracht zieht, so ergeben sich 2 Typen von Kegelschnittbüscheln. Der Halbierungspunkt der isotropen Diagonale ist der Mittelpunkt:

$\mathbb{I}_B 3$ einer Hyperbel 2. Art;

$\mathbb{I}_B 4$ eines sich schneidenden Geradenpaares, wobei eine Gerade im Kegelschnittbüschel isotrop ist.

In beiden Fällen ist der Hauptteil der Brennpunktskurve eine Hyperbel 1. Art.

$\mathbb{I}_B 5$, 6. Das Grundviereck $DABC$ ist ein zulässiges konvexes Viereck (kein Trapez), wobei die Diagonale DB isotrop ist. Die Mittelpunktskurve ist eine Hyperbel 2. Art. Wenn man die Lage der Mittelpunktshyperbel bezüglich der nichtisotropen Diagonale des Grundvierecks in Betracht zieht, so ergeben sich 2 Typen von Kegelschnittbüscheln. Der Halbierungspunkt der nicht isotropen Diagonale ist entweder

$\mathbb{I}_B$5 der Mittelpunkt einer Hyperbel 2. Art, oder

$\mathbb{I}_B$6 der Mittelpunkt eines sich schneidenden Geradenpaares, wobei eine Gerade im Kegelschnittbüschel isotrop ist.

In beiden Fällen ist der Hauptteil der Brennpunktskurve eine Hyperbel 1. Art.

$\mathbb{I}_B$7, 8, 9, 10, 11, 12, 13, 14. Die konvexe Hülle der vier Grundpunkte D, A, B, C ist ein zulässiges Dreieck DAC in der isotropen Ebene I_2. Es enthält den zu D parallelen Grundpunkt B. Die Mittelpunktskurve ist eine Ellipse. Wenn man die Lage der Mittelpunktsellipse zu den Seiten der konvexen Hülle in Betracht zieht, so ergeben sich 8 Typen von Kegelschnittbüscheln. Die Grundgerade AC schneidet die Mittelpunktsellipse in zwei Punkten und die Halbierungspunkte der anderen zwei Seiten der konvexen Hülle sind:

$\mathbb{I}_B$7 die Mittelpunkte einer Hyperbel 1. Art;

$\mathbb{I}_B$8 die Mittelpunkte einer Hyperbel 2. Art;

$\mathbb{I}_B$9 einer ist der Mittelpunkt einer Hyperbel 1. Art, der andere der Mittelpunkt einer Hyperbel 2. Art;

$\mathbb{I}_B$10 einer ist der Mittelpunkt eines reellen, sich schneidenden Geradenpaares, der andere der Mittelpunkt einer Hyperbel 1. Art;

$\mathbb{I}_B$11 einer ist der Mittelpunkt eines reellen, sich schneidenden Geradenpaares, der andere der Mittelpunkt einer Hyperbel 2. Art im Kegelschnittbüschel.

Die Grundgerade AC berührt die Mittelpunktsellipse und die Halbierungspunkte der anderen zwei Seiten der konvexen Hülle sind die Mittelpunkte:

$\mathbb{I}_B$12 einer Hyperbel 1. Art;

$\mathbb{I}_B$13 einer Hyperbel 2. Art;

$\mathbb{I}_B$14 eines reellen, sich schneidenden Geradenpaares im Kegelschnittbüschel.

In allen 8 Fällen ist der Hauptteil der Brennpunktskurve eine Ellipse.

2. Klassifikation der Kegelschnittbüschel

Auf den umfangreichen Beweis des Satzes 1.1 muß hier verzichtet werden. Durch exemplarische Beschreibung der Typen $\mathbb{I}_A$2, 3, 4, 5, 6, 7 soll jedoch die Untersuchungsmethode vorgeführt werden.

Wir bezeichnen mit D, A, C, B die Grundpunkte des Kegelschnittbüschels, wobei D zu B parallel ist, mit A_1, B_1, C_1, E_1, G_1, D_1 die Halbierungspunkte der Grundstrecken BD, DA, AC, CB, DC, AB und mit P, Q, R die Schnittpunkte der Grundgeraden BC, DA; BD, CA und BA, DC.

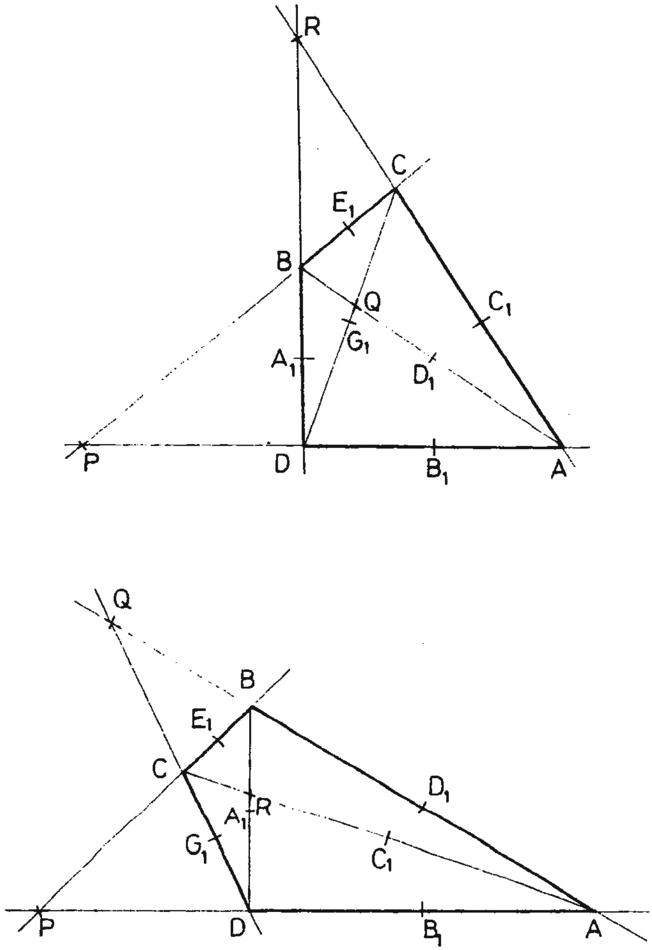

Fig. 1.

Wir gehen stets von der allgemeinen Gleichung

$$a_{11}x^2 + a_{22}y^2 + 2a_{12}xy + 2a_{01}x + 2a_{02}y + a_{00} = 0 \tag{2.1}$$

eines Kegelschnittes in affinen Koordinaten aus und versuchen zunächst eine Normalform des jeweiligen Büscheltyps mittels geometrischer überlegungen herzuleiten.

Die Typen $\mathbb{I}_A 2$, 3, 4, 5, 6, 7

Das Grundviereck $BDAC$ ist ein nichtzulässiges konvexes Viereck der isotropen Ebene (kein Trapez), wobei für die isotrope Spanne $s(DB) > 0$ gilt.

Man kann dann durch eine isotrope Bewegung erreichen, daß D in den Ursprung und die Grundgerade DA in die x-Achse des zugrundegelegten Koordinatensystems fällt. Dann sind die Koordinaten der Grundpunkte durch

(2.2) $$D(0,0); \quad A(a,0); \quad B(0,b) \quad \text{und}$$

(2.3) $$C(c_1, c_2) \quad \text{mit} \quad b, c_2 \in \mathbb{R}^+; \quad a, c_1 \in \mathbb{R}, \text{ aber } a \neq c_1, b \neq c_2 \text{ mit}$$

(2.4) $$ac_1 > 0 \quad \text{und} \quad c_2 > -\frac{b}{a}c_1 + b \quad \text{gegeben.}$$

Die Gleichung (2.1) reduziert sich wegen (2.2) auf

(2.5) $$a_{11}x^2 + a_{22}y^2 + 2a_{01}x + 2a_{02}y + 2a_{12}xy = 0.$$

Da die Koordinaten von A, B, C die Gleichung (2.5) erfüllen müssen, ergeben sich folgende Gleichungen:

(2.6) $$a_{11}a^2 + 2a_{01}a = 0; \quad a_{22}b^2 + 2a_{02}b = 0$$
$$a_{11}c_1^2 + a_{22}c_2^2 + 2a_{01}c_1 + 2a_{02}c_2 + 2a_{12}c_1c_2 = 0.$$

Aus (2.6) folgt wegen $ab \neq 0$

(2.7) $$2a_{01} = -a_{11}a; \; 2a_{02} = -a_{22}b \text{ und } 2a_{12} = a_{11}\frac{(a-c_1)}{c_2} + a_{22}\frac{(b-c_2)}{c_1}.$$

Es sei

(2.8) $$k_1 = \frac{c_2}{c_1 - a} \qquad \text{und} \qquad k_2 = \frac{c_2 - b}{c_1}.$$

Aus (2.5) gewinnt man mittels (2.7) mit den Umbezeichungen (2.8) und $\lambda = -\frac{a_{11}}{k_1 a_{22}}$ als <u>Normalform</u> dieses Büscheltyps:

(2.9) $$F \equiv \lambda(xy - k_1x^2 + k_1ax) + y^2 - k_2xy - by = 0.$$

Aus (2.2), (2.3) und (2.8) sieht man sofort, daß a der isotrope Abstand und b die isotrope Spanne zwischen D, A bzw. D, B ist. k_1 und k_2 sind die isotropen Winkeln zwischen den Grundgeraden DA, AC bzw. DA, BC, d.h.

(2.10) $$a = d(DA), \quad b = s(DB), \quad k_1 = \sphericalangle(DA, AC), \quad k_2 = \sphericalangle(DA, BC).$$

Damit sind a, b, k_1, k_2 als isotrope Bewegungsinvarianten erkannt.

Durch eine einfache Rechnung ergeben sich aus (2.3), (2.4) und (2.8) die folgenden Gleichungen

$$c_1 = \frac{k_1 a + b}{k_1 - k_2}; \quad c_2 = \frac{k_1 k_2 a + k_1 b}{k_1 - k_2} \quad \text{bzw.} \tag{2.11}$$

Ungleichungen

$$a k_1 (a k_1 + b) > 0, \quad b > 0; \quad k_2 a + b > 0 \tag{2.12}$$

$$k_1^2 - k_1 k_2 > 0 \text{ für die Invarianten } a, b, k_1, k_2. \tag{2.13}$$

Aus (2.9) kann man mittels der Determinante

$$\begin{vmatrix} -\lambda k_1 & \frac{\lambda - k_2}{2} \\ \frac{\lambda - k_2}{2} & 1 \end{vmatrix} \tag{2.14}$$

erkennen, daß das Kegelschnittbüschel für

$$\lambda_1 = -2k_1 + k_2 - 2\sqrt{k_1^2 - k_1 k_2}; \quad \lambda_2 = -2k_1 + k_2 + 2\sqrt{k_1^2 - k_1 k_2} \tag{2.15}$$

reelle Parabeln, für $\lambda_1 < \lambda < \lambda_2$ Ellipsen und für $\lambda < \lambda_1$, $\lambda_2 < \lambda$ Hyperbeln enthält.

Nach einiger Rechnung ergibt sich, daß die Durchmesserrichtungen der Parabeln durch

$$\left(1,\ k_1 + \sqrt{k_1^2 - k_1 k_2}\right); \quad \left(1,\ k_1 - \sqrt{k_1^2 - k_1 k_2}\right) \tag{2.16}$$

gegeben sind.

Aus (2.16) folgt für die Winkeln zwischen den Durchmesserrichtungen dieser Parabeln und der Grundgeraden DA

$$\varphi_1 = k_1 + \sqrt{k_1^2 - k_1 k_2}; \quad \varphi_2 = k_1 - \sqrt{k_1^2 - k_1 k_2}. \tag{2.17}$$

Aus (2.17) findet man

$$\varphi_1 \varphi_2 = k_1 k_2 \quad \text{und} \quad \varphi_1 + \varphi_2 = 2k_1. \tag{2.18}$$

Aus (2.9) kann man mittels $\frac{\partial F}{\partial y} = 0$ die Typen der Hyperbeln im Kegelschnittbüschel bestimmen. Das Büschel enthält für

$$\lambda_3 = 0 \quad \text{und} \quad \lambda_4 = \frac{k_2 a b + k_1 b^2}{k_1 a^2 + a b} \tag{2.19}$$

reelle, schneidende Geradenpaare mit den Mittelpunkten P und Q, für $\lambda \in (\lambda_3, \lambda_4)$ Hyperbeln 2. Art, und schließlich für $\lambda \in \mathbb{R} \setminus \{[\lambda_1, \lambda_2] \cup [\lambda_3, \lambda_4]\}$ Hyperbeln 1. Art.

Im Büschel liegt auch ein schneidendes Geradenpaar mit dem Mittelpunkt R, wobei eine Gerade isotrop ist ($\lambda = \infty$).

Mittels $\frac{\partial F}{\partial x}=\frac{\partial F}{\partial y}=0$ und (2.9) bestimmt man die Mittelpunktskurve. Man findet

(2.20) $-2k_1k_2x^2-2y^2+4k_1xy+(k_1k_2a-2k_1b)x+(b-2k_1a)y+k_1ab=0.$

Diese Kurve ist wegen (2.13) eine Hyperbel.

Eine andere Darstellung der Mittelpunktskurve ist:

(2.21) $x(\lambda)=\dfrac{k_2b-\lambda b-2\lambda k_1a}{-k_2^2+2\lambda k_2-4\lambda k_1-\lambda^2};\quad y(\lambda)=\dfrac{-2\lambda k_1b-\lambda k_1k_2a+\lambda^2k_1a}{-k_2^2+2\lambda k_2-4\lambda k_1-\lambda^2},$

wobei λ der Büschelparameter ist.

Die Funktionen $x(\lambda)$, $y(\lambda)$ sind stetig in den Intervalle (λ_1,λ_2); $(-\infty,\lambda_1)$ und (λ_2,∞). Weiter kann man aus (2.21) sehen, daß $\lim\limits_{|\lambda|\to\infty}x(\lambda)=0$ und $\lim\limits_{|\lambda|\to\infty}y(\lambda)=-k_1a$ gilt. Dieser Punkt mit den Koordinaten $(0,-k_1a)$ ist R. Aus (2.21) folgt $\lim\limits_{\lambda\to\lambda_1}|x(\lambda)|=\infty$ und $\lim\limits_{\lambda\to\lambda_1}|y(\lambda)|=\infty$ $i=1,2$. Genau so findet man: Im Fall $\lambda\in(\lambda_1,\lambda_2)$ gehört λ zu einer Ellipse des Kegelschnittbüschels und (2.21) beschreiben einen Zweig der Mittelpunktshyperbel. Im Fall $\lambda\in\mathbb{R}\setminus\{[\lambda_1,\lambda_2]\}$ und $\lambda=\infty$ gehört λ zu einer Hyperbel 1. oder 2. Art oder zu einer degenerierten Hyperbel. (2.21) beschreiben dann den anderen Zweig der Mittelpunktshyperbel, welcher die Punkte P, Q, R enthält.

Weiter untersuchen wir den Differentialquotienten der Funktion $x(\lambda)$

(2.22) $$\dot{x}(\lambda)=\frac{\lambda^2(-b-2k_1a)+2\lambda k_2b+2k_1k_2^2a+4k_1k_2b-k_2^2b}{(-k_2^2+2\lambda k_2-4\lambda k_1-\lambda^2)^2}.$$

Aus (2.22) kann man sehen: $x(\lambda)$ hat zwei lokale Extremwerte, wenn der Ausdruck:

(2.23) $$S\equiv k_1^2k_2^2a^2+2k_1^2k_2ab+k_1k_2b^2$$

positiv ist, und es gibt keinen Extremwert, wenn (2.23) Null oder negativ ist. In diesem letzteren Fall ist $x(\lambda)$ eine strikt monotone Funktion in den Intervallen (λ_1,λ_2) bzw. $(-\infty,\lambda_1)\cup(\lambda_2,\infty)$.

Wir bilden

(2.24) $$S^*=k_1k_2,$$

(2.25) $$S^{**}=k_1k_2a^2+2k_1ab+b^2,\ \text{d.h. } S=S^*S^{**}.$$

Wenn $S=0$, d.h. $S^*=0$ oder $S^{**}=0$ gilt $(k_1\neq0)$, dann ist die Mittelpunktskurve ein schneidendes Geradenpaar $(\mathbb{I}_A1;\mathbb{I}_B1,2)$.

Im folgenden sind zwei Hauptfälle mit den entsprechenden Unterfällen zu diskutieren:

Hauptfall 1: $S>0$, Hauptfall 2: $S<0$.

Zuerst untersuchen wir den Hauptfall 1. In diesem Fall hat $x(\lambda)$ zwei Extremwerte und die Mittelpunktskurve ist eine Hyperbel 1. Art. Man kann leicht sehen, daß die Mittelpunktshyperbel die Halbierungspunkte A_1, B_1, D_1, C_1, E_1, G_1 der Grundstrecken enthält (Fig. 1.). Aus (2.2), (2.11) und (2.21) kann man die Koordinaten bzw. die λ-Werten berechnen, die zu den Halbierungspunkten gehören:

$$A_1\left(0,\frac{b}{2}\right),\qquad \lambda_{A_1}=\frac{k_2b}{b+2k_1a}, \tag{2.26}$$

$$B_1\left(\frac{a}{2},0\right),\qquad \lambda_{B_1}=\frac{2b+k_2a}{a}, \tag{2.27}$$

$$D_1\left(\frac{a}{2},\frac{b}{2}\right),\qquad \lambda_{D_1}=k_2, \tag{2.28}$$

$$C_1\left(\frac{2k_1a+b-k_2a}{2(k_1-k_2)},\frac{k_1k_2a+k_1b}{2(k_1-k_2)}\right),\qquad \lambda_{C_1}=\frac{k_2b+ak_2^2}{ak_2-2ak_1-b}, \tag{2.29}$$

$$G_1\left(\frac{b+k_1a}{2(k_1-k_2)},\frac{k_1b+k_1k_2a}{2(k_1-k_2)}\right),\qquad \lambda_{G_1}=-k_2, \tag{2.30}$$

$$E_1\left(\frac{b+k_1a}{2(k_1-k_2)},\frac{k_1k_2a+2k_1b-k_2b}{2(k_1-k_2)}\right),\qquad \lambda_{E_1}=\frac{k_2b-2k_1b-ak_1k_2}{k_1a+b}. \tag{2.31}$$

Hieraus ergibt sich nach einiger Rechnung, daß der Punkt A_1 der Mittelpunkt einer Hyperbel 2. Art und der Punkt C_1 der Mittelpunkt einer Ellipse des Kegelschnittbüschels ist. Man zeigt leicht: Die Punkte B_1, E_1 und ebenso G_1, D_1 gehören zu verschiedenen Zweigen der Mittelpunktshyperbel. Es gilt genauer: Ist $S^*>0$ und damit wegen $S>0$ $S^{**}>0$, dann ist B_1 der Mittelpunkt einer Hyperbel 1. Art und damit ist E_1 der Mittelpunkt einer Ellipse; gilt andererseits $S^*<0$ und damit $S^{**}<0$, dann ist B_1 der Mittelpunkt einer Ellipse und E_1 ist der Mittelpunkt einer Hyperbel 1. Art. Es sei zum Beispiel $S^*>0$. In diesem Fall gehört G_1 zu einer Ellipse, aber D_1 kann der Mittelpunkt einer Hyperbel 1. Art, eines schneidenden Geradenpaares oder einer Hyperbel 2. Art sein, je nachdem der Ausdruck

$$S_1\equiv k_1k_2-\frac{b^2}{a^2} \tag{2.32}$$

positiv, 0, oder negativ ist. Wenn $S^*<0$ gilt, dann kann man unschwer sehen, daß D_1 der Mittelpunkt einer Ellipse ist und G_1 der Mittelpunkt einer Hyperbel 1. Art, eines schneidenden Geradenpaares oder einer Hyperbel 2. Art sein kann, je nachdem der Ausdruck

$$S_2\equiv k_1k_2a^2+2k_2ab+b^2 \tag{2.33}$$

positiv, 0, oder negativ ist.

Durch Untersuchung der Lage der 6 Halbierungspunkte der 6 Grundstrecken, gelangt man zu drei neuen Typen des Hauptfalls 1 $S>0$ und zwar zu den Unterfällen:

$\mathbb{I}_A2$. $S^*>0$, $S_1>0$ oder $S^*<0$, $S_1<0$.

Eine Diagonale des Grundvierecks schneidet den hyperbolischen Zweig der Mittelpunktshyperbel in 2 Punkten, wobei der Halbierungspunkt zu einer Hyperbel 1. Art gehört.

$\mathbb{I}_A3$. $S^*>0$, $S_1=0$ oder $S^*<0$, $S_2=0$.

Eine Diagonale des Grundvierecks berührt den hyperbolischen Zweig der Mittelpunktshyperbel. Der Berührungspunkt — der Halbierungspunkt der Diagonale — ist der Mittelpunkt eines reellen, schneidenden Geradenpaares des Büschels ($Q\equiv G_1$ oder $Q\equiv D_1$).

$\mathbb{I}_A4$. $S^*>0$, $S_1<0$ oder $S^*<0$, $S_2>0$.

Eine Diagonale des Grundvierecks schneidet den hyperbolischen Zweig der Mittelpunktshyperbel in 2 Punkten, wobei der Halbierungspunkt zu einer Hyperbel 2. Art gehört.

SATZ 2.1: *Ein Kegelschnittbüschel der isotropen Ebene vom Typ* $\mathbb{I}_A2$, *3, 4 ist bis auf isotrope Bewegungen durch vier Invarianten* a, b, k_1, k_2 *eindeutig bestimmt. In allen drei Fällen ist die Mittelpunktskurve eine Hyperbel 1. Art. Vier Grundpunkte mit den isotropen Invarianten* a, b, k_1, k_2, *wobei die Bedingungen* (2.12), (2.13) *und* $S>0$ *gelten, bestimmen ein Kegelschnittbüschel vom Typ*

$\mathbb{I}_A2$ *wenn* $S^*>0$, $S_1>0$ *oder* $S^*<0$, $S_2<0$,

$\mathbb{I}_A3$ *wenn* $S^*>0$, $S_1=0$ *oder* $S^*<0$, $S_2=0$,

$\mathbb{I}_A4$ *wenn* $S^*>0$, $S_1<0$ *oder* $S^*<0$, $S_2>0$ *gilt.*

BEMERKUNGEN: 1. Aus Satz 2.1 kann man sofort sehen, daß es wegen $S_1=0$ bzw. $S_2=0$ beim Typ $\mathbb{I}_A3$ drei unabhängige Invarianten gibt.

2. Man kann unschwer sehen, daß die algebraischen Bedingungen bei den Typen $\mathbb{I}_A2$, 3, 4 eine gleichwertige geometrische Bedeutung besitzen.

Wir untersuchen den Fall $\mathbb{I}_A3$. Die Mittelpunktskurve berührt die Diagonale, wenn zum Beispiel

$$S^*>0, \quad S_1=0 \quad \text{gilt.} \tag{2.34}$$

An Fig. 3. erkennt man, daß PAC ein zulässiges Dreieck der isotropen Ebene ist. Bekanntlich ist die Summe der orientierten Winkel eines Dreiecks gleich Null (vgl. [1]), d.h. im Dreieck PAC gilt

$$-k_2+k_1-k_3=0 \tag{2.35}$$

Es seien

$$S^*<0, \qquad S_2=0. \tag{2.36}$$

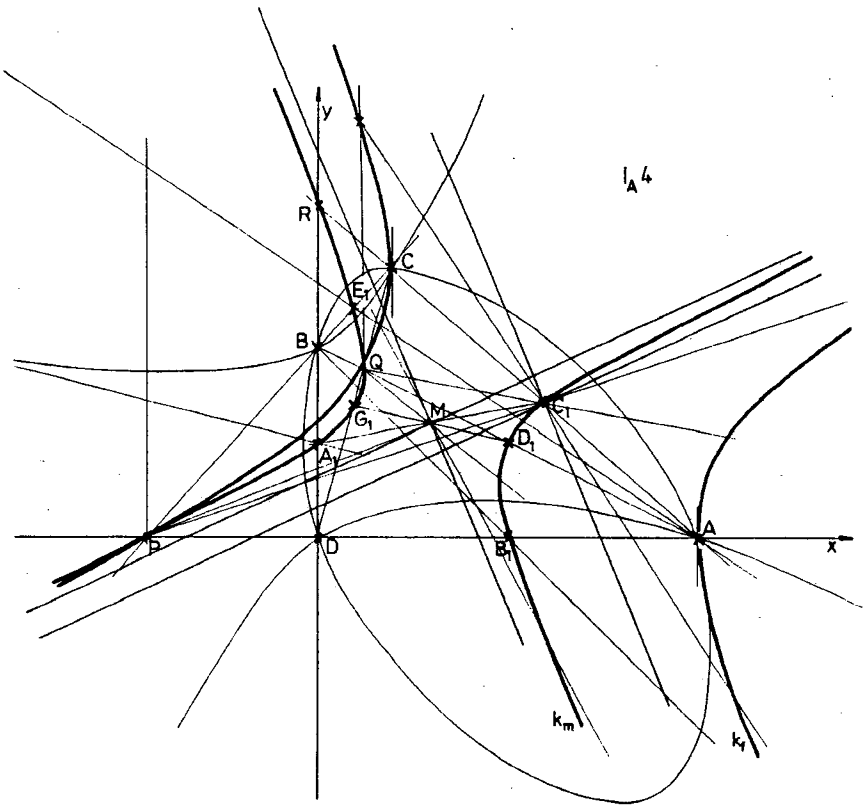

Fig. 2.

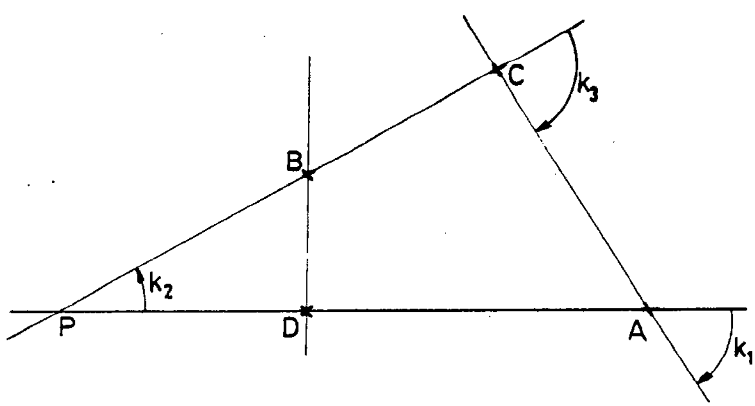

Fig. 3.

Ersichtlich gelten die folgenden Gleichungen

$$\frac{(k_1^2-k_1k_2)}{(b+k_1a)^2}\left(b^2+2k_2ab+a^2k_1k_2\right)=0, \quad \text{d.h.} \tag{2.37}$$

$$\frac{(k_1-k_2)^2b^2}{(b+ak_1)^2}=-k_2(k_1-k_2). \tag{2.38}$$

Aus (2.2) und (2.11) berechnet man dann der Reihe nach

$$d(BC)=\frac{b+k_1a}{k_1-k_2}; \quad s(BD)=-b; \tag{2.39}$$

$$-k_2=\sphericalangle(BC,DA); \quad k_3=\sphericalangle(BC,CA)=k_1-k_2 \quad \text{aus (2.35).} \tag{2.40}$$

Aus diesen Beziehungen kann man sofort sehen

$$\left(\frac{s(BD)}{d(BC)}\right)^2=-k_2k_3, \quad \text{wobei } -k_2k_3>0 \text{ gilt.} \tag{2.41}$$

Es ist evident, daß die Bedingungen (2.41) und (2.38) gleichwertig sind.

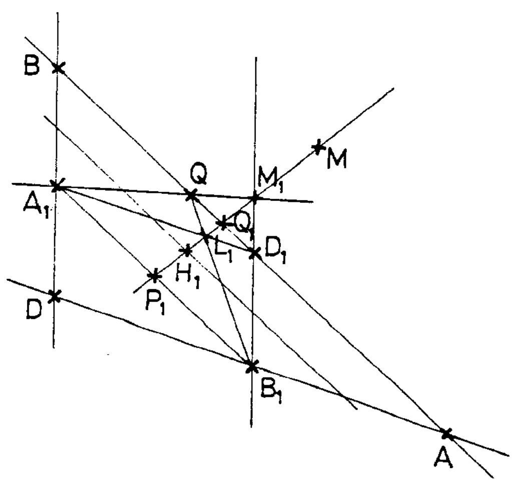

Fig. 4.

Zur geometrischen Erzeugung eines Büschels vom Typ $\mathbb{I}_A2$, 3, 4 sei BDA ein nicht zulässiges Dreieck, wobei B, D parallele Punkte der isotropen Ebene sind. Bezeichnen A_1, B_1, D_1 die Halbierungspunkte der Grundstrecken BD, DA und AB und ist Q ein Punkt der Strecke BD_1, dann bestimmen die Punkte A_1, B_1, D_1, Q ein Kegelschnittbüschel vom Typ $\mathbb{I}_A1$. Es seien Q_1, P_1 die Halbierungspunkte der Strecken QD_1 bzw.

A_1B_1 und sei L_1 der Schnittpunkt der Geraden QB_1 und A_1D_1. (Fig. 4.) Wie aus den Untersuchungen des Typs $\mathbb{I}_A1$ bekannt, liegen die Punkte P_1, L_1, Q_1 auf dem Hauptteil der Mittelpunktskurve des Kegelschnittbüschels und die Punkte A_1, B_1, D_1, Q sind eindeutig bestimmt. Die Halbgerade mit L_1 als Ursprung, welche Q_1 enthält, trägt die Mittelpunkte der Hyperbeln 1. Art. Diese Halbgerade schneidet die isotrope Gerade D_1B_1 im Punkt M_1. Es sei M ein Punkt auf der Halbgeraden mit dem Ursprung M_1, welche nicht den Punkt Q_1 enthält. Es ist klar, daß es eine Hyperbel 1. Art durch die Punkte $A_1B_1D_1Q$ gibt, welche M zum Mittelpunkt besitzt, wobei die Punkte A_1, B_1, D_1, Q auf demselben Zweig liegen. Diese Hyperbel 1. Art — bezeichnen wir sie mit k_m — schneidet die isotrope Gerade BD im Punkt R. Es sei C der Schnittpunkt der Geraden RA und DK_1. Man kann hierbei M stets so wählen, daß der Schnittpunkt C existiert. Aus der Konstruktion kann man leicht erkennen, daß die Punkte B, D, A, C gerade ein Kegelschnittbüschel vom Typ $\mathbb{I}_A2$ bestimmen, wobei die Mittelpunktskurve K_m ist.

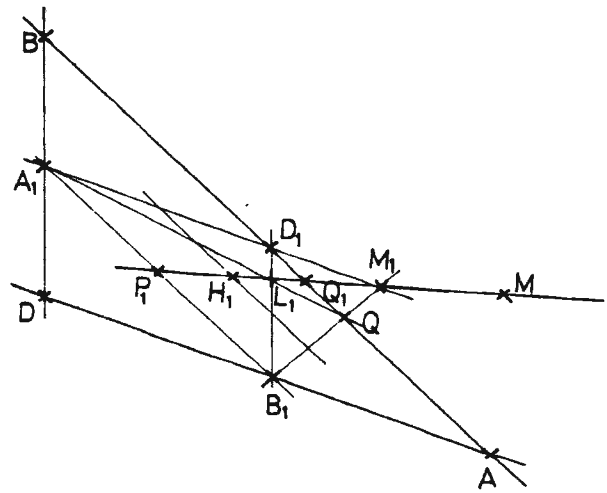

Fig. 5.

Ähnlich kann man die Grundpunkte B, D, A, C des Typs $\mathbb{I}_A4$ konstruieren, wobei Q ein Punkt der Strecke D_1A ist. Die Punkte A_1, B_1, D_1, Q bestimmen ein Kegelschnittbüschel vom Typ $\mathbb{I}_B2$. (Fig. 5.) Wie beim Typ $\mathbb{I}_A2$ kann man den Punkt M auf dem Hauptteil der Mittelpunktskurve des vorliegenden Kegelschnittbüschels vom Typ $\mathbb{I}_B2$ so wählen, daß die Punkte A_1, B_1, D_1, Q eine Hyperbel 1. Art mit dem Mittelpunkt M bestimmen — wir bezeichnen diese mit k_m —, wobei die Punkte A_1, B_1, D_1 und Q auf ein und demselben Zweig von k_m liegen. Mittels k_m kann man den Punkt

C konstruieren. Man zeigt leicht durch Konstruktion, daß die Punkte B, D, A, C ein Kegelschnittbüschel vom Typ $\mathbb{I}_A 4$ bestimmen.

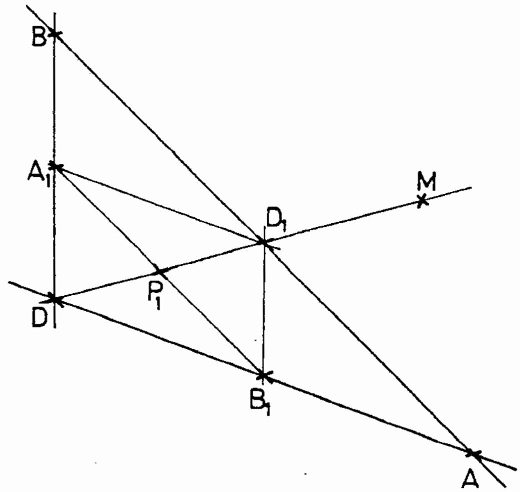

Fig. 6.

Zur geometrischen Erzeugung eines Büschels vom Typ $\mathbb{I}_A 3$ betrachten wir noch einmal das nicht zulässige Dreieck BDA mit den Halbierungspunkten A_1, B_1, D_1 der Seiten. (Fig. 6.) Man kann das Kegelschnittbüschel durch A_1, B_1, D_1 untersuchen, wobei die Gerade BA die gemeinsame Tangente aller Kegelschnitte mit D_1 als Berührungspunkt ist. Gemäß der Untersuchung des Kegelschnittbüschels gibt es eine Hyperbel 1. Art — wir bezeichnen sie mit k_m —, welche die Punkte A_1, B_1, D_1 enthält und die Gerade BA im Punkt D_1 berührt. Mittels k_m kann man den Punkt C konstruieren. Man zeigt wieder leicht durch Konstruktion, daß die Punkte B, D, A, C ein Kegelschnittbüschel vom Typ $\mathbb{I}_A 3$ bestimmen, wobei k_m die Mittelpunktshyperbel ist.

Im folgenden untersuchen wir den Hauptfall 2, d.h. es gilt $S < 0$. In diesem Fall hat $x(\lambda)$ keinen Extremwert und die Mittelpunktskurve ist eine Hyperbel 2. Art.

Analog wie im Hauptfall 1 untersuchen wir die Punkte A_1, B_1, C_1, D_1, E_1, G_1. Mittels (2.26), (2.29) kann man leicht sehen, daß der Punkt A_1 der Mittelpunkt einer Ellipse ist, und daß C_1 der Mittelpunkt einer Hyperbel 2. Art ist. Man zeigt unschwer mittels (2.27) und (2.31), daß die Punkte B_1, E_1 und ebenso mittels (2.30), (2.28) daß die Punkte G_1, D_1 zu verschiedenen Zweigen der Mittelpunktshyperbel gehören. Es gilt genauer: Ist $S^* > 0$ und damit wegen $S < 0$ $S^{**} < 0$, so ist B_1 der Mittelpunkt einer Ellipse und E_1 der Mittelpunkt einer Hyperbel 1. Art; gilt andererseits $S^* < 0$ und damit $S^{**} > 0$, so ist B_1 der Mittelpunkt einer Hyperbel 1. Art und E_1 der Mittelpunkt einer Ellipse. Man zeigt leicht, daß im Fall $S^* > 0$ der Punkt G_1 und im Fall $S^* < 0$ der Punkt D_1 der Mittelpunkt einer Ellipse im Büschel ist. Da der Punkt D_1 bzw. G_1 (der Halbierungspunkt der anderen Diagonale) zum hyperbolischen Zweig der Mittelpunktskurve gehört — welche so wie bei den Typen $\mathbb{I}_A 2$, 3, 4 — der Mittelpunkt einer Hyperbel 1. Art, eines schneidenden Geradenpaares oder einer Hyperbel 2. Art ist, je nachdem die Ausdrücke S_1, S_2 im (2.32) bzw. (2.33) positiv, Null oder negativ sind, unterscheiden wir demgemäß die Typen $\mathbb{I}_A 5$, $\mathbb{I}_A 6$, $\mathbb{I}_A 7$ des Kegelschnittbüschels.

So ergeben sich die Unterfälle:

$\mathbb{I}_A 5$. $S^* > 0$, $S_1 > 0$ oder $S^* < 0$, $S_2 < 0$

Eine Diagonale des Grundvierecks schneidet den hyperbolischen Zweig der Mittelpunktshyperbel in 2 Punkten, wobei der Halbierungspunkt zu einer Hyperbel 1. Art gehört.

$\mathbb{I}_A 6$. $S^* > 0$, $S_1 = 0$ oder $S^* < 0$, $S_2 = 0$

Eine Diagonale des Grundvierecks berührt den hyperbolischen Zweig der Mittelpunktshyperbel. Der Berührungspunkt — der Halbierungspunkt der Diagonale — ist der Mittelpunkt eines reellen schneidenden Geradenpaares des Büschels ($Q \equiv G_1$ oder $Q \equiv D_1$).

$\mathbb{I}_A 7$. $S^* > 0$, $S_1 < 0$ oder $S^* < 0$, $S_2 > 0$

Eine Diagonale des Grundvierecks schneidet den hyperbolischen Zweig der Mittelpunktshyperbel in 2 Punkten, wobei der Halbierungspunkt zu einer Hyperbel 2. Art gehört.

SATZ 2.2: *Ein Kegelschnittbüschel der isotropen Ebene vom Typ* $\mathbb{I}_A 5$, *6, 7 ist bis auf isotrope Bewegungen durch die vier Invarianten* a, b, k_1, k_2 *eindeutig bestimmt. In allen drei Fällen ist die Mittelpunktskurve eine Hyperbel 2. Art. Die vier Grundpunkte mit den isotropen Invarianten* a, b, k_1, k_2, *für die* (2.12), (2.13) *und* $S < 0$ *gelten, bestimmen ein Kegelschnittbüschel vom Typ*

$\mathbb{I}_A 5$ *wenn* $S^* > 0$, $S_1 > 0$ *oder* $S^* < 0$, $S_2 < 0$,

$\mathbb{I}_A 6$ *wenn* $S^* > 0$, $S_1 = 0$ *oder* $S^* < 0$, $S_2 = 0$,

$\mathbb{I}_A 7$ *wenn* $S^* > 0$, $S_1 < 0$ *oder* $S^* < 0$, $S_2 > 0$ *gilt.* (Fig. 7.)

BEMERKUNGEN: 1. Aus Satz 2.2 kann man sofort sehen, daß es beim Typ $\mathbb{I}_A 6$ drei unabhängige Invarinten gibt; dies folgt aus $S_1 = 0$ bzw. $S_2 = 0$.

2. Analog wie bei den Typen $\mathbb{I}_A 2$, 3, 4 kann man sehen, daß die algebraischen Bedingungen bei den Typen $\mathbb{I}_A 5$, 6, 7 eine gleichwertige geometrische Bedeutung besitzen.

Zur geometrischen Erzeugung eines Büschels vom Typ 5, 6, 7 betrachten wir wieder das nicht zulässige Dreieck BDA mit den Halbierungs-

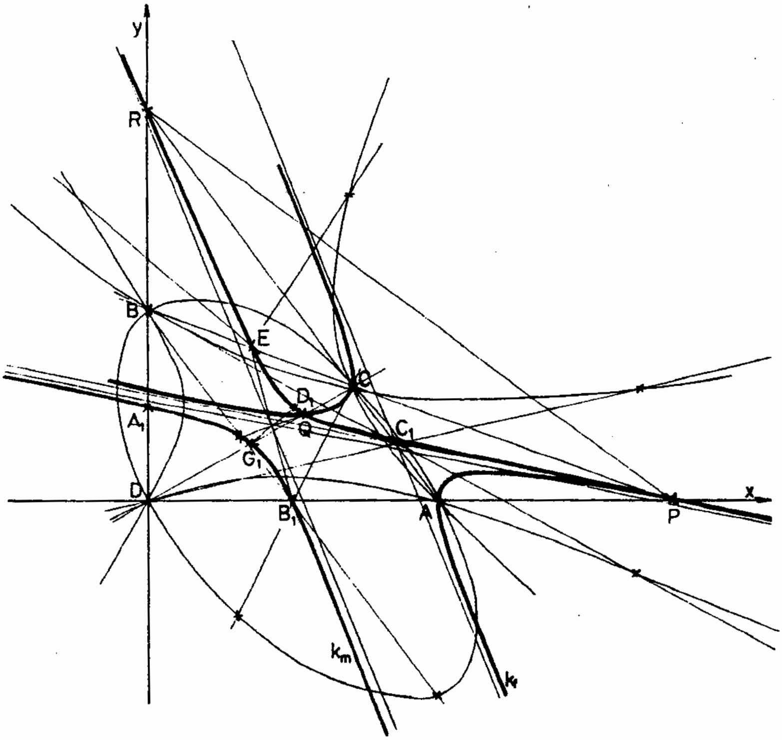

Fig. 7.

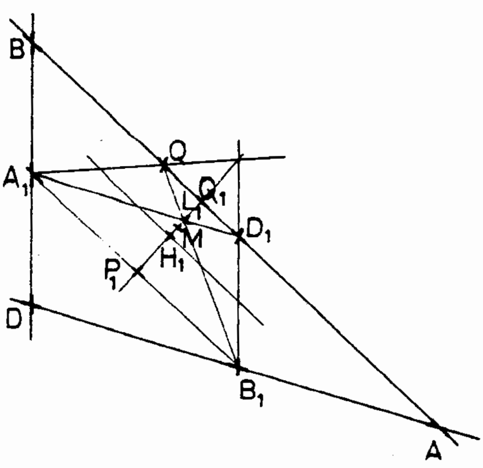

Fig. 8.

punkten A_1, B_1, D_1 der Seiten. Sei Q ein Punkt der Strecke BD_1. (Fig. 8.) Wie beim Typ II_A2 kann man sehen, daß die Punkte A_1, B_1, D_1, Q ein Kegelschnittbüschel vom Typ II_A1 bestimmen. Wir benützen die Fig. 4. Sei H_1 der Mittelpunkt der Mittelpunktskurve des Kegelschnittbüschels $A_1B_1D_1Q$. Wie aus der Untersuchung des Typs II_A1 bekannt, enthält die Strecke L_1H_1 die Mittelpunkte der durch die Punkte A_1, B_1, D_1, Q legbaren Hyperbeln 2. Art; sei M ein Punkt der Strecke L_1H_1. Wir bezeichnen mit k_m die Hyperbel 2. Art, bestimmt durch die Punkte A_1, B_1, D_1, Q mit dem Mittelpunkt M. Wie beim Typ II_A2 kann man mittels k_m und den Punkten B, D, A den Punkt C bestimmen. Aus der Konstruktion ergibt sich, daß die Punkte B, D, A, C ein Kegelschnittbüschel vom Typ II_A5 bestimmen, wobei die Mittelpunktskurve k_m ist.

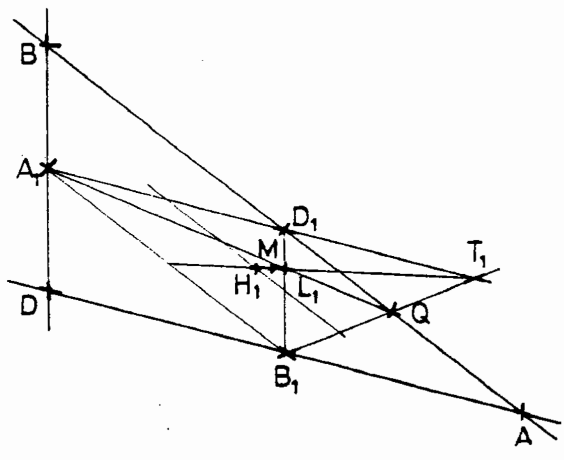

Fig. 9.

Ähnlich kann man die Grundpunkte B, D, A, C beim Typ II_A7 konstruieren; der Punkt Q liegt aber auf der Strecke D_1A_1 (Fig. 9.). Wie beim Typ II_A4 kann man sehen, daß die Punkte A_1, B_1, D_1, Q ein Kegelschnittbüschel vom Typ II_B2 bestimmen.

Es sei H_1 der Mittelpunkt der Mittelpunktskurve des Kegelschnittbüschels $A_1B_1D_1Q$ und T_1 der Schnittpunkt der Geraden A_1D_1 und B_1Q. Wie beim Typ II_A4 kann man den Punkt M ($M \in H_1L_1$) so wählen, daß die Punkte A_1, B_1, D_1, Q eine Hyperbel 2. Art mit dem Mittelpunkt M bestimmen — wir bezeichnen sie mit k_m —, wobei die Punkte D_1, K_1 bzw. die Punkte A_1, B_1 auf demselben Zweig von k_m liegen. Mittels k_m kann man den Punkt C konstruieren. Aus der Konstruktion ist ersichtlich, daß die Punkte B, D, A, C ein Kegelschnittbüschel vom Typ $\mathbb{I}_A7$ bestimmen, wobei k_m die Mittelpunktskurve ist.

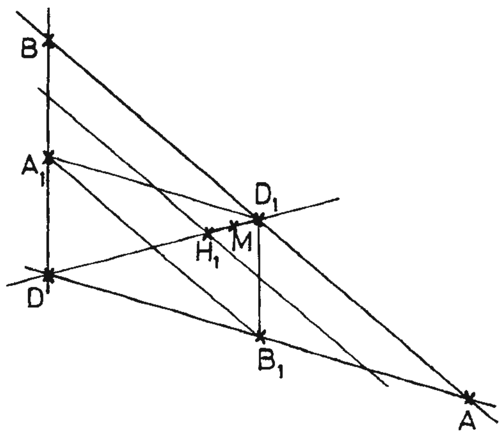

Fig. 10.

Zur geometrischen Erzeugung eines Büschels vom Typ $\mathbb{I}_A6$ betrachten wir noch einmal das nicht zulässige Dreieck BDA mit den Halbierungspunkten A_1, B_1, D_1 der Seiten. (Fig. 10.) Wir untersuchen das Kegelschnittbüschel durch A_1, B_1, D_1, wobei die Gerade BA die gemeinsamen Tangente aller Kegelschnitte mit dem Berührungspunkt D_1 ist.

Aus der Untersuchung des vorliegenden Kegelschnittbüschels kann man sehen, daß es eine Hyperbel 2. Art gibt (ihr Mittelpunkt liegt auf H_1D_1), welche die Punkte A_1, B_1, D_1 enthält und die Gerade BA im Punkt D_1 berührt; wir bezeichnen sie mit k_m. Mittels k_m kann man den Punkt C konstruieren. Die Punkte B, D, A, C bestimmen ein Kegelschnittbüschel vom Typ $\mathbb{I}_A6$, wobei k_m die Mittelpunktskurve ist.

Wir untersuchen noch die isotropen Brennpunktskurven k_f der Büscheltypen $\mathbb{I}_A2, 3, 4, 5, 6, 7$. Aus (2.9) und $\frac{\partial F}{\partial y}=0$ findet man als Gleichung von k_f

(2.42) $-k_1k_2x^2-y^2+2k_1xy+(k_1k_2a-bk_1)x-2k_1ay+bk_1a=0$ für $\lambda\in\mathbb{R}$ und
(2.43) $x=0$ für $\lambda=\infty$.

Damit ist die Brennpunktskurve k_f eine Kurve 3. Ordnung

$$(2.44)\qquad \left[-k_1k_2x^2-y^2+2k_1xy+(k_1k_2a-bk_1)x-2k_1ay+bk_1a\right]x=0.$$

Sie zerfällt in 2 Teile, wobei der Hauptteil eine Hyperbel 1. Art (2.42) ist und der andere Teil die isotrope Grundgerade (2.43) ist. Nach einiger Rechnung ergeben sich die isotropen Brennpunkte A und C; diese liegen nicht auf der isotropen Grundgeraden. Man zeigt leicht, daß die Asymptoten des Hauptteils der Brennpunktskurve (2.42) durch

$$(2.45)\quad y=\left(k_1+\sqrt{k_1^2-k_1k_2}\right)\left(x-\frac{b+2k_1a-k_2a}{2(k_1-k_2)}\right)+\frac{k_1b+k_2k_1a}{2(k_1-k_2)} \quad \text{bzw.}$$

$$(2.46)\quad y=\left(k_1-\sqrt{k_1^2-k_1k_2}\right)\left(x-\frac{b+2k_1a-k_2a}{2(k_1-k_2)}\right)+\frac{k_1b+k_2k_1a}{2(k_1-k_2)}$$

gegeben sind.

Man kann sofort sehen, daß diese Geraden zu den Asymptoten der Mittelpunktshyperbel parallel sind und mit den Durchmesserrichtungen der Büschelparabeln übereinstimmen. Nach einiger Rechnung zeigt man, daß die Punkte P, Q, R auf der Brennpunktskurve liegen.

SATZ 2.3: *Die Brennpunktskurve k_f eines Kegelschnittbüschels vom Typ* $\mathbb{I}_A 2, 3, 4, 5, 6, 7$ *ist eine Kurve 3. Ordnung, welche in 2 Teile zerfällt; nämlich in eine Hyperbel 1. Art bzw. in die isotrope Grundgerade. Sie enthält die Grundpunkte, sowie die Diagonalecken P, Q und R, wobei A und C die isotropen Brennpunkte der Brennpunktshyperbel sind. Sie schneidet die absolute Gerade in den gemeinsamen Punkten mit der Mittelpunktshyperbel und den Parabeln des Kegelschnittbüschels.*

3. Zusammenfassung

In dieser Arbeit untersuchen wir jene Kegelschnittbüschel der isotropen Ebene, welche vier reelle und verschiedene Grundpunkte bezitzen, wobei mindestens 2 Grundpunkte parallel sind, d.h. auf einer isotropen Geraden liegen. Mit analytischen Methoden wird eine vollständige Klassifikation aller Büscheltypen gegeben, wobei jeder Typ durch ein vollständiges Invariantensystem beschrieben wird. Die auftretenden algebraischen Beziehungen werden in der isotropen Ebene geometrisch gedeutet.

Literaturverzeichnis

[1] SACHS, H: *Ebene Isotrope Geometrie*, Vieweg-Verlag, Braunschweig/Wiesbaden, 1987.

[2] SACHS, H: Oskulierende und hyperoskulierende Kegelschnittbüschel der isotropen Ebene I, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, **196** (1987), 337–375.

[3] SACHS, H: Oskulierende und hyperoskulierende Kegelschnittbüschel der isotropen Ebene II, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, (im Druck).

[4] ŠČURIČ, VL.: Kegelschnittbüschel der isotropen Ebene I., *RAD JAZU* **450** (1990), 41–51.

[5] ŠČURIČ, VL. und SACHS, H: Klassifikationstheorie der Kegelschnittbüschel vom Typ IV. in der isotropen Ebene, *RAD JAZU*, (1991, im Druck).

[6] COXETER H. S. M.: *Projektív geometria*, Gondolat, Budapest, 1986.

[7] JAGLOM, I. M.: Galilei relativitási elve és egy nemeuklideszi geometria, Gondolat, Budapest, 1985.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 59–67

ON THE CONJUGATE FUNCTION OF DIRICHLET SERIES

By

I. JOÓ

Department of Analysis, L. Eötvös University, Budapest

(Received July 31, 1991)

Consider a sequence of numbers

$$\lambda_n \in \mathbb{C}, \quad \lambda_n = \varrho_n + i\nu_n, \quad 0 < \nu_n \le H \quad (n \in \mathbb{N})$$

and suppose that the real parts ϱ_n increase. Suppose further that the system

$$e(\Lambda) := \{e^{i\lambda_n x} : n \in \mathbb{Z}\}$$

forms a Riesz basis in $L^2(0,2\pi)$ (see [2]). Denote $v_n^* \in L^2(0,2\pi)$ the biorthogonal functions, i.e.

$$\langle e^{i\lambda_n x}, v_k^* \rangle_{L^2} = \delta_{n,k}.$$

The expansion of a function $f \in L^1 = L^1(0,2\pi)$ is given by

$$f \sim \sum \langle f, v_n^* \rangle e^{i\lambda_n x}.$$

In what follows we suppose that

(1) $$\|v_n^*\|_{L^\infty} \le c \quad (n \in \mathbb{Z}).$$

It is the case when the generating function $F(z)$ of the system $e(\Lambda)$ is of sine type with separated zeros and can be given in the form

(2) $$F(z) = \int_0^{2\pi} e^{izt} \mathrm{d}\sigma(t), \quad \sigma \in BV[0,2\pi]$$

(the function σ has bounded variation on $[0,2\pi]$). Indeed, we can ensure that

$$\delta_{n,k} = \langle v_n^*, e^{i\lambda_k x} \rangle = \sqrt{2\pi}\mathcal{F}(v_n^*)(\overline{\lambda}_k)$$

($\mathcal{F}$ denotes the Fourier transform) if we set

$$\mathcal{F}(v_n^*) := \frac{1}{\sqrt{2\pi}} \frac{G(z)}{z - \overline{\lambda}_n} \cdot \frac{1}{G'(\overline{\lambda}_n)}, \qquad G(z) := \overline{F(\overline{z})} = \mathcal{F}(d\sigma).$$

Here $\frac{F(x)}{x-\lambda_n} \in L^2(\mathbb{R})$, hence $v_n^* \in L^2(0,2\pi)$ and

$$v_n^* = \frac{1}{\sqrt{2\pi}}\frac{1}{G'(\overline{\lambda}_n)}\mathcal{F}^{-1}\left(\mathcal{F}(d\sigma)\cdot\frac{1}{z-\overline{\lambda}_n}\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{G'(\overline{\lambda}_n)}\mathcal{F}^{-1}\left(\frac{1}{z-\overline{\lambda}_n}\right) * d\sigma.$$

We know that

$$\mathcal{F}^{-1}\left(\frac{1}{z-\overline{\lambda}_n}\right) = \sqrt{2\pi} i e^{i\overline{\lambda}_n t}\Big|_{(0,\infty)},$$

hence

$$\left\|\mathcal{F}^{-1}\left(\frac{1}{z-\overline{\lambda}_n}\right) * d\sigma\right\|_\infty \le c$$

uniformly in n. On the other hand if the zeros of a sine type function F are separated then $|F'(\lambda_n)| \ge c > 0$ see ([4], [5]), which shows (1).

DEFINITION. Let $f \in L^1$, $f \sim \sum \langle f, v_n^*\rangle e^{i\lambda_n x}$. The series

$$-i \sum \operatorname{sgn} \varrho_k \langle f, v_k^*\rangle e^{i\lambda_k x}$$

is called the Dirichlet conjugate series, its partial sums

$$\tilde{S}_\mu(f,x) = -i \sum_{|\varrho_k|<\mu} \operatorname{sgn} \varrho_k \langle f, v_k^*\rangle e^{i\lambda_k x}$$

are called the conjugate partial sums. If there exists $\tilde{f} \in L^1$ with $\tilde{f} \sim$ $\sim -i\sum \operatorname{sgn} \varrho_k \langle f, v_k^*\rangle e^{i\lambda_k x}$ then $\tilde{f}$ is the Dirichlet conjugate function of f. Recall the formula

(3)

$$\tilde{S}_n^T(f,x) = \frac{2}{\pi}\int_0^\pi \psi_x(t)\frac{\cos\frac{t}{2} - \cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}\mathrm{d}t, \quad \psi_x(t) = \frac{f(x-t)-f(x+t)}{2}$$

for the trigonometric conjugate partial sum, see [1], ch. II. 5. We use here the decomposition

$$\frac{\cos\frac{t}{2} - \cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}} = \frac{1-\cos nt}{t} + \frac{\sin nt}{2} + (1-\cos nt)\left(\frac{1}{2\operatorname{tg}\frac{t}{2}} - \frac{1}{t}\right)$$

and the fact that

$$\frac{1}{2\operatorname{tg}\frac{t}{2}} - \frac{1}{t} = O(t) \qquad \text{on} \quad [0,\pi].$$

Fix a number $0 < R < \pi$. Then

(4) $$\tilde{S}_n^T(f,x)=\frac{2}{\pi}\int_0^R \psi_x(t)\frac{1-\cos nt}{t}\mathrm{d}t+f_R(x)+g_{n,R}(x)$$

where

(5) $$f_R(x)=\frac{2}{\pi}\int_R^\pi \psi_x(t)\frac{1}{t}\mathrm{d}t+\frac{2}{\pi}\int_0^\pi \psi_x(t)\left(\frac{1}{2\operatorname{tg}\frac{t}{2}}-\frac{1}{t}\right)\mathrm{d}t$$

is a continuous function independent of n and

(6) $$g_{n,R}(x)=\frac{1}{\pi}\int_0^\pi \psi_x(t)\sin nt\mathrm{d}t-\frac{2}{\pi}\int_R^\pi \frac{\psi_x(t)}{t}\cos nt\mathrm{d}t-$$
$$-\frac{2}{\pi}\int_0^\pi \psi_x(t)\left(\frac{1}{2\operatorname{tg}\frac{t}{2}}-\frac{1}{t}\right)\cos nt\mathrm{d}t$$

which tends to zero uniformly in x since these are Fourier coefficients of functions which form a precompact family of L^1. We prove the following

THEOREM 1.
$$\left|\tilde{S}_\mu(f,x)-\tilde{S}_{[\mu]}^T(f,x)\right|\le c\|f\|_1,\quad f\in L^1,\quad x\in[0,2\pi],$$
where the constant c is independent of f, μ and x and $[\mu]$ denotes the largest integer $l\le\mu$.

PROOF. From (4)–(6) we see that
$$\left|\tilde{S}_n^T(f,x)-\frac{2}{\pi}\int_0^R \psi_x(t)\frac{1-\cos nt}{t}\mathrm{d}t\right|\le\frac{c}{R}\|f\|_1,$$
where c is independent also of R. We can write
$$\frac{2}{\pi}\int_0^R \psi_x(t)\frac{1-\cos nt}{t}\mathrm{d}t=\langle f,\tilde{w}_{R,n}\rangle,$$
$$\tilde{w}_{R,n}(x+t):=\begin{cases}-\frac{1-\cos nt}{\pi t} & \text{if } |t|<R\\ 0 & \text{if } |t|\ge R.\end{cases}$$
Let further
$$\tilde{w}_R^\mu(x+t):=\begin{cases}-\frac{1-\cos \mu t}{\pi t} & \text{if } |t|<R\\ 0 & \text{if } |t|\ge R,\end{cases}$$

then

$$\left|\langle f,\tilde{w}_R^\mu\rangle-\langle f,\tilde{w}_{R,[\mu]}\rangle\right|\le\int\limits_0^R|f(x-t)-f(x+t)|\,\frac{|\cos\mu t-\cos[\mu]t|}{\pi t}\mathrm{d}t=$$

$$=\int\limits_0^R|f(x-t)-f(x+t)|2\cdot\left|\frac{\sin(\mu-[\mu])}{\pi t}\sin(\mu+[\mu])\frac{t}{2}\right|\mathrm{d}t\le c\|f\|_1.$$

Finally fix a number $0<R_0<\pi$ and let

$$\tilde{w}=S_{R_0}(\tilde{w}_R^\mu),\quad S_{R_0}(g(R)):=\frac{2}{R_0}\int\limits_{R_0/2}^{R_0}g(R)\mathrm{d}R.$$

Since $\tilde{w}=\tilde{w}_R^\mu$ for $|t|\le R/2$, hence $|\tilde{w}-\tilde{w}_R^\mu|\le\frac{c}{R_0}$, so

$$|\langle f,\tilde{w}\rangle-\langle f,\tilde{w}_R^\mu\rangle|\le c\frac{\|f\|_1}{R_0}.$$

Consequently the statement of Theorem 1 follows from the estimate

$$|\langle f,\tilde{w}\rangle-\tilde{S}_\mu(f,x)|\le c\frac{\|f\|_1}{R_0};\qquad f\in L^1,\quad x\in[0,2\pi],\tag{7}$$

where the constant c is independent of f, x and R_0. To get (7) we expand the conjugate kernel $\tilde{w}$ in the Riesz basis (v_n^*). The coefficients are

$$\langle\tilde{w},e^{i\lambda_n t}\rangle=S_{R_0}\left[\langle\tilde{w}_R^\mu,e^{i\lambda_n t}\rangle\right]=$$

$$=S_{R_0}\left[\int\limits_0^R\left(e^{-i\overline{\lambda}_n(x-t)}-e^{-i\overline{\lambda}_n(x+t)}\right)\frac{1-\cos\mu t}{\pi t}\mathrm{d}t\right]=$$

$$=\frac{2i}{\pi}e^{-i\overline{\lambda}_n x}S_{R_0}\left[\int\limits_0^R\sin\overline{\lambda}_n t\frac{1-\cos\mu t}{t}\mathrm{d}t\right]=$$

$$=\frac{2i}{\pi}e^{-i\overline{\lambda}_n x}S_{R_0}\left[\int\limits_0^R(\sin\overline{\lambda}_n t-\sin\varrho_n t)\frac{1-\cos\mu t}{t}\mathrm{d}t\right]+$$

$$+\int\limits_0^\infty\sin\varrho_n t\frac{1-\cos\mu t}{t}\mathrm{d}t-\int\limits_R^\infty\sin\varrho_n t\frac{1-\cos\mu t}{t}\mathrm{d}t=$$

$$=:\frac{2i}{\pi}e^{-i\overline{\lambda}_n x}S_{R_0}[I_1+I_2+I_3].$$

Using the fact that

$$\int_0^\infty \frac{\sin \alpha t}{t} \mathrm{d}t = \frac{\pi}{2} \operatorname{sgn} \alpha, \quad \alpha \in \mathbb{R}$$

we get that

$$I_2 = \frac{\pi}{2} \operatorname{sgn} \varrho_n \cdot \delta(\mu, |\varrho_n|), \quad \delta(\mu, |\varrho_n|) := \begin{cases} 0 & \text{if } |\varrho_n| > \mu \\ \frac{1}{2} & \text{if } |\varrho_n| = \mu \\ 1 & \text{if } |\varrho_n| < \mu. \end{cases}$$

We assert that

$$|S_{R_0} I_3| \le \frac{c}{1 + (\mu - \varrho_n)^2 R_0^2}. \tag{8}$$

Indeed, let $\alpha > 0$ be arbitrary; then

$$\left| \int_R^\infty \frac{\sin \alpha t}{t} \mathrm{d}t \right| = \left| \int_{R\alpha}^\infty \frac{\sin \xi}{\xi} \mathrm{d}\xi \right| \le \frac{c}{1 + R\alpha},$$

on the other hand

$$\int_R^\infty \frac{\sin \alpha t}{t} \mathrm{d}t = -\left[\frac{\cos \alpha t}{\alpha t} \right]_R^\infty - \int_R^\infty \frac{\cos \alpha t}{\alpha t^2} =$$

$$= \frac{\cos \alpha R}{\alpha R} - \left[\frac{\sin \alpha t}{\alpha^2 t^2} \right]_R^\infty + 2 \int_R^\infty \frac{\sin \alpha t}{\alpha^2 t^3} \mathrm{d}t = \frac{\cos \alpha R}{\alpha R} + O(\alpha^2 R_0^2),$$

$$S_{R_0} \left[\int_R^\infty \frac{\sin \alpha t}{t} \mathrm{d}t \right] = O(\alpha^2 R_0^2) + \frac{2}{R_0} \int_{\frac{R_0}{2}}^{R_0} \frac{\cos \alpha R}{\alpha R} \mathrm{d}R =$$

$$= O(\alpha^2 R_0^2) + \frac{2}{R_0} \left[\frac{\sin \alpha R}{\alpha^2 R} \right]_{\frac{R_0}{2}}^{R_0} + \frac{2}{R_0} \int_{\frac{R_0}{2}}^{R_0} \frac{\cos \alpha R}{\alpha R} \mathrm{d}R = O(\alpha^2 R_0^2)$$

which proves (8). Next we estimate I_1. Take the decomposition

$$I_1 = \int_0^R \frac{\operatorname{ch} \nu_n t - 1}{t} \sin \varrho_n t (1 - \cos \mu t) \mathrm{d}t - i \int_0^R \frac{\sin \nu_n t}{t} \cos \varrho_n t (1 - \cos \mu t) \mathrm{d}t =$$

$$=: I_{11} + I_{12}.$$

Taking into account that in the Taylor series of $(\operatorname{ch}\nu_n t-1)/t$ and $\operatorname{sh}\nu_n t/t$ all coefficients are nonnegative, we get

$$|I_{11}|\le 2\int_0^R \frac{\operatorname{ch}\nu_n t-1}{t}\mathrm{d}t\le 2(\operatorname{ch}\nu_n R-1),$$

$$|I_{12}|\le 2\int_0^R \frac{\operatorname{sh}\nu_n t}{t}\mathrm{d}t\le 2\operatorname{sh}\nu_n R.$$

Integrating by parts we obtain with positive constants α, ν:

$$\int_0^R \frac{\operatorname{ch}\nu t-1}{t}\sin\alpha t\mathrm{d}t=-\left[\frac{\operatorname{ch}\nu t-1}{t}\cdot\frac{\cos\alpha t}{\alpha}\right]_0^R+\int_0^R\frac{\cos\alpha t}{\alpha}\left(\frac{\operatorname{ch}\nu t-1}{t}\right)'\mathrm{d}t=$$

$$=-\frac{\operatorname{ch}\nu R-1}{R}\cdot\frac{\cos\alpha R}{\alpha}+\left[\frac{\sin\alpha t}{\alpha^2}\left(\frac{\cos\nu t-1}{t}\right)'\right]_{t=0}^R-$$

$$-\int_0^R\frac{\sin\alpha t}{\alpha^2}\left(\frac{\operatorname{ch}\nu t-1}{t}\right)''\mathrm{d}t=$$

$$=-\frac{\operatorname{ch}\nu R-1}{R}\cdot\frac{\operatorname{ch}\alpha R}{\alpha}+\frac{\sin\alpha R}{\alpha^2}\left(\frac{\operatorname{ch}\nu R-1}{R}\right)'+O\left(\frac{1}{\alpha^2}\right)\int_0^R\left(\frac{\operatorname{ch}\nu t-1}{t}\right)''\mathrm{d}t$$

hence

$$S_{R_0}\left[\int_0^R\frac{\operatorname{ch}\nu t-1}{t}\sin\alpha t\mathrm{d}t\right]=-\frac{2}{R_0}\left[\frac{\sin\alpha R}{\alpha^2}\cdot\frac{\operatorname{ch}\nu R-1}{R}\right]_{\frac{R_0}{2}}^{R_0}+$$

$$+\frac{2}{R_0}\int_{R_0/2}^{R_0}\frac{\sin\alpha R}{\alpha^2}\left(\frac{\operatorname{ch}\nu R-1}{R}\right)'\mathrm{d}R+O\left(\frac{1}{\alpha^2}\right)\frac{2}{R_0}\left[\frac{\operatorname{ch}\nu R-1}{R}\right]_{\frac{R_0}{2}}^{R_0}+$$

$$+O\left(\frac{1}{\alpha^2}\right)\frac{2}{R_0}\left[\frac{\operatorname{ch}\nu R-1-\nu^2R^2/2}{R}\right]_{\frac{R_0}{2}}^{R_0},$$

consequently

$$\left|S_{R_0}\left[\int_0^R\frac{\sin\nu t-1}{t}\sin\alpha t\mathrm{d}t\right]\right|\le c\frac{e^{\nu R_0}}{1+\alpha^2R_0^2}. \tag{9}$$

Similar argument lead to the estimate

$$(10)\qquad \left|S_{R_0}\left[\int_0^R \frac{\sin\nu t}{t}\cos\alpha t\,dt\right]\right| \le c\frac{e^{\nu R_0}}{1+\alpha^2 R_0}.$$

Applying (9) to I_{11} and (10) to I_{12} we get

$$(11)\qquad |I_1| \le c\frac{e^{\nu_n R_0}}{1+(\mu-\varrho_n)^2 R_0^2}.$$

Summarize the above estimates in

$$\left|\left\langle \tilde{w}, e^{i\lambda_n t}\right\rangle - i\operatorname{sgn}\varrho_n \delta(\mu,|\varrho_n|)e^{-i\bar{\lambda}_n x}\right| \le c\frac{e^{\nu_n R_0}}{1+(\mu-\varrho_n)^2 R_0^2} \le c\frac{e^{H\pi}}{1+(\mu-\varrho_n)^2 R_0^2}.$$

Since there are finitely many ϱ_n with $|\varrho_n| \le \mu$, it implies that

$$\sum\left|\left\langle \tilde{w}, e^{i\lambda_n t}\right\rangle\right| < \infty \quad \forall x,$$

and hence the series

$$\tilde{w}(y) = \sum \langle \tilde{w}, e^{i\lambda_n t}\rangle v_n^*(y)$$

converging in L^2 by the Riesz basis property of (v_n^*), converge also uniformly by (1). But then for any $f \in L^1$

$$\langle f, \tilde{w}\rangle = \sum \langle f, v_n^*\rangle \overline{\langle \tilde{w}, e^{i\lambda_n t}\rangle} =$$

$$= -i\sum_{|\varrho_n|<\mu} \operatorname{sgn}\varrho_n e^{i\lambda_n x}\langle f, v_n^*\rangle - \frac{i}{2}\cdot\sum_{|\varrho_n|<\mu} \operatorname{sgn}\varrho_n e^{i\lambda_n x}\langle f, v_n^*\rangle +$$

$$+O\left(\sum |\langle f, v_n^*\rangle| \frac{1}{1+(\mu-\varrho_n)^2 R_0^2}\right)$$

and

$$\left|\langle f, \tilde{w}\rangle - \tilde{S}_\mu(f,x)\right| = O\left(\sum_{|\varrho_n|=\mu} |\langle f, v_n\rangle| + \sum \frac{|\langle f, v_n\rangle|}{1+(\mu-\varrho_n)^2 R_0^2}\right) \le$$

$$\le c\|f\|_1 \left(\sum_{|\varrho_n|=\mu} 1 + \sum \frac{1}{1+(\mu-\varrho_n)^2 R_0^2}\right).$$

By the separability of the (λ_n) we see that

$$\sum_{|\varrho_n|=\mu} 1 \le c$$

and

$$\sum \frac{1}{1+(\mu-\varrho_n)^2 R_0^2} \le \sum_{k=0}^{\infty} \frac{1}{1+k^2} \sum_{k\le|\mu-\varrho_n|R_0\le k+1} 1 \le \frac{c}{R_0}\sum_0^{\infty}\frac{1}{1+k^2} \le \frac{c}{R_0}$$

which proves Theorem 1. ∎

As a consequence we get the Riesz Marcell theorem for $p \le 2$:

THEOREM 2. *Let $1<p\le 2$; then for every $f \in L^p$ there exists $\tilde{f} \in L^p$ with*

$$\tilde{f} \sim -i \sum \operatorname{sgn} \varrho_k \langle f, v_k^* \rangle e^{i\lambda_k x}$$

and

$$\|\tilde{f}\|_p \le c(p)\|f\|_p, \quad f \in L^p.$$

PROOF. Introduce the operators

$$A_\mu : f \to \tilde{S}_\mu f - \tilde{S}_{[\mu]}^T f.$$

As we have seen in Theorem 1, $A_\mu : L^1 \to C[0,2\pi] \subset L^1$ are uniformly bounded and by the Riesz basis property $A_\mu : L^2 \to L^2$ are also uniformly bounded. We see by interpolation that $A_\mu : L^p \to L^p$ are uniformly bounded for $1 \le p \le 2$. The finite sums $f = \sum a_k e^{i\lambda_k x}$ form a dense set in L^2 and hence in L^p too. For such f we have $\tilde{S}_\mu f = f$ for sufficiently large μ. From the trigonometric theory [1], Ch VII. (6.4) we know that

$$\tilde{S}_{[\mu]}^T f \to \tilde{f}^T \quad \text{in} \quad L^p.$$

So on a dense set the uniformly bounded operators A_μ converge. Then the Banach–Steinhaus theorem states that $A_\mu f \to Af$ in L^p for any $f \in L^p$ and the limit operator is bounded as well. Now

$$\tilde{S}_\mu f = A_\mu f + \tilde{S}_{[\mu]}^T f \xrightarrow{L^p} Af + \tilde{f}^T, \quad f \in L^p$$

and then

$$\langle Af + \tilde{f}^T, v_k^* \rangle = \lim_{\mu\to\infty} \langle \tilde{S}_\mu f, v_k^* \rangle = -i \operatorname{sgn} \varrho_k \langle f, v_k^* \rangle$$

which implies that $\tilde{f} = Af + \tilde{f}^T$. Finally

$$\|\tilde{f}\|_p \le \|Af\| + \|\tilde{f}^T\|_p \le c(p)\|f\|_p.$$ ∎

THEOREM 3. *For every $f \in L^1$ the partial sums $\tilde{S}_\mu f$ and $\tilde{S}_{[\mu]}^T f$ are equiconvergent locally uniformly in $(0,2\pi)$ in the sense that $\tilde{S}_\mu(f,x) - \tilde{S}_{[\mu]}^T(f,x)$ converges locally uniformly to a continuous function (a bounded* *element of $C(0,2\pi)$). If $f \in L^p$ for some $p > 1$, then the limit function is equal to $\tilde{f} - \tilde{f}^T$ a.e. (in particular $\tilde{f} - \tilde{f}^T$ is a.e. a bounded continuous function). In this case $\tilde{S}_\mu f \to \tilde{f}$ a.e.*

REMARK. If $p > 2$ then we consider the conjugate function as $\tilde{f} \in L^2$ (because $f \in L^p \subset L^2$). We do not know wether $\tilde{f}$ belongs to L^p or not.

PROOF. We consider again the operators $A_\mu : L^1 \to C[\varepsilon, 2\pi - \varepsilon]$ for some $\varepsilon > 0$. They are uniformly bounded (even in case $\varepsilon = 0$) by Theorem 1. If $f = \sum c_k e^{i\lambda_k x}$ is a finite sum then $\tilde{S}_\mu f = \tilde{f}$ for large μ and by [1], Ch II. (6.8) $\tilde{S}^T_{[\mu]} f \to \tilde{f}^T$ locally uniformly on $(0, 2\pi)$. By the Banach–Steinhaus theorem $A_\mu f \to Af \in C[\varepsilon, 2\pi - \varepsilon]$ uniformly in $[\varepsilon, 2\pi - \varepsilon]$ for every $f \in L^1$. If $f \in L^p$ for some $2 \geq p > 1$ then, as we have seen in the proof of the Theorem 2, $\tilde{S}_\mu f \to \tilde{f}$ in L^p, hence an appropriate sequence $\tilde{S}_{\mu_n} f$ tends to $\tilde{f}$ a.e.. By the famous theorem of R. Hunt, we know that $\tilde{S}_{[\mu_n]} f \to \tilde{f}^T$ a.e. and then the limit function Af equals to $\tilde{f} - \tilde{f}^T$ a.e.. If $p > 2$ then $\tilde{S}_\mu f \to \tilde{f}$ (in L^2) implies the a.e. convergence of $\tilde{S}_{\mu_n} f$ and the other steps remain the same. ∎

REMARK Riesz Marcell-type theorems are proved also in [6] and [7].

References

[1] A. ZYGMUND, *Trigonometric Series*, Cambridge, 1959, vol. I.

[2] R. M. YOUNG, *An Introduction to Nonharmonic Fourier Series*, Academic Press, 1980.

[3] B. JA. LEVIN, *Distribution of Zeros of Entire Functions*, Translations of Mathematical Monographs, Vol. 5. Am. Math. Soc., Providence, Rhode Island, 1964.

[4] V. E. KATZNELSON, Exponential bases in L^2, *Funct. Anal. Appl.*, **5** (1971), 31–38.

[5] B. JA. LEVIN, and I. V. OSTROVSKII, Small perturbations of the set of roots of sine-type functions, *Izv. Akad. Nauk. SSSR*, Ser. Mat., **43** (1979), 87–110.

[6] I. JOÓ, Saturation theorems for Hermite–Fourier series, *Acta Math. Hung.*, **57** (1991), 169–179.

[7] M. HORVÁTH, Some saturation theorems for classical orthogonal expansions II, *Acta Math. Hung.*, **58** (1991), 139–169.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 69–76

ON RIESZ MEANS OF HERMITE–FOURIER SERIES OF FUNCTIONS FROM LIPSCHITZ CLASS

By

I. JOÓ

Department of Analysis, L. Eötvös University, Budapest

(Received August 8, 1991)

Denote h_n the normed Hermite polynomials and

$$R_n^r(f,x)=\int_{-\infty}^{\infty} f(t)e^{-t^2}\sum_{s=0}^{n}\frac{(s+1)^r-s^r}{(n+1)^r}K_s(x,t)\mathrm{d}t$$

the Riesz means of parameter r of the Hermite–Fourier series of f, where

$$K_s(x,t)=\sqrt{\frac{s+1}{2}}\cdot\frac{h_{s+1}(x)h_s(t)-h_s(x)h_{s+1}(t)}{x-t},\quad h_s(x)=\frac{H_s(x)}{\sqrt[4]{\pi}\sqrt{2^s\cdot s!}}.$$

We prove here the following

THEOREM. *Let*

$$g(t)=\begin{cases} t & \textit{if } 1\geq t\geq 0\\ 2-t & \textit{if } 2\geq t\geq 1\\ 0 & \textit{otherwise,}\end{cases}$$

then

$$R_n^{\frac{1}{2}}(g,0)=\frac{1}{\pi\sqrt{8}}\cdot\frac{\log n}{\sqrt{n}}+O\left(\frac{\log\log n}{\sqrt{n}}\right).$$

For the Fejér means of trigonometric Fourier series the analogous result was proved by S. BERNSTEIN (see [2] p. 215). The corresponding upper estimate for the trigonometric case is due to S. BERSTEIN [2] p. 117 and for the Hermite–Fourier series it is proved in [1]. The Alexits theorem for Laguerre and Jacobi–Fourier series were proved by M. HORVÁTH [3], [4].

PROOF. We need the following.

LEMMA. Let $a>0$ be any fixed real number, then we have

$$\sum_{m=1}^{n}\frac{\cos a\sqrt{m}}{m}=O(1)\log\log n. \tag{1}$$

PROOF. Let $2\le y<n$ be arbitrary (it will be chosen later). Then

$$\sum_{m=1}^{n}\frac{\cos a\sqrt{m}}{m}=\sum_{m=y}^{n}\frac{\cos a\sqrt{m}}{m}+O(1)\log y. \tag{2}$$

Now investigate the distribution of $a\sqrt{m}\mod 2\pi$. Let j be arbitrary fixed positive integer (it may depend on n) that will be chosen later, $0\le k<j-1$ be an integer and $0\le l\le \frac{a}{2\pi}\sqrt{n}$ be also an integer. Then

$$2l\pi+k\cdot\frac{2\pi}{j}\le a\sqrt{m}<2l\pi+(k+1)\frac{2\pi}{j}, \tag{3}$$

with suitable l and k. If m satisfies (3) then

$$\cos a\sqrt{m}=\cos\frac{2\pi k}{j}+O\left(\frac{1}{j}\right). \tag{4}$$

Obviously

$$\sum_{m=y}^{n}\frac{\cos a\sqrt{m}}{m}=\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}}\sum_{k=0}^{j-1}\sum_{m=\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2}\frac{\cos a\sqrt{m}}{m}= \tag{5}$$

$$=\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}}\sum_{k=0}^{j-1}\sum_{m=\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2}\frac{1}{m}\cos\frac{2\pi k}{j}+$$

$$+O(1)\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}}\sum_{k=0}^{j-1}\sum_{m=\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2}\frac{1}{m}\cdot\frac{1}{j}.$$

Here

$$\sum_{m=\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2}\frac{1}{m}=2\log\left(l+\frac{k+1}{j}\right)-2\log\left(l+\frac{k}{j}\right)+O(1)\frac{a^2}{l^2}.$$

We can assume that $l\ge\frac{a}{2\pi}\sqrt{y}\ge 2$, so

$$\frac{1+\frac{k+1}{lj}}{1+\frac{k}{lj}}=1+\frac{1}{lj}+O(1)\frac{1}{l^2},$$

thus

$$\sum_{m=\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2} \frac{1}{m} = \frac{2}{lj} + O(1)\frac{1}{l^2} + O(1)\frac{a^2}{l^2}.$$

Using this estimate we obtain from (5)

$$(6)\qquad \sum_{m=y}^{n} \frac{\cos a\sqrt{m}}{m} = \sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}} \sum_{k=0}^{j-1} \left(\frac{2}{lj} + O(1)\frac{1}{l^2} + O(1)\frac{a^2}{l^2}\right)\cos\frac{2\pi k}{j} +$$

$$+O(1)\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}} \left(\frac{1}{lj} + \frac{1}{l^2} + \frac{a^2}{l^2}\right) = \frac{2}{j}\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}} \frac{1}{l}\sum_{k=0}^{j-1}\cos\frac{2\pi k}{j} +$$

$$+O(1)\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}} \left(\frac{1}{lj} + \frac{j}{l^2} + \frac{ja^2}{l^2}\right).$$

Since

$$\sum_{k=0}^{j-1}\cos\frac{2\pi k}{j} = 0$$

hence

$$\sum_{m=y}^{n} \frac{\cos a\sqrt{m}}{m} = O(1)\sum_{l=\frac{a}{2\pi}\sqrt{y}}^{\frac{a}{2\pi}\sqrt{n}} \left(\frac{1}{lj} + \frac{j}{l^2}\right) = O(1)\left(\frac{\log n}{j} + \frac{j}{\sqrt{y}}\right).$$

Now let $y = (\log n)^2$, $j = \log n$. Then we get from (2)

$$\sum_{m=1}^{n} \frac{\cos a\sqrt{m}}{m} = O(1) + O(1)\log\log n$$

and the Lemma is proved.

REMARK. We can prove similarly that for any fixed positive real number $a > 0$ the estimate

$$\sum_{m=1}^{n} \frac{\sin a\sqrt{m}}{m} = O(1)\log\log n$$

holds.

PROOF OF THEOREM. First we give an asymptotic formula for $K_s(0,t)$, $|t| \le c_0$. Since $K_0(x,t) = \frac{1}{\sqrt{\pi}}$, in what follows assume that $s \ge 1$. Suppose

$s=2m$. Then $K_{2m}(0,t)=K_{2m+1}(0,t)$ and

(7)
$$K_{2m}(0,t)=\sqrt{\frac{2m+1}{2}}h_{2m}(0)\frac{h_{2m+1}(t)}{t}=$$
$$=\frac{(-1)^m}{\sqrt[4]{\pi}}\sqrt{\frac{2m+1}{2}}\cdot\frac{\sqrt{(2m)!}}{m!\cdot 2^m}\cdot\frac{h_{2m+1}(t)}{t}.$$

Using the asymptotic formula of the Γ function we obtain:

(8)
$$K_{2m}(0,t)=K_{2m+1}(0,t)=$$
$$=\frac{(-1)^m}{\sqrt{\pi}}\sqrt{\frac{2m+1}{2}}\cdot(m-1)^{-\frac{1}{4}}\cdot\left(1+O\left(\frac{1}{m}\right)\right)\frac{h_{2m+1}(t)}{t}.$$

Here we have to use an asymptotic formula for $h_{2m+1}(t)$. It is well know ([5] (8.22.8)) that

$$\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma(n+1)}e^{-\frac{x^2}{2}}H_n(x)=\cos\left(\sqrt{2n+1}x-\frac{n\pi}{2}\right)+$$
$$+\frac{x^3}{6}(2n+1)^{-\frac{1}{2}}\cdot\sin\left(\sqrt{2n+1}x-\frac{n\pi}{2}\right)+O\left(\frac{1}{n}\right),\quad(|x|\le c_0),$$

where the error term is uniform in x. From this we get

(9)
$$h_n(x)=\frac{e^{\frac{x^2}{2}}\sqrt[4]{2}}{\sqrt{\pi}}\cdot\left(1+O\left(\frac{1}{n}\right)\right)\frac{1}{\sqrt[4]{n}}\left\{\cos\left(\sqrt{2n+1}x-\frac{n\pi}{2}\right)+\right.$$
$$\left.+\frac{x^3}{6}(2n+1)^{-\frac{1}{2}}\cdot\sin\left(\sqrt{2n+1}x-\frac{n\pi}{2}\right)+O\left(\frac{1}{n}\right)\right\}.$$

Using (9) we obtain

(10)
$$K_{2m}(0,t)=K_{2m+1}(0,t)=$$
$$=\frac{\sqrt[4]{2}e^{\frac{t^2}{2}}}{\pi t}\left\{\sin\sqrt{4m+3}t-\frac{t^3}{6}(4m+3)^{-\frac{1}{2}}\cdot\cos\sqrt{4m+3}t\right\}+O\left(\frac{1}{tm}\right),$$

and hence

(11)
$$R_n^{\frac{1}{2}}(g,0)=\int_0^2 g(t)e^{-t^2}\cdot\frac{1}{\sqrt{n+1}}\sum_{s=1}^{n}\left((s+1)^{\frac{1}{2}}-s^{\frac{1}{2}}\right)K_s(0,t)\mathrm{d}t+O\left(\frac{1}{\sqrt{n}}\right)=$$
$$=\frac{\sqrt[4]{2}}{\pi\sqrt{n+1}}\int_0^1 e^{-\frac{t^2}{2}}\sum_{m=1}^{\frac{n}{2}}\left\{\frac{\sin\sqrt{4m+3}t}{\sqrt{2m}}-\frac{t^3(4m+3)^{-\frac{1}{2}}}{6\sqrt{2m}}\cos\sqrt{4m+3}t\right\}\mathrm{d}t-$$

$$-\frac{\sqrt[4]{2}}{\pi\sqrt{n+1}}\int_1^2 e^{-\frac{t^2}{2}}\sum_{m=1}^{\frac{n}{2}}\left\{\frac{\sin\sqrt{4m+3}t}{\sqrt{2m}}-\frac{t^3(4m+3)^{-\frac{1}{2}}}{6\sqrt{2m}}\cos\sqrt{4m+3}t\right\}dt+$$

$$+\frac{2\sqrt[4]{2}}{\sqrt{n+1}}\int_1^2 e^{-t^2}\sum_{m=1}^{\frac{n}{2}}\frac{K_{2m}(0,t)}{\sqrt{2m}}dt+O\left(\frac{1}{\sqrt{n}}\right)=$$

$$=\frac{\sqrt[4]{2}}{\pi\sqrt{n+1}}(I_1-I_2)+\frac{2\sqrt[4]{2}}{\sqrt{n+1}}I_3+O\left(\frac{1}{\sqrt{n}}\right).$$

a) $$I_1=\int_0^1 e^{-\frac{t^2}{2}}\sum_{m=1}^{\frac{n}{2}}\left\{\frac{\sin\sqrt{4m+3}t}{\sqrt{2m}}-\frac{t^3(4m+3)^{-\frac{1}{2}}}{6\sqrt{2m}}\cos\sqrt{4m+3}t\right\}dt.$$

Here integrating by parts

$$\int_0^1 e^{-\frac{t^2}{2}}t^3\cos\sqrt{4m+3}t\,dt=\left[e^{-\frac{t^2}{2}}\cdot\int_0^t s^3\cos\sqrt{4m+3}s\,ds\right]_0^1+$$

$$+\int_0^1 te^{-\frac{t^2}{2}}\left(\int_0^t s^3\cos\sqrt{4m+3}s\,ds\right)dt,$$

but

$$\int_0^t s^3\cos\sqrt{4m+3}s\,ds=O\left(\frac{1}{\sqrt{m}}\right),\quad (|t|\le 1),$$

hence

$$\int_0^1 e^{-\frac{t^2}{2}}t^3\cos\sqrt{4m+3}t\,dt=O\left(\frac{1}{\sqrt{m}}\right),$$

thus

$$I_1=\int_0^1 e^{-\frac{t^2}{2}}\sum_{m=1}^{\frac{n}{2}}\frac{\sin\sqrt{4m+3}t}{\sqrt{2m}}dt+O(1).$$

Here

$$\int_0^1 e^{-\frac{t^2}{2}}\sin\sqrt{4m+3}t\,dt=\frac{1}{\sqrt{4m+3}}\left(1-e^{-\frac{1}{2}}\cos\sqrt{4m+3}\right)+O\left(\frac{1}{m}\right),$$

consequently

$$I_1=\sum_{m=1}^{\frac{n}{2}}\frac{1-e^{-\frac{1}{2}}\cdot\cos\sqrt{4m}}{\sqrt{8}m}+O(1).$$

Using (1) we get

$$I_1=\frac{\log n}{\sqrt{8}}+O(\log\log n). \tag{12}$$

b) $$I_2=\int_1^2 e^{-\frac{t^2}{2}}\sum_{m=1}^{\frac{n}{2}}\left\{\frac{\sin\sqrt{4m+3}t}{\sqrt{2m}}-\frac{t^3(4m+3)^{-\frac{1}{2}}}{6\sqrt{2m}}\cos\sqrt{4m+3}t\right\}\mathrm{d}t.$$

Similarly as above we obtain

$$I_2=\sum_{m=1}^{\frac{n}{2}}\frac{e^{-\frac{1}{2}}\cdot\cos\sqrt{4m}-e^{-2}\cdot\cos 2\sqrt{4m}}{\sqrt{8}m}+O(1),$$

i.e.

$$I_2=O(\log\log n). \tag{13}$$

c) $$I_3=\int_1^2 e^{-t^2}\sum_{m=1}^{\frac{n}{2}}\frac{K_{2m}(0,t)}{\sqrt{2m}}\mathrm{d}t.$$

From (7) we get

$$K_{2m}(0,t)=\frac{1}{\sqrt{\pi}}L_m^{\left(\frac{1}{2}\right)}(t^2), \tag{14}$$

using the well-known relation $H_{2m+1}(x)=(-1)^m\cdot 2^{2m+1}\cdot m!\cdot x\cdot L_m^{\left(\frac{1}{2}\right)}(x^2)$. We need an asymptotic formula for $L_m^{(\alpha)}(x)$. We have ([5] (8.22.2)):

$$L_n^{(\alpha)}(x)=\frac{1}{\sqrt{\pi}}e^{\frac{x}{2}}\cdot x^{-\frac{\alpha}{2}-\frac{1}{4}}\cdot n^{\frac{\alpha}{2}-\frac{1}{4}}\cdot\cos\left\{2(nx)^{\frac{1}{2}}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right\}\cdot$$
$$\cdot\left\{\sum_{\nu=0}^{p-1}A_\nu(x)n^{-\frac{\nu}{2}}+O\left(n^{-\frac{p}{2}}\right)\right\}+$$
$$+\frac{1}{\sqrt{\pi}}e^{\frac{x}{2}}\cdot x^{-\frac{\alpha}{2}-\frac{1}{4}}\cdot n^{\frac{\alpha}{2}-\frac{1}{4}}\cdot\sin\left\{2(nx)^{\frac{1}{2}}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right\}\cdot$$
$$\cdot\left\{\sum_{\nu=0}^{p-1}B_\nu(x)n^{-\frac{\nu}{2}}+O\left(n^{-\frac{p}{2}}\right)\right\},$$

where $A_\nu(x)$, $B_\nu(x)$ are functions of x, independent of n, regular for $x>0$ and the error terms are uniform on $[\varepsilon, w]$, where ε, $w>0$ are arbitrary fixed numbers. Here $A_0(x)=1$, $B_0(x)=O$. From this we obtain

$$(15)\quad L_m^{(\frac{1}{2})}(t^2)=\frac{1}{\sqrt{\pi}}e^{\frac{t^2}{2}}\cdot t^{-1}\cdot\sin 2\sqrt{m}t\cdot\left\{1+A_1(t^2)m^{-\frac{1}{2}}+O(m^{-1})\right\}-$$
$$-\frac{1}{\sqrt{\pi}}e^{\frac{t^2}{2}}\cdot t^{-1}\cdot\cos 2\sqrt{m}t\cdot\left\{B_1(t^2)m^{-\frac{1}{2}}+O(m^{-1})\right\},$$

$(1\le t\le 2)$.

Using (15) we have

$$I_3=\frac{1}{\pi}\sum_{m=1}^{\frac{n}{2}}\frac{1}{\sqrt{2m}}\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot\left\{\sin 2\sqrt{m}t+\frac{A_1(t^2)\sin 2\sqrt{m}t}{\sqrt{m}}+\right.$$
$$\left.+\frac{B_1(t^2)\cos 2\sqrt{m}t}{\sqrt{m}}\right\}\mathrm{d}t+O(1).$$

Integrating by parts

$$\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot A_1(t^2)\sin 2\sqrt{m}t\,\mathrm{d}t=\left[e^{-\frac{t^2}{2}}\cdot\int_1^t s^{-1}\cdot A_1(s^2)\sin 2\sqrt{m}s\,\mathrm{d}s\right]_1^2+$$
$$+\int_1^2 te^{-\frac{t^2}{2}}\cdot\left(\int_1^t s^{-1}A_1(s^2)\sin 2\sqrt{m}s\,\mathrm{d}s\right)\mathrm{d}t,$$

and taking into account

$$\int_1^t s^{-1}\cdot A_1(s^2)\sin 2\sqrt{m}s\,\mathrm{d}s=O\left(\frac{1}{\sqrt{m}}\right),$$

(this is true because $s^{-1}A_1(s^2)$ is regular on $[1,2]$) we get

$$\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot A_1(t^2)\cdot\sin 2\sqrt{m}t\,\mathrm{d}t=O\left(\frac{1}{\sqrt{m}}\right),$$

and similarly

$$\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot B_1(t^2)\cos 2\sqrt{m}t\,\mathrm{d}t=O\left(\frac{1}{\sqrt{m}}\right).$$

Thus

$$I_3=\frac{1}{\pi}\sum_{m=1}^{\frac{n}{2}}\frac{1}{\sqrt{2m}}\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot\sin 2\sqrt{m}t\,\mathrm{d}t+O(1).$$

Here

$$\int_1^2 e^{-\frac{t^2}{2}}\cdot t^{-1}\cdot\sin 2\sqrt{m}t\,\mathrm{d}t=\frac{1}{2\sqrt{m}}\left(e^{-\frac{1}{2}}\cdot\cos 2\sqrt{m}-\frac{e^{-2}}{2}\cos 4\sqrt{m}\right)+O\left(\frac{1}{m}\right),$$

consequently

$$I_3=\frac{1}{\pi}\sum_{m=1}^{\frac{n}{2}}\frac{1}{\sqrt{8m}}\left(e^{-\frac{1}{2}}\cdot\cos 2\sqrt{m}-\frac{e^{-2}}{2}\cos 4\sqrt{m}\right)+O(1),$$

and using (1) we obtain

(16) $$I_3=O(\log\log n).$$

From (11), (12), (13), (16) the statement of the Theorem follows.

References

[1] I. Joó, Saturation theorems for Hermite–Fourier series, *Acta Math. Hung.*, **57** (1991), 169–179.

[2] I. P. Natanson, *Constructive theory of functions* (in Hungarian), Akadémiai Kiadó, Budapest, 1952.

[3] M. Horváth, Some saturation theorems for classical orthogonal expansions I, *Per. Math. Hung.*, **22** (1991), 27–69.

[4] M. Horváth, Some saturation theorems for classical orthogonal expansions II, *Acta Math. Hung.*, **58** (1991), 139–169.

[5] G. Szegő, *Orthogonal Polynomials*, New York, 1939.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 77–98

ON SOME NOTIONS OF HARMONIC ANALYSIS FOR STURM–LIOUVILLE EXPANSIONS

By

I. JOÓ

Department of Analysis, L. Eötvös University, Budapest

(Received August 29, 1991)

The notion of analytic function, the Cauchy–Riemann equations, the conjugate function can be transformed from the trigonometrical case to other orthogonal expansions. First we mention the paper [5] of E. M. STEIN and B. MUCKENHOUPT where some basic facts of harmonic analysis are proved for the ultraspherical and Bessel expansions. The Jacobi and Bessel conjugate are introduced and investigated (see also the end of the book [1]). MUCKENHOUPT applied then these ideas to the Hermite and Laguerre expansions, proved the norm boundedness of the conjugate function operator. Based on this result the author of this paper proved saturation theorems of Alexits-type and investigated also the Abel–Poisson means of Hermite expansions, see [10]–[15]; after M. HORVÁTH studied analogous problems for the Jacobi and Laguerre expansions [16], [17].

In this paper we aim to give a common generalization of some of the above mentioned concepts and results for the case of Sturm–Liouville expansions. Consider an interval $G=(a,b)$ finite or infinite and the eigenvalue problem

$$-y''+b(x)y'=\lambda y, \qquad x\in G, \quad b\in C^1(G), \quad b'>0. \tag{1}$$

Introduce the weight function $w=e^{-\int^x b}$.

We make the following assumptions:

(2) There exists a system $(y_n w^{1/2})\subset L^2(G)$ complete in $L^2(G)$, where y_n is a solution of (1), the corresponding eigenvalues λ_n are non-negative and increase strictly to infinity further

$$\lim_{a+} y_n y_k w = \lim_{b-} y_n y_k w = 0 \text{ for all } n,k.$$

We can transform (1) to a Schrödinger equation in two ways. A short counting shows that

$$-v_n''+\frac{(w^{1/2})''}{w^{1/2}}v_n=\lambda_n v_n, \quad v_n:=y_n w^{1/2}, \tag{3}$$

$$-v_{1,n}''+\left(\frac{(w^{1/2})''}{w^{1/2}}+b'\right)v_{1,n}=\lambda_n v_{1,n}, \quad v_{1,n}:=\frac{y_n'}{\sqrt{\lambda_n}}w^{1/2}, \quad (\lambda_n\neq 0). \tag{4}$$

Remark that

$$\frac{(w^{1/2})''}{w^{1/2}}=\frac{b^2}{4}-\frac{b'}{2}. \tag{5}$$

We fix first the differentiation rules.

LEMMA 1.

$$[v_n w^{-1/2}]'=\sqrt{\lambda_n}v_{1,n}w^{-1/2}, \tag{6}$$

$$[v_{1,n}w^{1/2}]'=-\sqrt{\lambda_n}v_n w^{1/2}. \tag{7}$$

PROOF. Indeed, both sides of (6) is equal to y_n' further

$$[v_{1,n}w^{1/2}]'=\left[\frac{y_n'}{\sqrt{\lambda_n}}w\right]'=\frac{(y_n''-y_n'b)w}{\sqrt{\lambda_n}}=\frac{-\lambda_n y_n w}{\sqrt{\lambda_n}}=-\sqrt{\lambda_n}v_n w^{1/2}$$

which proves (7). ∎

LEMMA 2. The systems (v_n) and $(v_{1,n})$ are orthonormal in $L^2(G)$ (after changing y_n by $c_n y_n$ if necessary).

PROOF. Let $\|v_n\|_{L^2}=1$, then for $n\neq k$ we have

$$\lambda_n\langle v_n,v_k\rangle=\left\langle -v_n''+\frac{(w^{1/2})''}{w^{1/2}},v_k\right\rangle=\lim_{\substack{a'\to a+\\ b'\to b-}}\int_{a'}^{b'}\left(-v_n''+\frac{(w^{1/2})''}{w^{1/2}}v_n\right)v_k=$$

$$=\lim_{\substack{a'\to a+\\ b'\to b-}}\left[-v_n'v_k+v_nv_k'\right]_{a'}^{b'}+\lim_{\substack{a'\to a\\ b'\to b-}}\int_{a'}^{b'}v_n\left(-v_k''+\frac{(w^{1/2})''}{w^{1/2}}v_k\right)=$$

$$=\lim_{b-}(wy_ny_k'-wy_ky_n')-\lim_{a+}(wy_ny_k'-wy_ky_n')+\langle v_n,\lambda_k v_k\rangle=\lambda_k\langle v_n,v_k\rangle.$$

Since $\lambda_n\neq\lambda_k$ for $n\neq k$, we obtain the orthonormality of (v_n). On the other hand we get for $\lambda_n\neq 0$

$$\delta_{nk}=\langle v_n,v_k\rangle=\int_G y_ny_kw=\frac{1}{\lambda_n}\int_G y_k(-y_n''+by_n')w=$$

$$=\frac{1}{\lambda_n}\lim_{\substack{a'\to a+\\ b'\to b-}}\int_{a'}^{b'}-y_k(y_n'w)'=\frac{1}{\lambda_n}\left(\lim_{b-}-\lim_{a+}\right)(-y_ky_n'w)+$$

$$+\frac{1}{\lambda_n}\int_G y_k'y_n'w=\frac{1}{\lambda_n}\left\langle y_k'w^{1/2},y_n'w^{1/2}\right\rangle$$

and then $\delta_{nk}=\left\langle v_{1,n},v_{1,k}\right\rangle$ as we asserted. ∎

Before continuing the investigations, we list some widely known special cases.

EXAMPLE A: *Hermite* polynomials and functions.

Denote h_n the normalized Hermite polynomials defined by

$$\int_{\mathbb{R}} h_n(x)h_k(x)e^{-x^2}\mathrm{d}x=\delta_{nk}.$$

Then $w=e^{-x^2}$ is the weight and

$$(8)\qquad \left|\begin{array}{ll} -y_n''+2xy_n'=2ny_n, & y_n=h_n(x)\\ -u_n''+2xu_n+2u_n=2nu_n, & u_n=h_n'(x)=\sqrt{2n}h_{n-1}(x)\\ -v_n''+(x^2-1)v_n=2nv_n, & v_n=h_n(x)e^{-x^2/2}\\ -v_{1,n}+(x^2+1)v_{1,n}=2nv_{1,n}, & v_{1,n}=\frac{y_n'}{\sqrt{2n}}w^{1/2}=v_{n-1},\ n\geq 1. \end{array}\right.$$

The Hermite functions are complete in $L^2(\mathbb{R})$ and

$$\lim_{\pm\infty} y_ny_k'w=\lim_{\pm\infty} h_nh_k'e^{-x^2}=0$$

so (2) fulfills.

EXAMPLE B: *Laguerre* polynomials and functions, $\alpha>-1$.

Denote $\ell_n^{(\alpha)}(x)$ the normed Laguerre polynomials

$$\int_0^\infty \ell_n^{(\alpha)}\ell_k^{(\alpha)}x^\alpha e^{-x}\mathrm{d}x=\delta_{n,k}.$$

Then $w=2x^{2\alpha+1}e^{-x^2}$ and

$$
(9)\quad
\left|
\begin{array}{l}
-x\left(\ell_n^{(\alpha)}\right)''+(x-\alpha-1)\left(\ell_n^{(\alpha)}\right)'=n\ell_n^{(\alpha)},\\
-y_n''+\frac{2x^2-2\alpha-2}{x}y_n'=4ny_n,\quad y_n:=\ell_n^{(\alpha)}(x^2),\\
-u_n''+\frac{2x^2-2\alpha-1}{x}u_n'+\left(2+\frac{2\alpha+1}{x^2}\right)u_n=4nu_n,\\
\qquad\qquad u_n:=\left[\ell_n^{(\alpha)}(x^2)\right]'=2\sqrt{n}x\ell_{n-1}^{(\alpha+1)}(x^2),\\
-v_n''+\left(x^2-2\alpha-2+\frac{\alpha^2-\frac14}{x^2}\right)v_n=4nv_n,\\
\qquad\qquad v_n=\sqrt{2}x^{\alpha+1/2}\cdot e^{-x^2/2}\ell_n^{(\alpha)}(x^2),\\
-v_{1,n}''+\left(x^2-2\alpha+\frac{\left(\alpha+\frac12\right)\left(\alpha+\frac32\right)}{x^2}\right)v_{1,n}=4nv_{1,n},\\
\qquad\qquad v_{1,n}=\sqrt{2}x^{\alpha+3/2}e^{-x^2/2}\ell_{n-1}^{(\alpha+1)}(x^2).
\end{array}
\right.
$$

We see that $v_{1,n}$ equals to $v_{n-1}^{(\alpha+1)}$, where $v_{n-1}^{(\alpha+1)}$ denotes the function v_{n-1} with the parameter $\alpha+1$ instead of α. Here the condition $\lim\limits_{\infty} y_n y_k' w=0$ obviously holds and

$$\lim_{0+} y_n y_k' w=2\sqrt{k}\ell_n^{(\alpha)}(x^2)\ell_{k-1}^{(\alpha+1)}(x^2)2x^{2\alpha+2}e^{-x^2}=0$$

is also true when $\alpha>-1$.

EXAMPLE C: *Jacobi* polynomials and functions; α, $\beta>-1$.

Denote $p_n^{(\alpha,\beta)}(x)$ the orthonormal Jacobi polynomials

$$\int_{-1}^{1} p_n^{(\alpha,\beta)}(x)p_k^{(\alpha,\beta)}(x)(1-x)^\alpha(1+x)^\beta \mathrm{d}x=\delta_{nk}.$$

Here we have

$$G=(0,\pi),\quad w=2^{\alpha+\beta+1}\sin^{2\alpha+1}\frac{\Theta}{2}\cos^{2\beta+1}\frac{\Theta}{2}$$

and

$$
(10)\quad
\left|
\begin{aligned}
&-(1-x^2)[p_n^{(\alpha,\beta)}]'' + [(\alpha+\beta+2)x-\beta+\alpha][p_n^{(\alpha,\beta)}]' = \\
&\qquad = n(n+\alpha+\beta+1)p_n^{(\alpha,\beta)},\\
&-y_n'' + \frac{\beta-\alpha-(\alpha+\beta+1)\cos\Theta}{\sin\Theta} y_n' = n(n+\alpha+\beta+1)y_n,\\
&\qquad y_n = p_n^{(\alpha,\beta)}(\cos\Theta),\\
&-u_n'' + \frac{\beta-\alpha-(\alpha+\beta+1)\cos\Theta}{\sin\Theta} u_n' + \left[(\beta-\alpha)\frac{\cos\Theta}{\sin^2\Theta} + \frac{\alpha+\beta+1}{\sin^2\Theta}\right] u_n = \\
&\qquad = n(n+\alpha+\beta+1)u_n, \quad u_n = [p_n^{(\alpha,\beta)}(\cos\Theta)]' = \\
&\qquad = \sqrt{n(n+\alpha+\beta+1)}\,p_{n-1}^{(\alpha+1,\beta+1)}(\cos\Theta)\cdot(-\sin\Theta),\\
&v_n'' + \left(\frac{\alpha^2-\frac14}{4\sin^2\frac{\Theta}{2}} + \frac{\beta^2-\frac14}{4\cos^2\frac{\Theta}{2}} - \left(\frac{\alpha+\beta+1}{2}\right)^2\right) v_n = n(n+\alpha+\beta+1)v_n,\\
&\qquad v_n = 2^{(\alpha+\beta+1)/2}\sin^{\alpha+1/2}\tfrac{\Theta}{2}\cos^{\beta+1/2}\tfrac{\Theta}{2}\,p_n^{(\alpha,\beta)}(\cos\Theta),\\
&-v_{1,n}'' + \left(\frac{\left(\alpha+\frac12\right)\left(\alpha+\frac32\right)}{4\sin^2\frac{\Theta}{2}} + \frac{\left(\beta+\frac12\right)\left(\beta+\frac32\right)}{4\cos^2\frac{\Theta}{2}} - \left(\frac{\alpha+\beta+1}{2}\right)^2\right) v_{1,n} = \\
&\qquad = n(n+\alpha+\beta+1)\cdot v_{1,n}\\
&v_{1,n} := -\sin\Theta\, w^{1/2}(\Theta)\,p_{n-1}^{(\alpha+1,\beta+1)}(\cos\Theta) = \\
&\qquad = -2^{(\alpha+\beta+3)/2}\sin^{\alpha+3/2}\tfrac{\Theta}{2}\cos^{\beta+3/2}\tfrac{\Theta}{2}\,p_{n-1}^{(\alpha+1,\beta+1)}(\cos\Theta).
\end{aligned}
\right.
$$

Here again $v_{1,n} = v_{1,n}^{(\alpha+1,\beta+1)}$ i.e. v_{n-1} with parameters $\alpha+1$, $\beta+1$. The boundary conditions (2) hold for every α, $\beta > -1$ since

$$\lim_{\substack{\Theta\to 0\\(\Theta\to\pi)}} y_n y_k' w = \lim_{\substack{\Theta\to 0\\(\Theta\to\pi)}} c\sin^{2\alpha+2}\frac{\Theta}{2}\cos^{2\alpha+2}\frac{\Theta}{2}p_n^{(\alpha,\beta)}(\cos\Theta)p_{n-1}^{(\alpha+1,\beta+1)}(\cos\Theta) = 0.$$

EXAMPLE D: *Bessel* functions, $\nu > -1$.

We know from [20], 7.2.8 (56), (57) and 7.14.1 (10) that if γ_n denotes the positive zeros of the Bessel function $J_\nu(x)$, i.e. $J_\nu(\gamma_n) = 0$, then

$$\int_0^1 x[J_\nu(\gamma_n x)]^2 \mathrm{d}x = \frac12[J_{\nu+1}(\gamma_n)]^2 = \frac12[J_\nu'(\gamma_n)]^2 =: \frac{1}{\alpha_n^2}.$$

We have the equations $w=x^{2\nu+1}$,

$$
(11)\qquad \left|\begin{aligned}
&-y_n''-\tfrac{2\nu+1}{x}y_n'=\gamma_n^2 y_n, \quad y_n=\alpha_n x^{-\nu}J_\nu(\gamma_n x),\\
&-u_n''-\tfrac{2\nu+1}{x}u_n'+\tfrac{2\nu+1}{x^2}u_n=\gamma_n^2 u_n,\\
&\qquad\qquad u_n=\alpha_n[x^{-\nu}J_\nu(\gamma_n x)]'=-\alpha_n\gamma_n x^{-\nu}\cdot J_{\nu+1}(\gamma_n x),\\
&-v_n''+\frac{\nu^2-\frac14}{x^2}v_n=\gamma_n^2 v_n, \quad v_n:=\alpha_n x^{1/2}J_\nu(\alpha_n x),\\
&-v_{1,n}''+\frac{\left(\nu+\frac12\right)\left(\nu+\frac32\right)}{x^2}v_{1,n}=\gamma_n^2 v_{1,n}, \quad v_{1,n}:=-\alpha_n x^{\frac12}J_{\nu+1}(\gamma_n x)
\end{aligned}\right.
$$

Now for $\nu>-1$ we have

$$\lim_{\substack{x\to 0\\(x\to 1)}} y_n y_k' w=\lim_{\substack{x\to 0\\(x\to 1)}}\left(-\alpha_n\alpha_k\gamma_k\cdot xJ_\nu(\gamma_n x)J_{\nu+1}(\gamma_k x)\right)=0.$$

REMARK. Since in the general setting the functions v_n form an orthonormal basis in $L^2(G)$, any function $f\in L^2(G)$ has an expansion

$$f\sim\sum a_k v_k, \qquad a_k=\int_G f v_k.$$

If we want to extend this definition to $f\in L^p(G)$, $1\le p\le\infty$, we need that $v_k\in L^1\cap L^\infty$ for every k. In the Hermite case it holds trivially; in the cases B, C, D we must assume that $\alpha\ge-1/2$; $\alpha,\beta\ge-1/2$; $\nu\ge-1/2$ respectively. In general we shall take the assumption

$$(12)\qquad v_n,\ v_{1,n}\in L^1(G)\cap L^\infty(G) \quad (\forall n).$$

DEFINITION. Let $1\le p\le\infty$. Take a function

$$f\in L^p(G), \quad f\sim\sum a_k v_k, \quad a_k=\int_G f v_k.$$

If there exists a function

$$\tilde f\in L^p(G), \quad \tilde f\sim-\sum a_k v_{1,k}, \quad -a_k=\int_G \tilde f v_{1,k},$$

then we say that $\tilde f$ is the conjugate function of f.

The function $f+i\tilde f$ is called analytic. In other words, the complex valued function $g=g_1+ig_2$

$$g_1\sim\sum a_n v_n, \quad a_n=\langle g_1,v_n\rangle, \quad g_2\sim\sum b_n v_{1,n}, \quad b_n=\langle g_2,v_{1,n}\rangle$$

is analytic if and only if $g_2 = \tilde{g}_1$ (or equivalently if $b_n = -a_n$ for all n with $\lambda_n \neq \neq 0$). Hence we consider the n-th analytic unit $v_n - iv_{1,n}$; roughly speaking the linear combinations of these members are the analytic functions.

REMARK. In the case C, $\alpha = \beta = -1/2$ we have

$$v_n = \cos n\Theta, \quad (n \geq 0), \quad v_{1,n}(\Theta) = -\sin n\Theta \quad (n \geq 1) \quad (v_n - iv_{1,n})(\Theta) = e^{in\Theta}$$

and

$$f(\Theta) \sim \sum a_n \cos n\Theta \Longrightarrow \tilde{f} \sim \sum a_n \sin n\Theta.$$

This is the direct motivation of the above definition. As the following examples show, the situation is asymptotically the same for the other expansions too.

EXAMPLE A. By Szegő [18], 8.22.8 we get that

$$(13) \quad \left| \begin{aligned} & v_n(x) = e^{-x^2/2} h_n(x) = \sqrt{\frac{\sqrt{2}}{\pi}} \frac{\cos(\sqrt{2n}x - n\frac{\pi}{2})}{n^{1/4}} + O(n^{-3/4}), \\ & v_{1,n}(x) = e^{-x^2/2} h_{n-1}(x) = -\sqrt{\frac{\sqrt{2}}{\pi}} \frac{\sin(\sqrt{2n}x - n\frac{\pi}{2})}{n^{1/4}} + O(n^{-3/4}), \\ & v_n(x) - iv_{1,n}(x) = \sqrt{\frac{\sqrt{2}}{\pi}} n^{-1/4} e^{i(\sqrt{2n}x - n\frac{\pi}{2})} + O(n^{-3/4}). \end{aligned} \right.$$

These estimates are uniform in $|x| \leq \omega$ where $\omega > 0$ is arbitrarily fixed.

EXAMPLE B. Using [18], 8.22.1 we obtain

$$(14) \quad \left| \begin{aligned} & v_n(x) = (-1)^n \sqrt{\frac{2}{\pi}} \frac{\cos(2x\sqrt{n} - \alpha\frac{\pi}{2} - \frac{\pi}{4})}{n^{1/4}} + O(n^{-3/4}), \\ & v_{1,n}(x) = (-1)^{n-1} \sqrt{\frac{2}{\pi}} \frac{\sin(2x\sqrt{n} - \alpha\frac{\pi}{2} - \frac{\pi}{4})}{n^{1/4}} + O(n^{-3/4}), \\ & v_n - iv_{1,n} = (-1)^n \sqrt{\frac{2}{\pi}} n^{-1/4} e^{i(2\sqrt{n}x - \alpha\frac{\pi}{2} - \frac{\pi}{4})} + O(n^{-3/4}), \\ & \text{uniformly in } 0 < \varepsilon \leq x \leq \omega < \infty. \end{aligned} \right.$$

EXAMPLE C. Here we use [18], 8.21.10

$$(15) \quad \left| \begin{aligned} & v_n(\Theta) = \sqrt{\frac{2}{\pi}} \cos(N\Theta + \gamma) + O(1/n), \\ & v_{1,n}(\Theta) = -\sqrt{\frac{2}{\pi}} \sin(N\Theta + \gamma) + O(1/n), \\ & v_n(\Theta) - iv_{1,n}(\Theta) = \sqrt{\frac{2}{\pi}} e^{i(N\Theta + \gamma)} + O(1/n), \\ & \qquad N := n + \frac{\alpha + \beta + 1}{2}, \quad \gamma := -\frac{\pi}{2}\left(\alpha + \frac{1}{2}\right) \\ & \text{and these estimates are uniform in } \varepsilon \leq \Theta \leq \pi - \varepsilon \text{ for any } \varepsilon > 0. \end{aligned} \right.$$

EXAMPLE D. We apply [20], 7.13.1 (3) to obtain that

$$
(16)\quad \left|\begin{array}{l}
\gamma_n = \pi\left(n+\frac{\nu}{2}+\frac{3}{4}\right)+O(1/n),\ \alpha_n=\pi\sqrt{n}+O(1/n),\ (n=1,2,\dots)\\
v_n(x)=\sqrt{2}\cos\left[\pi\left(n+\frac{\nu}{2}+\frac{3}{4}\right)x-\frac{\pi\nu}{2}-\frac{\pi}{4}\right]+O(1/n),\\
v_{1,n}(x)=-\sqrt{2}\cos\left[\pi\left(n+\frac{\nu}{2}+\frac{3}{4}\right)x-\frac{\pi\nu}{2}-\frac{3\pi}{4}\right]+O(1/n)=\\
\qquad =-\sqrt{2}\sin\left[\pi\left(n+\frac{\nu}{2}+\frac{3}{4}\right)x-\frac{\pi\nu}{2}-\frac{\pi}{4}\right]+O(1/n),\\
v_n(x)-iv_{1,n}(x)=\sqrt{2}e^{i\left[\pi\left(n+\frac{\nu}{2}+\frac{3}{4}\right)x-\frac{\pi\nu}{2}-\frac{\pi}{4}\right]}+O\left(\frac{1}{n}\right),\\
\text{and the } O\text{-term is uniform for } \varepsilon\le x\le 1 \text{ if } \varepsilon>0 \text{ is arbitrarily fixed.}
\end{array}\right.
$$

The classical theorem of G. ALEXITS has been generalized for the case of some classical orthonormal expansions in the papers [10]–[17]. We give here the analogous result for the Jacobi and Bessel expansions. The present method works also in Hermite and Laguerre case, see [13], [17]. We hope that it will be a basis for a general proof.

THEOREM 1. *Let α, $\beta>-1/2$, $1<p\le\infty$, $f\in L^p(-1,1)$. Then the following statements are equivalent for the Jacobi expansions:*

$$
\text{(a)}\qquad \|R_nf-f\|_p=O\left(\frac{1}{\sqrt{\lambda_n}}\right),\quad R_nf:=\sum_{k=0}^{n}\left(1-\frac{\sqrt{\lambda_k}}{\sqrt{\lambda_{n+1}}}\right)a_kv_k,
$$

$$
f\sim\sum a_kv_k,
$$

(b) *There exists $\tilde f\in L^p(-1,1)$, $\tilde f$ is locally absolutely continuous and*

$$
w^{-1/2}[\tilde f w^{1/2}]'\in L^p(-1,1).
$$

THEOREM 2. *Let $\nu>-1/2$, $1<p\le\infty$ and suppose that*

$$
(17)\qquad \|R_nf\|\le c(p)\|f\|_p,\quad f\in L^p(0,1)\quad\text{if}\quad 1<p\le\infty
$$

holds for the Bessel expansion. Then the statements (a) *and* (b) *of* Theorem 1 *are equivalent for the Bessel expansions.*

PROOF OF THE THEOREMS 1 AND 2.

Remark first that (17) holds for the Jacobi expansion, see [16]. The first step of the proof can be applied for general Sturm–Liouville expansions, too. By a generalization of the Alexits lemma [14],

$$
\|R_nf-f\|_p=O(1/\sqrt{\lambda_n})\Leftrightarrow\left\|R_n\left(\sum\sqrt{\lambda_k}a_kv_k\right)\right\|_p=O(1).
$$

Next we show that

$$\text{(18)}\qquad \left\| R_n\left(\sum \sqrt{\lambda_k} a_k v_k\right)\right\|_p = O(1) \Leftrightarrow$$
$$\Leftrightarrow \text{there exists } g \in L^p(0,1) \text{ with } g \sim \sum \sqrt{\lambda_k} a_k v_k.$$

The implication "$\Leftarrow$" follows from (17). The converse can be proved by introducing the functional sequence

$$T_n h := \int_G h R_n, \qquad R_n = R_n\left(\sum \sqrt{\lambda_k} a_k v_k\right).$$

These are uniformly bounded functionals on L^q, $p^{-1}+q^{-1}=1$ and for $h=v_k \in L^q$ we have $T_n v_k = \left(1-\frac{\sqrt{\lambda_k}}{\sqrt{\lambda_{n+1}}}\right) a_k \sqrt{\lambda_k} \to a_k\sqrt{\lambda_k}$ as $n\to\infty$. Consequently the operator sequence T_n converges in the linear space spanned by the (v_k)-s to a (necessarily bounded) functional T. By the Hahn–Banach theorem T can be extended to a continuous functional of the whole L^q, hence it has the form $Th = \int_G gh$ with some $g \in L^p$. In particular we have $a_k\sqrt{\lambda_k} = Tv_k = \int_G g v_k$ so $g \sim \sum \sqrt{\lambda_k} a_k v_k$ which proves (18). In the next step we suppose the existence of such g and prove (b). Consider the function

$$f_1(x) := w^{-1/2}(x) \int_0^x g w^{1/2}.$$

We prove that

$$\text{(19)}\qquad f_1 \in L^p, \quad \lim_{\substack{a\\(b)}} y_k(x) \int_0^x w^{1/2} = 0 \quad (k=0,1,2,\ldots).$$

Since f_1 is continuous (locally absolute continuous) in G, we need only the boundary behaviour. Consider first the Jacobi case. Suppose that x is nearly zero, then

$$|f_1(x)| \le w^{-1/2}(x)\|g\|_p \left(\int_0^x w^{q/2}\right)^{1/q} \le cx^{-\alpha-1/2}\left(\int_0^x t^{(\alpha+1/2)q}\right)^{1/q} \le$$
$$\le cx^{1/q} \to 0 \quad (\text{as } x\to 0),$$

$$\left| y_k(x) \int_0^x g w^{1/2}\right| \le c_k \left(\int_0^x w^{q/2}\right)^{1/q} \le c_k x^{\alpha+1/2+1/q} \to 0 \quad (\text{as } x \to 0).$$

Since the constant coefficient of g is zero, $0=\int_0^\pi g v_0=\int_0^\pi g w^{1/2}$, hence for x near π we have

$$|f_1(x)|\le w^{-1/2}(x)\int_x^\pi |g|w^{1/2}\le c w^{-1/2}(x)\|g\|_p\left(\int_x^\pi w^{q/2}\right)^{1/q}\le$$

$$\le c(\pi-x)^{-\beta-1/2}\left(\int_x^\pi(\pi-t)^{(\beta+1/2)q}\mathrm{d}t\right)^{1/q}\le c(\pi-x)^{1/q}\to 0\quad(\text{as } x\to\pi),$$

$$\left|y_k(x)\int_0^x g w^{1/2}\right|\le c_k\left(\int_x^\pi w^{q/2}\right)^{1/q}\to 0\quad(\text{as } x\to\pi).$$

For the Bessel expansion we have analogous estimates for $x\to 0$ but here $\int_0^1 g w^{1/2}=0$ does not hold. So for $x\to 1$ we can estimate by using $y_k(1)=0$:

$$|f_1(x)|\le w^{-1/2}(x)\|g\|_p\left(\int_0^x w^{q/2}\right)^{1/q}\le c x^{-\nu-1/2}\left(\int_0^x t^{(\nu+1/2)q}\mathrm{d}t\right)^{1/q}\le c$$

$$(\text{as } x\to 1),$$

$$\left|y_k(x)\int_0^x g w^{1/2}\right|\le |y_k(x)|\|g\|_p\left(\int_0^x w^{q/2}\right)^{1/q}\to 0\quad(\text{as } x\to 1).$$

So (19) holds indeed and we can compute the coefficients by (6), $k\ge 1$

$$\langle f_1,v_{1,k}\rangle=\int_G(w^{-1/2}v_{1,k})\int_0^x(g w^{1/2})\mathrm{d}x=\left[\frac{1}{\sqrt{\lambda_n}}w^{-1/2}v_k\int_0^x g w^{1/2}\right]_a^b-$$

$$-\int_G\frac{1}{\sqrt{\lambda_k}}\left(-w^{-1/2}v_k g w^{1/2}\right)=\frac{1}{\sqrt{\lambda_k}}\left[y_k(x)\int_0^x g w^{1/2}\right]_a^b-\frac{1}{\sqrt{\lambda_k}}\int_G g v_k=-a_k.$$

Consequently $f_1\sim-\sum a_k v_{1,k}$, $f_1=\tilde f$, $\tilde f$ is locally absolute continuous and

$$w^{-1/2}[\tilde f w^{1/2}]'=g\in L^p,$$

so (b) is proved. Conversely suppose the existence of $h \in L^p$, $h = w^{-1/2} \cdot [\tilde{f} w^{1/2}]'$ and compute its coefficients:

$$\langle h, v_k \rangle = \int_G w^{-1/2} v_k [\tilde{f} w^{1/2}]' = [\tilde{f} v_k]_a^b - \sqrt{\lambda_k} \int_G w^{-1/2} v_{1,k} \tilde{f} w^{1/2} =$$

$$= [\tilde{f} v_k]_a^b + \sqrt{\lambda_k} a_k.$$

We shall prove

$$[\tilde{f} v_k]_a^b = 0. \tag{20}$$

In fact this is always true if $w \in L^p$, $1 \le p < \infty$. Indeed,

$$\tilde{f}(x) w^{1/2}(x) - \tilde{f}(a') w^{1/2}(a') = \int_{a'}^{x} h w^{1/2} = O(1)$$

uniformly in x and a'. Taking the limit $a' \to a$ we see that there exists $\lim_{a' \to a} \tilde{f}(a') w^{1/2}(a') = c_0$ and then

$$\tilde{f}(x) w^{1/2}(x) = c_0 + \int_a^x h w^{1/2} = O(1).$$

Since $v_k(a) = v_k(b) = 0$, (20) follows and then $\langle h, v_k \rangle = \sqrt{\lambda_k} a_k$. Theorems 1 and 2 are proved. ∎

THEOREM 1'. *Let $\alpha, \beta > -1/2$, $1 < p \le \infty$, $f \in L^p$. Then for the Jacobi expansions the following statements are equivalent:*

(a') *There exists $\tilde{f} \in L^p$, $\|R_n \tilde{f} - \tilde{f}\|_p = O(1/\sqrt{\lambda_n})$,*

(b') *f is locally absolutely continuous and $w^{1/2}[f w^{-1/2}]' \in L^p$.*

THEOREM 2'. *Let $\nu > -1/2$, $1 < p \le \infty$, $f \in L^p$ and suppose that (17) is true for the Bessel expansion. Then the Bessel variant of* (a') $\Leftrightarrow$ (b') *holds, if we assume that (v_k) is complete in L^p for $1 < p \le \infty$.*

PROOF OF THEOREMS 1' AND 2'. By the above mentioned Alexits lemma, (a') holds if and only if $\left\| R_n \left(\sum \sqrt{\lambda_k} a_k v_{1,k} \right) \right\| = O(1)$ and this is equivalent to the existence of $h \in L^p$ with $h \sim \sum \sqrt{\lambda_k} a_k v_{1,k}$. Consider the function

$$f_2(x) := w^{1/2}(x) \int_{1/2}^{x} h w^{-1/2};$$

we pick the point $1/2\in G$. We shall show

$$\text{(19')}\qquad f_2\in L^p,\qquad \lim_{\substack{a\\(b)}}\left[v_{1,k}w^{1/2}\int_{1/2}^{x} hw^{-1/2}\right]=0.$$

Indeed, for the Jacobi case and $x\to b=\pi$

$$w^{1/2}(x)\left|\int_{1/2}^{x} hw^{-1/2}\right|\le \|h\|_p w^{1/2}(x)\left(\int_{1/2}^{x} w^{-q/2}\right)^{1/q}\le$$

$$\le c(\pi-x)^{\alpha+1/2}\left(\int_{1/2}^{x}(\pi-t)^{(-\alpha-1/2)q}\mathrm{d}t\right)^{1/q}\le$$

$$\le\begin{cases} c(\pi-x)^{\alpha+\frac12}, & \left(\alpha+\frac12\right)q<1\\ c(\pi-x)^{\alpha+\frac12}\ln\frac{1}{\pi-x}, & \left(\alpha+\frac12\right)q=1\\ c(\pi-x)^{1/q}, & \left(\alpha+\frac12\right)q>1\end{cases}$$

so f_2 is bounded and from $v_{1,k}(0)=v_{1,k}(\pi)=0$ it follows (19'). For the Bessel case

$$w^{1/2}(x)\left|\int_{1/2}^{x} hw^{-1/2}\right|\le \|h\|_p x^{\nu+1/2}\left(\int_{1/2}^{x} t^{-(\nu+1/2)q}\mathrm{d}t\right)$$

is bounded in case $x\to 1$ and tends to zero if $x\to 0$; so $f_2\in L^p$ and since $v_{1,k}(0)=v_{1,k}(1)=0$, (19') follows also in this case. Count the coefficients of $f_2\in L^p$ for $\lambda_k\ne 0$

$$\langle f_2,v_k\rangle=\int_G v_k(x)w^{1/2}(x)\left(\int_{1/2}^{x} hw^{-1/2}\right)\mathrm{d}x=$$

$$=\left[-\frac{1}{\sqrt{\lambda_k}}v_{1,k}(x)w^{1/2}(x)\int_{1/2}^{x} hw^{-1/2}\right]_a^b+\frac{1}{\sqrt{\lambda_n}}\int_g hv_{1,k}=a_k.$$

In the Bessel case the completeness of (v_n) implies that $f_2=f$; in the Jacobi case the coefficients of f_2-f vanish except for the first one, thus $f = f_2 + cw^{1/2}$. In both cases f is locally absolutely continuous and $w^{1/2}[fw^{-1/2}]' = h \in L^p$ which is (b'). Conversely let $h := w^{1/2}[fw^{-1/2}]' \in L^p$ and compute the coefficients

$$\langle h, v_{1,k}\rangle = \int_G v_{1,k} w^{1/2}[fw^{-1/2}]' = [v_{1,k} f]_a^b + \sqrt{\lambda_k}\int_G v_k f.$$

We shall prove

$$[v_{1,k} f]_a^b = 0. \tag{20'}$$

Indeed,

$$f(x)w^{-1/2}(x) = f(1/2)w^{-1/2}(1/2) + \int_{1/2}^{x} hw^{-1/2},$$

$$|f(x)| \le cw^{1/2}(x)\left[1 + \left(\int_{\frac{1}{2}}^{x} w^{-q/2}\right)^{1/q}\right]$$

which shows that $f(x)$ is bounded; from $v_{1,k}(a) = v_{1,k}(b) = 0$ the statement (20') follows and then $\langle h, v_{1,k}\rangle = \sqrt{\lambda_k} a_k$, $h \sim \sum \sqrt{\lambda_k} a_k v_{1,k}$. The proof is complete. ∎

REMARK. The above theorems are saturation type statements. For the general Sturm–Liouville case we can argue as follows. Suppose that

$$\|R_n f - f\|_p = o(1/\sqrt{\lambda_n}).$$

Then

$$|\langle R_n f - f, v_k\rangle| \le \|v_k\|_q \|R_n f - f\|_p = o(1/\sqrt{\lambda_n}),$$

$$o(1/\sqrt{\lambda_n}) = \left|\frac{\sqrt{\lambda_k}}{\sqrt{\lambda_n}} a_k\right|, \quad \sqrt{\lambda_k} a_k = o(1), \quad \sqrt{\lambda_k} a_k = 0.$$

Hence $a_k = 0$ except for the case $\sqrt{\lambda_k} = 0$. Hence

$$\|R_n f - f\|_p = o(1/\sqrt{\lambda_n}) \Leftrightarrow f = 0 \quad \text{or} \quad f = cv_k \quad \text{if} \quad \lambda_k = 0.$$

Analogously

$$\|R_n \tilde{f} - \tilde{f}\|_p = o(1/\sqrt{\lambda_n}) \Leftrightarrow f = 0 \quad \text{or} \quad f = cv_k \quad \text{if} \quad \lambda_k = 0 \Leftrightarrow \tilde{f} = 0.$$

In the next part of this paper we prove the equiconvergence a.e. between the trigonometric conjugate and Sturm–Liouville conjugate partial sums. For the nonorthogonal Dirichlet expansions the corresponding theorem was proved in [15].

Introduce the notation

$$\tilde{S}_\mu(f,x) := \sum_{\sqrt{\lambda_n}<\mu} a_n v_{1,n}.$$

Let $0 < R_0 < \min\{x-a, b-x\}$ and

$$\tilde{w}_R(x+t) := \begin{cases} -\frac{1-\cos\mu t}{\pi t}, & |t| < R \\ 0 & |t| > R. \end{cases}$$

Define the average operator

$$S_{R_0}[g(R)] := \frac{2}{R_0} \int_{R_0/2}^{R_0} g(R)\,\mathrm{d}R$$

and let

$$\tilde{S}_\mu^T(f,x) := \langle f, \tilde{w} \rangle, \qquad \tilde{w} := S_{R_0}[\tilde{w}_R].$$

Then

$$\tilde{S}_\mu^T(f,x) = S_{R_0}\left[\int_0^R \left(\frac{f(x-t)-f(x+t)}{\pi} \cdot \frac{1-\cos\mu t}{t}\right) \mathrm{d}t\right].$$

THEOREM 3. *Suppose* (2) *holds. Then for every* $f \in L^2$ *the difference* $\tilde{S}_\mu(f,x) - \tilde{S}_\mu^T(f,x)$ *converges locally uniformly on* G *to a continuous limit function from* $C(G)$.

In other words, the conjugate partial sums $\tilde{S}_\mu$ and $\tilde{S}_\mu^T$ are equiconvergent locally uniformly.

REMARK. The orthogonality of (v_n) and $(v_{1,n})$ implies that for $f \in L^p$ there exists $\tilde{f} \in L^2$. By the equiconvergence theorem of JOÓ and KOMORNIK [21], $\tilde{S}_\mu(f,x)$ and $S_\mu^T(\tilde{f},x)$ are equiconvergent in the strict sense (the difference converges to zero locally uniformly), hence we can reformulate Theorem 3 by stating that $S_\mu^T(\tilde{f},x)$ and $\tilde{S}_\mu^T(f,x)$ are equiconvergent, or, roughly speaking, that the trigonometric expansions of the Sturm–Liouville conjugate $\tilde{f}$ and of the trigonometric conjugate $\tilde{f}^T$ are equiconvergent.

PROOF OF THEOREM 3. Recall the asymmetric Titchmarsh formulas [21]. If $-v''+q(x)v=\lambda v(x)$ on G and x, $x+t\in G$ then

$$v(x+t)=v(x)\cos\sqrt{\lambda}t+v'(x)\frac{\sin\sqrt{\lambda}t}{\sqrt{\lambda}}+\int_x^{x+t} q(\xi)v(\xi)\frac{\sin\sqrt{\lambda}(t-|x-\xi|)}{\sqrt{\lambda}}\mathrm{d}\xi$$

if $x-t$, $x\in G$, then

$$v(x-t)=v(x)\cos\sqrt{\lambda}t-v'(x)\frac{\sin\sqrt{\lambda}t}{\sqrt{\lambda}}+\int_{x-t}^{x} q(\xi)v(\xi)\frac{\sin\sqrt{\lambda}(t-|x-\xi|)}{\sqrt{\lambda}}\mathrm{d}\xi$$

and by subtraction we obtain

$$v(x-t)-v(x+t)=$$
$$=-2v'(x)\frac{\sin\sqrt{\lambda}t}{\sqrt{\lambda}}+\left(\int_{x-t}^{x}-\int_{x}^{x+t}\right)q(\xi)v(\xi)\frac{\sin\sqrt{\lambda}(t-|x-\xi|)}{\sqrt{\lambda}}\mathrm{d}\xi.$$

Compute the coefficients of $\tilde{w}$ as follows

$$\langle v_n,\tilde{w}\rangle=S_{R_0}\left[\int_0^R[v_n(x-t)-v_n(x+t)]\frac{1-\cos\mu t}{\pi t}\mathrm{d}t\right]=$$
$$=-2v_n'(x)S_{R_0}\left[\int_0^R\frac{\sin\sqrt{\lambda_n}t}{\sqrt{\lambda_n}}\frac{1-\cos\mu t}{\pi t}\mathrm{d}t\right]+$$
$$+S_{R_0}\left[\int_0^R\left(\int_{x-t}^{x}-\int_{x}^{x+t}\right)q(\xi)v_n(\xi)\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{\sqrt{\lambda_n}}\mathrm{d}\xi\frac{1-\cos\mu t}{t}\mathrm{d}t\right]=$$
$$=:\Delta_n\frac{v_n'(x)}{\sqrt{\lambda_n}}+B_n.$$

Recall that here

$$q=\frac{(w^{1/2})''}{w^{1/2}}+b'.$$

In [21] is proved that

$$|\Delta_n+\delta(\mu,\sqrt{\lambda_n})|\leq\frac{c}{1+(\sqrt{\lambda_n}-\mu)^2R_0^2}\tag{21}$$

with a constant c independent of μ, n, R_0. In the next step we estimate B_n. Changing the order of the inner two integrals we get

(22) $B_n =$

$$= S_{R_0}\left[\left(\int_{x-R}^{x} - \int_{x}^{x+R}\right) q(\xi)v_n(\xi)\int_{|x-\xi|}^{R}\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)\cdot(1-\cos\mu t)}{\sqrt{\lambda_n}\cdot\pi\cdot t}\mathrm{d}t\mathrm{d}\xi\right].$$

Denote

$$\tilde{\gamma}_n^\mu := \int_{|x-\xi|}^{R}\frac{1-\cos\mu t}{t}\sin\sqrt{\lambda_n}(t-|x-\xi|)\mathrm{d}t$$

(the inner integral in (22)). To estimate it, consider first the case $\sqrt{\lambda_n} > \mu$. Integrating by parts we obtain

$$\tilde{\gamma}_n^\mu = \int_{|x-\xi|}^{R}\sin\sqrt{\lambda_n}(t-|x-\xi|)\mathrm{d}t\frac{1-\cos\mu|x-\xi|}{|x-\xi|} - \tag{23}$$

$$-\int_{|x-\xi|}^{R}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1-\cos\mu t}{t}\right)\int_{t}^{R}\sin\sqrt{\lambda_n}(\tau-|x-\xi|)\mathrm{d}\tau\mathrm{d}t =$$

$$= O\left(\frac{\mu}{\sqrt{\lambda_n}}\right) + \int_{|x-\xi|}^{R}\left|\frac{t\mu\sin\mu t-(1-\cos\mu t)}{t^2}\right|\mathrm{d}t\cdot O\left(\frac{1}{\sqrt{\lambda_n}}\right) =$$

$$O\left(\frac{\mu}{\sqrt{\lambda_n}}\right)\left(1+\int_{|x-\xi|}^{R}t^{-1}\mathrm{d}t\right) = O\left(\frac{\mu}{\sqrt{\lambda_n}}\ln\frac{1}{|x-\xi|}\right)$$

(we can suppose that $R<1$ and then $\ln\frac{1}{|x-\xi|}>0$). In case $\sqrt{\lambda_n}\le\mu$ we consider another integration by parts:

$$\int_{|x-\xi|}^{R}\cos\mu t\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{t}\mathrm{d}t = \frac{\sin\mu R}{\mu}\frac{\sin\sqrt{\lambda_n}(R-|x-\xi|)}{R} - \tag{24}$$

$$-\frac{\sqrt{\lambda_n}}{\mu}\int_{|x-\xi|}^{R}\sin\mu t\frac{\cos\sqrt{\lambda_n}(t-|x-\xi|)}{t}\mathrm{d}t +$$

$$+\frac{\sqrt{\lambda_n}}{\mu}\int\limits_{|x-\xi|}^{R}\sin\mu t\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{t\sqrt{\lambda_n}t}\mathrm{d}t=$$

$$=O(1/\mu)+O\left(\frac{\sqrt{\lambda_n}}{\mu}\right)\int\limits_{|x-\xi|}^{R}t^{-1}\mathrm{d}t=O\left(\frac{\sqrt{\lambda_n}}{\mu}\ln\frac{1}{|x-\xi|}\right);$$

$$\left|\int\limits_{|x-\xi|}^{R}\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{t}\mathrm{d}t\right|\le\int\limits_{|x-\xi|}^{R}t^{-1}\mathrm{d}t\le\ln\frac{1}{|x-\xi|},$$

$$\left|\int\limits_{|x-\xi|}^{R}\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{t}\mathrm{d}t\right|=-\left[\frac{\cos\sqrt{\lambda_n}(t-|x-\xi|)}{\sqrt{\lambda_n}}\frac{1}{t}\right]_{|x-\xi|}^{R}-$$

$$-\int\limits_{|x-\xi|}^{R}\frac{\cos\sqrt{\lambda_n}(t-|x-\xi|)}{\sqrt{\lambda_n}\cdot t^2}\mathrm{d}t=O\left(\frac{1}{\sqrt{\lambda_n}|x-\xi|}\right),$$

hence

$$\int\limits_{|x-\xi|}^{R}\frac{\sin\sqrt{\lambda_n}(t-|x-\xi|)}{t}\mathrm{d}t=O\left(\sqrt{\ln\frac{1}{|x-\xi|}\cdot\frac{1}{\sqrt{\lambda_n}|x-\xi|}}\right)=$$

$$=O\left(|x-\xi|^{-3/4}\cdot\lambda_n^{-1/4}\right).$$

Next we shall prove that the series

$$\sum\left|\left[\Delta_n+\delta(\mu,\sqrt{\lambda_n})\right]\frac{v_n'(x)}{\sqrt{\lambda_n}}\langle f,v_n\rangle\right|\le c\sum\frac{|v_n'(x)||\langle f,v_n\rangle|}{\sqrt{\lambda_n}(1+(\sqrt{\lambda_n}-\mu)^2R_0^2)}$$

converges for every fixed μ to a function bounded by a bound independent of μ and locally uniformly in $x\in G$. Indeed,

$$v_n'=(y_nw^{1/2})'=\sqrt{\lambda_n}v_{1,n}-\frac{b}{2}v_n.$$

We refer to the general square sum estimation of [22]. If

$$-u_n''+qu_n=\lambda_nu_n,\qquad q\in L^1_{\mathrm{loc}}(G)$$

and the (u_n) form a Bessel system in $L^2(G)$ then

$$\sup_{\mu>0}\sum_{|\mu-\sqrt{\lambda_n}|\le 1}\|u_k\|_{L^2(K)}<\infty$$

for every compact subinterval $K\subset G$. We can apply this for (v_n) and $(v_{1,n})$ to obtain

$$\sum\frac{|v_{1,n}(x)||\langle f,v_n\rangle|}{1+(\sqrt{\lambda_n}-\mu)^2R_0^2}=\sum_{l=0}^{\infty}\sum_{l\le\sqrt{\lambda_n}\le l+1}\frac{|v_{1,n}(x)||\langle f,v_n\rangle|}{1+(\sqrt{\lambda_n}-\mu)^2R_0^2}\le$$

$$\le\sum_{l=0}^{\infty}\frac{1}{1+l^2R_0^2}\left(\sum_{l\le\sqrt{\lambda_n}\le l+1}|v_{1,n}(x)|^2\right)^{\frac{1}{2}}\left(\sum_{l\le\sqrt{\lambda_n}\le l+1}|\langle f,v_n\rangle|^2\right)^{\frac{1}{2}}\le$$

$$\le c(K)\|f\|_2\sum_{l=0}^{\infty}\frac{1}{1+l^2R_0^2}\le\frac{c(K)}{R_0^2}\|f\|_2\quad\text{if}\quad x\in K.$$

Next we estimate the convergence of $\sum\langle f,v_n\rangle B_n$. We know that

$$|B_n|=\left|S_{R_0}\left[\left(\int_{x-R}^{x}-\int_{x}^{x+R}\right)q(\xi)v_n(\xi)\frac{\tilde{\gamma}_n^{\mu}(x,\xi)}{\pi\sqrt{\lambda_n}}\mathrm{d}\xi\right]\right|\le$$

$$\le\int_{x-R_0}^{x+R_0}|q(\xi)v_n(\xi)\tilde{\gamma}_n^{\mu}(x,\xi)|\mathrm{d}\xi\cdot\frac{1}{\sqrt{\lambda_n}}$$

so by (23)

$$\sum_{\sqrt{\lambda_n}>\mu}|\langle f,v_n\rangle B_n|\le\sum_{\sqrt{\lambda_n}>\mu}\frac{\mu}{\lambda_n}\int_{x-R_0}^{x+R_0}q(\xi)v_n(\xi)\ln\frac{1}{|x-\xi|}\mathrm{d}\xi\cdot\langle f,v_n\rangle\le$$

$$\le\sum_{l\ge\mu-1}\frac{\mu}{l^2}\int_{x-R_0}^{x+R_0}\left|q(\xi)\ln\frac{1}{|x-\xi|}\right|\sum_{l\le\sqrt{\lambda_n}\le l+1}|v_n(\xi)||\langle f,v_n\rangle|\mathrm{d}\xi\le$$

$$\le c(K,R_0)\|f\|_2\sum_{l\ge\mu-1}\frac{\mu}{l^2}\int_{x-R_0}^{x+R_0}\ln\frac{1}{|x-\xi|}\mathrm{d}\xi\le c(K,R_0)\|f\|_2.$$

We use (24) to estimate $\sum\limits_{\sqrt{\lambda_n}\le\mu}|\langle f,v_n\rangle B_n|$:

$$\sum_{\sqrt{\lambda_n}\le\mu}|\langle f,v_n\rangle|\frac{1}{\sqrt{\lambda_n}}\int_{x-R_0}^{x+R_0}|q(\xi)v_n(\xi)|\frac{\sqrt{\lambda_n}}{\mu}\ln\frac{1}{|x-\xi|}\mathrm{d}\xi\le$$

$$\le \frac{1}{\mu}\sum_{l\le\mu}\int_{x-R_0}^{x+R_0} q(\xi)\ln\frac{1}{|x-\xi|}\sum_{l\le\sqrt{\lambda_n}\le l+1}|v_n(\xi)||\langle f,v_n\rangle|\mathrm{d}\xi\le$$

$$\le\frac{c(K)\|f\|_2}{\mu}\mu\int_{x-R_0}^{x+R_0}|q(\xi)|\ln\frac{1}{|x-\xi|}\mathrm{d}\xi\le c(K,R_0)\|f\|_2,$$

$$\sum_{\sqrt{\lambda_n}\le\mu}|\langle f,v_n\rangle|\frac{1}{\sqrt{\lambda_n}}\int_{x-R_0}^{x+R_0}|q(\xi)v_n(\xi)||x-\xi|^{-3/4}\lambda_n^{-1/4}\mathrm{d}\xi\le$$

$$\le\sum_{l\le\mu}\frac{1}{l^{3/2}}\int_{x-R_0}^{x+R_0}|q(\xi)||x-\xi|^{-3/4}\sum_{l\le\sqrt{\lambda_n}\le l+1}|\langle f,v_n\rangle v_n(\xi)|\mathrm{d}\xi\le$$

$$\le c(K,R_0)\|f\|_2\sum_{l\le\mu}\frac{1}{l^{3/2}}\le c(K,R_0)\|f\|_2.$$

By the above estimations we get that

$$\left|\langle f,\tilde{w}\rangle+\sum\langle f,v_n\rangle\,\delta(\mu,\sqrt{\lambda_n})\frac{v_n'(x)}{\sqrt{\lambda_n}}\right|=$$

$$=\sum\langle f,v_n\rangle\left|B_n+\left(\Delta_n+\delta(\mu,\sqrt{\lambda_n})\right)\frac{v_n'(x)}{\sqrt{\lambda_n}}\right|\le c(K,R_0)\|f\|_2.$$

Since

$$\sum_{\sqrt{\lambda_n}=\mu}\left|\langle f,v_n\rangle\frac{v_n'(x)}{\sqrt{\lambda_n}}\right|=\sum_{\sqrt{\lambda_n}=\mu}|\langle f,v_n\rangle v_{1,n}(x)|+$$

$$+\frac{b(x)}{2}\sum_{\sqrt{\lambda_n}=\mu}|\langle f,v_n\rangle v_n(x)|\le c(K)\|f\|_2,$$

hence we finally obtain that

$$\left|\langle f,\tilde{w}\rangle+\sum_{\sqrt{\lambda_n}<\mu}\langle f,v_n\rangle\frac{v_n'(x)}{\sqrt{\lambda_n}}\right|\le c(K,R_0)\|f\|_2. \tag{25}$$

Our last estimate needed is

$$\left|\sum_{\sqrt{\lambda_n}<\mu}\langle f,v_n\rangle\frac{v_n(x)}{\sqrt{\lambda_n}}\right|\le\sum_{l\le\mu}\sum_{l\le\sqrt{\lambda_n}\le l+1}\left|\langle f,v_n\rangle\frac{v_n(x)}{\sqrt{\lambda_n}}\right|\le$$

$$\le c(K)\sum_{l\le\mu}\frac{1}{l+1}\left(\sum_{l\le\sqrt{\lambda_n}\le l+1}|\langle f,v_n\rangle|^2\right)^{1/2}\le$$

$$\le c(K)\left(\sum_{l\le\mu}\frac{1}{(l+1)^2}\right)^{\frac12}\left(\sum_{l\le\mu}\sum_{l\le\sqrt{\lambda_n}\le l+1}|\langle f,v_n\rangle|^2\right)^{\frac12}\le c(K)\|f\|_2$$

i.e.

$$(26)\quad |\tilde S^T(f,x)-\tilde S_\mu(f,x)|=\left|\langle f,\tilde w\rangle+\sum_{\sqrt{\lambda_n}<\mu}\langle f,v_n\rangle v_{1,n}\right|\le c(K,R_0)\|f\|_2,$$

in other words, the operators

$$A_\mu: f\to \tilde S^T_\mu f-\tilde S_\mu f,\qquad \left(A_\mu: L^2\to C(K)\right)$$

are uniformly bounded. If $f=\sum c_k v_k$ is a finite sum, then $\tilde S_\mu f=\tilde f$ for large μ, and since f is smooth, $\tilde S^T_\mu f$ tends to $\tilde f^T$ locally uniformly, see [23], Ch. II 6.8. Since the linear subspace spanned by the (v_k)-s is dense in L^2, by the Banach–Steinhaus theorem $A_\mu f$ tends uniformly on K to a function $Af\in C(K)$. Theorem 3 is proved. ∎

REMARK. We introduced above the notion of the analytic or power type function. Here we deduce (formally) the Cauchy–Riemann equations. Let $f\sim\sum a_k v_k$, $g\sim\sum b_k v_{1,k}$. Then introduce the functions

$$G_1(t,x):=\sum a_k e^{-\sqrt{\lambda_k}t}\cdot v_k(x),\qquad G_2(t,x):=-\sum b_k e^{-\sqrt{\lambda_k}t}\cdot v_{1,k}(x).$$

Then the function $f+ig$ is analytic if and only if the Cauchy–Riemann type equations

$$\frac{\partial G_1}{\partial t}=w^{-1/2}(x)\frac{\partial(w^{1/2}xG_2)}{\partial x},\qquad \frac{\partial G_2}{\partial t}=w^{1/2}(x)\frac{\partial(w^{-1/2}xG_1)}{\partial x}$$

hold. Indeed, both sides of the first equality are equal to

$$-\sum a_k\sqrt{\lambda_k}e^{-\sqrt{\lambda_k}t}v_k(x)$$

and both sides of the second one are equal to

$$-\sum b_k\sqrt{\lambda_k}e^{-\sqrt{\lambda_k}t}v_{1,k}(x).$$

References

[1] E. M. STEIN, *Topics in harmonic analysis* (related to the Littlewood–Paley theory), Annals of Mathematical Studies, Number 63, Princeton, New Jersey, 1970.

[2] A WAWRZYNCZYK, *Group representations and special functions*, Reidel, Dortrecht, 1984.

[3] M. A. NEUMARK, *Lineare differentialoperatoren,* Akademie-Verlag Berlin, 1967.

[4] E. HEWITT and K. ROSS, *Abstract harmonic analysis I, II,* Springer-Verlag, Berlin, 1963, 1970.

[5] B. MUCKENHOUPT and E. M. STEIN, Classical expansions and their relation to conjugate harmonic functions, *Trans. Amer. Math. Soc.,* **118** (1965), 17–92.

[6] B. MUCKENHOUPT, Mean convergence of Jacobi series, *Proc. Amer. Math. Soc.,* **23** (1969), 306–310.

[7] B. MUCKENHOUPT, Poisson integrals for Hermite and Laguerre expansions, *Trans. Amer. Math. Soc.,* **139** (1969), 231–242.

[8] B. MUCKENHOUPT, Hermite conjugate expansions, *Trans. Amer. Math. Soc.,* **139** (1969), 243–260.

[9] B. MUCKENHOUPT, Conjugate functions for Laguerre expansions, *Trans. Amer. Math. Soc.,* **147** (1970), 403–418.

[10] I. JOÓ, On the order of approximation by Fejér means of Hermite–Fourier and Laguerre–Fourier series, *Acta Math. Hung.,* **51** (1989), 365–370.

[11] I. JOÓ, On the divergence of eigenfunction expansions, *Annales Univ. Sci. Budapest.,* Sect. Math., **32** (1989), 3–36.

[12] I. JOÓ, Answer to some problems of M. Horváth, *Annales Univ. Sci. Budapest.,* Sect. Math., **31** (1988), 243–260.

[13] I. JOÓ, Saturation theorems for Hermite–Fourier series, *Acta Math. Hung.,* **57** (1991), 169–179.

[14] I. JOÓ, On Hermite–Fourier series, *Per. Math. Hungar.,* **24** (1992), 87–117.

[15] I. JOÓ, On the conjugate function of Dirichlet series, *Annales Univ. Sci. Budapest.,* Sect. Math., **35** (1992), 59–67.

[16] M. HORVÁTH, Some saturation theorems for classical orthogonal expansions I, *Per. Math. Hung.,* **22** (1991), 27–69.

[17] M. HORVÁTH, Some saturation theorems for classical orthogonal expansions II. *Acta. Math. Hung.,* **58** (1991), 139–169.

[18] G. SZEGŐ, *Orthogonal Polynomials,* New York, 1939.

[19] G. N. WATSON, *A treatise on the theory of Bessel functions,* Cambridge, University Press, 1952.

[20] H. BATEMAN and A. ERDÉLYI, *Hingher transcendental functions,* Vol. 1–2, New York, McGraw Hill, 1953.

[21] I. Joó and V. KOMORNIK, On the equiconvergence of expansions by Riesz bases formed by eigenfunctions of the Schrödinger operator, *Acta Sci. Math. Szeged*, **46** (1983), 357–375.

[22] A ZYGMUND, *Trigonometric series*, Cambridge Univ. Press, 1959.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 99–102

BESSEL-EXPONENTIAL PARTIAL DIFFERENTIAL EQUATION AND FOX'S H-FUNCTION

By

S. D. BAJPAI and SADHANA MISHRA

Department of Mathematics, University of Bahrain, Isa Town, BAHRAIN
V. B. R. I. Polytechnic Vidya Bhawan, Udaipur, INDIA

(Received August 29, 1991)

1. Introduction

The object of this paper is to formulate a two dimensional Bessel-Exponential partial differential equation and obtain its double series solution. We further present a particular solution of our Bessel-Exponential equation involving Fox's H-function. It is interesting to note that the particular solution also yields a new two dimensional series expansion for Fox's H-function involving Bessel functions and exponential functions.

The H-function introduced by FOX [6], pp. 408, will be represented as follows:

$$H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1,e_1), & \dots, & (a_p,e_p)\\ (b_1,f_1), & \dots, & (b_q,f_q)\end{matrix}\right.\right]\equiv H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_p,e_p)\\ (b_q,f_q)\end{matrix}\right.\right]. \tag{1.1}$$

The following formulae are required in the proof:

Theg H-function analogue of the integral [3] pp. 37, (2.2):

$$\int_0^\infty x^{w_1-1}J_a(x)K_a(x)H_{p,q}^{m,n}\left[zx^{4h}\left|\begin{matrix}(a_p,e_p)\\ (b_q,f_q)\end{matrix}\right.\right]\mathrm{d}x= \tag{1.2}$$

$$=2^{w_1-2}H_{p+3,q}^{m,n+2}\left[2^{4h}\left|\begin{matrix}(1-\frac{w_1}{2},2h), & (1-\frac{w_1}{2}-\frac{a}{2},h), & (a_p,e_p),\\ (1-\frac{w_1}{2}+\frac{a}{2},h) & & \\ (b_q,f_q) & & \end{matrix}\right.\right],$$

where $h>0$, $\sum\limits_{j=1}^{p}e_j-\sum\limits_{j=1}^{q}f_j\equiv A\le 0$, $\sum\limits_{j=1}^{n}e_j-\sum\limits_{j=n+1}^{p}e_j+\sum\limits_{j=1}^{m}f_j-\sum\limits_{j=m+1}^{q}f_j\equiv B>0$, $|\arg z|<\frac{1}{2}B\pi$, $Re(w_1+2a)+4h\min\limits_{1\le j\le m}[Re b_j/f_j]>0$.

The integral [2] pp. 46, (2.1):

$$\int_0^{\pi} e^{-2ivy}(\sin y)^{w_2-1} H_{p,q}^{m,n}\left[z(\sin y)^{-2k}\middle|\begin{matrix}(a_p,e_p)\\(b_q,f_q)\end{matrix}\right]\mathrm{d}y= \tag{1.3}$$

$$=\frac{\pi e^{-\pi iv}}{2^{w_2-1}} H_{p+2,q+1}^{m+1,n}\left[z2^{2k}\middle|\begin{matrix}(a_p,e_p), & \left(\frac{w_2+1}{2}\pm v,k\right)\\(w_2,2k), & (b_q,f_q)\end{matrix}\right],$$

where $k>0$, $A\leq 0$, $B>0$; $|\arg z|<\frac{1}{2}B\pi$, $Re w_2-2k\max\limits_{1\leq j\leq n}[Re(a_j-1)/e_j]:=0$.

The orthogonality property of the Bessel functions [7] pp. 291, (6):

$$\int_0^{\infty} x^{-1}J_{v+2n+1}(x)J_{v+2m+1}(x)\mathrm{d}x= \tag{1.4}$$

$$=\begin{cases}0, & \text{if } m\neq n;\\ (4n+2v+2)^{-1}, & \text{if } m=n,\ Rev+m+n>-1.\end{cases}$$

2. Two dimensional Bessel-Exponential partial differential Equation

Let us consider

$$\frac{\partial u}{\partial t}=x^2\frac{\partial^2 u}{\partial x^2}+x\frac{\partial u}{\partial x}+x^2u+c\frac{\partial^2 u}{\partial y^2}, \tag{2.1}$$

where $u\equiv u(x,y,t)$ and $u(x,y,0)=f(x,y)$.

To solve (2.1), we assume that (2.1) has a solution of the form:

$$u(x,y,t)=e^{(v+2r+1)^2t+4cs^2t}X(x)Y(iy). \tag{2.2}$$

The substitution of (2.2) into (2.1) yields:

$$Y[x^2X''+xX'+\{x^2-(v+2r+1)^2\}Y]-cX[Y''+4s^2Y]=0. \tag{2.3}$$

We see that $x^2X''+xX'+\{x^2-(v+2r+1)^2\}X=0$ is Bessel equation [1] pp. 200, (6.25), with solution $X=J_{v+2r+1}(x)$, and $Y''+4s^2Y=0$ has a solution $Y=e^{2isy}$. Therefore the solution of (2.1) is of the form:

$$u(x,y,t)=e^{(v+2r+1)^2t+4cs^2t}J_{v+2r+1}(x)e^{2isy}. \tag{2.4}$$

In view of the principle of superposition, the general solution of (2.1) is given by

$$u(x,y,t)=\sum_{r=0}^{\infty}\sum_{s=-\infty}^{\infty}A_{r,s}e^{(v+2r+1)^2t+4cs^2t}J_{v+2r+1}(x)e^{2isy}. \tag{2.5}$$

In (2.5), putting $t=0$, we get

$$f(x,y)=\sum_{r=0}^{\infty}\sum_{s=-\infty}^{\infty} A_{r,s}J_{v+2r+1}(x)e^{2isy}. \tag{2.6}$$

Multiplying both sides of (2.6) by $x^{-1}J_{v+2u+1}(x)e^{-2iwy}$, integrating with respect to y from 0 to π and with respect to x from 0 to ∞, then using (1.4) and the orthogonality property of exponential functions, the Fourier Bessel-Exponential coefficient are given by

$$A_{r,s}=\frac{2}{\pi}(v+2r+1)\int_0^{\infty}\int_0^{\pi} f(x,y)x^{-1}J_{v+2r+1}(x)e^{-2isy}\,\mathrm{d}y\mathrm{d}x. \tag{2.7}$$

In view of the theory of double and multiple Fourier series given by CARSLAW and JAEGER [4] pp. 180–183, and many other references, such as ERDÉLYI [5] pp. 64–65 etc., the double series (2.6) is convergent provided the function $f(x,y)$ is defined in the region $0<x<\infty$, $0<y<\pi$. In brief, the double series (2.6) normally converges, if the double integral on the right hand side of (2.7) exists.

In the subsequent section, we take $f(x,y)$ as Fox's H-function and present another method to obtain Fourier Bessel-Exponential coefficients $A_{r,s}$.

3. Particular solution involving Fox's H-function

The particular solution to be obtained is

$$u(x,y,t)= \tag{3.1}$$

$$=2^{w_1-w_2}\sum_{r=0}^{\infty}\sum_{s=-\infty}^{\infty}(v+2r+1)e^{(v+2r+1)^2t+4cs^2t+2is(y-\pi/2)}\times H^{m+1,n+2}_{p+5,q+1}\cdot$$

$$\cdot\left[2^{2(2h+k)}z\left|\begin{matrix}(1-\frac{w_1}{2},2h), & \left(\frac{1}{2}-\frac{w_1}{2}-\frac{v}{2}-r,h\right), & \\ (a_p,e_p), & (3/2-\frac{w_1}{4}+\frac{v}{2}+r,h), & \left(\frac{w_2+1}{2}\pm s,k\right)\\ (w_2,2k), & (b_q,f_q) & \end{matrix}\right.\right]J_{v+2r+1}$$

valid under the conditions of (1.2), (1.3) and (1.4).

PROOF. Let

$$f(x,y)=x^{w_1}K_{v+2u+1}(x)(\sin y)^{w_2-1}H^{m,n}_{p,q}\left[zx^{4h}(\sin y)^{-2k}\left|\begin{matrix}(a_p,e_p)\\(b_q,f_q)\end{matrix}\right.\right]= \tag{3.2}$$

$$=\sum_{r=0}^{\infty}\sum_{s=-\infty}^{\infty}A_{r,s}J_{v+2r+1}(x)e^{2isy}.$$

Equation (3.2) is valid, since $f(x,y)$ is defined in the region $0 < x < \infty$, $0 < y < \pi$.

Multiplying both sides of (3.2) by e^{-2iwy} and integrating with respect to y from 0 to π, then using (1.3) and the orthogonality property of exponential functions. Now multiplying both sides of the resulting expression by $x^{-1} J_{v+2u+1}(x)$ and integrating with respect to x from 0 to ∞, then using (1.2) and (1.4), we obtain the value of $A_{r,s}$. Substituting this value of $A_{r,s}$ in (3.2), the solution (3.1) is obtained.

NOTE. If we put $t=0$ in (3.1), it reduces to a new two dimensional series expansion for Fox's H-function involving Bessel functions and Exponential functions.

Since on specializing the parameters Fox's H-function yields almost all special functions appearing in applied mathematics and physical sciences. Therefore, the result (3.1) presented in this paper is of a general character and hence may encompass several cases of interest.

References

[1] L. C. ANDREWS, *Special functions for engineers and applied mathematicians*, Macmillan Publishing Co., New York, (1985).

[2] S. D. BAJPAI, Some results involving Fox's H-function, *Portugal. Math.*, **30** (1971), 45–52.

[3] S. D. BAJPAI, Expansion formulae for the products of Meijer's G-functions and Bessel functions, *Portugal. Math.*, **33** (1974), 35–41.

[4] H. S. CARSLAW, and J. C. JAEGER, *Conduction of heat in solids* (2nd Ed.), Clarendon Press, Oxford, 1986.

[5] A. ERDÉLYI, *Higher transcendental functions*, Vol. 2., McGraw–Hill, New York, (1953).

[6] C. FOX, The G and H-functions as symmetrical Fourier kernels, *Trans. Amer. Math. Soc.*, **98** (1961), 395–429.

[7] Y. L. LUKE, *Integrals of Bessel functions*, McGraw–Hill, New York, (1962).

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 103–107

LAGRANGIANS AND SPRAYS

By

J. SZENTHE

Department of Geometry, L. Eötvös University, Budapest

(Received September 4, 1991)

Let M be a C^∞ manifold and $L: TM \to \mathbb{R}$ a *Lagrangian function* which is C^∞ on $\check{T}M = TM - \{O_{TM}\}$ where O_{TM} is the zero section of TM. The Lagrangian function L gives rise to the *Euler-Lagrange equation*

$$\iota_X d\hat{d}L - d(L - AL) = 0$$

where $\hat{d}: \wedge(TM) \to \wedge(TM)$ is the vertical differential, $A: TM \to TTM$ the Liouville field of TM, and the "unknown" is a C^∞ vector field $X: \check{T}M \to TTM$ ([B] pp.14–25; [G] pp.159–168). If a vector field X exists which satisfies the above equation then X is said to be a *Lagrangian field* of the Lagrangian function L. If the 2-form $d\hat{d}L$ is non-degenerate then L is said to be *regular*; in this case L has obviously a unique Lagrangian field. Moreover, the Lagrangian field of a regular Lagrangian function is a second-order differential equation according to a fundamental result ([B] p.26; [AM] pp.213–216). A vector field $Z: \check{T}M \to TTM$ which is a second-order differential equation is called a *spray* provided that it is homogeneous of degree 2, i.e. it satisfies the condition $\mathcal{L}_A Z = [A, Z] = Z$. The kinetic energy function of a Riemannian metric as a Lagrangian function has a Lagrangian field which is a spray ([B] pp.28–29) and analogous statement holds in case of a Finsler metric [W]. The problem, to find the Lagrangian functions having Lagrangian fields which are sprays, is studied below.

First a concept is introduced which will be applied subsequently. Let $X: \check{T}M \to TTM$ be a vector field then $\mathbf{L}(X)$ is the set of those Lagrangian function which have X as Lagrangian field. Since the Euler–Lagrange equation

$$\iota_X d\hat{d}L - d(L - AL) = 0$$

is linear in L, the set $\mathbf{L}(X)$ is a vector space over $\mathbb{R}$. However as the following Proposition 1 shows $\mathbf{L}(X)$ can be endowed canonically with a richer structure as well. In the proof of Proposition 1 an equivalent form of the Euler–Lagrange equation is utilized; this form is as follows:

$$(\mathcal{L}_A - \iota_X \hat{d})dL = dL.$$

The above form of the Euler–Lagrange equation can be obtained by taking into account the facts that d and $\hat{d}$ anticommute ([G] pp.163–164) and that $\mathcal{L}_A$ and d commute.

PROPOSITION 1. *Let the C^∞ vector field $X:\check{T}M \to TTM$ be a second-order differential equation. If L, $L' \in \mathbf{L}(X)$ then $LL' \in \mathbf{L}(X)$ holds as well.*

PROOF. Since L and L' are elements of $\mathbf{L}(X)$, the vector field X satisfies the Euler–Lagrange equations for L and L':

$$(\mathcal{L}_A - \iota_X \hat{d})dL = dL, \quad (\mathcal{L}_A - \iota_X d)dL' = dL'.$$

Since X is a second-order differential equation, the values of X are jets ([B] pp.25–26) and consequently

$$(\mathcal{L}_A - \iota_X \hat{d})L = 0, \quad (\mathcal{L}_A - \iota_X \hat{d})L' = 0.$$

But then the following equalities are valid as well

$$\begin{aligned}(\mathcal{L}_A - \iota_X \hat{d})d(LL') &= (\mathcal{L}_A - \iota_X \hat{d})(LdL' + L'dL) = \\ &= \left\{(\mathcal{L}_A - \iota_X \hat{d})L\right\} dL' + L\left\{(\mathcal{L}_A - \iota_X \hat{d})dL'\right\} + \left\{(\mathcal{L}_A - \iota_X \hat{d})L'\right\} dL + \\ &\quad + L'\left\{(\mathcal{L}_A - \iota_X \hat{d})dL\right\} = LdL' + L'dL = d(LL').\end{aligned}$$

Therefore, X satisfies the Euler–Lagrange equation for LL' which means that $LL' \in \mathbf{L}(X)$ holds.

Let $X:\check{T}M \to TTM$ be a second-order differential equation. Then $\mathbf{L}(X)$ is called the *algebra of Lagrangian functions associated with* X by reason of the preceding proposition.

PROPOSITION 2. *Let the C^∞ vector field $X:\check{T}M \to TTM$ be a spray. Then $L \in \mathbf{L}(X)$ implies that $AL \in \mathbf{L}(X)$. Conversely, if $X:\check{T}M \to TTM$ is a C^∞ vector field such that there is a regular Lagrangian function $L \in \mathbf{L}(X)$ with $AL \in \mathbf{L}(X)$ then X is a spray.*

PROOF. Let $Z:\check{T}M \to TTM$ be a C^∞ vector field and $L \in \mathbf{L}(Z)$. Then in consequence of some well-known basic identities and of the identity $\hat{d}\mathcal{L}_A - \mathcal{L}_A\hat{d} = [\hat{d}, \mathcal{L}_A] = \hat{d}$ ([G] pp.163–164) the following holds:

$$\begin{aligned}\iota_{[A,Z]}d\hat{d}L &= [\mathcal{L}_A, \iota_Z]d\hat{d}L = \mathcal{L}_A\iota_Z d\hat{d}L - \iota_Z\mathcal{L}_A d\hat{d}L = \\ &= \mathcal{L}_A(d((L - AL)) - \iota_Z d\mathcal{L}_A\hat{d}L = d(AL - A(AL)) - \iota_Z d\hat{d}(AL) + \\ &\quad + \iota_Z d(\hat{d}\mathcal{L}_A L - \mathcal{L}_A\hat{d}L) = d(AL - A(AL)) - \iota_Z d\hat{d}(AL) + \iota_Z d\hat{d}L.\end{aligned}$$

Assume now that Z is a spray; then $[A,Z]=Z$ holds. Consequently, the above equality yields the following one

$$d(AL-A(AL))-\iota_Z d\hat{d}(AL)=0.$$

But then $AL\in \mathbf{L}(Z)$ is valid. Thus the first assertion of the proposition is proved.

Assume now that Z is a C^∞ vectorfield and that L, $AL\in \mathbf{L}(Z)$ where L is regular. Then the above equality implies now that

$$\iota_{[A,Z]}d\hat{d}L=\iota_Z d\hat{d}L.$$

But since L is regular, now $[A,Z]=Z$ must be valid; which means that Z is homogeneous of degree 2. Since $L\in \mathbf{L}(Z)$ is regular, Z must be a second-order differential equation by a basic result already cited above. Thus Z is a spray.

COROLLARY. *Let $X:\check{T}M\to TTM$ be a C^∞ vector field such that $\mathbf{L}(X)$ contains a regular Lagrangian function which is homogeneous. Then X is a spray.*

The following theorem yields a converse to the preceding Corollary under a strong assumption postulating the existence of analytic Lagrangian function. The concept of higher order differential is applied here in the sense of Ambrose–Palais–Singer [APS].

THEOREM. *Let M be a real analytic manifold and $X:\check{T}M\to TTM$ a C^∞ vector field which is a spray. If the algebra $\mathbf{L}(X)$ of Lagrangian functions associated with X contains a Lagrangian function L which is analytic on TM and its k-th differential does not reduce to the $(k-1)$-th on O_{TM} then it contains also such one which is homogeneous of degree k.*

PROOF. Let $(x^1,\dots,x^m)$ be a coordinate system of M and $(x^1,\dots,x^m;\dot{x}^1,\dots,\dot{x}^m)$ the induced coordinate system of TM. Then the analytic Lagrangian function L is given by a function $L(x^1,\dots,x^m;\dot{x}^1,\dots,\dot{x}^m)$ in the above coordinate system. Consider now the functions $\tilde{P}_n$ defined in the above induced coordinate system as follows:

$$\tilde{P}_n(x^1,\dots,x^m;\dot{x}^1,\dots,\dot{x}^m)=\sum_{i_1,\dots,i_n=1}^{m}\dot{x}^{i_1}\dots\dot{x}^{i_n}\frac{\partial^n\tilde{L}}{\partial x^{i_1}\dots\partial x^{i_n}}$$

where $n\in\mathbf{N}$. These function are the coordinate expression of the functions $P_n:U\to\mathbb{R}$ on the domain $U\in TM$ of the induced coordinate system. A simple direct calculation yields that the derivatives of the above functions with respect to the Liouville field A can be given as follows:

$$AP_n=nP_n+P_{n+1},\qquad n\in\mathbf{N}.$$

But then the functions P_n, $n \in \mathbf{N}$ can be defined directly everywhere on TM independently of any coordinate system; actually, $P_1 = AL$ and $P_n = AP_{n-1} - (n-1)P_{n-1}$ for $n \geq 2$. Moreover, $P_n \in \mathrm{L}(X)$ for $n \in \mathbf{N}$, since $P_1 = AL \in \mathrm{L}(X)$ by Proposition 2 and an obvious inductive argument yields that $P_n \in \mathrm{L}(X)$ for $n \geq 2$ considering the above expression of P_n and Proposition 2. Consider now for each $k \in \mathbf{N}$ the function H_k which is defined as follows:

$$H_k = \sum_{n=k}^{\infty} \frac{(-1)^{n-k}}{(n-k)!} P_n.$$

In order to justify the above definition it will be shown that since L is analytic, the series in the definition of H_k is uniformly convergent, and therefore H_k is also analytic. Actually, since the function $\tilde{L}$ is analytic, an absolutely and uniformly convergent power series is given by

$$\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{i_1,\ldots,i_n=1}^{m} \left| \frac{\partial^n \tilde{L}}{\partial \dot{x}^{i_1} \ldots \partial \dot{x}^{i_n}} \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m\right) \right| r^n$$

and consequently its sum $s(r)$ is an analytic function. But then the analytic function

$$r^k \frac{d^k s}{dr^k} = \sum_{n=k}^{\infty} \frac{1}{(n-k)!} \left| \frac{\partial^n \tilde{L}}{\partial \dot{x}^{i_1} \ldots \partial \dot{x}^{i_n}} \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m\right) \right| r^n$$

is given by an absolutely and uniformly convergent power series. Consequently, the function series

$$\tilde{S}_k \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m;\xi^1,\ldots,\xi^m\right) =$$

$$= \sum_{n=k}^{\infty} \sum_{i_1,\ldots,i_n=1} \frac{1}{(n-k)!} \frac{\partial^n \tilde{L}}{\partial \dot{x}^{i_1} \ldots \partial \dot{x}^{i_n}} \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m\right) \xi^{i_1} \ldots \xi^{i_n}$$

is absolutely and uniformly convergent in the variables $x^1,\ldots,x^m$; $\dot{x}^1,\ldots,\dot{x}^m$; $\xi^1,\ldots,\xi^m$, therefore $\tilde{S}_k$ is analytic. But then by

$$\tilde{H}_k \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m\right) = \tilde{S}_k \left(x^1,\ldots,x^m;\dot{x}^1,\ldots,\dot{x}^m;-\dot{x}^1,\ldots,-\dot{x}^m\right)$$

the existence of H_k follows. By the same reason the following holds as well:

$$AH_k = \sum_{n=k}^{\infty} \frac{(-1)^{n-k}}{(n-k)!} (nP_n + P_{n+1}) = k \sum_{n=k}^{\infty} \frac{(-1)^{n-k}}{(n-k)!} P_n = kH_k.$$

The fact that H_k is not identically 0 follows by the assumption that the k-th differential of L does not reduce to the $(k-1)$-th on O_{TM}. In fact assume that H_k is identically zero, then for its coordinate expression $\tilde{H}_k$ in an induced coordinate system

$$\frac{\partial^k \tilde{H}_k}{\partial x^{i_1} \dots \partial x^{i_k}} = 0$$

holds for any $1 \leq i_1, \dots, i_k \leq m$. But on the intersection of the domain of the induced coordinate system with the zero section O_{TM} the following holds

$$\frac{\partial^k \tilde{H}_k}{\partial x^{i_1} \dots \partial x^{i_k}} \left(x^1, \dots, x^m; 0, \dots, 0\right) = C \frac{\partial^k \tilde{L}}{\partial x^{i_1} \dots \partial x^{i_k}} \left(x^1, \dots, x^m; 0, \dots, 0\right).$$

Thus if H_k vanishes identically on TM then the k-th differential of L on O_{TM} reduces to its $(k-1)$-th differential in contradiction with the assumption. Consequently, the function H_k is homogeneous of degree k. The fact that

$$(\mathscr{L}_A - \iota_X \hat{d}) dH_k = dH_k$$

holds, is a consequence of the expression of H_k in terms of the functions P_n and of $P_n \in \mathbf{L}(X)$, $n \in \mathbf{N}$.

References

[AM] R. Abraham, and J. Marsden, *Foundations of Mechanics*, Reading, Massachusetts, 1978.

[APS] W. Ambrose, R. S. Palais, and I. M. Singer, Sprays, *An. Acad. Brasil.*, **32** (1960), 163–178.

[B] A. L. Besse, *Manifolds all of whose Geodesics are Closed*, Berlin, 1978.

[G] C. Godbillon, *Géométrie Différentielle et Mécanique Analytique*, Paris, 1969.

[W] F. W. Warner, The conjugate locus of a Riemannian manifold, *Amer. J. Math.*, **87** (1965), 575–604.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 109–127

METRISCHE RESULTATE IN KEGELSCHNITTBÜSCHELN DER ISOTROPEN EBENE MIT VIER REELLEN GRUNDPUNKTEN VON DENEN MINDESTENS ZWEI PARALLEL SIND

Von

KATHARINA MOLNÁR-SÁSKA

"Ybl Miklós" Technische Fachhochschule, Budapest

(Eingegangen am 4. Oktober, 1991)

1. Einleitung

Die Klassifikation der Kegelschnittbüschel mit vier reellen Grundpunkten ist in der isotropen Ebene mit synthetischen Methoden erstmals von VL. ŠČURIČ [4] und später mit analytischen Methoden von der Verfasserin dieses Artikels [14] in Angriff genommen worden. In der letzteren Arbeit kann man die vollständige Klassifikation und die Bestimmung von Invariantensystemen der Kegelschnittbüschel mit vier reellen Grundpunkten nachlesen, wobei mindesten zwei Grundpunkte parallel sind.

Bei der Untersuchung der Kegelschnittbüschel der euklidischen Ebene bestimmte KANTOR die Längen der Achsen des Mittelpunktskegelschnittes [9]; weiters haben STEINER [8] und BOBEK [12] folgende schöne Sätze gefunden:

1. Die Büschelkurven sind paarweise ähnlich.
2. Je 6 Kurven haben gleiches Achsenprodukt.
3. Die Brennpunktskurve enthält die Ecken des allen Kegelschnitten konjugierten Poldreiecks sowie die Höhenfußpunkte dieses Dreiecks.

Analoge Fragestellungen sind auch hier naheliegend.

In diesem Artikel verwenden wir die Bezeichnungen der metrischen Klassifikation aus [14] und untersuchen nur die Typen $\mathbb{I}_A 8$, 9, 10, 11, 12, 13, 14, 15, d.h. jene wo die konvexe Hülle der vier Grundpunkte ein nichtzulässiges Dreieck BDA der isotropen Ebene I_2 ist, die Punkte B, D parallel sind und $s(DB) > 0$ gilt.

Wir bezeichnen die Halbierungspunkte der Grundstrecken BD, DA, AC, AB, BC, DC mit A_1, B_1, C_1, D_1, E_1, G_1.

Man kann dann durch eine isotrope Bewegung erreichen, daß D in den Ursprung und die Grundgerade DA in die x-Achse des zugrundegelegten

Koordinatensystems fällt. Dann sind die Koordinaten der Grundpunkte

(1.1) $D(0,0);\quad A(a,0);\quad B(0,b)\quad \text{und}\quad C(c_1,c_2)$

wobei $a,\ c_1 \in \mathbb{R}\setminus\{0\}$, $b,\ c_2 \in \mathbb{R}^+$, aber $a \neq c_1$, $b \neq c_2$ ist; weiter gelten die Bedingungen $ac_1 > 0$ und $c_2 < -\frac{b}{a}c_1 + b$.

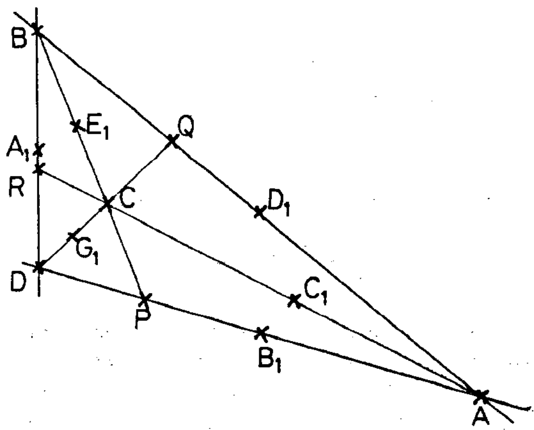

Fig. 1.

Wie beim Typ $\mathbb{I}_A 2$, 3, 4, 5, 6, 7 zeigt man leicht, daß sich als Normalform der Typen $\mathbb{I}_A 8$, 9, 10, 11, 12, 13, 14, 15 die Gleichung

$$(1.2)\qquad F \equiv \lambda(xy - k_1x^2 + k_1ax) + y^2 - k_2xy - by = 0$$

einstellt, wobei die Büschelinvarianten a, b, k_1, k_2 die gleichen geometrischen Bedeutungen haben wie bei den Typen $\mathbb{I}_A 2$, 3, 4, 5, 6, 7, d.h. es gelten die Gleichungen

$$(1.3)\qquad k_1 = \frac{c_2}{c_1 - a} \quad \text{und} \quad k_2 = \frac{c_2 - b}{c_1}.$$

Wegen (1.1) und (1.3) bestehen jedoch die folgenden Bedingungen

$$(1.4)\qquad c_1 = \frac{k_1a + b}{k_1 - k_2}; \qquad c_2 = \frac{k_1k_2a + k_1b}{k_1 - k_2} \qquad \text{bzw.}$$

Ungleichungen

$$(1.5)\qquad ak_1(ak_1 + b) < 0, \quad b > 0, \quad ak_2 + b < 0$$

$$(1.6)\qquad k_1^2 - k_1k_2 < 0.$$

Mittels (1.2) und (1.6) kann man sehen, daß die genannten Kegelschnittbüschel Hyperbeln 1. Art, Hyperbeln 2. Art und schneidende Geradenpaaren enthalten. Allerdings enthält das Büschel jetzt keine Ellipsen und auch keine reellen Parabeln. Genauer erkennt man aus (1.2) mittels $\frac{\partial F}{\partial y} = 0$, daß zu den Parameterwerten $\lambda = \infty$, $\lambda_3 = 0$ und $\lambda_4 = \frac{k_1 k_2 ab + k_1 b^2}{k_1^2 a^2 + k_1 ab}$ die schneidenden Geradenpaare mit den Mittelpunkten R, P, Q gehören. Zu $\lambda \in (\lambda_3, \lambda_4)$ gehören Hyperbeln 1. Art, zu allen anderen Werten von $\lambda \in \mathbb{R}$ gehören Hyperbeln 2. Art.

Mittels $\frac{\partial F}{\partial x} = \frac{\partial F}{\partial y} = 0$ bestimmt man aus (1.2) die Mittelpunktskurve. Man findet

$$(1.7) \quad -2k_1k_2x^2 - 2y^2 + 4k_1xy + (k_1k_2a - 2k_1b)x + (b - 2k_1a)y + k_1ab = 0,$$

und diese Kurve ist wegen (1.6) eine Ellipse. Ein anderer Ausdruck für die Mittelpunktsellipse ist

$$(1.8) \quad x(\lambda) = \frac{k_2b - \lambda b - 2\lambda k_1 a}{-k_2^2 + 2\lambda k_2 - 4\lambda k_1 - \lambda^2}; \quad y(\lambda) = \frac{-2\lambda k_1 b - \lambda k_1 k_2 a + \lambda^2 k_1 a}{-k_2^2 + 2\lambda k_2 - 4\lambda k_1 - \lambda^2},$$

wobei wegen (1.6) für $\lambda \in \mathbb{R}$

$$(1.9) \qquad -k_2^2 + 2\lambda k_2 - 4\lambda k_1 - \lambda^2 < 0 \quad \text{gilt.}$$

Wegen (1.9) sind $x(\lambda)$ bzw. $y(\lambda)$ für $\lambda \in \mathbb{R}$ stetige Funktionen und es gilt $\lim_{|\lambda| \to \infty} x(\lambda) = 0$, $\lim_{|\lambda| \to \infty} y(\lambda) = -k_1 a$. Dieser Punkt hat die Koordinaten $(0, -k_1 a)$ und ist gerade R. Man zeigt leicht, daß die Mittelpunktsellipse auch die Punkte P, Q enthält. Aus (1.8) kann man sehen, daß die Mittelpunkte der Hyperbeln 1. Art auf einem Bogen und jene der Hyperbeln 2. Art auf dem anderen Bogen der Mittelpunktsellipse liegen; diese Bögen werden durch die Punkte P und Q begrenzt.

Nach einiger Rechnung ergibt sich, daß die Punkte E_1, G_1 bzw. C_1 die Mittelpunkte von Hyperbeln 2. Art bzw. 1. Art im Büschel sind. Wir untersuchen die Halbierungspunkte A_1, B_1, D_1. Man kann sofort sehen, daß A_1 der Mittelpunkt einer Hyperbel 2. Art ist, falls

$$(1.10) \qquad S_4 \equiv k_1 + \frac{b}{2a}$$

ungleich Null ist. Für $S_4 = 0$ ist A_1 der Mittelpunkt eines schneidenden Geradenpaares, denn es ist $A_1 \equiv R$. Weiter kann man sehen, daß D_1 der Mittelpunkt des Kegelschnittes

$$(1.11) \qquad K \equiv y^2 - by - k_1k_2x^2 + k_1k_2ax = 0$$

des Büschels ist. Mittels (1.11) und $\frac{\partial K}{\partial y} = 0$ zeigt man leicht, daß der vorliegende Kegelschnitt eine Hyperbel 1. Art, ein schneidendes Geradenpaar $(D_1 \equiv Q)$ oder eine Hyperbel 2. Art ist, je nachdem der Ausdruck

$$S_1^* \equiv k_1 k_2 a^2 - b^2 \tag{1.12}$$

positiv, Null oder negativ ist. Ebenso zeigt man leicht, daß B_1 der Mittelpunkt des Kegelschnittes

$$K_1 \equiv y^2 - k_2 xy - by + \frac{k_2 a + 2b}{a}(xy - k_1 x^2 + k_1 ax) = 0 \tag{1.13}$$

des Büschels ist. Mittels $\frac{\partial K_1}{\partial y} = 0$ und (1.13) sieht man sofort, daß dieser Kegelschnitt eine Hyperbel 1. Art, ein schneidendes Geradenpaar $(P \equiv B_1)$ oder eine Hyperbel 2. Art ist, je nachdem der Ausdruck

$$S_3 \equiv k_2 a + 2b \tag{1.14}$$

negativ, Null oder positiv ist.

Durch die Diskussion der 6 Halbierungspunkte ergeben sich die folgende 8 Untertypen des Typs $\mathbb{I}_A$

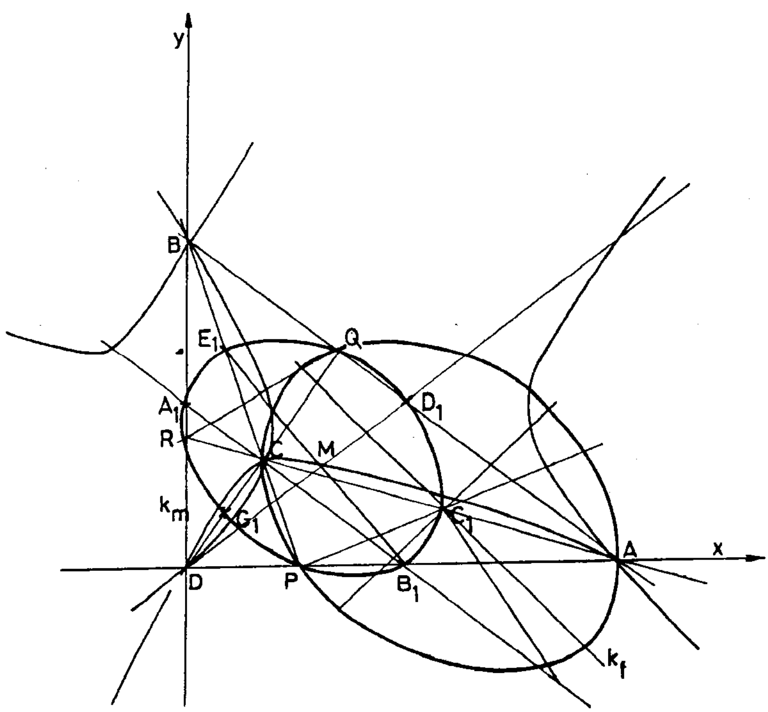

Fig. 2.

Unterfälle:

$\mathrm{II}_A 8.\ S_4 \neq 0,\ S_1^* > 0,\ S_3 < 0.$

Die isotrope Grundgerade schneidet die Mittelpunktsellipse in 2 verschiedenen Punkten und die beiden Punkte B_1, D_1 sind die Mittelpunkte von Hyperbeln 1. Art. (Fig. 2.)

$\mathrm{II}_A 9.\ S_4 \neq 0,\ S_1^* < 0,\ S_3 > 0.$

Die isotrope Grundgerade schneidet die Mittelpunktsellipse in zwei verschiedenen Punkten und die beiden Punkte B_1, D_1 sind die Mittelpunkte von Hyperbeln 2. Art. (Fig. 3.)

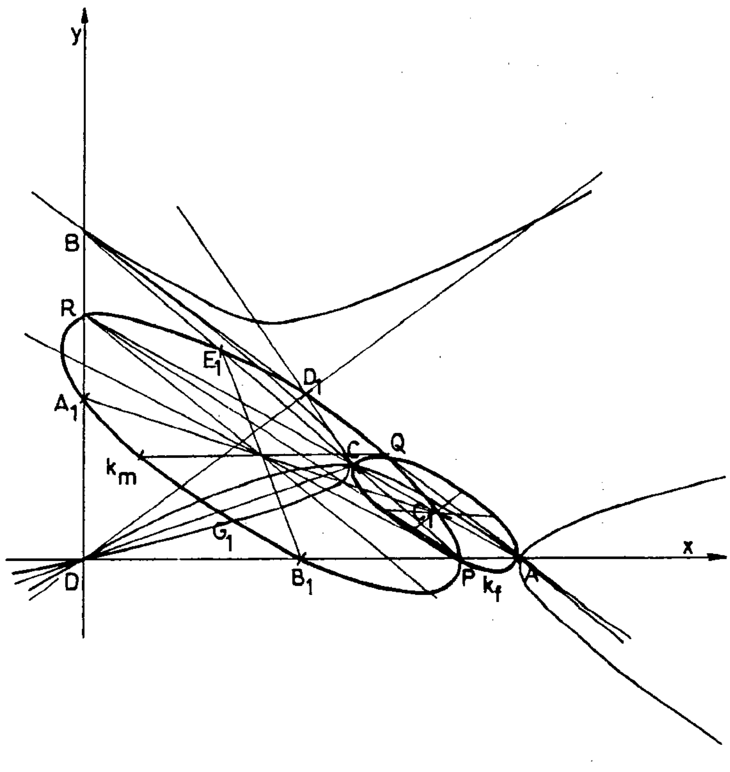

Fig. 3.

$\mathrm{II}_A 10.\ S_4 \neq 0,\ S_1^* > 0,\ S_3 > 0$ oder $S_4 \neq 0,\ S_1^* < 0,\ S_3 < 0.$

Die isotrope Grundgerade schneidet die Mittelpunktsellipse in zwei verschiedenen Punkten B_1 bzw. D_1. Einer davon ist der Mittelpunkt einer Hyperbel 1. Art, während der andere Punkt der Mittelpunkt einer Hyperbel 2. Art im Büschel ist.

$\mathrm{II}_A 11$. $S_4 \neq 0$, $S_1^* > 0$, $S_3 = 0$ oder $S_4 \neq 0$, $S_1^* = 0$, $S_3 < 0$.

Die isotrope Grundgerade schneidet die Mittelpunktsellipse in 2 Punkten B_1 und D_1. Einer davon ist der Mittelpunkt eines schneidenden Geradenpaares ($B_1 \equiv P$ oder $D_1 \equiv Q$), während der andere Punkt der Mittelpunkt einer Hyperbel 1. Art ist.

$\mathrm{II}_A 12$. $S_4 \neq 0$, $S_1^* < 0$, $S_3 = 0$ oder $S_4 \neq 0$, $S_1^* = 0$, $S_3 > 0$.

Die isotrope Grundgerade schneidet die Mittelpunktsellipse in 2 verschiedenen Punkten B_1 und D_1. Einer davon ist der Mittelpunkt eines schneidenden Geradenpaares ($B_1 \equiv P$ oder $D_1 \equiv Q$), während der andere Punkt der Mittelpunkt einer Hyperbel 2. Art ist.

$\mathrm{II}_A 13$. $S_4 = 0$, $S_1^* > 0$, $S_3 < 0$.

Die isotrope Grundgerade berührt die Mittelpunktsellipse im Punkt A_1, wobei $A_1 \equiv R$ ist, und die beiden Punkte B_1, D_1 Mittelpunkte von Hyperbeln 1. Art sind. (Fig. 4.)

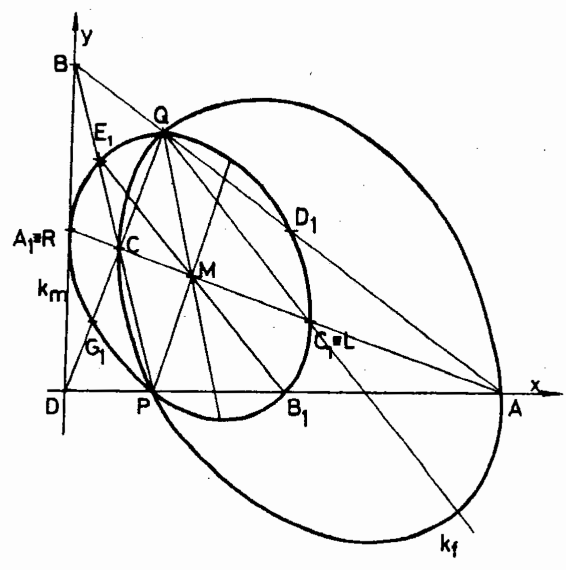

Fig. 4.

$\mathrm{II}_A 14$. $S_4 = 0$, $S_1^* < 0$, $S_3 > 0$.

Die isotrope Grundgerade berührt die Mittelpunktsellipse im Punkt A_1 ($A_1 \equiv R$) und die beiden Punkte B_1, D_1 sind Mittelpunkte von Hyperbeln 2. Art. (Fig. 5.)

$\mathrm{II}_A 15$. $S_4 = 0$, $S_1^* = 0$, $S_3 = 0$.

Alle drei Grundgeraden berühren die Mittelpunktsellipse in den Punkten A_1, B_1, D_1, d.h. es ist $A_1 \equiv R$, $B_1 \equiv P$, $D_1 \equiv Q$. (Fig. 6.)

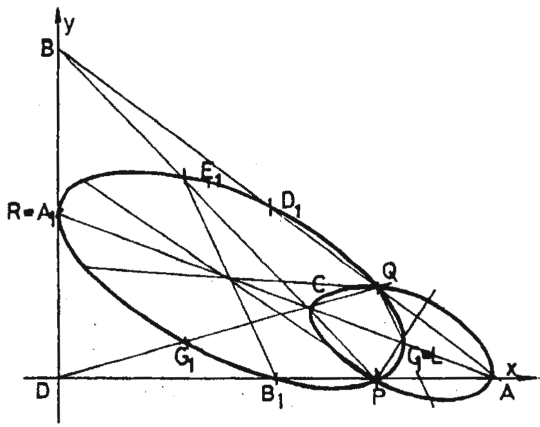

Fig. 5.

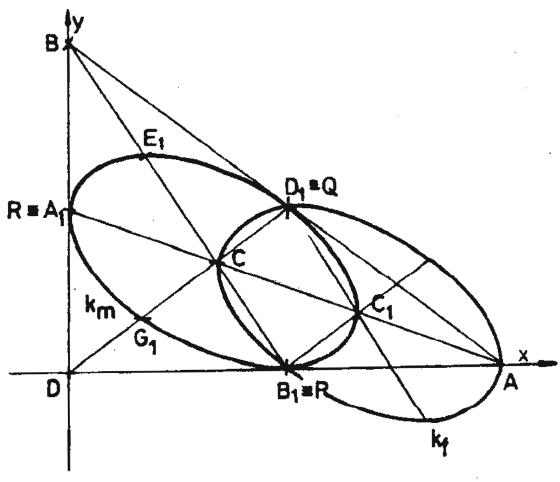

Fig. 6.

BEMERKUNGEN: 1. Man kann sofort sehen, daß es bei den typen $\mathrm{I\!I}_A 13$, 14 drei Invarianten gibt; dies folgt aus $S_4 = 0$. Beim Typ $\mathrm{I\!I}_A 15$ gibt es 2 Invarianten, zufolge der Gleichungen $S_4 = 0$ und $S_1^* = 0$.

2. Analog wie bei den Typen $\mathrm{I\!I}_A 2$, 3, 4, 5, 6, 7 zeigt man wieder, daß die algebraischen Bedingungen bei den Typen $\mathrm{I\!I}_A 10$, 11, 12 äquivalent sind.

Wir untersuchen noch die isotropen Brennpunktskurven k_f der Büschel vom Typ $\mathrm{I\!I}_A 8$, 9, 10, 11, 12, 13, 14, 15. Aus (1.2) und $\frac{\partial F}{\partial y} = 0$ findet man als Gleichung von k_f

$$-k_1k_2x^2-y^2+2k_1xy+(k_1k_2a-bk_1)x-2k_1ay+bk_1a=0 \tag{1.15}$$
$$\text{für } \lambda\in\mathbb{R} \quad \text{und}$$
$$x=0 \quad \text{für } \lambda=\infty. \tag{1.16}$$

Damit ist die Brennpunktskurve k_f eine Kurve 3. Ordnung

$$[-k_1k_2x^2-y^2+2k_1xy+(k_1k_2a-bk_1)x-2k_1ay+bk_1a]x=0. \tag{1.17}$$

Diese zerfällt in 2 Teile, wobei der Hauptteil (1.15) — wegen (1.6) — eine Ellipse ist und der andere Teil die isotrope Grundgerade darstellt. Nach einiger Rechnung ergeben sich die Grundpunkte A, C als isotrope Brennpunkte dieser Ellipse. Sie liegen nicht auf der isotropen Grundgeraden. Man kann zeigen, daß k_f die Punkte P, Q und R enthält.

2. Untersuchung der Mittelpunktskurven und Brennpunktskurven

Wegen (1.6) ist die Mittelpunktskurve und der Hauptteil der Brennpunktskurve eine Ellipse in allen 8 Fällen.

Wir bezeichnen mit a^*_M bzw. a^*_B die nicht isotropen Halbachsenlängen und mit b^*_M bzw. b^*_B die isotropen Halbachsenlängen der Mittelpunktsellipse bzw. der Brennpunktsellipse. (Fig. 7.)

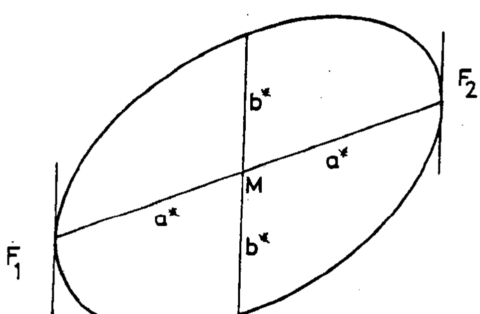

Fig. 7.

Weiters bezeichnen wir mit F_{1M}, F_{2M} bzw. F_{1B}, F_{2B} die isotropen Brennpunkte der Mittelpunktsellipse bzw. der Brennpunktsellipse. Nach einiger Rechnung ergibt sich aus (1.7) und (1.15)

$$a^*_M=\frac{\sqrt{k_1^2k_2^2a^2+2k_1^2k_2ab+k_1k_2b^2}}{4(k_1k_2-k_1^2)}, \tag{2.1}$$

(2.2) $$a_B^* = \frac{-|k_1|(b+ak_2)}{2(k_2k_1-k_1^2)},$$

(2.3) $$b_M^* = \frac{\sqrt{k_2^2b^2-k_1k_2b^2-k_2^2k_1^2a^2-2k_1^2k_2ab+k_2^3k_1a^2+2k_2^2k_1ab}}{4|k_1-k_2|}$$

und

(2.4) $$b_B^* = \frac{-(ak_2+b)\sqrt{k_2k_1-k_1^2}}{2|k_1-k_2|}.$$

Aus (2.1), (2.2), (2.3) und (2.4) kann man erkennen, daß

(2.5) $$\frac{a_M^*}{a_B^*} = \frac{b_M^*}{b_B^*} = c$$

gilt, wobei $c \in \mathbb{R}^+$ eine Konstante ist. Damit ist gezeigt, daß die Mittelpunktskurve und die Brennpunktskurve bezüglich der Gruppe

(2.6) $$\overline{x} = c_1 + \frac{1}{c}x; \qquad \overline{y} = c_2 + c_3x + \frac{1}{c}y$$

der winkeltreuen isotropen Ähnlichkeiten ähnlich sind, wobei c die Konstante in (2.5) ist [1].

SATZ 2.1. *Bei den Büscheltypen* $\mathbb{I}_A$*8, 9, 10, 11, 12, 13, 14, 15 ist jeweils die Mittelpunktsellipse zur Brennpunktsellipse des Kegelschnittbüschels ähnlich bezüglich der Gruppe der winkeltreuen isotropen Ähnlichkeiten.*

Wir untersuchen den Quotienten c in (2.5). Im Fall des Typs $\mathbb{I}_A10$ kann $c>1$, $c=1$, $0<c<1$ sein, in den Fällen des Typs $\mathbb{I}_A8$, 13, 11 gilt aber $0<c<1$. In den Fällen $\mathbb{I}_A9$, 12, 14 gilt $c>1$ schließlich gilt $c=1$ im Fall $\mathbb{I}_A15$.

Weiters folgt aus (2.1), (2.3) mittels (2.5)

(2.7) $$\frac{a_M^*}{b_M^*} = \frac{a_B^*}{b_B^*} = \frac{1}{\sqrt{k_1k_2-k_1^2}},$$

und für die Achsenprodukte der Brennpunktsellipse bzw. der Mittelpunktsellipse gilt

(2.8)
$$a_B^*b_B^* = \frac{(ak_2+b)^2\sqrt{k_1k_2-k_1^2}}{4(k_1-k_2)^2} \quad \text{bzw.} \quad a_M^*b_M^* = \frac{1}{c^2}\frac{(ak_2+b)^2\sqrt{k_1k_2-k_1^2}}{4(k_1-k_2)^2},$$

wobei c die Konstante in (2.5) bezeichnet.

Aus (1.7), (1.15) kann man mittels $\frac{\partial F}{\partial y}=0$ die Gleichungen jener Geraden bestimmen, welche die Brennpunkte der Mittelpunktskurve bzw. der Brennpunktsellipse enthalten. Man gewinnt

$$y=k_1x+\frac{b-2k_1a}{4} \quad \text{bzw.} \quad y=k_1x-k_1a. \tag{2.9}$$

Es ist klar, daß diese Geraden parallel sind.

Die Koordinaten des Mittelpunkts der Mittelpunktsellipse bzw. der Brennpunktsellipse sind

$$M\left(\frac{2k_1a+b-k_2a}{4(k_1-k_2)},\frac{2k_1b-k_2b+k_1k_2a}{4(k_1-k_2)}\right) \quad \text{bzw.} \tag{2.10}$$
$$L\left(\frac{b+2k_1a-k_2a}{2(k_1-k_2)},\frac{k_1b+k_1k_2a}{2(k_1-k_2)}\right).$$

Aus (2.10) kann man sofort ersehen, daß $2d(DM)=d(DL)$ gilt, d.h. der isotrope Abstand zwischen dem Grundpunkt D (D liegt auf der isotropen Grundgeraden) und dem Mittelpunkt der Brennpunktsellipse ist zweimal so groß wie der isotrope Abstand zwischen dem Grundpunkt D und dem Mittelpunkt der Mittelpunktsellipse.

Mittels der Koordinaten von M und L kann man die Werten c_1, c_2, c_3 in (2.6) bestimmen. Man erhält

(2.11)
$$c_1=-\frac{b+2k_1a-k_2a}{4c(k_1-k_2)}+\frac{b+2k_1a-k_2a}{2(k_1-k_2)},\; c_2=\frac{k_2b-2k_1b-k_1k_2a}{4c(k_1-k_2)}+\frac{k_1b+k_1k_2a}{2(k_1-k_2)}.$$

und $c_3=0$. Die isotrope Ähnlichkeit (2.6) ist für $c=1$ eine isotrope Bewegung und für $c\neq 1$ eine winkeltreue isotrope Ähnlichkeit mit einem Fixpunkt. Die Koordinaten dieses Fixpunktes sind

(2.12)
$$x_0=\frac{-\frac{b+2k_1a-k_2a}{4c(k_1-k_2)}+\frac{b+2k_1a-k_2a}{2(k_1-k_2)}}{1-\frac{1}{c}},\quad y_0=\frac{\frac{k_2b-2k_1b-k_1k_2a}{4c(k_1-k_2)}+\frac{k_1b+k_1k_2a}{2(k_1-k_2)}}{1-\frac{1}{c}}.$$

Hinsichtlich der Gruppe der winkeltreuen isotropen Ähnlichkeiten sind einige interessante Untersuchungen bekannt [1].

3. Untersuchung des gemeinsamen Poldreiecks des Kegelschnittbüschels

Wie schon in **1.** erwähnt, liegen die drei Diagonalecken P, Q, R — d.h. die Mittelpunkte der schneidenden Geradenpaare des Kegelschnittbüschels

— auf der Brennpunktskurve 3. Ordnung k_f. Hierbei liegt R auf der isotropen Grundgeraden, während die Punkte P und Q der Brennpunktsellipse angehören. Die Koordinaten diesen 3 Punkte sind

(3.1)
$$R(0,-k_1a);\ P\left(-\frac{b}{k_2},0\right);\ Q\left(\frac{ab(b+k_1a)}{b^2+2abk_1+a^2k_1k_2},\frac{ab(k_1b+ak_1k_2)}{b^2+2abk_1+a^2k_1k_2}\right).$$

An Hand des bekannten Ausdrucks

$$R^*=\frac{\begin{vmatrix}1&1&1\\x_1&x_2&x_3\\y_1&y_2&y_3\end{vmatrix}}{(x_2-x_1)(x_3-x_1)(x_3-x_2)} \tag{3.2}$$

[1] für den isotropen Umkreisradius eines Dreiecks, bestimmen wir den Radius des isotropen Umkreises des Dreiecks P, Q, R, d.h. des gemeinsamen Poldreiecks des Kegelschnittbüschels. Nach längerer Rechnung ergibt sich

$$R^*=\frac{2k_1k_2(b^2+2abk_1+a^2k_1k_2)}{b^2(b+2ak_1)}. \tag{3.3}$$

Aus (3.3) ziehen wir einige Folgerungen: 1. Im Falle $b+2ak_1=0$, d.h. bei den Typen $\mathbb{I}_A13$, 14, 15 existiert R^* nicht, denn die Punkte P und Q sind parallel. Somit ist das Poldreieck PQR kein zulässiges Dreieck der isotropen Ebene I_2.

2. Im Falle $b+2ak_1\neq 0$, d.h. bei den Typen $\mathbb{I}_A8$, 9, 10, 11, 12 sind die Punkte P, Q, R die Ecken eines zulässigen Dreiecks in der isotropen Ebene.

Mittels der Parameterwerte a, b, k_1, k_2 kann man den Flächeninhalt des Dreiecks PQR beschreibt, wobei man den bekannten Ausdruck

$$F=-\frac{|a||b||c|}{4p} \tag{3.4}$$

benützt, wobei $|a|$, $|b|$, $|c|$ die Seitenlängen des Dreieck sind und p der Parameter des Umkreises ist [1]. Man erhält den Flächeninhalt des Poldreiecks PQR zu

$$F=-\frac{abk_1(b+k_1a)(b+k_2a)}{k_2(b^2+2abk_1+a^2k_2k_1)}. \tag{3.5}$$

Bei den Typen $\mathbb{I}_A8$, 9, 10, 11, 12, d.h. wenn das Poldreieck PQR ein zulässiges Dreieck der isotropen Ebene I_2 ist, kann man die Höhenfußpunkte des Poldreiecks PQR untersuchen. Es ist klar, daß ein Fußpunkt von ihnen, nämlich T_3 in der isotropen Grundgeraden DB liegt. Für die weiteren Untersuchungen sind die beiden anderen Fußpunkte T_1 und T_2 interessant. Nach einiger Rechnung ergeben sich ihre Koordinaten zu

$$(3.6)\quad T_1\left(\frac{ab(b+k_1a)}{b^2+2abk_1+a^2k_2k_1}, \frac{-k_1a(b+2k_1a)(b+k_2a)}{b^2+2abk_1+a^2k_1k_2}\right) \quad \text{und}$$

$$T_2\left(-\frac{b}{k_2}, -\frac{2k_1}{k_2}(b+k_2a)\right).$$

Man kann sofort mittels (1.15) erkennen, daß diese Punkte T_1, T_2 auf dem Hauptteil der Brennpunktskurve, d.h. auf der Brennpunktsellipse liegen.

SATZ 3.1. *Ist das gemeinsame Poldreieck PQR eines Kegelschnittbüschels der Typen $\mathbb{I}_A 8$, 9, 10, 11, 12 ein zulässiges Dreieck, dann liegen außer den Punkte P, Q, R auch noch die isotropen Höhenfußpunkte des Dreiecks PQR auf der Brennpunktskurve.*

4. Metrische Beziehungen für die Achsen der Kegelschnitte im Büschel

Wir betrachten nur Büschel der Typen $\mathbb{I}_A 15$, 14, 13.

Der Typ $\mathbb{I}_A 15$.

Zunächst gelten die Bedingungen $2ak_1+b=0$ und $ak_2+2b=0$, d.h., die Normalform lautet

$$(4.1)\qquad F \equiv \frac{\lambda b}{2a}x^2+y^2+2\left(\frac{b}{a}+\frac{\lambda}{2}\right)xy-\frac{\lambda b}{2}x-by=0.$$

In der folgenden Untersuchung sei $a>0$. Wir gewinnen den Fall $a<0$ aus diesem durch die Spiegelung $\overline{x}=-x$; $\overline{y}=y$.

Aus (2.1), (2.2), (2.3) und (2.4) kann man leicht sehen, daß die folgenden Gleichungen

$$(4.2)\qquad a_M^* = a_B^* = \frac{a}{3},\quad b_M^* = b_B^* = \frac{\sqrt{3}}{6}b \quad \text{gelten.}$$

Mittels (2.10) erhält man die Koordinaten des Mittelpunkts der Mittelpunktsellipse bzw. der Brennpunktsellipse zu

$$(4.3)\qquad M\left(\frac{a}{3},\frac{b}{3}\right),\quad L\left(\frac{2a}{3},\frac{b}{6}\right).$$

Die Koordinaten der isotropen Brennpunkte der Mittelpunktsellipse bzw. der Brennpunktsellipse sind

(4.4)

$$F_{1M}\equiv A_1\left(0,\frac{b}{2}\right);\quad F_{2M}\equiv C_1\left(\frac{2a}{3},\frac{b}{6}\right);\quad F_{1B}\equiv C\left(\frac{a}{3},\frac{b}{3}\right);\quad F_{2B}\equiv A(a,0).$$

Aus diesen Beziehungen ergibt sich der

SATZ 4.1. *Die Mittelpunktsellipse und die Brennpunktsellipse eines Kegelschnittbüschels vom Typ* $\mathbb{I}_A 15$ *sind kongruent; beide Kurven lassen sich durch die Schiebung* $\left\{\overline{x}=x+\frac{a}{3},\ \overline{y}=y-\frac{b}{6}\right\}$ *aufeinander abbilden.*

BEMERKUNG: Man kann leicht sehen, daß die Punkte M, L, F_{1M}, F_{2M}, F_{1B}, F_{2B} auf der Geraden $y=-\frac{b}{2a}x+\frac{b}{2}$ liegen.

Aus (4.2) folgen die Gleichungen

$$\frac{a_M^*}{b_M^*}=\frac{a_B^*}{b_B^*}=\frac{2a}{\sqrt{3}b} \quad \text{und} \quad a_M^* b_M^* = a_B^* b_B^* = \frac{ab}{6\sqrt{3}}. \tag{4.5}$$

Nach einiger Rechnung ergeben sich aus (4.1)

$$(a^*)^2 = \mp\frac{a^4b^2\lambda^2+2\lambda a^2b^2}{(4b^2+\lambda^2a^2+2\lambda ab)^2}, \quad (b^*)^2=\mp\frac{a^4b^2\lambda^2+2\lambda a^3b^3}{4a^2(4b^2+\lambda^2a^2+2\lambda ab)}, \tag{4.6}$$

wobei wir mit a^* bzw. b^* die nichtisotrope bzw. isotrope Halbachsenlänge der Hyperbeln 1. Art $(-)$ bzw. der Hyperbeln 2. Art $(+)$ des Büschels bezeichnen; λ ist der zugehörige Büschelparameter.

Wir untersuchen die Funktion

$$g(\lambda)=\frac{b^*(\lambda)}{a^*(\lambda)}=\frac{\sqrt{4b^2+\lambda^2a^2+2\lambda ab}}{2a}. \tag{4.7}$$

Durch Differenzieren von $g(\lambda)$ ergibt sich, daß stets $g(\lambda)\geq\frac{\sqrt{3}b}{2a}$ ist und Gleichheit für $\lambda=-\frac{b}{a}$ gilt. Dieser Parameterwert gehört zu einer Hyperbel 1. Art. Ihre Halbachsenlängen sind

$$a_0^*=\frac{a}{3} \quad \text{und} \quad b_0^*=\frac{b}{2\sqrt{3}}. \tag{4.8}$$

Aus (4.2) und (4.8) folgt $a_M^*=a_B^*=a_0^*=\frac{a}{3}$ und $b_M^*=b_B^*=b_0^*=\frac{b}{2\sqrt{3}}$.

Sei k' eine beliebige reelle Zahl mit

$$k'>\frac{\sqrt{3}b}{2a}. \tag{4.9}$$

Nach einiger Rechnung ergibt sich die Gleichung

$$g(\lambda)=\frac{\sqrt{4b^2+\lambda^2a^2+2\lambda ab}}{2a}=k' \tag{4.10}$$

und diese hat genau zwei verschiedene Nullstellen, nämlich

$$\lambda_1^*=-\frac{b}{a}+\sqrt{-\frac{3b^2}{a^2}+4k'^2} \quad \text{und} \quad \lambda_2^*=-\frac{b}{a}-\sqrt{-\frac{3b^2}{a^2}+4k'^2}. \tag{4.11}$$

Diese Parameterwerte gehören entweder zu Hyperbeln 1. Art oder zu Hyperbeln 2. Art. Man kann leicht zeigen, daß die Gleichungen

$$a^*(\lambda_1^*) = a^*(\lambda_2^*) \quad \text{und} \quad b^*(\lambda_1^*) = b^*(\lambda_2^*) \tag{4.12}$$

gelten.

Damit haben wir den

SATZ 4.2. *In einem Kegelschnittbüschel vom Typ* $\mathbb{I}_A 15$ *der isotropen Ebene existieren stets Paare kongruenter Hyperbeln 1. Art bzw. 2. Art. Es existiert eine einzige Ausnahmehyperbel 1. Art, zu der es kein verschiedenes kongruentes Exemplar im Büschel gibt; für diese Hyperbel ist der Quotient der Halbachsenlängen maximal. Die Halbachsenlängen dieser Ausnahmehyperbel stimmen mit jenen der Mittelpunktsellipse bzw. der Brennpunktsellipse überein.*

BEMERKUNG: Man kann auch leicht zeigen, daß das Verhältnis der Halbachsenlängen der Kegelschnitte zwischen $\frac{2a}{\sqrt{3}b}$ und $\frac{a}{b}$ Hyperbeln 1. Art kennzeichnet. Ist dieses Verhältnis kleiner als $\frac{a}{b}$, dann liegen Hyperbeln 2. Art im Kegelschnittbüschel vor.

Im folgenden untersuchen wir noch die Halbachsenprodukten der Kegelschnitte des Kegelschnittbüschels. Aus (4.6) erhält man die Funktion

$$f(\lambda) = a^* b^* = \frac{|a^4 b^2 \lambda^2 + 2\lambda a^3 b^3|}{2a\sqrt{(4b^2 + \lambda^2 a^2 + 2\lambda ab)^3}}. \tag{4.13}$$

Mittels $f(\lambda)$ bestimmen wir die Extremwerten von $f(\lambda)$. Man erhält Extremwerte für die Büschelparameter

$$\lambda_1' = \frac{2b}{a}, \quad \lambda_2' = -\frac{4b}{a}, \quad \lambda_3' = -\frac{b}{a}. \tag{4.14}$$

Nach einiger Rechnung ergibt sich

$$f(\lambda_1') = f(\lambda_2') = f(\lambda_3') = \frac{ab}{6\sqrt{3}}. \tag{4.15}$$

Ersichtlich gehören λ_1', λ_2' zu Hyperbeln 2. Art, λ_3' aber zu einer Hyperbeln 1. Art. Man zeigt leicht:

$$a^*(\lambda_1') = a^*(\lambda_2') = \frac{a}{3\sqrt{2}}, \quad b^*(\lambda_1') = b^*(\lambda_2') = \frac{b}{\sqrt{6}}, \tag{4.16}$$

$$a^*(\lambda_3') = \frac{a}{3}, \quad b^*(\lambda_3') = \frac{b}{2\sqrt{3}}. \tag{4.17}$$

Sei $d \in \mathbb{R}$ mit $0 < d < \frac{ab}{6\sqrt{3}}$. Durch Untersuchung der Gleichung

$$f(\lambda) = \frac{|a^4 b^2 \lambda^2 + 2\lambda a^3 b^3|}{2a\sqrt{(4b^2 + \lambda^2 a^2 + 2\lambda ab)^3}} = d \tag{4.18}$$

ergibt sich, daß diese Gleichung genau 6 verschiedene reelle Nullstellen hat, und diese Parameterwerte zu 2 kongruenten Hyperbeln 1. Art und 2 Paaren kongruenter Hyperbeln 2. Art gehören.

SATZ 4.3. *In einem Kegelschnittbüschel vom Typ* $\mathbb{I}_A 15$ *der isotropen Ebene existieren stets 6 Kegelschnitte zu fest vorgegebenem Halbachsenprodukt* $< \frac{ab}{6\sqrt{3}}$; *hierbei 2 Kegelschnitte kongruente Hyperbeln 1. Art, die beiden anderen Kegelschnittpaare sind kongruente Hyperbeln 2. Art. Im Büschel liegen drei Hyperbeln — eine Hyperbel 1. Art und 2 kongruente Hyperbeln 2. Art — so, daß ihr Halbachsenprodukt den Wert* $\frac{ab}{6\sqrt{3}}$ *annimt. Die Hyperbel 1. Art stimmt hierbei mit der Hyperbel 1. Art aus Satz 4.2 überein.*

Die Typen $\mathbb{I}_A 13$, 14.

Aus (1.2) und der Gleichung $2ak_1 + b = 0$ erhält man die Normalform der Typen $\mathbb{I}_A 13$, 14 zu

$$\frac{\lambda b}{2a}x^2 + y^2 + 2xy\left(\frac{\lambda}{2} - \frac{k_2}{2}\right) - \frac{\lambda bx}{2} - by = 0. \tag{4.19}$$

In der folgenden Untersuchung sei $a > 0$ wie beim Typ $\mathbb{I}_A 15$. Gemäß (2.1), (2.2), (2.3) und (2.4) gelten die Gleichungen

$$a_M^* = \frac{k_2 a^2}{4ak_2 + 2b}, \quad b_M^* = -\frac{ak_2}{4}\sqrt{\frac{b}{-b-2ak_2}}, \tag{4.20}$$

$$a_B^* = \frac{ab + a^2 k_2}{2ak_2 + b}, \quad b_B^* = -\frac{b + ak_2}{4}\sqrt{\frac{b}{-b-2ak_2}}. \tag{4.21}$$

Aus diesen folgert man leicht die folgenden Beziehungen

$$\frac{a_M^*}{a_B^*} = \frac{b_M^*}{b_B^*} = \frac{k_2 a}{2(b+ak_2)}, \quad \frac{a_M^*}{b_M^*} = \frac{a_B^*}{b_B^*} = \frac{2a}{\sqrt{b}}\frac{1}{\sqrt{-b-2ak_2}}, \tag{4.22}$$

$$a_M^* b_M^* = \frac{k_2^2 a^3 \sqrt{b}}{8\sqrt{(-2ak_2 - b)^3}}, \quad a_B^* b_B^* = \frac{(ak_2 + b)^2 a\sqrt{b}}{2\sqrt{(b - 2ak_2)^3}}. \tag{4.23}$$

Die Mittelpunkte der Mittelpunktsellipse bzw. der Brennpunktsellipse haben die Koordinaten

(4.24)

$$M\left(\frac{2a^2k_2}{4(b+2ak_2)}, \frac{2b^2 + 3abk_2}{4(b+2ak_2)}\right) \quad \text{bzw.} \quad L\left(\frac{2a^2k_2}{2(b+2ak_2)}, \frac{b^2 + abk_2}{2(b+2ak_2)}\right).$$

An Hand von (4.24) erkennt man, daß die Mittelpunkte und die Brennpunkte der Mittelpunktsellipse bzw. der Brennpunktsellipse auf der Grundgeraden AC liegen.

Nach einiger Rechnung ergeben sich für die Halbachsenlängen der Kegelschnitte des Büschel die Werte

$$(a^*)^2 = \pm\frac{-\lambda^2a^2b^2+2\lambda ab^3+2\lambda a^2b^2k_2}{(\lambda^2a-2\lambda(b+ak_2)+k_2^2a)^2}, \tag{4.25}$$
$$(b^*)^2 = \pm\frac{-\lambda^2a^2b^2+2\lambda ab^3+2\lambda a^2b^2k_2}{4a(\lambda^2a-2\lambda(b+ak_2)+k_2^2a)}.$$

wobei wir wie in (4.6) mit a^* und b^* die nichtisotrope bzw. isotrope Halbachsenlänge der Hyperbeln 1. Art (+) bzw. der Hyperbeln 2. Art (−) bezeichnet haben und λ der Büschelparameter ist.

Wir untersuchen wieder das Verhältnis der Halbachsenlängen der Kegelschnitte. Diese Funktion lautet

$$g(\lambda) = \frac{b^*}{a^*}(\lambda) = \sqrt{\frac{\lambda^2a-2\lambda(b+ak_2)+k_2^2a}{4a}}. \tag{4.26}$$

Durch Differenzieren von (4.26) ergibt sich, daß $g(\lambda)$ ein lokales Minimun für

$$\lambda = \lambda_0^* = \frac{b+ak_2}{a} \tag{4.27}$$

hat. Dieser Parameterwert gehört zu einer Hyperbel 1. Art, i.f. bezeichnet mit h_0. Die Halbachsenlängen von h_0 sind

$$a^*(\lambda_0^*) = \frac{a(b+ak_2)}{b+2ak_2} \quad \text{und} \quad b^*(\lambda_0^*) = \frac{\sqrt{b}}{2}\frac{-(b+ak_2)}{\sqrt{-b-2ak_2}}. \tag{4.28}$$

Aus (4.21) und (4.28) kann man ersehen, daß

$$a_B^* = a^*(\lambda_0^*) \quad \text{und} \quad b_B^* = b^*(\lambda_0^*) \quad \text{ist.} \tag{4.29}$$

Aus (4.28) berechnen wir schließlich das Verhältnis der Halbachsenlängen der Hyperbel 1. Art h_0 zu

$$\frac{a^*}{b^*}(\lambda_0^*) = \frac{2a}{\sqrt{b}\sqrt{-b-2ak_2}}. \tag{4.30}$$

Sei $k' \in \mathbb{R}$ mit $k' > \frac{b^*}{a^*}(\lambda_0^*)$. Nach einiger Rechnung ergibt sich, daß die Gleichung

$$g(\lambda) = \sqrt{\frac{\lambda^2a-2\lambda(b+ak_2)+k_2^2a}{4a}} = k' \tag{4.31}$$

genau 2 verschiedene Nullstellen hat, nämlich

(4.32)

$$\lambda_1^* = \frac{b}{a}+k_2+\sqrt{\frac{b^2}{a^2}+2\frac{b}{a}k_2+4k'^2} \quad \text{und} \quad \lambda_2^* = \frac{b}{a}+k_2-\sqrt{\frac{b^2}{a^2}+2\frac{b}{a}k_2+4k'^2}.$$

Diese Parameterwerte gehören entweder zu einer Hyperbel 1. Art oder einer Hyperbel 2. Art. Kann man zeigen, daß die Gleichungen

$$(4.33) \qquad a^*(\lambda_1^*) = a^*(\lambda_2^*), \quad b^*(\lambda_1^*) = b^*(\lambda_2^*)$$

gelten.

Damit haben wir den

SATZ 4.4. *In einem Kegelschnittbüschel vom Typ* $\mathbb{I}_A$*13, 14 der isotropen Ebene existieren stets Paare kongruenter Hyperbeln 1. Art bzw. 2. Art. Es existiert eine einzige Ausnahmehyperbel 1. Art* h_0*, zu der es kein verschiedenes kongruentes Exemplar im Büschel gibt. Für* h_0 *ist der Quotient der Halbachsenlängen maximal. Die Halbachsenlängen von* h_0 *stimmen mit den Halbachsenlängen der Brennpunktsellipse überein.*

Wir untersuchen schließlich die Halbachsenprodukte der Kegelschnitte eines Büschels der Typen $\mathbb{I}_A$13, 14. Aus (4.25) folgt für die Funktion $f(\lambda) = a^*b^*(\lambda)$

$$(4.34) \qquad f(\lambda) = a^*b^*(\lambda) = \frac{|\lambda^2 a^2 b^2 - 2\lambda a b^3 - 2\lambda a^2 b^2 k_2|}{2\sqrt{a}\sqrt{(\lambda^2 a - 2\lambda(b + ak_2) + k_2^2 a)^3}}.$$

Mittels $f'(\lambda)$ kann man die lokalen Extremwerte der Funktion $f(\lambda)$ bestimmen. Sie hat ein lokales Maximum für

$$(4.35) \qquad \lambda = \lambda_0^*, \quad \lambda = \lambda_1' = \frac{b}{a} + k_2 + \sqrt{\frac{b^2}{a^2} + 2\frac{b}{a}k_2 + 3k_2^2}$$

$$\text{bzw.} \quad \lambda = \lambda_2' = \frac{b}{a} + k_2 - \sqrt{\frac{b^2}{a^2} + 2\frac{b}{a}k_2 + 3k_2^2}.$$

Nach einiger Rechnung ergibt sich

$$(4.36) \qquad f(\lambda_1') = f(\lambda_2') = -\frac{b^2}{3\sqrt{3}k_2} \quad \text{und} \quad f(\lambda_0^*) = \frac{a\sqrt{b}(b + ak_2)^2}{2\sqrt{(-b - 2ak_2)^3}}.$$

Man zeigt leicht, daß λ_1', λ_2' zu kongruenten Hyperbeln 2. Art gehören. Wir berechnen die Halbachsenlängen der vorliegenden Hyperbeln 2. Art zu

$$(4.37) \qquad a^*(\lambda_1') = a^*(\lambda_2') = -\frac{\sqrt{2}}{3}\frac{b}{k_2} \quad \text{und} \quad b^*(\lambda_1') = b^*(\lambda_2') = \frac{b}{\sqrt{6}}.$$

Durch Untersuchung von $f(\lambda)$ (4.34) kann man unschwer sehen, daß $f(\lambda_0^*) \neq f(\lambda_1')$ gilt. Genauer hat man die folgenden Ungleichungen:

Beim Typ $\mathbb{I}_A$13 $f(\lambda_0^*) > f(\lambda_1')$; Beim Typ $\mathbb{I}_A$14 $f(\lambda_0^*) < f(\lambda_1')$.

Sei $d \in \mathbb{R}$ mit $0 < d < d^*$, wobei $d^* = \max(f(\lambda_0^*), f(\lambda_1'))$. Wir untersuchen die Nullstellen der Gleichung.

$$f(\lambda) = a^* b^*(\lambda) = \frac{ab^2}{2\sqrt{a}} \frac{|\lambda^2 a - 2\lambda b - 2\lambda a k_2|}{\sqrt{(\lambda^2 a - 2\lambda b - 2\lambda a k_2 + k_2^2 a)^3}} = d. \tag{4.38}$$

Nach einigen Überlegungen ergibt sich der interessante

SATZ 4.5. *In einem Kegelschnittbüschel vom Typ* $\mathbb{I}_A 13$, *14 der isotropen Ebene existieren stets 6 Kegelschnitte — nämlich 2 kongruente Hyperbeln 1. Art und 2 Paare kongruenter Hyperbeln 2. Art — derart, daß ihr Halbachsenprodukt einer festen Zahl* $< d^{**}$ *gleich ist, wobei* $d^{**} = \min(f(\lambda_0^*), f(\lambda_1'))$ *bedeutet. In einem Büschel vom Typ* $\mathbb{I}_A 13$ *existiert eine Hyperbel 1. Art* h_0 *derart, daß ihr Halbachsenprodukt gleich* $f(\lambda_0^*)$ *ist; es existieren 2 Hyperbeln 1. Art, mit einem Halbachsenprodukt zwischen* $f(\lambda_0^*)$ *und* $f(\lambda_1')$, *und es gibt 4 Hyperbeln — genauer 2 kongruente Hyperbeln 1. Art und 2 kongruente Hyperbeln 2. Art — so, daß ihr Halbachsenprodukt gleich* $f(\lambda_1')$ *ist. In einem Büschel vom Typ* $\mathbb{I}_A 14$ *gibt es 2 kongruente Hyperbeln 2. Art mit maximalen Halbachsenprodukt* $f(\lambda_1')$; *es existiert 4 Hyperbeln 2. Art — wobei 2 Paare je kongruent sind — so, daß ihr Halbachsenprodukt zwischen* $f(\lambda_1')$ *und* $f(\lambda_0^*)$ *liegt. Schließlich existieren 5 Hyperbeln, deren Halbachsenprodukt den Wert* $f(\lambda_0^*)$ *annimt; unter den 4 Hyperbeln von 2. Art sind je zwei Paare kongruent. Eine Hyperbel ist von 1. Art.*

5. Zusammenfassung

In dieser Arbeit untersuchen wir mit analytischen Methoden jene Kegelschnittbüschel der isotropen Ebene, welche vier reelle und verschiedene Grundpunkte bezitzen, wobei mindestens zwei parallel sind und die konvexe Hülle der vier Grundpunkte ein nichtzulässiges Dreieck ist. Die auftretenden algebraischen Beziehungen werden in der isotropen Ebene geometrisch gedeutet. Für diese Büscheltypen werden metrische Resultate angegeben.

Literaturverzeichnis

[1] SACHS, H: *Ebene Isotrope Geometrie*, Vieweg–Verlag, Braunschweig/Wiesbaden, 1987.

[2] SACHS, H: Oskulierende und hyperoskulierende Kegelschnittbüschel der isotropen Ebene I, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, **196** (1987), 337–375.

[3] SACHS, H: Oskulierende und hyperoskulierende Kegelschnittbüschel der isotropen Ebene II, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, (im Druck).

[4] ŠČURIČ, VL.: Kegelschnittbüschel der isotropen Ebene I, *RAD JAZU*, **450** (1990), 41–51.

[5] SACHS, H. u. ŠČURIČ, VL.: Klassifikationstheorie der Kegelschnittbüschel vom Typ IV. in der isotropen Ebene, *RAD JAZU*, (im Druck).

[6] COXETER, H. S. M.: *Projektiv Geometria*, Gondolat, Budapest, 1986.

[7] JAGLOM, I. M.: *Galilei relativitási elve és egy nemeuklideszi geometria*, Gondolat, Budapest, 1985.

[8] STEINER, J.: Vermischte Sätze und Aufgaben, *Journal für die reine und angew. Math.*, **55** (1858), 356–378.

[9] KANTOR S.: Metrische Formeln für das Kegelschnittbüschel mit vier reellen Grundpunkten, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, **78** (1878), 905–915.

[10] TREBITSCHER, M.: Über die Reduction eines Büschels von Kurven zweiter Ordnung ein Strahlenbüschel, *Sitz.-Ber. d. Öst. Akad. Wiss. Wien*, **80** (1879), 913–943.

[11] DINGELDEY, F.: Kegelschnitte und Kegelschnittsysteme, *Encykl. d. Math. Wiss. III*, C1, 46., 88–139.

[12] BOBEK, K.: Die Brennpunktscurve des Kegelschnittbüschels, *Monatsh. Math. Phys.*, **3** (1892), 309–317.

[13] TREBITSCHER, M.: Über Beziehungen zwischen Kegelschnittbüscheln und rationalen Kurven dritter Klasse, *Sitz. Ber. d. Öst. Akad. Wiss. Wien*, **81** (1880), 1080–1091.

[14] MOLNÁR-S., K.: Klassifikation der Kegelschnittbüschel mit vier reellen Grundpunkten und einem parallelen Grundpunktepaar in der isotropen Ebene, *Annales Univ. Scient. Budapest.* Sect. Math., **35** (1992), 37–57.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 129–132

ON THE NUMBER OF EXPANSIONS $1=\sum q^{-n_i}$

By

P. ERDŐS and I. JOÓ

Mathematical Institute of the Hungarian Academy of Sciences, Budapest

(Received September 4, 1991)

Let $1<q<2$ and consider the expansions

$$1=\sum_{i=1}^{\infty}\varepsilon_i/q^i,\qquad \varepsilon_i=\begin{cases}0\\1\end{cases}. \tag{1}$$

Such an expansion is not unique in general. The unique expansions were investigated in [1]. The aim of the present note is to prove the following

THEOREM. *For every $n\geq 1$ there exists $2^{\aleph_0}$ many $q\in(1,2)$ such that 1 has exactly $n+1$ expansions of the form* (1).

PROOF. It is a development of the ideas of the paper [1]. We define a desired q by the digits (ε_i) as follows: $\varepsilon_k=1$ if $k\leq 9$, $\varepsilon_k=0$ if $9+(i-1)10<k<$ $<9+10i$ $(i=1,\ldots,n)$, $\varepsilon_k=1$ if $k=9+10i$ $(i=1,\ldots,n)$, $\varepsilon_k=1$ if $k=9+10n+5j$ $(j\geq 0)$ and $\varepsilon_k=0$ for each k with $9+10n+5(j-1)<k\leq 9+10n+5j$ $(j\geq 1)$ except one such k where ε_k can be 0 or 1 as we want, i.e. the digits are

$$\underbrace{11\ldots1}_{9}\ \underbrace{\boxed{\underbrace{00\ldots01}_{9}}}_{n\text{ times}}\ \underbrace{\boxed{0a001}}_{\infty\text{ times}},\qquad a=\begin{cases}0\\1\end{cases}. \tag{2}$$

Here a is arbitrary in each cycle.

One can obtain the remaining n expansions as follows: we substitute by zero one of the 1's before some 9 tuple of zeros and shifting the digit sequence (2) until the digit following that 1 we add to (2) (instead of the omitted digit) the whole shifted sequence (2). In this case, obviously, we will not have two 1-digit on the same place. For example, in case $n=2$ the three expansions are characterized by the digit sequences (ε_k) as follows:

1) $\underbrace{11\ldots1}_{9}\,\underbrace{00\ldots0}_{9}1\,\underbrace{00\ldots0}_{9}1\,\underbrace{0a001}_{\infty}$

2)

$$\underbrace{11\ldots\not{1}}_{9}\underbrace{00\ldots0}_{9}1\underbrace{00\ldots0}_{9}1\underbrace{0a001}_{\infty}$$

$$+\qquad \underbrace{11\ldots1}_{9}\underbrace{00\ldots0}_{9}1\underbrace{00\ldots0}_{9}1\underbrace{0a001}_{\infty}$$

$$\underbrace{11\ldots11}_{8}0\underbrace{11\ldots11}_{10}\underbrace{0\ldots0}_{8}110a0010a\underbrace{011aa}_{\infty}011aa011$$

3)

$$\underbrace{1\ldots1}_{9}\underbrace{0\ldots0}_{9}\not{1}\underbrace{0\ldots0}_{9}10a0010a0010a0010a0010a0010a0010a001$$

$$+\qquad \underbrace{1\ldots1}_{9}000000000100000000010a0010a0010a0010$$

$$\underbrace{1\ldots1}_{9}\underbrace{00\ldots0}_{10}\underbrace{1\ldots.1}_{10}0a0010a0110a0010a\underbrace{011aa}_{\infty}011aa011aa011$$

Next we show in three steps that this construction satisfies the requirements of the Theorem. To this consider any q defined by (2) and investigate any expansion of 1 in terms of such a q.

a) If $\varepsilon_k=1$ and on some of the places $k+1, \ldots, k+8$ we have 1, then we can not change ε_k to 0 (if we do not change $\varepsilon_1, \ldots \varepsilon_{k-1}$).

Indeed, omit $\varepsilon_k=1$ i.e. replace with zero and add from the $k+1$-th digit to the shifted expansion itself. There is a $k<i<k+9$ such that there are two 1's on the i-th place. Take the first i with this property and replace one of the 1's by the shifted equation (2). Again, we will have two 1's at the same place; and we can repeat this process indefinitely. We arrive in this way to the following state: we have an expansion of 1 such that for every $n \geq k+1$-th place there is one 1, there are places with more than one 1, but we have $\leq n$ of 1 in the n-th position. Hence

$$\varepsilon_1, \ldots, \varepsilon_{k-1}, 0, 1, \ldots < 1$$

i.e.

$$1 > \sum_{i=1}^{k-1} \varepsilon_i q^{-i} + \sum_{i=k+1}^{\infty} q^{-i},$$

and this means that there is no expansion of 1 beginning with ε_1, ε_{k-1}, 0.

b) If $\varepsilon_k=0$ and we have 0 on some of the places $k+1, \ldots, k+8$, then there is no expansion of 1 beginning with $\varepsilon_1, \ldots, \varepsilon_{k-1}, 1$. The proof is dual in some sense to that of a): write 1 in place of $\varepsilon_k=0$ and beginning from the $k+1$-th place write the shifted expansion (2) with "$-$" sign. Then we have one -1 on some of the places $\leq k+8$, which is not canceled by a 1, further we have on the $k+9$-th place ≤ 0. We omit the -1 digits at positions $\leq k+8$ and in order to get still an expansion of 1, compensate them by adding the correspondingly shifted (2) with $-$ sign. Thus we will have ≤ -1 on the $k+9$-th position, we omit it but compensate by adding the corresponding shifted (2) with $-$ sign; after on the $k+10$-th, $k+11$-th position we will have -1 and continuing this process we arrive to the following final state: at the positions $\geq k+1$ we have ≤ 0, on some places <0 and the n-th digit is $\geq -n$. This will be an expansion of 1, hence

$$\varepsilon_1,\ldots,\varepsilon_{k-1},1,0,0,\ldots > 1,$$

i.e. there is no expansion of 1 beginning with $\varepsilon_1,\ldots,\varepsilon_{k-1},1$.

c) Now consider (2) and look for the first digit which we can change. According to b) this can be only 1 and according to a) this can be only among those n digits, which are followed by nine zeros. Look at the k-th such digit, which is ε_{10k-1}. We call k-th expansion that one in which we omit ε_{10k-1} and after, beginning from the $10k$-th place we write the correspondingly shifted expansion (2) in order to get an expansion of 1 again. Now we show that the k-th expansion is unique if we fix its first $10k-1$ digits. Indeed, in places of indices $\geq 10k$ the 0 is not followed by eight 1 digits, hence by b), we can not change 0; according to a) we can change a digit 1 only in the case it is followed by eight 0. This is possible only if $\varepsilon_n=1$, $\varepsilon_{n+1}=\ldots=\varepsilon_{n+8}=0$, $\varepsilon_{n+9}=\varepsilon_{n+10}=1$. Change $\varepsilon_n=1$ by the shifted digit sequence (2). Then write the shifted digit sequence in place of ε_{n+10}, after in place of ε_{n+11} etc.; we can follow the procedure in a) and we get that it is not possible to change $\varepsilon_n=1$. This proves that the k-th expansion is unique if we fix its first $10k-1$ digits. At last it is enough to remark that there is no other expansion of 1 than the given k-th expansions and (2). Indeed, consider any expansion of 1 in terms of q defined by (2). Consider the first place where it differs from (2). According to the a) and b) this can be only a 1 digit of (2) which is followed by nine zeros. Replace this digit of (2) by 0 compensating it by adding a suitably shifted form of (2). This coincides with the arbitrary expansion considered in the first $10k$ digits and according to the uniqueness of the k-th expansion (proved above in c)) they must coincidence in every digits. The Theorem is proved. ∎

REMARK. In the definition of q given in (2) we have used the fact that for any given sequence (ε_i) with at least two 1 digits and at least one 0 digit there exists a unique q satisfying (1); hence our definition in (2) has meaning. The proof is easy and is given in [1].

References

[1] P. ERDŐS, M. HORVÁTH and I. JOÓ, On the uniqueness of the expansions $1 = \sum q^{-n_i}$, *Acta Math. Hung.* (To appear).

[2] A. BOGMÉR, M. HORVÁTH and A. SÖVEGJÁRTÓ, On some problems of I. Joó, *Acta Math. Hungar.*, **58** (1991), 153–155.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 133–146

ON LAGUERRE FUNCTIONS (ESTIMATE FOR THE SUMS OF THE SQUARES)

By

I. JOÓ

Department of Analysis, L. Eötvös University, Budapest

(Received October 18, 1991)

The aim of the present paper is to prove the following theorem, which makes possible to extend the results of M. HORVÁTH [1] for $-1<\alpha<-\frac{1}{2}$ (among others).

THEOREM 1. *Let $\alpha>-1$ denote $\ell_k^{(\alpha)}(x)$ the normed Laguerre polynomials and $v_n(x):=\sqrt{2}x^{\alpha+\frac{1}{2}}\cdot e^{-\frac{x^2}{3}}\cdot\ell_n^{(\alpha)}(x^2)$, $x>0$, the normed Laguerre functions. Let $d_1>0$ be an arbitrary fixed absolute constant and $\beta\geq 0$. Then we have*

$$(1)\qquad x^\beta\sum_{a\leq n<b}v_n^2(x)\leq c\sqrt{b-1}\cdot\min(x^\beta,b^\beta),\qquad (x>d_1>0),$$

$$(2)\qquad \max_{x>d_1>0}x^\beta\sum_{a\leq n<b}v_n^2(x)\asymp c\sqrt{b-a}\cdot b^\beta;$$

$$(3)\qquad \max_{0<x\leq d_1}x^\beta\sum_{a\leq n<b}v_n^2(x)\asymp b^\alpha(b-a),\qquad \beta=-2\alpha-1,\quad \alpha\geq-\frac{1}{2};$$

$$(4)\qquad \max_{0<x\leq d_1}x^\beta\sum_{a\leq n<b}v_n^2(x)\asymp\frac{b-a}{\sqrt{b}},\qquad \beta=-2\alpha-1,\quad -1<\alpha<-\frac{1}{2}.$$

In (2), (3), (4) we suppose that $b-a\geq c\sqrt{b}$ with a constant $c=c(\alpha,\beta,d_1)$ sufficiently large.

PROOF. We know [1] that for $\alpha>-1$

$$(5)\qquad x^\alpha e^{-x}\sum_{a\leq n<b}(\ell_n^{(\alpha)}(x))^2=$$

$$=\frac{i}{4\pi}\int_{-\pi}^{\pi}\frac{e^{-ix\operatorname{ctg}\frac{\varphi}{2}}}{\sin\frac{\varphi}{2}}\cdot\frac{J_\alpha\left(-\frac{x}{\sin\frac{\varphi}{2}}\right)}{\left(ie^{i\frac{\varphi}{2}}\right)^\alpha}\cdot e^{-i(a+O(1))\varphi}\cdot\frac{1-e^{-i(b-a)\varphi}}{1-e^{-i\varphi}}\,\mathrm{d}\varphi.$$

Hence

(6) $$x^\beta\sum_{a\le n<b}v_n^2(x)=$$

$$=x^{\beta+1}\cdot\frac{i}{2\pi}\int_{-\pi}^{\pi}\frac{e^{-ix^2\operatorname{ctg}\frac{\varphi}{2}}}{\sin\frac{\varphi}{2}}\cdot\frac{J_\alpha\left(-\frac{x^2}{\sin\frac{\varphi}{2}}\right)}{\left(ie^{i\frac{\varphi}{2}}\right)^\alpha}\cdot e^{-i(a+O(1))\varphi}\cdot\frac{1-e^{-i(b-a)\varphi}}{1-e^{-i\varphi}}\,\mathrm{d}\varphi.$$

First we prove (1). Obviously

$$x^\beta\sum_{a\le n<b}v_n^2(x)\le c\cdot x^{\beta+1}\cdot\int_{-\pi}^{\pi}\frac{\left|J_\alpha\left(-\frac{x^2}{\sin\frac{\varphi}{2}}\right)\right|}{\left|\sin\frac{\varphi}{2}\right|^2}\cdot\left|\sin\frac{b-a}{2}\varphi\right|\mathrm{d}\varphi.$$

Taking into account

$$\frac{x^2}{\left|\sin\frac{\varphi}{2}\right|}\ge c>0\qquad\text{if}\qquad x>d_1>0\quad\text{and}\quad|\varphi|\le\pi$$

and that ([2], p.168, (6))

$$\left|J_\alpha\left(-\frac{x^2}{\sin\frac{\varphi}{2}}\right)\right|\le c\frac{\sqrt{|\varphi|}}{x},$$

we have

(7) $$x^\beta\sum_{a\le n<b}v_n^2(x)\le c\cdot x^{\beta+1}\cdot\int_{-\pi}^{\pi}\frac{1}{x\cdot|\varphi|^{3/2}}\cdot\left|\sin\frac{b-a}{2}\varphi\right|\mathrm{d}\varphi=$$

$$=c\cdot x^\beta\int_0^\pi\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\mathrm{d}\varphi\le cx^\beta\cdot\sqrt{b-a}.$$

Using the "infinite–finite range inequality" or the asymptotic of the Laguerre functions we get

$$\max_{x>d_1>0}x^\beta\sum_{a\le n<b}v_n^2=\max_{d_1<x<c_0\cdot\sqrt{b}}x^\beta\sum_{a\le n<b}v_n^2(x).$$

hence (1) follows.

Now we prove the lower estimate in (2). It is well known that

$$J_\alpha(x)=\sqrt{\frac{2}{\pi x}}\cdot\left[\cos\left(x-\frac{\pi}{2}\left(\alpha+\frac{1}{2}\right)\right)+O\left(\frac{1}{x}\right)\right],\qquad (x>0),$$

and

$$J_\alpha(-x)=(-1)^\alpha J_\alpha(x),$$

([2], p. 168, (6)). Applying (5) and (6) for $\int_0^{2\pi}$ in place of $\int_{-\pi}^{\pi}$ we can estimate the part corresponding to the remaining part of (5) as follows:

$$c\int_0^{2\pi}\left(\frac{\sin\frac{\varphi}{2}}{x}\right)^{3/2}\cdot\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\sin^2\frac{\varphi}{2}}\mathrm{d}\varphi\le\frac{c}{x^{3/2}}\cdot\int_0^{\pi}\varphi^{-\frac{1}{2}}\mathrm{d}\varphi\le\frac{c}{x^{3/2}}\le c,$$

and the main term is:

$$I=\frac{i^{\alpha+1}}{4\pi}\cdot\frac{\sqrt{2}}{\sqrt{\pi}}\cdot\int_0^{2\pi}e^{-i\left(x\mathrm{ctg}\frac{\varphi}{2}+\frac{b+a}{2}+O(1)\right)\varphi}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\cos\left(\frac{x}{\sin\frac{\varphi}{2}}-\frac{\pi\alpha}{2}-\frac{\pi}{4}\right)\mathrm{d}\varphi.$$

Consider the partition

$$\int_0^{2\pi}=\int_0^{\frac{1}{(b-a)\varepsilon}}+\int_{\frac{1}{(b-a)\varepsilon}}^{2\pi-\frac{1}{(b-a)\varepsilon}}+\int_{2\pi-\frac{1}{(b-a)\varepsilon}}^{2\pi}=I_1+I_2+I_3,$$

where $\varepsilon=\varepsilon(\alpha)>0$ will be chosen to be sufficiently small in below.

Here we have

$$|I_2|\le c\int_{\frac{1}{(b-a)\varepsilon}}^{\pi}\varphi^{-\frac{3}{2}}\mathrm{d}\varphi\le c\sqrt{(b-a)\varepsilon}. \tag{8}$$

We shall apply the identity $\cos t=\frac{e^{it}+e^{-it}}{2}$ in I_1 and I_3. For this define

$$I_{11}=\int_0^{\frac{1}{(b-a)\varepsilon}}e^{-i\left(\mathrm{ctg}\frac{\varphi}{2}+\frac{x}{\sin\frac{\varphi}{2}}-\frac{\pi\alpha}{2}-\frac{\pi}{4}+\left(\frac{b+a}{2}+O(1)\right)\varphi\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi, \tag{9}$$

$$I_{12}=\int_0^{\frac{1}{(b-a)\varepsilon}}e^{-i\left(x\mathrm{ctg}\frac{\varphi}{2}-\frac{x}{\sin\frac{\varphi}{2}}+\frac{\pi\alpha}{2}+\frac{\pi}{4}+\left(\frac{b-a}{2}+O(1)\right)\varphi\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi, \tag{10}$$

$$(11)\ I_{31}=\int_{2\pi-\frac{1}{(b-a)\varepsilon}}^{2\pi} e^{-i\left(x\operatorname{ctg}\frac{\varphi}{2}+\frac{x}{\sin\frac{\varphi}{2}}-\frac{\pi\alpha}{2}-\frac{\pi}{4}+\left(\frac{b+a}{2}+O(1)\right)\varphi\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi,$$

$$(12)\ I_{32}=\int_{2\pi-\frac{1}{(b-a)\varepsilon}}^{2\pi} e^{-i\left(x\operatorname{ctg}\frac{\varphi}{2}-\frac{x}{\sin\frac{\varphi}{2}}+\frac{\pi\alpha}{2}+\frac{\pi}{4}+\left(\frac{b+a}{2}+O(1)\right)\varphi\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi,$$

We know that

$$\frac{1}{\sin\frac{\varphi}{2}}-\operatorname{ctg}\frac{\varphi}{2}=\frac{1-\cos\frac{\varphi}{2}}{\sin\frac{\varphi}{2}}=\frac{\frac{\varphi^2}{8}+O(\varphi^4)}{\frac{\varphi}{2}+O(\varphi^3)}=\frac{\varphi}{4}+O(\varphi^3).$$

Let $x_0:=4\left(\alpha-1+\frac{b+a}{2}+O(1)\right)$. Then we have

$$(13)\ \ I_{11}=\int_0^{\frac{1}{(b-a)\varepsilon}} e^{i\left(\frac{2x_0}{\sin\frac{\varphi}{2}}-\frac{x_0\varphi}{4}-\frac{\pi\alpha}{2}+O(x_0\varphi^3)+\left(\frac{b+a}{2}+O(1)\right)\varphi\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi.$$

Here $O(x_0\varphi^3)$ gives a small term after integration, in I_{11}, namely

$$(14)\qquad \int_0^{\frac{1}{(b-a)\varepsilon}} x_0\varphi^3\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\mathrm{d}\varphi\le c(b+a)(b-a)^{-\frac{5}{2}}\cdot\varepsilon^{-\frac{7}{2}}.$$

We know

$$\frac{1}{\left(\sin\frac{\varphi}{2}\right)^{3/2}}=\frac{1}{\left(\frac{\varphi}{2}\right)^{3/2}}+O(\sqrt{\varphi}),$$

hence $O(\sqrt{\varphi})$ gives the following remainder term in I_{11}

$$(15)\qquad \int_0^{\frac{1}{(b-a)\varepsilon}} \sqrt{\varphi}\cdot\left|\sin\frac{b-a}{2}\varphi\right|\mathrm{d}\varphi\le c(b-a)^{-\frac{3}{2}}\cdot\varepsilon^{-\frac{5}{2}},$$

i.e. the main part of I_{11} is:

$$\tilde{I}_{11}:=\int_0^{\frac{1}{(b-a)\varepsilon}} e^{-i\left(\frac{2x_0}{\sin\frac{\varphi}{2}}-\frac{\pi\alpha}{2}-\frac{\pi}{4}\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi.$$

Because

$$\frac{1}{\sin\frac{\varphi}{2}}-\frac{1}{\frac{\varphi}{2}}=\frac{\frac{\varphi}{2}-\sin\frac{\varphi}{2}}{\frac{\varphi}{2}\sin\frac{\varphi}{2}}=\frac{\varphi}{12}+O(\varphi^3),$$

hence

$$\tilde{I}_{11}=\int_0^{\frac{1}{(b-a)\varepsilon}} e^{-i\left(\frac{4x_0}{\varphi}-x_0\frac{\varphi}{6}+O(x_0\varphi^3)\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi.$$

The remainder term gives after integration

$$\int_0^{\frac{1}{(b-a)\varepsilon}} x_0\varphi^3\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\mathrm{d}\varphi\le c(b+a)(b-a)^{-\frac{5}{2}}\cdot\varepsilon^{-\frac{7}{2}}, \tag{16}$$

hence we have to estimate

$$\hat{I}_{11}:=\int_0^{\frac{1}{(b-a)\varepsilon}} e^{-ix_0\left(\frac{4}{\varphi}+\frac{\varphi}{6}\right)}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi=\int_0^{\frac{\varepsilon}{b-a}}+\int_{\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}}.$$

Here

$$\left|\int_0^{\frac{\varepsilon}{b-a}}\right|\le c\int_0^{\frac{\varepsilon}{b-a}}\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\mathrm{d}\varphi\le c\sqrt{b-a}\sqrt{\varepsilon}, \tag{17}$$

and

$$\int_{\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}}=\int_{\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}}\left(\frac{1}{6}-\frac{4}{\varphi^2}\right)e^{-ix_0\left(\frac{4}{\varphi}+\frac{\varphi}{6}\right)}\cdot\frac{6\varphi^2}{\varphi^2-24}\cdot\frac{\sin\frac{b-a}{2}\varphi}{\left(\frac{\varphi}{2}\right)^{3/2}}\mathrm{d}\varphi.$$

Using

$$\frac{6\varphi^2}{\varphi^2-24}=-\frac{\varphi^2}{4}+O(\varphi^4),$$

we can estimate the contribution of the remainder term as before. Estimate the main term integrating by part:

$$\left|\int_{\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}}\left(\frac{1}{6}-\frac{4}{\varphi^2}\right)e^{-ix_0\left(\frac{4}{\varphi}+\frac{\varphi}{6}\right)}\cdot\varphi^{\frac{1}{2}}\cdot\sin\frac{b-a}{2}\varphi\mathrm{d}\varphi\right|\le$$

$$\leq \left| \left[\frac{e^{-ix_0\left(\frac{4}{\varphi}+\frac{\varphi}{6}\right)}}{-ix_0} \cdot \varphi^{\frac{1}{2}} \cdot \sin \frac{b-a}{2}\varphi \right]_{\varphi=\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}} \right| +$$

$$+ \left| \int_{\frac{\varepsilon}{b-a}}^{\frac{1}{(b-a)\varepsilon}} \frac{e^{-ix_0\left(\frac{4}{\varphi}+\frac{\varphi}{6}\right)}}{-ix_0} \cdot \left[\frac{1}{2}\varphi^{-\frac{1}{2}} \cdot \sin \frac{b-a}{2}\varphi + \frac{b-a}{2}\varphi^{\frac{1}{2}} \cdot \cos \frac{b-a}{2}\varphi \right] \mathrm{d}\varphi \right| \leq$$

$$\leq c\frac{1}{b+a} \cdot \frac{1}{(b-a)^{1/2}} \cdot \varepsilon^{-3/2} + c\frac{1}{b+a} \cdot \frac{1}{(b-a)^{1/2}} \cdot \varepsilon^{-3/2}.$$

Summarizing our estimates we obtain

$$|I_{11}| \leq c(b+a)(b-a)^{-\frac{5}{2}} \cdot \varepsilon^{-\frac{7}{2}} + c(b-a)^{-\frac{3}{2}} \cdot \varepsilon^{-\frac{5}{2}} + c\sqrt{b-a} \cdot \sqrt{\varepsilon} + \tag{18}$$
$$+ c\frac{1}{b+a}(b-a)^{-\frac{1}{2}} \cdot \varepsilon^{-\frac{3}{2}} \leq c\sqrt{b-a} \cdot \sqrt{\varepsilon}.$$

Similarly, if we write $(2\pi - \varphi)$ in place of φ:

$$I_{32} = -e^{-i\pi\alpha} \cdot \int_0^{\frac{1}{(b-a)\varepsilon}} e^{-i\left(-x_0 \operatorname{ctg}\frac{\varphi}{2} - x_0\frac{1}{\sin\frac{\varphi}{2}} + \frac{\pi\alpha}{2} + \frac{\pi}{4} - \left(\frac{b+a}{2} + O(1)\right)\right)\varphi} \cdot$$

$$\cdot \frac{\sin\left(\frac{b-a}{2} + O(1)\right)\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}} \mathrm{d}\varphi = -e^{-i\pi\alpha} \cdot \overline{I}_{11},$$

we get

$$|I_{32}| \leq c\sqrt{b-a} \cdot \sqrt{\varepsilon}. \tag{19}$$

Now consider I_{12}. Because $x_0\left(\operatorname{ctg}\frac{\varphi}{2} - \frac{1}{\sin\frac{\varphi}{2}}\right) = -\frac{x_0\varphi}{4} + O(x_0\varphi^3)$ and

$$\int_0^{\frac{1}{(b-a)\varepsilon}} x_0\varphi^3 \frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}} \mathrm{d}\varphi \leq c(b+a)(b-a)^{-\frac{5}{2}} \cdot \varepsilon^{-\frac{7}{2}} \leq c\sqrt{b-a} \cdot \sqrt{\varepsilon}, \tag{20}$$

hence

$$\left| I_{12} - e^{-i\left(\frac{\pi\alpha}{2}+\frac{\pi}{4}\right)} \cdot \int_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{\left(\sin\frac{\varphi}{2}\right)^{3/2}} \mathrm{d}\varphi \right| \leq c\sqrt{b-a} \cdot \sqrt{\varepsilon}. \tag{21}$$

Using $(\sin\frac{\varphi}{2})^{-\frac{3}{2}} = (\frac{\varphi}{2})^{-\frac{3}{2}} + O(\sqrt{\varphi})$, we get by similar ideas as used above:

$$\left| I_{12} - e^{-i(\frac{\pi\alpha}{2}+\frac{\pi}{4})} \cdot \int\limits_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{(\frac{\varphi}{2})^{3/2}} \mathrm{d}\varphi \right| \le c\sqrt{b-a}\cdot\sqrt{\varepsilon}. \tag{22}$$

Because $I_{31} = -e^{-i\pi\alpha}\cdot \overline{I}_{12}$, hence

$$\left| I_{31} + e^{-i(\frac{\pi\alpha}{2}-\frac{\pi}{4})} \cdot \int\limits_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{(\frac{\varphi}{2})^{3/2}} \mathrm{d}\varphi \right| \le c\sqrt{b-a}\cdot\sqrt{\varepsilon},$$

consequently

$$\left| I_{12} + I_{31} + i\sqrt{2}e^{-i\frac{\pi\alpha}{2}} \cdot \int\limits_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{(\frac{\varphi}{2})^{3/2}} \mathrm{d}\varphi \right| \le c\sqrt{b-a}\cdot\sqrt{\varepsilon}. \tag{23}$$

On the other hand

$$\left(\frac{b-a}{2}\right)^{-\frac{1}{2}} \cdot \int\limits_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{\varphi^{3/2}} \mathrm{d}\varphi = \int\limits_0^{\frac{1}{2\varepsilon}\cdot\frac{b-a}{b-a}} \frac{\sin u}{u^{3/2}} \mathrm{d}u \to \int\limits_0^{+\infty} \frac{\sin u}{u^{3/2}} \mathrm{d}u, \quad \varepsilon \to 0+,$$

hence for $0 < \varepsilon < \varepsilon_0$ we have

$$\int\limits_0^{\frac{1}{(b-a)\varepsilon}} \frac{\sin\frac{b-a}{2}\varphi}{\varphi^{3/2}} \mathrm{d}\varphi \ge c\sqrt{b-a}, \quad (c > 0).$$

Taking into account (23) we obtain $|I_{12}+I_{31}| \ge c(\alpha)\sqrt{b-a}$. Because $|I_{11} + I_{32}| \le c\sqrt{b-a}\cdot\sqrt{\varepsilon}$, hence we obtain $|I| \ge c(\alpha)\sqrt{b-a}$ if $\varepsilon > 0$ is small enough. Thus (2) is proved.

Now prove (3) and (4). Because

$$x^\beta v_n^2(x) = 2x^{\beta+2\alpha+1}\cdot e^{-x^2}\cdot \left(\ell_n^{(\alpha)}(x^2)\right)^2$$

the singularity disappears at zero at $\beta+2\alpha+1=0$, i.e. $\beta = -2\alpha-1$ and in this case we have

$$x^\beta \sum_{a\le n<b} v_n^2(x) = 2e^{-x^2} \sum_{a\le n<b} \left[\Gamma(\alpha+1)\binom{n+\alpha}{n}\right]^{-1} \cdot \left(L_n^{(\alpha)}(x^2)\right)^2. \tag{24}$$

First consider the case

1. $a \geq qb$ for some $0 < q < 1$.

a) If $0 < x \leq \frac{c}{\sqrt{b}}$ for some small constant $0 < c = c(\alpha)$ then use ([3], (8.22.4), (8.22.5), p.206)

$$(25)\quad e^{-\frac{x}{2}} \cdot x^{\frac{\alpha}{2}} \cdot L_n^{(\alpha)} = N^{-\frac{\alpha}{2}} \frac{\Gamma(n+\alpha+1)}{n!} J_\alpha \left\{2(Nx)^{\frac{1}{2}}\right\} + x^{\frac{\alpha}{2}+2} \cdot O(n^\alpha),$$
$$\left(\alpha > -1, \quad N = n + \frac{\alpha+1}{2}, \quad 0 < x \leq \frac{c}{n}\right)$$

and the remainder term is uniform in x; further take into account $J_\alpha u \asymp u^\alpha$, $u > 0$, $u \approx 0$. Because $N \asymp b$ and $x^2 \leq \frac{c}{b}$, hence we can apply the last estimate in (25), so we get from (24)

$$(26)\quad x^\beta \sum_{a \leq n < b} v_n^2(x) \asymp c(\alpha) \sum_{a \leq n < b} \frac{1}{n^\alpha} b^{2\alpha} \left(1 + x^{\alpha+4} \cdot O(1)\right) \asymp b^\alpha (b-a).$$

b) In this case $\frac{c}{\sqrt{b}} \leq x < d_1$, use ([3], (8.22.6), p.207)

$$(27)\quad L_n^{(\alpha)}(x) = \pi^{-\frac{1}{2}} e^{\frac{x}{2}} x^{-\frac{\alpha}{2}-\frac{1}{4}} \cdot n^{\frac{\alpha}{2}-\frac{1}{4}} \cdot$$
$$\cdot \left\{\cos\left[2(nx)^{\frac{1}{2}} - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right] + (nx)^{-\frac{1}{2}} O(1)\right\}, \qquad \left(\alpha > -1, \quad \frac{c}{n} \leq x \leq c\right),$$

we get

$$x^\beta \sum_{a \leq n < b} v_n^2(x) \asymp$$
$$\asymp \frac{1}{b^\alpha} \sum_{a \leq n < b} x^{-2\alpha-1} b^{\alpha-\frac{1}{2}} \cdot \left\{\cos^2\left(2x\sqrt{n} - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + \frac{1}{x\sqrt{b}} \cdot O(1)\right\}.$$

Here the contribution of the remainder term is $x^{-2\alpha-2} \cdot \frac{(b-a)}{b}$. Because $-2\alpha - 2 < 0$ hence $x^{-2\alpha-2} \cdot \frac{(b-a)}{b} \leq cb^\alpha(b-a)$. In the case of $-2\alpha - 1 \leq 0$ i.e. $-\frac{1}{2} \leq \alpha$ we have $x^{-2\alpha-1} \leq b^{\alpha+\frac{1}{2}}$ and so

$$(28)\quad x^\beta \sum_{a \leq n < b} v_n^2(x) \leq cb^\alpha(b-a).$$

In the case of $-2\alpha - 1 > 0$, because $-1 < \alpha < -1/2$, we have

$$(29)\quad x^\beta \sum_{a \leq n < b} v_n^2(x) \leq c\frac{(b-a)}{\sqrt{b}}.$$

Taking into account (26):

$$\max_{0<x<d_1} x^{\beta} \sum_{a\le n<b} v_n^2(x) \asymp b^{\alpha}(b-a), \quad \left(a\ge qb,\ \alpha\ge -\frac{1}{2},\ \beta=-2\alpha-1\right). \tag{30}$$

In order to prove that (29) is exact, we need the following

LEMMA. *Let $a\neq 0$ be any fixed real number and $d>0$. Then*

$$\sum_{d\le m<b} \cos a\sqrt{m} = O(1)\sqrt[4]{b}(\sqrt{b}-\sqrt{d})+O(1)\sqrt{b} \tag{31}$$

where $O(1)$ depends (continuously) only on a (independent on b and on d).

PROOF. Investigate the distribution of $a\cdot\sqrt{m}$ mod 2π. We may suppose that $a>0$. Let $j>0$ be any fixed integer (which may depend on b and on d) which will be chosen later and let $0\le k<j-1$, $0\le l\le \frac{a}{2\pi}\sqrt{b}$ integers. Then

$$2l\pi + k\frac{2\pi}{j} \le a\sqrt{m} < 2l\pi+(k+1)\frac{2\pi}{j}, \tag{32}$$

for suitably chosen k and l. If m satisfies (32) then

$$\cos a\sqrt{m} = \cos k\frac{2\pi}{j} + O\left(\frac{1}{j}\right) \tag{33}$$

i.e.

$$\sum_{d\le m<b} \cos a\sqrt{m} = \sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} \sum_{m=\left(\frac{2\pi}{a}\right)^2\cdot\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2} \cos a\sqrt{m} + O(\sqrt{b}) =$$

$$= \sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} \sum_{m=\left(\frac{2\pi}{a}\right)^2\cdot\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2} \cos k\frac{2\pi}{j} +$$

$$+O(1) \sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} \sum_{m=\left(\frac{2\pi}{a}\right)^2\cdot\left(l+\frac{k}{j}\right)^2}^{\left(\frac{2\pi}{a}\right)^2\left(l+\frac{k+1}{j}\right)^2} \frac{1}{j} + O(\sqrt{b}).$$

The contribution of the remainder term is

$$\sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} \frac{1}{j}\left[\frac{1}{j}\left(2l+\frac{2k+1}{j}\right)+O(1)\right] =$$

$$= \frac{1}{j} \sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} 2l + O(1) \sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} 1 \asymp \frac{b-d}{j} + O(1)(\sqrt{b}-\sqrt{d}).$$

The contribution of the main term is

$$\sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} \cos k\frac{2\pi}{j} \left(\frac{2\pi}{a}\right)^2 \left[\frac{1}{j}\left(2l+\frac{2k+1}{j}\right) + O(1)\right].$$

It is known that
$$\sum_{k=0}^{j-1} \cos k\frac{2\pi}{j} = 0,$$

hence the main term can be estimated as follows:

$$\sum_{l=\frac{a}{2\pi}\sqrt{d}}^{\frac{a}{2\pi}\sqrt{b}} \sum_{k=0}^{j-1} O(1) = O(1)j(\sqrt{b}-\sqrt{d}),$$

i.e.

$$\sum_{d\le m<b} \cos a\sqrt{m} = O(1)j(\sqrt{b}-\sqrt{d}) + O(1)\frac{b-d}{j} + O(\sqrt{b}).$$

If $j = \sqrt[4]{b}$ then the statement of the Lemma follows.

REMARK. We can prove similarly

$$\sum_{d\le m<b} \sin a\sqrt{m} = O(1)\sqrt[4]{b}(\sqrt{b}-\sqrt{d}) + O(\sqrt{b}). \tag{34}$$

Now we are in the position to prove the exactness of (29). To this let $x_0>0$ be any fixed real number, $\frac{d_1}{2} < x_0 < d_1$, then we have by the Lemma

$$x_0^\beta \sum_{a\le n<b} v_n^2(x_0) \asymp \frac{1}{\sqrt{b}} \sum_{a\le n<b} \cos^2\left(2x_0\sqrt{n} - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + O(1)b^\alpha(b-a) \asymp$$

$$\asymp \frac{b-a}{\sqrt{b}} + \frac{1}{\sqrt{b}} \sum_{a\le n<b} \cos\left(4x_0\sqrt{n} - \alpha\pi - \frac{\pi}{2}\right) + O(1)b^\alpha(b-a) \asymp$$

$$\asymp \frac{b-a}{\sqrt{b}} + \frac{1}{\sqrt{b}} \sum_{a\le n<b} (\sin 4x_0\sqrt{n}\cdot\cos\alpha\pi - \cos 4x_0\sqrt{n}\cdot\sin\alpha\pi) \asymp$$

$$\asymp \frac{b-a}{\sqrt{b}} + \frac{1}{\sqrt{b}}\sqrt[4]{b}(\sqrt{b}-\sqrt{a}) + O(1) \asymp \frac{b-a}{\sqrt{b}}.$$

Here we have used (31), (34) and the fact that $-1<\alpha<-\frac{1}{2}$.

Now investigate the case

2. $a \le qb$.

a) If $0 < x \le \frac{c_0}{\sqrt{b}}$, c_0 is small then we can follow the calculations of 1. a):

(35)
$$x^{\beta} \sum_{a \le n < b} v_n^2(x) \asymp \sum_{a \le n < b} \frac{1}{n^{\alpha}} n^{2\alpha} \left(1 + x^{\alpha+4} \cdot O(1)\right) \asymp$$
$$\asymp \sum_{a \le n < b} n^{\alpha} \asymp b^{\alpha+1} \asymp b^{\alpha}(b-a).$$

b) In the case of $\frac{c}{\sqrt{a}} \le x < d_1$ we can follow the calculations of 1. b):

$$x^{\beta} \sum_{a \le n < b} v_n^2(x) \asymp$$
$$\asymp \sum_{a \le n < b} \frac{1}{n^{\alpha}} x^{-2\alpha-1} \cdot n^{\alpha - \frac{1}{2}} \cdot \left\{ \cos^2 \left(2x\sqrt{n} - \frac{\alpha\pi}{2} - \frac{\pi}{4} \right) + \frac{1}{x\sqrt{n}} O(1) \right\}.$$

The contribution of the remainder term is $x^{-2\alpha-2}(\log b - \log a)$. Because $-2\alpha - 2 < 0$, hence $x^{-2\alpha-2}(\log b - \log a) \le ca^{\alpha+1}(\log b - \log a)$.

In the case of $-2\alpha - 1 \le 0$, i.e. $-\frac{1}{2} \le \alpha$, we have $x^{-2\alpha-1} \le ca^{\alpha+\frac{1}{2}}$ so

(28′)
$$x^{\beta} \sum_{a \le n < b} v_n^2(x) \le ca^{\alpha+\frac{1}{2}} \sqrt{b} + ca^{\alpha+1}(\log b - \log a).$$

In the case of $-2\alpha - 1 > 0$ i.e. $-1 < \alpha < -\frac{1}{2}$ we get

(29′)
$$x^{\beta} \sum_{a \le n < b} v_n^2(x) \le c\sqrt{b} + ca^{\alpha+1}(\log b - \log a).$$

c) If $\frac{c}{\sqrt{b}} < x < \frac{c}{\sqrt{a}}$ then

$$x^{\beta} \sum_{a \le n < b} v_n^2(x) = x^{\beta} \sum_{a \le n < \frac{c}{x^2}} v_n^2(x) + x^{\beta} \sum_{\frac{c}{x^2} \le n < b} v_n^2(x) =: S_1 + S_2.$$

Apply for S_1 the method used in 1. a):

(36)
$$S_1 = x^{\beta} \sum_{a \le n < \frac{c}{x^2}} v_n^2(x) \asymp \sum_{a \le n < \frac{c}{x^2}} n^{\alpha} \asymp \left(\frac{c}{x^2}\right)^{\alpha+1} = cx^{-2\alpha-2} \le cb^{\alpha+1} \asymp b^{\alpha}(b-a)$$

and for S_2 the method used in 1. b):

$$S_2 = x^{\beta} \sum_{\frac{c}{x^2} \le n < b} v_n^2(x) \asymp$$

$$\asymp x^{-2\alpha-1} \sum_{\frac{c}{x^2}\le n<b} \frac{1}{\sqrt{n}} \left\{ \cos^2\left(2x\sqrt{n}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) + \frac{1}{x\sqrt{n}}O(1) \right\}.$$

The contribution of the remainder term is $x^{-2\alpha-2}\left(\log b - \log \frac{c}{x^2}\right) \le b^{\alpha+1}$ because the function $f(x) = x^{-2\alpha-2}\left(\log b - \log\frac{c}{x^2}\right)$ attains its maximum at $x_0 = \frac{c}{\sqrt{b}}$.

We can estimate the main term as follows:

$$(28'') \qquad x^{-2\alpha-1} \sum_{\frac{c}{x^2}\le n<b} \frac{1}{\sqrt{n}} \cos^2\left(2x\sqrt{n}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) \le cb^{\alpha+1},$$

if $-2\alpha-1\le 0$, i.e. $-\frac{1}{2}\le\alpha$;

$$(29'') \qquad x^{-2\alpha-1} \sum_{\frac{c}{x^2}\le n<b} \frac{1}{\sqrt{n}} \cos^2\left(2x\sqrt{n}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) \le ca^{\alpha+\frac{1}{2}}\sqrt{b},$$

if $-2\alpha-1>0$ i.e. $-1<\alpha<-\frac{1}{2}$.

That is the case of $\alpha\ge -1/2$, taking into account (35), (28′), (36), (28″), $a\le qb$

$$(37) \qquad \max_{0<x<d_1} x^{\beta} \sum_{a\le n<b} v_n^2(x) \asymp b^{\alpha}(b-a), \quad (\beta=-2\alpha-1)$$

follows. In the case of $-1<\alpha<-1/2$, take into account (35), (29′), (36), (29″), $a\le qb$

$$(38) \qquad x^{\beta} \sum_{a\le n<b} v_n^2(x) \le c\frac{b-a}{\sqrt{b}}.$$

In order to prove the exactness of (38) it is enough to prove that of (29′). Let $x_0>0$ be any fixed real number $\frac{d_1}{2}<x_0<d_1$, then:

$$x^{\beta} \sum_{a\le n<b} v_n^2(x) \asymp \sum_{a\le n<b} \frac{1}{n^{\alpha}} x_0^{-2\alpha-1}\cdot n^{\alpha-\frac{1}{2}} \cos^2\left(2x_0\sqrt{n}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right) +$$

$$+O(1)a^{\alpha+1}(\log b-\log a) \asymp \sqrt{b}-\sqrt{a} + \sum_{a\le n<b} \frac{1}{\sqrt{n}} \cos\left(4x_0\sqrt{n}-\alpha\pi-\frac{\pi}{2}\right) =$$

$$= \sqrt{b}-\sqrt{a} + \sum_{a\le n<b} \frac{1}{\sqrt{n}} (\sin 4x_0\sqrt{n}\cdot\cos\alpha\pi - \cos 4x_0\sqrt{n}\cdot\sin\alpha\pi).$$

By (31) and (34) we get $\sum_{n\le x}\cos t\sqrt{n}=O(1)x^{3/4}$, $\sum_{n\le x}\sin t\sqrt{n}=O(1)x^{3/4}$ hence summing by part we get

$$\sum_{a\le n<b}\frac{\cos t\sqrt{n}}{\sqrt{n}}=O(1)\frac{b^{3/4}}{\sqrt{b}}-O(1)\frac{a^{3/4}}{\sqrt{a}}+O(1)\int_a^b x^{3/4}\cdot x^{-3/2}\,\mathrm{d}x=O(1)b^{1/4},$$

and

$$\sum_{a\le n<b}\frac{\sin t\sqrt{n}}{\sqrt{n}}=O(1)b^{1/4},$$

consequently

$$x_0^{\beta}\sum_{a\le n<b}v_n^2(x_0)\asymp\sqrt{b}-\sqrt{a}+\sqrt[4]{b}\asymp\frac{b-a}{\sqrt{b}}.$$

Thus we have proved that (29′) and (38) are exact. Hence (3) and (4) is proved also in the case of $a=o(b)$. Our Theorem is completely proved.

At last we remark that

$$\int_0^{\infty}x^{\alpha}e^{-x}\sum_{k=1}^{n}\frac{(L_n^{(\alpha)}(x))^2}{|x-x_k|}x_k^{\gamma}e^{-\tau x_k}\mathrm{d}x\asymp n^{\alpha},\quad(\alpha>-1,\ n=1,2,\ldots),\tag{39}$$

$0<x_1<\ldots<x_n$ denote the zeros of the Laguerre polynomial $L_n^{(\alpha)}(x)$ ([3]); where $\tau>0$ and $\gamma>0$ are arbitrary real numbers and the implicit constants depend only on α, γ, τ, but independent on n. The upper estimate is proved in [4], hence we prove here the lower one. We will need this estimate in our next papers.

THEOREM 2. *We have*

$$I:=\int_0^{\infty}x^{\alpha}e^{-x}\sum_{k=1}^{n}\frac{|L_n^{(\alpha)}(x)|^2}{|x-x_k|}\cdot x_k^{\gamma}e^{-\tau x_k}\mathrm{d}x\ge cn^{\alpha}\tag{40}$$

$$(\alpha>-1;\ n=0,1,2,\ldots).$$

PROOF. We may suppose that $n>n_0(\alpha,\gamma,\tau)$. We know: $x_k\asymp\frac{k^2}{n}$, i.e. in case of $\sqrt{n}\le k\le 2\sqrt{n}$ we have $x_k\asymp 1$, hence

$$I\ge c\int_1^2\sum_{k=\sqrt{n}}^{2\sqrt{n}}\frac{|L_n^{(\alpha)}(x)|^2}{|x-x_k|}x_k^{\gamma}e^{-\tau x_k}\mathrm{d}x\ge$$

$$\geq c\int_1^2 \sum_{k=\sqrt{n}}^{2\sqrt{n}} |L_n^{(\alpha)}(x)|^2 \mathrm{d}\,x = c\sqrt{n}\int_1^2 |L_n^{(\alpha)}(x)|^2 \mathrm{d}\,x.$$

We know from (27) that

$$|L_n^{(\alpha)}(x)|^2 \asymp n^{\alpha-1/2}\left\{\cos\left(2\sqrt{nx}-\frac{\alpha\pi}{2}-\frac{\pi}{4}\right)+O\left(\frac{1}{\sqrt{n}}\right)\right\}^2 =$$
$$= n^{\alpha-1/2}\left\{\cos^2\left(4\sqrt{nx}-\alpha\pi-\frac{\pi}{2}\right)+O\left(\frac{1}{\sqrt{n}}\right)\right\}.$$

Using this estimate we can continue our lower estimate as follows:

$$cn^\alpha\int_1^2\left(\cos^2\left(4\sqrt{nx}-\alpha\pi-\frac{\pi}{2}\right)+O\left(\frac{1}{\sqrt{n}}\right)\right)\mathrm{d}\,x \geq cn^\alpha,$$

and thus the Theorem 2 is proved.

References

[1] M. Horváth, On eigenfunction expansions, *Annales Univ. Sci. Budapest.*, Sect Math., **32** (1989), 159–190.

[2] A. F. Nikiforov, V. B. Uvarov, *Special functions of Mathematical Physics*, (in Russian), Moscow, "Nauka", 1984.

[3] G. Szegő, *Orthogonal Polynomials*, New York, 1939.

[4] I. Joó, On the order of mean convergence of interpolating processes, *Annales Univ. Sci. Budapest.*, Sect. Math., **33** (1990), 147–160.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 147–150

NOTE ON CONTINUITY OF FUNCTIONS DERIVED FROM EQUIVALENCE RELATIONS

By

L. HORVÁTH

Department of Mathematics, University of Veszprém, Veszprém

(Received October 11, 1991)

1. Introduction

Throughout this paper X denotes a finite-dimensional real normed space.

Let U be a non-empty open subset of $\mathbb{R}\times X$, and let E be an equivalence relation in U satisfying the properties (P1) and (P2).

(P1) The equivalence classes are continuous functions whose domains are open intervals.

We denote the equivalence class containing $(t,x)\in U$ by $\phi_{t,x}$ and the domain of $\phi_{t,x}$ by $I_{t,x}\cdot \phi_{t,x}$ is said to be trajectory passing through (t,x).

(P2) If $(t_0,x_0)\in U$ and $\mathbb{J}\subseteq I_{t_0,x_0}$ is a compact interval, then there exists a neighbourhood $V(t_0,x_0,\mathbb{J})$ of (t_0,x_0) in U such that $\mathbb{J}\subseteq I_{t,x}$ for every $(t,x)\in V(t_0,x_0,\mathbb{J})$.

Let $U^*\subseteq\mathbb{R}\times\mathbb{R}\times X$ and $\Phi: U^*\to X$ be defined by

$$U^*=\left\{(s,t,x):(t,x)\in U,\ s\in I_{t,x}\right\},$$
$$\Phi(s,t,x)=\phi_{t,x}(s).$$

We study the continuity of Φ in this paper. It is proved that Φ is continuous if $X=\mathbb{R}$. This problem comes from the problem of the dependence of the solutions of an ordinary differential equation on the initial conditions.

Let U be a non-empty open subset of $\mathbb{R}\times X$, and let $f:U\to X$ be a continuous function such that the initial value problem

$$\begin{cases} x'(t)=f(t,x(t))\\ x(t_0)=x_0\end{cases}$$

has unique solution for every $(t_0,x_0)\in U$. The solution reaching to the boundary of U in both directions is denoted by ϕ_{t_0,x_0}. Let the domain of ϕ_{t_0,x_0} be the interval I_{t_0,x_0}. Consider the equivalence relation E in

U whose equivalence class containing $(t,x)\in U$ is $\phi_{t,x}$. Then E has the properties (P1) and (P2) (see [1]). It is well known that U^* is open and Φ is continuous in this case (see [1]).

2. The main theorem

The following well known result (see [1]) will be useful.

LEMMA 1. Let X, Z be topological spaces, let X be compact, let Y be a metric space, and let $f: X\times Z\to Y$. Then f is continuous if and only if both $x\to f(x,z)$ is continuous on X for each $z\in Z$ and also, for each $z_0\in Z$, $f(x,z)\to f(x,z_0)$ uniformly for $x\in X$ as $z\to z_0$.

LEMMA 2. Let U be a non-empty open subset of $\mathbb{R}\times X$, and let E be an equivalence relation in U satisfying (P1) and (P2).

(i) U^* is open.

(ii) If K is a compact subset of U, then there exists a compact interval $\mathbb{J}\subseteq I_{t_0,x_0}$ such that $(t,\phi_{t_0,x_0}(t))\in U\setminus K$ for every $t\in I_{t_0,x_0}\setminus\mathbb{J}$.

PROOF. (i) It is obvious.

(ii) Since I_{t_0,x_0} is an open interval in $\mathbb{R}$, I_{t_0,x_0} is the union of countable many compact intervals $[a_n,b_n]$ which can be chosen so that $[a_n,b_n]$ lies in the interior of $[a_{n+1},b_{n+1}]$ for $n=1,2,\ldots$. If the result is false we can find a sequence $\{t_n\}$ such that $t_n\in I_{t_0,x_0}\setminus[a_n,b_n]$ and $(t_n,\phi_{t_0,x_0}(t_n))\in K$ for every n. We can suppose that $t_n\in]a,a_n[$ for each n, where a is the left-hand endpoint of I_{t_0,x_0}.

Since K is compact there exists a subsequence $\{t_{n_k}\}$ such that $\{(t_{n_k},\phi_{t_0,x_0}(t_{n_k}))\}$ converges to a point (a,x) of K.

$I_{a,x}$ is an open interval in $\mathbb{R}$, hence there is a real number $c<a$ such that $\overline{\mathbb{J}}=[c,a]\subset I_{a,x}$. (P2) shows that there exists a neighbourhood V of (a,x) in U such that $\overline{\mathbb{J}}\subset I_{\overline{t},\overline{x}}$ for every $(\overline{t},\overline{x})\in V$. Since $\{(t_{n_k},\phi_{t_0,x_0}(t_{n_k}))\}$ converges to (a,x), the trajectory passing through (t_0,x_0) has some points in V. Hence $\overline{\mathbb{J}}\subset I_{t_0,x_0}$, and this gives the required contradiction. ∎

THEOREM 3. *Let U be a non-empty open subset of $\mathbb{R}\times X$, and let E be an equivalence relation in U satisfying (P1) and (P2).*

(i) If $X=\mathbb{R}$, then Φ is continuous.

(ii) If $X\neq\mathbb{R}$, then in general Φ is not continuous.

PROOF. (i) Let $(t_1,t_0,x_0)\in U^*$, and let $\mathbb{J}=[a,b]\subset I_{t_0,x_0}$ be a compact interval containing t_0 and t_1 as interior points. (P2) shows that there exists a neighbourhood V of (t_0,x_0) in U such that $\mathbb{J}\subset I_{t,x}$ for every $(t,x)\in V$. Hence $\mathbb{J}\times V$ is a neighbourhood of (t_1,t_0,x_0) in U^*. If we prove that $\Phi(s,t,x) \to \Phi(s,t_0,x_0)$ as $(t,x) \to (t_0,x_0)$ uniformly for s in $\mathbb{J}$, then Φ is continuous at (t_1,t_0,x_0), by the continuity of the trajectory passing through (t_0,x_0) and Lemma 1.

Let ε_0 be a positive number. Since $\{(t,\Phi(t,t_0,x_0)) : t \in \mathbb{J}\}$ is a compact set contained in U and since the complement of U is closed, there is a positive number $\varepsilon < \varepsilon_0$ such that

$$K_\varepsilon = \{(t,x) : t \in \mathbb{J} \quad \text{and} \quad |\Phi(t,t_0,x_0) - x| \leq \varepsilon\}$$

is a subset of U. We can suppose that $\{(t_0,x) : |x - x_0| \leq \varepsilon\} \subset V$. Since all the trajectories are continuous and since the trajectories are pairwise disjoint, it is enough to prove the existence of states x_1 and x_2 such that $x_0 - \varepsilon \leq x_1 < x_0 < x_2 \leq x_0 + \varepsilon$ and $\{(t,\Phi(t,t_0,x_i)) : t \in \mathbb{J}\} \subset K_\varepsilon$ for $i = 1,2$. We prove only the existence of x_2, the proof of the other case is similar.

Let Ψ be defined on $\mathbb{J}$ by $\Psi(t) = \Phi(t,t_0,x_0) + \varepsilon$. Suppose on the contrary that for every $x \in]x_0, x_0+\varepsilon[$ there exists $\bar{t}_x \in]a,b[$ such that $\Phi(\bar{t}_x,t_0,x) > > \Psi(\bar{t}_x)$. Then for each $x \in]x_0,x_0+\varepsilon[$ we can choose $t_x \in]a,b[$ such that $\Phi(t_x,t_0,x) = \Psi(t_x)$ and $\Phi(t,t_0,x) < \Psi(t)$ for every t between t_x and t_0. Since the trajectories are pairwise disjoint we can suppose that $t_x \in]t_0,b[$ for every $x \in]x_0,x_0+\varepsilon[$. If $x_0 < x < \hat{x} < x_0 + \varepsilon$, then $t_{\hat{x}} < t_x$. Hence $t_x \to \bar{t}$ as $x \to x_0$, where $\bar{t} \in]t_0,b]$. Let $\Phi(\bar{t},t_0,x_0) < \bar{x} < \Psi(\bar{t})$, and let $\bar{a}$ be the left-hand endpoint of $I_{\bar{t},\bar{x}}$. $t_0 < \bar{a} < \bar{t}$ by the choice of $\bar{t}$ and $\bar{x}$. Since all the trajectories are continuous and since the trajectories are pairwise disjoint, $\{(t,\Phi(t,\bar{t},\bar{x})) : \bar{a} < t < \bar{t}\} \subset K_\varepsilon$. This contradicts Lemma 2. (ii), since K_ε is compact.

(ii) We give a concrete example, when $X = \mathbb{R}^2$. Suppose $0 < p \leq 1$ and $q > 0$. Let $f_{p,q}$ be defined on $\mathbb{R}$ by

$$f_{p,q}(t) = \begin{cases} q, & \text{if } t \leq 0 \text{ or } t > 2p \\ q + \frac{1-p}{p^2} qt, & \text{if } 0 < t \leq p \\ \frac{q(2-p)}{p} + \frac{p-1}{p^2} qt, & \text{if } p < t \leq 2p \end{cases}$$

The graph of $f_{p,q}$ is in the half plane $S = \{(t,u) : u > 0\}$ for every possible p and q. Suppose $(t_0,u_0) \in S$, and let $0 < p \leq 1$ be fixed. It easy to see that there is exactly one $q > 0$ such that $f_{p,q}(t_0) = u_0$. Thus the graphs of the functions $f_{p,q}$ $(q > 0)$ induce a classification of S, which is called the classification of S according to p.

Let S_0 be the half plane $\{(t,u,0) : u > 0\}$. Let S_ω denote S_0 rotated about the t-axis through ω, where $\omega \in [0,2\pi[$. Consider the classification of S_ω according to $1 - \frac{\omega}{2\pi}$. The classes in S_ω ($\omega \in [0,2\pi[$) and the class $\{(t,0,0) : t \in \mathbb{R}\}$ induce a classification of $\mathbb{R}^3$. This classification of $\mathbb{R}^3$ generates an equivalence relation E in $\mathbb{R}^3$. The equivalence classes of E are functions from $\mathbb{R}$ into $\mathbb{R}^2$. It follows that E satisfies (P1) and (P2).

Let $x(\omega) = \left(\left(1-\frac{\omega}{2\pi}\right)\sin\omega, \left(1-\frac{\omega}{2\pi}\right)\cos\omega\right)$, where $\omega \in [0, 2\pi[$. Then

$$\left(1-\frac{\omega}{2\pi}, 0, x(\omega)\right) \in U^* \quad \text{and} \quad \left(1-\frac{\omega}{2\pi}, 0, x(\omega)\right) \to (0,0,(0,0))$$

as $\omega \to 2\pi - 0$. Since $\Phi\left(1-\frac{\omega}{2\pi}, 0, x(\omega)\right) = (\sin\omega, \cos\omega) \to (0,1)$ as $\omega \to 2\pi - 0$ and since $\Phi(0,0,(0,0)) = (0,0)$, Φ is not continuous at $(0,0,(0,0))$. ∎

Reference

[1] T.M. Flett, *Differential analysis*, Cambridge University Press, Cambridge, 1980.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 151–155

ARITHMETICAL FUNCTIONS WITH CONGRUENCE PROPERTIES

By

I. JOÓ and B. M. PHONG

Department of Analysis, L. Eötvös University, Budapest

(Received February 25, 1992)

1. Multiplicative functions with congruence properties

Let $\mathcal{M}$ denote the set of all integer-valued multiplicative functions. For a fixed positive integer k let N_k be the arithmetical function defined in [2] by $N_k(n) := m$ if $n = m^k r$, where r is k-free. It is obvious that N_k is a multiplicative function and $N_1(n) = n$ for all positive integers n.

We prove the following theorem:

THEOREM 1. *Let A, B and k be fixed positive integers with the condition $(A, B) = 1$. If a function $g \in \mathcal{M}$ and an integer $C \neq 0$ satisfy the congruence*

$$g(An+B) \equiv C \pmod{N_k(n)} \tag{1}$$

for every positive integer n, then there are a non-negative integer α and a real-valued Dirichlet character χ (mod A) such that

$$g(n) = \chi(n) n^{\alpha}$$

holds for all positive integers n which are prime to A.

COROLLARY. Let A and k be fixed positive integers. If the functions $g_1 \in \mathcal{M}$ and $g_2 \in \mathcal{M}$ satisfy the congruence

$$g_1(An+m) \equiv g_2(m) \pmod{N_k(n)}$$

for every positive integer n and m, then there are a non-negative integer α and a real-valued Dirichlet character χ (mod A) such that

$$g_1(n) = g_2(n) = \chi(n) n^{\alpha}$$

holds for all positive integers n which are prime to A.

We note that for $k = 1$ this theorem was proved in [5] and for $A = 1$ was obtained in [2] and it was stated in this form in [2] with indication of a different (complicated) proof.

Similarly as in [2] one can prove Lemma 1 and Lemma 2 below.

LEMMA 1. Assume that (1) holds for every positive integer n. Then

$$g(ab)=g(a)g(b) \tag{2}$$

for all positive integers a and b which are prime to A, furthermore $g(B)=C$.

LEMMA 2. Assume that (1) holds for every positive integer n. Then for each prime $Q\nmid A$ there is a non-negative integer $\alpha(Q)$ such that

$$|g(Q)|=Q^{\alpha(Q)}. \tag{3}$$

LEMMA 3. (D. R. HEATH-BROWN [1]) Let q, r and s be three non-zero integers which are multiplicatively independent, i.e. if $q^e r^f s^g=1$ with integers e, f and g then $e=f=g=0$. Suppose that none of q, r, s, $-3qr$, $-3qs$, $-3rs$ or qrs is a square. Then the number $N(x)$ of primes $\le x$ for which at least one of q, r and s is a primitive root satisfies $N(x)\gg x/(\log x)^2$.

LEMMA 4. Assume that (1) holds for every positive integer n. Then there is a non-negative integer α such that for $\alpha(Q)$ (given in (3)) $\alpha(Q)=\alpha$ holds for all primes $Q\nmid A$.

PROOF. Lemma 3 implies that there is a $g\in\mathbb{N}$ such that g is a primitive root (mod p) for infinitely many primes p. Let $p>A$ be a prime for which g is a primitive root (mod p). Then $(A,p)=1$ and we can find an integer t such that

$$P=pt+g \qquad \text{is prime,} \tag{4}$$

$$P\equiv 1 \pmod{A} \tag{5}$$

and

$$P^{p-1}\not\equiv 1 \left(\mathrm{mod}\ p^2\right). \tag{6}$$

Let

$$R_n(P):=\frac{P^n-1}{P-1} \qquad (n=1,2,3,\ldots).$$

For each positive integer $(m,P)=1$ let $r(m)$ denote the rank of apparation of m in the sequence $R_n(P)$.

We get easily that $r(p)=p-1$ and $r(p)\ne r(p^2)$, furthermore we have

$$r(p^h)=(p-1)p^{h-1} \qquad (h=1,2,\ldots). \tag{7}$$

Let $Q\ne p$ be an arbitrary prime for which $(Q,A)=1$ and

$$W_n:=Q^{\varphi(A)}P^n-1 \qquad (n=1,2,\ldots).$$

Using Theorem 4.1 in [4] there is a positive integer s_h for each positive integer h such that

(8) $$W_{s_h} \equiv 0 \quad \left(\bmod p^h\right).$$

By (1) we have

$$g\left[Q^{2\varphi(A)}P^{2s_{kh}}B\right] = g\left[BA\left(Q^{2\varphi(A)}P^{2s_{kh}}-1\right)/A+B\right] \equiv C$$
$$\left(\bmod N_k\left[\left(Q^{2\varphi(A)}P^{2s_{kh}}-1\right)/A\right]\right).$$

Thus (8) yields

(9) $$g\left[Q^{2\varphi(A)}P^{2s_{kh}}B\right] \equiv C \quad \left(\bmod p^h\right).$$

On the other hand by Lemma 1 and Lemma 2 we get

(10) $$g\left[Q^{2\varphi(A)}P^{2s_{kh}}B\right] = Q^{2\varphi(A)\alpha(Q)}P^{2s_{kh}\alpha(P)}g(B).$$

Therefore (9) and (10) imply

(11) $$Q^{2\varphi(A)\alpha(Q)}P^{2s_{kh}\alpha(P)}g(B) \equiv C \ \left(\bmod p^h\right)$$

for every positive integer h. Since by (8)

$$\left[Q^{2\varphi(A)}P^{2s_{kh}}\right]^{\alpha(P)} \equiv 1 \quad \left(\bmod p^h\right),$$

we get immediately from (11) that

$$Q^{2\varphi(A)\alpha(Q)}g(B) = Q^{2\varphi(A)\alpha(P)}C.$$

$g(B) = C$ yields $\alpha(Q) = \alpha(P)$ for all pair of primes (Q,P) where $Q \nmid A$ and P satisfies the conditions (4)–(6). Thus we have proved that $\alpha(Q) = \alpha$ (constant).

LEMMA 5. Let $G(n) := g(n)/n^\alpha$ for all positive integers n which are prime to A, where α is defined in Lemma 4. Then $G(n) = \chi(n)$, where χ denotes a real-valued Dirichlet character (mod A).

PROOF. The proof of Lemma 5 is similar to that of Theorem 6 in [3], hence we omit it.

PROOF OF THEOREM 1. The proof of Theorem 1 follows from Lemmas 4 and 5 easily.

PROOF OF THE COROLLARY. Assume that the functions $g_1 \in \mathcal{M}$ and $g_2 \in \mathcal{M}$ satisfy the congruence

(12) $$g_1(An+m) \equiv g_2(m) \quad (\bmod N_k(n))$$

for every positive integer n and m. The choice $m=1$ in (12) by Theorem 1 yields that there are a non-negative integer α and a real-valued Dirichlet character χ (mod A) such that $g_1(n)=\chi(n)n^{\alpha}$ holds for all positive integers n which are prime to A. Therefore (12) gives

$$\chi(An+m)(An+m)^{\alpha} \equiv \chi(m)m^{\alpha} \equiv g_2(m) \pmod{N_k(n)},$$

for all $(m,A)=1$. Thus we have $g_2(m)=\chi(m)\cdot m^{\alpha}$.

2. Additive functions with congruence properties

Let $\mathcal{A}$ denote the set of all real-valued additive functions. For a fixed positive integer A let $\mathcal{D}(A)$ denote the class of all arithmetical functions $D:\mathbb{N}\to\mathbb{N}$ for which the following conditions are satisfied:

(a) $D(n)|D(nm)$ for all positive integers n and m

(b) for each positive integer m we have

$$\limsup_{s\to\infty} \frac{D\{[(Am+1)^s-1]/A\}}{(s,D\{[(Am+1)^s-1]/A\})}=\infty.$$

For example the functions $n\to n$ and $n\to\varphi(n)$ are elements of $D(A)$ for any $A\in\mathbb{N}$.

We prove the following theorem:

THEOREM 2. *Let A, B be positive integers and let C be a real number. Assume that $f\in\mathcal{A}$ and $D\in\mathcal{D}(A)$ satisfy the congruence*

$$f(An+B)\equiv C \pmod{D(n)} \tag{13}$$

for all positive integers n. Then $f(n)=0$ holds for all positive integers n which are prime to A.

We note that this result improves Theorem 1 of [3].

LEMMA 6. Assume that the conditions of Theorem 2 are satisfied. Then $f(B)=C$ and

$$f(ab)=f(a)+f(b) \tag{14}$$

holds for all positive integers a and b with $(ab,A)=1$.

PROOF. Let a and b be positive integers with $(ab,A)=1$. From the condition b) of $\mathcal{D}(A)$, for each $k>0$ there exists an s such that

$$D(m)>k \tag{15}$$

with $$m:=\frac{(ab)^{\varphi(A)s}-1}{A}$$

Since $(m,ab)=1$ we can choose positive integers x, y, u and v such that

$$ax=1+Amy, \qquad (x,abB)=1 \tag{16}$$

and

$$bu=B+Amv, \qquad (u,abx)=1. \tag{17}$$

By using (13), (16), (17) and the condition a) of $\mathcal{D}(A)$ we have

$$f(aB)+f(x)=f(axB)=f(ABmy+B)\equiv C \pmod{D(m)},$$
$$f(b)+f(u)=f(bu)=f(Amv+B)\equiv C \pmod{D(m)}$$

and

$$f(ab)+f(x)+f(u)=f(axbu)=f(AmT+B)\equiv C \pmod{D(m)},$$

where $T:=Amyv+By+v$. Therefore

$$f(ab)-f(aB)-f(b)+C\equiv 0 \pmod{D(m)},$$

Then (15) implies

$$f(ab)-f(aB)-f(b)+C=0 \tag{18}$$

for all $(ab,A)=1$. Applying (18) with $a=b=1$ we have $f(B)=C$. Replacing $b=1$ in (18) we have $f(aB)=f(a)+C$. Thus (18) yields $f(ab)=f(a)+f(b)$.

Proof of Theorem 2. By Lemma 1 we get $f(ab)=f(a)+f(b)$ for all $(a,b,A)=1$. Therefore $n=Bm$ in (13) implies

$$f(Am+1)\equiv 0 \pmod{D(m)}. \tag{19}$$

Thus we have

$$sf(Am+1)=f\left[(Am+1)^s\right]\equiv 0 \quad \left(\bmod\, D\left[\frac{(Am+1)^s-1}{A}\right]\right).$$

This yields $f(Am+1)=0$ from the condition b) of $D(A)$. It is easy to verify that $f(n)=0$ for all $(n,A)=1$.

References

[1] D. R. Heath-Brown, Artins conjecture for primitive roots, *Q. J. Math. Oxford* II. Ser., **37** (1986), 27–38.

[2] I. Joó, Note on multiplicative functions satisfying a congruence property II, *Acta Math. Hungar.*, (to appear).

[3] I. Joó, Arithmetical functions satisfying a congruence property, *Acta Math. Hungar.*, (to appear).

[4] P. Kiss and B. M. Phong, Divisibility properties in the second order recurrences, *Publ. Math. Debrecen*, **26** (1979), 187–197.

[5] B. M. Phong, Multiplicative functions satisfying a congruence property V, *Acta Math. Hungar.*, (to appear).

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 157–188

ON THE FEJÉR SUMMABILITY OF EIGENFUNCTION EXPANSIONS

By

S. SZABÓ and M. B. TAHIR

Department of Analysis, L. Eötvös University, Budapest

(Received May 5, 1992)

The aim of the present paper is to prove general equiconvergence theorems for Fejér means. We consider the Schrödinger operator with any complex potential function $q: G \to \mathbb{C}$ on any (finite or infinite) interval G, with arbitrary (complex) eigenvalues λ_n. These investigations are started in JOÓ's and SÖVEGJÁRTÓ's recent work [1], and in JOÓ's papers ([2], [12]) where $\lambda_n \geq 0$ and the Hermite–Fourier expansions are considered.

1. Let G be an arbitrary (finite or infinite) open interval on the real line, q, $\hat{q} \in L^1_{\mathrm{loc}}(G)$ arbitrary complex functions. Let (u_k) (resp. $(\hat{u}_k)$) be a Riesz-basis in $L^2(G)$ consisting of eigenfunctions of the operator $Lu = -u'' + qu$ (resp. $\hat{L}u = -u'' + \hat{q}u$) and having the following properties:

(1) $\quad \sup o_k < \infty, \qquad \sup \hat{o}_k < \infty$

(2) $\quad$ in case $o_k > 0 \qquad$ (resp. $\hat{o}_k > 0$)

$$\lambda_k u_k - L u_k = u_{k-1} \qquad (\text{resp. } \hat{\lambda}_k \hat{u}_k - \hat{L}\hat{u}_k = \hat{u}_{k-1}),$$

where λ_k and o_k (resp. $\hat{\lambda}_k$ and $\hat{o}_k$) are the eigenvalue and the order of u_k (resp. $\hat{u}_k$).

Now let us introduce some notations:

$$(3) \quad R_\mu(f,x) := \sum_{|\mathrm{Re}\sqrt{\lambda_k}|<2\mu} \langle f, v_k\rangle u_k(x)\left(1 - \frac{\mu_k}{2\mu}\right), \qquad (\mu_k = \sqrt{\lambda_k}),$$

$$\hat{R}_\mu(f,x) := \sum_{|\mathrm{Re}\sqrt{\hat{\lambda}_k}|<2\mu} \langle f, \hat{v}_k\rangle \hat{u}_k(x)\left(1 - \frac{\hat{\mu}_k}{2\mu}\right), \qquad (\hat{\mu}_k = \sqrt{\hat{\lambda}_k});$$

($f \in L^2(G)$, $x \in G$, $\mu > 0$), where (v_k) (resp. $(\hat{v}_k)$) is the dual system of (u_k) (resp $(\hat{u}_k)$), i.e. (v_k), $(\hat{v}_k) \subset L^2(G)$ and $\langle v_k, u_j\rangle = \langle \hat{v}_k, \hat{u}_j\rangle = \delta_{k,j}$.

The following result holds:

THEOREM 1: *Given any compact interval $K \subset G$, for all $f \in L^2(G)$*

$$\lim_{\mu\to\infty} \sup_{x\in K} |R_\mu(f,x) - \hat{R}_\mu(f,x)| = 0.$$

For $f \in L^2(G)$, $\mu > 0$ and $x \pm R \in G$, define

$$F_\mu(f,x) = F_\mu(f,x,R) := \frac{1}{\mu\pi} \int_{x-R}^{x+R} \left(\frac{\sin\mu(y-x)}{y-x}\right)^2 f(y)\,\mathrm{d}y. \tag{4}$$

The theorem will follow obviously from the following assertion:

PROPOSITION 1: Given any compact interval $K \subset G$, for any sufficiently small $R > 0$, and for all $f \in L^2(G)$, we have

$$\lim_{\mu\to\infty} \sup_{x\in K} |F_\mu(f,x) - R_\mu(f,x)| = 0.$$

Indeed, an analogous result holds for $\hat{R}_\mu(f,x)$, too, and it remains only to apply the triangle inequality. For the sake of brevity, from now on we shall denote by μ_k a square root of λ_k with $\operatorname{Re}\mu_k > 0$ and we set $\varrho_k := \operatorname{Re}\mu_k$, $\nu_k := \operatorname{Im}\mu_k$.

REMARK 1: If we modify the definition of R_μ,

$$R^*_\mu(f,x) := \sum_{|\operatorname{Re}\sqrt{\lambda_k}|<2\mu} \langle f, v_k\rangle\, u_k(x) \left(1 - \frac{\varrho_k}{2\mu}\right),$$

the Proposition 1 and Theorem 1 remain true.

First we prove the Remark 1. Denote

$$S_\mu(f,x) := \frac{1}{\pi} \int_{x-R}^{x+R} \frac{\sin\mu(y-x)}{y-x} f(y)\,\mathrm{d}y,$$

$$\sigma_\mu(f,x) := \sum_{\varrho_k<\mu} \langle f, v_k\rangle\, u_k(x).$$

Then we have

$$R^*_\mu(f,x) = \frac{1}{2\mu} \int_0^{2\mu} \sigma_t(f,x)\,\mathrm{d}t,$$

$$F_\mu(f,x) = \frac{1}{2\mu} \int_0^{2\mu} S_t(f,x)\,\mathrm{d}t.$$

We know ([3], (43)) that

$$\lim_{\mu\to\infty}\sup_{x\in K}|S_\mu(f,x)-\sigma_\mu(f,x)|=0, \qquad \text{for all } f\in L^2(G).$$

Since

$$|F_\mu(f,x)-R^*_\mu(f,x)|\le\frac{1}{2\mu}\int_0^{2\mu}|S_t(f,x)-\sigma_t(f,x)|\,\mathrm{d}t$$

therefore we obtain the statement of Remark 1 at once.

Now we prove the Proposition 1. For the proof we need a lemma.

LEMMA 1. ([3], Proposition 1, p.359). Given any compact interval $K\subset G$, there exists an $R>0$ with

$$\sup_{\mu>0}\sum_{|\mu-|\operatorname{Re}\sqrt{\lambda_k}||\le 1}\left(\|u_k\|_{L^\infty(K)}\operatorname{ch}\left(R\operatorname{Im}\sqrt{\lambda_k}\right)\right)^2<\infty.$$

Let us prove the Proposition 1.

We know ([3], (66)), that

$$\sup_{\mu>0}\sup_{x\in K}|S_\mu(f,x)-\sigma_\mu(f,x)|\le D\|f\|_{L^2(G)}$$

for all $f\in L^2(G)$, where $D>0$ is a constant, $K\subset G$ is an arbitrary compact interval. Therefore

$$\sup_{\mu>0}\sup_{x\in K}|F_\mu(f,x)-R^*_\mu(f,x)|\le D\|f\|_{L^2(G)}. \tag{5}$$

Fixing $x\in K$ and $\mu>0$ arbitrary, we define $w:G\to\mathbb{R}$ by

$$w(x+t)=\begin{cases}\frac{1}{\mu\pi}\left(\frac{\sin\mu t}{t}\right)^2, & \text{if } |t|\le R,\\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

Since (u_k) is a Riesz-basis and (v_k) is the dual system of (u_k), we have

$$F_\mu(f,x)=\sum_k\langle f,v_k\rangle\langle u_k,w\rangle. \tag{7}$$

Denote

$$\delta^*(\mu,\varrho_k)=\begin{cases}1-\frac{\varrho_k}{2\mu}, & \varrho_k<2\mu,\\ 0 & ,\varrho_k\ge 2\mu.\end{cases} \tag{8}$$

Then we obtain from (5)

$$\sup_{\mu>0}\sup_{x\in K}\left|\left\langle f,\sum_k\left(\overline{\langle u_k,w\rangle}-\delta^*(\mu,\varrho_k)\overline{u_k(x)}\right)v_k\right\rangle\right|\le D\|f\|_{L^2(G)}.$$

Here taking sup in f we get

$$\sup_{\mu>0}\sup_{x\in K}\left\|\sum_k\left(\overline{\langle u_k,w\rangle}-\delta^*(\mu,\varrho_k)\overline{u_k(x)}v_k\right)\right\|_{L^2(G)}\leq D.$$

Since (v_k) is a Riesz-basis, therefore

$$\sup_{\mu>0}\sup_{x\in K}\sum_k|\langle u_k,w\rangle-\delta^*(\mu,\varrho_k)u_k(x)|^2\leq D_1. \tag{9}$$

Denote

$$\delta(\mu,\varrho_k)=\begin{cases}1-\frac{\mu_k}{2\mu}, & \varrho_k<2\mu,\\ 0, & \varrho_k\geq 2\mu.\end{cases} \tag{10}$$

Then we have

$$\sum_k|\langle u_k,w\rangle-\delta(\mu,\varrho_k)u_k(x)|^2\leq$$

$$\leq c\sum_k|\langle u_k,w\rangle-\delta^*(\mu,\varrho_k)u_k(x)|^2+c\sum_k|(\delta^*(\mu,\varrho_k)-\delta(\mu,\varrho_k))u_k(x)|^2\leq$$

$$\leq c+c\sum_{\varrho_k<2\mu}\left|\frac{\nu_k}{\mu}u_k(x)\right|^2$$

where we have used (9). But

$$\sum_{\varrho_k<2\mu}\left|\frac{\nu_k}{\mu}u_k(x)\right|^2\leq\frac{c}{\mu^2}\sum_{\varrho_k<2\mu}\left(\|u_k\|_{L^\infty(K)}e^{|\nu_k|R}\right)^2\leq$$

$$\leq\frac{c}{\mu^2}\sum_{i=1}^{2\mu}\sum_{2\mu-i\leq\varrho_k\leq 2\mu-i+1}\left(\|u_k\|_{L^\infty(K)}e^{|\nu_k|R}\right)^2\leq\frac{c}{\mu^2}\sum_{i=1}^{2\mu}1\leq\frac{c}{\mu}$$

where we used Lemma 1.

Hence

$$\sum_k|\langle u_k,w\rangle-\delta(\mu,\varrho_k)u_k(x)|^2\leq c. \tag{11}$$

Since (u_k) is a Riesz-basis there exists a constant c_0 such that for all $f\in L^2(G)$

$$\sum_k|\langle f,u_k\rangle|^2\leq c_0\|f\|^2_{L^2(K)}. \tag{12}$$

Taking into account (7), (12) and applying the Cauchy–Schwarz inequality, (11) and (12) give

$$\sup_{\mu>0}\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|\leq c\|f\|_{L^2(G)}\quad(\text{for any } f\in L^2(G)).$$

Now it suffices to show that

$$\lim_{\mu\to\infty}\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|=0$$

for any f from a dense subset of $L^2(G)$. But this last property is satisfied for any finite linear combination f of the eigenfunctions u_k because then f is continuously differentiable and $R_\mu(f,x)-f(x)=O\left(\frac{1}{\mu}\right)$ for μ sufficiently large, therefore one can apply a classical result of the theory of Fourier-series. Hence Proposition 1 and Theorem 1 is proved.

2. In the next part of this paper we investigate the case $f\in L^1(G)$.

Let G be an arbitrary finite open interval in the real line. Let (u_k) be a Bessel-system i.e. for any $f\in L^2(G)$

$$\sum_k |\langle u_k,f\rangle|^2\le c_0\|f\|^2_{L^2(G)}$$

and assume (1), (2). Furthermore assume that (v_k) consists of the eigenfunctions of the operator $L^*v:=-v''+\bar{q}v$ with eigenvalues $\bar{\lambda}_k$ and (v_k) is Bessel-system.

THEOREM 2: *If q, $\hat{q}\in L^p(G)$, $p>1$, then for any compact interval $K\subset G$ and for all $f\in L^1(G)$ we have*

$$\lim_{\mu\to\infty}\sup_{x\in K}|R_\mu(f,x)-\hat{R}_\mu(f,x)|=0.$$

If q, $\hat{q}\in L^1(G)$, then for all $f\in L^1(G)$ we have

$$\lim_{\mu\to\infty}|R_\mu(f,x)-\hat{R}_\mu(f,x)|=0 \quad \text{a.e.} \quad x\in G.$$

PROPOSITION 2: If $q\in L^p(G)$, $p>1$, then for any compact interval $K\subset G$, for any sufficiently small $R>0$, and for all $f\in L^1(G)$, we have

$$\lim_{\mu\to\infty}\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|=0.$$

If $q\in L^1(G)$ then for any sufficiently small $R>0$ and for all $f\in L^1(G)$, we have

$$\lim_{\mu\to\infty}|F_\mu(f,x)-R_\mu(f,x)|=0 \quad \text{a.e.} \quad x\in G.$$

REMARK 2: The Proposition 2 and Theorem 2 remain true for R^*_μ also.

Introduce the operator S_{R_0}

$$f \to S_{R_0}[f] := \frac{2}{R_0} \int\limits_{\frac{R_0}{2}}^{R_0} f(R)\,\mathrm{d}R.$$

First we prove the Remark 2.

We consider only the case $q \in L^p(G)$, $p > 1$, because the case $p = 1$ is similar. We know ([6], (3.8)) that

$$\lim_{\mu\to\infty} \sup_{x\in K} |S_\mu(f,x) - \sigma_\mu(f,x)| = 0, \quad \text{for all } f \in L^1(G).$$

Since

$$|F_\mu(f,x) - R^*_\mu(f,x)| \le \frac{1}{2\mu} \int\limits_0^{2\mu} |S_t(f,x) - \sigma_t(f,x)|\,\mathrm{d}t$$

therefore we obtain the statement of Remark 2 at once.

Now we prove the Proposition 2. For the proof we need a lemma.

LEMMA 2. ([6], Lemma 3.5) We have for any Bessel-system (u_k) of eigenfunctions of order $\le m < \infty$ of the operator $Lu = -u'' + qu$ ($q \in L^1(G)$, $|G| < \infty$) with eigenvalues $\{\lambda_k\} \subset \mathbb{C}$

$$\sup_{\mu>0} \sum_{|\mu-\varrho_k|\le 1} \|u_k\|^2_{L^\infty(G)} < \infty.$$

Let us prove the Proposition 2. We know ([6], (3.26)) that

$$\sup_{\mu>0} \sup_{x\in K} \sum_{k=1}^{\infty} \left|\left\langle u_k, S_{R_0}[w_0]\right\rangle - \delta_0(\mu,\varrho_k) u_k(x)\right| \cdot \|v_k\|_{L^\infty(G)} < \infty$$

where $K \subset G$ is an arbitrary compact interval and

$$w_0 = w_{R,\mu}(x+t) = \begin{cases} \frac{1}{\pi} \cdot \frac{\sin \mu t}{t}, & \text{if } |t| \le R, \\ 0, & \text{otherwise.} \end{cases}$$

$$\delta_0(\mu,\varrho_k) = \begin{cases} 1, & \mu > \varrho_k, \\ \frac{1}{2}, & \mu = \varrho_k, \\ 0, & \mu < \varrho_k. \end{cases}$$

Obviously, we have

$$\delta^*(\mu,\varrho_k) = \frac{1}{2\mu} \int\limits_0^{2\mu} \delta_0(t,\varrho_k)\,\mathrm{d}t, \qquad w = \frac{1}{2\mu} \int\limits_0^{2\mu} w_{R,t}\,\mathrm{d}t.$$

Therefore we get

$$\left|\left\langle u_k, S_{R_0}[w]\right\rangle - \delta^*(\mu,\varrho_k)u_k(x)\right| = \left|S_{R_0}\left[\langle u_k, w\rangle - \delta^*(\mu,\varrho_k)u_k(x)\right]\right| =$$

$$= \left|\frac{1}{2\mu}\int_0^{2\mu} S_{R_0}\left[\left\langle u_k, w_{R,t}\right\rangle - \delta_0(t,\varrho_k)u_k(x)\right]\,\mathrm{d}t\right| =$$

$$= \left|\frac{1}{2\mu}\int_0^{2\mu} \left(\left\langle u_k, S_{R_0}[w_0]\right\rangle - \delta_0(t,\varrho_k)u_k(x)\right)\,\mathrm{d}t\right|.$$

Using this we obtain

$$\sup_{\mu>0}\sup_{x\in K}\sum_{k=1}^{\infty}\left|\left\langle u_k, S_{R_0}[w]\right\rangle - \delta^*(\mu,\varrho_k)u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)} \le \tag{13}$$

$$\le \sup_{\mu>0}\sup_{x\in K}\sum_{k=1}^{\infty}\frac{1}{2\mu}\int_0^{2\mu}\left|\left\langle u_k, S_{R_0}[w_0]\right\rangle - \delta_0(t,\varrho_k)u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)}\,\mathrm{d}t < \infty.$$

We have

$$\sum_{k=1}^{\infty}\left|\left\langle u_k, S_{R_0}[w]\right\rangle - \delta(\mu,\varrho_k)u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)} \le$$

$$\le \sum_{k=1}^{\infty}\left|\left\langle u_k, S_{R_0}[w]\right\rangle - \delta^*(\mu,\varrho_k)u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)} +$$

$$+ \sum_{k=1}^{\infty}\left|(\delta^*(\mu,\varrho_k) - \delta(\mu,\varrho_k))u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)} \le$$

$$\le c + c\sum_{\varrho_k<2\mu}\left|\frac{\nu_k}{\mu}u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)}$$

where we have used (13). But

$$\sum_{\varrho_k<2\mu}\left|\frac{\nu_k}{2\mu}u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)} \le \frac{c}{\mu}\sum_{\varrho_k<2\mu}\left(\|u_k\|_{L^\infty(K)}e^{|\nu_k|R}\right)\cdot\|v_k\|_{L^\infty(G)} \le$$

$$\le \frac{c}{\mu}\sum_{i=1}^{2\mu}\sum_{2\mu-i\le\varrho_k\le 2\mu-i+1}\left(\|u_k\|_{L^\infty(K)}e^{|\nu_k|R}\right)\cdot\|v_k\|_{L^\infty(G)} \le$$

$$\leq \frac{c}{\mu}\sum_{i=1}^{2\mu}\left(\sum_{2\mu-i\leq\varrho_k\leq 2\mu-i+1}\left(\|u_k\|_{L^\infty(K)}e^{|\nu_k|R}\right)^2\right)^{\frac{1}{2}}\cdot$$

$$\cdot\left(\sum_{2\mu-i\leq\varrho_k\leq 2\mu-i+1}\|v_k\|^2_{L^\infty(G)}\right)^{\frac{1}{2}}\leq\frac{c}{\mu}\sum_{i=1}^{2\mu}1<\infty,$$

where we used Lemma 1 and Lemma 2.

Hence

$$\sup_{\mu>0}\sup_{x\in K}\sum_{k=1}^{\infty}\left|\left\langle u_k, S_{R_0}[w]\right\rangle-\delta(\mu,\varrho_k)u_k(x)\right|\cdot\|v_k\|_{L^\infty(G)}<\infty. \tag{14}$$

Denote $w_{R_0}:=S_{R_0}[w]$. Because of (14) the series

$$\sum_{k=1}^{\infty}\left(\left\langle u_k, w_{R_0}\right\rangle-\delta(\mu,\varrho_k)u_k(x)\right)\bar{v}_k(t)$$

converges in $L^\infty(G)$ for any fixed $x\in K$ and is equal to

$$\Theta_\mu(x,t):=\bar{w}_{R_0}(t)-\sum_{\varrho_k<2\mu}u_k(x)\bar{v}_k(t)\left(1-\frac{\mu_k}{2\mu}\right).$$

Consequently, by (14)

$$\sup_{\mu>0}\sup_{x\in K}\|\Theta_\mu(x,\cdot)\|_{L^\infty(G)}<\infty. \tag{15}$$

On the other hand it is easy to see that

$$S_{R_0}[w]=\begin{cases} w(x+t), & \text{if } |t|\leq\frac{R_0}{2},\\ \frac{2(R_0-t)}{R_0}w(x+t), & \text{if } \frac{R_0}{2}<|t|\leq R_0,\\ 0, & \text{if } |t|>R_0.\end{cases} \tag{16}$$

Therefore

$$\sup_{\mu>0}\sup_{x\in K}\|\bar{w}_{R_0}-w\|_{L^\infty(G)}<\infty. \tag{17}$$

From (15), (17) we obtain

$$M:=\sup_{\mu>0}\sup_{x\in K}\left\|w-\sum_{\varrho_k<2\mu}u_k(x)\bar{v}_k\left(1-\frac{\mu_k}{2\mu}\right)\right\|_{L^\infty(G)}<\infty,$$

hence

$$\sup_{\mu>0}\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|\le M\|f\|_{L^1(G)}.$$

On the other hand the linear hull of $\{u_k\}$ form a dense set in $L^1(G)$. Consequently

$$\lim_{\mu\to\infty}\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|=0.$$

Hence Proposition 2 and Theorem 2 is proved.

3. Now we investigate a special case.

Denote $G=(0,+\infty)$,

$$u_k:=\sqrt{2}x^{\alpha+\frac{1}{2}}e^{-\frac{x^2}{2}}l_k^{(\alpha)}(x^2),\quad q(x):=-x^2-2\alpha-2+\frac{\alpha^2-\frac{1}{4}}{x^2},\quad \lambda_k:=4k,$$

where $\alpha\ge-\frac{1}{2}$, $l_k^{(\alpha)}(x)$ is the normed Laguerre polynomial. Then we have $-u_k''+qu_k=\lambda_k u_k$. We consider three cases:

a) $f\in L^2(G)$, b) $f\in L^1(G)$, c) $f\in L^1(G)$, $f'(t)(1+t^2)\in L^1(G)$, $\lim\limits_{+\infty} f=0$.

a) In this case $q\in L^1_{\text{loc}}(G)$, hence the conditions are satisfied and the statements Proposition 1 and Remark 1 remain true.

b) In this case $q\notin L^p(G)$, but the statements of Proposition 2 and Remark 2 remain true. For the proof we need some Lemmas.

LEMMA 3: For any $R_0>0$ there exists $C(R_0)$ such that

$$\left|S_{R_0}\left[\frac{2}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\cos\mu_k t\,dt\right]-\delta(\mu,\varrho_k)\right|\le$$

$$\le\frac{c}{\mu}\left(\frac{1}{1+(2\mu-\varrho_k)^2}+\frac{1}{1+\varrho_k^2}\right)\operatorname{ch}\nu_k R_0,$$

for any $\mu>0$ and k.

PROOF. We can write

$$\frac{2}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\cos\mu_k t\,\mathrm{d}t=I_1+I_2+I_3+I_4,$$

$$I_1:=\frac{2}{\mu\pi}\int_0^\infty\left(\frac{\sin\mu t}{t}\right)^2\cos\varrho_k t\,\mathrm{d}t,$$

$$I_2:=\frac{2}{\mu\pi}\int_R^\infty\left(\frac{\sin\mu t}{t}\right)^2\cos\varrho_k t\,\mathrm{d}t,$$

$$I_3:=\frac{2}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\cos\varrho_k t(\operatorname{ch}\nu_k t-1)\,\mathrm{d}t,$$

$$I_4:=\frac{-2i}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\sin\varrho_k t\operatorname{sh}\nu_k t\,\mathrm{d}t.$$

It is well-known

(18) $$I_1=\begin{cases}1-\frac{\varrho_k}{2\mu}, & \varrho_k<2\mu,\\ 0, & \varrho_k\ge 2\mu.\end{cases}$$

Now estimate I_2.

Obviously

$$I_2=\frac{\varrho_k}{\mu\pi}\int_{\varrho_k R}^\infty\frac{\cos t}{t^2}\,\mathrm{d}t-\frac{|2\mu-\varrho_k|}{2\mu\pi}\int_{|2\mu-\varrho_k|R}^\infty\frac{\cos t}{t^2}\,\mathrm{d}t-\frac{2\mu+\varrho_k}{2\mu\pi}\int_{(2\mu+\varrho_k)R}^\infty\frac{\cos t}{t^2}\,\mathrm{d}t,$$

where we used $\sin^2\alpha\cos\gamma=\frac14(-\cos(2\alpha-\gamma)+2\cos\gamma-\cos(2\alpha+\gamma))$.

Here

$$\int_x^\infty\frac{\cos t}{t^2}\,\mathrm{d}t=\begin{cases}O\left(\frac1x\right), & 0<x\le c,\\ -\frac{\sin x}{x^2}+O\left(\frac{1}{x^3}\right), & x>c.\end{cases}$$

If $\varrho_k<1$ then

(19) $$S_{R_0}\left[\varrho_k\int_{\varrho_k R}^\infty\frac{\cos t}{t^2}\,\mathrm{d}t\right]=O(1).$$

If $\varrho_k \geq 1$ then

$$S_{R_0}\left[\varrho_k \int\limits_{\varrho_k R}^{\infty} \frac{\cos t}{t^2}\,\mathrm{d}t\right] = -\frac{1}{\varrho_k} S_{R_0}\left[\frac{\sin \varrho_k R}{R^2}\right] + O\left(\frac{1}{\varrho_k^2}\right).$$

Here

$$S_{R_0}\left[\frac{\sin \varrho_k R}{R^2}\right] = \left[\frac{-\cos \varrho_k R}{\varrho_k} \cdot \frac{1}{R^2}\right]_{\frac{R_0}{2}}^{R_0} - 2 \int\limits_{\frac{R_0}{2}}^{R_0} \frac{\cos \varrho_k R}{\varrho_k R^3}\,\mathrm{d}R = O\left(\frac{1}{\varrho_k}\right),$$

therefore

$$S_{R_0}\left[\varrho_k \int\limits_{\varrho_k R}^{\infty} \frac{\cos t}{t^2}\,\mathrm{d}t\right] = O\left(\frac{1}{\varrho_k^2}\right). \tag{20}$$

From (19) and (20) we obtain

$$S_{R_0}\left[\varrho_k \int\limits_{\varrho_k R}^{\infty} \frac{\cos t}{t^2}\,\mathrm{d}t\right] = O(1)\frac{1}{1+\varrho_k^2}. \tag{21}$$

Similarly

$$S_{R_0}\left[|2\mu - \varrho_k| \int\limits_{|2\mu-\varrho_k|R}^{\infty} \frac{\cos t}{t^2}\,\mathrm{d}t\right] = O(1)\frac{1}{1+(2\mu-\varrho_k)^2} \tag{22}$$

and

$$S_{R_0}\left[(2\mu + \varrho_k) \int\limits_{(2\mu+\varrho_k)R}^{\infty} \frac{\cos t}{t^2}\,\mathrm{d}t\right] = O(1)\frac{1}{1+(2\mu+\varrho_k)^2}. \tag{23}$$

From (21), (22), (23) we get

$$S_{R_0}[I_2] = O(1)\frac{1}{\mu}\left(\frac{1}{1+\varrho_k^2} + \frac{1}{1+(2\mu-\varrho_k)^2}\right). \tag{24}$$

We have

$$I_3 = \frac{1}{2\mu\pi} \int\limits_0^R (-\cos(2\mu-\varrho_k)t + 2\cos\varrho_k t - \cos(2\mu+\varrho_k)t)\,\frac{\operatorname{ch}\nu_k t - 1}{t^2}\,\mathrm{d}t.$$

If $|2\mu-\varrho_k|<1$ or $\varrho_k<1$ then

$$S_{R_0}\left[\int\limits_0^R \frac{\operatorname{ch}\nu_k t-1}{t^2}\,\mathrm{d}t\right]=O(1)\operatorname{ch}\nu_k R_0. \tag{25}$$

If $|2\mu-\varrho_k|\geq 1$ then twofold integration by parts gives

$$\int\limits_0^R \cos(2\mu-\varrho_k)t\cdot\frac{\operatorname{ch}\nu_k t-1}{t^2}\,\mathrm{d}t=$$

$$=\frac{\sin(2\mu-\varrho_k)R}{2\mu-\varrho_k}\cdot\frac{\operatorname{ch}\nu_k R-1}{R^2}+\frac{1}{(2\mu-\varrho_k)^2}\cdot\left[\cos(2\mu-\varrho_k)t\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)'\right]_0^R-$$

$$-\frac{1}{(2\mu-\varrho_k)^2}\int\limits_0^R \cos(2\mu-\varrho_k)t\cdot\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)''\mathrm{d}t.$$

Here

$$\left|\int\limits_0^R \cos(2\mu-\varrho_k)t\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)''\mathrm{d}t\right|\leq\left[\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)'\right]_0^R$$

because $\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)''$ doesn't change sign.

$$\left[\left(\frac{\operatorname{ch}\nu_k t-1}{t^2}\right)'\right]_0^R=\sum_{n=2}^{\infty}\frac{(2n-2)\nu_k^{2n}R^{2n-3}}{(2n)!}=$$

$$=\frac{\nu_k^2}{R}\sum_{n=2}^{\infty}\frac{\nu_k^{2n-2}R^{2n-2}}{(2n-3)!(2n-1)2n}\leq\frac{\nu_k^2}{R}\sum_{n=2}^{\infty}\frac{\nu_k^{2n-2}R^{2n-2}}{(2n-2)!}=\frac{\nu_k^2}{R}\operatorname{ch}\nu_k R.$$

Hence

$$S_{R_0}\left[\int\limits_0^R \cos(2\mu-\varrho_k)t\cdot\frac{\operatorname{ch}\nu_k t-1}{t^2}\,\mathrm{d}t\right]=$$

$$=\frac{1}{2\mu-\varrho_k}S_{R_0}\left[\sin(2\mu-\varrho_k)R\frac{\operatorname{ch}\nu_k R-1}{R^2}\right]+O(1)\frac{1}{(2\mu-\varrho_k)^2}\nu_k^2\operatorname{ch}\nu_k R_0.$$

Here

$$S_{R_0}\left[\sin(2\mu-\varrho_k)R\frac{\operatorname{ch}\nu_k R-1}{R^2}\right]=\left[-\frac{\cos(2\mu-\varrho_k)R}{2\mu-\varrho_k}\cdot\frac{\operatorname{ch}\nu_k R-1}{R^2}\right]_{\frac{R_0}{2}}^{R_0}+$$

$$+\frac{1}{2\mu-\varrho_k}\int\limits_{\frac{R_0}{2}}^{R_0}\cos(2\mu-\varrho_k)R\left(\frac{\operatorname{ch}\nu_k R-1}{R^2}\right)'\mathrm{d}R=O\left(\frac{1}{2\mu-\varrho_k}\right)\operatorname{ch}\nu_k R_0.$$

Therefore

(26)

$$S_{R_0}\left[\int\limits_0^R\cos(2\mu-\varrho_k)t\frac{\operatorname{ch}\nu_k t-1}{t^2}\,\mathrm{d}t\right]=O\left(\frac{1}{(2\mu-\varrho_k)^2}\right)(1+\nu_k^2)\operatorname{ch}\nu_k R_0.$$

From (25) and (26) we obtain

(27)

$$S_{R_0}\left[\int\limits_0^R\cos(2\mu-\varrho_k)t\frac{\operatorname{ch}\nu_k t-1}{t^2}\,\mathrm{d}t\right]=O(1)\frac{1}{1+(2\mu-\varrho_k)^2}(1+\nu_k^2)\operatorname{ch}\nu_k R_0.$$

Similarly

(28)

$$S_{R_0}\left[\int\limits_0^R\cos(2\mu+\varrho_k)t\frac{\operatorname{ch}\nu_k-1}{t^2}\,\mathrm{d}t\right]=O(1)\frac{1}{1+(2\mu+\varrho_k)^2}(1+\nu_k^2)\operatorname{ch}\nu_k R_0$$

and

$$\text{(29)}\qquad S_{R_0}\left[\int\limits_0^R\cos\varrho_k t\frac{\operatorname{ch}\nu_k-1}{t^2}\,\mathrm{d}t\right]=O(1)\frac{1}{1+\varrho_k^2}(1+\nu_k^2)\operatorname{ch}\nu_k R_0.$$

From (27), (28), (29) we get

$$\text{(30)}\qquad S_{R_0}[I_3]=O(1)\frac{1}{\mu}\left(\frac{1}{1+\varrho_k^2}+\frac{1}{1+(2\mu-\varrho_k)^2}\right)(1+\nu_k^2)\operatorname{ch}\nu_k R_0.$$

We have

$$I_4=\frac{i}{2\mu\pi}\int\limits_0^R(\sin(\varrho_k-2\mu)t+\sin(\varrho_k+2\mu)t-2\sin\varrho_k t)\frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t.$$

First investigate the case $\varrho_k>2\mu$. We have

$$\int\limits_0^R\sin(\varrho_k-2\mu)t\frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t=$$

$$=\nu_k\left(\int_0^R \frac{\sin(\varrho_k-2\mu)t}{t}\,\mathrm{d}t+\sum_{l=1}^{\infty}\frac{\nu_k^{2l}}{(2l+1)!}\int_0^R \sin(\varrho_k-2\mu)t\cdot t^{2l-1}\,\mathrm{d}t\right).$$

If $\varrho_k-2\mu<1$ then

$$\int_0^R \frac{\sin(\varrho_k-2\mu)t}{t}\,\mathrm{d}t=O(1)(\varrho_k-2\mu)=O(1)\frac{1}{1+(\varrho_k-2\mu)^2}.$$

If $\varrho_k-2\mu\geq 1$ then

$$\int_0^R \frac{\sin(\varrho_k-2\mu)t}{t}\,\mathrm{d}t=\frac{\pi}{2}-\frac{\cos(\varrho_k-2\mu)R}{(\varrho_k-2\mu)R}+\int_{(\varrho_k-2\mu)R}^{\infty}\frac{\cos s}{s^2}\,\mathrm{d}s=$$

$$=\frac{\pi}{2}-\frac{\cos(\varrho_k-2\mu)R}{(\varrho_k-2\mu)R}+O(1)\frac{1}{1+(\varrho_k-2\mu)^2}.$$

Since

$$S_{R_0}\left[\frac{\cos(\varrho_k-2\mu)R}{R}\right]=O\left(\frac{1}{\varrho_k-2\mu}\right),$$

therefore if $\varrho_k>2\mu$ then

$$S_{R_0}\left[\int_0^R \frac{\sin(\varrho_k-2\mu)t}{t}\,\mathrm{d}t\right]=\frac{\pi}{2}+O(1)\frac{1}{1+(\varrho_k-2\mu)^2}. \tag{31}$$

Similarly, if $\varrho_k-2\mu<1$ then

$$\int_0^R \sin(\varrho_k-2\mu)t\cdot t^{2l-1}\,\mathrm{d}t=O(1)\frac{R^{2l+1}}{1+(\varrho_k-2\mu)^2}.$$

If $\varrho_k-2\mu\geq 1$ then

$$\int_0^R \sin(\varrho_k-2\mu)t\cdot t^{2l-1}\,\mathrm{d}t=-\frac{\cos(\varrho_k-2\mu)R}{\varrho_k-2\mu}R^{2l-1}+\frac{(2l-1)}{\varrho_k-2\mu}\cdot$$

$$\cdot\int_0^R \cos(\varrho_k-2\mu)t\cdot t^{2l-2}\,\mathrm{d}t=-\frac{\cos(\varrho_k-2\mu)R}{\varrho_k-2\mu}R^{2l-1}+O(1)\frac{(2l-1)R^{2l-1}}{1+(\varrho_k-2\mu)^2}.$$

Since

$$S_{R_0}[\cos(\varrho_k-2\mu)R\cdot R^{2l-1}]=O(1)\frac{R_0^{2l+1}}{\varrho_k-2\mu},$$

Therefore if $\varrho_k - 2\mu \geq 1$ then

$$S_{R_0}\left[\int_0^R \sin(\varrho_k - 2\mu)t \cdot t^{2l-1}\,\mathrm{d}t\right] = O(1)\frac{R_0^{2l+1}}{1+(\varrho_k-2\mu)^2}.$$

Hence if $\varrho_k > 2\mu$ then

$$S_{R_0}\left[\int_0^R \sin(\varrho_k - 2\mu)t \cdot t^{2l-1}\,\mathrm{d}t\right] = O(1)\frac{R_0^{2l+1}}{1+(\varrho_k-2\mu)^2}. \tag{32}$$

Using (31) and (32) we obtain

$$S_{R_0}\left[\int_0^R \sin(\varrho_k - 2\mu)t \cdot \frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t\right] = \nu_k\frac{\pi}{2} + O(1)\frac{\operatorname{sh}|\nu_k|R_0}{1+(\varrho_k-2\mu)^2}. \tag{33}$$

Similar calculation gives that for $\varrho_k > 0$

$$S_{R_0}\left[\int_0^R \sin\varrho_k t \cdot \frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t\right] = \nu_k\frac{\pi}{2} + O(1)\frac{\operatorname{sh}|\nu_k|R_0}{1+\varrho_k^2}. \tag{34}$$

Finally

$$S_{R_0}\left[\int_0^R \sin(\varrho_k + 2\mu)t \frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t\right] = \nu_k\frac{\pi}{2} + O(1)\frac{\operatorname{sh}|\nu_k|R_0}{(\varrho_k+2\mu)^2}. \tag{35}$$

From (33)–(35) we have if $\varrho_k \geq 2\mu$ then

$$S_{R_0}[I_4] = O(1)\frac{1}{\mu}\left(\frac{1}{1+(\varrho_k-2\mu)^2} + \frac{1}{1+\varrho_k^2}\right)\operatorname{sh}|\nu_k|R_0. \tag{36}$$

Now investigate the case $\varrho_k < 2\mu$. Then

$$\int_0^R \frac{\sin(\varrho_k-2\mu)t}{t}\,\mathrm{d}t = -\int_0^R \frac{\sin|\varrho_k-2\mu|t}{t}\,\mathrm{d}t,$$

hence

$$S_{R_0}\left[\int_0^R \sin(\varrho_k - 2\mu)t \frac{\operatorname{sh}\nu_k t}{t^2}\,\mathrm{d}t\right] = -\nu_k\frac{\pi}{2} + O(1)\frac{\operatorname{sh}|\nu_k|R_0}{1+(\varrho_k-2\mu)^2}. \tag{37}$$

Thus from (34), (35), (37) we have if $\varrho_k < 2\mu$ then

$$(38)\qquad S_{R_0}[I_4] = \frac{-i\nu_k}{2\mu} + O(1)\frac{1}{\mu}\left(\frac{1}{1+(\varrho_k-2\mu)^2} + \frac{1}{1+\varrho_k^2}\right)\operatorname{sh}|\nu_k|R_0.$$

Using (36), (38) we obtain
(39)

$$S_{R_0}[I_4] = \begin{cases} O(1)\frac{1}{\mu}\left(\frac{1}{1+(\varrho_k-2\mu)^2} + \frac{1}{1+\varrho_k^2}\right)\operatorname{sh}|\nu_k|R_0, & \text{if } \varrho_k \geq 2\mu \\ -\frac{i\nu_k}{2\mu} + O(1)\frac{1}{\mu}\left(\frac{1}{1+(\varrho_k-2\mu)^2} + \frac{1}{1+\varrho_k^2}\right)\operatorname{sh}|\nu_k|R_0, & \text{if } \varrho_k < 2\mu. \end{cases}$$

From (18), (24), (30), (39) the Lemma 3 follows.

Lemma 4. Denote

$$\gamma_k^* := \int\limits_{|x-\xi|}^{R} \left(\frac{\sin\mu t}{t}\right)^2 \sin\mu_k(t-|x-\xi|)\,\mathrm{d}t,$$

where $x \in K$, $0 < \frac{R_0}{2} < R < R_0$, $0 < |x-\xi| < R$, $R_0 < \operatorname{dist}(K, \partial G)$. Then we have

$$\gamma_k^* = O(1)\left(\frac{1}{\mu} + \frac{1+\varrho_k^{1-\varepsilon}}{|x-\xi|^\varepsilon}\right)\operatorname{ch}4\nu_k R, \qquad (0 \leq \varrho_k < 2\mu),$$

$$\gamma_k^* = O(1)\frac{\mu^{2-\varepsilon}}{|\mu_k|\cdot|x-\xi|^\varepsilon}\operatorname{ch}\nu_k R, \qquad (0 < 2\mu \leq \varrho_k).$$

Proof. First investigate the case $0 \leq \varrho_k < 2\mu$. We will use several times the following relations

$$\sin(x \pm iy) = \sin x \operatorname{ch} y \pm i \operatorname{sh} y \cos x,$$
$$\cos(x \pm iy) = \cos x \operatorname{ch} y \mp i \sin y \operatorname{sh} y,$$
$$|\sin\alpha| \leq |\sin\alpha|^{1-\varepsilon} \leq |\alpha|^{1-\varepsilon}, \qquad (\alpha \in \mathbb{R},\ 1 \geq \varepsilon \geq 0),$$
$$\left|\frac{\sin z}{z}\right| \leq 2\operatorname{ch}(\operatorname{Im} z), \qquad (z \in \mathbb{C}).$$

Integrating by parts we get

$$\gamma_k^* = \int\limits_{|x-\xi|}^{R} \sin\mu t\left(\frac{\sin\mu t \sin\mu_k(t-|x-\xi|)}{t^2}\right)\mathrm{d}t =$$

$$= -\int\limits_{|x-\xi|}^{R}\left(\int\limits_0^R \sin\mu\tau\,\mathrm{d}\tau\right)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\sin\mu t \sin\mu_k(t-|x-\xi|)}{t^2}\right)\mathrm{d}t =$$

$$=\frac{\cos\mu R}{\mu}\int\limits_{|x-\xi|}^{R}\frac{\mathrm{d}}{\mathrm{d}t}(\cdot)\,\mathrm{d}t+\frac{1}{\mu}\int\limits_{|x-\xi|}^{R}\cos\mu t\cdot\frac{\mathrm{d}}{\mathrm{d}t}(\cdot)\,\mathrm{d}t=$$

$$=\frac{\sin 2\mu R\sin\mu_k(R-|x-\xi|)}{2\mu R^2}+\int\limits_{|x-\xi|}^{R}\frac{\cos^2\mu t\sin\mu_k(t-|x-\xi|)}{t^2}\,\mathrm{d}t+$$

$$+\frac{\mu_k}{2\mu}\int\limits_{|x-\xi|}^{R}\frac{\sin 2\mu t\cos\mu_k(t-|x-\xi|)}{t^2}\,\mathrm{d}t-\frac{1}{\mu}\int\limits_{|x-\xi|}^{R}\frac{\sin 2\mu t\sin\mu_k(t-|x-\xi|)}{t^3}\,\mathrm{d}t.$$

Since $\cos^2\mu t=1-\sin^2\mu t$ therefore we obtain

(40)
$$2\gamma_k^*=2\int\limits_{|x-\xi|}^{R}\left(\frac{\sin\mu t}{t}\right)^2\sin\mu_k(t-|x-\xi|)\,\mathrm{d}t=\frac{\sin 2\mu R\sin\mu_k(R-|x-\xi|)}{2\mu R^2}+$$

$$+\int\limits_{|x-\xi|}^{R}\frac{\sin\mu_k(t-|x-\xi|)}{t^2}\,\mathrm{d}t+\frac{\mu_k}{2\mu}\int\limits_{|x-\xi|}^{R}\frac{\sin 2\mu t\cos\mu_k(t-|x-\xi|)}{t^2}\,\mathrm{d}t-$$

$$-\frac{1}{\mu}\int\limits_{|x-\xi|}^{R}\frac{\sin 2\mu t\sin\mu_k(t-|x-\xi|)}{t^3}\,\mathrm{d}t.$$

First we estimate

$$I_1:=\int\limits_{|x-\xi|}^{R}\frac{\sin\mu_k(t-|x-\xi|)}{t^2}\,\mathrm{d}t.$$

Obviously

$$I_1=\cos\mu_k|x-\xi|\int\limits_{|x-\xi|}^{R}\frac{\sin\mu_k t}{t^2}\,\mathrm{d}t-\sin\mu_k|x-\xi|\int\limits_{|x-\xi|}^{R}\frac{\cos\mu_k t}{t^2}\,\mathrm{d}t.$$

Here

$$\int\limits_{|x-\xi|}^{R}\frac{\sin\mu_k t}{t^2}\,\mathrm{d}t=\int\limits_{|x-\xi|}^{R}\frac{\sin\varrho_k t\,\mathrm{ch}\,\nu_k t}{t^2}\,\mathrm{d}t+i\int\limits_{|x-\xi|}^{R}\frac{\mathrm{sh}\,\nu_k t\cos\varrho_k t}{t^2}\,\mathrm{d}t=:I_{11}+iI_{12}$$

and

$$\int\limits_{|x-\xi|}^{R} \frac{\cos \mu_k t}{t^2} \,\mathrm{d}t = \int\limits_{|x-\xi|}^{R} \frac{\cos \varrho_k t \operatorname{ch} \nu_k t}{t^2} \,\mathrm{d}t - i \int\limits_{|x-\xi|}^{R} \frac{\sin \varrho_k t \operatorname{sh} \nu_k t}{t^2} \,\mathrm{d}t =: I_{13} - iI_{14}.$$

$$I_{11} = \int\limits_{|x-\xi|}^{R} \frac{\sin \varrho_k t \operatorname{ch} \nu_k t}{t^2} \,\mathrm{d}t = O(1) \operatorname{ch} \nu_k R \int\limits_{|x-\xi|}^{R} \frac{\varrho_k^{1-\varepsilon} t^{1-\varepsilon}}{t^2} \,\mathrm{d}t = O(1) \frac{\varrho_k^{1-\varepsilon} \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}.$$

$$I_{12} = \int\limits_{|x-\xi|}^{R} \frac{\operatorname{sh} \nu_k \cos \varrho_k t}{t^2} \,\mathrm{d}t = \nu_k \int\limits_{|x-\xi|}^{R} \frac{\operatorname{sh} \nu_k t}{\nu_k t} \cdot \frac{\cos \varrho_k t}{t} \,\mathrm{d}t =$$

$$= O(1) |\nu_k| \operatorname{ch} \nu_k R \int\limits_{|x-\xi|}^{R} \frac{1}{t} \,\mathrm{d}t = O(1) \frac{|\nu_k| \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}},$$

(where the implicit constant depends on ε).

$$I_{13} = \int\limits_{|x-\xi|}^{R} \frac{\cos \varrho_k t \operatorname{ch} \nu_k t}{t^2} \,\mathrm{d}t = O(1) \frac{1}{|x-\xi|}.$$

$$I_{14} = \int\limits_{|x-\xi|}^{R} \frac{\sin \varrho_k t \operatorname{sh} \nu_k t}{t^2} \,\mathrm{d}t = O(1) \operatorname{ch} \nu_k R \int\limits_{|x-\xi|}^{R} \frac{\varrho_k^{1-\varepsilon} t^{1-\varepsilon}}{t^2} \,\mathrm{d}t = O(1) \frac{\varrho_k^{1-\varepsilon} \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}.$$

Hence

$$I_1 = O(1) \frac{\varrho_k^{1-\varepsilon} \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}} + O(1) \frac{|\nu_k| \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}} +$$
$$+ O(1) \left(\frac{\sin \varrho_k |x-\xi| \operatorname{ch} \nu_k |x-\xi| + i \operatorname{sh} \nu_k |x-\xi| \cos \varrho_k |x-\xi|}{|x-\xi|} \right) +$$
$$+ O(1) \frac{\varrho_k^{1-\varepsilon} \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}} = O(1) \frac{(\varrho_k^{1-\varepsilon} + |\nu_k|) \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}},$$

i.e.

(41) $$I_1 = O(1) \frac{(1 + \varrho_k^{1-\varepsilon}) \operatorname{ch} 3\nu_k R}{|x-\xi|^{\varepsilon}}.$$

Now we estimate

$$I_2 := \int_{|x-\xi|}^{R} \frac{\sin 2\mu t \cos \mu_k (t-|x-\xi|)}{t^2}\,\mathrm{d}t.$$

Obviously

$$I_2 = \cos \mu_k |x-\xi| \int_{|x-\xi|}^{R} \frac{\sin 2\mu t \cos \mu_k t}{t^2}\,\mathrm{d}t + \sin \mu_k |x-\xi| \int_{|x-\xi|}^{R} \frac{\sin 2\mu t \sin \mu_k t}{t^2}\,\mathrm{d}t.$$

Here

$$\int_{|x-\xi|}^{R} \frac{\sin 2\mu t \sin \mu_k t}{t^2}\,\mathrm{d}t =$$

$$= \int_{|x-\xi|}^{R} \sin 2\mu t \frac{\sin \varrho_k t \operatorname{ch} \nu_k t}{t^2}\,\mathrm{d}t + i \int_{|x-\xi|}^{R} \sin 2\mu t \frac{\operatorname{sh} \nu_k t \cos \varrho_k t}{t^2}\,\mathrm{d}t =: I_{21} + i I_{22}$$

and

$$\int_{|x-\xi|}^{R} \frac{\sin 2\mu t \cos \mu_k t}{t^2}\,\mathrm{d}t =$$

$$= \int_{|x-\xi|}^{R} \frac{\sin 2\mu t}{t^2} \cos \varrho_k t \operatorname{ch} \nu_k t\,\mathrm{d}t - i \int_{|x-\xi|}^{R} \sin 2\mu t \frac{\sin \varrho_k t \operatorname{sh} \nu_k t}{t^2}\,\mathrm{d}t =: I_{23} - i I_{24}.$$

Using the estimates of I_{11}, I_{12} and I_{14} we obtain

$$I_{21} = O(1) \frac{\varrho_k^{1-\varepsilon} \cdot \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}, \qquad I_{22} = O(1) \frac{|\nu_k| \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}},$$

$$I_{23} = O(1) \frac{\mu^{1-\varepsilon} \cdot \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}, \qquad I_{24} = O(1) \frac{\varrho_k^{1-\varepsilon} \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}.$$

Hence

$$I_2 = O(1) \frac{\mu^{1-\varepsilon} \operatorname{ch} 3\nu_k R}{|x-\xi|^{\varepsilon}}. \tag{42}$$

Investigate now

$$I_3 := \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t \sin \mu_k (t - |x-\xi|)}{t^3} \mathrm{d}t.$$

Obviously

$$I_3 = \cos \mu_k |x-\xi| \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t \sin \mu_k t}{t^3} \mathrm{d}t - \sin \mu_k |x-\xi| \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t \cos \mu_k t}{t^3} \mathrm{d}t.$$

Here

$$\int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t \sin \mu_k t}{t^3} \mathrm{d}t =$$

$$= \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t}{t} \cdot \frac{\sin \varrho_k t \operatorname{ch} \nu_k t}{t^2} \mathrm{d}t + i \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t}{t} \cdot \frac{\operatorname{sh} \nu_k t \cos \varrho_k t}{t^2} \mathrm{d}t =: I_{31} + i I_{32}$$

and

$$\int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t \cos \mu_k t}{t^3} \mathrm{d}t =$$

$$= \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t}{t^3} \cos \varrho_k t \operatorname{ch} \nu_k t \,\mathrm{d}t - i \int\limits_{|x-\xi|}^{R} \frac{\sin 2\mu t}{t} \cdot \frac{\sin \varrho_k t \operatorname{sh} \nu_k t}{t^2} \mathrm{d}t =: I_{33} - i I_{34}.$$

Using the estimates of I_{21}, I_{22} and I_{24} we obtain

$$I_{31} = O(1) \frac{\mu \cdot \varrho_k^{1-\varepsilon} \cdot \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}, \qquad I_{32} = O(1) \frac{\mu |\nu_k| \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}},$$

$$I_{33} = O(1) \frac{\mu}{|x-\xi|}, \qquad I_{34} = O(1) \frac{\mu \cdot \varrho_k^{1-\varepsilon} \cdot \operatorname{ch} \nu_k R}{|x-\xi|^{\varepsilon}}.$$

Hence

$$I_3 = O(1) \frac{\mu \cdot \varrho_k^{1-\varepsilon} \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}} + O(1) \frac{\mu \cdot |\nu_k| \operatorname{ch} 2\nu_k R}{|x-\xi|^{\varepsilon}} +$$

$$+ O(1) \mu \left(\frac{\sin \varrho_k |x-\xi| \operatorname{ch} \nu_k |x-\xi| + i \operatorname{sh} \nu_k |x-\xi| \cos \varrho_k |x-\xi|}{|x-\xi|} \right) +$$

$$+O(1)\frac{\mu\cdot\varrho_k^{1-\varepsilon}\cdot\operatorname{ch}2\nu_k R}{|x-\xi|^{\varepsilon}}=O(1)\frac{\mu(\varrho_k^{1-\varepsilon}+|\nu_k|)\operatorname{ch}2\nu_k R}{|x-\xi|^{\varepsilon}},$$

i.e.

$$I_3=O(1)\frac{\mu(1+\varrho_k^{1-\varepsilon})\operatorname{ch}3\nu_k R}{|x-\xi|^{\varepsilon}}. \tag{43}$$

From (40)–(43) we obtain

$$\gamma_k^*=O(1)\left(\frac{1}{\mu}+\frac{1+\varrho_k^{1-\varepsilon}}{|x-\xi|^{\varepsilon}}\right)\operatorname{ch}4\nu_k R,\qquad (0\le\varrho_k<2\mu). \tag{44}$$

Now investigate the case $0<2\mu\le\varrho_k$. Integrating by parts we get

$$\gamma_k^*=\left[\frac{\sin^2\mu t}{t^2}\int\limits_{|x-\xi|}^{t}\sin\mu_k(\tau-|x-\xi|)\,\mathrm{d}\tau\right]_{t=|x-\xi|}^{R}-$$

$$-\int\limits_{|x-\xi|}^{R}\left(\int\limits_{|x-\xi|}^{t}\sin\mu_k(\tau-|x-\xi|)\,\mathrm{d}\tau\right)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\sin^2\mu t}{t^2}\right)\mathrm{d}t=$$

$$=\frac{\sin^2\mu|x-\xi|}{\mu_k|x-\xi|^2}-\frac{\sin^2\mu R\cos\mu_k(R-|x-\xi|)}{\mu_k R^2}+$$

$$+\frac{1}{\mu_k}\int\limits_{|x-\xi|}^{R}\cos\mu_k(t-|x-\xi|)\cdot\left(\mu\frac{\sin2\mu t}{t^2}-\frac{2\sin^2\mu t}{t^3}\right)\mathrm{d}t.$$

Using the trivial estimates we obtain

$$\gamma_k^*=O(1)\frac{\mu^{2-\varepsilon}}{|\mu_k|\cdot|x-\xi|^{\varepsilon}}\cdot\operatorname{ch}\nu_k R,\qquad (0<2\mu\le\varrho_k). \tag{45}$$

The Lemma 4 is proved. ∎

LEMMA 5 ([8], Lemma 1): If $\alpha\ge-\frac{1}{2}$ then

$$\sup_{x>0}\sum_{a\le k\le b}u_k^2(x)\le c\sqrt{b-a}.$$

Let us prove the Proposition 2. First we prove that (analogously to (14))

$$\sup_{\mu>0}\sup_{x\in K}\sum_{k=1}^{\infty}\left|\left\langle u_k,w_{R_0}\right\rangle-\delta(\mu,\varrho_k)u_k(x)\right|\cdot|u_k(y)|<\infty, \tag{46}$$

uniformly in y on G.

We know the Titchmarsh-formula

(47)
$$\frac{u_k(x+t)+u_k(x-t)}{2}=u_k(x)\cos\mu_k t+\frac{1}{2\mu_k}\int_{x-t}^{x+t} q(\xi)u_k(\xi)\cdot\sin\mu_k(t-|x-\xi|)\,\mathrm{d}\xi.$$

Using (47) we obtain

$$\sum_{k=1}^{\infty}\left|\left\langle u_k,w_{R_0}\right\rangle-\delta(\mu,\varrho_k)u_k(x)\right|\cdot|u_k(y)|\le$$

(48)
$$\le\sum_{k=1}^{\infty}\frac{1}{\mu\mu_k\pi}\left|S_{R_0}\left[\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right]\right|\cdot|u_k(y)|+$$
$$+\sum_{k=1}^{\infty}\left|S_{R_0}\left[\frac{2}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\cos\mu_k t\,\mathrm{d}t-\delta(\mu,\varrho_k)\right]\right|\cdot\|u_k\|_{L^\infty(K)}\cdot|u_k(y)|.$$

Using Lemma 3, Lemma 1 and Lemma 5 we obtain

(49)
$$\sum_{k=1}^{\infty}\left|S_{R_0}\left[\frac{2}{\mu\pi}\int_0^R\left(\frac{\sin\mu t}{t}\right)^2\cos\mu_k t\,\mathrm{d}t-\delta(\mu,\varrho_k)\right]\right|\cdot$$
$$\cdot\|u_k\|_{L^\infty(K)}\cdot|u_k(y)|=O(1)\frac{1}{\mu^{3/4}}.$$

Using Lemma 4, Lemma 1 and Lemma 5 we obtain

(50)
$$\sum_{k=1}^{\infty}\frac{1}{\mu\mu_k\pi}\left|S_{R_0}\left[\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right]\right|\cdot|u_k(y)|=O(1)\mu^{\frac{1}{4}-\varepsilon},$$

where $0<\varepsilon<1$ is an arbitrary number. Hence from (48)–(50) follows (46).

The proof of the Proposition 2 can be finished as in **2.**

c) We prove the following statement

THEOREM 3: *If $f\in L^1(G)$, $f'(t)(1+t^2)\in L^1(G)$, $\lim_{+\infty} f=0$, then for any compact interval $K\subset G$ and for any sufficiently small $R>0$ we have*

$$\sup_{x\in K}|F_\mu(f,x)-R_\mu(f,x)|=O\left(\frac{1}{\mu}\right).$$

For the proof we need a Lemma.

LEMMA 6: If $\alpha>-1$ then

$$\sum_{a\le k<b}\left(\int_{x_1}^{x_2}u_k(x)\,\mathrm{d}x\right)^2\le c\frac{\sqrt{b-a}}{a}(x_2^4+1),\quad (0\le x_1\le x_2\le\infty).$$

PROOF. By the Mehler-formula

$$\sum_{n=0}^{\infty}\ell_n^{(\alpha)}(x)\ell_n^{(\alpha)}(y)z^n=\frac{\exp\left\{-(x+y)\frac{z}{1-z}\right\}}{1-z}\cdot\frac{J_\alpha\left(2i\frac{\sqrt{xyz}}{1-z}\right)}{(i\sqrt{xyz})^\alpha},$$

where $|z|<1$ and J_α is the Bessel-function. From this

$$\ell_k^{(\alpha)}(x)\ell_k^{(\alpha)}(y)=\frac{1}{2\pi i}\int_\Gamma\frac{\exp\left\{-(x+y)\frac{z}{1-z}\right\}}{1-z}\cdot\frac{J_\alpha\left(2i\frac{\sqrt{xyz}}{1-z}\right)}{(i\sqrt{xyz})}\cdot\frac{1}{z^{k+1}}\,\mathrm{d}z,$$

where $\Gamma=S(0,r)$, $r<1$. Taken $r\to 1-0$, i.e. $z=e^{i\varphi}$, we obtain

$$\ell_k^{(\alpha)}(x)\ell_k^{(\alpha)}(y)=\frac{1}{2\pi}\int_0^{2\pi}\frac{\exp\left\{-(x+y)\frac{e^{i\varphi}}{1-e^{i\varphi}}\right\}}{1-e^{i\varphi}}\cdot\frac{J_\alpha\left(2i\sqrt{xy}\frac{e^{i\frac{\varphi}{2}}}{1-e^{i\varphi}}\right)}{(i\sqrt{xy})^\alpha e^{i\varphi\frac{\alpha}{2}}}\cdot\frac{1}{e^{ik\varphi}}\,\mathrm{d}\varphi.$$

From this

$$x^\beta\cdot x^{\alpha+\frac12}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\cdot y^\beta\cdot y^{\alpha+\frac12}e^{-\frac{y^2}{2}}\ell_k^{(\alpha)}(y^2)=$$

$$=\frac{1}{2\pi i^\alpha}\int_0^{2\pi}\frac{\exp\left\{-i\frac{x^2+y^2}{2}\operatorname{ctg}\frac{\varphi}{2}\right\}}{1-e^{i\varphi}}\cdot x^{\beta+\frac12}\cdot y^{\beta+\frac12}J_\alpha\left(-\frac{xy}{\sin\frac{\varphi}{2}}\right)e^{-i(k+\frac{\alpha}{2})\varphi}\,\mathrm{d}\varphi.$$

Hence

$$\sum_{a\le k<b}\left(\int_d^t x^\beta u_k(x)\,\mathrm{d}x\right)^2=\frac{1}{4\pi i^{\alpha-1}}\int_0^{2\pi}\int_d^t\int_d^t\frac{\exp\left\{-i\frac{x^2+y^2}{2}\operatorname{ctg}\frac{\varphi}{2}\right\}}{\sin\frac{\varphi}{2}}\cdot \tag{51}$$

$$\cdot x^{\beta+\frac12}\cdot y^{\beta+\frac12}\cdot J_\alpha\left(-\frac{xy}{\sin\frac{\varphi}{2}}\right)e^{-i\frac{\alpha}{2}\varphi-i\frac{b+a}{2}\varphi}\frac{\sin\frac{b-a}{2}\varphi}{\sin\frac{\varphi}{2}}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\varphi,$$

where $d>0$, $t>0$. In what follows we use many times the formulas of Szegő's book [9]. We have

$$\left[x^{\alpha+\frac12}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\right]'=$$

$$\left(\alpha+\frac{1}{2}\right)x^{\alpha-\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}x^2-x^{\alpha+\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)-2\sqrt{k}x^{\alpha+\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_{k-1}^{(\alpha+1)}(x^2).$$

From this we get

$$(52)\quad \int_d^t x^{\alpha+\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_{k-1}^{(\alpha+1)}(x^2)\,\mathrm{d}x=\left(\alpha+\frac{1}{2}\right)\frac{1}{2\sqrt{k}}\int_d^t x^{\alpha-\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\,\mathrm{d}x-$$

$$-\frac{1}{2\sqrt{k}}\int_d^t x^{\alpha+\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\,\mathrm{d}x-\frac{1}{2\sqrt{k}}\left[x^{\alpha+\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\right]_d^t.$$

Investigate the case $t\geq d$, where $d>0$ is an arbitrary fixed number. Then we have from (52)

$$(53)\qquad \sum_{a\leq k<b}\left(\int_d^t x^{\alpha+\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_k^{(\alpha)}(x^2)\,\mathrm{d}x\right)^2\leq$$

$$\leq c\frac{\left(\alpha-\frac{1}{2}\right)^2}{a}\sum_{a\leq k<b}\left(\int_d^t x^{\alpha-\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\,\mathrm{d}x\right)^2+$$

$$+\frac{c}{a}\sum_{a\leq k<b}\left(\int_d^t x^{\alpha+\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\,\mathrm{d}x\right)^2+$$

$$+\frac{c}{a}\sum_{a\leq k<b}\left(\left[x^{\alpha-\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\right]_d^t\right)^2.$$

If $x>d_1>0$ where d_1 is an arbitrary fixed number then the result of Lemma 1 can be extended for $\alpha>-1$ (see: [10]), further following the calculation in [10] we see that it can be extended for every $\alpha\in\mathbb{R}$. The definition of $L_k^{(\alpha)}$ for $\alpha\leq-1$ (see: [9], p.111). The formula (51) also holds for every $\alpha\in\mathbb{R}$. Hence

$$(54)\qquad \sum_{a\leq k<b}\left(\left[x^{\alpha-\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\right]_d^t\right)^2\leq c\sqrt{b-a}.$$

Applying (51) with $\beta=-1$ and $\alpha-1$ we obtain

$$\sum_{a\leq k<b}\left(\int_d^t x^{\alpha-\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\,\mathrm{d}x\right)^2\leq$$

$$\leq c\int_0^{2\pi}\int_d^t\int_d^t \frac{1}{\left|\sin\frac{\varphi}{2}\right|} x^{-\frac{1}{2}}\cdot y^{-\frac{1}{2}}\cdot\left|J_{\alpha-1}\left(-\frac{xy}{\sin\frac{\varphi}{2}}\right)\right|\cdot\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\left|\sin\frac{\varphi}{2}\right|}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\varphi.$$

Since $\frac{xy}{\sin\frac{\varphi}{2}}\geq c>0$ therefore using the estimate of $J_{\alpha-1}$ ([11], p.168, (6)) we get

$$\text{(55)}\qquad \sum_{a\leq k<b}\left(\int_d^t x^{\alpha-\frac{3}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\,\mathrm{d}x\right)^2\leq$$

$$\leq c\int_0^{2\pi}\int_d^t\int_d^t\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\cdot\frac{1}{x}\cdot\frac{1}{y}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\varphi\leq c\sqrt{b-a}(\log t)^2.$$

Similarly, with $\beta=1$ and $\alpha-1$, we have

$$\text{(56)}\qquad \sum_{a\leq k<b}\left(\int_d^t x^{\alpha+\frac{1}{2}}e^{-\frac{x^2}{2}}\ell_{k+1}^{(\alpha-1)}(x^2)\,\mathrm{d}x\right)^2$$

$$\leq c\int_0^{2\pi}\int_d^t\int_d^t\frac{\left|\sin\frac{b-a}{2}\varphi\right|}{\varphi^{3/2}}\cdot xy\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\varphi\leq c\sqrt{b-a}t^4.$$

From (53)–(56) we obtain

$$\text{(57)}\qquad \sum_{a\leq k<b}\left(\int_d^t u_k(x)\,\mathrm{d}x\right)^2\leq c\frac{\sqrt{b-a}}{a}t^4.$$

Now investigate the case $0<t\leq d$. We estimate directly using the asymptotic formulas for $L_k^{(\alpha)}(x^2)$. We distinguish two cases: a) $\frac{c}{\sqrt{k}}\leq t\leq d$, b) $0<t\leq\frac{c}{\sqrt{k}}$.

a) It is known ([9], Theorem 8.22.4)

$$\text{(58)}\qquad e^{-\frac{x}{2}}x^{\frac{\alpha}{2}}L_k^{(\alpha)}(x)=N^{-\frac{\alpha}{2}}\frac{\Gamma(k+\alpha+1)}{k!}J_\alpha\left\{2(Nx)^{\frac{1}{2}}\right\}+x^{\frac{5}{4}}O\left(k^{\frac{\alpha}{2}-\frac{3}{4}}\right),$$

where $N=k+\frac{\alpha+1}{2}$, $\frac{c}{k}\leq x\leq w$, w is an arbitrary constant. From (58) we get

$$\text{(59)}\qquad e^{-\frac{x^2}{2}}x^{\alpha+\frac{1}{2}}L_k^{(\alpha)}(x^2)=N^{-\frac{\alpha}{2}}\frac{(k+\alpha+1)}{k!}xJ_\alpha\{2\sqrt{N}x\}+x^3O\left(k^{\frac{\alpha}{2}-\frac{3}{4}}\right).$$

Therefore the error term of $\sum\limits_{a\le k<b}\left(\int\limits_t^d u_k(x)\,\mathrm{d}x\right)^2$ is

$$(60)\quad \sum_{a\le k<b}\left(k^{-\frac{\alpha}{2}}\cdot k^{\frac{\alpha}{2}-\frac{3}{4}}\cdot\int\limits_t^d x^3\,\mathrm{d}x\right)^2\le c\sum_{a\le k<b}k^{-\frac{3}{2}}\le c\frac{\sqrt{b}-\sqrt{a}}{\sqrt{ab}}\le c\frac{\sqrt{b-a}}{a}.$$

It is known ([11], p.168, (6))

$$J_\nu(z)=\sqrt{\frac{2}{\pi z}}\left\{\sum_{j=0}^{n-1}\frac{A_j}{z^j}\cos\left[z-\frac{\pi}{2}\left(\nu-k+\frac{1}{2}\right)\right]+O\left(\frac{e^{|\operatorname{Im} z|}}{z^n}\right)\right\},$$

where $A_0=1$, $A_j=\prod\limits_{l=1}^{j}\left[\frac{4\nu^2-(2l-1)^2}{8l}\right]$, $z\neq 0$, $|\arg z|<\pi$. Using this we have

$$\int\limits_t^d\sqrt{x}J_\alpha\{2\sqrt{N}x\}\,\mathrm{d}x=\frac{N^{-\frac{1}{4}}}{\sqrt{\pi}}\int\limits_t^d\left(\cos\left[2\sqrt{N}x-\frac{\pi}{2}\left(\alpha+\frac{1}{2}\right)\right]+\right.$$
$$\left.+\frac{A_1}{2\sqrt{N}x}\cos\left[2\sqrt{N}x-\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right]+O(1)\frac{1}{Nx^2}\right)\mathrm{d}x.$$

Integrating by part

$$\int\limits_t^d\frac{1}{x}\cos\left[2\sqrt{N}x-\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right]\mathrm{d}x=\frac{\sin\left[2\sqrt{N}d-\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right]}{2d\sqrt{N}}-$$
$$-\frac{\sin\left[2\sqrt{N}t-\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right]}{2t\sqrt{N}}+\frac{1}{2\sqrt{N}}\int\limits_t^d\frac{\sin\left[2\sqrt{N}x-\frac{\pi}{2}\left(\alpha-\frac{1}{2}\right)\right]}{x^2}\mathrm{d}x=O(1)\frac{1}{t\sqrt{N}}$$

Hence

$$(61)\quad \int\limits_t^d\sqrt{x}J_\alpha\{2\sqrt{N}x\}\,\mathrm{d}x=\frac{\sin\left[2\sqrt{N}d-\frac{\pi}{2}\left(\alpha+\frac{1}{2}\right)\right]}{2\sqrt{\pi}\cdot N^{3/4}}-$$
$$-\frac{\sin\left[2\sqrt{N}t-\frac{\pi}{2}\left(\alpha+\frac{1}{2}\right)\right]}{2\sqrt{\pi}\cdot N^{3/4}}+O(1)\frac{1}{t\cdot N^{5/4}}=O(1)\frac{1}{N^{3/4}}.$$

From (59), (60), (61) we obtain

$$(62)\quad \sum_{a\le k<b}\left(\int\limits_t^d u_k(x)\,\mathrm{d}x\right)^2\le c\sum_{a\le k<b}k^{-\frac{3}{2}}\le c\frac{\sqrt{b-a}}{a},$$

where $\frac{c}{\sqrt{k}} \le t \le w$.

b) It is known ([9], Theorem 8.22.4)

$$(63)\qquad e^{-\frac{x}{2}} x^{\frac{\alpha}{2}} L_k^{(\alpha)}(x) = N^{-\frac{\alpha}{2}} \frac{\Gamma(k+\alpha+1)}{k!} J_\alpha\left\{2(Nx)^{\frac{1}{2}}\right\} + x^{\frac{\alpha}{2}+2} O(k^\alpha),$$

where $N = k + \frac{\alpha+1}{2}$, $0 < x \le \frac{c}{k}$. From (63) we get

$$(64)\qquad e^{-\frac{x^2}{2}} x^{\alpha+\frac{1}{2}} L_k^{(\alpha)}(x^2) = N^{-\frac{\alpha}{2}} \frac{\Gamma(k+\alpha+1)}{k!} \sqrt{x} J_\alpha(2\sqrt{N}x) + x^{\alpha+\frac{9}{2}} O(k^\alpha).$$

Therefore the error term of $\sum\limits_{a \le k < b} \left(\int\limits_t^{\frac{c}{\sqrt{k}}} u_k(x)\,\mathrm{d}x \right)^2$ is

$$(65)\qquad \sum_{a \le k < b} \left(x^{-\frac{\alpha}{2}} \cdot k^\alpha \cdot \int_t^{\frac{c}{\sqrt{k}}} k^{\alpha+\frac{9}{2}}\,\mathrm{d}x \right)^2 \le c\frac{\sqrt{b-a}}{a}.$$

It is known ([11], p.170) $J_\nu(z) \asymp \frac{(z/2)^\nu}{\Gamma(\nu+1)}$ if $z \to 0$. Using this we have

$$(66)\qquad \int_t^{\frac{c}{\sqrt{k}}} \sqrt{x} J_\alpha\{2\sqrt{N}x\}\,\mathrm{d}x \asymp \int_t^{\frac{c}{\sqrt{k}}} N^{\frac{\alpha}{2}} x^{\alpha+\frac{1}{2}}\,\mathrm{d}x = O(1)\frac{1}{k^{3/4}}.$$

From (64), (65), (66) we obtain

$$(67)\qquad \sum_{a \le k < b} \left(\int_t^{\frac{c}{\sqrt{k}}} u_k(x)\,\mathrm{d}x \right)^2 \le c\frac{\sqrt{b-a}}{a}.$$

From (62) and (67) we obtain

$$(68)\qquad \sum_{a \le k < b} \left(\int_t^{d} u_k(x)\,\mathrm{d}x \right)^2 \le c\frac{\sqrt{b-a}}{a},$$

where $0 < t \le d$. From (57) and (68) the Lemma 6 follows.

Let us prove the Theorem 3. Let K be an arbitrary fixed compact interval $K \subset G$, $R_0 > 0$, $\frac{R_0}{2} < R < R_0$, $x \in K$, $y \in G$. Count the Fourier–

coefficients of the function $w(x+\cdot)$ according to the system $\{u_k\}$

$$\langle u_k, w\rangle = \frac{1}{\mu\pi}\int_0^R \left(\frac{\sin\mu t}{t}\right)^2 (u_k(x+t)+u_k(x-t))\,\mathrm{d}t.$$

Using (47)

$$\langle u_k, w\rangle = u_k(x)\frac{2}{\mu\pi}\int_0^R \left(\frac{\sin\mu t}{t}\right)^2 \cos\mu_k t\,\mathrm{d}t + \frac{2}{\mu\pi}\int_0^R \left(\frac{\sin\mu t}{t}\right)^2 \cdot$$

$$\cdot\frac{1}{2\mu_k}\int_{x-t}^{x+t} q(\xi)u_k(\xi)\cdot\sin\mu_k(t-|x-\xi|)\,\mathrm{d}\xi\,\mathrm{d}t =$$

$$= u_k(x)\delta(\mu,\varrho_k) - u_k(x)I_R^*(\mu,\varrho_k) + \frac{1}{\mu\mu_k\pi}\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi,$$

where

$$I_R^*(\mu,\varrho_k) := \frac{2}{\mu\pi}\int_R^\infty \left(\frac{\sin\mu t}{t}\right)^2 \cos\mu_k t\,\mathrm{d}t.$$

Hence

$$w(x+y) = \sum_{\varrho_k<2\mu} u_k(x)u_k(y)\left(1-\frac{\mu_k}{2\mu}\right) - \sum_{k=1}^\infty u_k(x)u_k(y)I_R^*(\mu,\varrho_k) +$$

$$+\frac{1}{\mu\pi}\sum_{k=1}^\infty \frac{1}{\mu_k}\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\cdot u_k(y)$$

where the equality is in $L^2(G)$-convergence in y. Applying S_{R_0} operator we obtain

$$(69)\qquad S_{R_0}[w(x+y)] = \sum_{\varrho_k<2\mu} u_k(x)u_k(y)\left(1-\frac{\mu_k}{2\mu}\right) -$$

$$-\sum_{k=1}^\infty u_k(x)u_k(y)S_{R_0}[I_R^*(\mu,\varrho_k)] + \frac{1}{\mu\pi}\sum_{k=1}^\infty \frac{1}{\mu_k} S_{R_0}\left[\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right] u_k(y)$$

where the equality is in $L^2(G)$-convergence in y. From (49) and (50) we have that in (69) the equality is true in the usual sense also. In (69) multiplying by $f(y)$ both sides and integrating on G we obtain

$$\int_0^\infty S_{R_0}[w(x+y)]f(y)\,\mathrm{d}y = \sum_{\varrho_k<2\mu} u_k(x)\langle f,u_k\rangle\left(1-\frac{\mu_k}{2\mu}\right) - \tag{70}$$

$$-\sum_{k=1}^\infty u_k(x)\cdot\int_0^\infty f(y)u_k(y)\,\mathrm{d}y\cdot S_{R_0}[I_R^*(\mu,\varrho_k)]+$$

$$+\frac{1}{\mu\pi}\sum_{k=1}^\infty \frac{1}{\mu_k} S_{R_0}\left[\int_{x-R}^{x+R} q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right]\cdot\int_0^\infty f(y)u_k(y).$$

Integrating by parts

$$\int_0^\infty f(y)u_k(y)\,\mathrm{d}y = \left[f(y)\int_0^y u_k(\tau)\,\mathrm{d}\tau\right]_0^\infty - \int_0^\infty f'(y)\int_0^y u_k(\tau)\,\mathrm{d}\tau\,\mathrm{d}y =$$

$$= -\int_0^\infty f'(y)\int_0^y u_k(\tau)\,\mathrm{d}\tau\,\mathrm{d}y.$$

Using Lemma 3, Lemma 1 and Lemma 6 we obtain

$$\left|\sum_{k=1}^\infty u_k(x)\cdot\int_0^\infty f(y)u_k(y)\,\mathrm{d}y\cdot S_{R_0}[I_R^*(\mu,\varrho_k)]\right| \le$$

$$c\int_0^\infty |f'(y)|\cdot\sum_{k=1}^\infty\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|\cdot\|u_k\|_{L^\infty(K)}\cdot\frac{1}{\mu}\left(\frac{1}{1+(2\mu-\varrho_k)^2}+\frac{1}{1+\varrho_k^2}\right)\mathrm{d}y \le$$

$$\le\frac{c}{\mu}\int_0^\infty |f'(y)|\cdot\left(\sum_{\substack{k=1\\|2\mu-\varrho_k|\le 1}}^\infty \|u_k\|^2_{L^\infty(K)}\right)^{\frac12}\cdot\left(\sum_{\substack{k=1\\|2\mu-\varrho_k|\le 1}}^\infty \left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|^2\right)^{\frac12}\mathrm{d}y+$$

$$+\frac{c}{\mu}\int_0^\infty |f'(y)|\cdot\sum_{l=0}^\infty\left(\sum_{l\le\varrho_k<l+1}\|u_k\|^2_{L^\infty(K)}\right)^{\frac12}\cdot$$

$$\cdot\left(\sum_{l\le\varrho_k<l+1}\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|^2\right)^{\frac12}\cdot\left(\frac{1}{1+l^2}+\frac{1}{1+(2\mu-l)^2}\right)\mathrm{d}y\le$$

$$\le\frac{c}{\mu}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y.$$

i.e.

(71)

$$\sum_{k=1}^\infty u_k(x)\int_0^\infty f(y)u_k(y),\,\mathrm{d}y\cdot S_{R_0}[I_R^*(\mu,\varrho_k)]=O(1)\frac{1}{\mu}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y.$$

Now we estimate the other sum. Using Lemma 4, Lemma 1 and Lemma 6 we obtain

$$\left|\sum_{k=1}^\infty\frac{1}{\mu\mu_k}S_{R_0}\left[\int_{x-R}^{x+R}q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right]\cdot\int_0^\infty f(y)u_k(y)\,\mathrm{d}y\right|\le$$

$$\frac{c}{\mu}\int_{x-R_0}^{x+R_0}\frac{|q(\xi)|}{|x-\xi|^\varepsilon}\mathrm{d}\xi\cdot\int_0^\infty|f'(y)|\sum_{\substack{k=1\\1\le\varrho_k<2\mu}}^\infty\frac{1}{\mu\varrho_k}\|u_k\|_{L^\infty(K_{R_0})}\cdot\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|\mathrm{d}y+$$

$$+\frac{c}{\mu}\int_{x-R_0}^{x+R_0}\frac{|q(\xi)|}{|x-\xi|^\varepsilon}\mathrm{d}\xi\cdot\int_0^\infty|f'(y)|\cdot\sum_{\substack{k=1\\1\le\varrho_k<2\mu}}\frac{1}{\varrho_k^\varepsilon}\|u_k\|_{L^\infty(K_{R_0})}\cdot\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|\mathrm{d}y+$$

$$+\frac{c}{\mu}\int_{x-R_0}^{x+R_0}\frac{|q(\xi)|}{|x-\xi|^\varepsilon}\mathrm{d}\xi\cdot\int_0^\infty|f'(y)|\cdot\sum_{\substack{k=1\\\varrho_k\ge2\mu}}^\infty\frac{\mu^{2-\varepsilon}}{\varrho_k^2}\cdot\|u_k\|_{L^\infty(K_{R_0})}\cdot\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\,\mathrm{d}y\right|\le$$

$$\le\frac{c}{\mu}\int_0^\infty|f'(y)|\sum_{l=1}^{2\mu}\frac{1}{l^\varepsilon}\left(\sum_{l\le\varrho_k\le l+1}\|u_k\|^2_{L^\infty(K_{R_0})}\right)^{\frac12}\cdot$$

$$\cdot\left(\sum_{l\le\varrho_k\le l+1}\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|^2\right)^{\frac12}\mathrm{d}y+\frac{c}{\mu}\int_0^\infty|f'(y)|\cdot\mu^{2-\varepsilon}\cdot\sum_{l=2\mu}^\infty\frac{1}{l^2}.$$

$$\cdot\left(\sum_{l\le\varrho_k\le l+1}\|u_k\|^2_{L^\infty(K_{R_0})}\right)^{\frac12}\cdot\left(\sum_{l\le\varrho_k\le l+1}\left|\int_0^y u_k(\tau)\,\mathrm{d}\tau\right|^2\right)^{\frac12}\mathrm{d}y\le$$

$$\le\frac{c}{\mu}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y\sum_{l=1}^{2\mu}\frac{1}{l^{3/4+\varepsilon}}+c\mu^{1-\varepsilon}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y\cdot\sum_{l=2\mu}^\infty\frac{1}{l^{2+3/4}}\le$$

$$\le\frac{c}{\mu}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y$$

where we choose ε such that $1>\varepsilon>\frac14$. Hence

$$\frac{1}{\mu\pi}\sum_{k=1}^\infty\frac{1}{\mu_k}S_{R_0}\left[\int_{x-R}^{x+R}q(\xi)u_k(\xi)\gamma_k^*\,\mathrm{d}\xi\right]\cdot\int_0^\infty f(y)u_k(y)\,\mathrm{d}y=\tag{72}$$

$$=O(1)\frac{1}{\mu}\int_0^\infty|f'(y)|(1+y^2)\,\mathrm{d}y.$$

Obviously

$$\int_0^\infty S_{R_0}[w(x+y)]f(y)\,\mathrm{d}y=$$

$$=\int_0^\infty\Big(S_{R_0}[w(x+y)]-w(x+y)\Big)f(y)\,\mathrm{d}y+\int_0^\infty w(x+y)f(y)\,\mathrm{d}y.$$

Hence

$$\int_0^\infty\Big(S_{R_0}[w(x+y)]\Big)-w(x+y)f(y)\,\mathrm{d}y=$$

$$=\left[\int_0^y\Big(S_{R_0}[w(x+\tau)]-w(x+\tau)\Big)\,\mathrm{d}\tau f(y)\right]_0^\infty-$$

$$-\int_0^\infty f'(y)\cdot\int_0^y\Big(S_{R_0}[w(x+\tau)]-w(x+\tau)\Big)\,\mathrm{d}\tau\,\mathrm{d}y=$$

$$= -\int_0^\infty f'(y)\cdot \int_0^y \Big(S_{R_0}[w(x+\tau)] - w(x+\tau)\Big)\,\mathrm{d}\tau\,\mathrm{d}y.$$

Taking into account of (16) we obtain

$$\int_0^\infty \Big(S_{R_0}[w(x+y)] - w(x+y)\Big) f(y)\,\mathrm{d}y = O(1)\frac{1}{\mu}\int_0^\infty |f'(y)|\,\mathrm{d}y.$$

Hence

$$\int_0^R S_{R_0}[w(x+y)]f(y)\,\mathrm{d}y = \int_0^\infty w(x+y)f(y)\,\mathrm{d}y + O(1)\frac{1}{\mu}\int_0^\infty |f'(y)|\,\mathrm{d}y. \tag{73}$$

From the definition of $w(x+y)$, (71), (72), (73) and (70) the Theorem follows.

References

[1] I. Joó and A. Sövegjártó, On the Fejér means of eigenfunction expansions, *Per. Math.* (to appear).

[2] I. Joó, On Fejér summability of Hermite–Fourier series, *Per. Math.* (to appear).

[3] Joó and V. Komornik, On the equiconvergence of expansions by Riesz bases formed by eigenfunctions of the Schrödinger operator, *Acta Sci. Math.* (Szeged), **46** (1983), 357–375.

[4] V. Komornik, On the equiconvergence of eigenfunction expansions associated with ordinary linear differential operators, *Acta Math. Acad. Sci. Hung.*, **47** (1986), 261–280.

[5] V. Komornik, Upper estimates for the eigenfunctions of higher order linear differential operator, *Acta Sci. Math. Szeged*, **45** (1983), 260–271.

[6] I. Joó, On the divergence of eigenfunction expansions, *Annales Univ. Sci. Budapest.*, Sect. Math., **32** (1989), 3–36.

[7] I. Joó, Upper estimates for the eigenfunctions of the Schrödinger operator, *Acta Sci. Math.*, 44 (1982), 87–93.

[8] M. Horváth, On eigenfunction expansions, *Annales Univ. Sci. Budapest.*, Sect. Math., **32** (1989), 159–190.

[9] G. Szegő, *Orthogonal Polynomials*, AMS Coll. Publ., New York, 1959.

[10] I. Joó, On Laguerre functions, *Annales Univ. Sci. Budapest.*, Sect. Math., **35** (1992), 133–146.

[11] A. F. Nikiforov and V. B. Uvarov, *Special functions of mathematical physics*, Moscow, 1984.

[12] I. Joó, Saturation theorems for Hermite–Fourier series. *Acta Math. Hung.*, **57** (1991), 169–179.

ANNALES UNIV. SCI. BUDAPEST., 35 (1992), 189–215

ON CONTROLLABILITY OF BILINEAR SYSTEMS I (CONTROLLABILITY IN FINITE DIMENSIONS)

By

I. JOÓ and N. M. TUAN

Department of Analysis, L. Eötvös University, Budapest

(Received August 14, 1992)

1. Introduction

The theory of bilinear systems proves to be a natural extension of the linear system theory and plays a more and more significant role in dealing with complicated models beyond the scope of the linear theory. A great development has been obtained in studying linear systems; we mention the description of the structure of systems, controllability, observability, constructibility theories, optimal control and their applications. Many efficient and beautiful principles are used as the rank conditions of Kalman for the controllability and observability, the optimal control principle of Pontriagin, the duality principle etc., see [1], [2].

However, linear systems are not general enough for the description of several phenomena in natural sciences; they can be investigated by nonlinear systems. In recent years many efforts have been made in order to transform the linear theory to the case of nonlinear systems. It presents several difficulties and needs the application of recently elaborated mathematical tools. The intermediate step between linear and general nonlinear systems is a class of dynamical systems, namely the bilinear systems. It is an area in active development, whose results are applied e.g. in directing rockets, in biology, in the problems of environment protection etc.

As in the case of linear systems, many problems arise here, as the structure of the system, controllability, observability, constructibility, optimal control and their applications in concrete problems. The first step is the study of the properties of the systems (controllability, observability); the present paper is a contribution to this part of the theory. The paper is the description of the reachability sets and the controllability properties of bilinear systems defined on a state space e.g. on a differentiable manifold, finite dimensional vector space, infinite dimensional Banach space or on a Lie group. Fundamental results are known in describing reachability sets of systems defined on finite dimensional vector space or finite dimensional smooth manifolds by the aid of differential geometry, algebra, analysis and functional analysis. However these results are still too "general", far from the nice rank criteria of Kalman in the linear theory.

Our plan is to get nice and more practical controllability criteria and to describe the reachability sets for some special bilinear systems in order to make the applications easier.

In what follows we intend to give a (necessarily incomplete) list of known results in controllability theory. Introduce the following systems

$$\dot{x}(t)=Ax(t)+N(x(t)),u(t))+Bv(t). \tag{1}$$

Here $0\leq t<\infty$ and $x(t)\in X$, $u(t)\in U$, $v(t)\in V$ for all $t\geq 0$, where X, U and V are fixed Banach spaces. Further

$$A:X\to X \qquad B:V\to X$$

are bounded linear operators.

$$N:X\times U\to X$$

is a bounded bilinear operator. Now X is called the state space, U and V are the spaces of the values of controls. We assume that the control functions $u(t)$, $v(t)$ are locally Bochner integrable, or equivalently that $u(t)$, $v(t)$ are strongly measurable and

$$\|u(t)\|_U,\ \|v(t)\|_V\in L_1^{\mathrm{loc}}([0,\infty)),$$

see HILLE, PHILLIPS [25], 3.7. By a solution $x(t)$ of (1) we mean a locally absolutely continuous function satisfying (1) for almost with $t>0$. If we require that $u(t)$, $v(t)$ be piecewise continuous control functions then (1) must hold in every continuity point. It is known (KREIN [26] II.2.1. II.3.2.) that for any given initial condition $x(0)=x_0\in X$ the system (1) has unique solution which will be denoted by $x(t,x_0,u,v)$. Introduce the reachability sets

$$R_t(x_0):=\{x(t,x_0,u,v):u(t),\ v(t)\text{ are control functions}\}$$

$$R(x_0):=\bigcup_{t>0}R_t(x_0).$$

Consider now the special case of (1) where $B=0$ and X, U, V are finite dimensional spaces.

In this case (1) can be rewritten as

$$\dot{x}(t)=\left(A+\sum_{i=1}^{m}u_i(t)B_i\right)x(t), \tag{2}$$

here $X=\mathbb{R}^n$, $U=\mathbb{R}^m$, A, $B_i\in\mathbb{R}^{n\times n}$. The control functions $u_i(t)$ are supposed to be piecewise continuous and the right and left limits exist in the discontinuity points. In this case the structure of the reachability set $R_t(x)$ is described in the works of H. J. SUSSMANN, V. JURDJEVIC, R. M. HIRSCHORN, J. KUCERA and R. W. BROCKETT. Here we mention only some basic facts. First introduce some notions. For two matrices A, $B \in \mathbb{R}^{n\times n}$ introduce their Lie product

$$[A,B] := AB - BA$$

A subset of $\mathbb{R}^{n\times n}$ is called Lie algebra if it is a linear subspace closed for the action of taking Lie products. The Lie algebra generated by the subset $\mathcal{N} \subset \mathbb{R}^{n\times n}$ is the smallest Lie algebra containing $\mathcal{N}$, we denote it by $\{\mathcal{N}\}_{\mathrm{LA}}$. $\mathcal{N}_1 \subset \mathcal{N}$ is called ideal in $\mathcal{N}$ if $N_1 \in \mathcal{N}_1$, $N \in \mathcal{N}$ imply $[N, N_1] \in \mathcal{N}_1$ ($\mathcal{N}_1, \mathcal{N}$ are Lie algebras). For a Lie algebra $\mathcal{N} \subset \mathbb{R}^{n\times n}$ let

$$e^{\mathcal{N}} := \left\{ e^{N_1} e^{N_2} \dots e^{N_p} : p \geq 1,\ N_i \in \mathcal{N}\ (i = 1, \dots, p) \right\}.$$

The following Lie algebras will be characteristic to the system (2)

$$\mathcal{L} := \{A, B_1, \dots, B_m\}_{\mathrm{LA}}$$
$$\mathcal{B} := \{B_1, \dots, B_m\}_{\mathrm{LA}}$$
$$\mathcal{L}_0 := \{ad_A^k B_i : i = 1, \dots, m;\ k = 0, 1, \dots\}_{\mathrm{LA}}$$

where

$$ad_A^0 B_i := B_i, \quad ad_A^1 B_i := [A, B_i], \quad ad_A^k B_i := \left[A, ad_A^{k-1} B_i\right].$$

Then $\mathcal{L}_0$ is the ideal generated by $B_1, \dots, B_m$ in $\mathcal{L}$. We know

THEOREM A. ([8] Theorem 4.5) *Let $x \in \mathbb{R}^n \setminus \{0\}$ and $t > 0$, then for the system* (2)

1. $R_t(x) \subset e^{At} e^{\mathcal{L}_0} x$
2. *$R_t(x)$ has nonempty interior in $e^{At} e^{\mathcal{L}_0} x$.*

THEOREM B. ([16] Theorem 3.6) *Take the additional assumption that*

$$[\mathcal{L}_0, \mathcal{B}](x) = \{[L,B]x : L \in \mathcal{L}_0,\ B \in \mathcal{B}\} \subset \mathcal{B}(x).$$

Then

$$R_t(x) = e^{At} e^{\mathcal{L}_0} x \qquad \text{for all} \quad t > 0.$$

THEOREM C. ([13], Theorem 2.5) *Suppose that $A = 0$ in (2), i.e. $\dot{x}(t) = \sum_{i=1}^{m} u_i(t) B_i x(t)$. Then*

1. *There exists $0 < t < \infty$ with $R_t(x) = R(x) = e^{\mathcal{L}_0} x$ for all $x \in \mathbb{R}^n \setminus \{0\}$.*
2. *The system is controllable in $\mathbb{R}^n \setminus \{0\}$, i.e.*

$$R(x) = \mathbb{R}^n \setminus \{0\} \quad \text{for all} \quad x \in \mathbb{R}^n \setminus \{0\}, \quad \text{if and only if}$$
$$\mathcal{L}_0(x) = \mathbb{R}^n \quad \text{for all} \quad x \in \mathbb{R}^n \setminus \{0\}.$$

In **2.** below we investigate the controllability of the system (1) with finite dimensional state space and control spaces, namely

$$\dot{x}(t)=\left(A+\sum_{i=1}^{m}u_i(t)B_i\right)x(t)+\sum_{j=1}^{r}v_j(t)d_j \tag{3}$$

here A, B_1, ..., $B_m\in\mathbb{R}^{n\times n}$, $d_1,\dots,d_r\in\mathbb{R}^n$, $d_j\neq 0$.

Using the extension techniques created by V. JURDJEVIC, I. KUPKA and G. SALLET [28]–[31] we manage to describe the reachability set of (3) from the origin (Theorem 1). Then we give necessary and sufficient controllability conditions for the system (3). Our result of this type presents the direct generalization of the following known result:

THEOREM D. ([1], 2.3, p.36) *The system*

$$\dot{x}(t)=Ax(t)=\sum_{i=1}^{t}v_j(t)d_j$$

is completely controllable in $\mathbb{R}^n$ *if and only if*

$$\operatorname{rank}[D,AD,\dots,A^{n-1}D]=n$$

where $D:=[d_1,\dots,d_r]$.

In the **3.** we study the controllability and the accessibility of the system (2). Here we give necessary and sufficient conditions for the controllability and for the strong accessibility of some special systems (Theorems 4, 5). The conditions given here depend only on the coefficients of the system, and they are easily checked. Then some examples are given, too.

In the **4.** we apply the results obtained in **2.** and **3.** to investigate the controllability of linear systems with output feedback. This explains why the case $\operatorname{rank}B_i=1$, $i=1,2,\dots,m$, considered in **3.**, is interesting.

2. The general case ($d_j\neq 0$)

Consider the system

$$\dot{x}(t)=\left(A+\sum_{i=1}^{m}u_i(t)B_i\right)x(t)+\sum_{j=1}^{r}v_j(t)d_j, \tag{3}$$

here $A,B_1,\dots,B_m\in\mathbb{R}^{n\times n}$, $d_1,\dots,d_r\in\mathbb{R}^n$ and the control functions $u(t)=$ $=(u_1(t),\dots,u_m(t))$, $v(t)=(v_1(t),\dots,v_r(t))$ are piecewise continuous. This means that for every $0<T<\infty$ there exists a finite decomposition of $(0,T)$ into disjoint intervals, on every of which $u(t)$ and $v(t)$ are continuous. The equation (3) holds for all $t \geq 0$ except for the discontinuity points of the control functions.

As usual, consider the following polysystem corresponding to (3)

$$\dot{x}(t) = \left(\overline{A} + \sum_{i=1}^{m} u_i(t)\overline{B}_i + \sum_{j=1}^{r} v_j(t)\overline{D}_j \right) x(t) \tag{4}$$

where $\overline{A}$, $\overline{B}_i$, $\overline{D}_j$ are vector fields on $\mathbb{R}^n$ defined by $\overline{A}(x) = Ax$, $\overline{B}_i(x) = B_i x$, $\overline{D}_j(x) = d_j$, $i = 1,\ldots,m$; $j = 1,\ldots,r$. Here $\overline{D}_j$ is a constant vector field corresponding to d_j.

It is clear that the solutions of (3) and (4) are the same.

Introduce the following notations:

$$\mathscr{L} := \{\overline{A}, \overline{B}_1, \ldots, \overline{B}_m, \overline{D}_1, \ldots, \overline{D}_r\}_{\mathrm{LA}}$$

$$\mathscr{L}_0 := \{ad^k_{\overline{A}}\overline{G}_i : \overline{G}_i \in \{\overline{B}_1, \ldots, \overline{B}_m, \overline{D}_1, \ldots, \overline{D}_r\},\ k = 0,1,\ldots\}_{\mathrm{LA}}$$

$$E := \{A^{k_0} B_{i_1} A^{k_1} \ldots B_{i_l} A^{k_l} d_j : k_s = 0,1,\ldots;\ s = 0,1,\ldots,l;$$
$$l = 0,1,\ldots;\ i_1,\ldots,i_l \in \{1,\ldots,m\},\ j \in \{1,\ldots,r\}\}.$$

Denote $\tilde{E}$ the set of the constant vector fields corresponding to all vectors of E; V_E (resp. $V_{\tilde{E}}$) the linear hull generated by E (resp. $\tilde{E}$).

REMARK: 1. $\mathscr{L}_0$ is the ideal generated by $\overline{B}_1, \ldots, \overline{B}_m, \overline{D}_1, \ldots, \overline{D}_r$ in $\mathscr{L}$.

2. From the definition of the Lie bracket of two vector fields f, g on $\mathbb{R}^n$:

$$[f,g](x) := \frac{\partial f(x)}{\partial x} g(x) - \frac{\partial g(x)}{\partial x} f(x)$$

(sometimes may be defined with the opposite sign) we have

$$[\overline{G}, \overline{D}](x) = Gd \quad \text{and} \quad [\overline{D}, \overline{F}](x) = 0 \quad \text{for all} \quad x \in \mathbb{R}^n$$

if $\overline{G}(x) = Gx$, $\overline{D}(x) = d$, $\overline{F}(x) = f$, d, $f \in \mathbb{R}^n$.

We need the following result:

LEMMA 1. ([8], Corollaries 4.6 and 4.7) *For the system* (4)

(a) *$R(x)$ has nonempty interior in $\mathbb{R}^n$ if and only if* $\dim \mathscr{L}(x) = \dim\{Lx : L \in \mathscr{L}\} = n$,

(b) *$R_t(x)$ has nonempty interior in $\mathbb{R}^n$ for all $t > 0$ if and only if* $\dim \mathscr{L}_0(x) = n$.

LEMMA 2. *For the system* (4), $\mathscr{L}(0) = \mathscr{L}_0(0) = V_E$.

PROOF. (a) In the first step we show that

$$\mathscr{L} = \{c\overline{A} + L_0 : c \in \mathbb{R},\ L_0 \in \mathscr{L}_0\}.$$

Denote $\mathscr{H}$ the set on the right. The inclusion $\mathscr{H}\subset\mathscr{L}$ is obvious and $\overline{A}$, $\overline{B}_i$, $\overline{D}_j\in\mathscr{H}$. Hence $\mathscr{H}=\mathscr{L}$ will follow if $\mathscr{H}$ is a Lie algebra. The linearity of $\mathscr{H}$ is obvious and

$$[c\overline{A}+L_0,c'\overline{A}+L_0']=c[\overline{A},L_0']-c'[\overline{A},L_0]+[L_0,L_0']$$

and the right hand side belongs to $\mathscr{L}_0$ since $\mathscr{L}_0$ is an ideal in $\mathscr{L}$. Since we did not prove the ideal property, we give a short proof of the fact that $L_0\in\mathscr{L}_0$ implies $[\overline{A},L_0]\in\mathscr{L}_0$. Denote $\tilde{\mathscr{L}}_0:=\{L_0\in\mathscr{L}_0:[\overline{A},L_0]\in\mathscr{L}_0\}$. Then $\tilde{\mathscr{L}}_0$ contains all generating elements $ad^k_{\overline{A}}\overline{G}_i$ of $\mathscr{L}_0$, and it is Lie algebra, too, since by the Jacobi identity

$$[[L_0,L_0'],\overline{A}]=[L_0,[L_0',\overline{A}]]+[L_0',[\overline{A},L_0]]\in\mathscr{L}_0.$$

Hence $\tilde{\mathscr{L}}_0=\mathscr{L}_0$ as we asserted.

From here the ideal property of $\mathscr{L}_0$ follows at once. Indeed, let $\mathscr{L}_1=\{L\in\in\mathscr{L}:[L,L_0]\in\mathscr{L}_0$ for all $L_0\in\mathscr{L}_0\}$. This is a Lie algebra (by the above used Jacobi identity) and contains all generating elements of $\mathscr{L}$ (in particular $\overline{A}$ by the above proof). Consequently $\mathscr{L}_1=\mathscr{L}$, hence $\mathscr{L}_0$ is ideal in $\mathscr{L}$.

(b) $\mathscr{L}_0(0)=\mathscr{L}(0)\subset V_E$.

By the remark 2), every element of $\mathscr{L}_0(x)$ is the finite sum of members

$$cA^{k_0}B_{i_1}A^{k_1}\dots B_{i_l}A^{k_l}x$$

or

$$cA^{k_0}B_{i_1}A^{k_1}\dots B_{i_l}A^{k_l}d_j.$$

Replacing $x=0$, for $L\in\mathscr{L}_0$

$$L(0)=\sum\gamma_{i_1\dots i_l}\cdot A^{k_0}B_{i_1}A^{k_1}\dots B_{i_l}A^{k_l}d_j\in V_E$$

i.e. $\mathscr{L}_0(0)\subset V_E$. By the point (a) for $L\in\mathscr{L}$, $L=c\overline{A}+L_0$ we have $L(0)==L_0(0)\in V_E$ hence $\mathscr{L}(0)\subset V_E$.

(c) $V_E\subset\mathscr{L}_0(0)$.

By the linearity of $\mathscr{L}_0$ it is enough to show that $E\subset\mathscr{L}_0(0)$. From the point 2) of the remark we have for all $x\in\mathbb{R}^n$

$$A^kd_j=ad^k_{\overline{A}}\overline{D}_j(x)\in\mathscr{L}_0(x)$$
$$B_iA^kd_j=[\overline{B}_i,ad^k_{\overline{A}}\overline{D}_j](x)\in\mathscr{L}_0(x).$$

We have seen in (a) that $\mathscr{L}_0$ is ideal in $\mathscr{L}$. Hence

$$A^{k_0}B_iA^{k_1}d_j=ad^{k_0}_{\overline{A}}ad_{\overline{B}_i}ad^{k_1}_{\overline{A}}\overline{D}_j(x)\in\mathscr{L}_0(x)$$

and by a simple induction we obtain finally

$$A^{k_0}B_{i_1}A^{k_1}\dots B_{i_l}A^{k_l}d_j\in\mathscr{L}_0(x)$$

hence $E \subset \mathcal{L}_0(x)$ for all $x \in \mathbb{R}^n$, specially $E \subset \mathcal{L}_0(0)$. From (b) and (c) the statement of Lemma 2 follows.

COROLLARY. (through Lemma 1) *For the system* (4) *the statements below are equivalent:*

(a) $R(0)$ *has nonempty interior in* $\mathbb{R}^n$

(b) $R_t(0)$ *has nonempty interior in* $\mathbb{R}^n$ *for all* $t > 0$

(c) $\dim V_E = n$.

The following theorem is based on an assertion of R. W. BROCKETT (see [23], Theorem 4 of 2.1). But his proof is rather sketchy, given from engineer's point of view. Here, using the enlargements technique of V. JURDJEVIC and I. KUPKA we present a "mathematical" proof.

THEOREM 1. *For the system* (3) *the reachability set at time* t *from the origin is the linear subspace generated by the set* E, *i.e.* $R_t(0) = V_E$ *for all* $t > 0$.

First we introduce the enlargements technique (for details see [28]–[31]).

Enlargement technique (in our case).

Now we consider only the system (3) with the control class $\mathcal{P}$ of all piecewise constant functions on $[0, \infty)$. Rather than working with (3) it will be convenient to work with an equivalent family of vector fields

$$\mathcal{F} := \left\{ \overline{A} + \sum_{i=1}^{m} u_i \overline{B}_i + \sum_{j=1}^{r} v_j \overline{D}_j : (u_1, \ldots, u_m, v_1, \ldots, v_r) \in \mathbb{R}^{m+r} \right\}.$$

For $F \in \mathcal{F}$ denote $\exp(tF)x$ the solution of

$$\begin{cases} \dot{x}(t) = Fx(t) \\ x(0) = x \end{cases}$$

As usual we denote

$$A_{\mathcal{F}}(t,x) := \{\exp(t_1F_1)\exp(t_2F_2)\ldots\exp(t_pF_p)x : F_1, \ldots, F_p \in \mathcal{F}, \quad t_i \geq 0, \quad t_1 + \ldots + t_p = t, \quad p = 1, 2, \ldots\},$$

the accessibility set of $\mathcal{F}$ at time t from the state x.

$$\mathbb{A}_{\mathcal{F}}(t,x) = \bigcup_{0 \leq s \leq t} A_{\mathcal{F}}(s,x).$$

We see that the accessibility sets of $\mathcal{F}$ and the reachability sets of (3) (with $\mathcal{P}$) are the same.

DEFINITION: Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be families of vector fields on $\mathbb{R}^n$. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are equivalent, written $\mathcal{F}_1 \sim \mathcal{F}_2$, if

$$\operatorname{cl} \mathbb{A}_{\mathcal{F}_1}(t,x) = \operatorname{cl} \mathbb{A}_{\mathcal{F}_2}(t,x) \quad \text{for all} \quad x \in \mathbb{R}^n \quad \text{and all} \quad t > 0.$$

If we denote $\mathrm{Sat}(\mathscr{F}) := \cup\{\mathscr{F}' : \mathscr{F}' \sim \mathscr{F}\}$ then $\mathrm{Sat}(\mathscr{F}) \sim \mathscr{F}$. We use $\mathscr{L}(\mathscr{F})$ to denote the Lie algebra generated by $\mathscr{F}$. The Lie saturate of $\mathscr{F}$, $\mathrm{LS}(\mathscr{F}) := \mathrm{Sat}(\mathscr{F}) \cap \mathscr{L}(\mathscr{F})$ will play an important role in our technique to determine the transitivity of a given system (see below). We say that a diffeomorphism $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is a normalizer of a polysystem $\mathscr{F}$ (i.e. a family of vector fields) if

$$\Phi\left(\mathrm{cl}\,\mathbb{A}_{\mathscr{F}}(t, \Phi^{-1}(x))\right) \subset \mathrm{cl}\,\mathbb{A}_{\mathscr{F}}(t,x) \quad \text{for all } x \in \mathbb{R}^n \text{ and all } t > 0.$$

LEMMA 3. ([30])

1. *Φ is a normalizer of $\mathscr{F}$ if both $\Phi(x) \in \mathrm{cl}\,\mathbb{A}_{\mathscr{F}}(t,x)$ and $\Phi^{-1}(x) \in \mathrm{cl}\,\mathbb{A}_{\mathscr{F}}(t,x)$ for all $x \in \mathbb{R}^n$ and all $t > 0$*

2. *Let $\mathrm{Norm}(\mathscr{F}) := \{\Phi \in \mathrm{Diff}(\mathbb{R}^n) : \Phi$ is a normalizer of $\mathscr{F}\}$ and for a vector field F and a diffeomorphism Φ let*

$$\Phi_* F(x) := \frac{\mathrm{d}}{\mathrm{d}t} \Phi\left(\exp(tF)\Phi^{-1}(x)\right)\Big|_{t=0}$$

If $\mathscr{F}$ is any polysystem, then $\displaystyle\bigcup_{\Phi \in \mathrm{Norm}(\mathscr{F})} \{\Phi_* F : F \in \mathscr{F}\} \sim \mathscr{F}$.

3. *If $\mathscr{F}$ is any smooth polysystem, then $\mathscr{F}$ is equivalent to the closed convex hull generated by $\{\lambda F : 0 \le \lambda \le 1,\ F \in \mathscr{F}\}$. The closure is taken in the C^∞ topology on compact subsets of $\mathbb{R}^n$.*

The technique to prove the transitivity is based on the following statement: A polysystem $\mathscr{F}$ is transitive (i.e. $\mathbb{A}_{\mathscr{F}}(t,x) = \mathbb{R}^n$ for all $x \in \mathbb{R}^n$ and all $t > 0$) if $\mathrm{LS}(\mathscr{F})(x) = \mathbb{R}^n$ for all $x \in \mathbb{R}^n$. (see [28]-[31]).

Now we return to our problem. First we prove a simple lemma.

LEMMA 4. *For the system* (3) *(with $\mathscr{P}$) or for the family*

$$\mathscr{F} = \left\{ \overline{A} + \sum_{i=1}^{m} u_i \overline{B}_i + \sum_{j=1}^{r} v_j \overline{D}_j : (u,v) \in \mathbb{R}^{n+r} \right\},$$

$$\mathbb{A}_{\mathscr{F}}(t,0) = A_{\mathscr{F}}(t,0) \quad \textit{for all} \quad t \ge 0.$$

PROOF. It is obvious that $A_{\mathscr{F}}(t,0) \subset \mathbb{A}_{\mathscr{F}}(t,0)$. Now let $x \in \mathbb{A}_{\mathscr{F}}(t,0)$. Then there exist $0 \le s \le t$, $s_i \ge 0$, $i = 1, \dots, p$, $s_1 + \dots + s_p = s$ and $F_1, \dots, F_p \in \mathscr{F}$ such that $x = (\exp s_1 F_1) \dots (\exp s_p F_p)(0)$. Let $s_{p+1} = t - s$ and let $F_{p+1} = \overline{A} \in \mathscr{F}$, then $\exp(s_{p+1} F_{p+1})(0) = 0$. So $x = (\exp s_1 F_1) \dots (\exp s_p F_p) \cdot (\exp s_{p+1} F_{p+1})(0)$ and since $s_1 + \dots + s_p + s_{p+1} = s + t - s = t$, $x \in A_{\mathscr{F}}(t,0)$, hence $\mathbb{A}_{\mathscr{F}}(t,0) \subset A_{\mathscr{F}}(t,0)$. ∎

THE PROOF OF THE THEOREM 1. We use $\mathscr{F}$ to denote the family in Lemma 4. We have only to show that $\mathbb{A}_{\mathscr{F}}(t,0) = V_E$ for all $t > 0$.

Indeed by the Lemma 4 $\mathbb{A}_{\mathcal{F}}(t,0)=V_E$ for all $t>0$. From the relation $R_t(0)=\mathbb{A}_{\mathcal{F}}(t,0)=V_E$ we obtain the Theorem.

(a) $V_{\tilde{E}}\subset \mathrm{LS}(\mathcal{F})$. We have seen in the proof of the Lemma 2 that if $d=A^{k_0}B_{i_1}\dots B_{i_l}A^{k_l}d_j\in E$, then $\overline{D}(x)=d$, $\overline{D}\in\tilde{E}$ and $\overline{D}=ad_{\bar{A}}^{k_0}ad_{\bar{B}_{i_1}}\dots ad_{\bar{B}_{i_l}}ad_{\bar{A}}^{k_l}\overline{D}_j\in\mathcal{L}(\mathcal{F})$, hence $V_{\tilde{E}}\subset\mathcal{L}(\mathcal{F})$. By induction on $i_d=k_0+\dots k_l+l$ we can show that $\lambda\overline{D}\in\mathrm{Sat}(\mathcal{F})$ for all $\lambda\in\mathbb{R}$.

For $i_d=0$, $\lambda\overline{D}_j=\lim_{n\to\infty}\frac{1}{n}(\overline{A}+\lambda n\overline{D}_j)$, but $\overline{A}+\lambda n\overline{D}_j\in\mathcal{F}$ and by Lemma 3, point 3., $\lambda\overline{D}_j\in\mathrm{Sat}(\mathcal{F})$. Similarly $\overline{B}_i\in\mathrm{Sat}(\mathcal{F})$.

Assume that for $\le i_d-1$ our assertion is true. If $k_0\ne0$ let $\overline{G}=\overline{A}$, if $k_0=0$ let $\overline{G}=\overline{B}_{i_1}$, then $\lambda\overline{D}=ad_{\bar{G}}(\lambda\overline{F})$ where $\lambda\overline{F}\in\mathrm{Sat}(\mathcal{F})$ by the induction assumption. We have for $t>0$

$$\pm\frac{1}{t}\lambda\overline{F}\in\mathrm{Sat}(\mathcal{F})\qquad\text{(induction assumption)}$$

hence

$$\exp t\left(\pm\frac{1}{t}\lambda\overline{F}\right)x\in\mathbb{A}_{\mathrm{Sat}(\mathcal{F})}(t,x)\subseteq\mathrm{cl}\,\mathbb{A}_{\mathrm{Sat}(\mathcal{F})}(t,x).$$

Since $\mathrm{cl}\,\mathbb{A}_{\mathrm{Sat}(\mathcal{F})}(t,x)=\mathrm{cl}\,\mathbb{A}_{\mathcal{F}}(t,x)$ we obtain

$$\exp(\lambda\overline{F})x\in\mathrm{cl}\,\mathbb{A}_{\mathcal{F}}(t,x)$$
$$\left(\exp(\lambda\overline{F})\right)^{-1}x\in\mathrm{cl}\,\mathbb{A}_{\mathcal{F}}(t,x)$$

for all $x\in\mathbb{R}^n$ and all $t>0$. By Lemma 3, point 1., $\exp(\lambda\overline{F})\in\mathrm{Norm}(\mathcal{F})$. By the point 2. of this Lemma

$$(\exp\lambda\overline{F})_*\overline{G}\in\bigcup_{\Phi\in\mathrm{Norm}(\mathrm{Sat}(\mathcal{F}))}\{\Phi_*V:V\in\mathrm{Sat}\,\mathcal{F}\}\sim\mathrm{Sat}\,\mathcal{F}\sim\mathcal{F}.$$

(Here we use the fact that $\mathrm{Norm}(\mathcal{F})=\mathrm{Norm}(\mathrm{Sat}(\mathcal{F}))$, so $(\exp\lambda\overline{F})\in$ $\in\mathrm{Norm}(\mathrm{Sat}(\mathcal{F}).)$

Now

$$(\exp\lambda\overline{F})_*\overline{G}(x)=\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}(\exp\lambda\overline{F})(\exp t\overline{G})(\exp\lambda\overline{F})^{-1}x=$$
$$=\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}\{e^{tG}(x-\lambda f)+\lambda f\}=Gx-\lambda Gf=$$
$$=\overline{G}(x)-\lambda\overline{D}(x)=(\overline{G}-\lambda\overline{D})(x).$$

(Here $\overline{G}(x)=Gx$, $\overline{F}(x)=f$.) Hence $\overline{G}-\lambda\overline{D}=(\exp\lambda\overline{F})_*\overline{G}\in\mathrm{Sat}\,\mathcal{F}$, for all λ. Using the point 3. of the Lemma 3 we obtain $\lambda\overline{D}\in\mathrm{Sat}\,\mathcal{F}$. Since $\mathrm{Sat}\,\mathcal{F}$ is convex, $V_{\tilde{E}}\subset\mathrm{Sat}\,\mathcal{F}$ and hence $V_{\tilde{E}}\subset\mathrm{LS}(\mathcal{F})$.

(b) $\operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0) = V_E$ for all $t > 0$.

Let $d = A^{k_0} B_{i_1} \dots B_{i_l} A^{k_l} d_j$, then $d = \left(\exp t \cdot \frac{1}{t}\overline{D}\right)(0)$ hence $d \in \mathbb{A}_{V_{\tilde{E}}}(t,0)$ for all $t > 0$. If $d \in V_E$, then $d = \sum_{i=1}^{k} \lambda_i f_i$, $f_i \in E$, and $f_i = \left(\exp t \cdot \frac{1}{t}\overline{F}_i\right)(0)$.

Let $\overline{D} = \frac{1}{t} \sum_{i=1}^{k} \lambda_i \overline{F}_i$, then $d = (\exp t\overline{D})(0) \in \mathbb{A}_{V_{\tilde{E}}}(t,0)$ for all $t > 0$. We have $\mathbb{A}_{V_{\tilde{E}}}(t,0) \supset V_E$. From this and from $V_{\tilde{E}} \subset \mathrm{LS}(\mathcal{F})$, we obtain

$$\mathbb{A}_{\mathrm{LS}(\mathcal{F})}(t,0) \supset V_E \qquad \text{for all} \quad t > 0.$$

Hence

$$\operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0) \supset V_E \qquad \text{for all} \quad t > 0.$$

On the other hand, $\mathbb{A}_{\mathcal{F}}(t,0) \subset V_E$ and V_E is a finite dimensional closed subspace in $\mathbb{R}^n$, $\operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0) = V_E$ for all $t > 0$.

(c) $\mathbb{A}_{\mathcal{F}}(t,0) = V_E$.

We show that V_E is $\mathcal{F}$-invariant, i.e. $\mathbb{A}_{\mathcal{F}}(t,x) \subseteq V_E$ for all $t \geq 0$ and all $x \in V_E$.

Assume indirectly that there is $x \in V_E$ and $s > 0$ such that $\mathbb{A}_{\mathcal{F}}(s,x) \not\subseteq V_E$, i.e. there is $y \in \mathbb{A}_{\mathcal{F}}(s,x)$ and $y \notin V_E$. Let $y = (\exp t_1 F_1) \dots (\exp t_p F_p)x$ for some $F_i \in \mathcal{F}$, $t_i \geq 0$, $\sum t_i \leq s$. Since $x \in V_E = \operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0)$ (for all $t > > 0$), there is a series $(x_n) \subset \mathbb{A}_{\mathcal{F}}(t,0)$ and $x_n \to x$ as $n \to \infty$. Let $y_n = = (\exp t_1 F_1) \dots (\exp t_p F_p) x_n$, then $y_n \in \mathbb{A}_{\mathcal{F}}(t+s,0)$ and since $x_n \to x$, $y_n \to y$ as $n \to \infty$, i.e. $y \in \operatorname{cl} \mathbb{A}_{\mathcal{F}}(t+s,0) = V_E$, in contradiction with assumption that $y \notin V_E$. We obtain that V_E is $\mathcal{F}$-invariant.

With this fact we can consider only the restriction of $\mathcal{F}$ to V_E. By Lemma 1, or more exactly, by Theorem 3.1 in [8] (Here $M = V_E$, $D = \mathcal{F}$, $\mathcal{F}(D) = \mathcal{L}$, $\mathbb{L}_x(D,t) = \mathbb{A}_{\mathcal{F}}(t,x)$), since $\mathcal{L}(0) = V_E$ (Lemma 2) we obtain that for all $t > 0$ the interior of $\mathbb{A}_{\mathcal{F}}(t,0)$ relative to V_E is dense in $\mathbb{A}_{\mathcal{F}}(t,0)$ [thus, in particular $\mathbb{A}_{\mathcal{F}}(t,0)$ has a nonempty interior (relative to V_E)]. The same is true for $-\mathcal{F}$.

From this and from $\operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0) = V_E$ and $\operatorname{cl}(\operatorname{int}_{V_E} \mathbb{A}_{\mathcal{F}}(t,0)) = \operatorname{cl} \mathbb{A}_{\mathcal{F}}(t,0)$ we have

$$\operatorname{int}_{V_E} \mathbb{A}_{\mathcal{F}}(t,0) \bigcap \operatorname{int}_{V_E} \mathbb{A}_{-\mathcal{F}}(t,0) \neq 0 \quad \text{for all} \quad t > 0$$

(here int_{V_E} denotes the interior in V_E).

Let x be a common point, $x \in \operatorname{int}_{V_E} \mathbb{A}_{\mathcal{F}}(t/2,0)$, $x \in \mathbb{A}_{-\mathcal{F}}(t/2,0)$. Then there exists a trajectory $x(s)$ of $\mathcal{F}$ such that $x(0) = 0$, $x(t) = 0$, $x\left(\frac{t}{2}\right) = x \in \in \operatorname{int}_{V_E} \mathbb{A}_{\mathcal{F}}(t/2,0)$. Here we can apply the maximum principle to show that $0 \in \operatorname{int}_{V_E} \mathbb{A}_{\mathcal{F}}(t,0)$. We can also prove this fact directly. Indeed,

$0=(\exp t_1\overline{F}_1)\dots(\exp t_p\overline{F}_p)x$; since $x\in \operatorname{int}_{V_E}\mathbb{A}_{\mathcal{G}}(t/2,0)$ and $(\exp t_1\overline{F}_1)\dots(\exp t_p\overline{F}_p)$ is a homeomorphism of $\mathbb{R}^n$ leaving V_E invariant, we get

$$0\in \operatorname{int}_{V_E}\{(\exp t_1\overline{F}_1)\dots(\exp t_p\overline{F}_p)\mathbb{A}_{\mathcal{G}}(t/2,0)\}\subseteq \operatorname{int}_{V_E}\mathbb{A}_{\mathcal{G}}(t,0).$$

On the other hand we observe that if x is reached at time t starting from $x=0$ at $t=0$ using the vector fields

$$\overline{F}_k=\overline{A}+\sum_{i=1}^{m}u_i^k\overline{B}_i+\sum_{j=1}^{r}v_j^k\overline{D}_j,\quad k=1\dots,p,$$

then λx is reached at time t starting from $x=0$ at $t=0$ using the vector fields

$$\overline{F}_k^\lambda=\overline{A}+\sum_{i=1}^{m}u_i^k\overline{B}_i+\sum_{j=1}^{r}\lambda v_j^k\overline{D}_j,\qquad k=1,\dots,p.$$

By this cone property and by $0\in\operatorname{int}_{V_E}\mathbb{A}_{\mathcal{G}}(t,0)$ we obtain $\mathbb{A}_{\mathcal{G}}(t,0)=V_E$ for all $t>0$. The theorem is proved. ∎

THEOREM 2. *For the system* (3) *the following statements are equivalent*

(i) $R(x)=\mathbb{R}^n$ *for every* $x\in\mathbb{R}^n$

(ii) $R_t(x)=\mathbb{R}^n$ *for every* $x\in\mathbb{R}^n$, $t>0$

(iii) $V_E=\mathbb{R}^n$.

PROOF. First we check the case $x=0$.

(a) $R(0)=\mathbb{R}^n\Longleftrightarrow R_t(0)=\mathbb{R}^n$ for all $t>0\Longleftrightarrow V_E=\mathbb{R}^n$. From the Theorem 1, we have only to show that $R(0)=\mathbb{R}^n\Longrightarrow V_E=\mathbb{R}^n$. But this follows from the Corollary of the Lemma 2.

(b) (ii)$\Longleftrightarrow$(iii). Let

$$\dot y(t)=\left(-A-\sum_{i=1}^{m}u_i^*(t)B_i\right)y(t)-\sum_{j=1}^{r}v_j^*(t)d_j. \tag{5}$$

If $x(t)$ satisfies (3) then $y(t):=x(T-t)$, $0\le t\le T$ satisfies (5) with $u^*(t)==u(T-t)$, $v^*(t)=v(T-t)$ $0\le t\le T$ and conversely. Consequently the control (u,v) takes $x=x(0)$ to $y=x(T)$ if and only if (u^*,v^*) takes $y=y(0)$ to $x=y(T)$. now suppose $V_E=\mathbb{R}^n$. Denote $R^{(3)}$ resp. $R^{(5)}$ the reachability sets of the systems (3) resp. (5). By (a), $V_E=\mathbb{R}^n$ implies $R^{(5)}_{t/2}(0)=\mathbb{R}^n$ and then there exists a control (u,v) with $y(0)=0$, $y(t/2)=x$, i.e. the control

(u^*,v^*) takes $x(0)=x$ to $x(t/2)=0$. On the other hand $R^{(3)}_{t/2}(0)=\mathbb{R}^n$ implies the existence of (u',v') taking $x(0)=0$ to $x(t/2)=y$. Now let

$$(u''(s),v''(s)):=\begin{cases}(u^*(s),v^*(s)) & 0\le s\le \frac{t}{2}\\ (u'\left(s-\frac{t}{2}\right),v'\left(s-\frac{t}{2}\right)) & \frac{t}{2}<s\le t,\end{cases}$$

it takes $x(0)=x$ to $x(t)=y$, hence $R_t(x)=\mathbb{R}^n$ for all $x\in\mathbb{R}^n$ and $t>0$. Conversely the only condition $R_t(0)=\mathbb{R}^n$ implies $V_E=\mathbb{R}^n$ by (a).

(c) (i)$\Longleftrightarrow$(iii).

If $V_E=\mathbb{R}^n$ then $R_t(x)=\mathbb{R}^n$, hence $R(x)=\mathbb{R}^n$. Conversely $R(0)=\mathbb{R}^n$ implies $V_E=\mathbb{R}^n$ by (a).

Theorem 2 is proved. ∎

REMARK. 1. Using the enlargements technique we can prove the transitivity of the system (3) in the following way:

In the proof of the Theorem 1 we have seen that $V_E\subset \mathrm{LS}(\mathcal{F})(x)$, so if $V_E=\mathbb{R}^n$, then $\mathrm{LS}(\mathcal{F})(x)=\mathbb{R}^n$ for all x, and hence $\mathbb{A}_{\mathcal{F}}(t,x)=\mathbb{R}^n$ for all $t>0$ and all $x\in\mathbb{R}^n$. The sufficient condition is proved. The necessary condition follows from the fact that $\mathbb{A}_{\mathcal{F}}(t,0)=\mathbb{R}^n\Longrightarrow V_E=\mathbb{A}_{\mathcal{F}}(t,0)=\mathbb{R}^n$.

2. The assertion in the Theorem 2 of V. JURDJEVIC and I. KUPKA [30] that $\mathbb{A}_{\mathcal{F}}(t,x)$ contains the affin space $x+V_E$ is not true. From the proof it follows only that $\mathrm{cl}\,\mathbb{A}_{\mathcal{F}}(t,x)\supset x+V_E$. We give a simple example.

EXAMPLE: Consider the system satisfying all conditions of the Theorem 2 in [30]:

$$\dot{x}=Ax+ub.$$

Let A have at least two different nonzero real eigenvalues, say α and β, and a and b be the eigenvectors corresponding to α and β resp. Then $V_E=\{\mu b:\mu\in\mathbb{R}\}$. Consider the reachability set $\mathbb{A}(t,a)$. Assume that $a+V_E\subset\mathbb{A}(t,a)$. Then for all $\lambda\in\mathbb{R}$ there exist $u\in\mathcal{P}$:

$$a+\lambda b=e^{At}a+\int_0^t e^{A(t-s)}u(s)b\,\mathrm{d}s=e^{\alpha t}\cdot a+b\int_0^t e^{(t-s)\beta}u(s)\mathrm{d}s$$

i.e.

$$a+\lambda b=e^{\alpha t}\cdot a+k\cdot b,\qquad a(1-e^{\alpha t})=b(k-\lambda).$$

This happens only if $1=e^{\alpha t}$ and $k=\lambda$ since a and b are linearly independent. But $\alpha\ne 0$, $t\ne 0$, so $1\ne e^{\alpha t}$. The assumption is contradictory.

In what follows we investigate the special case of (3) when the B_i have rank 1, i.e. $B_i=b_ic_i^T$ for some $0\ne b_i$, $c_i\in\mathbb{R}^n$. Denote

$$B:=[b_1,\dots,b_m],\quad D:=[d_1,\dots,d_r],$$

$$\Gamma(A,B,D) := [B, AB, \ldots, A^{n-1}, D, AD, \ldots, A^{n-1}D]$$

(i.e. $B \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{n\times r}$, $\Gamma(A,B,D) \in \mathbb{R}^{n\times n(m+r)}$).

THEOREM 3. *Let*

$$\dot{x}(t) = \left(A + \sum_{i=1}^{m} u_i(t) b_i c_i^T\right) x(t) + \sum_{j=1}^{r} v_j(t) d_j \tag{6}$$

The following statements are equivalent

a) $R(x) = \mathbb{R}^n$ *for all* $x \in \mathbb{R}^n$

b) $R_t(x) = \mathbb{R}^n$ *for all* $x \in \mathbb{R}^n$, $t > 0$

c) (i) $\operatorname{rank}\Gamma(A,B,D) = n$ *and*

(ii) *For any* $1 \le i \le m$ *there exist* $1 \le j \le r$ *and* $l \ge 0$, $k_0, \ldots, k_l \ge 0$, $i_1, \ldots, i_l \in \{1, \ldots, m\}$ *such that* $c_i^T A^{k_0} b_{i_1} c_{i_1}^T A^{k_1} \ldots b_{i_l} c_{i_l}^T A^{k_l} d_j \ne 0$.

PROOF. By Theorem 2, a)$\Longleftrightarrow$b)$\Longleftrightarrow V_E = \mathbb{R}^n$. It remains to show that $V_E = \mathbb{R}^n \Longleftrightarrow$c). Suppose first $V_E = \mathbb{R}^n$ and prove c).

Since $c_{i_p}^T A^{k_p} b_{i_{p+1}} \in \mathbb{R}$, the elements of E are $A^{k_0} d_j$ or $cA^{k_0} b_{i_1}$, $c \in \mathbb{R}$. Here we can restrict ourselves to $0 \le k_0 \le n-1$, since A^{k_0}, $k_0 \ge n$ can be expressed as a linear combination of lower powers by the Cayley–Hamilton equality. So V_E is generated by the vectors $A^{k_0} d_j$, $A^{k_0} b_i$, $0 \le k_0 \le n-1$, $1 \le i \le m$, $1 \le j \le r$, and then $V_E = \mathbb{R}^n$ implies (i). For any $1 \le i \le m$ there exists $g \in E$ with $c_i^T g \ne 0$ (otherwise $c_i \perp V_E = \mathbb{R}^n$) which implies (ii). We proved that $V_E = \mathbb{R}^n \Longrightarrow$c). To prove the converse it is enough to show by (i) that $A^{k_0} d_j$, $A^{k_0} b_i \in V_E$. The first inclusion is trivial, and the second one follows from (ii) since for fixed i there exists an expression

$$\alpha = c_i^T A^{k_1} b_{i_2} c_{i_2}^T \ldots A^{k_{l-1}} b_{i_l} c_{i_l}^T A^{k_l} d_j \ne 0$$

and hence

$$\alpha \cdot A^{k_0} b_i = A^{k_0} b_i c_i^T A^{k_1} \ldots b_{i_l} c_{i_l}^T A^{k_l} d_j \in E.$$

The proof is complete. ∎

COROLLARIES OF THEOREMS 2 AND 3. 1. *The system*

$$\dot{x}(t) = \left(\sum_{i=1}^{m} u_i(t) B_i\right) x(t) + \sum_{j=1}^{r} v_j(t) d_j$$

is completely controllable in R^n, *i.e.* $R(x) = \mathbb{R}^n\ \forall x$, *if and only if the set*

$$\{B_{i_1} \ldots B_{i_l} d_j : l \ge 0,\ 1 \le j \le r,\ 1 \le i_1, \ldots, i_l \le m\}$$

generates linearly the whole $\mathbb{R}^n$.

2. *The system*

$$\dot{x}(t)=\sum_{i=1}^{m}u_i(t)b_ic_i^Tx(t)+\sum_{j=1}^{r}v_j(t)d_j$$

is completely controllable in $\mathbb{R}^n$ *if and only if*

a) $\operatorname{rank}[b_1,\dots,b_m,d_1,\dots,d_r]=n$ *and*

b) *for every* $1\le i\le m$ *there exist* $1\le j\le r$ *and* $1\le i_1,\dots,i_l\le m$ *such that*

$$c_i^Tb_{i_1}c_{i_1}^T\dots b_{i_l}c_{i_l}^Td_j\neq 0.$$

3. *The system*

$$\dot{x}(t)=(A+u(t)bc^T)x(t)+v(t)d$$

is completely controllable in $\mathbb{R}^n$ *if and only if*

a) $\operatorname{rank}[b,d,Ab,Ad,\dots,A^{n-1}b,A^{n-1}d]=n$ *and*

b) $\exists\ l: c^TA^ld\neq 0$.

3. The homogeneous-in-the-state case ($d_j=0\ \forall\ j$)

Next we take a further restriction on (3) namely that $A=0$, and $d_j=0$. In this case the solution starting from zero is identically zero, so we consider only $R_t(x)$, $x\neq 0$.

THEOREM 4. *For the system*

$$\dot{x}(t)=\sum_{i=1}^{m}u_i(t)b_ic_i^Tx(t) \tag{7}$$

the following statements are equivalent.

1° $R(x)=\mathbb{R}^n\setminus\{0\}$ *for every* $x\in\mathbb{R}^n\setminus\{0\}$

2° (i) $\operatorname{rank}B=\operatorname{rank}C=n$ ($B=[b_1,\dots,b_m]$, $C=[c_1,\dots,c_m]$) *and*

(ii) *For every* $1\le i,\ j\le m$ *there exist* $l\ge 0$, $1\le i_1,\dots,i_l\le m$ *such that*

$$c_i^Tb_{i_1}c_{i_1}^T\dots b_{i_l}c_{i_l}^Tb_j\neq 0.$$

PROOF. 1° $\Longrightarrow$ 2° By Theorem C, 1° $\Longleftrightarrow\mathscr{L}_0(x)=\mathbb{R}^n\ \forall x\in\mathbb{R}^n\setminus\{0\}$ and here $\mathscr{L}_0$ is the Lie algebra generated by B_i, $1\le i\le m$, since $A=0$. Any element of $\mathscr{L}_0(x)$ is a linear combination of members of type

$$b_{i_1}\left(c_{i_1}^Tb_{i_2}\right)\left(c_{i_2}^Tb_{i_3}\right)\cdots\left(c_{i_{k-1}}^Tb_{i_k}\right)\left(c_{i_k}^Tx\right)=\alpha b_{i_1}$$

hence $\mathscr{L}_0(x)$ is linearly generated by the vectors b_i, so $\operatorname{rank} B = n$. By the indirect assumption $\operatorname{rank} C < n$ it follows the existence of $x \neq 0$ with $c_i^T x = 0$ for all $i =$ and then $\mathscr{L}_0(x)$ would be $\{0\}$, so (i) is checked. We know $\mathscr{L}_0(b_j) = \mathbb{R}^n$ hence $c_i^T \mathscr{L}_0(b_j) \neq \{0\}$ which is the statement (ii).

$2° \Longrightarrow 1°$. It is enough to show that every b_i belongs to $\mathscr{L}_0(x)$, $x \neq 0$. For fixed $x \neq 0$, $\operatorname{rank} C = n$ implies the existence of c_j with $c_j^T x \neq 0$. Consequently $b_j c_j^T x \in \mathscr{L}_0(x)$, $b_j \in \mathscr{L}_0(x)$. Now take any other index i and a chain between c_i^T and b_j by (ii). We can suppose that l is the minimal length possible and use induction on l. If $l = 0$, i.e. $c_i^T b_j \neq 0$, then

$$\mathscr{L}_0(x) \ni [b_i c_i^T, b_j c_j^T] x = b_i (c_i^T b_j)(c_j^T x) - b_j (c_j^T b_i)(c_i^T x).$$

From $b_j \in \mathscr{L}_0(x)$ we have

$$\alpha b_i = b_i (c_i^T b_j)(c_j^T x) \in \mathscr{L}_0(x), \qquad \alpha \neq 0.$$

Now suppose that for minimal length $< l$ we are ready and take an index i with a sequence (ii). As we know

$$\begin{aligned} \mathscr{L}_0(x) \ni & \left[b_i c_i^T, \left[b_{i_1} c_{i_1}^T \left[\ldots \left[b_{i_l} c_{i_l}^T, b_j c_j^T\right] \ldots\right]\right]\right] x = \\ & = b_i c_i^T \left[b_{i_1} c_{i_1}^T, \ldots \left[b_{i_l} c_{i_l}^T, b_j c_j^T\right] \ldots\right] x - \\ & - \left[b_{i_1} c_{i_1}^T, \ldots, \left[b_{i_l} c_{i_l}^T, b_j c_j^T\right] \ldots\right] b_i c_i^T x = I_1 - I_2. \end{aligned}$$

Expand the Lie brackets in I_2. If $b_j c_j^T$ is the first left member, the summand belongs to $\mathscr{L}_0(x)$ since $b_j \in \mathscr{L}_0(x)$. If the product starts with $b_{i_k} c_{i_k}^T$, consider only the product from $b_{i_k} c_{i_k}^T$ to $b_j c_j^T$. If it is nonzero then by the induction hypothesis $b_{i_k} \in \mathscr{L}_0(x)$ and then the whole member belongs to $\mathscr{L}_0(x)$. Thus we proved that $I_2 \in \mathscr{L}_0(x)$, and then $I_1 \in \mathscr{L}_0(x)$. Expand the Lie brackets in I_1; if $b_j c_j^T$ is not the last factor, then the product from $b_i c_i^T$ until $b_j c_j^T$ must be zero (as any chain from c_i^T to b_j of length $< l$). Hence

$$I_1 = b_i (c_i^T b_{i_1})(c_{i_1}^T b_{i_2}) \ldots (c_{i_{l-1}}^T b_{i_l})(c_{i_l}^T b_j)(c_j^T x) = \alpha b_i, \quad \alpha \neq 0$$

so $b_i \in \mathscr{L}_0(x)$. Completing the induction we obtain $b_i \in \mathscr{L}_0(x)$ for all i whence $\mathscr{L}_0(x) = \mathbb{R}^n$. Theorem 4 is proved. ∎

EXAMPLES. The following examples show that the conditions $\operatorname{rank} B = n$, $\operatorname{rank} C = n$ and 2° (ii) of the Theorem 4 are independent of each other.

1. We consider the system (7) with $m = n = 2$

$$\dot{x} = \left(u_1(t) b_1 c_1^T + u_2(t) b_2 c_2^T\right) x(t). \tag{7'}$$

If we choose $b_1=c_1=\binom{1}{0}$, $b_2=c_2=\binom{0}{1}$, then the system can be written as

$$\dot{x}_1(t)=u_1(t)x_1(t)$$
$$\dot{x}_2(t)=u_2(t)x_2(t).$$

It is clear that $\operatorname{rank}B=\operatorname{rank}C=2$ and the condition (i) is satisfied but (ii) is not. Our system is not completely controllable on $\mathbb{R}^2\setminus\{0\}$, for example, one can't steer the state $\binom{1}{0}$ into the state $\binom{0}{1}$.

2. Consider the system (7′) with $b_1=\binom{1}{0}$, $b_2=\binom{0}{1}$, $c_1=c_2=\binom{1}{1}$. Then $\operatorname{rank}B=2$, the condition (ii) is satisfied, but $\operatorname{rank}C=1$. In this case the system (7′) can be written as

$$\dot{x}_1(t)=u_1(t)(x_1(t)+x_2(t))$$
$$\dot{x}_2(t)=u_2(t)(x_1(t)+x_2(t)).$$

Hence $(\dot{x}_1(t)+\dot{x}_2(t))=(u_1(t)+u_2(t))(x_1(t)+x_2(t))$. Putting $y(t)=x_1(t)+$ $+x_2(t)$ we have $\dot{y}(t)=(u_1(t)+u_2(t)y(t)$. The reachable set from 0 for this system is $\{0\}$. Thus if we choose $(x_1(0)+x_2(0))=0$ then for any t, $x_1(t)+x_2(t)=0$, this means that our system is not completely controllable on $\mathbb{R}^2\setminus\{0\}$.

3. If in the system (7′) we take $b_1=b_2=\binom{1}{1}$, $c_1=\binom{1}{0}$, $c_2=\binom{0}{1}$ then we have

$$\dot{x}_1(t)=u_1(t)x_1(t)+u_2(t)x_2(t)$$
$$\dot{x}_2(t)=u_1(t)x_1(t)+u_2(t)x_2(t).$$

For this system $\operatorname{rank}C=2$, the condition (ii) is satisfied, but $\operatorname{rank}B=1$. Putting $y(t)=x_1(t)-x_2(t)$ we have $\dot{y}(t)=0$. Hence $y(t)=y(0)=x_1(0)-x_2(0)$. This means that our system is not completely controllable on $\mathbb{R}^2\setminus\{0\}$.

Consider again the system

$$\dot{x}(t)=\left(A+\sum_{i=1}^{m}u_i(t)B_i\right)x(t) \tag{2}$$

with piecewise continuous controls $u_i(t)$.

DEFINITION ([19], p.18). The system (2) is *strongly accessible in small time* from $x\neq 0$ if for every $t>0$ the reachability set $R_t(x)$ has nonempty interior in $\mathbb{R}^n\setminus\{0\}$.

THEOREM E ([8], Corollary 4.7). *The system (2) is strongly accessible in small time from $x\neq 0$ if and only if $\mathcal{L}_0(x)=\mathbb{R}^n$. Here*

$$\mathcal{L}_0:=\{ad_A^k B_i : i=1,\dots,m;\ k\geq 0\}_{\mathrm{LA}}.$$

We shall restrict ourselves to the case $\operatorname{rank}B_i=1$, $B_i=b_ic_i^T$.

THEOREM 5. *In case $B_i = b_i c_i^T$ the system (2) is strongly accessible in small time from any $x \neq 0$ if and only if the following three conditions hold*

(i) $\operatorname{rank}[B, AB, \ldots, A^{n-1}B] = n$, *where* $B := [b_1, \ldots, b_m]$.

(ii) $\operatorname{rank}[C, A^T C, \ldots, (A^{n-1})^T C]^T = n$, *where* $C := [c_1, \ldots, c_m]$.

(iii) *For every* $1 \le i, j \le m$ *there exist* $l \ge 1$, $1 \le i_1, \ldots, i_{l-1} \le m$ *and* $k_0 \ge 0$, $\ldots$, $k_{l-1} \ge 0$ *such that*

$$c_i^T A^{k_0} b_{i_1} c_{i_1}^T A^{k_1} \ldots b_{i_{l-1}} c_{i_{l-1}}^T A^{k_{l-1}} b_j \neq 0.$$

PROOF. By Theorem E we have to show that $\mathcal{L}_0(x) = \mathbb{R}^n \Longleftrightarrow$(i), (ii), (iii).

a) $\mathcal{L}_0(x) = \mathbb{R}^n \Longrightarrow$(i).

Since any element of $\mathcal{L}_0$ is a polynomial in the variables A, B_1, $\ldots$, B_m where there is no monomial A^k, $k \ge 0$, hence any element of $\mathcal{L}_0(x)$ is a finite sum

$$y = \sum \alpha_{ik} A^k b_i.$$

Consequently the vectors $b_i, Ab_i, \ldots, A^{n-1}b_i$, $i = 1, \ldots, m$, span $\mathcal{L}_0(x) = \mathbb{R}^n$ whence (i) (we used the Cayley–Hamilton theorem for A).

b) $\mathcal{L}_0(x) = \mathbb{R}^n \quad \forall x \neq 0 \Longrightarrow$(ii).

The set

$$W := \{x \in \mathbb{R}^n : c_i^T A^k x = 0 \quad \text{for all} \quad 1 \le i \le m,\ 0 \le k \le n-1\}$$

is a subspace of $\mathbb{R}^n$ and $W \neq \{0\}$ if and only if (ii) does not hold. Suppose indirectly that there exists $0 \neq x \in W$. In $\mathcal{L}_0$ any element is a polynomial of $A, B_1, \ldots, B_m$ with no monomials A^k, $k \ge 0$, hence any monomial has the form

$$D b_i c_i^T A^k, \qquad k \ge 0$$

where D is some product of A and the B_j's. Since $D b_i c_i^T A^k x = 0$, we get $\mathcal{L}_0(x) = \{0\}$ contrarily to what is supposed.

c) $\mathcal{L}_0(x) = \mathbb{R}^n\ \forall x \neq 0 \Longrightarrow$(iii).

Let $1 \le i, j \le m$. By $\mathcal{L}_0(b_j) = \mathbb{R}^n$ there exists $L \in \mathcal{L}_0$ with $c_i^T L b_j \neq 0$. Here Lb_j is a finite sum

$$Lb_j = \sum \alpha_{i_1, \ldots, i_{l-1}, k_0, \ldots, k_{l-1}} A^{k_0} b_{i_1} c_{i_1}^T A^{k_1} \ldots b_{i_{l-1}} c_{i_{l-1}}^T A^{k_{l-1}} b_j$$

hence there exists a summand for which

$$c_i^T A^{k_0} b_{i_1} c_{i_1}^T \ldots A^{k_{l-1}} b_j \neq 0$$

whence (iii).

d) (i), (ii), (iii)$\Longrightarrow \mathcal{L}_0(x)=\mathbb{R}^n\ \forall x\neq 0$.

We shall show that

$$A^k b_i\in\mathcal{L}_0(x)\qquad \forall x\neq 0,\quad k\geq 0,\ 1\leq i\leq m,$$

it implies $\mathcal{L}_0(x)=\mathbb{R}^n$ by (i).

From (ii) it follows the existence of $1\leq j\leq m$, $0\leq k\leq n-1$ with $c_j^T A^k x\neq 0$.

We can suppose that (for fixed x) the number k is the smallest for which there exists such an index j. Then

$$\left(ad_A^k B_j\right)x=\sum_{s=0}^{k}(-1)^s\binom{k}{s}A^{k-s}b_jc_j^TA^sx=(-1)^kb_jc_j^TA^kx\in\mathcal{L}_0(x)$$

i.e. $b_j\in\mathcal{L}_0(x)$. We can show by induction on p that $A^pb_j\in\mathcal{L}_0(x)$. Suppose this for $0,\dots,p-1$. As we know

$$\left(ad_A^{k+p}b_jc_j^T\right)x=\sum_{s=0}^{k+p}(-1)^s\binom{k+p}{s}A^{k+p-s}b_jc_j^TA^sx\in\mathcal{L}_0(x).$$

Here the members $s<k$ vanish since $c_j^TA^sx=0$. The members $s>k$ belong to $\mathcal{L}_0(x)$ because $A^{k+p-s}b_j\in\mathcal{L}_0(x)$ by the inductional hypothesis. Consequently $A^pb_jc_j^TA^kx\in\mathcal{L}_0(x)$, $A^pb_j\in\mathcal{L}_0(x)$. Now we shall prove that $A^pb_i\in\mathcal{L}_0(x)$ for arbitrary other index i. Take a product of (iii) of the form

$$c_i^TA^{k_0}b_{i_1}c_{i_1}^TA^{k_1}\dots b_{i_{l-1}}c_{i_{l-1}}^TA^{k_{l-1}}b_j\neq 0.$$

We can suppose that the length $l\geq 1$ is minimal (for i, j fixed) and that for fixed $b_{i_1},\dots,b_{i_{l-1}}$ the number k_0 is minimal for which $c_i^TA^{k_0}b_{i_1}\neq 0$, $\dots,k_{l-2}$ is minimal for which $c_{i_{l-2}}^TA^{k_{l-2}}b_{i_{l-1}}\neq 0$, k_{l-1} is minimal for which $c_{i_{l-1}}^TA^{k_{l-1}}b_j\neq 0$. Now $A^pb_i\in\mathcal{L}_0(x)$ will be checked by induction on l.

d$_1$) $l=1$.

This means that $c_i^TA^{k_0}b_j\neq 0$ and $c_i^TA^rb_j=0$ $(r<k_0)$. Now

$$\mathcal{L}_0(x)\ni\left[ad_A^kb_jc_j^T,ad_A^{k_0}b_ic_i^T\right]x=$$

$$=\left[\sum_{s=0}^{k}(-1)^s\binom{k}{s}A^{k-s}b_jc_j^TA^s,\sum_{r=0}^{k_0}(-1)^r\binom{k_0}{r}A^{k_0-r}b_ic_i^TA^r\right]x=$$

$$=\sum_s\sum_r(-1)^{r+s}\binom{k}{s}\binom{k_0}{r}A^{k-s}b_j\left(c_j^TA^{s+k_0-r}b_i\right)\left(c_i^TA^rx\right)-$$

$$-\sum_s\sum_r(-1)^{r+s}\binom{k}{s}\binom{k_0}{r}A^{k_0-r}b_i\left(c_i^TA^{r+k-s}b_j\right)\left(c_j^TA^sx\right)=:I_1-I_2.$$

In I_1 every member is a constant multiple of some $A^{k-s}b_j$. hence $I_1\in\mathcal{L}_0(x)$. In I_2 $c_j^TA^sx\neq 0$ only for $s=k$ and then $c_i^TA^{r+k-s}b_j=c_i^TA^rb_j\neq 0$ only for $r=k_0$. Thus $I_2=(-1)^{k+k_0}b_i(c_i^TA^{k_0}b_j)(c_j^TA^kx)$ whence $b_i\in\mathcal{L}_0(x)$. Then $A^pb_i\in\mathcal{L}_0(x)$ follows by induction on p expanding $[ad_A^kb_jc_j^T,ad_A^{k_0+p}b_ic_i^T]x$ just as above.

d$_2$) $l>1$.

Suppose that the statement is proved for $1,\dots,l-1$. Consider the product of (iii) with the above listed minimality properties. Its subproducts are also nonvanishing

$$c_{i_s}^TA^{k_s}b_{i_{s+1}}\dots b_{i_{l-1}}c_{i_{l-1}}^TA^{k_1}b_j\neq 0$$

i.e. the chain between $c_{i_s}^T$ and b_j is shorter than l. By the induction hypothesis

$$A^pb_{i_s}\in\mathcal{L}_0(x)\qquad s=1,\dots,l-1;\quad p\geq 0.$$

Consider the following element of $\mathcal{L}_0(x)$:

$$\left[ad_A^{k_0}b_ic_i^T,\left[ad_A^{k_1}b_{i_1}c_{i_1}^T,\dots,\left[ad_A^{k_{l-1}}b_{i_{l-1}}c_{i_{l-1}}^T,ad_A^kb_jc_j^T\right]\dots\right]\right]x=$$
$$=ad_A^{k_0}b_ic_i^T\left[ad_A^{k_1}b_{i_1}c_{i_1}^T,\dots,\left[ad_A^{k_{l-1}}b_{i_{l-1}}c_{i_{l-1}},ad_A^kb_jc_j^T\right]\dots\right]x-$$
$$-\left[ad_A^{k_1}b_{i_1}c_{i_1}^T,\dots,\left[ad_A^{k_1}b_{i_{l-1}}c_{i_{l-1}}^T,ad_A^kb_jc_j^T\right]\dots\right]ad_A^{k_0}b_ic_i^Tx=:I_3-I_4.$$

In I_4 we expand all Lie brackets and expressions ad_A^k; we get a sum of products and each product begins with some $A^sb_{i_r}$ or A^sb_j. Since these belong to $\mathcal{L}_0(x)$, we get $I_4\in\mathcal{L}_0(x)$. Now expand the written Lie brackets in I_3. If there exists a bracket where we take the opposite ordered product of the components, we get chains between c_i^T and b_j of length $<l$ which are all zero. Consequently

$$I_3=ad_A^{k_0}\left(b_ic_i^T\right)ad_A^{k_1}\left(b_{i_1}c_{i_1}^T\right)\dots ad_A^{k_{l-1}}\left(b_{i_{l-1}}c_{i_{l-1}}^T\right)ad_A^k\left(b_jc_j^T\right)x\in$$
$$\in\mathcal{L}_0(x).$$

Finally expand all expressions $ad_A^{k_r}$, ad_A^k, we get a sum of products where the total A-power is $A^{k_0+\dots+k_{l-1}+k}$. Using the minimality property of $k_0,\dots,k_{l-1}$ and k we see that

$$I_3=b_i\left(c_i^TA^{k_0}b_{i_1}\right)\left(c_{i_1}^TA^{k_1}b_{i_2}\right)\dots\left(c_{i_{l-2}}^TA^{k_{l-2}}b_{i_{l-1}}\right)\cdot$$

$$\cdot\left(c_{i_{l-1}}^T A^{k_l-1} b_j\right)\left(c_j^T A^k x\right)=\alpha b_i, \quad \alpha\neq 0$$

whence $b_i \in \mathcal{L}_0(x)$. Finally $A^p b_i \in \mathcal{L}_0(x)$ can be checked by induction on p as above.

Theorem 5 is completely proved. ∎

REMARK. The chain condition (iii) can be checked by the following simple algorithm. Let

$$A\in\mathbb{R}^{n\times n}, \qquad 0\neq b_i,\ c_i\in\mathbb{R}^n$$

be arbitrary. We have to know for any pair $i, j\in\{1,\dots,m\}$ whether there exists a chain

$$c_i^T A^{k_0} b_{i_1} c_{i_1}^T \dots b_{i_{l-1}} c_{i_{l-1}} A^{k_l-1} b_j \neq 0$$

(in symbols $i \Longrightarrow j$) or not. The algorithm goes as follows.

First step. Introduce the notation for p, $q\in\{1,\dots,m\}$

$$p\rightarrow q: \quad \Longleftrightarrow \text{there exists } r\geq 0 \text{ with } c_p^T A^r b_q\neq 0.$$

To decide whether $p\rightarrow q$ or not, it is enough to compute A^r, $0\leq r\leq n-1$ by the Cayley–Hamilton theorem. Denote

$$I_p:=\{q\in\{1,\dots,m\}: p\rightarrow q\} \qquad J_q:=\{p\in\{1,\dots,m\}: p\rightarrow q\}.$$

Second step. We compute the sets I_i^l, J_j^l as follows. Let $I_i^1:=I_i$, $J_j^1:=J_j$ and we define recursively the sets

$$I_i^l:=\{q\in\{1,\dots,m\}:\exists\, p\in I_i^{l-1} \text{ with } p\rightarrow q\},$$
$$J_j^l:=\{p\in\{1,\dots,m\}:\exists\, q\in J_j^{l-1} \text{ with } p\rightarrow q\}.$$

a) If we have computed the sets I_i^r, J_j^s for $0\leq s$, $r\leq l$ and for some r, s we have $I_i^r\cap J_j^s\neq\emptyset$ then we conclude $i\Longrightarrow j$.

b) If there exists an $1\leq r\leq l-1$ with $I_i^l\subset I_i^r$ then $I_i^l=I_i^{l+1}=I_i^{l+2}=\dots$

c) If not, then we compute I_i^{l+1}.

We make analogous investigations a'), b'), c') related to the sets J_j^s. This algorithm decides after finitely many steps whether $i\Longrightarrow j$. The number of steps is of exponential order in the parameter m. We do not know, whether there exists an algorithm with polynomial number of steps.

EXAMPLES (see [23] p.62).

1. Consider the system

$$\begin{pmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\end{pmatrix}=\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}x_1(t)\\x_2(t)\end{pmatrix}+u(t)\begin{pmatrix}0&0\\0&1\end{pmatrix}\begin{pmatrix}x_1(t)\\x_2(t)\end{pmatrix}$$

which has the form $\dot{x} = Ax + ubc^T x$.

For this system the conditions (i), (ii), (iii) of Theorem 5 are satisfied, thus it is strongly accessible in small time from any $x \neq 0$. This system is not controllable on $\mathbb{R}^2 \setminus \{0\}$ (one can see that the set $\{x \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x \neq 0\}$ is invariant). This means that the conditions (i), (ii), (iii) of Theorem 5 are not enough for the controllability of the system (3) in the case $\operatorname{rank} B_i = 1$.

2. Consider the system

$$\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} + u(t) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}.$$

This system is controllable on $\mathbb{R}^2 \setminus \{0\}$, i.e. $R(x) = \mathbb{R}^2 \setminus \{0\}$ $\forall x \neq 0$ (By Theorem 1 in [46]).

In [38], D. H. DAO and al. showed that the system

$$x_{k+1} = \left(A + \sum u_k^i b_i c_i^T\right) x_k \quad \text{with} \quad \det A \neq 0$$

is controllable in $\mathbb{R}^n \setminus \{0\}$ if and only if the conditions (i), (ii) of Theorem 5 and the conditions

(iii') There exists $\delta > 0$ such that the condition (iii) of Theorem 5 is satisfied with $k_0 + k_1 + \ldots + k_{l-1} = \delta$.

(iv)Let $J = \{k : C^T A^k B \neq 0,\ 0 < k \leq n^2\}$ then $\sigma =$ g.c.d. of $J = 1$ holds.

For the discrete version of Example 2., $J = \{2,4\}$, thus $\sigma = 2$. The condition (iv) is not satisfied. Consequently in our case (for the continuous systems) the condition (iv) is not necessary. The continuous version is controllable, the discrete one is not. This means that in this sense the discrete systems are not more general than the continuous ones.

4. Controllability of Linear Systems with Output Feedback

It seems that the case of $\operatorname{rank} B_i = 1$ happens rarely and is not natural because of its instability (with a small change of the elements of B_i the conditions $\operatorname{rank} B_i = 1$ may become fail). Here we want to show (partly) that the case of $\operatorname{rank} B_i = 1$ is very interesting and important. It is known that the study of bilinear systems with $\operatorname{rank} B_i = 1$ came from the linear regular problem with output feedback (see [38]–[42] for discrete case).

Consider the linear system defined by the following equations:

$$(8) \qquad \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) + Dv(t) \\ y(t) = C^T x(t) \end{cases}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $v(t) \in \mathbb{R}^r$, $y(t) \in \mathbb{R}^q$ are the values of state, control and output, respectively. Here the controls are assumed to be piecewise continuous functions. A, B, D, C are real constant matrices of order $n\times n$, $n\times p$, $n\times r$ and $n\times q$, respectively.

We assume that the control $u(t)$ is defined by an output feedback mechanism given by the equation

$$u(t)=U(t)y(t) \tag{9}$$

$U(t)$ is $p\times q$-matrix control.

From the equations (8)–(9) we obtain the following system

$$\dot{x}(t)=(A+BU(t)C^T)x(t)+Dv(t) \tag{10}$$

which can be written as

$$\dot{x}(t)=Ax(t)+\sum_{kl} u_{kl}(t)b_k c_l^T x(t)+\sum_{j=1}^{r} d_j v_j(t), \qquad \text{or}$$

$$\dot{x}(t)=Ax(t)+\sum_{i=1}^{m} u_i(t)\bar{b}_i\bar{c}_i^T x(t)+\sum_{j=1}^{r} d_j v_j(t). \tag{11}$$

Here $m=p\times q$, $\bar{b}_i=b_\alpha$, $\bar{c}_i=c_\beta$, $u_i(t)=u_{\alpha\beta}(t)$ with $i=(\alpha-1)q+\beta$ ($\alpha=1,\dots,p$, $\beta=1,\dots,q$).

One can see that the system (11) here is just the system (6) in the Theorem 3, and the system (11) in case of $D=0$ is just the system mentioned in the Theorem 5.

The question is to find the conditions which insure the controllability, the accessibility, the stability or the pole placement, ... of the linear systems with output feedback defined by the equations (1)–(2). Applying the obtained results for the bilinear systems with $\operatorname{rank} B_i=1$ (Theorem 3 and 5) we give here only some results for the controllability and the accessibility of the linear systems with output feedback mentioned above. The stability and a more detailed study in this area will be considered in a later paper.

First we need a lemma:

LEMMA 5 (see [39]). 1. *Let* $\overline{B}=[\bar{b}_1,\dots,\bar{b}_m]$, $\overline{C}=[\bar{c}_1,\dots\bar{c}_m]$ *then*

$$\operatorname{rank}[\overline{B},A\overline{C},\dots,A^{n-1}\overline{B}]=\operatorname{rank}[B,AB,\dots,A^{n-1}B]$$

$$\operatorname{rank}[\overline{C},A^T\overline{C},\dots,(A^T)^{n-1}\overline{C}]^T=\operatorname{rank}[C,A^TC,\dots,(A^T)^{n-1}C]^T$$

$$\operatorname{rank}[\overline{B},A\overline{B},\dots,A^{n-1}\overline{B};D,AD,\dots,A^{n-1}D]=$$
$$=\operatorname{rank}[B,AB,\dots,A^{n-1}B;D,\dots,A^{n-1}D].$$

2. *For any* i, $j\in\{1,\dots,m\}$ *there exists* $k\in\{1,\dots,m\}$ *such that* $\bar{b}_i=\bar{b}_k$, $\bar{c}_j=\bar{c}_k$.

PROOF. 1. By definition of $\bar{b}_i$ and $\bar{c}_j$ we see that $\overline{B}$ and $\overline{C}$ are obtained from B and C by repeating every column q times and p times, respectively. From this we get the result.

2. We have

a) $\bar{b}_i = \bar{b}_j$ if and only if $\left[\frac{i-1}{q}\right] = \left[\frac{j-1}{q}\right]$.

b) $\bar{c}_i = \bar{c}_j$ if and only if $i \equiv j \pmod{q}$.

If $i = (\alpha - 1)q + \beta$, $j = (\alpha' - 1)q + \beta'$ then let $k = (\alpha - 1)q + \beta'$. ∎

THEOREM 6. *The linear system with output feedback* (8)–(9) *is completely controllable if and only if the following conditions are satisfied*

(a) $\operatorname{rank}[B, AB, \ldots, A^{n-1}B, D, AD, \ldots, A^{n-1}D] = n$

(b) $\exists\, l \in \{1, \ldots, q\}$, $\exists\, j \in \{1, \ldots, r\}$, $\exists\, k \geq 0$ *such that* $c_l A^k d_j \neq 0$.

PROOF. The system (8)–(9) is equivalent to the system (11), and by Theorem 3, it is controllable of and only if

(i) $\operatorname{rank}\Gamma(A, \overline{B}, D) = n$

(ii) for any $1 \leq i \leq m$, there exist $1 \leq j \leq r$, $l \geq 0$, $k_0, \ldots, k_l \geq 0$, $i_1, \ldots, i_l \in$ $\in \{1, \ldots, m\}$ such that

$$\bar{c}_i^T A^{k_0} \bar{b}_{i_1} \bar{c}_{i_1}^T A^{k_1} \ldots \bar{b}_{i_l} \bar{c}_{i_l}^T A^{k_l} d_j \neq 0.$$

By the Lemma 5, the conditions (a) and (i) are equivalent.

Necessity: If the system (8)–(9) is completely controllable we have the conditions (i) and (ii). The condition (a) is obviously true. From the (ii), $c_{i_l}^T A^{k_l} d_j \neq 0$ and $\bar{c}_{i_l} = c_\beta$ for some $\beta \in \{1, \ldots, q\}$ hence the condition (b) is true.

Sufficiency: If the conditions (a) and (b) are true, we have immediately (i). Now let $i \in \{1, \ldots, m\}$, since

$$\operatorname{rank}\{B, AB, \ldots, A^{n-1}B; D, AD, \ldots, A^{n-1}D] = n, \qquad \text{we have}$$

$$\bar{c}_i^T A^\lambda b_\alpha \neq 0 \quad \text{for some } \lambda \text{ and } \alpha \quad \text{or} \quad \bar{c}_i^T A^\mu d_\beta \neq 0 \quad \text{for some } \mu \text{ and } \beta.$$

In the second case we are ready. In the first case we have $\bar{c}_i^T A^\lambda b_\alpha c_l^T A^k d_j \neq$ $\neq 0$. By the Lemma 5, there exists $\gamma \in \{1, \ldots, m\}$ such that $b_\alpha = \bar{b}_\gamma$, $c_l = \bar{c}_\gamma$, and from this

$$\bar{c}_i^T A^\lambda \bar{b}_\gamma \bar{c}_\gamma^T A^k d_j \neq 0$$

that is the condition (ii) is proved. We have proved the Theorem. ∎

In what follows we consider the case $D = 0$. Because of the incompleteness of the Theorem 5 we obtain only the result for the accessibility of the system. Besides we show that the multidimensional control $U(t)$ can be replaced by one dimensional control.

THEOREM 7. *For the linear system with output feedback*

$$\dot{x}(t)=Ax(t)+Bu(t),\quad y(t)=C^Tx(t),\quad u(t)=U(t)y(t) \tag{12}$$

the following statements are equivalent:

1. *The system has strong accessibility property in small time on* $\mathbb{R}^n\setminus\{0\}$

2. (a) $\operatorname{rank}[B,AB,\dots,A^{n-1}B]=n$.

   (b) $\operatorname{rank}[C,A^TC,\dots,(A^T)^{n-1}C]^T=n$

In this case its feedback matrix $U(t)$ *can be chosen such that it depends linearly on a* $\mathbb{R}\to\mathbb{R}$ *function.*

PROOF. We need the following result (see [43], Theorem 1).
$(*)$ *Assume that the condition 2. above is satisfied then there is a vector* $b=Bg$ *and a vector* $c=Ch$, *a matrix* K *such that for* $\overline{A}=A+BKC^T$ *we have*

$$\operatorname{rank}[b,\overline{A}b,\dots,\overline{A}^{n-1}b]=n \qquad \operatorname{rank}[c,\overline{A}^Tc,\dots,(\overline{A}^T)^{n-1}c]^T=n.$$

Now the system (12) is equivalent to the system

$$\dot{x}(t)=Ax(t)+\sum_{i=1}^{m}u_i(t)\bar{b}_i\bar{c}_i^Tx(t). \tag{13}$$

By the Theorem 5, the system (12) has strong accessibility property if and only if the conditions (i), (ii), (iii) in the Theorem 5 are satisfied for A, $\bar{b}_i$, $\bar{c}_i$, $i=1,\dots,m$.

Necessity: Assume that the conditions (i), (ii), (iii) in Theorem 5 are true, then by the Lemma 5 the condition 2. is satisfied.

Sufficiency: Assume that the condition 2. is true, then by the Lemma 5, the conditions (i) and (ii) in Theorem 5 are satisfied for system (13). We have only to prove the condition (iii). Let $i,j\in\{1,\dots,m\}$. From the condition (2) there are α, λ and β, μ such that $\bar{c}_i^TA^\lambda\bar{b}_\alpha\neq0$, $\bar{c}_\beta^TA^\mu\bar{b}_j\neq0$, hence $\bar{c}_i^TA^\lambda\bar{b}_\alpha^T\bar{c}_\beta^TA^\mu\bar{b}_j\neq0$. By the Lemma 5 there is an index l such that $\bar{b}_\alpha=\bar{b}_l$, $\bar{c}_\beta=\bar{c}_l$ that is $\bar{c}_i^TA^\lambda\bar{b}_l\bar{c}_l^TA^\mu\bar{b}_j\neq0$, which proves the truth of the condition (iii). By the Theorem 5 the system has strong accessibility in small time on $\mathbb{R}^n\setminus\{0\}$.

Now assume that the system has strong accessibility property in small time, that is the condition 2) is satisfied. By $(*)$, there is a matrix K, a vector $b=Bg$ and a vector $c=Ch$ such that for $\overline{A}=A+BKC^T$ we have

$$\operatorname{rank}[b,\overline{A}b,\dots,\overline{A}^{n-1}b]=n \qquad \operatorname{rank}[c,\overline{A}^Tc,\dots,(\overline{A}^T)^{n-1}c]^T=n$$

By the proved result, the system

$$\dot{x} = (A + BKC^T)x(t) + bu(t)c^T x(t)$$

has strong accessibility property in small time. It can be written as

$$\dot{x}(t) = Ax(t) + BKC^T x(t) + u(t)Bgh^T C^T x(t)$$

$$\text{or} \quad \dot{x}(t) = Ax(t) + B\{K + u(t)gh^T\}C^T x(t)$$

From this one can see that the output feedback matrix $U(t)$ can be chosen as

$$U(t) = K + u(t)gh^T$$

which depends on only one parameter $u(t)$ and the system has still strong accessibility property in small time. ∎

References

[1] R. Kalman, P. Falb, M. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill, New York, 1969.

[2] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, New York, 1976.

[3] A. A. Sagle and R. E. Walde, *Introduction to Lie Groups and Lie Algebras*, Academic Press, New York and London, 1973.

[4] J. Kučera, Solution in Large of Control Problem $\dot{x} = (A(1-u) + Bu)x$, *Czech. Math. J.*, **16** (1966), 600–623.

[5] J. Kučera, Solution in Large of Control Problem $\dot{x} = (Au + Bv)x$, *Czech. Math. J.*, **17** (1967), 91–96.

[6] G. W. Haynes and H. Hermes, Nonlinear Controllability vie Lie Theory, *SIAM J. Control Optim.*, **8** (1970).

[7] C. Lobry, Contrôlabilité des Systèmes non Linéaires, *SIAM J. Control Optim.*, **8**, (1970), 573–605.

[8] H. J. Sussmann and V. Jurdjevic, Controllability of Nonlinear Systems, *J. Diff. Equats.*, **12** (1972), 96–116.

[9] V. Jurdjevic and H. J. Sussmann, Control Systems on Lie Groups, *J. Diff. Equats.*, **12** (1972), 313–329.

[10] D. L. Lukes, Global Controllability of nonlinear Systems, *SIAM J. Control Optim.*, **10** (1972), 112–126.

[11] R. W. Brockett, System Theory on Group manifold and Coset Space, *SIAM J. Control Optim.*, **10** (1972), 265–284.

[12] R. W. Brockett, Nonlinear Systems and Differential Geometry, *Proceedings of The IEEE*, **64** (1976), 61–72.

[13] C. Bruni, G. Di Pillo and G. Koch, Bilinear Systems: An Appealing Class of "Nearly Linear" Systems in Theory and Applications, *IEEE Trans. on Aut. Control*, AC **19**, (1974), 334–348.

[14] R. E. Rink and R. R. Mohler, Completely controllable bilinear systems, *SIAM J. Control Optim.*, **6**, (1968).

[15] R. M. Hirschon, Controllability in Nonlinear Systems, *J. Diff. Equats*, **19** (1975), 46–61.

[16] R. M. Hirschon, Global Controllability of Nonlinear Systems, *SIAM J. Control Optim.*, **14** (1976), 700–711.

[17] A. Kumpera, Remarks on Paper by R. M. Hirschon, *SIAM J. Control Optim.*, **26** (1988), 465–467.

[18] R. Hermann and A. J. Krener, Nonlinear Controllability and Observability, *IEEE Trans Aut. Control*, AC **22** (1977), 728–740.

[19] H. J. Sussmann, Lie Brackets, Real Analyticity and Geometric Control, in *Differential Geometric Control Theory*, R. W. Brockett, R. S. Millmann, H. J. Sussmann eds., Birkhäuser, Boston–Basel–Stuttgart, 1983, 1–116.

[20] J. B. Gonçales, Sufficient conditions for local controllability with unbounded control, *SIAM J. Control Optim.*, **25** (1987), 1371–1378.

[21] J. P. Gauthier et G. Bornard, Contrôlabilité des systèmes bilinéaries, *SIAM J. Control Optim.*, **20** (1982), 377–384.

[22] R. W. Brockett, On the algebraic structure of Bilinear Systems in *Variable Structure Control Systems*, R. Mohler and A. Ruberti eds., Academic Press, 1972, 153–168.

[23] R. W. Brockett, Lie algebras and Lie groups in control theory, in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett Eds., Reidel, Dordrecht, 1973, 43–82.

[24] V. V. Amel'kin, *Differential equations in applications*, Science for everyone, Mir, Moscow, 1990.

[25] E. Hille, R. S. Phillips, Functional Analysis and Semi-Groups, *AMS Coll. Publ., XXXI,* Providence, 1957.

[26] S. G. Krein, *Linear differential equations in Banach spaces* (in Russian), Nauka, Moscow, 1967.

[27] R., Hermann, On the Accessibility Problem in Control Theory in *Nonlinear Differential Equations and Nonlinear Mechanics* J. P. LaSalle and S. Lefschetz eds., Academic Press, N. Y., 1963.

[28] V. Jurdjevic and I. Kupka, Control Systems Subordinated to a Group Action: Accessibility, *J. Diff. Equats.*, **39** (1981), 186–211.

[29] V. Jurdjevic and I. Kupka, Control Systems on Semi-simple Lie Groups and their Homogeneous Spaces, *Ann. Inst. Fourier*, Grenoble **31** (1981), 151–179.

[30] V. Jurdjevic and I. Kupka, Polynomial Control Systems, *Math. Ann.*, **272** (1985), 361–368.

[31] G. SALLET, Extension Techniques, in *Systems and Control Encyclopedia*, (1987) II. 1581–1583.

[32] V. JURDJEVIC and SALLET, Controllability properties of affine Systems, *SIAM J. Control and Optim.*, **22** (1984), 501–508.

[33] P. BRUNOVSKÝ, Local Controllability of odd Systems, *Banach Center Publications*, Vol 1, Warszawa, 1976. 39–44.

[34] U. PIECHOTTKA and P. M. FRANK, Controllability of Bilinear Systems: A survey and some new results, in *Nonlinear Control Systems design*, ed. A. Isidori, Pergamon Press, 1990, 23–28.

[35] B. JAKUBCZYK and E. D. SONTAG, Controllability of Nonlinear Discrete-time Systems: A Lie-algebraic approach, *SIAM J. Control Optim.*, **28** (1990), 1–33.

[36] R. M. HIRSCHON, Strong Controllability of Nonlinear Systems, *SIAM J. Control Optim.*, **24** (1986), 264–275.

[37] H. J. SUSSMANN, A general Theorem on Local Controllability, *SIAM J. Contr. Optim.*, **25** (1987), 158–194.

[38] D. H. DAO, H. M. HAI, PH. D. DIEU, Necessary and Sufficient Conditions for the Controllability of a Class of Discrete Bilinear Systems, *Syst. Anal. Model. Simul.*, **2** (1985), 3–17.

[39] D. H. DAO, H. M. HAI, PH. D. DIEU, Controllability of Discrete Linear Systems with Output Feedback, *Syst. Anal. Model. Simul.*, **2** (1985), 121–127.

[40] M. E. EVANS and D. N. P. MURTHY, Controllability of Discrete time Bilinear Systems, *IEEE Trans. Amer. Aut. Cont.*, AC. **22** (1977), 78–83.

[41] CLIFFORD T. MULLIS, On the Controllability of Discrete Linear System with Output Feedback, *IEEE Trans. Aut. Contr.*, AC. **18** (1973), 608–615.

[42] T. GOKA, T. J. TARN, J. ZABORSZKY, On the Controllability of a Class of discrete bilinear systems, *Automatica*, **9** (1973), 615–622.

[43] F. M. BRASCH, JR. and J. B. PEARSON, Pole Placement using dynamic compensators, *IEEE Trans. Aut. Contr.*, AC. **15** (1970), 34–43.

[44] R. W. BROCKETT, Asymptotic Stability and Feedback Stabilization, in *Differential Geometric Control Theory*, R. W. Brockett, R. S. Millman, H. J. Sussmann eds., Birkhäuser, 1983, 181–191.

[45] H. NIJMEIJER and A. J. VAN DER SCHAFT, *Nonlinear Dynamical Control Systems*, Springer-Verlag, 1990.

[46] I. JOÓ and N. M. TUAN, On controllability of Bilinear Systems II, *Annales Univ. Sci. Budapest.*, Sect. Math., **35** (1992), 217–265.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 217–265

ON CONTROLLABILITY OF BILINEAR SYSTEMS II (CONTROLLABILITY IN TWO DIMENSIONS)

By

I. JOÓ and N. M. TUAN

Department of Analysis, L. Eötvös University, Budapest

(Received August 14, 1992)

1. Introduction

We consider the following bilinear control systems

$$\dot{x}(t)=Ax(t)+u(t)Bx(t) \tag{1}$$

and

$$\dot{x}(t)=Ax(t)+u(t)Bx(t)+u(t)b. \tag{2}$$

where $x(t)\in\mathbb{R}^2$, A and B are nonzero matrices in $\mathbb{R}^{2\times 2}$, b is a nonzero vector in $\mathbb{R}^2$, the control $u(\cdot)$ is assumed to belong to a class of admissible controls.

We mention some basic concepts for controllability theory. As usual we denote $x(t,x_0,u(\cdot))$ the solution of (1) (or of (2)) with the initial condition $x(0)=x_0$. The accessibility sets are denoted by $R_t(x)$ or $R(x)$, where

$$R_t(x):=\{x(t,x,u(\cdot)):\ u(\cdot) \text{ is admissible control}\}$$

$$R(x)=\bigcup_{t\geq 0}R_t(x).$$

We say that the system (1) is (completely) controllable on $\mathbb{R}^2\setminus\{0\}$ if for each $x\neq 0\in\mathbb{R}^2$, $R(x)=\mathbb{R}^2\setminus\{0\}$. Similarly the system (2) is (completely) controllable (on $\mathbb{R}^2$) if for all $x\in\mathbb{R}^2$, $R(x)=\mathbb{R}^2$.

The system (1) is sometimes called homogeneous system and the system (2) is called affine system.

The controllability of homogeneous systems (1) in a state of dimension n was examined in a number of publications (see for example [4], [5], [13], [14], [19], [29], [43], [45], [46], [48]). For linear systems $\dot{x}=Ax+Bu$, the necessary and sufficient controllability condition is well known, so-called Kalman controllability rank condition ([1], [2]), which depends only on the coefficient matrices. It is very natural to try to find similar controllability conditions for bilinear systems. In general it seems to be very difficult. The reason is the following: the behaviour of the accessibility sets depends strongly on the Lie-algebra of the system at each state. General non-linear systems in the plane were studied by L. R. HUNT [36] and O. HÁJEK [47]. In [36] a geometric sufficient controllability condition and a necessary condition for the system $\dot{x}=f(x)+ug(x)$ have been given. In O. HÁJEK's book [47] the local behaviour of the system $\dot{x}=uf(x)+(1-u)g(x)$ was examined: the classification of critical and non-critical points of the system, the point controllability, attainable set boundaries and a version of the bang-bang theorem were stated (chapters 4, 5). The homogeneous bilinear systems (1) in a low-dimensional state space were discussed in more detail [43]. For the two-dimension case LEPE [29] gave necessary and sufficient controllability conditions for the system (1) in the case $\operatorname{rank} B=2$ and other results have been given by KODITSCHEK and NARENDRA [45]. Only a few papers consider affine systems (2) ([30], [41], [44]). In [41] it is shown in which cases the system (2) is controllable provided the system (1) is controllable. In [45] necessary and sufficient conditions for complete controllability of (2) are derived. In [46] it is shown that the conditions in [45] are only sufficient and a counterexample is given.

In the **2.** below we consider the systems (1) in $\mathbb{R}^2\setminus\{0\}$. A Lepe-type necessary and sufficient controllability condition is stated for the system (1) in the case $\operatorname{rank} B=1$. A constructive proof for the Lepe's theorem is given, too. It is shown that we need only two constant control values to steer any state to other one, provided the system is controllable. The method seems to be useful to study the system (2) in **3.** The tools used in the proofs are elementary facts in differential equations theory.

In **3.** we consider the system (2). Using Theorem B [41] and the results obtained in **2.** we state necessary and sufficient controllability conditions for the system (2) with $\operatorname{rank} B=2$ (Theorem 2). For the case $\operatorname{rank} B=1$ the situation is more complicated. Here we use the idea in papers by HUNT ([35], [36]): The integral curves of $(B+b)$ divide $\mathbb{R}^2$ into two components. For controllability the orientation of A on these integral curves must be toward the first component for some points and towards the second component for some other points. We manage to consider all the cases (Theorems 3, 4, 5).

In **4.** we give two structure theorems (Theorems 6, 7).

2. Controllability of the homogeneous systems

We consider the homogeneous system

$$\dot{x}(t)=Ax(t)+u(t)Bx(t) \tag{1}$$

where $x(t)\in\mathbb{R}^2\setminus\{0\}$, A and B are nonzero matrices in $\mathbb{R}^{2\times 2}$, control $u(t)$ is a piecewise continuous scalar function with values in $\mathbb{R}$. In case $\operatorname{rank} B=2$ easily calculable controllability conditions have been given for the system (1) by N. L. LEPE [29]. They are the following:

THEOREM A [29]. *Consider the system* (1) *with* $\operatorname{rank} B=2$. *Then the necessary and sufficient condition of controllability is*

1. $\det[A,B]<0$ *if* $B\neq kE$ *and* B *has real eigenvalues.*
2. $\operatorname{Sp}^2 A-4\det A<0$ *if* $B=kE$.
3. $\det[A,B]+\det B\cdot\operatorname{Sp}^2 A<0$ *if* B *has purely imaginary conjugate eigenvalues.*
4. $B\neq kA$ *if* B *has complex conjugate eigenvalues with nonzero real part.*

In particular, in case $\det[A,B]>0$ *the system is not controllable.*

In what follows we consider the system (1) with $\operatorname{rank} B=1$, i.e. $B=bc^T$, b, c are nonzero vectors in $\mathbb{R}^2$. An analogous controllability condition is given. First we need the following lemma.

LEMMA 1. *Let* $B=bc^T$ *with* $0\neq b$, $c\in\mathbb{R}^2$. *Then in a suitable basis* B *has the form*

a) $\begin{pmatrix} c^Tb & 0\\ 0 & 0\end{pmatrix}$ *if* $c^Tb\neq 0$,

b) $\begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix}$ *if* $c^Tb=0$.

PROOF. Let $b=\begin{pmatrix} b_1\\ b_2\end{pmatrix}$, $c=\begin{pmatrix} c_1\\ c_2\end{pmatrix}$.

a) If $c^Tb\neq 0$ let $P=\begin{pmatrix} c_1 & c_2\\ -b_2 & b_1\end{pmatrix}$, then $P^{-1}:=\frac{1}{c^Tb}\begin{pmatrix} b_1 & -c_2\\ b_2 & c_1\end{pmatrix}$. It is obvious that $\det P=c^Tb\neq 0$ and

$$PBP^{-1}=\begin{pmatrix} c^Tb & 0\\ 0 & 0\end{pmatrix}.$$

b) $c^Tb=0\Longrightarrow c_1b_1=-c_2b_2$.

If $c_1\neq 0\Longrightarrow b_2\neq 0$, then let $P=\begin{pmatrix} 0 & \frac{1}{b_2}\\ c_1 & c_2\end{pmatrix}\Longrightarrow P^{-1}=\begin{pmatrix} b_1 & \frac{1}{c_1}\\ b_2 & 0\end{pmatrix}$.

If $c_2\neq 0\Longrightarrow b_1\neq 0$, then let $P=\begin{pmatrix} \frac{1}{b_1} & 0\\ c_1 & c_2\end{pmatrix}\Longrightarrow P^{-1}=\begin{pmatrix} b_1 & 0\\ b_2 & \frac{1}{c_2}\end{pmatrix}$.

In both cases

$$PBP^{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \qquad \blacksquare$$

Now we can prove the following result:

THEOREM 1. *Consider the system* (1) *with* $B = bc^T$, $b \neq 0$, $c^T \neq 0$. *Then it is controllable in* $\mathbb{R}^2 \setminus \{0\}$ *if and only if* $\det[A,B] < 0$.

PROOF. We mention first that we can change arbitrarily the basis in which the data are given. Indeed, if P is a nonsingular matrix and $A' := PAP^{-1}$, $B' = PBP^{-1}$ then the system

$$(1') \qquad \dot{y}(t) = A'y(t) + u(t)B'y(t)$$

is controllable if and only if (1) is ($x(t)$ is a solution of (1) if and only if $y(t) = Px(t)$ satisfies (1')). Further $\det[A,B] = \det[A',B']$. Hence and by Lemma 1 we have only to prove the theorem for the cases a) and b) mentioned in the Lemma 1.

a) $B = \begin{pmatrix} c^Tb & 0 \\ 0 & 0 \end{pmatrix}$, $c^Tb \neq 0$.

Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then $\det[A,B] = (c^Tb)^2 a_{12}a_{21}$. Hence $\det[A,B] <$ $< 0 \Longleftrightarrow a_{12}a_{21} < 0$.

Since the values of $u(t)$ are in $\mathbb{R}$, we can assume that $c^Tb = 1$. We have

$$(1a) \qquad \dot{x}(t) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} x(t) + u(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} x(t).$$

I. Necessity of the condition:

Assume that $a_{12}a_{21} \geq 0$. We assert that

α) if $a_{12} > 0$, $a_{21} > 0$ then the set $H := \{x \neq 0 : x_1 > 0,\ x_2 > 0\}$ is invariant set of (1a),

β) if $a_{12} < 0$, $a_{21} < 0$ then the set $H := \{x \neq 0 : x_1 > 0,\ x_2 < 0\}$ is invariant set of (1a),

γ) if $a_{12} = 0$ then $\{0\} \times (\mathbb{R} \setminus \{0\}) = H$ is invariant,

δ) if $a_{21} = 0$ then $(\mathbb{R} \setminus \{0\}) \times \{0\} = H$ is invariant,

i.e. if a solution $x(t)$ of (1a) starts in H then the whole trajectory remains in H. This proves the non-controllability of the system, from which the necessity follows. We prove the case α). The cases β), γ), δ can be proved similarly. Assume indirectly that $x(t)$ reaches the boundary of H (from inside into outside) first time. If $x(t) = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ with some $x_1 > 0$ then

$\dot{x}_2(t) = a_{21}x_1 > 0$ which shows that the trajectory goes back into the inside of H, in contradiction with the assumption that it reaches first time the boundary (from inside into outside). On the other hand, if $x(t) = \begin{pmatrix} 0 \\ x_2 \end{pmatrix}$ with $x_2 > 0$, then $\dot{x}_1(t) = a_{12}x_2 > 0$ which shows again the solution goes from outside into inside. The contradiction completes the proof of our assertion.

II. Sufficiency of the condition:

Let x^0, x^1 be arbitrary in $\mathbb{R}^2 \setminus \{0\}$. We assert (with a constructive proof) that there exists an admissible control $w(t)$ which steers x^0 into x^1, i.e.

$$\dot{x}(t) = (A + w(t)B)x(t)$$
$$x(0) = x^0, \quad x(T) = x^1 \quad \text{with some } T \geq 0.$$

For arbitrary matrices A, B let $A_u := A + uB$. Denote $f_u(\lambda)$ the characteristic polynomial of A_u, Δ_u the discriminant of $f_u(\lambda) = 0$, $-b_u$ the coefficient of λ, c_u the free coefficient in $f_u(\lambda)$. In our case we have

$$\begin{aligned} f_u(\lambda) &= \det(A_u - \lambda E) = \lambda^2 - \lambda(u + a_{11} + a_{22}) + a_{22}(a_{11} + u) - a_{12}a_{21} \\ \Delta_u &= (u + a_{11} - a_{22})^2 + 4a_{12}a_{21} \\ b_u &= u + a_{11} + a_{22} \\ c_u &= a_{22}(a_{11} + u) - a_{12}a_{21}. \end{aligned}$$

Now assume that $\det[A, B] < 0$, i.e. $a_{12}a_{21} < 0$. Then there is a number u such that $\Delta_u < 0$ and $b_u \neq 0$. With this u the eigenvalues of A_u are complex conjugate numbers with real part $b_u/2$.

α) $b_u < 0$.

In this case the trajectory $\dot{x}(t) = A_u x(t)$ tends to the origin as $t \to \infty$, goes clockwise around the origin if $a_{12} > 0$ and goes counterclockwise if $a_{12} < 0$, it makes infinitely many complete rotations as $t \to \infty$. There are two possibilities:

α_1) There is v such that $b_v > 0$ and $\Delta_v < 0$. The solution of $\dot{y} = -A_v y$, $y(0) = x^1$ tends to the origin and makes infinitely many complete rotation as $t \to \infty$. This trajectory goes counterclockwise if $a_{12} > 0$ and goes clockwise if $a_{12} < 0$. By the opposite directions of the rotations, this trajectory and the trajectory $\dot{x} = A_u x$, $x(0) = x^0$ meet infinitely often near the origin in both cases ($a_{12} > 0$ or $a_{12} < 0$). Let $x(T_1) = y(T_2)$ be such a common point. Taking

$$w(t) = \begin{cases} u, & 0 \leq t \leq T_1 \\ v, & T_1 < t \leq T_1 + T_2 \end{cases}$$

the solution of

$$\begin{cases} \dot{x}(t)=(A+w(t)B)x(t) \\ x(0)=x^0 \end{cases}$$

satisfies $x(T_1+T_2)=x^1$.

α_2) If $b_v>0$ then $\Delta_v\geq 0$. We assert that in this case $b_v>0$ implies also $c_v<0$. Indeed, $b_v>0$ for $v>-a_{11}-a_{22}$, $\Delta_v\geq 0$ for $|v+a_{11}-a_{22}|\not<\sqrt{-4a_{12}a_{21}}$ hence $b_v>0\Longrightarrow\Delta_v\geq 0$ means that $-a_{11}-a_{22}\geq -a_{11}+a_{22}+\sqrt{-4a_{12}a_{21}}$ i.e. $a_{22}\leq-\sqrt{-a_{12}a_{21}}$. From here we get the inequality $-a_{11}-a_{22}\geq -a_{11}+ +\frac{a_{12}a_{21}}{a_{22}}$. Now if $b_v>0$, then

$$v>-a_{11}-a_{22}\geq -a_{11}+\frac{a_{12}a_{21}}{a_{22}}, \qquad a_{22}(v+a_{11})<a_{12}a_{21}$$

i.e. $c_v<0$ as we asserted.

So take any v with $b_v>0$. Then $\Delta_v\geq 0$ and $c_v<0$ follows, hence one of the eigenvalues of A_v is negative, the other is positive.

Let f be the eigenvector of A_v which corresponds to the negative eigenvalue, i.e. $A_vf=\mu f$, $\mu<0$. If $x^0\neq\lambda f$ then $\dot{x}=A_vx$, $x(0)=x^0$ tends to infinity as $t\to\infty$, but it does not go around the origin (its angle tends to that of the other eigenvector). The trajectory $\dot{y}=-A_uy$, $y(0)=x^1$ tends to infinity as $t\to\infty$ and goes around the origin, makes infinitely many rotations as $t\to\infty$. These two trajectories meet infinitely often as $t\to\infty$, x^1 can be reached from x^0 by some admissible control $w(t)$ by the same way in α_1) and α_2). If $x^0=\lambda f$, consider $\dot{z}=A_uz$, $z(0)=x^0$, then for some $T_0>0$, $z(T_0)\neq kf$, and by the above, x^1 can be reached from $z(T_0)$, and hence can be reached from x^0.

β) $b_u>0$.

By the similar way we can prove this case. Here we give only a short sketch.

β_1) $\exists v$, $b_v<0$, $\Delta_v<0$. The trajectory $\dot{x}=A_ux$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other.

β_2) For all v, $b_v<0\Longrightarrow\Delta_v\geq 0$, In this case we have $\sqrt{-a_{12}a_{21}}\leq a_{22}$ and hence $b_v<0$ implies also $c_v<0$ (similarly as in α_2)). So let v be arbitrary with $b_v<0$. Since $\Delta_v\geq 0$, $c_v<0$, A_v has a positive and a negative eigenvalue. If $-A_vf=\mu f$, $\mu<0$, and $x^1\neq\lambda f$, then the trajectories $\dot{x}=A_ux$, $x(0)=x^0$ and $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other. If $x^1=\lambda f$, then there is $z^1\neq kf$, $T_0>0$ $\dot{z}=A_uz$, $z(0)=z^1$, $z(T_0)=x^1$ and since z^1 can be reached from x^0, x^1 can be reached from x^0, too.

b) $B=\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Let $A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then $\det[A,B]=-a_{21}^2$. Here $\det[A,B]<0 \Longleftrightarrow$ $\Longleftrightarrow a_{21}\neq 0$. In the case b) the system (1) is of the form

$$\text{(1b)} \qquad \begin{cases} \dot{x}_1(t)=a_{11}x_1(t)+a_{12}x_2(t)+u(t)x_2(t) \\ \dot{x}_2(t)=a_{21}x_1(t)+a_{22}x_2(t). \end{cases}$$

If $a_{21}=0$, then the half line $\{x : x_1>0,\ x_2=0\}$ is an invariant set of (1b). Indeed, the solution of $\dot{x}_2(t)=a_{22}x(t)$, $x_2(0)=0$ is $x_2(t)\equiv 0$ for all $t\geq 0$. And then $\dot{x}_1(t)=a_{11}x_1(t)$, the solution of this equation is $x_1(t)=e^{a_{11}t}x_1(0)>0$ if $x_1(0)>0$. This shows the non-controllability of the system. Conversely, let $a_{21}\neq 0$. Then

$$\begin{aligned} f_u(\lambda) &= \lambda^2-\lambda(a_{11}+a_{22})+a_{11}a_{22}-a_{21}(a_{12}+u) \\ \Delta_u &= (a_{11}-a_{22})^2+4a_{21}(a_{12}+u) \\ b_u &\equiv a_{11}+a_{22} \\ c_u &= a_{11}a_{22}-a_{21}(a_{12}+u). \end{aligned}$$

Since $a_{21}\neq 0$ there is a number u such that $\Delta_u<0$ and a number v such that $c_v<0$. then $\Delta_v=(a_{11}+a_{22})^2-4c_v\geq 0$. One of the eigenvalues of A_v (and $-A_v$) is negative, the other is positive. Let $A_v e=\lambda e$, $A_v f=\mu f$, $\lambda>0$, $\mu<0$. If $x^0\in \mathbb{R}e\cup\mathbb{R}f$ then there is $T_0>0$ such that $\dot{z}=A_u z$, $z(0)=x^0$ $z(T_0)\notin \mathbb{R}e\cup\mathbb{R}f$. If $x^1\in\mathbb{R}e\cup\mathbb{R}f$ then there is $T_0>0$ such that $\dot{z}=A_u z$, $z(T_0)=x^1$, $z(0)\notin\mathbb{R}e\cup\mathbb{R}f$. Hence we have only to show that for any x^0, $x^1\notin\mathbb{R}e\cup\mathbb{R}f$ x^1 can be reached from x^0. Now we assume that x^0, $x^1\notin\mathbb{R}e\cup\mathbb{R}f$. There are three possibilities.

1. $a_{11}+a_{22}<0$.

By analogous way as in α_2) above, we see that the trajectories $\dot{x}=A_v x$, $x(0)=x^0$ and $\dot{y}=-A_u y$, $y(0)=x^1$ meet in some point. Hence x^1 can be reached from x^0.

2. $a_{11}+a_{22}>0$.

In this case we can also show that the trajectory $\dot{x}=A_u x$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_v y$, $y(0)=x^1$ meet in some point, hence x^1 can be reached from x^0.

3. $a_{11}+a_{22}=0$.

In this case the eigenvalues of A_u are purely imaginary, the trajectory $\dot{x}=A_u x$ is an ellipse. If x^0 is in inside of the ellipse $\dot{y}=A_u y$, $y(0)=x^1$, then the trajectory $\dot{x}=A_v x$, $x(0)=x^0$ goes to infinity, so it must meet the ellipse in some point and x^1 can be reached from this point along the ellipse. If x^1 is in inside of the ellipse $\dot{x}=A_u x$, $x(0)=x^0$, then the trajectory $\dot{y}=-A_v y$, $y(0)=y^1$ goes to infinity and it meets the ellipse in a point which can be reached from x^0 and from which x^1 can be reached along the trajectory $\dot{y}=A_v y$. If x^0 and x^1 are on the same ellipse $\dot{x}=A_u x$, $x(0)=x^0$ then on this ellipse $x(T)=x^1$ for some $T>0$.

The Theorem 1 is completely proved. ∎

The technique used in the above proof seems to be useful. Applying the idea of this technique in the following we give a constructive proof for Lepe's theorem.

PROOF OF LEPE'S THEOREM.

SOME REMARKS BEFORE THE PROOF: 1. By the remark at the beginning of the proof of the Theorem 1 we have only to prove the theorem for the system (1) in a suitable basis, in which the calculation is more convenient. We remark also that when changing the basis the values $\det(\cdot)$, $\mathrm{Sp}(\cdot)$ do not change, hence the conditions in the theorem are unchanged.

2. Here we use also the notations defined in the proof of the Theorem 1: $f_u(\lambda)$, A_u, Δ_u, b_u, c_u ...

3. The proof of the sufficiency of the conditions is based on the following idea: Let x^0, x^1 be arbitrary in $\mathbb{R}^2\setminus\{0\}$. We try to find values u and v such that the trajectory $\dot{x}=A_u x$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_v(y)$, $y(0)=x^1$ meet each other in $x(T_1)=y(T_2)$. Hence the control

$$w(t)=\begin{cases} u, & 0\le t\le T_1 \\ v, & T_1<t\le T_1+T_2 \end{cases}$$

steers x^0 into x^1.

Now we prove the theorem.

1. There are two cases.

1.a) B has distinct eigenvalues.

In a suitable basis B is of the form

$$B=\begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}.$$

(Here $P=(y,z)$ where $By=b_1 y$, $Bz=b_2 z$, $P^{-1}BP=\mathrm{diag}(b_1,b_2)$.) Since $\mathrm{rank}\,B=2$, we have $0\ne b_1\ne b_2\ne 0$. Let

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

then $\det[A,B]=(b_1-b_2)^2 a_{12}a_{21}$. Hence $\det[A,B]<0 \Longleftrightarrow a_{12}a_{21}<0$ (because $b_1\ne b_2$).

Necessary condition: Assume that $a_{12}a_{21}\ge 0$. By the same way in a) of the proof of the Theorem 1 one can see that the system (1) is not controllable in $\mathbb{R}^2\setminus\{0\}$.

Sufficient condition: Assume that $\det[A,B]<0$, i.e. $a_{12}a_{21}<0$, then

$$f_u(\lambda)=\lambda^2-\lambda(a_{11}+a_{22}+u(b_1+b_2))+(a_{11}+ub_1)(a_{22}+ub_2)-a_{12}a_{21}$$
$$\Delta_u=(a_{11}-a_{22}+u(b_1-b_2))^2+4a_{12}a_{21}.$$

α) $b_1b_2>0$.

Since $b_1\neq b_2$, $a_{12}a_{21}<0$, $b_1+b_2\neq 0$, we can choose u such that $\Delta_u<0$ and $b_u\neq 0$

α_1) $b_u<0$. If there is a number v such that $b_v>0$, $\Delta_v<0$ then by the same way in α_1) of the proof of the Theorem 1 one can see that the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other. Otherwise $b_v>0\Longrightarrow\Delta_v\geq 0$. Since $b_1b_2>0$ there is a number v such that $c_v>0$ and $b_v>0$. In this case $\Delta_v\geq 0$, the eigenvalues of $-A_v$ are negative. The trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ tends to the origin as $t\rightarrow\infty$ but it does not go around the origin, it must meet the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ in some point near the origin.

α_2) $\Delta_v<0\Longrightarrow b_v\geq 0$, i.e. $b_v<0\Longrightarrow\Delta_v\geq 0$. In this case we can choose v such that $c_v>0$, $b_v<0$, $\Delta_v\geq 0$. The trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meets the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ in some point.

β) $b_1b_2<0$.

We can choose u and v such that $\Delta_u<0$, and $\Delta_v>0$, $c_v<0$. In this case the eigenvalues of A_v ($-A_v$) are opposite real. By the same reasons as in the part b) of the proof of the Theorem 1 we have only to show that for any x^0, $x^1\notin\mathbb{R}e\cup\mathbb{R}f$ (here e and f are distinct eigenvectors of A_v) x^1 can be reached from x^0. Hence we assume that x^0, $x^1\notin\mathbb{R}e\cup\mathbb{R}f$.

β_1) $b_1+b_2\neq 0$. Then we can obtain $b_u\neq 0$. If $b_u<0$ then the trajectories $\dot{x}=A_vx$, $x(0)=x^0$ and $\dot{y}=-A_uy$, $y(0)=x^1$ meet each other. If $b_u>0$, then the trajectories $\dot{x}=A_ux$, $x(0)=x^0$ and $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other.

β_2) $b_1+b_2=0$.

If $a_{11}+a_{12}=0$ then the trajectory $\dot{x}=A_ux$ is an ellipse. In the case if x_0 is in the interior of the ellipse $\dot{y}=A_uy$, $y(0)=x^1$, the trajectory $\dot{x}=A_vx$, $x(0)=x^0$ meets the ellipse in a point from which x^1 can be reached along the ellipse. In the case x^1 is in the inside of the ellipse $\dot{x}=A_ux$, $x(0)=x^0$, the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meets the ellipse in a point, which can be reached from x^0. In both cases x^1 can be reached from x^0. If $a_{11}+a_{22}=b_u\neq 0$, we have the cases $b_u<0$ and $b_u>0$ in β_1).

1.b) B has a twofold real eigenvalue and $B\neq kE$.

Let b be the eigenvalue of B. Then $b=\frac{b_{11}+b_{22}}{2}$ and $\det(B-bE)=0$. Since $B\neq kE$, $b_{12}^2+b_{21}^2\neq 0$. If $b_{12}\neq 0$ let $P=\begin{pmatrix}0 & b_{12}\\ 1 & b-b_{11}\end{pmatrix}$, and if $b_{21}\neq 0$, let $P=\begin{pmatrix}1 & b-b_{22}\\ 0 & b_{21}\end{pmatrix}$. One can see that $P^{-1}BP=\begin{pmatrix}b & 0\\ 1 & b\end{pmatrix}$. Hence we can assume that

$$B=\begin{pmatrix}b & 0\\ 1 & b\end{pmatrix},\qquad b\neq 0,\quad (\operatorname{rank} B=2).$$

Then $\det[A,B]=-a_{12}^2$. $\det[A,B]<0\Longleftrightarrow a_{12}\neq 0$.

Necessary condition: If $a_{12}=0$ then one can see that the set $\{x:x_1==0,\ x_2>0\}$ is an invariant set of (1). This proves the non-controllability of the system (1).

Sufficient condition: Assume that $\det[A,B]<0$, i.e. $a_{12}\neq 0$. Then

$$f_u(\lambda)=\lambda^2-\lambda(2ub+a_{11}+a_{22})+(a_{11}+ub)(a_{22}+ub)-a_{12}(a_{21}+u)$$
$$\Delta_u=(a_{11}-a_{22})^2+4a_{12}a_{21}+4a_{12}u.$$

Here we can choose u such that $\Delta_u<0$ and $b_u\neq 0$.

α) $b_u<0$.

If there is a number v such that $b_v>0$ and $\Delta_v<0$ then the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other (see α_1) in the proof of the Theorem 1). If $b_v>0\Longrightarrow\Delta_v\geq 0$, then there is a number v such that $c_v>0$, $b_v>0$. In this case the eigenvalues of $-A_v$ are negative. The trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ tends to origin as $t\to\infty$ but it does not go around the origin, it must meet the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ in some point near the origin.

β) $b_u>0$.

If there is v such that $b_v<0$ and $\Delta_v<0$, then the trajectory $\dot{x}=A_ux$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other in some point (see 1.a) above). Otherwise $b_v<0\Longrightarrow\Delta_v\geq 0$. Let v be a number such that $c_v>0$, $b_v<0$. In this case the eigenvalues of A_v are negative. The trajectory $\dot{x}=A_vx$, $x(0)=x^0$ tends to the origin as $t\to\infty$, does not go around the origin, it must meet the trajectory $\dot{y}=-A_ux$, $y(0)=x^1$ in a point near the origin. We completed the proof of the case 1.

2. $B=kE$: $(k\neq 0)$.

Denote $\varrho=\operatorname{Sp}^2A-4\det A$, then $\varrho=(a_{11}-a_{22})^2+4a_{12}a_{21}$. Since the values of $u(t)$ can be arbitrary in $\mathbb{R}$, we can assume that $k=1$, i.e. $B=E$.

Necessary condition: Assume that $\varrho\geq 0$. We calculate $\det A_u$:

$$\det A_u=u^2+u\cdot\operatorname{Sp}A+\det A.$$

The discriminant of $\det A_u = 0$ is of the form

$$\Delta = \mathrm{Sp}^2 A - 4\det A = \varrho.$$

Since $\varrho \geq 0$ there is a real number α such that $\det A_\alpha = 0$. We can write the system (1) in the form

$$\dot{x} = A_\alpha x + ux.$$

Since $\det A_\alpha = 0$, A_α can be written as $A_\alpha = bc^T$, b, $c \in \mathbb{R}^2$. Then

$$\dot{x} = bc^T x + ux.$$

If $A_\alpha = 0$ then $\dot{x} = ux$ is not controllable in $\mathbb{R}^2 \setminus \{0\}$. If $A_\alpha \neq 0$ then $c^T \neq 0$ and

$$c^T \dot{x} = c^T bc^T x + uc^T x.$$

From this if $c^T x(0)=0$ then $c^T x(t) \equiv 0$ for all $t \geq 0$, the half line $\{x \neq 0 : c^T x = 0\}$ is an invariant set of (1). Hence the system is not controllable in $\mathbb{R}^2 \setminus \{0\}$.

Sufficient condition: Assume that $\varrho < 0$

$$f_u(\lambda) = \lambda^2 - \lambda(2u + a_{11} + a_{22}) + (a_{11} + u)(a_{22} + u) - a_{12}a_{21}$$
$$\Delta_u = (a_{11} - a_{22})^2 + 4a_{12}a_{21} = \varrho < 0 \quad \text{for all } u.$$

Hence we can choose u and v such that $b_u < 0$ and $b_v > 0$. The trajectory $\dot{x} = A_u x$, $x(0) = x^0$ and the trajectory $\dot{y} = -A_v y$, $y(0) = x^1$ tend to the origin as $t \to \infty$, go around the origin and make infinitely many rotations as $t \to \infty$, but here the rotations are in the opposite directions (because from $\varrho < 0$ we have $a_{12}a_{21} < 0$). The trajectories must meet in some point near the origin.

We finished the proof of the part 2.

3. B has purely imaginary conjugate eigenvalues.

Then $B = \begin{pmatrix} k & b_{12} \\ b_{21} & -k \end{pmatrix}$ with $\det B > 0$, i.e. $b_{12}b_{21} < -k^2$. Since $b_{12}b_{21} < -k^2$, $b_{12}b_{21} \neq 0$. Let $P = \begin{pmatrix} k & 1 \\ b_{21} & 0 \end{pmatrix}$ then $P^{-1}BP = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}$ with $b = -\det B < 0$. Hence we have only to consider the system (1) with

$$B = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}, \qquad b < 0.$$

Now let

$$\varrho = \det[A, B] + \det B \cdot \mathrm{Sp}^2 A, \quad \text{then}$$
$$\varrho = -(a_{12}b - a_{21})^2 - 4ba_{11}a_{22}.$$

Necessary condition: Assume that $\varrho \geq 0$.

Consider the system (1) $\dot{x}=A_{u(t)}x(t)$, which can be written as

$$\begin{cases} \dot{x}_1(t)=a_{11}x_1(t)+(a_{12}+u(t))x_2(t) \\ \dot{x}_2(t)=(a_{21}+u(t)b)x_1(t)+a_{22}x_2(t). \end{cases}$$

By a simple calculation we obtain that the solution of (1) satisfies the following:

$$\frac{\mathrm{d}}{\mathrm{d}t}\{-bx_1^2(t)+x_2^2(t)\}=$$
$$=2\{a_{22}x_2^2(t)-ba_{11}x_1^2(t)+(a_{21}-ba_{12})x_1(t)x_2(t)\}=:U(x_1(t))=:V(x_2(t)).$$

If $a_{11}=a_{22}=0$ then $\varrho=-(a_{12}b-a_{21})^2\le 0$ and the assumption that $\varrho\ge 0$ holds only in the case $\varrho=0$. Then $\frac{\mathrm{d}}{\mathrm{d}t}\{-bx_1^2+x_2^2\}=0$ and hence $-bx_1^2(t)+$ $+x_2^2(t)=$ constant, the ellipses $-bx_1^2+x_2^2=c$ with $c>0$ are invariant sets of (1).

If $a_{11}\ne 0$ then the discriminant of $\frac{1}{2}U(x_1(t))=0$ is calculated as

$$\Delta=x_2^2(t)\{(a_{21}-ba_{12})^2+4ba_{11}a_{22}\}=-\varrho x_2^2(t)\le 0.$$

$U(x_1(t))\ge 0$ if $-ba_{11}>0$ and in this case the set $\{x\ne 0:-bx_1^2+x_2^2\ge c\}$ with some $c>0$ is an invariant set of (1). In the case $-ba_{11}<0$ the set $\{x\ne 0:-bx_1^2+x_2^2\le c\}$ is invariant.

Similarly if $a_{22}>0$ then the set $\{x\ne 0:-bx_1^2+x_2^2\ge c\}$ is invariant and if $a_{22}<0$ then the set $\{x\ne 0:-bx_1^2+x_2^2\le c\}$ is invariant.

Sufficient condition: Assume that $\varrho<0$. Then

$$f_u(\lambda)=\lambda^2-\lambda(a_{11}+a_{22})+a_{11}a_{22}-(a_{12}+u)(a_{21}+bu)$$
$$\Delta_u=(a_{11}+a_{22})^2-4a_{11}a_{22}+4(a_{12}+u)(a_{21}+bu)=$$
$$=4bu^2+4(a_{12}b+a_{21})u+4a_{12}a_{21}+(a_{11}-a_{22})^2$$
$$b_u=a_{11}+a_{22}$$
$$c_u=-bu^2-(a_{12}b+a_{21})u+a_{11}a_{22}-a_{12}a_{21}.$$

The discriminant of $\Delta_u=0$ is the following:

$$\Delta_{\Delta_u}=16(a_{12}b+a_{21})^2-64ba_{12}a_{21}+64ba_{11}a_{22}-16b(a_{11}+a_{22})^2=$$
$$=16\left((a_{12}b-a_{21})^2+4ba_{11}a_{22}-b(a_{11}+a_{22})^2\right)=$$
$$=16(\{-\varrho-b(a_{11}+a_{22})^2\}>0.$$

The discriminant of c_u is

$$\Delta_{c_u}=(a_{11}b+a_{21})^2-4ba_{21}a_{12}+4ba_{11}a_{22}=-\varrho>0.$$

Since $4b<0$ there is a number u such that $\Delta_u<0$. Since $(a_{11}-a_{22})^2\geq \geq -4a_{11}a_{22}$, we see that $c_v<0$ implies $\Delta_v>0$. From $\Delta_{c_v}<0$ we get the existence of v with $c_v>0$ and $\Delta_v>0$. The eigenvalues of A_v (and of $-A_v$) are opposite real numbers. By analogous reasons as in the part b) of the proof of the Theorem 1 or in the part 1.a) we can assume that x^0, $x^1\notin \mathbb{R}e\cup\mathbb{R}f$ (e, f are eigenvectors of A_v). For any $z^0\notin\mathbb{R}e\cup\mathbb{R}f$ the trajectory $\dot{z}=A_vz$ ($\dot{z}=-A_vz$), $z(0)=z^0$ tends to infinity as $t\to\infty$ and does not go around the origin.

There are three cases

α) $a_{11}+a_{22}=0$.

In this case $b_u=a_{11}+a_{22}=0$. The trajectory $\dot{x}=A_ux$ is an ellipse. If x^0 and x^1 are on the same ellipse $\dot{x}=A_ux$, $x(0)=x^0$, then x^1 can be reached from x^0 along this trajectory. If x^0 is in the inside of the ellipse $\dot{y}=-A_uy$, $y(0)=x^1$, then the trajectory $\dot{x}=A_vx$, $x(0)=x^0$ goes from inside into outside, meets the ellipse in some point, from which x^1 can be reached. If x^1 is in the inside of the ellipse $\dot{x}=A_ux$, $x(0)=x^0$, then the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ goes to infinity, it meets the ellipse in some point.

β) $a_{11}+a_{22}<0$.

In this case the trajectory $\dot{x}=A_vx$, $x(0)=x^0$ and the trajectory $\dot{y}= =-A_uy$, $y(0)=x^1$ meet each other.

γ) $a_{11}+a_{22}>0$.

The trajectory $\dot{x}=A_ux$, $x(0)=x^0$ and the trajectory $\dot{y}=-A_vy$, $y(0)=x^1$ meet each other (in infinity).

We completed the proof of the part 3.

4. B has complex conjugate eigenvalues with nonzero real parts.

Necessary condition: Assume that $B=kA$, then the system (1) is of the form

$$\dot{x}(t)=(1+ku(t))Ax(t)=v(t)Ax(t)$$

which is not controllable on $\mathbb{R}^2\setminus\{0\}$.

Sufficient condition: Assume that $B\neq kA$. Let $B=\begin{pmatrix} b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix}$. From the conditions of B we have

$$b_{11}+b_{22}\neq 0 \qquad (b_{11}-b_{22})^2+4b_{12}b_{21}<0.$$

Hence $b_{21}\neq 0$. Let $P=\begin{pmatrix} -b_{22} & 1\\ b_{21} & 0\end{pmatrix}$, then $P^{-1}BP=\begin{pmatrix} 0 & 1\\ b_1 & b_2\end{pmatrix}$ where $b_2=b_{11}+b_{22}$, $b_1=b_{12}b_{21}-b_{11}b_{12}$. We can assume that

$$B=\begin{pmatrix} 0 & 1\\ b_1 & b_2\end{pmatrix}, \qquad b_2\neq 0,\quad b_2^2+4b_1<0 \quad\Longrightarrow\quad b_1<0.$$

Now

$$A_u=\begin{pmatrix} a_{11} & a_{12}+u \\ a_{21}+b_1u & a_{22}+b_2u \end{pmatrix}$$

$$f_u(\lambda)=\lambda^2-\lambda(a_{11}+a_{22}+b_2u)+a_{11}(a_{22}+b_2u)-(a_{12}+u)(a_{21}+b_1u)$$

$$\Delta_u=(b_2u+a_{22}-a_{11})^2+4(a_{12}+u)(a_{21}+b_1u)$$

$$\Delta_u=(b_2^2+4b_1)u^2+u\cdot k+l \quad (k,\ l \text{ are independent of } u).$$

$$b_u=b_2u+a_{11}+a_{22}$$

$$c_u=-(a_{12}+u)(a_{21}+b_1u)+a_{11}(a_{22}+b_2u).$$

First we show that there is v such that $\Delta_v\geq 0$ and $c_v\neq 0$.

a) $a_{12}b_1\neq a_{21}$. Since $b_1<0$, $(a_{12}+u)(a_{21}+b_1u)>0$ for all $u\in\left(a_{12},\frac{a_{21}}{b_1}\right)$ if $a_{12}<\frac{a_{21}}{b_1}$, and for all $u\in\left(\frac{a_{21}}{b_1},a_{12}\right)$ if $\frac{a_{21}}{b_1}<a_{12}$. In both cases there is v such that $c_v\neq 0$ and v is between a_{12} and $\frac{a_{12}}{b_1}$. With this v we have $\Delta_v>0$.

b) $a_{12}b_1=a_{21}$. Let $u=-a_{12}=-\frac{a_{21}}{b_1}$, then $\Delta_u=(-b_2a_{12}+a_{22}-a_{11})^2$.

If $-b_2a_{12}+a_{22}-a_{11}\neq 0$ then there is $\varepsilon>0$ such that for $|u+a_{12}|<\varepsilon$ $\Delta_u>0$. Hence there is v such that $|v+a_{12}|<\varepsilon$ and $c_v\neq 0$, and $\Delta_v>0$.

If $-b_2a_{12}+a_{22}-a_{11}=0$, then $\Delta_{-a_{12}}=0$. In this case $a_{11}\neq 0$. Indeed, if $a_{11}=0$ then $b_2a_{12}=a_{22}$, i.e. $a_{12}B=A$, in contradiction with the assumption that $B\neq kA$ (Here $a_{12}\neq 0$ because if $a_{12}=0\Longrightarrow a_{21}=a_{22}=0\Longrightarrow A=0$). $a_{11}\neq 0$ and $a_{11}=-b_2a_{12}+a_{22}$, we have $c_{-a_{12}}=a_{11}^2\neq 0$. Hence for $v=-a_{12}$, $\Delta_v=0$ and $c_v\neq 0$.

Now we can show that x^1 can be reached from x^0. Consider the trajectory $\dot{x}=A_vx$, $x(0)=x^0$. Since the eigenvalues of A_v are non-zero real numbers, this trajectory tends to the origin or tends to infinity as $t\to\infty$, but it does not go around are origin. Assume first that it tends to the origin. Since $b_2^2+4b_1<0$ and $b_2\neq 0$ we can choose u such that $b_u>0$ and $\Delta_u<0$. The trajectory $\dot{y}=-A_uy$, $y(0)=x^1$ tends to the origin as $t\to\infty$, goes around the origin, it must meet the trajectory $\dot{x}=A_vx$, $x(0)=x^0$. In the case the trajectory $\dot{x}=A_vx$, $x(0)=x^0$ tends to infinity as $t\to\infty$, then we can choose u such that $b_u<0$ and $\Delta_u<0$. The trajectories $\dot{x}=A_vx$, $x(0)=x^0$ and $\dot{y}=-A_uy$, $y(0)=x^1$ meet each other.

We finished the proof of the part 4.

Now we show that in the case $\det[A,B]>0$ the system is not controllable.

In the case 1. it is obvious that $\det[A,B]\geq 0$ implies the non-controllability of of (1). In the case 2. we have always $\det[A,B]=0$. In the case

3. if $\det[A,B] \geq 0$ then $\det[A,B] + \det B \cdot \mathrm{Sp}^2 A \geq 0$ because here $\det B > 0$. Now we consider the case 4.

$\det[A,B] =$

$$= b_1(a_{22}-a_{11})^2 - (a_{22}-a_{11})(b_1 b_2 a_{12} + b_2 a_{21}) - (b_1 a_{12} - a_{21})^2 + b_2^2 a_{12} a_{21}.$$

This is a second degree polynomial of $(a_{22}-a_{11})$ whose discriminant is non-positive

$$\Delta = (b_2^2 + 4b_1)(b_1 a_{12} - a_{21})^2 \leq 0.$$

Since $b_1 < 0$, $\det[A,B] \leq 0$ for all A. $\det[A,B] > 0$ does not hold in this case.

The Theorem is completely proved. ∎

REMARK: From the proof of the Theorems 1 and A we can see that if the system (1) is controllable on $\mathbb{R}^2 \setminus \{0\}$ then we need only two constant control values (here u and v) to steer any state to other one.

3. Controllability of affine systems

In this part we study the controllability of the affine system

$$\dot{x}(t) = Ax(t) + u(t)Bx(t) + u(t)d \tag{2}$$

where $x(t) \in \mathbb{R}^2$, A and B are nonzero matrices in $\mathbb{R}^{2\times 2}$, d is a nonzero vector in $\mathbb{R}^2$. Here we assume that $u(t)$ is a piecewise constant scalar function with value in $\mathbb{R}$.

Instead of working with (2) it will sometimes be convenient to work with an equivalent family of vector fields

$$\mathcal{F}_0 = \{A + u(B+d) : u \in \mathbb{R}\}$$

where the vector field $A + u(B+d)$ at a point x is defined by

$$(A + u(B+d))(x) = Ax + uBx + d.$$

For a vector field F on $\mathbb{R}^2$ we denote by $\exp(tF)x$ the solution of

$$\begin{cases} \dot{x}(t) = F(x(t)) \\ x(0) = x. \end{cases}$$

For an arbitrary family $\mathcal{F}$ of vector fields on $\mathbb{R}^2$, denote

$$\mathrm{A}_{\mathcal{F}}(x) := \{\exp(t_1 F_1)\ldots\exp(t_p F_p)x : t_i \geq 0,\ F_i \in \mathcal{F},\ p = 1,2,\ldots\}$$

This is the accessibility set of $\mathcal{F}$ from x. We see that the accessibility sets of $\mathcal{F}_0$ and of (2) are equal. We say that the system $\mathcal{F}$ is controllable in $\mathbb{R}^2$ if $\mathrm{A}_{\mathcal{F}}(x) = \mathbb{R}^2$ for all $x \in \mathbb{R}^2$. We need the following lemma:

LEMMA 2. *Let* $\mathcal{F}_1 := \{F+uG : u\in\mathbb{R}\}$ *and* $\mathcal{F}_2 = \{F, \pm G\}$. *Then* $\mathcal{F}_1$ *is controllable if and only if* $\mathcal{F}_2$ *does. (*F*,* G *are vector fields on* $\mathbb{R}^2$*).*

PROOF. This result is well known in system theory. Here we give only a sketch.

Using extension techniques we can see that the systems $\mathcal{F}_1$ and $\mathcal{F}_2$ are equivalent (somewhere it is called weakly equivalent) (see [37], [38], [39], [40]), i.e. $\operatorname{cl}\mathbb{A}_{\mathcal{F}_1}(x) = \operatorname{cl}\mathbb{A}_{\mathcal{F}_2}(x)$ for all $x\in\mathbb{R}^2$. (cl: closure of). The Lie algebras of these systems are the same, generated by F and G. Denote it by $\mathcal{L}$. Assume that $\mathcal{F}_1$ is controllable, then $\mathcal{L}(x) = \mathbb{R}^2$ for all x (see [8]) and then $\operatorname{cl}\mathbb{A}_{\mathcal{F}_2}(x) = \operatorname{cl}\mathbb{A}_{\mathcal{F}_1}(x) = \mathbb{R}^2$ for all x. $\mathcal{L}(y) = \mathbb{R}^2 \Longrightarrow \operatorname{int}\mathbb{A}_{-\mathcal{F}_0}(y) \neq \emptyset$ (for all y), hence $\mathbb{A}_{\mathcal{F}_0}(x) \cap \operatorname{int}\mathbb{A}_{-\mathcal{F}_0}(y) \neq \emptyset$ and this shows that y can be reached from x by $\mathcal{F}_0$. ∎

We will also use the following result about a sufficient condition of the controllability of affine systems stated by V. JURDJEVIC and G. SALLET (see [41], Theorem 2). In our case we have:

THEOREM B [41]. *Consider the system* $\mathcal{F}_0$ *above. Assume that*

a) $\{A+uB : u\in\mathbb{R}\}$ *is controllable in* $\mathbb{R}^2 \setminus \{0\}$

b) $\mathcal{F}_0$ *has no fixed points in* $\mathbb{R}^2$*, i.e. for all* $x\in\mathbb{R}^2$ *there exists* $F\in\mathcal{F}_0$ *such that* $F(x)\neq 0$.

Then $\mathcal{F}_0$ *is controllable on* $\mathbb{R}^2$.

Now we can prove the following theorem.

THEOREM 2. *Consider the system* (2) *with* $\operatorname{rank} B = 2$. *Then the necessary and sufficient conditions of the controllability of* (2) *are:*

1. $\det[A,B] < 0$ *and* $AB^{-1}d \neq 0$ *if* $B\neq kE$ *and* B *has real eigenvalues*
2. $\operatorname{Sp}^2 A - 4\det A < 0$ *and* $d\neq 0$ *if* $B = kE$
3. $\det[A,B] + \det B\cdot\operatorname{Sp}^2 A \leq 0$ *and* $AB^{-1}d\neq 0$ *if* B *has purely imaginary conjugate eigenvalues.*
4. $AB^{-1}d \neq 0$ *if* B *has complex conjugate eigenvalues with nonzero real part.*

In particular, in the case $\det[A,B] > 0$ *the system is not controllable.*

PROOF. We mention first that here we can also work on a suitable basis, in which the calculation is more convenient. In fact if P is nonsingular matrix and $A' = PAP^{-1}$, $B' = PBP^{-1}$, $d' = Pd$ then the system

$$\dot{y}(t) = A'y(t) + u(t)B'y(t) + u(t)d' \tag{2'}$$

is controllable if and only if (2) does ($x(t)$ is solution of (2) if and only if $y(t) = Px(t)$ satisfies (2').). Further if $X' = PXP^{-1}$ then $\det X' = \det X$, $\mathrm{Sp}\, X' = \mathrm{Sp}\, X$, and $AB^{-1}d \neq 0 \Longleftrightarrow A'B'^{-1}d \neq 0$.

The condition $AB^{-1}d \neq 0$ is equivalent to the condition that (2) has no fixed points in $\mathbb{R}^2$, which is a necessary condition of the controllability. We prove this fact:

If $AB^{-1}d = 0$, then for $x = -B^{-1}d$, $Ax = u(Bx + d) = 0$ for all $u \in \mathbb{R}$, i.e. x is a fixed point of $\mathcal{F}_0$ (or of (2)). If $Ax + u(Bx + d) = 0$ for all $u \in \mathbb{R}$, then $Ax = 0$ and $Bx = -d$, hence $x = -B^{-1}d$ and $AB^{-1}d = 0$. This proves the equivalence above. Now if x is a fixed point, then the system is not controllable: $AB^{-1}d \neq 0$ is a necessary condition of the controllability.

We remark that the Theorem A is true for the case when the controls are piecewise constant. This fact can be seen from the proof of the Theorem A.

We consider the following system

$$\dot{y}(t) = Ay(t) + u(t)By(t) + k \tag{2*}$$

where $k = -AB^{-1}d$.

We see that the system (2) is controllable if and only if the system (2*) is. Indeed $x(t)$ is a solution of (2) if and only if $y(t) = x(t) + B^{-1}d$ satisfies (2*).

Now we prove the cases stated in our theorem.

1. $B \neq kE$ and B has real eigenvalues.

By the Theorems A and B and by the remarks above, $\det[A, B] < 0$ and $AB^{-1}d \neq 0$ are sufficient conditions of the controllability of (2).

Necessity of the condition:

It is known that $AB^{-1}d \neq 0$ is a necessary condition of the controllability of (2). Assume that $\det[A, B] \geq 0$.

a) B has distinct real eigenvalues.

As in the proof of the Theorem A, we can assume that

$$B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} \qquad 0 \neq b_1 \neq b_2 \neq 0.$$

Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad k = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}.$$

Then $\det[A, B] \geq 0 \Longleftrightarrow a_{12}a_{21} \geq 0$.

a$_1$) $a_{12}a_{21} = 0$.

If $a_{12}=0$, then for the system (2*)

$$\dot{y}_1(t)=a_{11}y_1(t)+u(t)b_1y_1(t)+k_1.$$

Hence the half plane $\{y:y_1\geq 0\}$ is an invariant set of (2*) in the case $k_1>0$, the half plane $\{y:y_1\leq 0\}$ is an invariant set of (2*) in the case $k_1<0$ and the line $\{y:y_1=0\}$ is an invariant set in the case $k_1=0$.

Similarly, if $a_{21}=0$ then the sets $\{y:y_2\geq 0\}$, $\{y:y_2\leq 0\}$, $\{y:y_2=0\}$ are invariant sets of (2*) in the cases $k_2>0$, $k_2<0$, $k_2=0$, resp.

a$_2$) $a_{12}a_{21}>0$, $a_{12}>0$, $a_{21}>0$.

The system (2*) can be written as

$$\dot{y}_1(t)=a_{11}y_1(t)+a_{12}y_2(t)+u(t)b_1y_1(t)+k_1$$

$$\dot{y}_2(t)=a_{21}y_1(t)+a_{22}y_2(t)+u(t)b_2y_2(t)+k_2.$$

If $k_1\geq 0$, $k_2\geq 0$ then the set $\{y:y_1\geq 0,\ y_2\geq 0\}$ is an invariant set of (2*). If $k_1\leq 0$, $k_2\leq 0$ then the set $\{y:y_1\leq 0,\ y_2\leq 0\}$ is an invariant set of (2*). If $k_1>0$, $k_2<0$, then the set $\mathbb{R}^2\setminus\{y:y_1<0,\ y_2>0\}$ is an invariant set of (2*). If $k_1<0$, $k_2>0$, then the set $\mathbb{R}^2\setminus\{y:y_1>0,\ y_2<0\}$ is an invariant set of (2*).

a$_3$) $a_{12}a_{21}>0$, $a_{12}<0$, $a_{21}<0$.

Similar to a$_2$). If $k_1>0$, $k_2>0$ then the set $\mathbb{R}^2\setminus\{y:y_1<0,\ y_2<0\}$ is an invariant set of (2*). If $k_1<0$, $k_2<0$ then the set $\mathbb{R}^2\setminus\{y:y_1>0,\ y_2>0\}$ is an invariant set of (2*). If $k_1\geq 0$, $k_2\leq 0$ then the set $\{y:y_1\geq 0,\ y_2\leq 0\}$ is an invariant set of (2*). If $k_1\leq 0$, $k_2\geq 0$ then the set $\{y:y_1\leq 0,\ y_2\geq 0\}$ is an invariant set of (2*).

In all the cases there is an invariant set of (2*), hence (2*) is not controllable. We ended the proof of a).

b) $B\neq kE$ and B has a twofold real eigenvalue.

We can assume that

$$B=\begin{pmatrix} b & 0\\ 1 & b\end{pmatrix},\qquad b\neq 0.$$

In this case $\det[A,B]\leq 0$, and $\det[A,B]=0\Longleftrightarrow a_{12}=0$. If $a_{12}=0$ then

$$\dot{y}_1(t)=a_{11}y_1(t)+u(t)by_1(t)+k_1.$$

If $k_1>0$ then the half plane $\{y:y_1\geq 0\}$ is an invariant set of (2*). If $k_1<0$ then the half plane $\{y:y_1\leq 0\}$ is an invariant set of (2*). If $k_1=0$ then the line $\{y:y_1=0\}$ is an invariant set of (2*). In the case b) if $\det[A,B]=0$ (≥ 0) the system (2*) is not controllable.

We finished the proof of the part 1.

2. $B=kE$.

Sufficient condition:

If $\varrho = \mathrm{Sp}^2 A - 4\det A < 0$ and $d \neq 0$, then $\det A \neq 0$, $AB^{-1}d \neq 0$, there is not any fixed point of (2). By Theorems A and B, the system (2) is controllable.

Necessary condition:

If $d=0$ then $x=0$ is a fixed point of (2) and (2) is not controllable in $\mathbb{R}^2$.

Assume that $\varrho \geq 0$. Let $A_u = A + uB$, here we can take $k=1$, i.e. $B=E$; $\det A_u = u^2 + \mathrm{Sp}\,A + \det A$. Since $\mathrm{Sp}^2 A - 4\det A \geq 0$ we can choose a number α such that $\det A_\alpha = 0$. We can write the system (2*) in the form:

$$\dot{y}(t) = A_\alpha y(t) + u(t)y(t) + k.$$

$\det A_\alpha = 0 \Longrightarrow A_\alpha = bc^T$, b, $c \in \mathbb{R}^2$, $c \neq 0$ (maybe $b=0$). Then

$$c^T \dot{y}(t) = c^T bc^T y(t) + u(t)c^T y(t) + c^T k.$$

If $c^T k > 0$ then the set $\{y : c^T y \geq 0\}$ is an invariant set of (2*). If $c^T k < 0$ then the set $\{y : c^T y \leq 0\}$ is an invariant set of (2*). If $c^T k = 0$, then $\{y : c^T y = 0\}$ is an invariant set of (2*). The system (2*) is not controllable.

3. B has purely imaginary conjugate eigenvalues.

As in the proof of the Theorem A we can assume

$$B = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}, \qquad b < 0.$$

and $$\varrho = \det[A, B] + \det B \cdot \mathrm{Sp}^2 A = -(a_{12}b - a_{21})^2 - 4ba_{11}a_{22}.$$

Necessary condition:

We have only to show that if $\varrho > 0$ then the system (2*) is not controllable (nor is (2)).

Assume that $\varrho > 0$. Let $y(t)$ be a solution of (2*). Then

$$\frac{\mathrm{d}}{\mathrm{d}t}(-by_1^2(t) + y_2^2(t)) = 2(-by_1(t)\dot{y}_1(t) + y_2(t)\dot{y}_2(t)) =$$
$$= 2\{-ba_{11}y_1^2(t) - (a_{12}b - a_{21})y_1(t)y_2(t) + a_{22}y_2^2(t) - bk_1 y_1(t) + k_2 y_2(t)\}.$$

Now we consider the corresponding second degree curve in the plane:

$$-ba_{11}y_1^2 - (a_{12}b - a_{21})y_1 y_2 + a_{22}y_2^2 - bk_1 y_1 + k_2 y_2 = \Gamma(y) = 0.$$

Since $\varrho > 0$, by a well-known fact from linear algebra the curve $\Gamma(y) = 0$ is either a real ellipse, or an imaginary ellipse, or a point ellipse. Here $\Gamma(y)=0$ is a real ellipse, because $0 in \{\Gamma(y)=0\}$ and since $b \neq 0$, $|k_1| + |k_2| > 0$. From this, if c is large enough then the ellipse $-by_1^2 + y_2^2 = c$ is in the outside of the ellipse $\Gamma(y)=0$. This means that $\Gamma(y) > 0$ for all $y \in \{y : -by_1^2 + y_2^2 = c\}$ or

$\Gamma(y)<0$ for all $y\in\{y:-by_1^2+y_2^2=c\}$. In the first case the set $\{y:-by_1^2+y_2^2\geq 0\}$ is invariant set of (2^*), in the second case the set $\{y:-by_1^2+y_2^2\leq c\}$ is invariant set of (2^*). We showed that if $\varrho>0$ then (2^*) is not controllable.

Sufficient condition:

In the case $\varrho<0$ and $AB^{-1}d\neq 0$ the system (2) is controllable by Theorems A and B.

Now we assume $\varrho=0$ and $AB^{-1}d\neq 0$. We will show that the system $\mathcal{F}=\{A+k,\pm B\}$ is controllable (k denotes here the corresponding constant vector field), hence and by Lemma 2, the system $\{A+k+uB:u\in\mathbb{R}\}$ is controllable, so are the systems (2^*) and (2).

At first we state that on every integral curve of B: $\dot y=By$, $y(0)\neq 0$ there is one point at which the integral curve of $A+k$ $(\dot y(t)=Ay(t)+k)$ goes from inside into outside of the integral curve of B and there is another point at which the integral curve of $A+k$ goes from outside into inside.

For an integral curve $y(t)$ $(t\geq 0)$ of $A+k$, consider

$$\frac{\mathrm{d}}{\mathrm{d}t}(-by_1^2(t)+y_2^2(t))=$$
$$=2(-ba_{11}y_1^2(t)-(a_{12}b-a_{21})y_1(t)y_2(t)+a_{22}y_2^2(t)-bk_1y_1(t)+k_2y_2(t))$$

and the corresponding curve

$$-ba_{11}y_1^2-(a_{12}b-a_{21})y_1y_2+a_{22}y_2^2-bk_1y_1+k_2y_2=\Gamma(y)=0.$$

Let $E(c):=\{y:-by_1^2+y_2^2=c\}$ then $E(c)$ is integral curve of B.

a) $a_{11}=a_{22}=0$, then $(a_{12}b-a_{21})=0$ because $\varrho=0$.

In this case $\Gamma(y)=-bk_1y_1+k_2y_2$.

Since $b\neq 0$, $k=AB^{-1}d\neq 0$, $\Gamma(y)=0$ is a line through the origin. Then for all $c>0$, $E(c)\cap\{\Gamma(y)>0\}\neq\emptyset$ and $E(c)\cap\{\Gamma(y)<0\}\neq\emptyset$. This means that for all $c>0$ there is a point of $E(c)$ at which the integral curve of $A+k$ goes from inside into outside of $E(c)$ and there is another point of $E(c)$ at which the integral curve of $A+k$ goes from outside into inside of $E(c)$.

b) $a_{22}\neq 0$.

Since $\varrho=0$ we can write

$$\Gamma(y)=a_{22}\left(y_2-\frac{a_{12}b-a_{21}}{2a_{22}}y_1\right)^2+k_2\left(y_2-\frac{a_{12}b-a_{21}}{2a_{22}}y_1\right)+$$
$$+\left(k_2\cdot\frac{a_{12}b-a_{21}}{2a_{22}}-bk_1\right)y_1=$$
$$=a_{22}z_2^2+k_2z_2+\left(k_2\cdot\frac{a_{12}b-a_{21}}{2a_{22}}-bk_1\right)y_1$$

where $z_2 = y_2 - \frac{a_{12}b - a_{21}}{2a_{22}} y_1$.

We see that if $k_2 \neq 0$ and $k_2 \cdot \frac{a_{12}b-a_{21}}{2a_{22}} = bk_1$, then $\Gamma(y)=0$ consists of two parallel lines: $z_2 = 0$ and $z_2 = -\frac{k_2}{a_{22}}$. The line $z_2 = 0$ goes through the origin and $z_2 = -\frac{k_2}{a_{22}}$ does not. Besides, sine $k_1^2 + k_2^2 \neq 0$, the curve $\Gamma(y)=0$ is a parabola which goes through the origin. In both cases, we have $E(c) \cap \{\Gamma(y) > 0\} \neq \emptyset$ and $E(c) \cap \{\Gamma(y) < 0\} \neq \emptyset$ for all $c > 0$.

c) $a_{11} \neq 0$. Similar to the case b).

Similarly we can prove the above assertion for $-(A+k)$, too.

Now we show that $\mathbb{A}_{\mathcal{F}}(0) = \mathbb{R}^2$.

We assume indirectly that $\mathbb{A}_{\mathcal{F}}(0) \neq \mathbb{R}^2$. Let $z \notin \mathbb{A}_{\mathcal{F}}(0)$. Then $z \in E(c(z))$: $-bz_1^2 + z_2^2 = c(z)$, $c(z) \geq 0$. Let $c = \inf\{c(z) : z \notin \mathbb{A}_{\mathcal{F}}(0)\}$. We mention that if $z \notin \mathbb{A}_{\mathcal{F}}(0)$ then $E(c(z)) \cap \mathbb{A}_{\mathcal{F}}(0) = \emptyset$. Indeed, if $z' \in E(c(z)) \cap \mathbb{A}_{\mathcal{F}}(0)$, then z can be reached from z' (along a trajectory $\dot{y} = By$, $y(0) = z'$, i.e. along $E(c(z))$ and hence z can be reached from the origin by $\mathcal{F}$, what is contradictory. Further, $\{y : -by_1^2 + y_2^2 \geq c(z)\} \cap \mathbb{A}_{\mathcal{F}}(0) = \emptyset$. Because if there is a common point, then the trajectory, which steers the origin into this point, meets surely the curve $E(c(z))$ and hence $E(c(z))$ can be reached from the origin, in contradiction with the assumption that $z \notin \mathbb{A}_{\mathcal{F}}(0)$.

If $c = 0$, then 0 is fixed point of $\mathcal{F}$, in contradiction with the assumption that $\mathcal{F}$ has no fixed point. Hence $c > 0$.

If $E(c) \subset \mathbb{A}_{\mathcal{F}}(0)$, then by the above, $E(c) \cap \{\Gamma(y) > 0\} \neq \emptyset$ and there is a point on $E(c)$, at which the trajectory $\dot{y} = Ay + k$ goes from inside into outside of $E(c)$, hence we can steer the origin to a point in $\{y : -by_1^2 + b_2^2 > c\}$, in contradiction with the assumption that c is inf.

If $E(c) \cap \mathbb{A}_{\mathcal{F}}(0) = \emptyset$, then there is a point on $E(c)$, at which the trajectory $\dot{y} = -Ay - k$ goes from outside into inside of $E(c)$, hence $E(c)$ can be reached from a point in $\{y : -by_1^2 + y_2^2 < c\}$ and so from the origin, which is contradicting with $E(c) \cap \mathbb{A}_{\mathcal{F}}(0) = \emptyset$.

We proved that $\mathbb{A}_{\mathcal{F}}(0) = \mathbb{R}^2$. We can similarly prove that $\mathbb{A}_{-\mathcal{F}}(0) = \mathbb{R}^2$.

Now let x, y be arbitrary in $\mathbb{R}^2$. Then x can be reached from 0 by $-\mathcal{F}$ and hence 0 can be reached from x by $\mathcal{F}$. y can be reached from 0, hence y can be reached from x by $\mathcal{F}$. This shows that the system $\mathcal{F}$ is controllable that we wanted to prove.

4. B has complex conjugate eigenvalues with nonzero real part.

In the case $B \neq \lambda A$ and $AB^{-1}d \neq 0$, the system is controllable by Theorems A and B.

Now we assume that $B=\lambda A$ and $AB^{-1}d\neq 0$, i.e. $d\neq 0$. The system (2) can be written as

$$\dot{x}(t)=\left(\frac{1}{\lambda}+u(t)\right)Bx(t)+u(t)d.$$

Write here $u(t)-\frac{1}{\lambda}$ instead of $u(t)$; then $y(t)=x(t)+B^{-1}d$ satisfies

$$(2^{**})\qquad \dot{y}(t)=u(t)By(t)-\frac{1}{\lambda}d=u(t)By(t)+k\qquad \left(k=-\frac{1}{\lambda}d\right).$$

We will show that for arbitrary points x and y in $\mathbb{R}^2$, x can be steered into the origin and y can be reached from the origin hence y can be reached from x (by (2^{**})).

If the eigenvalues of B have positive real part, then the curve $e^{Bs}k$ tends to infinity as $s\to\infty$ ($k\neq 0$), goes around the origin, makes infinitely many complete rotations around the origin. The curve $e^{Bs}k$ ($s>0$) meets the half line $\{uBy+k:u>0\}$ in some point, if $y\neq 0$: $uBy+k=e^{Bs}k$ for some $u>0$, $s>0$. Let $y(t)=\frac{1}{u}B^{-1}e^{uBt}k-\frac{1}{u}B^{-1}k$. Then $\dot{y}(t)=uBy(t)+k$ and for $t=\frac{1}{u}\cdot s>0$

$$y(t)=\frac{1}{u}B^{-1}\cdot e^{Bs}k-\frac{1}{u}B^{-1}k=\frac{1}{u}B^{-1}(uBy+k)-\frac{1}{u}B^{-1}k=y.$$

This shows that y can be reached from 0.

If the eigenvalues of B have negative real part, then the curve $\{e^{-Bs}k;\ s>0\}$ meets the half line $\{uBy+k:u<0\}$: $uBy+k=e^{-Bs}k$ for some $u<0$, $s>0$. The curve $y(t)=\frac{1}{u}B^{-1}e^{uBt}k-\frac{1}{u}B^{-1}k$ satisfies

$$\dot{y}(t)=uBy(t)+k,\qquad y\left(-\frac{s}{u}\right)=y,\qquad \left(-\frac{s}{u}>0\right);$$

y can be reached from 0.

If $y=0$, then 0 can be reached from 0, too. Similarly, we can see that x can be reached from 0 by the system $\dot{z}(t)=-u(t)Bz(t)-k$. Hence x can be steered into 0 by (2^{**}).

This completed the proof of the Theorem 2. ∎

COROLLARY. *Consider the following system*

$$(3)\qquad \dot{x}(t)=Ax(t)+a+u(t)(Bx(t)+d)$$

in $\mathbb{R}^2$ *with* $\operatorname{rank}B=2$. *The necessary and sufficient conditions of controllability of* (3) *are:*

1. $\det[A,B]<0$ *and* $AB^{-1}d\neq a$ *if* $B\neq kE$ *and* B *has real eigenvalues*
2. $\operatorname{Sp}^2A-4\det A<0$ *and* $Ad\neq\lambda a$ *if* $B=\lambda E$

3. $\det[A,B]+\det B\cdot \mathrm{Sp}^2 A\leq 0$ *and* $AB^{-1}d\neq a$ *if* B *has purely imaginary conjugate eigenvalues*

4. $AB^{-1}d\neq a$ *if* B *has complex conjugate eigenvalues with nonzero real part.*

PROOF. We can write the system (3) as

$$\dot{y}(t)=Ay(t)+u(t)By(t)+k \tag{3*}$$

where $y(t)=x(t)+B^{-1}d$ and $k=a-AB^{-1}d$.

Then we can apply the proof of the Theorem 2 to this case. ∎

In what follows we study the controllability of the system (2) when $\operatorname{rank}B=1$. We mention that in the case $\operatorname{rank}B=0$ ($B=0$) the system (2) is a linear system, and the well-known Kalman rank condition has been given for the controllability: The system (2) with $B=0$ is controllable in $\mathbb{R}^n$ if and only if $\operatorname{rank}[d,Ad,\ldots,A^{n-1}d]=n$.

If $\operatorname{rank}B=1$, then $B=bc^T$ for some $b\neq 0$, $c\neq 0$ in $\mathbb{R}^2$ and in a suitable basis B can be written as

$$B=\begin{pmatrix} c^Tb & 0\\ 0 & 0\end{pmatrix} \quad \text{if } c^Tb\neq 0, \qquad \text{or} \quad B=\begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix} \quad \text{if } c^Tb=0.$$

We prove first the following result:

THEOREM 3. *Consider the system* (2)

$$\dot{x}(t)=Ax(t)+u(t)Bx(t)+u(t)d \tag{2}$$

in $\mathbb{R}^2$ *with* $\operatorname{rank}B=1$, $B=bc^T$.

I. *Assume that* $\det[A,B]<0$. *Then the system* (2) *is controllable in* $\mathbb{R}^2$ *if and only if either*

1. d *and* b *are linearly independent, i.e.* $d\neq\lambda b$; *or*
2. $0\neq d$ *and* b *are linearly dependent and* A *is nonsingular; or*
3. $0\neq d$ *and* b *are linearly dependent and* A *is singular with* $A=gc^T$ *for some* $g\in\mathbb{R}^2$.

II. *If* $\det[A,B]>0$ *then the system* (2) *is not controllable in* $\mathbb{R}^2$.

PROOF. I. By the Theorems 1 and B in the case $\det[A,B]<0$ the system (2) is controllable in $\mathbb{R}^2$ if and only if (2) has no fixed points in $\mathbb{R}^2$.

(2) has a fixed point $\Longleftrightarrow$ there exists an x such that $Ax=0$ and $Bx=-d$.

1. If d and b are linearly independent then $Bx=bc^Tx\neq -d$ for all $x\in\mathbb{R}^2$, what shows that (2) has no fixed points.

2. If A is nonsingular then $Ax=0\Longrightarrow x=0$ and $Bx=0\neq d$ because $d\neq 0$. In this case (2) has no fixed points.

3. If A is singular and $A=gc^T$ for some $g\in\mathbb{R}^2$, then $Ax=0\Longrightarrow c^Tx=0$ (here $g\neq 0$ since we supposed that $\det[A,B]<0$) with this x $Bx=0\neq d$, hence (2) has no fixed points.

If $A=gf^T$ with $f\neq\lambda c$, then there is x such that $Ax=0$ and $c^Tx\neq 0$, i.e. $f^Tx=0$, $x\neq 0\Longrightarrow c^Tx\neq 0$. If $-d=\mu b$, $\mu\neq 0$ then for $y=\frac{\mu}{c^Tx}x$ $Ay=0$ and $By=bc^Ty=\mu b=-d$. This means that y is a fixed point of (2) and the system (2) is not controllable in $\mathbb{R}^2$.

We finished the proof of the part I.

II. Since $\operatorname{rank} B=1$, by the remark at the beginning of the proof of the Theorem 2 and by the Lemma 1, we can assume that

$$B=\begin{pmatrix} c^Tb & 0\\ 0 & 0\end{pmatrix}\quad \text{if } c^Tb\neq 0 \quad\text{or}\quad B=\begin{pmatrix} 0 & 1\\ 0 & 0\end{pmatrix}\quad \text{if } c^Tb=0.$$

Assume that $\det[A,B]>0$. Let

$$A=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}.$$

Then $\det[A,B]=a_{12}a_{21}$ in the first case and $\det[A,B]=-a_{21}^2$ in the second case. The assumption that $\det[A,B]>0$ holds only in the case $c^Tb\neq 0$. Since $c^Tb\neq 0$ and $u(t)$ takes the values on $\mathbb{R}$, we can also assume that $c^Tb=1$. We have

$$\dot{x}(t)=Ax(t)+u(t)\begin{pmatrix} 1 & 0\\ 0 & 0\end{pmatrix}x(t)+u(t)d.$$

By Lemma 2 this system is controllable in $\mathbb{R}^2$ if and only if the system $\mathcal{F}=\{A,\ \pm(B+d)\}$ is.

We will show that if $\det[A,B]>0$ then the system $\mathcal{F}$ is not controllable.

a) $d=\begin{pmatrix} d_1\\ 0\end{pmatrix}$.

We can suppose $d_1\neq 0$; otherwise the system is homogeneous, i.e. not controllable. We calculate the trajectory of the vector field $B+d$:

From
$$\begin{cases} \dot{x}_1(t)=x_1(t)+d_1\\ \dot{x}_2(t)=0\end{cases}$$

we have
$$x_1(t)=e^t(x_1(0)+d_1)-d_1$$
$$x_2(t)=x_2(0).$$

Here we can take $t\in\mathbb{R}$ because $\pm(B+d)\in\mathcal{F}$. The trajectories of $\pm(B+d)$ are parallel half lines:

$$\{x: x_2=\text{const},\ x_1>-d_1\}\qquad \{x: x_2=\text{const},\ x_1<-d_1\}$$

The points of the line $\{x_1 = -d_1\}$ are fixed points of $(B+d)$. Since $\det[A,B] > 0$, $a_{12}a_{21} > 0$. The integral curves of A satisfy

$$\begin{cases} \dot{x}_1(t) = a_{11}x_1(t) + a_{12}x_2(t) \\ \dot{x}_2(t) = a_{21}x_1(t) + a_{22}x_2(t). \end{cases}$$

There are eight cases we have to consider.

a$_1$) $a_{12} > 0$, $a_{21} > 0$, $\frac{a_{11}d_1}{a_{12}} \geq 0$, $d_1 > 0$.

From the first equation, at the point x for which $x_1 = -d_1$, $x_2 \geq \frac{a_{11}d_1}{a_{12}}$, the trajectory of A doesn't go into the left side of the line $\{x_1 = -d_1\}$. If $x \in \left\{x_1 = -d_1,\ x_2 < \frac{a_{11}d_1}{a_{12}}\right\}$, the trajectory of A doesn't go from x into the right side of the line $\{x_1 = -d_1\}$. Similarly from the second equation, if $x \in \{x_1 \leq 0,\ x_2 = 0\}$ then the trajectory of A does not go from x into the half plane $\{x : x_2 > 0\}$. Since $d_1 > 0$, $\frac{a_{11}d_1}{a_{12}} \geq 0$ the set $\{x : x_1 < -d_1,\ x_2 < 0\}$ is an invariant set of $\mathcal{F}$.

a$_2$) $d_1 > 0$, $a_{12} > 0$, $a_{21} > 0$, $\frac{a_{11}d_1}{a_{12}} < 0$.

The set $\mathbb{R}^2 \setminus \{x : x_1 < -d_1,\ x_2 > 0\}$ is an invariant set of $\mathcal{F}$.

a$_3$) $d_1 > 0$, $a_{12} < 0$, $a_{21} < 0$, $\frac{a_{11}d_1}{a_{12}} \geq 0$.

The set $\mathbb{R}^2 \setminus \{x_1 < -d_1,\ x_2 < 0\}$ is an invariant set of $\mathcal{F}$.

a$_4$) $d_1 > 0$, $a_{12} < 0$, $a_{21} < 0$, $\frac{a_{11}d_1}{a_{12}} < 0$.

The set $\{x : x_1 \leq -d_1,\ x_2 \geq 0\}$ is an invariant set of $\mathcal{F}$.

a$_5$) $d_1 < 0$, $a_{12} > 0$, $a_{21} > 0$, $\frac{a_{11}d_1}{a_{12}} \geq 0$.

The set $\mathbb{R}^2 \setminus \{x_1 > -d_1,\ x_2 < 0\}$ is an invariant set of $\mathcal{F}$.

a$_6$) $d_1 < 0$, $a_{12} > 0$, $a_{21} > 0$, $\frac{a_{11}d_1}{a_{12}} < 0$.

The set $\{x : x_1 \geq -d_1,\ x_2 \geq 0\}$ is $\mathcal{F}$-invariant.

a$_7$) $d_1 < 0$, $a_{12} < 0$, $a_{21} < 0$, $\frac{a_{11}d_1}{a_{12}} \geq 0$.

The set $\{x : x_1 > -d_1,\ x_2 < 0\}$ is $\mathcal{F}$-invariant.

a$_8$) $d_1 < 0$, $a_{12} < 0$, $a_{21} < 0$, $\frac{a_{11}d_1}{a_{12}} < 0$.

The set $\mathbb{R}^2 \setminus \{x : x_1 > -d_1,\ x_2 > 0\}$ is $\mathcal{F}$-invariant.

We showed that if $d_2 = 0$ then the system $\mathcal{F}$ is not controllable.

b) $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$, $d_2 \neq 0$.

The integral curves of $(B+d)$ satisfy the equations

$$\begin{cases} \dot{x}(t)=x_1(t)+d_1 \\ \dot{x}_2(t)=d_2 \end{cases}$$

from which we obtain

$$x_1(t)=e^t(x_1(0)+d_1)-d_1 \qquad x_2(t)=d_2t+x_2(0).$$

Since we take $t\in\mathbb{R}$, we can assume always that $x_2(0)=0$. The integral curve of $(B+d)$ through $(x_1(0),0)$ satisfies

$$x_1(t)=e^{\frac{x_2(t)}{d_2}}(x_1(0)+d_1)-d_1.$$

Every integral of $(B+d)$ divides the plane into two parts

$$\Gamma_+(x_1(0)):=\left\{x: x_1\geq e^{\frac{x_2}{d_2}}(x_1(0)+d_1)-d_1\right\}$$

and

$$\Gamma_-(x_1(0)):=\left\{x: x_1\leq e^{\frac{x_2}{d_2}}(x_1(0)+d_1)-d_1\right\}.$$

We denote also

$$\Gamma(x_1(0)):=\left\{x_1=e^{\frac{x_2}{d_2}}(x_1(0)+d_1)-d_1\right\}.$$

Now we consider integral curves of A:

$$\text{(4)} \qquad \begin{cases} \dot{x}(t)=Ax(t), \quad t\geq 0 \\ x(0)\in\Gamma(\alpha). \end{cases}$$

We will show that if $\det[A,B]>0$ then there is α such that the set $\Gamma_+(\alpha)$ or the set $\Gamma_-(\alpha)$ is an invariant set of A, i.e. if $x(0)\in\Gamma(\alpha)$ and $x(t)$ satisfies (4) then $x(t)\in\Gamma_+(\alpha)$ or $x(t)\in\Gamma_-(\alpha)$ resp. for all $t\geq 0$.

Let $x(t)$ be a solution of (4) and

$$y(t)=x_1(t)-e^{\frac{x_2(t)}{d_2}}(\alpha+d_1)+d_1.$$

Then $y(0)=0$

$$\dot{y}(t)=\dot{x}_1(t)-\frac{\dot{x}_2(t)}{d_2}\cdot e^{\frac{x_2(t)}{d_2}}(\alpha+d_1).$$

Since $\dot{x}=Ax$, we obtain

$$\dot{y}(t)=a_{11}x_1(t)+a_{12}x_2(t)-\frac{a_{21}x_1(t)+a_{22}x_2(t)}{d_2}\cdot e^{\frac{x_2(t)}{d_2}}\cdot K$$

where $K=\alpha+d_1$. Since $y(0)=0$, with $x(0)=(x_1,x_2)$ we have

$$\dot y(0)=a_{11}e^{\frac{x_2}{d_2}}K-a_{11}d_1+a_{12}x_2-\frac{a_{21}}{d_2}e^{2\cdot\frac{x_2}{d_2}}K^2+\frac{a_{21}d_1}{d_2}e^{\frac{x_2}{d_2}}\cdot K-\frac{a_{22}x_2}{d_2}e^{\frac{x_2}{d_2}}\cdot K$$

$$(5)\quad \dot y(0)=-\frac{a_{21}}{d_2}e^{\frac{2x_2}{d_2}}K^2+\left\{-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\right\}e^{\frac{x_2}{d_2}}K+a_{12}x_2-a_{11}d_1.$$

We assert that there is K (hence α) such that $\dot y(0)>0$ for all x_2 (or $\dot y(0)<0$ for all x_2), hence at all points of $\Gamma(\alpha)$ the trajectories of A go into $\Gamma_+(\alpha)$ (or go into $\Gamma_-(\alpha)$), i.e. $\Gamma_+(\alpha)$ is A-invariant (or $\Gamma_-(\alpha)$ is A-invariant).

b$_1$) $d_2>0$, $a_{12}>0$, $a_{21}>0$.

(*) $a_{22}>0$: Let $\varrho=\frac{a_{11}d_2+a_{21}d_1}{a_{22}}$, $\quad \nu=\frac{a_{11}d_1}{a_{12}}$.

then for all $x_2\le\min(\varrho,\nu)$

$$-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\ge 0,\qquad a_{12}x_2-a_{11}d_1\le 0.$$

Now we consider $\dot y(0)$ as a polynomial of K then its discriminant can be calculated by

$$\Delta(x_2)=e^{\frac{2x_2}{d_2}}\left\{\left(-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\right)^2+4\frac{a_{21}}{d_2}(a_{12}x_2-a_{11}d_1)\right\}.$$

If $x_2\ge\min\{\varrho,\nu\}$ and $\Delta(x_2)<0$, then $\dot y(0)<0$ for all K. If $x_2\ge\min\{\varrho,\nu\}$ and $\Delta(x_2)\ge 0$, then $\dot y(0)<0$ for all $K<K_1(x_2)$ and $K>K_2(x_2)$, where $K_1(x_2)\le K_2(x_2)$ are solutions of the equation $\dot y(0)=0$.

We show that there is $M>0$ such that $|K_i(x_2)|<M$, $i=1,\,2$ for all $x_2\in\{x_2\ge\min\{\varrho,\nu\}$ and $\Delta(x_2)\ge 0\}=H$. Indeed

$$K_{1,2}(x_2)=\frac{-e^{\frac{x_2}{d_2}}\left\{a_{11}+\frac{a_{21}d_1}{d_2}-\frac{a_{22}}{d_2}x_2\right\}\pm\sqrt{\Delta}}{-2\frac{a_{21}}{d_2}\cdot e^{\frac{2x_2}{d_2}}}$$

and $K_{1,2}(x_2)\to 0$ as $x_2\to\infty$ ($\Delta>0$ as $x_2\to\infty$). For $\varepsilon>0$ $\exists N$: $x_2\ge N$ $\Delta(x_2)>0$ and $|K_{1,2}(x_2)|<\varepsilon$ and $K_{1,2}(x_2)$ is bounded on $\{x_2\le N\}\cap H$. Hence there is $M>0$ such that $|K_{1,2}(x_2)|<M$ for all $x_2\in H$. Now let $K<-M$, then for all $x_2\in H$, $K<K_{1,2}(x_2)$ and in this case $\dot y(0)<0$. With this K, if $x_2<\min(\varrho,\nu)$, then $\dot y(0)<0$, too. This shows that the trajectories of A started at the points of $\Gamma(K-d_1)$ go into the set $\Gamma_-(K-d_1)$, the set $\Gamma_-(K-d_1)$ is A-invariant and hence $\mathcal{F}$-invariant, too.

(*) $a_{22}<0$: if $x_2\le\min(\varrho,\nu)$ then

$$-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\le 0,\qquad a_{12}x_2-a_{11}d_1\le 0.$$

In this case we choose $K > M$ and with this K the set $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22} = 0$.

In this case we take $\varrho=\nu=\frac{a_{11}d_1}{a_{12}}$ and we choose $K<-M$ if $a_{11}+\frac{a_{21}d_1}{d_2}\geq 0$, and $K > M$ if $a_{11}+\frac{a_{21}d_1}{d_2} < 0$. Then the set $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant.

b$_2$) $d_2 > 0$, $a_{12} < 0$, $a_{21} < 0$.

(∗) $a_{22} > 0$.

If $x_2 \leq \min\{\varrho,\nu\}$ then

$$-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\geq 0, \qquad a_{12}x_2-a_{11}d_1\geq 0.$$

In this case we can show that (similarly to the case b$_1$)) there is $M > 0$ such that $x_2 \in H \Longrightarrow |K_{1,2}(x_2)| < M$. Hence if $K > M$ then $\dot{y}(0) > 0$ for all $x_2 \in H$. If $x_2 \notin H$, since $K>0$, $\dot{y}(0)>0$. This shows that the set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22} < 0$.

Here we choose $K < -M$, then $\dot{y}(0) > 0$ for all x_2, this means that $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22} = 0$.

Let $\varrho=\nu=\frac{a_{11}d_1}{a_{12}}$, and let $K > M$ if $a_{11}+\frac{a_{21}d_1}{d_2}\geq 0$ and let $K < -M$ if $a_{11}+\frac{a_{21}d_1}{d_2} < 0$. Then the set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

b$_3$) $d_2 < 0$, $a_{12} > 0$, $a_{21} > 0$.

Now $e^{\frac{x_2}{d_2}} \to \infty$ as $x_2 \to \infty$. In this case we can show similarly that there is $M>0$ such that $|K_{1,2}(x_2)|<M$ for all $x_2\in H=\{x_2\leq\max\{\varrho,\nu\},\Delta(x_2)\geq 0\}$

(∗) $a_{22} > 0$.

If $x_2 \geq \max\{\varrho,\nu\}$ then

$$-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\geq 0 \qquad a_{12}x_2-a_{11}b_1\geq 0.$$

We choose $K > M$. Then $\dot{y}(0) > 0$ for all $x(0)=(x_1,x_2)\in\Gamma(K-d_1)$. The set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22} < 0$.

If $x_2 \geq \max\{\varrho,\nu\}$ then

$$-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\geq 0 \qquad a_{12}x_2-a_{11}b_1\leq 0.$$

We choose $K < -M$. The set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22}=0$.

Let $\varrho=\nu=\frac{a_{11}b_1}{a_{12}}$.

We choose $K>M$ if $a_{11}+\frac{a_{21}d_1}{d_2}\geq 0$ and $K<-M$ if $a_{11}+\frac{a_{21}d_1}{d_2}<0$. Then the set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

b4) $d_2<0$, $a_{12}<0$, $a_{21}<0$.

(∗) $a_{22}>0$.

We choose $K<-M$, then $\dot{y}(0)<0$ for all $x(0)=(x_1,x_2)\in\Gamma(K-d_1)$. The set $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22}<0$.

We choose $K>M$. The set $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant.

(∗) $a_{22}=0$.

We choose $K<-M$ if $a_{11}+\frac{a_{21}d_1}{d_2}\geq 0$ and $K>M$ if $a_{11}+\frac{a_{21}d_1}{d_2}<0$. Then the set $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant.

In every case we showed that there is an $\mathcal{F}$-invariant set. This means that the system $\mathcal{F}$, so the system (2) is not controllable in $\mathbb{R}^2$ if $\det[A,B]>0$.

We finished completely the proof of the Theorem 3. ∎

In the case of homogeneous systems $(d=0)$ it was showed that if $\operatorname{rank}B=1$ and $\det[A,B]=0$ then the system is not controllable in $\mathbb{R}^2\setminus\{0\}$. Here, when the system is not homogeneous $(d\neq 0)$, in the case $\operatorname{rank}B=1$ and $\det[A,B]=0$ the system may be controllable, may be not controllable in $\mathbb{R}^2$.

THEOREM 4. *Consider the system*

(2) $$\dot{x}(t)=Ax(t)+u(t)Bx(t)+u(t)d$$

with $B=bc^T\neq 0$, $d\neq 0$, $\det[A,B]=0$.

Assume that $c^Tb\neq 0$, *then the system* (2) *is controllable in* $\mathbb{R}^2$ *if and only if either*

1. $\det[b,d]\neq 0$, $\det\begin{bmatrix}c^T\\ c^TA\end{bmatrix}\neq 0$ *and* $\det A=\operatorname{Sp}A=0$; *or*

2. $\det[b,d]\neq 0$, $\det\begin{bmatrix}c^T\\ c^TA\end{bmatrix}\neq 0$, $\det A\neq 0$ *and*

$$\frac{\left(c^TAb\right)^2}{\det A}>\frac{\left(c^Tb\right)^2c^TABd}{\det\begin{bmatrix}c^T\\ c^TA\end{bmatrix}\cdot\det[b,d]}.$$

PROOF. As usual we can assume that $B=\begin{pmatrix} c^T b & 0 \\ 0 & 0 \end{pmatrix}$.

Let $A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

Then $\det[A,B]=0 \Longleftrightarrow a_{12}a_{21}=0$. We can also assume that $c^T b=1$ Here we consider the system $\mathcal{F}=\{A,\pm(B+d)\}$ instead of (2), too.

I. $d=\begin{pmatrix} d_1 \\ 0 \end{pmatrix}$.

a) $a_{12}=0$.

Consider trajectories of (2) starting from the line $\{x:x_1=-d_1\}$. Then $\dot{x}_1(0)=-a_{11}d_1$. One can see that if $-a_{11}d_1>0$ then the half plane $\{x:x_1\geq -d_1\}$ is $\mathcal{F}$-invariant; if $-a_{11}d_1<0$ then the half plane $\{x:x_1\leq -d_1\}$ is $\mathcal{F}$-invariant; if $-a_{11}d_1=0$ then the line $\{x:x_1=-d_1\}$ is $\mathcal{F}$-invariant.

b) $a_{12}\neq 0$.

Then $a_{21}=0$. In this case the line $\{x:x_2=0\}$ is $\mathcal{F}$-invariant. Hence if $d=\begin{pmatrix} d_1 \\ 0 \end{pmatrix}$ and $\det[A,B]=0$ then the system (2) is not controllable

II. $d=\binom{d_1}{d_2}, d_2\neq 0$.

It was shown in the proof of the Theorem 3 that the integral curve of $(B+d)$ through the point $(\alpha,0)$ satisfies

$$x_1(t)=e^{\frac{x_2(t)}{d_2}}(\alpha+d_1)-d_1.$$

Here we use also the notations $\Gamma(\alpha)$, $\Gamma_+(\alpha)$, $\Gamma_-(\alpha)$, $y(t)$, K. We mention that

$$y(t)=x_1(t)-e^{\frac{x_2(t)}{d_2}}K+d_1$$

here $K=\alpha+d_1$.

We consider trajectories of A which start from the curve $\Gamma(\alpha)$.

$$(4)\qquad \begin{cases} \dot{x}(t)=Ax(t), & t\geq 0 \\ x(0)\in\Gamma(\alpha). \end{cases}$$

If $x(t)$ satisfies (4) then the corresponding $y(t)$ satisfies

$$\dot{y}(0)=-\frac{a_{21}}{d_2}e^{\frac{2x_2}{d_2}}K^2+\left\{-\frac{a_{22}}{d_2}x_2+a_{11}+\frac{a_{21}d_1}{d_2}\right\}e^{\frac{x_2}{d_2}}K+a_{12}x_2-a_{11}d_1$$

a) $a_{12}=0$.

Let $\alpha=-d_1$, then $K=0$ and $\dot{y}(0)=-a_{11}d_1$. From this we have:

If $a_{11}d_1 = 0$ then the line $\{x_1 = -d_1\}$ is $\mathcal{F}$-invariant.
If $a_{11}d_1 > 0$ then the half plane $\{x_1 \leq -d_1\}$ is $\mathcal{F}$-invariant.
If $a_{11}d_1 < 0$ then the half plane $\{x_1 \geq -d_1\}$ is $\mathcal{F}$-invariant.
In this case the system (2) is not controllable.

b) $a_{12} \neq 0$.

Then $a_{21} = 0$. We have

$$\dot{y}(0) = e^{\frac{x_2}{d_2}} K \left\{ -\frac{a_{22}}{d_2} x_2 + a_{11} \right\} + a_{12}x_2 - a_{11}d_1.$$

b$_1$) $a_{22} = 0$, $a_{11} \neq 0$. $\dot{y}(0) = e^{\frac{x_2}{d_2}} K a_{11} + a_{12}x_2 - a_{11}d_2$.

(*) $a_{12}d_2 > 0$.

We can choose $Ka_{11} \ll 0$ (i.e. $Ka_{11} \sim -\infty$) such that $\dot{y}(0) < 0$ for all x_2, with this K the set $\Gamma_-(K - d_1)$ is $\mathcal{F}$-invariant.

(*) $a_{12}d_2 < 0$.

We can choose $Ka_{11} \gg 0$ (i.e. $Ka_{11} \sim +\infty$) such that $\dot{y}(0) > 0$ for all x_2, with this K the set $\Gamma_+(K - d_1)$ is $\mathcal{F}$-invariant.

In the case b$_1$) the system (2) is not controllable.

b$_2$) $a_{11} = 0$, $a_{22} \neq 0$.

Let $K = \frac{a_{12}d_2}{a_{22}}$, then

$$\dot{y}(0) = a_{12}x_2 \left(1 - e^{\frac{x_2}{d_2}} \right).$$

Here we can see that if $a_{12}d_2 > 0$ then $\dot{y}(0) < 0$ for all $x_2 \neq 0$ and $\Gamma_-(K - d_1) \setminus \Gamma(K - d_1)$ is $\mathcal{F}$-invariant; if $a_{12}d_2 < 0$ then $\dot{y}(0) > 0$ for all $x_2 \neq 0$ and $\Gamma_+(K - d_1) \setminus \Gamma(K - d_1)$ is $\mathcal{F}$-invariant (we took into consideration that the solution of (4) in case $x_2 = 0$ is a point: $(\alpha, 0)$).

In this case the system (2) is not controllable.

b$_3$) $a_{11} = 0$, $a_{22} = 0$.

In this case the integral curves of A satisfy

$$\begin{cases} x_1(t) = a_{12}x_2(0)t + x_1(0) \\ x_2(t) = x_2(0). \end{cases}$$

They are parallel half lines $\{x_2 = \text{const}\}$, going from left to right and $x_1(t) \to \infty$ as $t \to \infty$ if $a_{12}x_2(0) > 0$, going from right to left and $x_1(t) \to -\infty$ as $t \to \infty$ if $a_{12}x_2(0) < 0$, the points of the line $\{x : x_2 = 0\}$ are fixed points of A. By this one can see that $\mathcal{F}$ is controllable in $\mathbb{R}^2$. Indeed, let x, y be arbitrary in $\mathbb{R}^2$. Then $x \in \Gamma(\alpha)$, $y \in \Gamma(\beta)$. If $\alpha = \beta$ then x can be steered into y on a trajectory of $\pm(B + d)$. If $\alpha < \beta$, then along the trajectory

$\Gamma(\alpha)$ of $\pm(B+d)\,x$ can be steered into z with $a_{12}z_2(0)>0$. The trajectory $\dot{x}=Ax$, $x(0)=z$ steers z into a point of $\Gamma(\beta)$ from which y can be reached along $\Gamma(\beta)$. If $\alpha>\beta$ then trajectory $\Gamma(\alpha)$ of $\pm(B+d)$ steers x into z with $a_{12}z_2(0)<0$. Hence y can similarly be reached from z.

In the case b$_3$) the system (2) is controllable.

b$_4$) $a_{11}\neq 0$, $a_{22}\neq 0$.

Here we assume $x(t)$ satisfies (4), then

$$\dot{y}(0)=e^{\frac{x_2}{d_2}}K\left\{-\frac{a_{22}}{d_2}x_2+a_{11}\right\}+a_{12}x_2-a_{11}d_1. \tag{6}$$

Now we consider $\dot{y}(0)$ as a function of x_2: $f(x_2):=\dot{y}(0)$ then

$$f'(x_2)=\frac{1}{d_2^2}e^{\frac{x_2}{d_2}}K\left\{-a_{22}x_2+a_{11}d_2-a_{22}d_2\right\}+a_{12}.$$

$f'(x_2)=0$ if and only if

$$K\left\{a_{22}x_2-a_{11}d_2+a_{22}d_2\right\}=d_2^2\cdot a_{12}e^{-\frac{x_2}{d_2}}.$$

b$_4\alpha$) $d_2>0$, $a_{12}>0$.

(*) $\frac{d_2a_{11}}{a_{22}}<\frac{a_{11}d_1}{a_{12}}$.

We assume that $Ka_{22}>0$. Then there is only one z such that $f'(z)=0$ because $K\{a_{22}x_2-a_{11}d_2+a_{22}d_2\}$ is strictly monotonously increasing, $d_2^2\cdot a_{12}e^{-\frac{x_2}{d_2}}$ is strictly monotonously decreasing. If $f'(z)=0$ then $f'(x_2)>0$ if $x_2<z$ and $f'(x_2)<0$ if $x_2>z$, hence $f(x_2)$ attains its global maximum at $x_2=z$.

Now we try to find K, z such that

$$Ka_{22}>0, \qquad f'(z)=0, \qquad f(z)<0.$$

Since $\frac{d_2a_{11}}{a_{22}}<\frac{a_{11}d_1}{a_{12}}$, there is z such that $\frac{d_2a_{11}}{a_{22}}<z<\frac{a_{11}d_1}{a_{12}}$ with this z $f(z)<0$, which follows from (6) and from the conditions of K, z, d_2, a_{12}. Now we calculate the value of K from

$$K\left\{a_{22}z-a_{11}d_2+a_{22}d_2\right\}=d_2^2\cdot a_{12}\cdot e^{-\frac{z}{d_2}} \tag{7}$$

and one can see here $Ka_{22}>0$ and $f'(z)=0$. We showed that we can choose $\frac{a_{11}d_2}{a_{22}}<z<\frac{a_{11}d_1}{a_{12}}$ and $Ka_{22}>0$ such that $f'(z)=0$, $f(z)<0$. Since $f(x_2)$ attains its global maximum at $x_2=z$, $f(x_2)<0$, i.e. $\dot{y}(0)<0$ for all x_2. This shows that $\Gamma_-(K-d_1)$ is $\mathcal{F}$-invariant, the system (2) is not controllable.

(*) $\frac{d_2a_{11}}{a_{22}}>\frac{d_1a_{11}}{a_{12}}$.

Then for $z=\frac{a_{11}d_2}{a_{22}}$, $f(z)=\frac{a_{12}a_{11}d_2}{a_{22}}-a_{11}d_1>0$ and $f(x_2)\to-\infty$ as $x_2\to-\infty$. Hence for all $K\in\mathbb{R}$, i.e. for all $\alpha\in\mathbb{R}$ on the curve $\Gamma(\alpha)$ there are points at which the integral curves of A go into the interior of $\Gamma_+(\alpha)$ and there are points at which the integral curves of A go into the interior of $\Gamma_-(\alpha)$. With this fact we can show that $\mathcal{F}$ is controllable. For this we prove that $\mathbb{A}_{\mathcal{F}}(0)=\mathbb{R}^2$ and $\mathbb{A}_{-\mathcal{F}}(0)=\mathbb{R}^2$, hence $\mathcal{F}$ is controllable. First we mention that $z\notin\mathbb{A}_{\mathcal{F}}(0)$, $z\in\Gamma(\alpha)$ implies $\Gamma(\alpha)\cap\mathbb{A}_{\mathcal{F}}(0)=\emptyset$. Consider the set $G=\{\alpha:\Gamma(\alpha)\subset\mathbb{A}_{\mathcal{F}}(0)\}$. This set is nonempty and open since $\Gamma(\alpha)$ can be crossed by trajectories of (4) from Γ_- to Γ_+ and from Γ_+ to Γ_-. Also G is closed. Indeed, let $\alpha_n\in G$, $\alpha_n\to\alpha$. Taking subsequences we can assume that α_n is monotone. If α_n increases, take a trajectory of (4) from $\Gamma_-(\alpha)$ to $\Gamma_+(\alpha)$; $\Gamma(\alpha_n)$ will meet it for sufficiently large n and then $\alpha\in G$ follows. The case of decreasing α_n is similar, hence G is closed. As an open closed and nonempty set, G must equal to $\mathbb{R}$ which shows $\mathbb{A}_{\mathcal{F}}(0)=\mathbb{R}^2$.

Similarly $\mathbb{A}_{-\mathcal{F}}(0)=\mathbb{R}^2$.

We proved that in this case (2) is controllable.

b$_4\beta$) $d_2>0$, $a_{12}<0$.

(*) $\frac{d_2a_{11}}{a_{22}}<\frac{a_{11}d_1}{a_{12}}$.

For $Ka_{22}<0$ there is only z such that $f'(z)=0$ (from (7)) with this z, $f'(x_2)<0$ if $x_2<z$ and $f'(x_2)>0$ if $x_2>z$, hence $f(x_2)$ attains its global minimum at $x_2=z$.

Here we find K and z such that

$$Ka_{22}<0,\qquad f'(z)=0,\qquad f(z)>0.$$

Let $\frac{d_2a_{11}}{a_{22}}<z<\frac{a_{11}d_1}{a_{12}}$, then $f(z)>0$ (see (6)).

We calculate K from the equation (7), then $f'(z)=0$ and $Ka_{22}<0$. With the chosen K and z, $f(x_2)>0$, i.e. $\dot{y}(0)>0$ for all x_2. This shows that $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(*) $\frac{d_2a_{11}}{a_{22}}>\frac{a_{11}d_1}{a_{12}}$.

Let $z=\frac{a_{11}d_2}{a_{22}}$, then $f(z)=\frac{a_{12}a_{11}d_2}{a_{22}}-a_{11}d_1<0$ and $f(x_2)\to\infty$ as $x_2\to-\infty$. Then for all $\alpha\in\mathbb{R}$ on the curve $\Gamma(\alpha)$ there are points at which the integral curves of A go into $\operatorname{int}\Gamma_+(\alpha)$ and there are points at which the integral curves of A go into $\operatorname{int}\Gamma_-(\alpha)$. From this fact, similarly to the case b4α), we can show that the system (2) is controllable

b$_4\gamma$) $d_2<0$, $a_{12}>0$.

(*) $\frac{d_2a_{11}}{a_{22}}>\frac{a_{11}d_1}{a_{12}}$.

For $Ka_{22}<0$ there is only z such that $f'(z)=0$. If $x_2<z$ then $f'(x_2)<0$ and if $x_2>z$ then $f'(x_2)>0$. Hence $f(x_2)$ attains its global minimum at $x_2=z$.

We find K and z such that

$$Ka_{22}<0, \qquad f'(z)=0, \qquad f(z)>0.$$

We choose $\frac{a_{11}d_2}{a_{22}}>z>\frac{d_1a_{11}}{a_{12}}$, then $f(z)>0$ in the case $Ka_{22}<0$.

We calculate K from the equation (7). We obtain that $f'(z)=0$ and $Ka_{22}<0$. With this K and z, $\dot{y}(0)>0$ for all x_2, the set $\Gamma_+(K-d_1)$ is $\mathcal{F}$-invariant.

(*) $\frac{d_2a_{11}}{a_{22}}<\frac{a_{11}d_1}{a_{12}}$.

Let $z=\frac{a_{11}d_1}{a_{22}}$ then $f(z)=\frac{a_{12}a_{11}d_2}{a_{22}}-a_{11}d_1<0$ and $f(x_2)\to+\infty$ as $x_2\to+\infty$. From this, similarly to the parts above, we see that in this case (2) is controllable.

b$_4\delta$) $d_2<0$, $a_{12}<0$.

(*) $\frac{d_2a_{11}}{a_{22}}>\frac{a_{11}d_1}{a_{12}}$.

For $Ka_{22}>0$ there is only z such that $f'(z)=0$. In this case if $x_2<z$ then $f'(x_2)>0$ and if $x_2>z$ then $f'(x_2)<0$. Hence $f(x_2)$ attains its global maximum at $x_2=z$.

We find K and z such that

$$Ka_{22}>0, \qquad f'(z)=0, \qquad f(z)<0.$$

We choose $\frac{a_{11}d_2}{a_{22}}>z>\frac{d_1a_{11}}{a_{12}}$. With this z, $f(z)<0$ (if $Ka_{22}>0$).

We calculate K from the equation (7). We obtain that $f'(z)=0$ and $Ka_{22}>0$. With this K and z, $\dot{y}(0)<0$ for all x_2, hence the set $\Gamma_-(K-d_2)$ is invariant for $\mathcal{F}$.

(*) $\frac{d_2a_{11}}{a_{22}}<\frac{a_{11}d_1}{a_{12}}$.

Let $z=\frac{a_{11}d_2}{a_{22}}$ then $f(z)=\frac{a_{12}a_{11}d_2}{a_{22}}-a_{11}d_1>0$ and $f(x_2)\to-\infty$ as $x_2\to+\infty$. From this and similarly to aboves, we see that in this case the system (2) is controllable.

b$_4\varepsilon$).

Now we assume that $\frac{d_2a_{11}}{a_{22}}=\frac{d_1a_{11}}{a_{12}}=z$. Then $d_1\neq0$ and $z\neq0$. Let $K=d_1e^{-\frac{z}{d_2}}$, we see that K satisfies (7).

We consider the integral curve $\Gamma(K-d_1)$ of $B+d$. We see that $(0,z)\in\in\Gamma(K-d_1)$.

Solving the differential equation $\dot{x}(t)=Ax(t)$, $x(0)=(0,z)$ we obtain

$$\begin{cases} x_1(t)=e^{a_{11}t}\cdot t\cdot a_{12}z \\ x_2(t)=e^{a_{22}t}z \end{cases} \qquad \text{if } a_{11}=a_{22}$$

and

$$\begin{cases} x_1(t)=\frac{a_{12}z}{a_{11}-a_{22}}\left(e^{a_{11}t}-e^{a_{22}t}\right) \\ x_2(t)=e^{a_{22}t}z \end{cases} \qquad \text{if } a_{11}\neq a_{22}.$$

Let

$$z_1(t)=e^{\frac{e^{a_{22}t}z-z}{d_2}}d_1-d_1.$$

Since the integral curve of $(B+d)$ through $(K-d_1,0)$ satisfies the equation $x_1=e^{\frac{x_2}{d_2}}K-d_1$, we have $(z_1(t),x_2(t))\in\Gamma(K-d_1)$.

We calculate now $\dot{x}_1(t)$, $\dot{z}_1(t)$.

$$\dot{z}_1(t)=\frac{a_{22}z}{d_2}e^{a_{22}t}\cdot e^{\frac{e^{a_{22}t}z-z}{d_2}}d_1=a_{12}z\cdot e^{a_{22}t}\cdot e^{\frac{e^{a_{22}t}z-z}{d_2}}.$$

(Here $\frac{a_{22}d_1}{d_2}=a_{12}$ by the assumption of z.)

$$\dot{x}_1(t)=a_{12}ze^{a_{11}t}(a_{11}t+1) \text{ if } a_{11}=a_{22}, \qquad \text{and}$$

$$\dot{x}_1(t)=\frac{a_{12}z}{a_{11}-a_{22}}e^{a_{22}t}\left(a_{11}e^{(a_{11}-a_{22})t}-a_{22}\right) \text{ if } a_{11}\neq a_{22}.$$

Consider now $\dot{z}_1(t)-\dot{x}_1(t)$ near $t=0$.

In the case $a_{11}=a_{22}$ we have $z=d_2$

$$\dot{z}_1(t)-\dot{x}_1(t)=a_{12}ze^{a_{11}t}\left(e^{e^{a_{11}t}-1}-a_{11}t-1\right).$$

The Taylor series expansion of $e^{e^{a_{11}t}-1}$ is the following.

$$e^{e^{a_{11}t}-1}=1+a_{11}t+a_{11}^2t^2+o(t^3).$$

We obtain

$$\dot{z}_1(t)-\dot{x}_1(t)=a_{12}z\cdot e^{a_{11}t}\left(a_{11}^2t^2+o(t^3)\right) \qquad (\text{if } a_{11}=a_{22}).$$

In the case $a_{11}\neq a_{22}$, one can see

$$e^{\frac{e^{a_{22}t}z-z}{d_2}}=1+\frac{z}{d_2}a_{22}t+\frac{z}{2d_2}a_{22}^2t^2+\frac{1}{2}\left(\frac{z}{d_2}\right)^2a_{22}^2t^2+o(t^3)=$$

$$=1+a_{11}t+\frac{1}{2}\left(a_{11}a_{22}+a_{11}^2\right)t^2+o(t^3).$$

$$\dot{z}_1(t)-\dot{x}_1(t)=a_{12}z\cdot e^{a_{22}t}\left\{e^{\frac{e^{a_{22}t}z-z}{d_2}}-\frac{1}{a_{11}-a_{22}}\left(a_{11}e^{(a_{11}-a_{22})t}-a_{22}\right)\right\}.$$

Calculating the Taylor series expansion of $\{\ \cdot\ \}$ we obtain

$$\dot z_1(t)-\dot x_1(t)=a_{12}z\cdot e^{a_{22}t}\left\{a_{11}a_{22}t^2+o(t^3)\right\} \qquad (\text{if } a_{11}\neq a_{22}).$$

In both of the cases we have

$$(*) \qquad \dot z_1(t)-\dot x_1(t)=a_{12}z\cdot e^{a_{22}t}\left\{a_{11}a_{22}t^2+o(t^3)\right\}= \\ =a_{11}^2a_{12}d_2e^{a_{22}t}\left(t^2+o(t^3)\right).$$

In what follows we will show that in the case $\frac{d_2a_{11}}{a_{22}}=\frac{d_1a_{11}}{a_{12}}$ the system (2) (with the assumption b$_4$)) is not controllable. In particular we show that either $\Gamma_-(K-d_1)$ or $\Gamma_+(K-d_1)$ is A-invariant.

We consider the case b$_4$). Similarly to there we can prove that with z and K defined above, $\dot y(0)<0$ (resp >0) for all $x_2\neq z$. For $x_2=z$ we have $\dot y(0)=f(z)=0$. We will show that the solution of $\dot x(t)=Ax(t)$, $x(0)=(0,z)$ satisfies $x_1(t)\leq z_1(t)$ (or $\geq$) for $t\in[0,\varepsilon)$ with some small positive number ε. From this and from $\dot y(0)<0$ (resp >0) $\forall x_2\neq z$ we obtain that $\Gamma_-(K-d_1)$ (resp $\Gamma_+(K-d_1)$) is A-invariant, thus the system (2) is not controllable. Indeed, in case b$_4\alpha$) and b$_4\delta$) above, $d_2a_{12}>0$ hence we get by $(*)$ that $z_1(t)>x_1(t)$ for $0<t<\varepsilon$, if $\varepsilon>0$ is sufficiently small. This means that the trajectory of A starting from $(0,z)$ goes into $\Gamma_-(K-d_1)$. Since $\dot y(0)<0$ for $x_2\neq z$, we get that $\Gamma_-(k-d_1)$ is invariant. In case b$_4\beta$) and b$_4\gamma$) we have $d_2a_{12}<0$, and then $z_1(t)<x_1(t)$ for $0<t<\varepsilon$, ε small. For $x_2\neq z$ we get here $\dot y(0)>0$, therefore $\Gamma_+(K-d_1)$ is invariant. So we have shown that (2) is never controllable when in case b$_4$) the equality $\frac{d_2a_{11}}{a_{22}}=\frac{d_1a_{11}}{a_{12}}$ holds.

Summarizing: (for b$_4$).

In all cases we showed that in the case b$_4$) the system (2) is controllable of and only if

$$\frac{a_{11}}{a_{22}}>\frac{a_{11}}{a_{12}}\frac{d_1}{d_2}.$$

Now we complete the proof of the Theorem 4.

We assume that our original system is written by

$$\dot x(t)=Ax(t)+u(t)bc^Tx(t)+u(t)d$$

and after changing basis and changing control ($v=c^Tbu$) the obtained system is written as

$$\dot x(t)=\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}x(t)+v(t)\begin{pmatrix}1&0\\0&0\end{pmatrix}x(t)+v(t)\begin{pmatrix}d_1\\d_2\end{pmatrix}.$$

Then we can calculate the values a_{ij}, d_i, $i=1,2$, $j=1,2$ by using Lemma 1.

Here

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = PAP^{-1}, \qquad \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \frac{1}{c^T b} Pd$$

where

$$P = \begin{pmatrix} c_1 & c_2 \\ -b_2 & b_1 \end{pmatrix}, \qquad P^{-1} = \frac{1}{c^T b} \begin{pmatrix} b_1 & -c_2 \\ b_2 & c_1 \end{pmatrix}$$

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \qquad c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

We obtain

$$a_{11} = \frac{c^T Ab}{c^T b}, \quad a_{12} = \frac{\det \begin{bmatrix} c^T \\ c^T A \end{bmatrix}}{c^T b}, \quad a_{21} = \frac{\det [b, Ab]}{c^T b},$$

$$d_1 = \frac{c^T d}{c^T b}, \qquad d_2 = \frac{\det[b, d]}{c^T b}.$$

We mention also that $\det(\cdot)$ and $\mathrm{Sp}(\cdot)$ do not change after changing basis.

From these facts we obtain:

I. $\Longleftrightarrow \det[b,d] = 0$.

II.a) $\Longleftrightarrow \det \begin{bmatrix} c^T \\ c^T A \end{bmatrix} = 0$.

II.b_1) or b_2) $\Longleftrightarrow \det[b,d] \neq 0$, $\det \begin{bmatrix} c^T \\ c^T A \end{bmatrix} \neq 0$, $\det A = 0$, $\mathrm{Sp} A \neq 0$.

II.b_3) $\Longleftrightarrow \det[b,d] \neq 0$, $\det \begin{bmatrix} c^T \\ c^T A \end{bmatrix} \neq 0$, $\det A = \mathrm{Sp} A = 0$.

II.b_4) $\Longleftrightarrow \det[b,d] \neq 0$, $\det \begin{bmatrix} c^T \\ c^T A \end{bmatrix} \neq 0$, $\det A \neq 0$.

In the case II.b_4), since $a_{21} = 0$, we can calculate a_{22}: $a_{22} = \frac{\det A}{a_{11}} = \frac{c^T b \det A}{c^T Ab}$. The inequality

$$\frac{a_{11}}{a_{22}} > \frac{a_{11}}{a_{12}} \cdot \frac{d_1}{d_2}$$

is equivalent to the inequality written in 2. of the theorem. With this we finished completely the proof of the Theorem 4. ∎

EXAMPLE [46]. Consider a system (2) with

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \qquad B = \begin{bmatrix} \nu & 0 \\ 0 & 0 \end{bmatrix}, \qquad d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

$$\det[A,B]=0, \quad c^T b \neq 0 \qquad \left(b=\begin{pmatrix}\nu\\0\end{pmatrix},\ c=\begin{pmatrix}1\\0\end{pmatrix}\right).$$

If $\lambda=0$, $\nu\neq 0$, $d_2\neq 0$ then the system is controllable.

If $\lambda\neq 0$, $\nu\neq 0$, $d_2\neq 0$, $1>\lambda\cdot\frac{d_1}{d_2}$ then the system is controllable. This contains the cases considered in [46].

THEOREM 5. *Consider the system*

$$\dot{x}(t)=Ax(t)+u(t)Bx(t)+u(t)d \tag{2}$$

with $B=bc^T\neq 0$, $d\neq 0$, $\det[A,B]=0$.

Assume that $c^T b=0$. *Then the system (2) is controllable in* $\mathbb{R}^2$ *if and only if either*

1. $\det[b,d]\neq 0$, $AB=0$ *and* $\det[Ad,d]\neq 0$; *or*
2. $\det[b,d]\neq 0$, $AB\neq 0$, $AB=2BA$ *and* $\det[Ad,d]\neq 0$.

PROOF. We mention first that we can change arbitrarily the basis in which the data are given. Indeed, if P is a nonsingular matrix and $A'=PAP^{-1}$, $B'=PBP^{-1}$, $d'=Pd$ then the system

$$\dot{y}(t)=A'y(t)+u(t)B'y(t)+u(t)d'$$

is controllable if and only if (2) does $(y(t)=Px(t))$. Further $B'=b'c'^T$ where $b'=Pb$, $c'^T=c^TP^{-1}$ and $\det[b,d]\neq 0\Longleftrightarrow\det[b',d']\neq 0$, $AB=0\Longleftrightarrow A'B'=0$, $AB=2BA\Longleftrightarrow A'B'=2B'A'$, $\det[Ad,d]\neq 0\Longleftrightarrow\det[A'd',d']\neq 0$.

From this and by Lemma 1 we can assume that

$$B=\begin{pmatrix}0&1\\0&0\end{pmatrix} \qquad (c^Tb=0).$$

Let $A=\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}$, $d=\begin{pmatrix}d_1\\d_2\end{pmatrix}$.

Then $\det[A,B]=0\Longleftrightarrow a_{21}=0$.

I. $d_2=0$ $(\Longleftrightarrow\det[b,d]=0)$.

From (2) we obtain

$$\dot{x}_2(t)=a_{22}x_2(t).$$

This shows that the line $\{x: x_2=0\}$ is an invariant set of (2).

The system is not controllable.

II. $d_2\neq 0$ $(\Longleftrightarrow\det[b,d]\neq 0)$.

Here we consider the system $\mathcal{F}=\{A,\pm(B+d)\}$ instead of (2), too.

The integral curves of $(B+d)$ satisfy

$$\begin{cases} \dot{x}_1(t)=x_2(t)+d_1 \\ \dot{x}_2(t)=d_2 \end{cases} \qquad t\in\mathbb{R} \qquad (\text{since } \pm(B+d)\in\mathcal{F}).$$

Solving that

$$\begin{cases} x_1(t)=\frac{d_2t^2}{2}+x_2(0)t+d_1t+x_1(0) \\ x_2(t)=d_2t+x_2(0). \end{cases}$$

We can take $x_2(0)=0$, $x(t)$ satisfies

$$x_1(t)=\frac{x_2^2(t)}{2d_2}+\frac{d_1}{d_2}x_2(t)+x_1(0).$$

This is the equation of the trajectory of $(B+d)$ which goes through the point $(x_1(0),0)$.

Every integral curve $\Gamma(\alpha)=\left\{x:x_1=\frac{x_2^2}{2d_2}+\frac{d_1}{d_2}x_2+\alpha\right\}$ divides the plane into two parts

$$\Gamma_+(\alpha):=\left\{x:x_1\geq\frac{x_2^2}{2d_2}+\frac{d_1}{d_2}x_2+\alpha\right\}$$

and

$$\Gamma_-(\alpha):=\left\{x:x_1\leq\frac{x_2^2}{2d_2}+\frac{d_1}{d_2}x_2+\alpha\right\}.$$

Now we consider integral curves of A:

$$(8) \qquad \begin{cases} \dot{x}(t)=Ax(t), & t\geq 0 \\ x(0)\in\Gamma(\alpha). \end{cases}$$

Let

$$y(t)=x_1(t)-\frac{x_2^2(t)}{2d_2}-\frac{d_1}{d_2}x_2(t)-\alpha.$$

If $x(t)$ satisfies (8), then $y(0)=0$ and

$$\dot{y}(t)=a_{11}x_1(t)+\left(a_{12}-\frac{d_1a_{22}}{d_2}\right)x_2(t)-\frac{a_{22}}{d_2}x_2^2(t)$$

Let $x(t)$ satisfy (8) and $x(0)=(x_1,x_2)$, then

$$\dot{y}(0)=\frac{a_{11}-2a_{22}}{2d_2}x_2^2+\frac{(a_{11}-a_{22})d_1+a_{12}d_2}{d_2}x_2+a_{11}\alpha.$$

a) $a_{11}=0$ ($\Longleftrightarrow AB=0$).

(*) $a_{12}=0$, $a_{22}\neq 0$.

In this case

$$\dot{y}(0)=-\frac{a_{22}}{d_2}x_2^2-\frac{a_{22}d_1}{d_2}x_2=-x_2\cdot(x_2+d_1)\cdot\frac{a_{22}}{d_2}.$$

If $d_1=0$ then $\dot{y}(0)\cdot\frac{a_{22}}{d_2}<0$ for $x_2\neq 0$, and $\Gamma_+(\alpha)$ is $\mathcal{F}$-invariant if $\frac{a_{22}}{d_2}<0$, $\Gamma_-(\alpha)$ is $\mathcal{F}$-invariant if $\frac{a_{22}}{d_2}>0$, for all α. The system is not controllable. Here $d_1a_{22}=d_2a_{12}$. If $d_1\neq 0$, there are values of x_2 for which $\dot{y}(0)>0$ and there are values of x_2 for which $\dot{y}(0)<0$. Then similarly to the proof of the Theorem 4 we can see that in this case the system $\mathcal{F}$ is controllable, so does (2). Here $d_1a_{22}\neq d_2a_{12}$.

(*) $a_{12}=0$, $a_{22}=0\Longrightarrow A=0$. The system is not controllable.

Here $d_1a_{22}=d_2a_{12}$.

(*) $a_{12}\neq 0$, $a_{22}=0$.

$\dot{y}(0)=a_{12}x_2$.

In this case $\dot{y}(0)>0$ if $a_{12}x_2>0$ and $\dot{y}(0)<0$ if $a_{12}x_2<0$. Similarly to the Theorem 4 we can see that the system is controllable. Here $d_1a_{22}\neq d_2a_{12}$.

(*) $a_{12}\neq 0$, $a_{22}\neq 0$.

$$\dot{y}(0)=x_2\left(-\frac{a_{22}}{d_2}x_2+a_{12}-\frac{a_{22}d_1}{d_2}\right).$$

If $a_{12}d_2=a_{22}d_1$ then $\dot{y}(0)=-\frac{a_{22}}{d_2}x_2^2$ and $\dot{y}(0)\cdot\frac{a_{22}}{d_2}<0$ for all $x_2\neq 0$, thus the system is not controllable (see above). If $a_{12}d_2\neq a_{22}d_1$ then $\dot{y}(0)>0$ for some x_2 and $\dot{y}(0)<0$ for some another x_2 (for all α). Hence we can see that the system is controllable.

Summarizing we obtain in the case II.a) that the system (2) is controllable if and only if $a_{12}d_2\neq a_{22}d_1$. This is equivalent to the fact that $\det[Ad,d]\neq 0$. Indeed, since $a_{11}=0$, $a_{21}=0$, $\det[Ad,d]=a_{12}d_2^2-a_{22}d_2d_1=$ $=d_2(a_{12}d_2-a_{22}d_1)$.

b) $a_{11}\neq 0$ ($\Longleftrightarrow AB\neq 0$).

b$_1$) $a_{11}\neq 2a_{22}$ ($\Longleftrightarrow AB\neq 2BA$).

In this case we can choose α such that $\dot{y}(0)\cdot\frac{a_{11}-2a_{22}}{2d_2}>0$ for all x_2. With this α if $\frac{a_{11}-2a_{22}}{2d_2}>0$ then $\Gamma_+(\alpha)$ is $\mathcal{F}$-invariant and if $\frac{a_{11}-2a_{22}}{2d_2}<0$ then $\Gamma_-(\alpha)$ is $\mathcal{F}$-invariant.

b$_2$) $a_{11}=2a_{22}$ ($\Longleftrightarrow AB=2BA$).

$$\dot{y}(0)=\frac{a_{22}d_1+a_{12}d_2}{d_2}x_2+a_{11}\alpha.$$

If $a_{22}d_1 = -a_{12}d_2$ then $\dot{y}(0) = a_{11}\alpha$. Hence $\dot{y}(0) > 0$ if $a_{11}\alpha > 0$ and with this α, $\Gamma_+(\alpha)$ is $\mathcal{F}$-invariant.

If $a_{22}d_1 \neq -a_{12}d_2$ then for all α, there are values of x_2 for which $\dot{y}(0) > 0$ and there are values of x_2 for which $\dot{y}(0) < 0$. Hence we can show that the system $\mathcal{F}$ is controllable by similar way as before.

Now we calculate $\det[Ad, d]$.

$$\det[Ad,d] = a_{11}d_1d_2 + a_{12}d_2^2 - a_{22}d_1d_2 = a_{22}d_1d_2 + a_{12}d_2^2 = \\ = d_2(a_{22}d_1 + a_{12}d_2).$$

Hence $a_{22}d_1 \neq -a_{12}d_2 \Longleftrightarrow \det[Ad, d] \neq 0$.

Thus we showed that in the case $AB \neq 0$ the system is controllable if and only if $AB = 2BA$ and $\det[Ad, d] \neq 0$.

We finished completely the proof of the Theorem 5. ∎

4. Two Structure theorems

Let A, B be matrices in $\mathbb{R}^{2\times 2}$. The pair (A, B) is said to be controllable if the corresponding system

$$\dot{x}(t) = Ax(t) + u(t)Bx(t)$$

is controllable in $\mathbb{R}^2 \setminus \{0\}$.

Denote

$$\mathcal{C} := \{(A,B) \in \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} : (A,B) \text{ is controllable}\}$$
$$\mathcal{K} := \{(A,B) \in \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} : \det[A,B] \leq 0\}.$$

As usual we denote the norm of a matrix A by

$$\|A\| = \left(\sum_{i,j} a_{ij}^2\right)^{\frac{1}{2}}.$$

We have the following structure theorem

THEOREM 6. *With the above notations $\mathcal{C} \subseteq \mathcal{K}$, $\mathcal{C}$ is open set in $\mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2}$ and $\mathcal{C}$ is dense in $\mathcal{K}$.*

PROOF. By Theorem A and Theorem 1 it is obvious that $\mathcal{C} \subseteq \mathcal{K}$.

I. $\mathcal{C}$ is open:

Let $(A, B) \in \mathcal{C}$. We show for there is $\varepsilon > 0$ such that if $\|A' - A\| < \varepsilon$, $\|B' - B\| < \varepsilon$ then $(A', B') \in \mathcal{C}$.

1. B has distinct real eigenvalues.

By Theorems A and 1, $\det[A,B]<0$.

Since B has distinct real eigenvalues, $\mathrm{Sp}^2 B-4\det B>0$. There is $\varepsilon>0$ such that $\|A'-A\|<\varepsilon$, $\|B'-B\|<\varepsilon$ (we write $(A',B')\in K_\varepsilon(A,B)$ for short) implies

$$\det[A',B']<0 \qquad \mathrm{Sp}^2 B'-4\det B'>0.$$

Then B' has distinct real eigenvalues, and by Theorems A and 1 $(A',B')\in\mathscr{C}$.

2. B has twofold real eigenvalue and $B\neq kE$

By the Theorems A and 1, $(A,B)\in\mathscr{C}\Longleftrightarrow\det[A,B]<0$. In this case $\mathrm{Sp}^2 B-4\det B=0$ and $B\neq kE$, $B\neq kA$.

Take a small $\varepsilon>0$ to ensure $\det[A',B']<0$, $B'\neq kE$, $B'\neq kA'$. Then $(A',B')\in\mathscr{C}$. Indeed, if $\mathrm{Sp}^2 B'-4\det B'\geq 0$ then B' has two real eigenvalues, hence $\det[A',B']<0$, $B'\neq kE$ is enough for the controllability. If $\mathrm{Sp}^2 B'=0$, $\det B'<0$, then B' has purely imaginary eigenvalues and $\det[A',B']+$ $+\det B'\cdot\mathrm{Sp}^2 A'\leq\det[A',B']<0$ implies $(A',B')\in\mathscr{C}$. Finally if $\mathrm{Sp}\,B'\neq 0$ and $\mathrm{Sp}^2 B'-4\det B'<0$ then B' has complex eigenvalues with nonzero real part, Hence $B'\neq kA'$ implies the controllability.

3. $B=kE$, $k\neq 0$

By Theorem A, $(A,B)\in\mathscr{C}\Longleftrightarrow\mathrm{Sp}^2 A-4\det A<0$

We need first the following fact:

"Let X be a matrix in $\mathbb{R}^{2\times 2}$ such that $\mathrm{Sp}^2 X-4\det X<0$ then for all $Y\in$ $\in\mathbb{R}^{2\times 2}$ $\det[X,Y]\leq 0$, and in the case $\det[X,Y]=0$ we have $\mathrm{Sp}^2 Y-4\det Y<0$ or $Y=kE$."

We prove this fact: Since $\det[X,Y]$ does not change under changing the basis, we can assume that

$$X=\begin{bmatrix}0 & 1\\ x_1 & x_2\end{bmatrix} \qquad \text{with} \quad 4x_1+x_2^2<0$$

(see 3. and 4. in the proof of the Theorem A)

$\det[X,Y]=$

$$=x_1(y_{22}-y_{11})^2-(x_1x_2y_{12}+x_2y_{21})(y_{22}-y_{11})-(x_1y_{12}-y_{21})^2+x_2^2y_{12}y_{21}$$

By the ending part of the proof of the Theorem A, $\det[X,Y]\leq 0$ and $\det[X,Y]=0\Longleftrightarrow x_1y_{12}=y_{21}$ and $y_{22}-y_{11}=\frac{x_2y_{12}}{x_1}$. With this and with $4x_1+x_2^2<0$ we obtain $\mathrm{Sp}^2 Y-4\det Y<0$, or $Y=kE$. Now, since $B=kE$, $k\neq 0$, $\mathrm{Sp}\,B\neq 0$.

There is $\varepsilon>0$ such that $(A',B')\in K_\varepsilon(A,B)$ implies

$$\mathrm{Sp}\,B'\neq 0, \qquad \mathrm{Sp}^2 A'-4\det A'<0, \qquad B'\neq\lambda A'.$$

a) If $\mathrm{Sp}^2 B' - 4\det B' \geq 0$ then by the above $\det[A',B'] < 0$ or $B' = kE$. In this case if $B' \neq kE$ then $(A',B') \in \mathscr{C}$ by the Theorems A and 1, if $B' = kE$ then $(A',B') \in \mathscr{C}$, too, by 2) of Theorem A.

b) If $\mathrm{Sp}^2 B' - 4\det B' < 0$, since $\mathrm{Sp}\, B' \neq 0$, then (A',B') satisfies the conditions of 4. in Theorem A, thus $(A',B') \in \mathscr{C}$.

4. B has purely imaginary conjugate eigenvalues

In this case $(A,B) \in \mathscr{C} \Longleftrightarrow \det[A,B] + \det B \cdot \mathrm{Sp}^2 A < 0$. B satisfies $\mathrm{Sp}\, B = 0$ and $\det B > 0$

$$\mathrm{Sp}^2 B - 4\det B < 0.$$

There is $\varepsilon > 0$ such that for all $(A',B') \in K_\varepsilon(A,B)$

$$\begin{cases} \det B' > 0 \\ \mathrm{Sp}^2 B' - 4\det B' < 0 \\ \det[A',B'] + \det B' \cdot \mathrm{Sp}^2 A' < 0. \end{cases}$$

Hence if $\mathrm{Sp}\, B' = 0$ then (A',B') satisfies the conditions of 3. in the Theorem A; if $\mathrm{Sp}\, B' \neq 0$, then (A',B') satisfies the conditions of 4. in the Theorem A ($\det B' > 0 \Longrightarrow \det[A',B'] < 0 \Longrightarrow B' \neq kA'$)

5. B has complex conjugate eigenvalues with nonzero real part.

In this case $(A,B) \in \mathscr{C} \Longleftrightarrow B \neq kA$ $(A \neq 0)$. B satisfies

$$\mathrm{Sp}^2 B - 4\det B < 0 \qquad \mathrm{Sp}\, B \neq 0.$$

There is $\varepsilon > 0$ such that for all $(A',B') \in K_\varepsilon(A,B)$

$$B' \neq kA', \quad \mathrm{Sp}\, B' \neq 0, \qquad \mathrm{Sp}^2 B' - 4\det B' < 0 \qquad \text{hold.}$$

This means that $(A',B') \in \mathscr{C}$.

We proved the part I.

II. $\mathscr{C}$ is dense in $\mathscr{K}$.

Let $(A,B) \in \mathscr{K}$. We will show that for all $\varepsilon > 0$ there is $(A',B') \in$ $\in K_\varepsilon(A,B) \cap \mathscr{C}$.

1. B has distinct real eigenvalues.

As usual, we can assume that

$$B = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix}, \qquad b_1 \neq b_2.$$

If $\det[A,B] < 0$ then $A' = A$, $B' = B$ and $(A',B') \in K_\varepsilon(A,B) \cap \mathscr{C}$ (by Theorems A and 1).

Assume that $\det[A,B] = 0$.

Then $\det[A,B] = (b_1 - b_2)^2 a_{12} a_{21} = 0$, where $A = (a_{ij})_{i,j}$. Since $b_1 \neq b_2$, $a_{12} a_{21} = 0$. Let $B' = B$.

(*) $a_{12}=0$, $a_{21}=0$.

Let

$$A'=\begin{pmatrix} a_{11} & a_{12}+\varrho \\ a_{21}-\varrho & a_{22} \end{pmatrix} \quad \text{with} \quad 0<\varrho^2<\frac{\varepsilon^2}{2}.$$

Then $\det[A',B']=(b_1-b_2)^2(-\varrho^2)<0$. Hence and by the Theorems A and 1, $(A',B')\in K_\varepsilon(A,B)\cap\mathscr{C}$.

(*) $a_{12}=0$, $a_{21}\neq 0$.

Let

$$A'=\begin{pmatrix} a_{11} & a_{12}+\varrho \\ a_{21} & a_{22} \end{pmatrix} \quad \text{with} \quad a_{21}\varrho<0 \text{ and } \varrho^2<\varepsilon^2.$$

Then $\det[A',B']=(b_1-b_2)^2a_{21}\varrho<0$ and $(A',B')\in K_\varepsilon(A,B)$. This means that $(A',B')\in K_\varepsilon(A,B)\cap\mathscr{C}$.

(*) $a_{12}\neq 0$, $a_{21}=0$.

Let

$$A'=\begin{pmatrix} a_{11} & a_{12} \\ a_{21}+\varrho & a_{22} \end{pmatrix} \quad \text{with} \quad a_{12}\varrho<0,\ \varrho^2<\varepsilon^2.$$

Then $(A',B')\in K_\varepsilon(A,B)\cap\mathscr{C}$.

2. B has a twofold real eigenvalue and $B\neq kE$.

We can assume that $B=\begin{pmatrix} b & 0 \\ 1 & b \end{pmatrix}$ (see the proof of the Theorem A).

Then $\det[A,B]=-a_{12}^2$.

(*) $a_{12}\neq 0$.

Let $A'=A$, $B'=B$. By the Theorems A and 1 $(A',B')\in\mathscr{C}$.

(*) $a_{12}=0$.

Let

$$A'=\begin{pmatrix} a_{11} & a_{12}+\varrho \\ a_{21} & a_{22} \end{pmatrix} \quad \text{with} \quad 0<\varrho<\varepsilon \text{ and } B'=B.$$

The $\det[A',B']=-\varrho^2<0$, hence $(A',B')\in K_\varepsilon(A,B)\cap\mathscr{C}$.

3. $B=kE$.

Let

$$B'=\begin{pmatrix} k+\varrho & \delta \\ -\delta & k+\varrho \end{pmatrix} \quad \text{with } \delta\neq 0,\ \varrho+k\neq 0,\ 2\varrho^2+2\delta^2<\varepsilon^2.$$

Then $\mathrm{Sp}^2B'-4\det B'<0$, $\mathrm{Sp}\,B'\neq 0$. Let

$$A'=\begin{pmatrix} a_{11}+\sigma & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{with } \sigma\neq a_{22}-a_{11} \text{ and } \sigma^2<\varepsilon^2.$$

Then $(A',B') \in K_\varepsilon(A,B)$ and $B' \neq \lambda A'$. Since B' has complex conjugate eigenvalues with nonzero real part (from above), by 4. of the Theorem A, $(A',B') \in \mathscr{C}$.

4. B has purely imaginary conjugate eigenvalues.

We can assume that

$$B = \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}, \qquad b<0.$$

Let

$$B' = \begin{pmatrix} \varrho & 1 \\ b & \varrho \end{pmatrix} \qquad \text{with} \quad 0 \neq 2\varrho^2 < \varepsilon^2.$$

Then $\mathrm{Sp}^2 B' - 4\det B' = 4b < 0$ and $\mathrm{Sp}\, B' \neq 0$.

Similarly to 3. above, we can choose A' such that $(A',B') \in K_\varepsilon(A,B) \cap \cap \mathscr{C}$.

5. B has complex conjugate eigenvalues with nonzero real part.

Let $B' = B$. Let

$$A' = \begin{pmatrix} a_{11}+\varrho & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \text{with} \quad \varrho^2 < \varepsilon^2 \text{ and such that } B' \neq kA'.$$

Then $(A',B') \in K_\varepsilon(A,B) \cap \mathscr{C}$.

We finished completely the proof of the Theorem 6. ∎

In the following we study the behaviour of the controllable triples (A,B,d) in $\mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} \times \mathbb{R}^2$.

The triple (A,B,d) is said to be controllable if the corresponding system

$$\dot{x} = Ax + uBx + ud$$

is controllable in $\mathbb{R}^2$. (Here A, B, d may be zero.)

Let

$$\mathscr{D} := \{(A,B,d) \in \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} \times \mathbb{R}^2 : (A,B,d) \text{ is controllable}\}$$
$$\mathscr{L} := \{(A,B,d) \in \mathbb{R}^{2\times 2} \times \mathbb{R}^{2\times 2} \times \mathbb{R}^2 : \det[A,B] \leq 0\}.$$

As a consequence of the Theorem 6 we can prove the following theorem

THEOREM 7. *With the notations above $\mathscr{D} \subseteq \mathscr{L}$, $\mathscr{D}$ is dense in $\mathscr{L}$, $\mathscr{D}$ is not open. Further, if $(A,B,d) \in \mathscr{D}$ with $(A,B) \in \mathscr{C}$ then $(A,B,d) \in \operatorname{int}\mathscr{D}$.*

PROOF. By the Theorems 2 and 3 if $\det[A,B] > 0$ then (A,B,d) is not controllable, this means that $\mathscr{D} \subset \mathscr{L}$.

Now let $(A,B,d) \in \mathscr{L}$.

By the Theorem 6 for each $\varepsilon>0$ there is $(A',B')\in K_\varepsilon(A,B)$ such that $(A',B')\in\mathscr{C}$, and we can suppose $B'\neq 0$.

1. B' has real eigenvalues with $\operatorname{rank} B'=2$ and $B'\neq kE$. Then for each $\varrho>0$ there is d' such that $\|d'-d\|<\varrho$ and $A'B'^{-1}d'\neq 0$ (since $(A',B')\in\mathscr{C}\Longrightarrow\det[A',B']<0\Longrightarrow A'\neq 0$). By the Theorem 2 $(A',B',d')\in\mathscr{D}$.

2. $B'=kE$, $k\neq 0$. Then $\operatorname{Sp}^2 A'-4\det A'<0$. For $\varrho>0$ there is $d'\neq 0$ such that $\|d'-d\|<\varrho$. By the Theorem 2 $(A',B',d')\in\mathscr{D}$.

3. B' has purely imaginary conjugate eigenvalues. In this case $\det[A',B']+\det B'\cdot\operatorname{Sp}^2 A'<0$. For $\varrho>0$ there is d' such that $\|d'-d\|<\varrho$ and $A'B'^{-1}d'\neq 0$. Hence and by the Theorem 2 $(A',B',d')\in\mathscr{D}$.

4. B' has complex conjugate eigenvalues with nonzereo real part. Then $B'\neq kA'$ and we can suppose $A'\neq 0$ (see the proof of Theorem 6). For $\varrho>0$ there is d' such that $\|d'-d\|<\varrho$ and $A'B'^{-1}d'\neq 0$. By the Theorem 2 $(A',B',d')\in\mathscr{D}$.

5. $\operatorname{rank} B'=1$ and $\det[A',B']<0$.

Let $B'=b'c'^T$. Then for $\varrho>0$ there is d' such that $\|d'-d\|<\varrho$ and $\det[b',d']\neq 0$. By the Theorem 3, $(A',B',d')\in\mathscr{D}$.

We showed that $\mathscr{D}$ is dense in $\mathscr{L}$.

Now we show that $\mathscr{D}$ is not open.

Consider a triple (A,B,d) with

$$A=\begin{pmatrix}0 & a_{12}\\ 0 & 0\end{pmatrix},\qquad B=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\qquad d=\begin{pmatrix}d_1\\ d_2\end{pmatrix}$$

and $a_{12}\neq 0$, $d_2\neq 0$.

Then $B=bc^T$, $b=\begin{pmatrix}1\\0\end{pmatrix}=c$, $d\neq 0$, $\det[A,B]=0$, $\det[b,d]\neq 0$, $\det\begin{bmatrix}c^T\\ c^TA\end{bmatrix}\neq 0$, $\det A=\operatorname{Sp} A=0$. By the Theorem 4 the triple (A,B,d) is controllable. For $\varepsilon>0$ let

$$A'=\begin{pmatrix}\frac{\varepsilon}{2} & a_{12}\\ 0 & 0\end{pmatrix},\qquad B'=B,\quad d'=d$$

then by the Theorem 4 the triple (A',B',d') is not controllable, and $(A',B')\in K_\varepsilon(A,B)$, $\|d'-d\|=0<\varepsilon$.

This shows that $\mathscr{D}$ fails to be open.

Now we prove the last assertion of the Theorem.

Let $(A,B,d)\in\mathscr{D}$ and $(A,B)\in\mathscr{C}$. Then (A,B,d) has no fixed points. By the Theorem 6 there is $\varepsilon>0$ such that $K_\varepsilon(A,B)\subset\mathscr{C}$. There is $\varrho>0$ such that $(A',B')\in K_\varrho(A,B)$ and $\|d'-d\|<\varrho$ implies that (A',B',d') has no fixed points. Indeed, assume indirectly that there are A_n, B_n, d_n,

x_n such that $A_n \to A$, $B_n \to B$, $d_n \to d$, $A_n x_n = 0$, $B_n x_n = -d_n$. If B is nonsingular then we can assume that B_n is non-singular for all n. Then $x_n \to -B_n^{-1} d_n \to -B^{-1} d = x$, thus $A_n x_n \to Ax$ and x is a fixed point of the system, which is contradictory. If $\operatorname{rank} B = 1$ then by Theorem 1 $\det[A,B] < 0$ and we can assume that $\det[A_n, B_n] < 0$ for all n. In this case $x_n = -[A_n, B_n]^{-1} A_n d_n \to -[A,B]^{-1} Ad$, and $A_n x_n \to Ax$, $B_n x_n \to Bx$, from which x is a fixed point of the system, and this is contradictory, too. We can choose $\varrho < \varepsilon$. Then for such a triple (A', B', d') we obtain that (A', B', d') is controllable. This means that the set $\{(A,B,d) \in \mathcal{D} : (A,B) \in \mathcal{E}\}$ is open. We completed the proof of the Theorem 7. ∎

References

[1] R. KALMAN, P. FALB, M. ARBIB, *Topics in Mathematical System Theory*, McGraw-Hill, New York, 1969.

[2] E. B. LEE and L. MARKUS, *Foundations of Optimal Control Theory*, Wiley, New York, 1976.

[3] A. A. SAGLE and R. E. WALDE, *Introduction to Lie Groups and Lie Algebras*, Academic Press, New York and London, 1973.

[4] J. KUČERA, Solution in Large of Control Problem $\dot{x} = (A(1-u) + Bu)x$, *Czech. Math. J.*, **16** (1966), 600–623.

[5] J. KUČERA, Solution in Large of Control Problem $\dot{x} = (Au + Bv)x$, *Czech. Math. J.*, **17** (1967), 91–96.

[6] G. W. HAYNES and H. HERMES, Nonlinear Controllability vie Lie Theory, *SIAM J. Control Optim.*, **8** (1970).

[7] C. LOBRY, Contrôlabilité des Systèmes non Linéaires, *SIAM J. Control Optim.*, **8** (1970), 573–605.

[8] H. J. SUSSMANN and V. JURDJEVIC, Controllability of Nonlinear Systems, *J. Diff. Equats.*, **12** (1972), 96–116.

[9] V. JURDJEVIC and H. J. SUSSMANN, Control Systems on Lie Groups, *J. Diff. Equats.*, **12** (1972), 313–329.

[10] D. L. LUKES, Global Controllability of nonlinear Systems, *SIAM J. Control Optim.*, **10** (1972), 112–126.

[11] R. W. BROCKETT, System Theory on Group manifold and Coset Space, *SIAM J. Control Optim.*, **10** (1972), 265–284.

[12] R. W. BROCKETT, Nonlinear Systems and Differential Geometry, *Proceedings of The IEEE*, **64** (1976), 61–72.

[13] C. BRUNI, G. DI PILLO and G. KOCH, Bilinear Systems: An Appealing Class of "Nearly Linear" Systems in Theory and Applications, *IEEE Trans. on Aut. Control*, AC **19** (1974), 334–348.

[14] R. E. RINK and R. R. MOHLER, Completely controllable bilinear systems, *SIAM J. Control Optim.*, **6**, (1968).

[15] R. M. HIRSCHON, Controllability in Nonlinear Systems, *J. Diff. Equats.*, **19** (1975), 46–61.

[16] H. J. SUSSMANN, Lie Brackets, Real Analyticity and Geometric Control, in *Differential Geometric Control Theory*, R. W. Brockett, R. S. Millmann, H. J. Sussmann eds., Birkhäuser, Boston–Basel–Stuttgart, 1983, 1–116.

[17] J. B. GONÇALES, Sufficient conditions for local controllability with unbounded control, *SIAM J. Control Optim.*, **25** (1987), 1371–1378.

[18] J. P. GAUTHIER et G. BORNARD, Contrôlabilité des systèmes bilinéaries, *SIAM J. Control Optim.*, **20** (1982), 377–384.

[19] R. W. BROCKETT, On the algebraic structure of Bilinear Systems in *Variable Structure Control Systems*, R. Mohler and A. Ruberti eds., Academic Press, 1972, 153–168.

[20] R. W. BROCKETT, Lie algebras and Lie groups in control theory, in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett Eds., Reidel, Dordrecht, 1973, 43–82.

[21] V. V. AMEL'KIN, *Differential equations in applications*, Science for everyone, Mir, Moscow, 1990.

[22] R. M. HIRSCHORN, Strong Controllability of Nonlinear Systems, *SIAM J. Contr. Optim.*, **24** (1986), 284–275.

[23] H. J. SUSSMANN, A general Theorem on Local Controllability *SIAM J. Contr. Optim.*, **25** (1987), 158–194.

[24] A. G. BUTKOVSKIY, *Phase portraits of dynamic systems with Control* (in Russian), Nauka, 1985.

[25] A. G. BUTKOVSKIY, Theory and Method of Phase Portrait for Dynamic Systems with Control (in Russian), *AIT*, (1985), No. 12., 43–53.

[26] A. G. BUTKOVSKIY, Dynamic Systems with Control and an opto-mechanical analogy (in Russian), *AIT*, 1988, No. 9. 34–47.

[27] A. G. BUTKOVSKIY, A differential-geometrical Method for constructive solution of Controllability and finite Control problems (in Russian), *AIT*, (1982), No. 1., 5–18.

[28] YU. N. ANDREEV, Differential-Geometrical methods in Control Theory (in Russian), *AIT*, (1982), No. 10., 5–46.

[29] N. L. LEPE, A geometrical approach to studying the Controllability of second-order Bilinear Systems, (in Russian) *AIT* (1984), No. 11., 19–25.

[30] B. BONNARD, Contrôlabilité des Systèmes Bilinéaires, *Math. Systems Theory*, **15** (1981), 79–92.

[31] R. HERMANN, On the Accessibility Problem in Control Theory in *Nonlinear Differential Equations and Nonlinear Mechanics J.*, P. LaSalle and S. Lefschetz eds., Academic Press, N. Y., 1963.

[32] R. M. HIRSCHON, Global Controllability of Nonlinear Systems, *SIAM J. Control Optim.*, **14** (1976), 700–711.

[33] A. KUMPERA, Remarks on Paper by R. M. Hirschon, *SIAM J. Control Optim.*, **26** (1988), 465–467.

[34] R. HERMANN and A. J. KRENER, Nonlinear Controllability and Observability, *IEEE Trans Aut. Control*, AC **22** (1977), 728–740.

[35] L. R. HUNT, Controllability of General Nonlinear Systems, *Math. Systems Theory*, **12** (1979), 361–370.

[36] L. R. HUNT, Global Controllability of Nonlinear Systems in two Dimensions, *Math. Systems Theory*, **13** (1980), 361–376.

[37] V. JURDJEVIC and I. KUPKA, Control Systems Subordinated to a Group Action: Accessibility, *J. Diff. Equats.*, **39** (1981), 186–211.

[38] V. JURDJEVIC and I. KUPKA, Control Systems on Semi-simple Lie Groups and their Homogeneous Spaces, *Ann. Inst. Fourier*, Grenoble, **31**, **4** (1981), 151–179.

[39] V. JURDJEVIC and I. KUPKA, Polynomial Control Systems, *Math. Ann.*, **272** (1985), 361–368.

[40] G. SALLET, Extension Techniques, in *Systems and Control Encyclopedia II.*, (1987), 1581–1583.

[41] V. JURDJEVIC and G. SALLET, Controllability properties of affine Systems, *SIAM J. Control and Optim.*, **22** (1984), 501–508.

[42] P. BRUNOVSKÝ, Local Controllability of odd Systems, *Banach Center Publications*, Vol 1, Warszawa, 1976, 39–44.

[43] U. PIECHOTTKA and P. M. FRANK, Controllability of Bilinear Systems: A survey and some new results, in *Nonlinear Control Systems design*, ed. A. Isidori, Pergamon Press, 1990, 23–28.

[44] PH. ADDA, Controllabilite des systemes bilineares generaux et homogenes dans $\mathbb{R}^2$, in *Lecture Notes in Control and Inform. Sciences*, INRIA, No 111 (1988), 205–214.

[45] D. E. KODITSCHEK and K. S. NARENDRA, The controllability of Planar Bilinear Systems, *IEEE Trans. Aut. Contr.*, AC **30** (1985), 87–89.

[46] U. PIESCHOTTKA, Comments on "The Controllability of Planar Bilinear Systems." *IEEE Trans. Aut. Contr.*, AC **35** (1990), 767–768.

[47] O. HÁJEK, Control Theory in the Plane, in *Lecture Notes in Control and Information Sciences*, Vol. 153, Springer Verlag, 1991.

[48] I. JOÓ and N. M. TUAN, On Controllability of Bilinear Systems I: (Controllability in finite dimensions), *Annales Univ. Sci. Budapest.*, Sect. Math., **35** (1992), 189–215.

ANNALES UNIV. SCI. BUDAPEST., **35** (1992), 267–270

A SIMPLE CLASS OF CUBIC SYSTEMS WITHOUT LIMIT CYCLES

By

I. SZÁNTÓ

Department of Analysis, L. Eötvös University, Budapest*

(Received January 20, 1992)

I. Introduction

A general technique to guarantee the non-existence of non-trivial closed orbit of second order dynamic systems is Bendixson criterion [1], [3].

This criterion was generalized by H. DULAC [1], [4] as follows. Let

$$\dot{X} = P(X,Y), \qquad \dot{Y} = Q(X,Y) \tag{1}$$

be an analytic dynamic system, G a simply connected subregion of the domain of definition of (1).

Let $B(X,Y)$ be a continuously differentiable function defined in G such that the function

$$\frac{\partial}{\partial X}(B(X,Y)P(X,Y)) + \frac{\partial}{\partial Y}(B(X,Y)Q(X,Y))$$

does not change sign in G, then there are no non-trivial closed orbits in G.

The procedure of construction of Dulac Functions for (1) concerns the solution to a partial differential equation

$$\operatorname{div}(B(X,Y)(P(X,Y),Q(X,Y))) = \Phi(X,Y)$$

where $\Phi(X,Y)$ is some suitable function positive or negative definite in G.

A different approach to obtain the function $B(X,Y)$ is based on the search of integrating factors as we show in the following.

THEOREM 1. *Let (1) be an analytic system on the plane, G a simply connected subregion of the domain of definition of the system. If there exists an analytic system on the plane*

$$\dot{X} = P_0(X,Y), \qquad \dot{Y} = Q_0(X,Y) \tag{2}$$

such that the system

* This research was supported by Fondecyt/Chile, under Grant 245/91, 92.

$$(3) \qquad \dot{X} = P(X,Y) - P_0(X,Y), \qquad \dot{Y} = Q(X,Y) - Q_0(X,Y)$$

has an integrating factor $\mu(X,Y) \in C^1(G)$ *and*

$$\operatorname{div}\left(\mu(X,Y)(P_0(X,Y), Q_0(X,Y))\right)$$

does not change sign in G, *then the system* (1) *does not admit any non-trivial closed orbit lying entirely in* G.

PROOF. If $\mu(X,Y)$ is an integrating factor for (3), then

$$\operatorname{div}\left(\mu(X,Y)(P(X,Y) - P_0(X,Y), Q(X,Y) - Q_0(X,Y))\right) \equiv 0$$

and this implies that $\operatorname{div}\left(\mu(X,Y)(P(X,Y), Q(X,Y))\right)$ does not change sign in G, and then $B(X,Y) = \mu(X,Y)$ is a Dulac Function for (1) in G.

2. Main result

Using the above procedure, we can generalize Bautin's theorem [2] as follows.

THEOREM 2. *Let us consider the cubic polynomial vector field*

$$(4) \qquad \begin{aligned} \dot{X} &= X\left(a_{30}X^2 + a_{21}XY + a_{12}Y^2 + a_{20}X + a_{11}Y + a_{10}\right), \\ \dot{Y} &= Y\left(b_{21}X^2 + b_{12}XY + b_{03}Y^2 + b_{11}X + b_{02}Y + b_{01}\right), \end{aligned}$$

where a_{ij}, $b_{ij} \in \mathbb{R}$. *If*

$$\delta = a_{20}b_{02} - a_{11}b_{11} \neq 0,$$

$$(4') \quad \begin{cases} [a_{30}(k+2) + hb_{21}][ka_{10} + hb_{01}] > 0, \text{ and} \\ [(k+1)a_{21} + (h+1)b_{12}]^2 - 4[a_{30}(k+2) + hb_{21}][ka_{12} + (h+2)b_{03}] < 0 \end{cases}$$

where $k = \frac{b_{02}(b_{11} - a_{20})}{\delta}$, $h = \frac{a_{20}(a_{11} - b_{02})}{\delta}$ *then the system* (4) *has no non-trivial closed orbit in the whole plane.*

PROOF. First of all, if there is a closed orbit not intersecting the coordinate axes, then the closed orbit must lie in some quadrant.

As $\delta \neq 0$, we can choose

$$\begin{aligned} P_0(X,Y) &= X\left(a_{30}X^2 + a_{21}XY + a_{12}Y^2 + a_{10}\right), \\ Q_0(X,Y) &= Y\left(b_{21}X^2 + b_{12}XY + b_{03}Y^2 + b_{01}\right). \end{aligned}$$

Then $\mu(X,Y) = X^{k-1}Y^{h-1}$ is an integrating factor for the system

$$\dot{X} = X(a_{20}X + a_{11}Y), \qquad \dot{Y} = Y(b_{11}X + b_{02}Y)$$

and

$$\operatorname{div}\left(X^{k-1}Y^{h-1}(P_0(X,Y),Q_0(X,Y))\right)=$$
$$=X^{k-1}Y^{h-1}\Big[X^2(a_{30}(k+2)+hb_{21})+XY((k+1)a_{21}+(h+1)b_{12})+$$
$$+Y^2(ka_{12}+(h+2)b_{03})+ka_{10}+hb_{01}\Big]$$

does not change sign in any quadrant if (4′) holds, and this proves our theorem.

COROLLARY. Let us consider the system

$$\text{(5)}\qquad \begin{aligned}\dot{X}&=X(a_1X+b_1Y+c_1)(a_2X+b_2Y+c_2),\\ \dot{Y}&=Y(a_3X+b_3Y+c_3)(a_4X+b_4Y+c_4).\end{aligned}$$

If

$$\delta=(c_2a_1+c_1a_2)(c_4b_3+c_3b_4)-(c_2b_1+c_1b_2)(c_4a_3+c_3a_4)\neq 0,$$
$$(kc_1c_2+hc_3c_4)(a_1a_2(k+2)+ha_3a_4)>0$$

and

$$[(k+1)(a_1b_2+a_2b_1)+(h+1)(a_3b_4+b_3a_4)]^2-$$
$$-4[(k+2)a_1a_2+ha_3a_4][kb_1b_2+(h+2)b_3b_4]<0$$

where

$$k=(c_4b_3+c_3b_4)(c_4a_3+c_3a_4-c_2a_1-c_1a_2)/\delta,$$
$$h=(c_2a_1+c_1a_2)(c_2b_1+c_1b_2-c_4b_3-c_3b_4)/\delta.$$

then (5) does not admit any non-trivial closed orbit in G.

REMARK. If $a_2=b_2=a_4=b_4=0$ and $c_4=c_2=1$ in (5) then we have a Bautin's result. In fact the system (5) has the form

$$\text{(6)}\qquad \dot{X}=X(a_1X+b_1Y+c_1),\qquad \dot{Y}=Y(a_3X+b_3Y+c_3)$$

and it has been shown by BAUTIN [2] that such a system cannot have a limit cycle. In fact, the Dulac function is $B(X,Y)=X^{k-1}Y^{h-1}$, where $k=\frac{b_3(a_3-a_1)}{\delta}$, $h=\frac{a_1(b_1-b_3)}{\delta}$ and $\delta=a_1b_3-b_1a_3\neq 0$. Then

$$\operatorname{div}(X^{k-1}Y^{h-1}(Xc_1,Yc_3)=X^{k-1}Y^{h-1}(kc_1+hc_3)$$

does not change sign in any quadrant if $kc_1+hc_3\neq 0$.

References

[1] ANDRONOV, A. A., LEONTOVICH, E. A., GORDON, I. I. and MAIER, A. G., *Qualitative theory of second order Dynamic Systems*, John Wiley and Sons, New York, 1973.

[2] BAUTIN, N. N., On the periodic solutions of a system of differential equations, (Russian), *Prikl. Mat. i. Mech.*, **18** (1954), 128.

[3] BENDIXSON, I., Sur les courbes defines par des equations differentielles, *Acta Math.*, **24** (1901), 1–88.

[4] DULAC, H., Recherches des cycles limits, *C. R. Acad. Sci. Paris*, Ser I. Math., **204** (1937), 1703–1706.

CORRECTIONS TO OUR PAPER "THE UNIQUE AMALGAMATION PROPERTY FOR LATTICES"

By

ERVIN FRIED and GEORGE GRÄTZER

The paper appeared in the same Journal **33** (1990), 167–176. Unfortunately, the galley proof of our paper did not reach us. Here we correct some of the "worse" misprints.

1. The correct address of the second author is: University of Manitoba, Winnipeg
2. "amalgamat … A, B, S" should be read "… A and B over S" (always).
3. "A. Slavík" is correctly "V. Slavík".
4. The second word on page 168 should be instance.
5. The upper index "^{1}" should be always "$^\ell$".
6. In Figure 7 (page 171) the "first element" in the "lower level" and all elements in the "upper level" are stripped (they look black-filled).
7. The first four references have already appeared:

[1] *Trans. Amer. Math. Soc.*, **286** (1984), 251–256.

[2] *Proc. Amer. Math. Soc.*, **106** (1989), 1–21.

[3] *J. Austral. Math. Soc.* (Series A), **47** (1989), 1–21.

[4] *Algebra Universalis*, **27** (1990), 270–278.

INDEX

Address:
MATHEMATICAL INSTITUTE, L. EÖTVÖS UNIVERSITY
Budapest, Múzeum krt. 6–8.
H - 1088
ISSN 0524–9007

Technikai szerkesztő:
Dr. SCHARNITZKY VIKTOR

A kiadásért felelős: az Eötvös Loránd Tudományegyetem rektora
A kézirat a nyomdába érkezett: 1992. szeptember. Megjelent: 1992. december
Terjedelem: 23,97 A/4 ív. Példányszám: 500
Készült az EMTEX szedőprogram felhasználásával
az MSZ 5601–59 és 5602–55 szabványok szerint
Az elektronikus tipografálás Juhász Lehel és Fried Katalin munkája
A sokszorosítás az ELTE sokszorosító üzemében készült
Felelős vezető: Arató Tamás
ELTE 92257