Алгоритмическая структура и функционал программного комплекса «CutGlueApproximation»

P. А. Нейдорф¹, В. В. Полях², О. Т. Ярахмедов³, И. В. Черногоров⁴ Донской государственный технический университет ¹ran_pro@mail.ru, ²silvervpolyah@gmail.com, ³orhashka@gmail.com, ⁴hintaivr@gmail.com

Аннотация. Для построения математических моделей существенно нелинейных технических объектов результатам эксперимента разработан альтернативный «Cut-Glue» существующим метол аппроксимации. использующий разбиение массива экспериментальных данных на фрагменты, имеющие хорошо описываемую полиномиальными функциями структуру ланных. фрагменты. Полиномы. описывающие эти мультипликативно преобразуются, получением фрагментарно изолированных импульсоподобных функций, описывающих данные в пределах границ фрагментов, и практически нулевые в остальной области определения. Эти свойства позволяют получать математическую модель всей зависимости аддитивным объединением фрагментов. В излагаемом методе выделяются три задачи: фрагментации, аппроксимации фрагментов и их мультипликативноаллитивной обработки. Метод решения залачи фрагментирования ланных основан логикокомбинаторных методах, позволяющих исследовать область В допустимых решений. задаче математического полиномиального описания фрагментов данных находятся их структурно и параметрически оптимальные варианты аппроксимирующих полиномов. При этом используется классического регрессионного анализа модификации эволюционно-генетического Решение результирующей задачи метода осуществляется двумя операциями: 1) вырезанием по границам с помощью специальных мультипликативно применяемых функций из полиномов их интервально изолированных фрагментов, 2) аддитивное объединение этих фрагментов в единую математическую модель исследуемой зависимости. При этом осуществляется параметрическая минимизация модели. Описываемый метод реализует принципиально новый подход и разработан для решения задач экспериментального описания нелинейных зависимостей. Для его реализации разработан программный комплекс "CutGlueApproximation". В статье описываются сущность и возможности реализованного в комплексе функционала.

Ключевые слова: экспериментальные данные; моделирование; нелинейность; аппроксимация; математическая модель; оптимизация; эвристические алгоритмы; комбинаторика; программные средства

І. Введение

Построение математических моделей (MM) технических объектов часто связано с математической

Работа выполнена при финансовой поддержке РФФИ, проект №18-08-01178A

обработкой экспериментальных данных (ЭД). Для экспериментального построения ММ таких объектов разработано много методов аппроксимации данных: [1-4],полиномиальной регрессии ортогональных разложений [5, 6], кусочных функций [7, 8], сплайнов [9, 10], радиальных базисных функций [11, 12] и т.д. Но из-за нелинейностей описание таких зависимостей известными методами затруднительна, и связана с погрешностями. В связи с этим проведены работы по созданию альтернативного метода «Cut-Glue» аппроксимации (CGA) Он использует разбиение моделируемой зависимости на участки со структурой данных, хорошо аппроксимируемые полиномиальными функциями. К ним применяется мультипликативное «вырезание» фрагментов границам участков. А завершается процедура «склеиванием» фрагментов в единую функцию, которая моделирует численную структуру исходного массива данных. В методе CGA выделяются три аппроксимации залачи: фрагментации, мультипликативно-аддитивной обработки результатов двух первых задач [13-15]. Первая задача - это разбиение ЭД на фрагменты ЭД (ФЭД). Методы ее решения рассматриваются в подразделе III.А. Второй задачей CGA является описание ФЭД с получением для фрагмента оптимального варианта локально аппроксимирующей функции (ЛАФ). Она исследуется в подразделе III.В. Результирующий этап метода CGA состоит из двух операций: преобразования ЛАФ в т.н. интервально изолированные функции (ИИФ), аппроксимирующие данные в пределах границ ФЭД, и их аддитивная их композиция в единой аналитической функции (ЕАФ), являющейся ММ объекта. Сущность и алгоритмы выполнения этих операций описываются в подразделе III.С. Метод CGA реализует принципиально новые подходы к аппроксимации и разработан для решения задач экспериментального описания нелинейных объектов.

II. Постановка задачи

В связи с существенной разноплановостью этапов метода СGA, и в идеологическом, и в алгоритмическом аспектах, возникла необходимость выработать перспективные подходы, и к теоретическим, и к практическим методам их решения, построить и проверить основные алгоритмы. Таким образом, в статье должны быть решены следующие задачи:

- 1. разработать алгоритм вариативной декомпозиции ЭД на ФЭД применительно к двумерным ЭД;
- решить задачу оптимальной полиномиальной аппроксимации ФЭД;
- 3. алгоритмически обеспечить и исследовать этапы мультипликативно-аддитивного преобразования ФЭД с получением ММ ЕАФ.

III. АЛГОРИТМЫ РЕАЛИЗАЦИИ И ОПТИМИЗАЦИИ ЭТАПОВ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ НЕЛИНЕЙНЫХ ОБЪЕКТОВ МЕТОДОМ «CGA»

А. Логико-комбинаторный алгоритм фрагментации ЭД в задаче CGA.

Метод CGA рассчитан решение на залач аппроксимации ЭД, представляемых гиперпараллелипипедами п-мерном факторном В пространстве. Для разбиения массива ЭД на фрагменты – ФЭД, факторное пространство покрывается сеткой, построенной по покоординатным шагам варьирования (ПШВ-сеткой) входных переменных, выбранный при планировании эксперимента.

Главным ограничением на алгоритм фрагментации (АФ), накладываемым методом CGA [14–15], являются следующие 2 основополагающих условия:

 объединение всех фрагментов представляет собой все исследуемое пространство данных:

$$S_1 \cup S_2 \cup \dots \cup S_n = S, \tag{1}$$

где $S_i - i$ -й фрагмент разбиения; n -количество фрагментов; S -все множество ЭД (S -от «snippet»);

 пересечение соседних фрагментов есть подмножество элементов, образующих их общую границу

$$S_i \cap S_j = B_{ij} = B_{ji}, \tag{2}$$

где S_i и S_j — соприкасающиеся гранями B_{ij} и B_{ji} фрагменты, (B — от «border»)

Условие (1) очевидно, а условие (2) введено в связи с тем, что в дальнейшем, при реализации аддитивного объединения итоговых ИИФ, необходимо обеспечить предельно «бесшовное» соединение между ними [16]. Поскольку нет предпосылок требовать от ФЭД полного совпадения по взаимным границам, что показано на рис. 1, фрагментированное множество ЭД не может быть описано математическими структурами. регулярными Можно условно совокупность ФЭД считать «гранично пересекающимися нерегулярно-клеточными матрицами». При этом результатом будет множество ФЭД, связанное с исходным множеством ЭД соотношениями:

$$M = \prod_{j=1}^{n} m_j = \sum_{k=1}^{N} n_k$$
 (a), $n_k = \prod_{j=1}^{n} m_{k_j}$ (b), (3)

где M — количество данных в ЭД; N — количество фрагментов его разбиения — гиперпараллелипипедов, каждые из которых имеют общие значения по граням, не являющимся пограничными; m_j , m_{kj} — количества значений

фактора по j-ой оси в ЭД и в k-м ФЭД; n_k – количество данных в структуре разбиения.

Фрагментация приводит к разделению ПШВ-сетки на множество гиперпараллелипипедов с общими смежными данными по не пограничным граням, а реализация описанного алгоритма сводится к последовательному выполнению следующих операций:

- задается диапазон допустимых размеров фрагментов в ПШВ: h_1 и h_2 . Длина фрагментов h генерируется в интервале $h_1 \le h \le h_2$ по случайному закону распределения;
- ПШВ-сетка заполняется фрагментами в случайной последовательности с логической проверкой ограничений на размеры ФЭД, пока все пространство не будет заполнено.

На рис. 1 пошагово показан процесс заполнения Φ ЭД в простой тестовой задаче размерностью 12x12. Литерой d обозначены данные (d – от «data»).

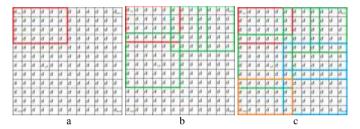


Рис. 1. Этапы работы программы: а – выделение первого фрагмента, b – промежуточный этап, с – финальный результат

Как первый вариант решения задач ограниченной размерности разработан логико-кодовый АФ (ЛКАФ), который показал довольно высокую эффективность. Спроектировано и разработано ПС «Experimental Data Fragmentation Method» (EDFM). В настоящее время ведутся изучение и дальнейшая модификация ЛКАФ для повышения его эффективности, а также для оценки границ применимости в рамках поставленной задачи.

В. Структурно-параметрическая оптимизация полиномиальной аппроксимации ФЭД

Методологической основой выполнения второго этапа CGA являются авторские алгоритмы, основанные на полиномиальной регрессии [8–11]. Степенные полиномы, имеющие регулярную структуру при любом порядке, приняты как универсальная форма ЛАФ для ФЭД. Это позволяет оценить сложность ЛАФ критериально по наличию и числу членов полинома произвольного порядка

$$Y(x) = b_0 + b_1 x_1 + \dots + b_n x_n + b_{11} x_1^2 + b_{12} x_1 x_2 + \dots$$

$$+ b_{1n} x_1 x_n + b_{22} x_2^2 + \dots + b_{2n} x_2 x_n + b_{(n-1)n} x_{n-1} x_n +$$

$$+ b_{nn} x_n^2 + b_{111} x_1^3 + \dots + b_{11n} x_1^2 x_n + \dots b_{122} x_1 x_2^2 + \dots$$

$$+ b_{nnn} x_n^3 + b_{1111} x_1^4 + \dots + b_{111n} x_1^3 x_n + b_{1122} x_1^2 x_2^2 + \dots$$

$$(4)$$

Регулярность полинома (4) упрощает структурное варьирование, а также возможность структурнопараметрической оптимизации (СПО) ЛАФ каждого

фрагмента. Для СПО аппроксимирующего степенного полинома (АСП) размерности n, разработан алгоритм, основанный на представления его псевдолинейным полиномом расширенной размерности n:

$$q(\tilde{x}) = \sum_{i=0}^{n} \tilde{b_i} \times \tilde{x_i} = \sum_{i=0}^{n} b_i \times x_i + \sum_{i=n+1}^{n} \hat{b_i} \times \hat{x_i} , \qquad (5)$$

где $\tilde{b_i}$ — коэффициенты псевдолинейного полинома \tilde{n} -ой размерности, а $\tilde{x_i}$ — обобщенные аргументы исследуемой зависимости, включающие, как исходные x_i , так и псевдоаргументы $\hat{x_i}$, которыми заменяются нелинейности степенного полинома.

Второй член полинома (5) содержит, фактически, зависимые аргументы. Поэтому его коэффициенты, вычисляемые по классической матричной формуле

$$\widetilde{b} = (X^T X)^{-1} X^T y, \tag{6}$$

являются лишь субоптимальными. Здесь $y\!-\!N_k$ - вектор значений k-го Φ ЭД; X -матрица входов \tilde{x}_{ji} ; $j \in [1, N_k]$ и i - номера опытов и псевдопеременных; $\tilde{x}_{i0} = 1$. Взаимозависимость части \tilde{x}_i обусловливает субоптимальность вычисляемых коэффициентов \tilde{b}_i , относительно используемого в КРА критерия МНК, но и он сам по себе является лишь эвристическим.

Поиск СПО варианта АСП повышает качество аппроксимации ФЭД, т.к. полный полином, зачастую, не дает наилучшей точности, из-за свойств ряда некоторых нелинейных членов, не согласующихся с особенностями кривизны аппроксимируемых зависимостей, задаваемых ФЭД. Алгоритм СПО АСП варьирует его структуру, оценивая, и ее сложность, и точности аппроксимации ФЭД. На множестве таких двумерных критериев решается задача дискретной СПО [14]. Если велика размерность объекта и/или количество ФЭД, комбинаторные методы становятся ресурсно не реализуемыми. Поэтому для СПО разработан модифицированный эволюционногенетический алгоритм (МЭГА). Им, например, решен отображенный ниже матрицами и табл. 1 пример решения задачи СПО АСП для конкретного ФЭД. Верхняя строка и левый столбец матриц (7) – значения x_1 и x_2 . Первая матрица – значения ФЭД, а вторая – его СПО ЛАФ.

$$\begin{pmatrix} 1500 & 2000 & 2500 & 3000 \\ 0 & 2438 & 4470 & 6632 & 8989 \\ 10 & 2238 & 4299 & 6550 & 8916 \\ 20 & 1732 & 3693 & 6060 & 8456 \\ 30 & 922,78 & 2911 & 5222 & 7884 \end{pmatrix} \Rightarrow \begin{pmatrix} 1500 & 2000 & 2500 & 3000 \\ 0 & 2441 & 4449 & 6672 & 8978 \\ 10 & 2236 & 4229 & 6360 & 8928 \\ 20 & 1731 & 3687 & 6064 & 8470 \\ 30 & 923 & 2902 & 5213 & 7901 \end{pmatrix}$$
 (7)

Верхняя строка табл. 1 содержит кодовые обозначения псевдопеременных, указывающие на номера Ниже приводятся структурные перемножаемых x_i . АСП в порядке vвеличения варианты ошибки аппроксимации. В столбцах таблицы представлены члены полного полинома 4-го порядка с соответствующим кодом, указанным в верхней строке. Их отсутствие обозначается символом "-". В двух последних колонках приведены оценки максимальной на множестве аппроксимируемых

данных абсолютной (Δ) и относительной (δ) ошибки, показывающие аппроксимационную точность вариантов полинома. Результаты табл. 1 убедительно демонстрируют эффективность структурной вариации АСП ФЭД при оптимизации его ММ.

ТАБЛИЦА І РЕЗУЛЬТАТЫ СПО АСП x_2

0	1	2	11	12	22	111	112	122	222
0	1	2	11	12		111		122	
0	1	2	11	12			112	122	
0	1	2					112		222

ПРОДОЛЖЕНИЕ ТАБЛИЦЫ 1

1111	1112	1122	1222	2222	Δ	δ
	1112		1222		21,85576	0,00523
1111			1222		22,94357	0,00549
1111		1122	1222	2222	31,73257	0,00608

C. Сущность и параметрическая оптимизация "Cut-Glue" этапа метода CGA

Последний этап метода CGA реализуется двумя операциями: «Cut the fragments» (CF) и «Glue the fragments» (GF) [13, 14]. Эти операции с достаточной общностью можно описать следующими выражениями:

CF:
$$\forall i = \overline{1, N} \to f_i(\vec{x}_i) = \phi_i(\vec{x}_i) \cdot \prod_{i=1}^n E_{i_i}(x_{i_i}, x_{i_{i_i}}, x_{i_{i_i}}, \varepsilon_{i_i}),$$
 (8)

GF:
$$F(\vec{x}) = \sum_{i=1}^{N} f_i(\vec{x}_i), \qquad (9)$$

где: $f_i(\vec{x}_i)-i$ -й ИИФ; N — количество фрагментов и их ИИФ; $\phi_i(\vec{x}_i)-i$ -я ЛАФ; n — факторная размерность объекта; $E_{i_j}(^*)-j$ -я одномерная функция для i-й ЛАФ; $x_{i_j}-i$ -й аргумент j-й $E_{i_j}(^*)$; $x_{i_{jL}}$, $x_{i_{jR}}$ — краевые значения диапазона аппроксимации i-го ФЭД по j-ой переменной; $\varepsilon_{i_{jL}}$ и $\varepsilon_{i_{jR}}$ — параметры крутизны фронтов i-й ИИФ по j-ой переменной.

Операция СF в (8) реализуется с использованием специальной мультипликативно выделяющей функции (МВФ или ε -функции) $E_{i_j}(*)$. Основой ее структуры являются ε -функции 1-го порядка или 1-МВФ:

$$E(x, x_L, x_R, \varepsilon_L, \varepsilon_R) = \frac{0.25 \cdot \sigma_L(x, x_L, \varepsilon_L) \cdot \sigma_R(x, x_R, \varepsilon_R)}{\delta(x, x_L, x_R, \varepsilon_L, \varepsilon_R)}, (10)$$

ГДЕ:
$$\delta(x, x_L, x_R, \varepsilon) = \sqrt{\left[\left(x - x_L\right)^2 + \varepsilon_L^2\right] \cdot \left[\left(x_R - x\right)^2 + \varepsilon_R^2\right]}$$
,
$$\sigma_R(x, x_R, \varepsilon) = x_R - x + \sqrt{\left(x_R - x\right)^2 + \varepsilon_R^2}$$
, $\sigma_L(x, x_L, \varepsilon) = x - x_L + \sqrt{\left(x - x_L\right)^2 + \varepsilon_L^2}$.

Свойства этой функции на границах фрагмента близки к сигнатурным. Это формирует фронты ИИФ $f_i(\vec{x}_i)$, отделяющие участки соседних ФЭД по координатам факторов (при многомерности $x{\sim}x_i$). Из ЛАФ $\phi_i(\vec{x}_i)$ 1-МВФ (10) «вырезает» ее участок на интервале фрагмента по x_i и приближает к нулю все значения ИИФ в (9) за границами «вырезанного» интервала. Это свойство МВФ позволяют следующей GF операции аддитивно объединить все ИИФ $f_i(\vec{x}_i)$ в единую функцию ЕАФ – ММ объекта.

Однако структура МВФ (10) не только обеспечивает эффективное объединение моделей фрагментов в единую модель [13–15]. Ее настройки ε_L , ε_R (в многомерном случае $\varepsilon_{i_{jL}}$ и $\varepsilon_{i_{jR}}$) позволяют снизить общую погрешность аппроксимации ЭД полученной ЕАФ. Возможность ее минимизации возникает из-за альтернативных тенденций зависимости пограничных значений 1-МВФ от $\varepsilon_{i_{j}}$. Исследования влияния свойств МВФ на результаты GF показали, что для влияния на точность аппроксимации массива ЭД, может понадобиться изменение значений $\varepsilon_{i_{jL}}$ и $\varepsilon_{i_{jR}}$ в широком диапазоне значений [13, 15].

Для решения задач параметрической минимизации погрешности аппроксимации ЭД итоговой ЕАФ любой размерности используется модифицированный под задачу эвристический метод роящихся частиц (МРЧ) [16-20]. В связи многофакторностью задачи, разработан алгоритм vсловной оптимизации (ПУО) методической погрешности совокупности операций СF-GF. Идея алгоритма ПУО состоит в том, чтобы найти близкие к оптимальным составляющие вектора ε в качестве стартовых, и максимально упростить задачу глобальной оптимизации ЕАФ. Согласно АПУМ на первом шаге осуществляется поиск условных оптимумов параметров $\mathcal{E}_{i_{j_{L}}}$ и $\mathcal{E}_{i-1_{j_{R}}}$ для пар соседних ИИФ с общей границей, при фиксации остальных варьируемых параметров. Далее осуществляется обход периметра фрагмента с заменой условно заданных значений ε_{i} на субоптимизированные, при необходимости, в цикле.

Для демонстрации эффективности операций СF и GF проведён эксперимент с использованием 4-х ФЭД. Массив 2-мерных ЭД разбит на 4 смежных фрагмента и каждый фрагмент аппроксимирован ЛАФ. Максимальная относительная погрешность составила \sim 7,84%. В табл. 2 показаны участки матрицы относительных ошибок (%) по границе склейки ИИФ и приграничные значения, полученные в ходе поэтапной оптимизации фрагментов. В каждом ФЭД вплотную к границам указаны полученные субэкстремальные значения вектора ε , для оптимизируемых границ.

ТАБЛИЦА II РЕЗУЛЬТАТ С**F** и **GF** операций и полученные экстремальные значения вектор ε

212	ElxR-	1,348	0.256	0,190	3,798	ε2xL -	1,440		A104.2
εlyL =	1,846	***	0,132	0,796	3,362	444	ε2vL =	-1,720	***
2.003	1,347	0,591	1,012	1,581	2.74	0,016	4,41	4,496	2,721
1,051	1.105	0.511	1,017	1.319	0.919	1,419	4.447	4,366	0.599
0,717	1,507	1,254	0,135	0,470	1,414	1,709	2,629	2,831	1,365
e3yR =	1,595	224			0.842		E4vR	1,352	124
	E3xR -	-0.133	1,849	1,796	1,854	ε4xL -	0.364	***	***
444		444	2,082	4,291	6,760	444	244	444	

Циклический обход границ стыковки ФЭД закончился после повторения значения максимальной относительной ошибки аппроксимации ~4,496%. Это доказывает эффективность предложенного алгоритма, упростившего задачу многомерной оптимизации до циклической двумерной, т.к. исходную максимальную ошибку первых этапов удалось снизить в 1,74 раза.

IV. ЗАКЛЮЧЕНИЕ

Разработанный комплекс ПС решения задач этапов СGA перспективность разработанного метода. Все этапы эффективно алгоритмизуются и программно реализуются. Несложный, но результативный логико-кодовый алгоритм фрагментации автоматизирует наиболее ресурсный этап СGA. Структурно-параметрическая оптимизация этапа аппроксимации фрагментов структуры минимизирует методическую погрешность первых двух этапов. При этом предложенный алгоритм начальной попарной 2-мерной оптимизации итоговой ЕАФ с финальной субоптимизации фрагмента с наибольшей ошибкой в процессе обхода, метод СGA доказал эффективность, т.к. исходная ошибка аппроксимации может быть существенно уменьшена.

Список литературы

- Rawlings J.O., Pantula S.O., Dickey D.A. Applied Regression Analysis: A Research Tool, Second Edition. 1998.
- [2] Bates D.M., Watts D.G., Wiley J. & Sons Nonlinear regression analysis and its applications / New York, 1988, 365 p.
- [3] Drapper N.R., Smith H., Wiley J. & Sons Applied regression analysis, vol. 1 / New York, 1981, 366 p.
- [4] Drapper N.R., Smith H., Wiley J. & Sons Applied regression analysis, vol. 2 / New York, 1981, 351 p.
- [5] Totik V. Orthogonal Polynomials. Surveys in Approximation Theory, vol. 1, 2005, PP. 70–125.
- [6] Khrushchev S. Orthogonal Polynomials and Continued Fractions From Euler's Point of View / Atilim University, Turkey: Cambridge University Press 2008.
- [7] Recent Developments In Generalized Analytic Functions And Their Applications. Ed. G.Giorgadze / Tbilisi, 2011.
- [8] Лоран П.Ж. Аппроксимация и оптимизация: Пер. с англ. Изд.-во "Мир", М.: 1975, 496 с.
- [9] Альберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения: Изд.-во "Мир", М.: 1972, 318 с.
- [10] Boor D.C. A practical guide to splines // Springer, 1978.
- [11] Sarra S.A., Kansa E.J., Davis Multiquadric Radial Basis Function Approximation Methods for the Numerical Solution of Partial Differential Equations // Marshall University and University of California, 2009, 220 p.
- [12] Buhmann, M.D. Radial Basis Functions: Theory and Implementations / Cambridge University Press, 2003.
- [13] Neydorf R., Bivariate "Cut-Glue" Approximation of Strongly Nonlinear Mathematical Models Based on Experimental Data // SAE Int. J. Aerosp. 2015. №8(1). PP. 47-54.
- [14] Neydorf R., Chernogorov I., Polyakh V., Yarakhmedov O., Goncharova J., Neydorf A. Formal Characterization and Optimization of Algorithm for the Modelling of Strongly Nonlinear Dependencies Using the Method "Cut-Glue" Approximation of Experimental Data // SAE Technical Paper. 2016. 2016-01-2033 issue. PP. 1-12.
- [15] Neydorf R., A. Neydorf. Technology of Cut-Glue Approximation Method for Modeling Strongly Nonlinear Multivariable Objects. Theoretical Bases and Prospects of Practical Application. SAE Technical Paper 2016-01-2035, 2016
- [16] Neydorf R., Chernogorov I., Polyakh V., Yarakhmedov O., Goncharova Y., Vucinic D. Comparative Analysis of Heuristic Algorithms for Solving Multiextremal Problems // International Journal on Advances in Systems and Measurements. vol. 10, №1&2. 2017. PP. 86-99.
- [17] Cieselski Z., Musielak J. Spline bases in function spaces / Approximation theory, Reidel, 1975, PP. 49–54.
- [18] Eberhart R.C., Kennedy J.A. New optimizer, using particle swarm theory // Proceedings of the 6th International Symposium on Micromachine and Human Science. Nagoya, Jap. 1995. PP. 39-43.
- [19] Kennedy J., Eberhart R. Particle Swarm Optimization // Proceedings of IEEE International Conference on Neural Networks. Piscataway, New Jersey. 1995. PP. 1942-1948.
- [20] Shi Y., Eberhart R.C. A modified particle swarm optimizer // Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway, New Jersey. 1998. PP. 69-73.