

УДК 519.95

О РЕГУЛЯРНЫХ ЯЗЫКАХ С ПОЛИНОМИАЛЬНЫМ РОСТОМ ЧИСЛА СЛОВ

А. С. Строгалов

Важной характеристикой регулярных множеств являются функции, описывающие распределение числа слов заданной длины в них и называемые функциями роста этих множеств. Устанавливается, что любой неотрицательный целозначный полином может являться функцией роста некоторого регулярного множества.

Введение

В работе исследуются функции роста регулярных языков, описывающие распределения числа слов заданной длины τ в них. Как установлено в [2], эти функции разделяются на два класса: в первый класс попадают все функции, ограниченные сверху некоторым полиномом от τ ; во второй класс — все остальные функции, причем они растут не медленнее, чем экспонента от τ .

Здесь рассматриваются функции роста из первого класса. Основным результатом статьи является доказательство того, что любой неотрицательный целозначный полином является функцией роста.

§ 1. Основные понятия и результаты

Пусть задан алфавит $A = \{\alpha_0, \dots, \alpha_{n-1}\}$, A^* — множество всех слов $\tilde{\alpha} = \alpha(1)\alpha(2)\dots\alpha(t)$ над A , где $\alpha(i) \in A$ при $i \in \{1, 2, \dots, t\}$; число $|\tilde{\alpha}| = t$ называем длиной $\tilde{\alpha}$. Множество $\mathcal{L} \subseteq A^*$ называем *регулярным языком*, если оно распознаваемо некоторым автоматом $\mathfrak{A}_{\mathcal{L}}$ [1].

Пусть $\mathcal{L}(\tau)$ — множество всех слов длины τ из \mathcal{L} , тогда $\mathcal{L} = \bigcup_{\tau=1}^{\infty} \mathcal{L}(\tau)$. Обозначая через $|B|$ мощность множества B , рассмотрим функцию $f_{\mathcal{L}}(\tau) = |\mathcal{L}(\tau)|$, которую назовем *функцией роста языка \mathcal{L}* .

В работе [2] показано, что эта функция либо не превосходит полинома, либо имеет экспоненциальный рост. В первом случае говорим, что язык имеет *полиномиальный тип (рост)*, а во втором — *экспоненциальный тип (рост)*.

В статье [3] описан следующий способ нахождения числа слов длины τ в регулярном языке \mathcal{L} . Сначала строится матрица M (называемая *скелетной*) для автомата, распознающего язык \mathcal{L} , элементы m_{ij} которой равны количеству дуг, соединяющих в диаграмме переходов i -е состояние с j -м. Затем к ней применяется теорема Гамильтона — Кэли [5] о том, что

всякая квадратная матрица удовлетворяет своему характеристическому многочлену, т. е.:

$$M^k + c_1 M^{k-1} + \dots + c_{k-1} M + c_k E \equiv \Theta. \quad (1)$$

Здесь E и Θ — единичная (диагональная) и, соответственно, нулевая матрицы, c_i есть i -й коэффициент характеристического многочлена матрицы M , а k — степень характеристического многочлена и одновременно порядок матрицы M , определяемый количеством состояний автомата, распознающего язык \mathcal{L} ; при этом можно рассматривать только приведенные автоматы.

Умножив обе части тождества (1) на $M^{\tau-k}$, где $\tau \geq k$, получим

$$M^\tau + c_1 M^{\tau-1} + \dots + c_{k-1} M^{\tau-(k-1)} + c_k M^{\tau-k} \equiv \Theta. \quad (2)$$

Расписав поэлементно (2), имеем

$$m_{ij}^{(\tau)} + c_1 m_{ij}^{(\tau-1)} + \dots + c_{k-1} m_{ij}^{(\tau-k+1)} + c_k m_{ij}^{(\tau-k)} = 0. \quad (3)$$

Здесь $m_{ij}^{(l)}$ ($\tau - k \leq l \leq \tau$) — соответствующий элемент матрицы M^l . Известно [4], что $m_{ij}^{(\tau)}$ есть число слов длины τ , ведущих из i -го в j -е состояние автомата со скелетной матрицей M . Осталось взять в качестве i номер инициального состояния, а в качестве j — номер финального состояния. Если финальных состояний несколько, то надо просуммировать соответствующие тождества, подобно тому, как это сделано в [2]. В результате мы получим линейное рекуррентное уравнение типа (3).

Структура общего решения линейного рекуррентного уравнения известна и имеет вид

$$m_{ij}^{(\tau)} = \sum_{i=1}^s (c_{i1} + c_{i2}\tau + \dots + c_{ir_i}\tau^{r_i-1}) \lambda_i^{\tau},$$

где λ_i — корни характеристического многочлена $\det |M - \lambda E| = 0$, а r_i — кратность соответствующего корня. Коэффициенты c_{ij} определяются из начальных условий для уравнения (3).

Для простоты изложения будем рассматривать случай $A = \{0, 1\}$. Обозначим через \mathbf{N} множество натуральных чисел.

Полином $P_s(\tau)$ называется *целозначным*, если при целых τ его значения целые числа. Например, любой полином с целыми коэффициентами является целозначным, а полином $P_s(\tau) = C_\tau^s$ целозначный, но не целочисленный; его коэффициенты рациональные числа [7]. Далее нас будут интересовать главным образом неотрицательные целозначные полиномы (н. ц. п.) и целозначные полиномы (ц. п.). Имеет место

Теорема 1. *Для любого н. ц. п. $P_s(\tau)$ существует $l \in \mathbf{N}$ такое, что при $\tau \geq l$*

$$P_s(\tau) = a_0 + a_1 C_{\tau-l}^1 + \dots + a_s C_{\tau-l}^s,$$

причем $\sum_{i=0}^s a_i \leq 2^l$ и $a_i \in \mathbf{N}$ при $i = 0, 1, \dots, s$.

Отметим, что l зависит только от полинома $P_s(\tau)$. Такое представление н. ц. п. $P_s(\tau)$ позволяет получить основной результат, который для случая, когда $A = \{0, 1\}$ формулируется следующим образом.

Теорема 2. *Для любого н. ц. п. $P_s(\tau)$ существует $l \in \mathbf{N}$ и регулярный язык \mathcal{L} такой, что $f_{\mathcal{L}}(\tau) = P_s(\tau)$ при $\tau \geq l$. Здесь l то же, что и в формулировке теоремы 1.*

Если ограничения на мощность входного алфавита A отсутствуют, то теорему 2 можно усилить.

Теорема 3. *Для любого н. ц. п. $P_s(\tau)$ существует регулярный язык \mathcal{L} такой, что $f_{\mathcal{L}}(\tau) = P_s(\tau)$ при $\tau \in \mathbf{N}$.*

Частным случаем регулярных языков полиномиального роста являются

те, функция роста которых ограничена сверху константой; их называем константно-ограниченными.

Теорема 4. Класс функций роста константно-ограниченных языков совпадает с классом почти-периодических функций $f: \mathbf{N} \rightarrow \mathbf{N} \cup \{0\}$.

§ 2. Доказательство вспомогательных утверждений

Лемма 1. Существует семейство регулярных языков \mathcal{L}_k , для которых $f_{\mathcal{L}_k}(\tau) = C_{\tau}^{k-1}$.

Доказательство. Заметим, что степень полинома C_{τ}^{k-1} равна $k-1$. Рассмотрим семейство регулярных языков \mathcal{L}_k , распознаваемых автоматами

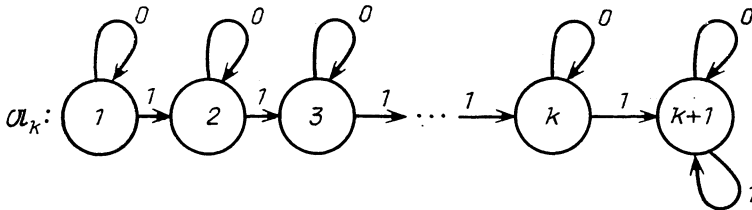


Рис. 1

\mathcal{A}_k , диаграммы которых изображены на рис. 1. Здесь состояние с номером 1 — инициальное, состояние с номером k — финальное. Как следует из работы [2], $f_{\mathcal{L}_k}$ имеет полиномиальный тип. Найдем ее явное выражение.

Скелетная матрица размера $(k+1) \times (k+1)$ автомата \mathcal{A}_k имеет вид

$$M = \begin{pmatrix} 1 & 1 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 1 & 1 \\ 0 & 0 & \dots & 0 & 0 & 2 \end{pmatrix}; \quad \det |M - \lambda E| = (\lambda - 1)^k (\lambda - 2).$$

Следовательно, общее решение уравнения (3) для этой матрицы есть

$$m_{ik}^{(\tau)} = (c_1 + c_2\tau + \dots + c_k\tau^{k-1}) + c_{k+1}2^\tau.$$

В силу того, что $f_{\mathcal{L}_k}$ полиномиального типа, $c_{k+1} = 0$. Отметим, что это верно именно для указанного распределения инициального и финального состояний. Нам остается определить константы c_1, \dots, c_k из начальных условий. Заметим, что из строения диаграммы автомата \mathcal{A}_k (рис. 1) вытекает, что

$$f_{\mathcal{L}_k}(1) = \dots = f_{\mathcal{L}_k}(k-2) = 0, \quad f_{\mathcal{L}_k}(k-1) = 1, \quad f_{\mathcal{L}_k}(k) = k.$$

Так как $f_{\mathcal{L}_k}(\tau)$ — полином степени $k-1$, то для его построения можно воспользоваться интерполяционной формулой Лагранжа [6] для узлов (x_i, y_i) , $1 \leq i \leq k$, вида $(i, 0)$ при $1 \leq i \leq k-2$ и $(k-1, 1)$, (k, k) . Имеем

$$\begin{aligned} f_{\mathcal{L}_k}(\tau) = m_{ik}^{(\tau)} &= 1 \cdot \frac{(\tau-1) \dots (\tau-(k-2)) (\tau-k)}{((k-1)-1) \dots ((k-1)-(k-2)) ((k-1)-k)} + \\ &+ k \cdot \frac{(\tau-1) \dots (\tau-(k-2)) (\tau-(k-1))}{(k-1) \dots (k-(k-2)) (k-(k-1))} = \\ &= \frac{(\tau-1) \dots (\tau-(k-2))}{(k-2)!} \left[k - \tau + \frac{k(\tau-(k-1))}{k-1} \right] = \frac{(\tau-1) \dots (\tau-(k-2)) \tau}{(k-2)! (k-1)} = C_{\tau}^{(k-1)}. \end{aligned}$$

Лемма 1 доказана.

Используя формулу интерполяционного многочлена Ньютона [8] (случай равноотстоящих узлов), можно получить следующее представление для

ц. п. $P_s(\tau)$:

$$P_s(\tau) = P_s(l) + \sum_{k=1}^s \frac{\Delta^k P_s(l)}{k!} (\tau-l)(\tau-l-1) \dots (\tau-l-(k-1)) = \\ = P_s(l) + \sum_{k=1}^s \Delta^k P_s(l) C_{\tau-l}^k, \quad (4)$$

где $\Delta^k P_s(l) = \sum_{i=0}^k (-1)^{k-i} C_k^i P_s(l+i)$ — k -я разделенная разность. Для нас существенно, чтобы в разложении (4) коэффициенты были целыми числами. Для этого достаточно заметить выполнение следующих очевидных равенств:

$$P_s(l+i) = P_s(l) + \sum_{k=1}^i C_i^k \Delta^k P_s(l), \quad i=0, 1, \dots, s.$$

Учитывая, что $P_s(\tau)$ — целозначный, из этих уравнений последовательно находим величины $\Delta^1 P_s(l)$, $\Delta^2 P_s(l)$, ..., $\Delta^s P_s(l)$ которые, очевидно, целые числа.

Лемма 2. Для всякого ц. п. $P_s(\tau)$ найдется $l_0 \in \mathbb{N}$ такое, что при всех $l \geq l_0$ в представлении (4) будет выполнено неравенство

$$\sum_{i=0}^s |\Delta^i P_s(l)| \leq 2^l.$$

Доказательство. Из определения k -й разделенной разности имеем

$$|\Delta^k P_s(l)| \leq \sum_{i=0}^k C_k^i |P_s(l+i)| \leq 2^k \sum_{i=0}^k |P_s(l+i)| \leq 2^k c (l+k)^s,$$

где $c > 0$. Далее, для некоторого $c_1 > 0$ имеем $\sum_{i=0}^s |\Delta^i P_s(l)| \leq 2^s s^2 c_1 (l+s)^s$.

Ясно, что при фиксированном s найдется такое l_0 , что при всех $l \geq l_0$ выполнено $2^s s^2 c_1 (l+s)^s \leq 2^l$. Лемма доказана.

Лемма 3. Для любого н. ц. п. $P_s(\tau)$ найдется $l_1 \in \mathbb{N}$ такое, что при всех $l \geq l_1$ в представлении (4) будет выполнено неравенство

$$\Delta^i P_s(l) \geq 0, \quad i=1, 2, \dots, s.$$

Доказательство. Используя интерполяционный многочлен Ньютона, имеем

$$P_s(\tau) = P_s(l) + P_s'(\xi_1) (\tau-l) + \frac{P_s''(\xi_2)}{2!} (\tau-l)(\tau-(l+1)) + \dots \\ + \frac{P_s^{(s)}(\xi_s)}{s!} (\tau-l)(\tau-(l+1)) \dots (\tau-(l+s)),$$

где $P_s^{(i)}(\xi_i)$ — i -я производная многочлена $P_s(\tau)$ в точке ξ_i , $i=1, 2, \dots, s$. Ясно, что с учетом неотрицательности многочлена можно выбрать такое l_1 , что при любых $\xi_i \in [l_1, l_1+s]$ выполнено $P_s^{(i)}(\xi_i) \geq 0$. Теперь остается воспользоваться теоремой Лагранжа для разделенных разностей [8]. Лемма доказана.

§ 3. Доказательство теорем 1—4

Доказательство теоремы 1. Выбрав $l(P_s)$ таким образом, чтобы оно удовлетворяло условиям лемм 2 и 3 как значения l_0 и l_1 соответственно, получим с учетом этих лемм справедливость утверждения теоремы 1.

Доказательство теоремы 2. Рассмотрим автомат \mathfrak{A}_k , диаграмма переходов которого изображена на рис. 2. Здесь $\alpha_1, \dots, \alpha_i \in \{0, 1\}$, $\bar{\alpha}_i$ —

отрицание α_i , начальное состояние имеет номер 1, финальное состояние имеет номер $l+k$.

Пусть \mathcal{L}_k — язык, распознаваемый этим автоматом, $\tilde{\alpha} \in \mathcal{L}_k$ тогда $\tilde{\alpha} = \alpha_1 \dots \alpha_l \tilde{\beta}$, где $|\tilde{\beta}| = \tau - l$. Число слов длины $\tau - l$, ведущих из $(l+1)$ -го в $(l+k)$ -е состояние в силу леммы 1 равно $C_{\tau-l}^{k-1}$. Отсюда следует, что $f_{\mathcal{L}_k}(\tau) = C_{\tau-l}^{k-1}$.

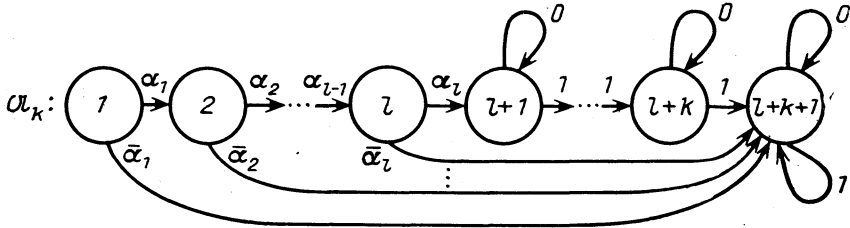


Рис. 2

Возьмем теперь дихотомическое дерево высоты l (l выбирается так, чтобы было выполнено условие теоремы 1) и выделим группы концевых вершин следующим образом (см. рис. 3): 1-я группа содержит a_0 вершин, 2-я группа содержит a_1 вершин, ..., $(s+1)$ -я группа содержит a_s вершин. Здесь a_0, a_1, \dots, a_s — коэффициенты разложения полинома $P_s(\tau)$ из теоремы 1, в силу которой $\sum_{i=0}^s a_i \leq 2^l$ и, следовательно, числа концевых вершин этого дерева достаточно для образования

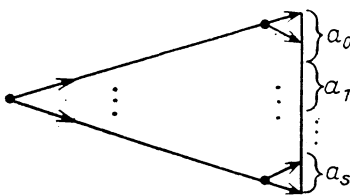


Рис. 3

из них таких групп вершин.

Вершины группы a_0 объявляются финальными, а к каждой вершине из группы i ($1 \leq i \leq s+1$) мы «приклеиваем» автомат, реализующий язык \mathcal{L}_i , для которого $f_{\mathcal{L}_i}(\tau) = C_{\tau-l}^i$, т. е. фактически мы отождествляем концевые вершины из группы i дерева с соответствующими инициальными состояниями автоматов,

реализующих язык \mathcal{L}_i . Финальные состояния этих автоматов объявляются финальными состояниями нового автомата. Все свободные ребра дерева подсоединяются к одному поглощающему (нефинальному) состоянию. С этим же состоянием отождествляются поглощающие состояния автоматов, реализующих языки \mathcal{L}_i . Из свойства префиксности дерева вытекает, что функция роста этого автомата \mathfrak{A} , реализующего язык \mathcal{L} , есть

$$f_{\mathcal{L}_i}(\tau) = a_0 + a_1 f_{\mathcal{L}_1}(\tau) + \dots + a_s f_{\mathcal{L}_s}(\tau) = a_0 + a_1 C_{\tau-l}^1 + \dots + a_s C_{\tau-l}^s,$$

что в свою очередь есть разложение полинома $P_s(\tau)$ из формулировки теоремы 1. Теорема 2 доказана.

Следствие. Если для $P_s(\tau)$ выполнено при всех $\tau \in \mathbb{N}$ соотношение $0 \leq P_s(\tau) \leq 2^\tau$, то существует регулярный язык \mathcal{L} такой, что $f_{\mathcal{L}}(\tau) = P_s(\tau)$.

Следствие вытекает из того, что ограничение $0 \leq P_s(\tau) \leq 2^\tau$ позволяет при $\tau < l$ «разбросать» по i -му ярусу дерева $P_s(i)$ финальных состояний.

Доказательство теоремы 3. Доказательство теоремы 2 распространяется для этого случая за счет выбора входного алфавита при условии совместности системы неравенств

$$P_s(1) \leq |A|, \quad P_s(2) \leq |A|^2, \quad \dots, \quad P_s(l-1) \leq |A|^{l-1}, \quad 2^l \leq |A|^l.$$

Доказательство теоремы 4. Пусть $f: \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$ — почти-периодическая функция. Тогда целочисленная последовательность $f(1), f(2), \dots, f(\tau), \dots$ имеет предпериод l и наименьший период T . Введем следу-

ющие функции:

$$f_i(\tau) = \begin{cases} 1, & \text{если } \tau = l + i + kT, \quad 1 \leq i \leq T, \quad k = 0, 1, 2, \dots, \\ 0 & \text{в остальных случаях;} \end{cases}$$

$$f_0(\tau) = \begin{cases} f(\tau), & 1 \leq \tau \leq l, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Очевидно следующее представление:

$$f(\tau) = f_0(\tau) + f(l+1)f_1(\tau) + \dots + f(l+T)f_T(\tau) = \sum_{i=1}^T f(l+i)f_i(\tau) + f_0(\tau). \quad (5)$$

Возьмем далее слово $\tilde{\beta}$ и семейство слов $\tilde{\alpha}_1^1, \dots, \tilde{\alpha}_{l+i}^1; \dots; \tilde{\alpha}_1^T, \dots, \tilde{\alpha}_{l+T}^T$ над входным алфавитом так, чтобы были выполнены следующие свойства.

1. Ни одно из этих слов (в том числе и $\tilde{\beta}$) не является префиксом другого слова.

2. Длины слов подчинены следующим соотношениям:

$$|\tilde{\alpha}_i^1| = \dots = |\tilde{\alpha}_{l+i}^1| = l + i; \quad |\tilde{\beta}| = T, \quad 1 \leq i \leq T. \quad (6)$$

Образует семейство языков $\mathcal{L}_i = \bigcup_{j=1}^{l+i} \mathcal{L}_{ij}$, где $\mathcal{L}_{ij} = \tilde{\alpha}_i^j \cup \tilde{\alpha}_j^i \cdot \{\tilde{\beta}\}^*$; здесь

$\cup, \cdot, *$ — операции объединения, конкатенации и, соответственно, итерации [1]. Учитывая свойство 2 выбранного семейства слов (6), имеем

$$f_{\mathcal{L}_{ij}}(\tau) = \begin{cases} 1, & \text{если } \tau = l + i + kT, \quad k = 0, 1, 2, \dots, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Значит, $f_{\mathcal{L}_{ij}}(\tau) = f_i(\tau)$, а из свойства 1 (все слова будут разные!) получаем, что

$$f_{\mathcal{L}_i}(\tau) = \sum_{j=1}^{l+i} f_{\mathcal{L}_{ij}}(\tau) = f(l+i)f_i(\tau).$$

В силу свойств 1–2 при $\mathcal{L}^1 = \bigcup_{i=1}^T \mathcal{L}_i$

$$f_{\mathcal{L}^1}(\tau) = \sum_{i=1}^T f_{\mathcal{L}_i}(\tau) = \sum_{i=1}^T f(l+i)f_i(\tau).$$

Нам осталось для получения представления (5) реализовать конечный язык \mathcal{L}_0 , в котором имеется $f(i)$ слов длины i при $i = 1, 2, \dots, l$; это делается с учетом того, что мы можем выбирать входной алфавит из любого конечного числа символов. Положим $\mathcal{L} = \mathcal{L}^1 \cup \mathcal{L}_0$ и заметим, что функция роста $f_{\mathcal{L}}(\tau)$ совпадает с $f(\tau)$. Константная ограниченность языка \mathcal{L} вытекает из свойства почти-периодичности функции. Докажем обратное. Пусть дан язык \mathcal{L} такой, что $0 \leq f_{\mathcal{L}}(\tau) \leq c$, где $c \geq 0$ — константа. Покажем, что $f_{\mathcal{L}}(\tau)$ — почти-периодическая функция. Для этого рассмотрим производящую функцию $f(\tau) = \sum_{\tau=1}^{\infty} a_{\tau}x^{\tau}$, где $a_{\tau} = f_{\mathcal{L}}(\tau)$. Известно (см., например, [9]), что $f(\tau)$ есть рациональная функция. В силу ограничения $0 \leq f_{\mathcal{L}}(\tau) \leq c$ и того, что $f_{\mathcal{L}}(\tau)$ — целое число, получаем, что коэффициенты степенного ряда $\sum_{\tau=1}^{\infty} a_{\tau}x^{\tau}$ принимают конечное число значений. Отсюда с учетом [10] получаем, что, начиная с некоторого номера, эти коэффициенты образуют почти-периодическую последовательность. Теорема 4 доказана.

В заключение автор пользуется случаем выразить признательность проф. В. Б. Кудрявцеву за внимание к работе и полезные замечания, способствовавшие улучшению работы.

СПИСОК ЛИТЕРАТУРЫ

1. Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов.— М.: Наука, 1985.
2. Строгалов А. С. Об одной мере близости для автоматных языков. // Проблемы теоретической кибернетики. Тезисы докладов VIII Всесоюзной конференции, июль 1988, ч. 2.— Горький.— С. 191.
3. Хомский М., Миллер Д. Языки с конечным числом состояний. // Киб. сборник, вып. 4.— М.: ИЛ, 1962.— С. 233—255.
4. Гилл А. Введение в теорию конечных автоматов.— М.: Наука, 1966.
5. Гантмахер Ф. Р. Теория матриц.— М.: Наука, 1988.
6. Бабенко К. И. Основы численного анализа.— М.: Наука, 1986.
7. Полиа Г., Сеге Г. Задачи и теоремы из анализа, ч. 2.— М.: Наука, 1985.
8. Гельфонд А. О. Исчисление конечных разностей.— М.: Наука, 1967.
9. Строгалов А. С. Об ε -моделировании поведения конечных автоматов. // Материалы I-го Всесоюзного семинара по дискретной математике и ее приложениям.— М.: Изд-во МГУ, 1986.— С. 42—56.
10. Бибербах Л. Аналитическое продолжение.— М.: Наука, 1967.

Статья поступила 01.02.90