ИНТЕРВАЛЬНЫЕ МЕТОДЫ ДЛЯ РЕШЕНИЯ УРАВНЕНИЙ И НЕОБХОДИМОСТЬ ПЕРЕФОРМУЛИРОВКИ ЗАДАЧИ

С. П. Шарый

Институт вычислительных технологий СО РАН
г. Новосибирск, Россия
(shary@ict.nsc.ru)

Аннотация

Работа представляет собой критический обзор современных интервальных методов для доказательного глобального решения уравнений и систем уравнений. Основываясь на теории топологической степени отображения и на понятии "эпсилон-решения", предложена переформулировка задачи решения уравнений, устойчивая к малым возмущениям в данных и, следовательно, более адекватно учитывающая приближённый характер как исходных данных задачи, так и процесса вычислений на цифровых ЭВМ с конечной разрядной сеткой. Намечены пути адаптации существующих интервальных методов для решения новой постановки задачи.

1 Постановка задачи

Предметом рассмотрения нашей работы является задача решения системы уравнений

$$\left\{\begin{array}{rcl} F_1(\,x_1, x_2, \ldots, x_n) & = & 0, \\ F_2(\,x_1, x_2, \ldots, x_n) & = & 0, \\ \vdots \qquad \ddots & & \vdots \\ F_n(\,x_1, x_2, \ldots, x_n) & = & 0, \end{array}\right. \tag{1}$$

над полем вещественных чисел $\mathbb{R}$, или, кратко,

$$F(x) = 0, \tag{2}$$

где $x = (\,x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ — вектор неизвестных переменных,
$F_i(x)$, $i = 1, 2, \ldots, n$ — некоторые вещественнозначные функции,
$F(x) = (\,F_1(x), F_2(x), \ldots, F_n(x)\,)^\top$ — вектор-столбец из F_i.

Всюду далее мы предполагаем, что функции $F_i(x)$ по меньшей мере непрерывны, а количество уравнений в системе (1)–(2) совпадает с количеством неизвестных переменных.

Задача решения уравнений и систем уравнений является одной из классических задач численного анализа, с изучения которой начинается, как правило, курс методов вычислений в высших учебных заведениях. Тогда же учащиеся узнают об основных подходах к решению этой задачи — методе простой итерации, методе Ньютона, их модификациях и т.п. Преимущества и недостатки этих классических методов хорошо известны, и мы не будем останавливаться на их подробном обсуждении. Для дальнейшего нам важны лишь два факта:

- для отображений F, не обладающих "хорошими" глобальными свойствами, все традиционные методы имеют *локальный характер*, т.е. обеспечивают сходимость к решению лишь из некоторой (иногда достаточно малой) окрестности этого решения;
- гарантированные оценки погрешности найденного приближения к решению в традиционных методах дать весьма непросто.

Указание приближённого значения величины и его максимальной погрешности равносильно тому, что мы знаем левую и правую границы возможных значений этой величины, и поэтому можно переформулировать нашу задачу в следующем усиленном виде —

Для каждого решения системы уравнений $F(x) = 0$ на данном множестве $D \subseteq \mathbb{R}^n$ найти гарантированные двусторонние границы

, (3)

— который будем называть *задачей доказательного глобального решения* системы уравнений. Она оказывается чрезвычайно сложной, а в классическом численном анализе почти полностью отсутствуют развитые методы для её решения. Из часто используемых подходов, имеющих ограниченный успех, следует упомянуть *аналитическое исследование*, *мультистарт*, *методы продолжения*.

2 Интервальные методы в решении уравнений

В то же время, в последние два десятилетия ушедшего века появились и быстро набрали силу *интервальные методы* решения задачи (3), основывающиеся на специфическом аппарате интервального анализа и смежных с ним дисциплин (в первую очередь, информатики).

2.1 Основы интервальной техники

Одномерные *интервалы* — это замкнутые отрезки вещественной оси $\mathbb{R}$, многомерные интервалы (называемые также *брусами*) — прямые произведения одномерных интервалов.

Говорят, что некоторая величина имеет *интервальную неопределённость*, если её точное значение не задано, но, тем не менее, известно, что оно лежит в пределах некоторого интервала.

Интервальный анализ может быть кратко определен как научная дисциплина на стыке математики и информатики, которая

- имеет дело с задачами, где присутствуют интервальные или, более общо, ограниченные по величине неопределённости в данных,

- рассматривает множества неопределённости как самостоятельные целостные объекты, устанавливая между ними арифметические и аналитические операции, отношения и т.п.

Типичными задачами интервального анализа являются, например, такие:

1) Найти *гарантированные двусторонние границы* решения (интервал, содержащий решение) для данной вещественной задачи.

2) Найти *внешнюю оценку* (объемлющий интервал, или семейство интервалов, или что-либо ещё) множества решений задачи с интервальными параметрами.

3) Найти *внутреннюю оценку* (интервальное или какое-либо другое подмножество) множества решений задачи с интервальными параметрами.

В частности, интересующая нас задача доказательного глобального решения систем уравнений (3) охватывается первым пунктом этого списка и находится в сфере применимости методов интервального анализа.

Одним из основных инструментов интервального анализа являются так называемые *интервальные арифметики* — алгебраические системы, формализующие операции между интервалами. В частности, наиболее популярная *классическая интервальная арифметика* — это алгебраическая система $\mathbb{IR}$, образованная интервалами $\mathbf{x} = [\underline{\mathbf{x}}, \overline{\mathbf{x}}] \subseteq \mathbb{R}$ так, что арифметические операции определены в соответствии со следующим фундаментальным принципом:

$$\mathbf{x} \star \mathbf{y} = \{\, x \star y \mid x \in \mathbf{x},\ y \in \mathbf{y} \,\} \qquad \text{для} \quad \star \in \{\, +\,, -\,, \cdot\,, /\,\},$$

— результирующий интервал любой операции есть множество всевозможных результатов этой операции между представителями интервалов-операндов. Явные формулы для интервальных сложения, вычитания, умножения и деления выглядят следующим образом:

$$\begin{aligned}
\mathbf{x} + \mathbf{y} &= [\,\underline{\mathbf{x}} + \underline{\mathbf{y}}, \overline{\mathbf{x}} + \overline{\mathbf{y}}\,], \\
\mathbf{x} - \mathbf{y} &= [\,\underline{\mathbf{x}} - \overline{\mathbf{y}}, \overline{\mathbf{x}} - \underline{\mathbf{y}}\,], \\
\mathbf{x} \cdot \mathbf{y} &= [\,\min\{\underline{\mathbf{x}}\,\underline{\mathbf{y}}, \underline{\mathbf{x}}\,\overline{\mathbf{y}}, \overline{\mathbf{x}}\,\underline{\mathbf{y}}, \overline{\mathbf{x}}\,\overline{\mathbf{y}}\}, \max\{\underline{\mathbf{x}}\,\underline{\mathbf{y}}, \underline{\mathbf{x}}\,\overline{\mathbf{y}}, \overline{\mathbf{x}}\,\underline{\mathbf{y}}, \overline{\mathbf{x}}\,\overline{\mathbf{y}}\}\,], \\
\mathbf{x} / \mathbf{y} &= \mathbf{x} \cdot [\,1/\overline{\mathbf{y}}, 1/\underline{\mathbf{y}}\,] \qquad \text{для} \;\; \mathbf{y} \not\ni 0.
\end{aligned}$$

Интервальные арифметические операции обладают важным свойством *монотонности по включению*:

$$\mathbf{x} \subseteq \mathbf{x}',\ \mathbf{y} \subseteq \mathbf{y}' \quad \Longrightarrow \quad \mathbf{x} \star \mathbf{y} \ \subseteq\ \mathbf{x}' \star \mathbf{y}' \qquad \text{для} \ \star \in \{\,+, -, \cdot, /\,\}.$$

Естественным образом определяются

$$\begin{aligned} \text{середина интервала — } \quad & \operatorname{mid}\mathbf{x} = \tfrac{1}{2}(\,\overline{\mathbf{x}} + \underline{\mathbf{x}}), \\ \text{ширина интервала — } \quad & \operatorname{wid}\mathbf{x} = \overline{\mathbf{x}} - \underline{\mathbf{x}}, \\ \text{абсолютное значение интервала — } \quad & |\mathbf{x}| = \max\{\,|\underline{\mathbf{x}}|, |\overline{\mathbf{x}}|\,\}. \end{aligned}$$

Интервальный вектор определяется как вектор, столбец или строка, с интервальными компонентами. Его геометрическим образом является прямоугольный параллелепипед в $\mathbb{R}^n$ со сторонами, параллельными координатным осям, который, как уже упоминалось, часто называют *брусом*. Сумма (разность) двух интервальных векторов одного размера — это вектор из поэлементных сумм (разностей) операндов. Топология на пространстве $\mathbb{IR}^n$ всех n-мерных интервальных векторов задаётся метрикой

$$\operatorname{dist}(\mathbf{x}, \mathbf{y}) = \max\{\|\underline{\mathbf{x}} - \underline{\mathbf{y}}\|, \|\overline{\mathbf{x}} - \overline{\mathbf{y}}\|\},$$

где $\underline{\mathbf{x}}$, $\underline{\mathbf{y}}$ — это векторы нижних концов, $\overline{\mathbf{x}}$, $\overline{\mathbf{y}}$ — векторы верхних концов для интервальных векторов $\mathbf{x}, \mathbf{y} \in \mathbb{IR}^n$ соответственно, а $\|\cdot\|$ — норма в $\mathbb{R}^n$. Операции взятия середины, ширины и абсолютного значения применяются к интервальным векторам покомпонентно. Аналогичным покомпонентным образом понимается и упорядочение интервальных векторов и матриц (по включению и т.п.).

Важнейшей задачей многих математических теорий является определение (описание, оценивание, приближение и т.п.) областей значений тех или иных конкретных функций, отображений и пр. на заданных множествах. В интервальном анализе эта задача принимает специфическую форму задачи о нахождении так называемого *интервального расширения* функции.

Определение [11, 32, 33]. *Интервальная функция* $\mathbf{f} : \mathbb{IR}^m \to \mathbb{IR}^n$ *называется интервальным расширением вещественной функции* $f : \mathbb{R}^m \to \mathbb{R}^n$, *если*

1) $\mathbf{f}(x) = f(x)$ *для* $x \in \mathbb{R}^m$,
2) $\mathbf{f}(\mathbf{x})$ *монотонна по включению, т.е.* $\mathbf{x} \subseteq \mathbf{x}' \;\Rightarrow\; \mathbf{f}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{x}')$.

Таким образом, если $\mathbf{f}(\mathbf{x})$ — интервальное расширение функции $f(x)$, то для области значений f на брусе $\mathbf{X} \subset \mathbb{R}^n$ мы получаем следующую внешнюю (с помощью объемлющего множества) оценку:

$$\{\, f(x) \mid x \in \mathbf{X}\,\} \subseteq \mathbf{f}(\mathbf{X}).$$

Эффективное построение интервальных расширений функций — это важнейшая задача интервального анализа, поиски различных решений которой продолжаются и в настоящее время. Тем не менее, чтобы у читателя не возникло ощущение сложности или неконструктивности рассматриваемого понятия, в рамках нашего беглого обзора уместно привести один из первых результатов в этом направлении. Иногда его высокопарно называют “основной теоремой интервальной арифметики”:

Теорема [2, 11, 27, 30, 32, 33]. *Если для рациональной функции* $f(x)$ *на интервале* $\mathbf{x}$ *определён результат* $\mathbf{f}(\mathbf{x})$ *подстановки вместо её аргументов интервалов их изменения и выполнения всех действий над ними по правилам интервальной арифметики, то*

$$\{\, f(x) \mid x \in \mathbf{x}\,\} \subseteq \mathbf{f}(\mathbf{x}),$$

т.е. $\mathbf{f}(\mathbf{x})$ *содержит множество значений функции* $f(x)$ *на* $\mathbf{x}$.

Интервальное расширение $\mathbf{f}(\mathbf{x})$ рациональной функции $f(x)$, о котором идёт речь в Теореме 1, называется *естественным интервальным расширением*, и вычисляется совершенно элементарно. За дальнейшей информацией мы отсылаем заинтересованного читателя к книгам [2, 11, 27, 30, 32, 33], развёрнуто излагающим проблему построения интервальных расширений функций.

2.2 Доказательное решение уравнений

Теоретическую основу доказательного решений систем уравнений интервальными методами составляют многие результаты математического анализа, среди которых типичными и наиболее часто используемыми являются следующие:

Теорема Лерэ-Шаудера[1]
Пусть $F : \mathbb{R}^n \supseteq D \to \mathbb{R}^n$ *— функция, непрерывная на области* D*, содержащей начало координат и имеющей границу* ∂D*. Если* $0 \in D$ *и из*

$$\tilde{x} \in \partial D,\ t \in \mathbb{R}, \quad F(\tilde{x}) = t\tilde{x},$$

следует $t \geq 0$*, то* F *имеет нуль в* D*, т.е. существует такая* $x^* \in D$*, что* $F(x^*) = 0$.

Теорема Брауэра о неподвижной точке
Пусть D *— выпуклый компакт в* $\mathbb{R}^n$*. Если непрерывное отображение* $T : \mathbb{R}^n \to \mathbb{R}^n$ *переводит* D *в себя, т.е.*

$$T(D) \subseteq D,$$

то оно имеет на D *неподвижную точку: существует такая* $x^* \in D$*, что*

$$x^* = T(x^*).$$

Теорема Миранды
Пусть функция $F : \mathbb{R}^n \to \mathbb{R}^n$ *непрерывна на брусе* $\mathbf{X} \subset \mathbb{R}^n$ *и такова, что при каждом* $i = 1, 2, \ldots, n$

$$F_i(x) \leq 0 \quad \textit{для} \quad x \in (\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \underline{\mathbf{X}}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n),$$

$$F_i(x) \geq 0 \quad \textit{для} \quad x \in (\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \overline{\mathbf{X}}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n).$$

Тогда $F(x)$ *имеет нуль в* $\mathbf{X}$*, т.е. существует такая* $x^* \in \mathbf{X}$*, что* $F(x^*) = 0$.

Теорема Миранды — это многомерный аналог теоремы Больцано-Коши о промежуточном значении [10], утверждающей существование нуля у непрерывной функции, которая принимает на концах отрезка разные знаки.

[1] Часто формулировку теоремы Лерэ-Шаудера относят к произвольным нормированным пространствам (см., например, [22]). Чтобы не отягощать изложение, мы не приводим максимально общую формулировку, так как работаем исключительно в конечномерных пространствах.

В классическом математическом анализе теоремы Лерэ-Шаудера, Брауэра и Миранды носили статус "чистых теорем существования", не указывающих конкретного способа нахождения того объекта — неподвижной точки или нуля отображения, — существование которого они утверждали. Даже сама проверка выполнения условий этих теорем могла быть практически выполнена лишь в ограниченном числе частных случаев, так как упиралась в необходимость нахождения области значений функции.

Имея в своем арсенале средства для вычисления интервальных расширений функций, мы можем обойти последнее затруднение, если

во-первых, ограничим себя рассмотрением областей D в виде интервальных векторов-брусов (это относится к теоремам Лерэ-Шаудера и Брауэра), и,

во-вторых, заменим точную область значений функции её внешней оценкой.

На этом пути получаются широко используемые в интервальном анализе теоремы Кравчика [27, 30, 33], Мура [31], Куи [36], Ньютона [33] и Хансена-Сенгупты [27, 33]. Первые три — это специализированные варианты теремы Брауэра, а последние — следствия теоремы Лерэ-Шаудера. Более того, с использованием интервальной техники возможно дальнейшее уточнение границ решения системы уравнений. Покажем, как это делается.

Итак, пусть к решению предъявлена система уравнений

$$F(x) = 0 \tag{2}$$

на брусе $\mathbf{X} \subset \mathbb{R}^n$. Если интервальное расширение $\mathbf{F}(\mathbf{X})$ функции F на $\mathbf{X}$ таково, что $0 \notin \mathbf{F}(\mathbf{X})$, то, очевидно, на $\mathbf{X}$ нет решений рассматриваемой системы уравнений.

Перепишем исходную систему (2) в равносильной рекуррентной форме:

$$x = T(x) \tag{4}$$

с некоторым отображением $T : \mathbb{R}^n \to \mathbb{R}^n$, которое может быть взято в виде

$$T(x) = x - F(x)$$

либо

$$T(x) = x - C \cdot F(x),$$

с невырожденной $n \times n$-матрицей C, либо как-нибудь ещё. Ясно, что решения системы (4) могут лежать лишь в пересечении $\mathbf{X} \cap T(\mathbf{X})$. Поэтому если $\mathbf{T}(\mathbf{X})$ — интервальное расширение функции T на брусе $\mathbf{X}$ и

$$\mathbf{T}(\mathbf{X}) \cap \mathbf{X} = \emptyset,$$

то в $\mathbf{X}$ решений системы уравнений (4) нет.

Напротив, если для бруса $\mathbf{X}$ выполнено

$$\mathbf{T}(\mathbf{X}) \subseteq \mathbf{X},$$

то в $\mathbf{X}$ по теореме Брауэра гарантированно находится решение системы (4). Более точно, это решение содержится и в $\mathbf{T}(\mathbf{X})$, и часто для дальнейшего уточнения мы можем организовать итерации с пересечением

$$\mathbf{X}^{(k+1)} := \mathbf{T}(\mathbf{X}^{(k)}) \cap \mathbf{X}^{(k)}, \qquad k = 1, 2, \dots . \tag{5}$$

Таким образом, наихудшим, с точки зрения уточнения информации о решении системы, является случай $\mathbf{T}(\mathbf{X}) \supsetneqq \mathbf{X}$.

Приведённую выше последовательность действий по обнаружению решений системы и уточнению его границ мы будем называть далее кратко *тестом существования.* В действительности, любой из изложенных выше приёмов уточнения решения допускает далеко идущие модификации и улучшения. Например, это относится к итерациям вида (5), которые могут быть последовательно применены не ко всему брусу $\mathbf{X}$ целиком, а к каждой отдельной его компоненте в комбинации с различными способами приведения исходной системы к рекуррентному виду (4). На этом пути мы приходим к чрезвычайно эффективным методам *распространения ограничений*[2] (см., к примеру, [25]).

Как простейший тест существования, так и его более продвинутые варианты без особых проблем реализуются на ЭВМ и работают тем лучше, чем более качественным является вычисление интервального расширения функций F в (2) и T в (4) и чем меньше ширина бруса $\mathbf{X}$. Последнее связано с тем, что погрешность оценивания области значений функции посредством практически любого интервального расширения убывает с уменьшением размеров бруса, на котором производится это оценивание.

2.3 Глобальное решение уравнений

Если же ширина бруса $\mathbf{X}$ не является достаточно малой, то на нём описанная выше в §2.2 методика может оказаться малоуспешной в том смысле, что она не приведёт ни к какому определенному заключению о существовании решения на $\mathbf{X}$ или же его отсутствии, и кроме того, сам этот брус нисколько не будет уточнён (уменьшен). Тогда практикуют принудительное дробление $\mathbf{X}$ на более мелкие подбрусы. Наиболее популярна при этом *бисекция* — разбиение бруса $\mathbf{X}$ на две (равные или неравные) части вдоль какой-нибудь грани, например, на половинки

$$
\begin{aligned}
\mathbf{X}' &= \big(\, \mathbf{X}_1, \ldots, [\, \underline{\mathbf{X}}_k, \operatorname{mid} \mathbf{X}_k \,], \ldots, \mathbf{X}_n \big), \\
\mathbf{X}'' &= \big(\, \mathbf{X}_1, \ldots, [\, \operatorname{mid} \mathbf{X}_k, \overline{\mathbf{X}}_k \,], \ldots, \mathbf{X}_n \big)
\end{aligned}
$$

для некоторого номера $k \in \{1, 2, \ldots, n\}$. При этом подбрусы $\mathbf{X}'$ и $\mathbf{X}''$ называются *потомками* бруса $\mathbf{X}$. Далее эти потомки можно разбить ещё раз, и ещё ... столько, сколько необходимо для достижения желаемой малости их размеров, при которой мы сможем успешно выполнять на этих брусах тесты существования из §2.2.

Если мы не хотим упустить при этом ни одного решения системы, то должны хранить все возникающие в процессе такого дробления подбрусы, относительно которых тестом существования не доказано строго, что они не содержат решений. Организуем поэтому *рабочий список* $\mathcal{L}$ из всех потомков начального бруса $\mathbf{X}$, подозрительных на содержание решений. В целом же алгоритм глобального доказательного решения системы уравнений организуем в виде повторяющейся последовательности следующих действий:

---

[2] В русской литературе этот подход часто называется также *недоопределёнными вычислениями*, термином, введённым А.С. Нариньяни [17].

– извлечение некоторого бруса из списка $\mathcal{L}$,

– дробление этого бруса на потомки,

– проверка существования решений в каждом из подбрусов-потомков, по результатам которой мы

либо выдаём этот подбрус в качестве ответа,

либо заносим его в рабочий список $\mathcal{L}$ для дальнейшей обработки,

либо исключаем из дальнейшего рассмотрения.

Кроме того, чтобы обеспечить ограниченность времени работы алгоритма, на практике имеет смысл задаться некоторым порогом мелкости (малости размеров) брусов δ, при достижении которого дальше дробить брус уже не имеет смысла. Псевдокод получающегося алгоритма приведён в Таблице 1.

Т а б л и ц а 1

Простейший интервальный алгоритм
доказательного глобального решения уравнений.

```
инициализируем список L  начальным брусом X
DO WHILE  ( L ≠ ∅ )
        извлекаем из L брус Y
        IF (  диаметр Y < δ ) THEN
                 выдаём брус Y в качестве ответа
                 в соответствующий из списков I–III
        ELSE
                 рассекаем брус Y на потомки Y′ и Y″
        END IF
        IF ( 0 ∈ F(Y′) ) THEN
                 применяем тест существования к Y′ ;
                 если результат непуст, заносим его в L
        END IF
        IF ( 0 ∈ F(Y″) ) THEN
                 применяем тест существования к Y″ ;
                 если результат непуст, заносим его в L
        END IF
END DO
```

Отметим, что неизбежные ограничения на вычислительные ресурсы ЭВМ могут воспрепятствовать решению этим алгоритмом задачи (3) “до конца”, поскольку могут возникнуть ситуации, когда

1) размеры обрабатываемого бруса уже $< \delta$, но нам ещё не удается ни доказать существование на нём решений, ни показать их отсутствие;

2) размеры обрабатываемого бруса ещё $> \delta$, но вычислительные ресурсы уже не позволяют производить его обработку дальше: исчерпались время, память и т.п.

В реальных вычислениях остановка алгоритма Таблицы 1 может происходить поэтому не только при достижении пустого рабочего списка $\mathcal{L}$ (когда исчерпана вся область поиска решений), и в общем случае результатом его работы является

I) список брусов шириной меньше δ, гарантированно содержащих решения,

а также

II) список брусов шириной меньше δ, подозрительных на содержание решения,

III) список недообработанных брусов, имеющих ширину больше δ,

такие что все решения системы уравнений, не принадлежащие списку I, содержатся в списках II и III.

Алгоритмы описанного выше типа, дополненные различными усовершенствованиями, получили большое развитие в интервальном анализе в последние десятилетия (см., например, книги [27, 30, 32, 33]). Но вычислительная практика выявила и ряд их слабых сторон и даже неожиданные и плохообъяснимые, на первый взгляд, феномены. В частности, на весьма простых примерах список II (а иногда даже и список III) получался неожиданно длинным, причём ни при уменьшении δ, ни при увеличении времени и точности вычислений этот эффект не исчезал. Чем это вызвано? Ответ на поставленный вопрос потребовал более глубокого анализа ситуации.

3 Вычислительно-корректные задачи

3.1 Формальные определения

Напомним общеизвестный факт: на вычислительных машинах (как электронных, так и механических, как цифровых, так и аналоговых) в условиях приближённого представления входных числовых данных и приближённого характера вычислений над полем вещественных чисел $\mathbb{R}$ мы в принципе можем решать лишь те постановки задач, ответы которых непрерывно зависят от входных данных, т.е. устойчивы по отношению к возмущениям в этих начальных данных. Дело в том, что решение задачи на любой вычислительной машине сопровождается неизбежными ошибками и погрешностями, вызванными конечным характером представления чисел, конечностью исполнительных устройств и т.п. Потенциально эти погрешности могут быть сделаны сколь угодно малыми, но в принципе избавиться от них не представляется возможным. Получается, что реально мы решаем на вычислительной машине не исходную математическую задачу, а более или менее близкую к ней.

Возникновение и бурное развитие компьютерной алгебры с её “безошибочными” вычислениями едва ли опровергает высказанный выше тезис, так как исходные постановки задач для систем символьных преобразований требуют *точную* представимость входных данных, которые поэтому подразумеваются целыми или, на худой конец, рациональными с произвольной длиной числителя и знаменателя (см. [1]), а все преобразования над ними не выводят за пределы поля рациональных чисел.

Под *массовой задачей* [9] будем понимать некоторый общий вопрос, формулировка которого содержит несколько свободных переменных — *параметров* — могущих изменяться в пределах предписанных им множеств. В целом задача П определяется

1) общим списком всех её параметров с областями их определения,

2) формулировкой тех свойств, которым должен удовлетворять *ответ*, т.е. решение задачи.

Индивидуальная задача I получается из массовой задачи П если всем параметрам задачи П присвоить конкретные значения. Наконец, *разрешающим отображением* задачи П назовём отображение, сопоставляющее каждому набору входных данных-параметров ответ соответствующей индивидуальной задачи.

Определение. *Станем говорить, что массовая математическая задача является вычислительно корректной, если её разрешающее отображение $\mathcal{P} \to \mathcal{A}$ из множества входных данных $\mathcal{P}$ во множество $\mathcal{A}$ ответов задачи непрерывно относительно некоторых топологий на $\mathcal{P}$ и $\mathcal{A}$, определяемых содержательным смыслом задачи.*

Те задачи, ответы на которые неустойчивы по отношению к возмущениям входных данных, могут решаться на ЭВМ с конечной разрядной сеткой лишь опосредованно, после проведения мероприятий, необходимых для защиты от этой неустойчивости или её нейтрализации.

Например, задача решения систем линейных уравнений с невырожденной квадратной матрицей является вычислительно-корректной. Если топология на евклидовом пространстве её решений задается обычной метрикой и подобным же традиционным образом задается расстояние между векторами правой части и матрицами, то существуют хорошо известные неравенства [6], оценивающие сверху границы изменения решений через изменения элементов матрицы, правой части и число обусловленности матрицы.

Для систем нелинейных уравнений, могущих иметь неединственное решение, топологию на множестве $\mathcal{A}$ нужно задавать уже каким-либо расстоянием между множествами, например, с помощью так называемой *хаусдорфовой метрики* (см. [5]).

Если ϱ — метрика на рассматриваемом пространстве, то *расстоянием* точки a до множества X называется величина $\varrho\,(a, X)$, определяемая как

$$\inf_{x\in X} \varrho\,(a, x).$$

Хаусдорфовым расстоянием между компактными множествами X и Y называют величину

$$\varrho\,(X, Y) = \max\Big\{ \max_{x\in X} \varrho\,(x, Y),\ \max_{y\in Y} \varrho\,(y, X) \Big\}.$$

При этом $\varrho\,(X, Y) = +\infty$, если $X = \emptyset$ или $Y = \emptyset$. Введённая таким образом величина действительно обладает всеми свойствами расстояния и может быть использована для задания топологии на пространствах решений тех задач, ответы к которым неединственны.

3.2 Задача решения уравнений не является вычислительно-корректной

Уже простейшие примеры показывают, что задача решения уравнений и систем уравнений не является вычислительно-корректной. Например, квадратное уравнение

$$x^2 + px + q = 0 \tag{6}$$

для

$$p^2 = 4q \tag{7}$$

имеет лишь одно решение $x = -p/2$. Но при любых сколь угодно малых возмущениях коэффициента p и свободного члена q, нарушающих равенство (7), уравнение (6) теряет это единственное решение или же приобретает ещё одно (см. Рис. 1).

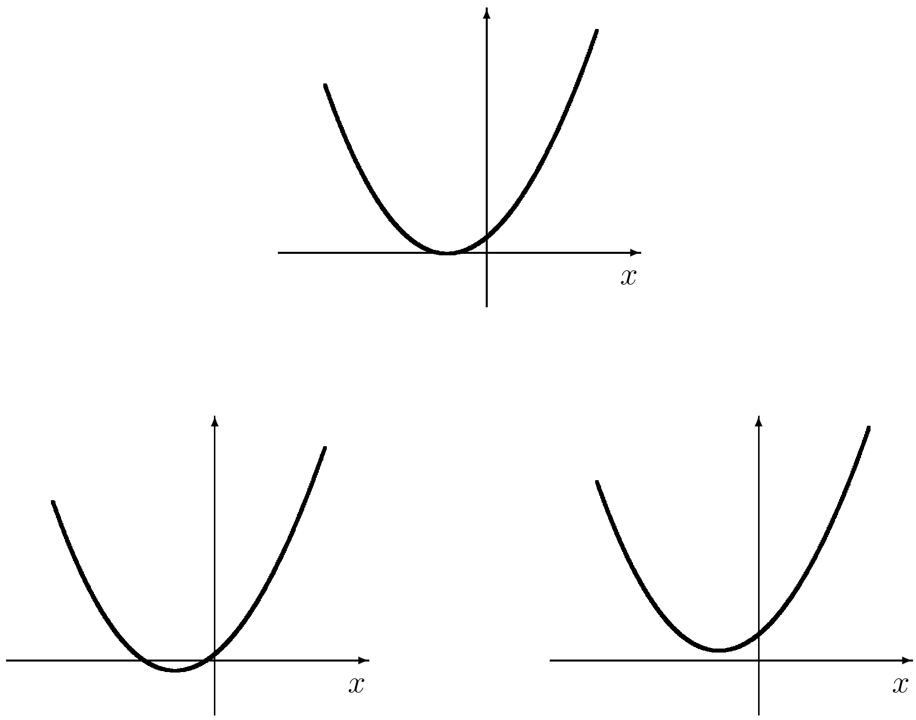

Рис. 1: Неустойчивая зависимость решений уравнения (6)–(7) от сколь угодно малых шевелений его коэффициентов.

Аналогичным образом ведёт себя решение эквивалентной двумерной системы

$$\begin{cases} x + y = r, \\ xy = s \end{cases}$$

при $s = r^2/4$. При этом раздвоение решения не является большим грехом, коль скоро мы можем рассматривать хаусдорфово расстояние между целостными множествами решений. Но вот исчезновение единственного решения — это чрезвычайное событие, однозначно указывающее на разрывность разрешающего отображения.

Итак, математическую постановку задачи нахождения решений уравнений нужно скорректировать, заменив какой-нибудь вычислительно-корректной постановкой задачи. Приступая к поиску ответа на этот чисто математический вопрос, отметим, прежде всего, что с точки зрения практических приложений задачи, которые мы обычно формулируем в виде решения уравнений или систем уравнений, традиционно выписывая

$$F(x) = 0, \tag{2}$$

имеют весьма различную природу. Это и будет отправной точкой нашей ревизии постановки задачи (3).

3.3 ε-решения уравнений

В ряде практических задач пользователю требуется не точное равенство некоторого выражения нулю, а лишь его “исчезающая малость” в сравнении с каким-то a priori установленным порогом.

Таковы, например, уравнения материального баланса в большинстве физических, химических и других естественнонаучных расчётов. Точное равенство нулю здесь неявным образом и не требуется, так как масса молекулы, размеры атома, заряд элементарной частицы, длина световой волны и т.п. — все это величины вполне конечные (хотя и весьма малые), обуславливающие точность тех или иных уравнений баланса и пр.

Например, не имеет смысла требовать, чтобы закон сохранения заряда выполнялся с погрешностью, меньшей чем величина элементарного электрического заряда (заряд электрона, равный $1.6 \cdot 10^{-19}$ Кл). Также бессмысленно требовать, чтобы погрешность изготовления деталей оптических систем была существенно меньшей длины световой волны (от $4 \cdot 10^{-7}$ м до $7.6 \cdot 10^{-7}$ м в зависимости от цвета). А что касается температуры, то при обычных земных условиях определение её с точностью превосходящей 10^{-3} K вообще проблематично в силу принципиальных соображений.

Наконец, ограниченная точность, с которой известны все физические константы[3], также воздвигает границы для требований равенства в физических соотношениях.

Во всех вышеприведенных примерах под решением уравнения понимается значение переменной, обращающее значение функции в пренебрежимо малую величину:

> Для заданных отображения $F : \mathbb{R}^n \to \mathbb{R}^n$ и $\varepsilon > 0$ найти значения неизвестной переменной x, такие что $F(x) \approx 0$ с абсолютной погрешностью ε, т.е. $\|F(x)\| < \varepsilon$.

Решением этой задачи является целое множество точек, которые мы будем называть *ε-решениями* или *почти решениями*, если порог этой пренебрежимой несуществен или не оговорен явно.

Нетрудно понять, что ε-решения устойчивы к малым возмущениям исходных данных, а задача “о нахождении почти решений” является вычислительно-корректной.

[3] В лучшем случае относительная погрешность известных на сегодняшний день значений физических констант равна 10^{-10}, см. [34].

Отметим, что в некоторых задачах система уравнений более естественно записывается не как (2), а в виде

$$G(x) = H(x), \tag{8}$$

и требуется обеспечить равенство левой и правой частей с относительной погрешностью ϵ:

Для заданных отображений $G,\, H : \mathbb{R}^n \to \mathbb{R}^n$ и $\epsilon > 0$ найти значения неизвестной переменной x, такие что $G(x) \approx H(x)$ с относительной погрешностью ϵ, т.е. $$\frac{\|G(x) - H(x)\|}{\max\{\|G(x)\|, \|H(x)\|\}} < \epsilon\ .$$

Решения этой задачи мы также будем называть *ϵ-решениями* системы уравнений вида (8).

Математические понятия, определения которых привлекают малый допуск ϵ, не являются новинкой. Таковы, к примеру, ε-энтропия, ε-субдифференциал, ε-оптимальные решения задач оптимизации и т.п. Наиболее близким, по-видимости, к нашим ϵ-решениям является концепция ϵ-спектра матрицы, которой интенсивно оперирует ряд западных авторов [35, 38], а также С.К. Годунов [7]. Говорят, что точка λ на комплексной плоскости принадлежит ϵ-спектру матрицы A, если существует комплексный вектор v единичной длины, такой что $\|(A - zI)v\| \le \epsilon$, где $\|\cdot\|$ — какая-то индуцированная матричная норма.[4]

3.4 Недостаточность ε-решений

Но есть и принципиально другой тип задач, которые образно могут быть названы задачами “об определении перехода через нуль” и не сводятся к нахождению ε-решений. Таковы задачи, в которых требуется гарантированно отследить переход функции к значениям противоположного знака (или, более общо, переход через некоторое критическое значение). При этом, в частности, в любой окрестности решения должны присутствовать как положительные значения функции, так и её отрицательные значения, тогда как в задачах нахождения почти решений это условие может и не выполняться.

Фазовый переход в физической (или химической системе) — типичная задача такого сорта, так как в процессе фазового перехода температура системы не меняется. Если мы хотим узнать, прошёл ли фазовый переход полностью, то нужно зафиксировать момент достижения противоположного множества состояний!

Другой пример. Рассмотрим систему линейных дифференциальных уравнений с постоянными коэффициентами

$$\frac{dx}{dt} = Ax, \tag{9}$$

[4]Определение ϵ-спектра, которое берет за основу С.К. Годунов в [7], формулируется неявным образом, так что его эквивалентность приведённому выше определению требует доказательства, а вычислительная корректность задачи определения ϵ-спектра из него напрямую не следует.

матрица которой зависит от параметра θ (возможно, векторного). Пусть при некотором начальном значении $\theta = \theta_0$ собственные значения матрицы A имеют отрицательные вещественные части, так что все решения системы (9) устойчивы по Ляпунову (и даже асимптотически устойчивы). При каких значениях параметра θ рассматриваемая система сделается неустойчивой?

Традиционный ответ на этот вопрос: "Срыв устойчивости в системе (9) произойдет при $\mathrm{Re}\ \lambda(A(\theta)) = 0$". Но он неправилен, так как для потери устойчивости необходимо не точное равенство нулю действительных частей некоторых собственных чисел матрицы, а переход их через нуль в область положительного знака. Без этого перехода через мнимую ось и "ещё чуть-чуть дальше" система останется устойчивой, сколь бы близко мы не придвинули собственные значения к мнимой оси или даже достигли бы её. Здесь важен именно переход "через и за" критическое значение, в отсутствие которого качественное изменение в поведении системы не совершится, и этот феномен совершенно не ухватывается понятиями ϵ-решения из §3.3 или ϵ-спектра из [7, 35, 38].

Рассмотренная ситуация, в действительности, весьма типична для динамических систем, где условием совершения многих типов бифуркаций является переход некоторого параметра через так называемое *бифуркационное значение*. К примеру, бифуркация Андронова-Хопфа (называемая также бифуркацией рождения цикла, см. [18]) происходит при переходе через мнимую ось пары комплексных собственных чисел матрицы линеаризованной системы. И здесь принципиален именно переход через некоторый порог, а не близость к нему, на которую делается упор в понятиях ϵ-решения и ϵ-спектра.

Нетрудно понять, что такое "переход через нуль" для непрерывной функции одного переменного $f : \mathbb{R} \to \mathbb{R}$. Но в многомерной ситуации мы сталкиваемся с методическими трудностями, возникающими из необходимости иметь для нестрогого понятия "прохождение функции через нуль" чисто математическое определение. Из требования вычислительной корректности следует, что в любой окрестности такого решения каждая из компонент $F_i(x)$ вектор-функции $F(x)$ должна принимать как положительные, так и отрицательные значения. Но как именно? Какими должны (или могут) быть значения компонент $F_j(x)$, $j \neq i$, если $F_i(x) > 0$ или $F_i(x) < 0$?

В разрешении этого затруднения нам на помощь приходят нелинейный анализ и алгебраическая топология.

4 Обзор теории вращения векторных полей

4.1 Векторные поля и их гомотопии

Если $M \subseteq \mathbb{R}^n$ и задано отображение

$$\Phi : M \to \mathbb{R}^n,$$

то часто удобно представлять значение $\Phi(x)$ как вектор, торчащий из точки x. При этом говорят, что на M задано *векторное поле* Φ. Любопытно, что это понятие было введено около 1830 года М. Фарадеем, затем соответствующий язык проник в математическую физику, теорию дифференциальных уравнений и теорию динамических систем (см. [3]), и в настоящее время не имеет тенденции к умиранию. Мы воспользуемся понятиями и результатами, наработанными в этой области знаний, для наших целей анализа решений систем уравнений и интервальных численных методов.

Векторное поле является *непрерывным*, если непрерывно $\Phi(x)$. Например, на Рис. 2 изображены поля

$$\Phi(x) = \Phi(\,x_1, x_2) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \text{и} \qquad \Psi(x) = \Psi(\,x_1, x_2) = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}, \tag{10}$$

которые непрерывны и даже дифференцируемы.

Определение. *Точки $x \in M$, в которых поле обращается в нуль, т.е. $\Phi(x) = 0$, будем называть нулями поля или же его особыми точками.*

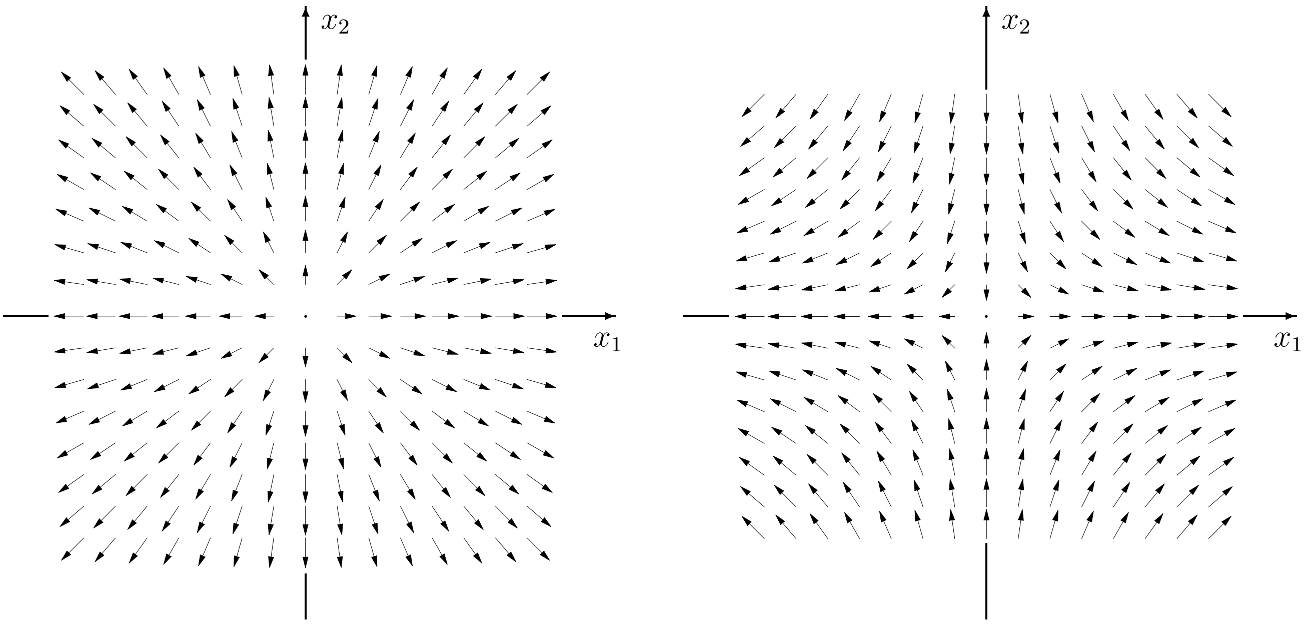

Рис. 2: Векторные поля $\Phi(x)$ и $\Psi(x)$, задаваемые формулами (10).

Связь векторных полей и их особых точек с основным предметом нашей статьи очевидна: особая точка поля Φ — это решение системы n уравнений

$$\begin{cases} \Phi_1(\,x_1, x_2, \cdots, x_n) &= \; 0, \\ \Phi_2(\,x_1, x_2, \cdots, x_n) &= \; 0, \\ \qquad\qquad \ddots & \qquad \vdots \\ \Phi_n(\,x_1, x_2, \cdots, x_n) &= \; 0. \end{cases}$$

Будем говорить, что векторное поле Φ *вырождено*, если у него есть особые точки. Иначе Φ называется *невырожденным*.

Определение. *Для множества $M \subseteq \mathbb{R}^n$ и $\lambda \in [0, 1]$ непрерывная функция*

$$\Delta(\lambda, x) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$$

называется деформацией векторного поля $\Delta(0, x)$ в векторное поле $\Delta(1, x)$.

Достаточно прозрачна связь деформаций с возмущениями векторного поля, т.е. отображения Φ

Определение. *Деформацию* $\Delta(\lambda, x)$ *назовём невырожденной, если* $\Delta(\lambda, x) \neq 0$ *для всех* $\lambda \in [0,1]$ *и* $x \in M$.

Невырожденные деформации могут соединять только невырожденные векторные поля. Примеры: растяжение, поворот, параллельный перенос.

Определение. *Если векторные поля можно соединить невырожденной деформацией, то они называются гомотопными.*

Отношение гомотопии векторных полей рефлексивно, симметрично и транзитивно, будучи поэтому *отношением эквивалентности*. Невырожденные на фиксированном множестве M непрерывные векторные поля распадается на классы гомотопных между собой полей.

4.2 Вращение векторных полей

Пусть D — ограниченная область в $\mathbb{R}^n$ с границей ∂D. Через $\operatorname{cl} D$ мы обозначим её топологическое замыкание. Оказывается, каждому невырожденному на ∂D векторному полю Φ можно сопоставить целочисленную характеристику — *вращение векторного поля* Φ на ∂D, — обозначаемую $\gamma(\Phi, D)$ и удовлетворяющую следующим условиям

A) Гомотопные на ∂D векторные поля имеют одинаковое вращение.

B) Пусть D_i, $i = 1, 2, \ldots$, — попарно непересекающиеся области, лежащие в D. Если непрерывное векторное поле Φ невырождено на множестве

$$\operatorname{cl} D \setminus \bigcup_{i=1}^{\infty} D_i,$$

то вращения $\gamma(\Phi, D_i)$ отличны от нуля лишь для конечного набора D_i и

$$\gamma(\Phi, D) = \gamma(\Phi, D_1) + \gamma(\Phi, D_2) + \ldots .$$

C) Если $\Phi(x) = x - a$ для некоторой точки $a \in D$, то вращение Φ на ∂D равно

$$\gamma(\Phi, D) = 1.$$

Ясно, что определенное так вращение на границе области устойчиво к малым шевелениям как области, так и векторного поля.

Условиями A)-B)-C) вращение векторного поля определяется однозначно, но можно показать [24], что это определение равносильно следующему конструктивному. Зафиксируем некоторую параметризацию поверхности ∂D, т.е. задание её в виде

$$\begin{aligned} x_1 &= x_1(u_1, u_2, \dots, u_{n-1}), \\ x_2 &= x_2(u_1, u_2, \dots, u_{n-1}), \\ &\vdots \\ x_n &= x_n(u_1, u_2, \dots, u_{n-1}), \end{aligned}$$

где $u_1, u_2, \dots, u_{n-1}$ — параметры, $x_i(u_1, u_2, \dots, u_{n-1})$, $i = 1, 2, \dots, n$ — функции, определяющие одноименные координаты. Тогда вращение поля $\Phi(x)$ на границе ∂D области D равно значению поверхностного интеграла

$$\frac{1}{n\sigma_n} \int_{\partial D} \frac{1}{\|\Phi(x)\|^n} \cdot \det \begin{pmatrix} \Phi_1(x) & \frac{\partial \Phi_1(x)}{\partial u_1} & \cdots & \frac{\partial \Phi_1(x)}{\partial u_{n-1}} \\ \Phi_2(x) & \frac{\partial \Phi_2(x)}{\partial u_1} & \cdots & \frac{\partial \Phi_2(x)}{\partial u_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_n(x) & \frac{\partial \Phi_n(x)}{\partial u_1} & \cdots & \frac{\partial \Phi_n(x)}{\partial u_{n-1}} \end{pmatrix} du_1 du_2 \dots du_n \ , \qquad (11)$$

где σ_n — площадь поверхности единичной сферы в $\mathbb{R}^n$. Интеграл (11) часто называют *интегралом Кронекера*.

В двумерном случае вращение векторного поля имеет простую геометрическую интерпретацию: это количество полных оборотов вектора поля, совершаемое при движении точки аргумента в положительном направлении по рассматриваемой границе области [3, 12, 13, 20]. В многомерном случае такой наглядности уже нет, но величина вращения векторного поля Φ всё равно может быть истолкована как "число раз, которое отображение $\Phi : \partial D \to \Phi(\partial D)$ накрывает образ $\Phi(\partial D)$".

Рассмотрим примеры. На любой окружности с центром в нуле изображенное на Рис. 2 поле имеет вращение $+1$. Векторные поля Рис. 3, которые задаются формулами

$$\begin{cases} x_1 = r \cos N\psi, \\ x_2 = r \sin N\psi, \end{cases}$$

где $r = \sqrt{x_1^2 + x_2^2}$ — длина радиус-вектора точки, ψ — его угол с положительным лучом оси абсцисс, при $N = 2$ и $N = 3$ имеют вращения $+2$ и $+3$ на окружностях с центром в нуле.

С вращением векторного поля тесно связана другая известная глобальная характеристика отображений — *топологическая степень* [12, 13, 19, 20, 21, 24]. Именно, вращение поля Φ на границе области D есть топологическая степень такого отображения ϕ границы ∂D в единичную сферу пространства $\mathbb{R}^n$, что

$$\phi(x) = \|\Phi(x)\|^{-1} \Phi(x).$$

Зачем нам понадобилось понятие вращения векторного поля? Мы собираемся использовать его для характеризации "прохождения через нуль" многомерной функции, и теоретической основой этого шага служат следующие результаты:

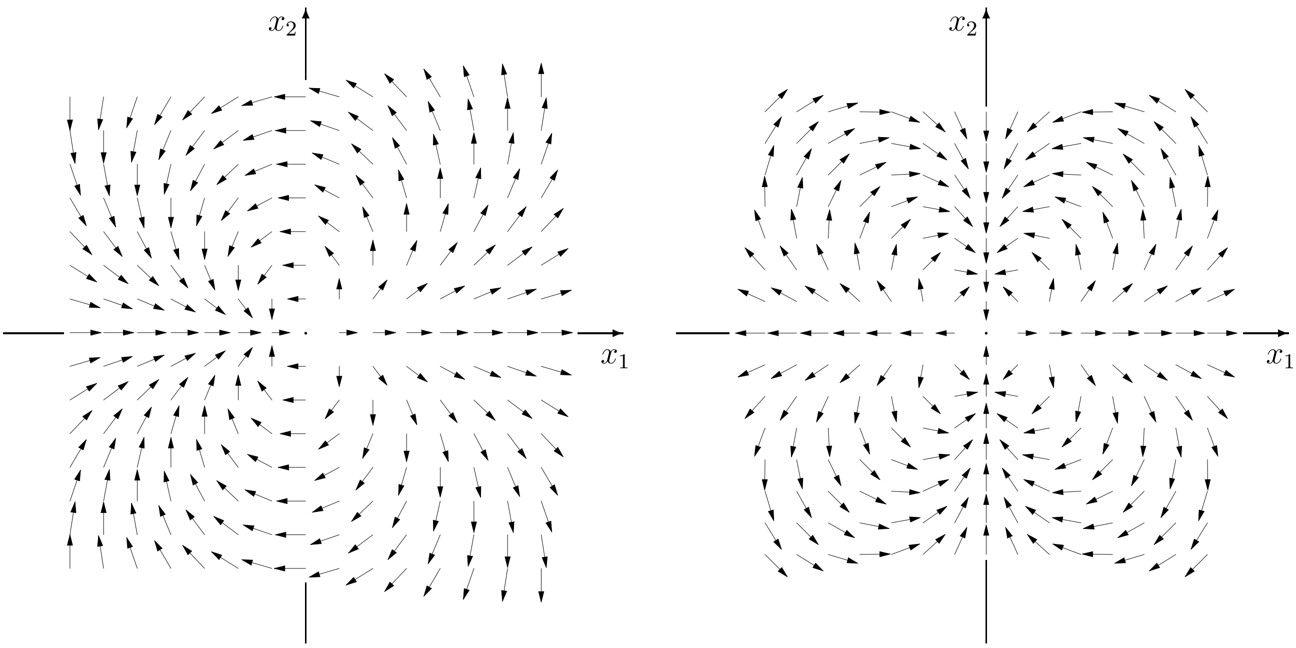

Рис. 3: Векторные поля, имеющие вращения +2 (левый чертёж) и +3 (правый чертёж) на любой окружности с центром в нуле.

Предложение [12, 13, 20]. *Если векторное поле* Φ *невырождено на замыкании ограниченной области* D*, то вращение* $\gamma(\Phi, D) = 0$.

Теорема Кронекера [12, 13, 20].

Пусть поле Φ *невырождено на границе ограниченной области* D *и непрерывно на её замыкании. Если* $\gamma(\Phi, D) \neq 0$*, то поле* Φ *имеет в* D *по крайней мере одну особую точку.*

Теорема Кронекера обладает очень большой общностью и часто применяется не напрямую, а служит основой для более конкретных достаточных условий существования нулей поля или решений систем уравнений. Например, доказательства приведенных в §2.2 теорем Лерэ-Шаудера, Брауэра и Миранды сводятся, фактически, к демонстрации того, что на границе области вращение векторного поля, соответствующего исследуемому отображению, равно ± 1.

4.3 Индексы особых точек

Станем говорить, что особая точка является *изолированной*, если в некоторой её окрестности нет других особых точек рассматриваемого векторного поля. Таким образом, вращение поля одинаково на сферах достаточно малых радиусов с центром в изолированной особой точке $\tilde{x}$.

Определение. *Это общее вращение называют индексом особой точки* $\check{x}$ *поля* Φ *или индексом нуля* $\check{x}$ *поля* Φ*, и обозначают* $\mathrm{ind}\,(\check{x}, \Phi)$.

Итак, оказывается, что особые точки векторных полей (и решения систем уравнений) могут быть существенно разными, отличаясь друг от друга своим индексом, и различных типов особых точек существует столько же, сколько и целых чисел, т.е. счётное множество. Какими являются наиболее часто встречающиеся особые точки и, соответственно, решения систем уравнений? Ответ на этот вопрос даётся следующими двумя результатами:

Предложение [12, 13, 20]. *Если A — невырожденное линейное преобразование пространства $\mathbb{R}^n$, то его единственная особая точка — нуль — имеет индекс* $\mathrm{ind}\,(0, A) = \mathrm{sgn}\ \det A$.

Определение. *Точка области определения отображения дифференцируемого отображения $F : \mathbb{R}^n \to \mathbb{R}^n$ называется критической, если в ней якобиан F' вырождается. Иначе говорят, что эта точка — регулярная.*

Предложение [12, 13, 20]. *Если $\tilde{x}$ — регулярная особая точка дифференцируемого векторного поля Φ, то* $\mathrm{ind}\,(\tilde{x}, \Phi) \;=\; \mathrm{sgn}\ \det \Phi'(\tilde{x})$.

Таким образом, регулярные (не критические) особые точки векторных полей имеют индекс ± 1, а в прочих случаях значение индекса может быть весьма произвольным.

Например, индексы расположенного в начале координат нуля векторных полей, которые изображены на Рис. 2, равны $+1$ и (-1), причём поля эти всюду дифференцируемы. Индексы нуля полей Рис. 3 равны $+2$ и $+3$, и в начале координат эти поля не дифференцируемы. Векторное поле на прямой, задаваемое рассмотренным в §3.2 квадратичным отображением $x \mapsto x^2 + px + q$ при $p^2 = 4q$ имеет особую точку $x = -p/2$ нулевого индекса.

Естественно спросить себя, насколько используемые нами при доказательном решении систем уравнений инструменты приспособлены для выявления особых точек различных типов. И здесь мы сталкиваемся с первым неприятным открытием: *интервальный метод Ньютона, методы Кравчика и Хансена-Сенгупты, тесты существования Мура и Куи, основывающиеся на теоремах Лерэ-Шаудера и Брауэра и наиболее часто используемые при практических доказательных вычислениях решений уравнений, охватывают только случаи индекса ± 1 особой точки F. Если же решение системы является критической точкой соответствующего отображения с индексом, не равным ± 1, то доказать его существование с помощью вышеупомянутых результатов принципиально не получится.* Это — одно из объяснений того, почему существующие практические интервальные алгоритмы для доказательного глобального решения систем уравнений не могут достичь “полного успеха” в общем случае.

4.4 Устойчивость особых точек

Определение. *Особая точка z поля Φ называется устойчивой, если для любого $\tau > 0$ можно найти такое $\eta > 0$, что всякое поле, отличающееся от Φ меньше чем на η, имеет особую точку, удаленную от z менее, чем на τ. Иначе особая точка z называется неустойчивой.*

Легко понять, что в связи с задачей решения систем уравнений нас интересуют именно устойчивые особые точки, поскольку задача поиска только таких точек является вычислительно-корректной.

Вторым основным результатом, ради которого мы затевали обзор теории вращения векторных полей, является следующее

Предложение [13]. *Изолированная особая точка непрерывного векторного поля устойчива тогда и только тогда, когда её индекс отличен от нуля.*

Например, неустойчивое решение квадратного уравнения (6)–(7) имеет индекс 0, а у векторных полей, изображённых на рисунках 2 и 3, начало координат является устойчивой особой точкой.

Отметим отдельно, что результат об устойчивости особой точки ненулевого индекса ничего не говорит о количестве особых точек, близких к возмущаемой особой точке. В действительности, путем шевеления одной устойчивой особой точки можно получить сразу *несколько* особых точек, и это легко видеть на примере полей Рис. 3. Любая сколь угодно малая постоянная добавка к полю, изображённому на левом чертеже Рис. 3, приводит к распадению нулевой особой точки индекса 2 на две особые точки индекса 1. Аналогично, любая сколь угодно малая постоянная добавка к полю, изображённому на правом чертеже Рис. 3, приводит к распадению нулевой особой точки на три особые точки индекса 1. Таким образом, свойство единственности решения неустойчиво и требовать его наличия нужно со специальными оговорками.

теорема Банаха о сжимающем отображении

Если в области D находится конечное число особых точек, то сумму их индексов называют *алгебраическим числом* особых точек.

Предложение. *Пусть непрерывное векторное поле Φ имеет в D конечное число особых точек x_1, x_2, ..., x_s и невырождено на границе ∂D. Тогда*

$$\gamma(\Phi, D) = \mathrm{ind}\,(x_1, \Phi) + \mathrm{ind}\,(x_2, \Phi) + \cdots + \mathrm{ind}\,(x_s, \Phi).$$

Алгебраическое число особых точек устойчиво к малым возмущениям области и поля, так как охватывает совокупную сумму индексов вне зависимости от рождения и уничтожения отдельных точек.

Теперь становится окончательно понятным и упоминавшийся в конце §2.3 феномен списка брусов-ответов с неопределённым статусом, которые никак не собираются исчезать в интервальных методах доказательного глобального решения уравнений ни при увеличении точности вычислений, ни при выделении дополнительного времени счета и т.п. Помимо причины, названой в предыдущем пункте, *эти списки могут соответствовать неустойчивым решениям системы уравнений, которые разрушаются при сколь угодно малых возмущениях и потому не могут быть идентифицированы никаким приближенным вычислительным алгоритмом с конечной точностью представления данных.*

К примеру, таковым является корень квадратного уравнения (6)–(7), и специалисты знают, что он чрезвычайно плохо находится численно, как традиционными, так и интервальными подходами.

Наконец, сделаем ещё одно важное замечание. Нередко на практике для решения систем нелинейных уравнений исходную задачу переформулируют как оптимизационную, пользуясь, например, тем, что справедливы следующие математические эквивалентности:

$$F(x) = 0 \qquad \Longleftrightarrow \qquad \min_x \ \|F(x)\| = 0$$

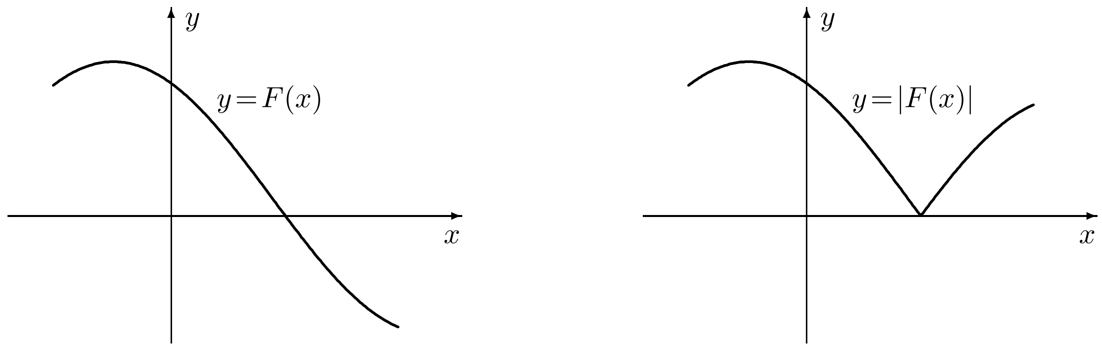

Рис. 4: Устойчивый нуль функции превращается
в неустойчивый после взятия нормы функции.

и

$$F(x) = 0 \qquad \Longleftrightarrow \qquad \min_x \|F(x)\|^2 = 0.$$

Далее имеющимися стандартными пакетами программ ищется решение задачи минимизации нормы $\|F(x)\|$ (или $\|F(x)\|^2$, чтобы обеспечить гладкость целевой функции) и результат сравнивается с нулём. С учётом наших знаний о задаче решения систем уравнений хорошо видна вычислительная неэквивалентность такого приведения: устойчивая особая точка *всегда* превращается при подобной трансформации в неустойчивое решение редуцированной задачи! Именно, любая сколь угодно малая добавка к $|F(x)|$ может приподнять график функции $y = |F(x)|$ над осью абсцисс (плоскостью нулевого уровня в общем случае), так что нуль функции исчезнет.

5 Модифицированная постановка задачи

Теперь все готово для вычислительно-корректной переформулировки задачи решения уравнений и систем уравнений. Она должна выглядеть следующим образом:

Для заданного $\varepsilon > 0$ и системы уравнений $F(x) = 0$ найти на данном множестве $D \subseteq \mathbb{R}^n$ 1) гарантированные двусторонние границы всех решений ненулевого индекса, 2) множество ε-решений.

(12)

Мы не требуем единственности решения в выдаваемых брусах, так как свойство решения быть единственным не является, вообще говоря, устойчивым к малым возмущениям задачи (см. §5.4).

Какими средствами мы располагаем для решения задачи в такой постановке?

Прежде всего отметим, что интервальные алгоритмы для доказательного глобального решения уравнений типа представленного в Таблице 1, вполне пригодны для решения модифицированной постановки задачи (12), если используемые в них "тесты существования" на брусах $\mathbf{Y}'$ и $\mathbf{Y}''$ усилить процедурой вычисления вращения поля $F(x)$ по границам $\partial\mathbf{Y}'$ и $\partial\mathbf{Y}''$. Эта задача действительно сложна и нетривиальна, являясь в самом общем случае NP-трудной (см. [37]), т.е. требующей для своего решения экспоненциальных вычислительных ресурсов [9]. Но она и не столь страшна, как может показаться вначале.

Если некоторая особая точка $\tilde{x}$ является регулярной (не критической), т.е. $\det F' \neq 0$, то, как отмечалось в §4.3, она надёжно находится с помощью интервального метода Ньютона, методов Кравчика, Хансена-Сенгупты и пр. Трудности представляет ситуация, когда индекс решения отличен от ± 1, возможная когда в рассматриваемой точке функция терпит разрыв гладкости, либо когда якобиан F' вырождается. В обоих этих случаях мы должны озаботиться более тщательным вычислением индекса решения, например, методами, изложенными в [24, 26, 29].

В целом результатом работы алгоритма доказательного глобального решения модифицированной постановки (12) является выдача пользователю нескольких списков, именно,

I) списка брусов с ненулевым вращением на границах,

II) списка почти решений,

а также

III) списка подозрительных на решения брусов,

IV) списка недообработанных брусов,

таких что все решения системы, не принадлежащие спискам I и II, содержатся в списках III и IV.

Реализация интервальных алгоритмов для решения модифицированной постановки задачи (12) ставит немало нетривиальных математических вопросов. Не претендуя здесь на полный анализ, дадим их краткий обзор.

Дробление брусов желательно производить так, чтобы решения уравнения не оказывались на границах подбрусов. В противном случае вычисление вращения векторного поля (топологической степени отображения) невозможно. Грубо говоря, чтобы убедиться в том, что решение является действительно решением, нужно рассмотреть его "со всех сторон".

Из создавшегося затруднения существует, по крайней мере, два выхода: 1) дробить брусы можно на такие потомки, которые немного налегают друг на друга, "перекрываются", а не просто касаются друг друга по грани меньшей размерности, 2) при необходимости дополнительного исследования подозрительных точек вокруг них можно искусственно раздуть некоторый брус малого размера, на границе которого функция гарантированно не занулялась бы.

Другая полезная операция — "склеивание" брусов решений. Цель — представить в удобном для пользователя виде протяжённые области решений одного статуса.

Отметим наконец, что наличие нескольких признаков решения уравнений может привести к их "конфликту" друг с другом. Величина ε из определения ε-решений может быть назначена такой, что

$$|\mathbf{F}(\mathbf{X})| \;<\; \varepsilon,$$

но

$$0 \notin \mathbf{F}(\mathbf{X}).$$

Должны ли мы считать при этом, что брус $\mathbf{X}$ представляет "почти решения" или нет? Однозначно хорошего способа разрешения вопроса нет. Традиционные рецепты:

1) Установление иерархии критериев решения.
2) Свертывание критериев решения в один "суперкритерий".

Но, как нам представляется, окончательное разрешение этого вопроса должно назначаться пользователем с учётом специфики решаемой задачи.

Весьма часто для анализа решений уравнений и систем уравнений привлекают понятие *кратности* решения (корня). Говорят, что решение $\tilde{x}$ уравнения (системы уравнений) $F(x) = 0$ имеет кратность k, если все производные (в общем случае частные) отображения F до k-ого порядка включительно зануляются в $\tilde{x}$ и производная порядка $(k+1)$ (по крайней мере одна в общем многомерном случае) не равна нулю в $\tilde{x}$. Иногда даже явление плохой определимости решений напрямую связывают с кратностью решений.

Корень квадратного уравнения в модельном примере (6)–(7) действительно кратный, но кратность сама по себе не служит причиной вычислительной некорректности, как показывает пример уравнения $x^3 = 0$. Единственное решение этого уравнения — нулевое, имеет кратность 2, устойчиво при малых шевелениях функции в левой части, и задача нахождения этого решения вычислительно корректна.

Но кратность решения действительно усугубляет вычислительные трудности при его нахождении, так как зануление нескольких производных подряд приводит к тому, что график отображения является очень пологим в окрестности решения, буквально "стелется" вдоль координатной плоскости нулевого уровня. При этом на ЭВМ с конечной и не очень высокой точностью представления данных решение не идентифицируется как изолированное или сосредоточенное на маленьком брусе, а "расплывается" в целое пятно[5]. Тем не менее, несмотря на большую неприятность этого явления, потенциально оно может быть преодолено путём повышения точности вычислений, увеличением длины разрядной сетки и т.п. в отличие от принципиальных ограничений на процесс решения, вызванных вычислительной некоррктностью задачи и различием индексов.

6 Вычислительная корректность и алгоритмическая разрешимость

Предлагаемый нами в §6 путь весьма непрост и получающиеся алгоритмы для решения систем уравнений, явно вычисляющие вращения векторных полей, могут быть очень трудоёмкими. Возникает естественный вопрос: нельзя ли обойтись чем-нибудь попроще?

В этом параграфе мы взглянем на проблему с другой стороны, именно, с точки зрения алгоритмической разрешимости исходной задачи решения уравнений и покажем, что альтернативы нашим "сложным" подходам весьма ограничены.

Известно, что теория алгоритмов и рекурсивных функций — это математическая дисциплина, исследующая конструктивные свойства различных математических объектов. Её основные понятия — это *алгоритм*, *конструктивный объект*, *вычислимость*, *разрешимость* и др.

[5]Такие решения А.Л. Семёнов и его сотрудники в ИСИ СО РАН называют "ползущими".

Алгоритм — это конечная последовательность инструкций, записанных на некотором языке и определяющих процесс переработки исходных данных в искомые результаты (ответ решаемой задачи и т.п.). Алгоритм принципиально конечен и определяет собой конечный процесс. Далее, *конструктивным объектом* называется объект, который может быть построен с помощью некоторой конечной последовательности действий над каким-то конечным алфавитом. Таковы, например, рациональные числа. Строго говоря, конструктивные объекты и только они могут быть получены в качестве ответов при решении задачи на реальных цифровых ЭВМ с конечными быстродействием и объёмом памяти.

В частности, конечными машинами являются широко распространенные ныне электронные цифровые вычислительные машины: они способны представлять, по сути дела, только конечные множества чисел. Таким образом, обречены на неудачу любые попытки использовать их для выполнения арифметических абсолютно точных операций над числовыми полями $\mathbb{R}$ и $\mathbb{C}$, которые являются бесконечными (и даже непрерывными) множествами, большинство элементов которых не представимы в цифровых ЭВМ.

Оказывается, что значительная часть объектов, с которыми работают современная математика и её приложения, не являются конструктивными. В частности, неконструктивным является традиционное понятие вещественного числа, подразумевающее бесконечную процедуру определения всех знаков его десятичного разложения (которое в общем случае непериодично). Факт неконструктивности вещественных чисел может быть обоснован строго математически (см. [15]), и он указывает на принципиальные границы возможностей алгоритмического подхода и ЭВМ в деле решения задач математического анализа.

Тем не менее, и в этом океане неконструктивности имеет смысл выделить объекты, которые могут быть “достаточно хорошо” приближены конструктивными объектами. На этом пути мы приходим к понятию *вычислимого вещественного числа* [15, 23][6]: вещественное число α называется вычислимым, если существует алгоритм, дающий по всякому натуральному числу n рациональное приближение к α с погрешностью $\frac{1}{n}$. Множество всех вычислимых вещественных чисел образует *вычислимый континуум*. Соответственно, *вычислимая вещественная функция* определяется как отображение из вычислимого континуума в вычислимый континуум, задаваемая алгоритмом преобразования программы аргумента в программу значений.

Важно помнить, что и вычислимое вещественное число, и вычислимая функция — это уже не конструктивные объекты. Но, как выясняется, даже ценой ослабления наших требований к конструктивности нельзя вполне преодолеть принципиальные алгоритмические трудности, связанные с задачей решения уравнений. Для вычислимых вещественных чисел и функций ряд традиционных постановок задач оказывается *алгоритмически неразрешимыми* в том смысле, что построение общих алгоритмов их решения принципиально невозможно.

Например, алгоритмически неразрешимыми являются задачи

[6]Совершенно аналогичными являются определения *рекурсивного вещественного числа* у Р.П. Гудстейна [8] и *конструктивного вещественного числа* у Б.А. Кушнера [14].

1) распознавания для произвольного вычислимого вещественного числа равно оно нулю или нет [14, 15, 16], распознавания равенства двух вычислимых вещественных чисел [8, 14, 15, 23, 24];

2) нахождения для каждой совместной системы линейных уравнений над полем конструктивных вещественных чисел какого-либо ее решения [14, 16];

3) нахождения нулей всякой непрерывной кусочно-линейной знакопеременной функции [16].

Приведённые выше результаты задают, как нам представляется, ту абсолютную и совершенно объективную мерку (в отличие от субъективных пристрастий), с которой мы должны подходить к оценке трудоёмкости тех или иных вычислительных методов. Получается, что необходимость переформулировки задачи решения уравнений и систем уравнений связана ещё и с тем, что в традиционной постановке эти задачи оказываются алгоритмически неразрешимыми! На фоне этого мрачного факта наличие даже экспоненциально трудного алгоритма с небольшим основанием "одноэтажной" экспоненты в оценке сложности (вроде 2^n) можно рассматривать как вполне приемлемый вариант разрешимости задачи, и именно это имеет место в ситуации с вычислением вращения векторного поля (степени отображения).

Заметим, что развитые в этом параграфе соображения не являются совершенно новыми, они высказывались в близкой форме О. Абертом в [24]. Но предпринятый нами в §§2–5 теоретический анализ ситуации опирался на другую основу — топологические свойства отображений, их нулей и соображения устойчивости. Он позволил даже глубже проникнуть в причину специфических явлений, возникающих при работе практических интервальных алгоритмов.

7 Итоги

Автор сочтёт свою задачу выполненной, если у читателя отложатся следующие выводы:

1. Задача (3) доказательного глобального решения уравнений и систем уравнений очень трудна.

2. В последние десятилетия разработаны и успешно применяются на практике интервальные методы для доказательного глобального решения уравнений и систем уравнений.

3. Задача нахождения решений уравнений и систем уравнений не является вычислительно-корректной.

   Причинами неудовлетворительного поведения интервальных методов для доказательного глобального решения уравнений являются

   1) неустойчивый характер решений этих уравнений и систем уравнений,

   2) непригодность тех средств, которые используются в интервальных алгоритмах: решение может иметь такой тип (индекс), на который в принципе не расчитаны существующие интервальные методы Ньютона, Кравчика, Хансена-Сенгупты и пр.

4. Привлечение понятия ε-решения и теории вращения векторных полей (или эквивалентной теории топологической степени отображений) позволяют дать вычислительно-корректную переформулировку (12) задачи решения уравнений и систем уравнений.

5. В настоящее время нарождаются интервальные методы для доказательного глобального решения систем уравнений с учётом новой вычислительно-корректной переформулировки (12) этой задачи. При этом, чтобы обеспечить доказательность процесса решения, привлечение понятий и методов интервального анализа оказывается необходимым по существу.

Список литературы

[1] А. Акритас, *Основы компьютерной алгебры с приложениями.* – М.: Мир, 1994.

[2] Г. Алефельд, Ю. Херцбергер, *Введение в интервальные вычисления.* – Москва: Мир, 1987.

[3] В.И. Арнольд, *Обыкновенные дифференциальные уравнения.* – Москва: Наука, 1984.

[4] Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельков, *Численные методы.* – Москва: Наука, 1987.

[5] М. Берже, *Геометрия. Т. 1, 2.* – Москва: Наука, 1984.

[6] В.В. Воеводин, Ю.А. Кузнецов, *Матрицы и вычисления.* – Москва: Наука, 1984.

[7] С.К. Годунов, *Современные аспекты линейной алгебры.* – Новосибирск: Научная книга, 1997.

[8] Р.П. Гудстейн, *Рекурсивный математический анализ.* – Москва: Наука, 1970.

[9] М. Гэри, Д. Джонсон, *Вычислительные машины и труднорешаемые задачи.* – Москва: Мир, 1982.

[10] В.А. Зорич, *Математический анализ. Том 1.* – Москва: Наука, 1981.

[11] С.А. Калмыков, Ю.И. Шокин, З.Х. Юлдашев, *Методы интервального анализа.* – Новосибирск: Наука, 1986.

[12] М.А. Красносельский, П.П. Забрейко, *Геометрические методы нелинейного анализа.* – Москва: Наука, 1975.

[13] М.А. Красносельский, А.И. Перов, А.И. Поволоцкий, П.П. Забрейко, *Векторные поля на плоскости.* – Москва: Физматлит, 1963.

[14] Б.А. Кушнер, *Лекции по конструктивному математическому анализу.* – Москва: Наука, 1973.

[15] П. Мартин-Лёф, *Очерки по конструктивной математике.* – М.: Наука, 1975.

[16] Математическая Энциклопедия, том. 2. – Москва: Наука, 1979.

[17] А.С. Нариньяни, Недоопределённость в системах представления и обработки знаний // *Известия АН СССР. Техническая Кибернетика.* – 1986. – №5. – С. 3–28.

[18] М. Холодниок, А. Клич, М. Кубичек, М. Марек, *Методы анализа нелинейных динамических моделей.* – Москва: Мир, 1991.

[19] Л. Ниренберг, *Лекции по нелинейному функциональному анализу.* – Москва: Мир, 1977.

[20] В.И. Опойцев, *Нелинейная системостатика.* – Москва: Наука, 1986.

[21] Дж. Ортега, В. Рейнболдт, *Итерационные методы решения нелинейных систем уравнений со многими неизвестными.* – Москва: Мир, 1975.

[22] В.А. Треногин, *Функциональный анализ.* – Москва: Наука, 1980.

[23] В.А. Успенский, А.Л. Семёнов, *Теория алгоритмов: основные открытия и приложения.* – Москва: Наука, 1987.

[24] O. Aberth, *Precise numerical methods using C++.* – San Diego: Academic Press, 1998.

[25] A.B. Babichev, O.B. Kadyrova, T.P. Kashevarova, A.S. Leshchenko, A.L. Semenov, UniCalc, a novel approach to solving systems of algebraic equations // *Interval Computations.* – 1993. – №2. – P. 29–47.

[26] J. Dian and R. B. Kearfott, Existence verification for singular and nonsmooth zeros of real nonlinear systems // *Mathematics of Computation.* – 2003. – Vol. 72 (242). – P. 757–766.

[27] E. Hansen, *Global optimization using interval analysis.* – New York: Marcel Dekker, 1992.

[28] N. Karcanias, Ch. Giannakopoulos, M. Hubbard, Almost zeros of a set of polynomials // *Intern. Journal of Control.* – 1983. – Vol. 38, №6. – P. 1213–1238.

[29] R.B. Kearfott, An efficient degree-computation method for a generalized method of bisection // *Numerische Mathematik.* – 1979. – Vol. 32. – P. 109–127.

[30] R.B. Kearfott, *Rigorous global search: Continuous problems.* – Dordrecht: Kluwer, 1996.

[31] R.E. Moore, A test for existence of solutions to nonlinear systems // *SIAM Journal on Numerical Analysis.* – 1977. – Vol. 14. – P. 611–615.

[32] R.E. Moore, *Methods and applications of interval analysis.* – Philadelphia: SIAM, 1979.

[33] A. Neumaier, *Interval methods for systems of equations.* – Cambridge: Cambridge University Press, 1990.

[34] `http://physics.nist.gov/cuu/Constants/index.html`

[35] `http://web.comlab.ox.ac.uk/projects/pseudospectra/`

[36] L. Qi, A note on the Moore test for nonlinear systems // *SIAM Journal on Numerical Analysis.* – 1982. – Vol. 19. – P. 851–857.

[37] K.A. Sikorski, *Optimal Solution of Nonlinear Equations.* – London: Oxford University Press, 2000.

[38] L.N. Trefethen, Approximation theory and numerical linear algebra // J.C. Mason and M.G. Cox, eds., *Algorithms for Approximation II*, Shrivenham, 1988. – Chapman and Hall, London, 1990. – P. 336–360.