ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2024, том 64, № 12, с. 2332–2354

ОБЫКНОВЕННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

УДК 519.62

О МИНИМАЛЬНОСТИ КВАДРАТИЧНОЙ ПОГРЕШНОСТИ РЕШЕНИЯ ПРЕОБРАЗОВАННЫХ К НАИЛУЧШЕМУ ПАРАМЕТРУ СИСТЕМ УРАВНЕНИЙ ПРИ МАЛЫХ ОДНОРОДНЫХ ВОЗМУЩЕНИЯХ

© 2024 г. Е.Б. Кузнецов[1,*], С.С. Леонов[1,2,]**

[1] 125993 Москва, Волоколамское ш., 4, Московский авиационный институт (национальный исследовательский университет), Россия
[2] 117198 Москва, ул. Миклухо-Маклая, 6, Российский университет дружбы народов имени Патриса Лумумбы, Россия
**e-mail: kuznetsov@mai.ru*
***e-mail: powerandglory@yandex.ru*

Поступила в редакцию 05.05.2024 г.
Переработанный вариант 05.08.2024 г.
Принята к публикации 23.08.2024 г.

В статье рассматривается решение систем нелинейных уравнений с одним скалярным параметром. Множеством решений подобных систем является кривая в пространстве неизвестных системы уравнений и параметра. Ее построение проводится, как правило, при помощи численных методов и сопряжено с многочисленными трудностями, возникающими вследствие наличия на кривой множества решений предельных и существенно особых точек. Для нахождения таких кривых используется метод продолжения решения по параметру и наилучшей параметризации, позволяющий свести решение к начальной задаче для системы дифференциальных уравнений продолжения решения. В данной работе исследуется устойчивость решения системы продолжения решения на вносимые в нее возмущения. Впервые полностью доказано сформулированное ранее утверждение о минимальности квадратичной ошибки решения системы продолжения решения при однородных малых возмущениях ее матрицы. Теоретические результаты проиллюстрированы на примере численного построения лемнискаты Бернулли. Библ. 10. Фиг. 2. Табл. 1.

Ключевые слова: системы нелинейных уравнений, продолжение решения по параметру, наилучший параметр, система продолжения решения, малые возмущения, квадратичная погрешность.

DOI: 10.31857/S0044466924120087, **EDN**: KBTDTO

1. ВВЕДЕНИЕ

При описании задач механики деформируемого твердого тела, механики жидкости и газа, биологии, химии и экономики важную роль играют системы алгебраических и трансцендентных уравнений, которые в общем случае называются нелинейными. Но получить решение подобных уравнений и их систем весьма сложно. Начиная с XVI века разрабатываются общие методы, как аналитического, так и приближенного решения нелинейных уравнений и их систем. Поскольку далеко не каждое нелинейное уравнение или система могут быть решены аналитически, на первый план выходят приближенные методы. При этом и приближенные методы решения имеют существенные недостатки. Наиболее известны такие методы решения нелинейных уравнений и систем, как метод простой итерации и метод Ньютона. Оба этих метода обладают своими достоинствами и недостатками. Метод Ньютона по сравнению с методом простой итерации имеет квадратичную скорость сходимости, однако, в использовании требует выполнения значительно больших условий: подбор начального приближения, вычисление первой производной функции уравнения (матрицы Якоби для систем уравнений), значительные ограничения на функцию уравнения или системы, необходимость в анализе сходимости метода. Наличие подобных недостатков и значительная скорость сходимости метода Ньютона привели к появлению его модификаций и обобщений.

В 30-е годы прошлого века бельгийским математиком М. Лаэем был предложен метод решения нелинейных уравнений с параметром на основе метода Ньютона [1]. Метод Лаэя заключается в разбиении интервала изменения параметра уравнения и построении для каждой точки разбиения итерационной последовательности по

методу Ньютона, принимая в качестве начального приближения последнюю точку, полученную для предыдущего значения параметра. Этот метод нацелен на устранение сразу двух недостатков метода Ньютона: выбор начальной точки и обеспечение сходимости итерационной последовательности к корню уравнения с квадратичной скоростью. По всей видимости метод Лаэя является одним из первых примеров применения параметризации в вычислительной математике. Позднее М. Лаэй обобщил предложенный им метод на случай систем нелинейных уравнений [2].

Другую идею в своих работах [3, 4] использовал советский математик Д. Ф. Давиденко. Он рассматривал параметр системы нелинейных уравнений как аргумент, а кривые множества решений отождествлял с интегральными кривыми задачи Коши для системы дифференциальных уравнений, полученной из исходной системы нелинейных уравнений путем ее дифференцирования по параметру (полагая, что неизвестные системы являются функциями параметра). Выбирая начальное условие при начальном значении параметра можно свести решение системы нелинейных уравнений к решению соответствующей задачи Коши для системы дифференциальных уравнений.

Существенным недостатком, как метода Лаэя, так и метода Давиденко, является необходимость смены параметра продолжения решения, если кривая множества решений содержит предельные или существенно особые точки (например, в случае замкнутых или самопересекающихся кривых множества решений). Сам факт необходимости смены параметра говорит о неединственности его способа выбора, что ставит вопрос о наличии оптимального в некотором плане параметра продолжения решения.

Впервые гипотеза о том, что параметр, отсчитываемый по касательной к кривой множества решений рассматриваемой нелинейной системы, является в некотором смысле наилучшим высказал акад. И. И. Ворович [5]. Идея доказательства этой гипотезы принадлежит Э. Риксу [6], однако полное доказательство оптимальности параметра, отсчитываемого по касательной к кривой множества решений рассматриваемой нелинейной системы, было дано лишь в работе В. И. Шалашилина и Е. Б. Кузнецова [7]. В работе Э. Рикса под оптимальностью параметра продолжения решения понимается наилучшая обусловленность линеаризованной системы продолжения решения, но это предположение не доказывается. В работах В. И. Шалашилина и Е. Б. Кузнецова помимо наилучшей обусловленности доказывается еще и минимальность квадратичной погрешности решения системы продолжения решения, возникающей при возмущении матрицы системы и вектора правой части. Это дополнение является существенным, так как позволяет утверждать, что при использовании наилучшего параметра система продолжения решения не только обладает наилучшей обусловленностью, но и минимизируется влияние на решение этой системы возмущений различного рода: ошибок округления, вычислительных погрешностей, неустранимых погрешностей. Для ряда задач последнее может быть важнее, чем обусловленность.

К сожалению, доказательство минимальности квадратичной погрешности в работе [7] дано не полностью – строго проверены лишь необходимые условия локального условного экстремума. В данной статье авторы ставят целью устранить этот недостаток, строго доказав достаточные условия локального условного минимума квадратичной погрешности решения системы продолжения решения, возникающей при возмущении матрицы системы малыми однородными возмущениями, в случае использования наилучшего параметра.

2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим систему n нелинейных уравнений относительно $n+1$ неизвестного

$$F_i\left(x_1,\dots,x_n,x_{n+1}\right)=0,\quad i=1,\dots,n. \tag{1}$$

Здесь и далее будем полагать, что функции F_i определены и непрерывны на всей рассматриваемой области изменения неизвестных $x_1,\dots,x_n,x_{n+1}$. Такими системами описываются многие задачи механики деформируемого твердого тела, физики, биологии, экономики. Тот факт, что количество неизвестных больше количества уравнений говорит о неединственности решения системы (1).

При определенных условиях можно построить множество решений системы (1). Примем одну из переменных системы (1) как параметр, для каждого значения которого (из некоторой области определения) система (1) является замкнутой. Без ограничения общности, в качестве параметра можно принять неизвестную x_{n+1}. Пусть известно какое-либо решение системы (1) для некоторого начального значения параметра $x_{(n+1)0}$, т. е. известна точка $M_0\left(x_{10},\dots,x_{n0},x_{(n+1)0}\right)$, удовлетворяющая всем уравнениям системы

$$F_i\left(x_{10},\dots,x_{n0},x_{(n+1)0}\right)=0,\quad i=1,\dots,n.$$

Если матрица Якоби относительно переменных $x_1,\dots,x_n$ невырожденна в окрестности точки M_0, то по теореме о неявной функции можно получить кривую множества решений в окрестности точки M_0, переходя по

параметру от значения параметра $x_{(n+1)0}$ к значению $x_{(n+1)1} = x_{(n+1)0} + h$, где h – заданная достаточно малая величина. При использовании метода Ньютона этот процесс определяется следующей итерационной процедурой:

$$\mathbf{x}_1^{(j)} = \mathbf{x}_1^{(j-1)} - J^{-1}\left(\mathbf{x}_1^{(j-1)}, x_{(n+1)1}\right) \cdot \mathbf{F}\left(\mathbf{x}_1^{(j-1)}, x_{(n+1)1}\right), \tag{2}$$

где $\mathbf{x} = (x_1, \ldots, x_n)^\top$, $\mathbf{x}_1^{(0)} = (x_{10}, \ldots, x_{n0})^\top$ – начальное приближение, $\mathbf{F} = (F_1, \ldots, F_n)^\top$, номер итерации задает индекс $j = 1, 2, \ldots$, а матрица Якоби относительно переменных $x_1, \ldots, x_n$ имеет вид

$$J(\mathbf{x}, x_{n+1}) = \left[\frac{\partial F_j}{\partial x_k}\right]_{j,k=1}^{n}.$$

Итерационный процесс (2) повторяется до момента выполнения условия $\left\|\mathbf{x}_1^{(j)} - \mathbf{x}_1^{(j-1)}\right\|_2 < \varepsilon$, где $\|\cdot\|_2$ – квадратичная векторная норма, которая для вектора $\mathbf{a} = (a_1, \ldots, a_n)^\top$ вычисляется по формуле $\|\mathbf{a}\|_2 = \sqrt{a_1^2 + \cdots + a_n^2}$, а ε задает величину требуемой погрешности. В результате будет получена точка $M_1\left(x_{11}, \ldots, x_{n1}, x_{(n+1)1}\right)$, также являющаяся решением системы (1). Продолжая этот процесс, можно продвигаться далее по кривой множества решений.

Описанный выше метод построения кривой множества решений системы (1) имеет значительный недостаток – он применим только в случае, когда матрица Якоби не вырождается. Если матрица Якоби вырождается в некоторой точке $M_k\left(x_{1k}, \ldots, x_{nk}, x_{(n+1)k}\right)$, то построение кривой множества решений из нее с использованием итерационной процедуры (2) невозможно. Используемый параметр x_{n+1} становится непригодным для продолжения решения и требует смены. Процедура смены параметра продолжения решения довольно трудоемка и слабо формализуема. Она состоит в следующем. На смену параметра x_{n+1} из переменных системы (1) выбирается новый параметр продолжения решения x_i, для которого в точке M_k будет невырожденной матрица Якоби для переменных $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, x_{n+1}$

$$\tilde{J}(\tilde{\mathbf{x}}, x_i) = \left[\frac{\partial F_j}{\partial x_k}\right]_{\substack{j,k=1 \\ j\neq n+1 \\ k\neq i}}^{n+1},$$

где $\tilde{\mathbf{x}} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, x_{n+1})^\top$. При наличии такого параметра x_i, итерационный процесс (2) преобразуется к виду

$$\tilde{\mathbf{x}}_{k+1}^{(j)} = \tilde{\mathbf{x}}_{k+1}^{(j-1)} - \tilde{J}^{-1}\left(\tilde{\mathbf{x}}_{k+1}^{(j-1)}, x_{i(k+1)}\right) \cdot \mathbf{F}\left(\tilde{\mathbf{x}}_{k+1}^{(j-1)}, x_{i(k+1)}\right),$$

где $\tilde{\mathbf{x}}_{k+1}^{(0)} = \left(x_{1k}, \ldots, x_{(i-1)k}, x_{(i+1)k}, \ldots, x_{nk}, x_{(n+1)k}\right)$, $x_{i(k+1)} = x_{ik} + h$, и процесс построения кривой множества решений продолжается. Описанная идея смены параметра продолжения решения впервые в численном анализе была отмечена в работах М. Лаэя.

Иную идею использовал в своих работах Д. Ф. Давиденко. Он полагал, что при выборе параметра, например, неизвестной x_{n+1}, остальные неизвестные $x_1, \ldots, x_n$ непрерывно зависят от него. Тогда дифференцируя уравнения системы (1) по переменной x_{n+1}, получим систему дифференциальных уравнений вида

$$\frac{\partial F_i}{\partial x_1} \cdot \frac{dx_1}{dx_{n+1}} + \cdots + \frac{\partial F_i}{\partial x_n} \cdot \frac{dx_n}{dx_{n+1}} + \frac{\partial F_i}{\partial x_{n+1}} = 0, \quad i = 1, \ldots, n, \tag{3}$$

или в векторно-матричной форме:

$$J\left(\mathbf{x}, x_{n+1}\right) \cdot \frac{d\mathbf{x}}{dx_{n+1}} + \frac{\partial \mathbf{F}}{\partial x_{n+1}} = 0. \tag{4}$$

Разрешая систему (4) относительно производной $\dfrac{d\mathbf{x}}{dx_{n+1}}$, перейдем к нормальной форме Коши

$$\frac{d\mathbf{x}}{dx_{n+1}} = -J^{-1}\left(\mathbf{x}, x_{n+1}\right) \cdot \frac{\partial \mathbf{F}}{\partial x_{n+1}}. \tag{5}$$

Дополняя систему уравнений (5) начальным условием

$$\mathbf{x}(x_{(n+1)0}) = \mathbf{x}_0, \quad \mathbf{x}_0 = (x_{10}, \ldots, x_{n0})^\top, \tag{6}$$

получим начальную задачу (5), (6).

В сравнении с методом Лаэя, метод Давиденко более привлекательный в вычислительном плане, поскольку не требует итерационного уточнения решения для каждого значения параметра. Однако метод Давиденко обладает тем же недостатком, что и метод Лаэя. Если матрица Якоби, входящая в правую часть системы (5), вырождается в некоторой точке $M_k\left(x_{1k},\ldots,x_{nk},x_{(n+1)k}\right)$, то дальнейшее движение вдоль кривой множества решений исходной системы (1) по параметру x_{n+1} невозможно. Для продолжения движения по кривой множества решений также необходима смена параметра продолжения. Эта идея впервые была применена Д. Ф. Давиденко. Для нового параметра продолжения решения x_i потребуем выполнения того же условия, что и в методе Лаэя: невырожденность матрицы Якоби $\tilde{J}(\tilde{\mathbf{x}},x_i)$. Преобразованная к новому параметру x_i задача (5), (6) примет вид системы

$$\frac{d\tilde{\mathbf{x}}}{dx_i}=-\tilde{J}^{-1}\left(\tilde{\mathbf{x}},x_i\right)\cdot\frac{\partial\mathbf{F}}{\partial x_i} \tag{7}$$

с начальным условием

$$\tilde{\mathbf{x}}(x_{ik})=\tilde{\mathbf{x}}_k,\quad \tilde{\mathbf{x}}_k=\left(x_{1k},\ldots,x_{(i-1)k},x_{(i+1)k}\ldots,x_{nk},x_{(n+1)k}\right)^{\top}. \tag{8}$$

Решая задачу (7), (8), мы продолжаем движение по кривой множества решений исходной системы (1) до момента, когда не возникнет новой точки, в которой вырождается определитель матрицы Якоби.

Метод Давиденко хоть и не преодолевает недостатков метода Лаэя, но дает идею построения кривой множества решений в случае, когда ни одна из неизвестных системы (1) не подходит на роль нового параметра. Эта идея состоит в использовании параметров продолжения иного вида.

Метод продолжения решения по параметру предполагает использование параметров продолжения более общего вида. Как правило они задаются локально, в каждой точке кривой множества решений. Определим параметр продолжения μ для системы (1) в форме

$$d\mu=\alpha_1\cdot dx_1+\cdots+\alpha_n\cdot dx_n+\alpha_{n+1}\cdot dx_{n+1}, \tag{9}$$

где $dx_1,\ldots,dx_{n+1}$ – дифференциалы неизвестных системы (1), $\alpha_1,\ldots,\alpha_{n+1}$ – заданные числовые коэффициенты. Стоит отметить, что при переходе от точки к точке множества решений системы (1), можно изменять параметр μ, т. е. придавать коэффициентам $\alpha_1,\ldots,\alpha_{n+1}$ другие значения, поэтому глобально $\alpha_1,\ldots,\alpha_{n+1}$ являются функциями неизвестных системы (1). Геометрически коэффициенты $\alpha_1,\ldots,\alpha_{n+1}$ можно интерпретировать как компоненты направляющего вектора $\bar{\alpha}$, вдоль которого отсчитывается параметр μ.

Полагая, что неизвестные системы (1) зависят от параметра μ, т. е.

$$x_1=x_1(\mu),\ldots,x_n=x_n(\mu),x_{n+1}=x_{n+1}(\mu),$$

можно продифференцировать уравнения системы (1) по параметру μ, получив систему дифференциальных уравнений

$$\frac{\partial F_i}{\partial x_1}\cdot\frac{dx_1}{d\mu}+\cdots+\frac{\partial F_i}{\partial x_n}\cdot\frac{dx_n}{d\mu}+\frac{\partial F_i}{\partial x_{n+1}}\cdot\frac{dx_{n+1}}{d\mu}=0,\quad i=1,\ldots,n. \tag{10}$$

Дополняя полученную систему уравнением, полученным из (9) делением на $d\mu$:

$$\alpha_1\cdot\frac{dx_1}{d\mu}+\cdots+\alpha_n\cdot\frac{dx_n}{d\mu}+\alpha_{n+1}\cdot\frac{dx_{n+1}}{d\mu}=1, \tag{11}$$

получим замкнутую систему обыкновенных уравнений с $n+1$ неизвестной $\dfrac{dx_1}{d\mu},\ldots,\dfrac{dx_{n+1}}{d\mu}$. Для удобства вычислений, наложим на коэффициенты $\alpha_1,\ldots,\alpha_{n+1}$ условие нормировки:

$$\alpha_1^2+\cdots+\alpha_{n+1}^2=1. \tag{12}$$

Последнее равенство отражает тот факт, что все направления отсчета параметра продолжения являются равноправными.

Можно видеть, что система (3) схожа по структуре с системой (10), поскольку использует ту же идею зависимости неизвестных системы (1) от параметра продолжения.

Запишем систему (10), (11) в векторно-матричном виде

$$\begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{n+1} \\ \dfrac{\partial F_1}{\partial x_1} & \dfrac{\partial F_1}{\partial x_2} & \cdots & \dfrac{\partial F_1}{\partial x_{n+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} & \dfrac{\partial F_n}{\partial x_2} & \cdots & \dfrac{\partial F_n}{\partial x_{n+1}} \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dx_1}{d\mu} \\ \dfrac{dx_2}{d\mu} \\ \vdots \\ \dfrac{dx_{n+1}}{d\mu} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{13}$$

Систему (13) будем называть системой *продолжения решения*.

Сам процесс решения системы (13) не отличается от используемого в методе Давиденко. Система (13) разрешается относительно производных неизвестных системы (1) и решается при начальных условиях

$$x_1(\mu_0) = x_{10}, \ldots, x_{n+1}(\mu_0) = x_{(n+1)0}.$$

Заметим, что система (13) не зависит явно от параметра μ, значит можно положить $\mu_0 = 0$.

Разрешение системы (13) и решение полученной начальной задачи проводится аналитически только в исключительных случаях. Эффективность же численного решения зависит существенно от обусловленности матрицы системы (13). Если выбрать параметр продолжения решения (т. е. коэффициенты $\alpha_1, \ldots, \alpha_{n+1}$) неудачно, то матрица системы (13) будет вырождаться, что приведет к необходимости смены параметра. Как видно из приводимых выше выкладок, эта процедура нежелательна. Стоит выбирать такой параметр продолжения решения, который не требуют смены в процессе решения. Очевидно, что подобный параметр будет не единственным, что ставит задачу выбора оптимального параметра продолжения решения. В работе [7] было доказано, что параметр продолжения решения, отсчитываемый по касательной к кривой множества решений, т. е. тот параметр, для которого коэффициенты

$$\alpha_1 = \frac{dx_1}{d\mu}, \ldots, \alpha_{n+1} = \frac{dx_{n+1}}{d\mu}, \tag{14}$$

будет доставлять матрице системы (13) наилучшую обусловленность. Такой параметр продолжения получил название наилучшего и обозначение λ.

Помимо наилучшей обусловленности параметр λ обладает тем свойством, что при его использовании становится минимальной квадратичная погрешность решения, возникающая при возмущении элементов матрицы системы (13). Это утверждение также было сформулировано в работе [7], однако полного его доказательства дано не было. Доказательство было ограничено лишь проверкой необходимых условий локального минимума, достаточные условия были лишь намечены. Целью данной работы является устранение этого недостатка.

3. ОСНОВНЫЕ ЛЕММЫ

Предварительно докажем две леммы.

Лемма 1. *Для любого натурального значения n, определитель n-го порядка*

$$\begin{vmatrix} a_1^2 + a_0 & -a_1 \cdot a_2 & a_1 \cdot a_3 & \cdots & (-1)^{n+1} a_1 \cdot a_n \\ -a_2 \cdot a_1 & a_2^2 + a_0 & -a_2 \cdot a_3 & \cdots & (-1)^{n+2} a_2 \cdot a_n \\ a_3 \cdot a_1 & -a_3 \cdot a_2 & a_3^2 + a_0 & \cdots & (-1)^{n+3} a_3 \cdot a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1} a_n \cdot a_1 & (-1)^{n+2} a_n \cdot a_2 & (-1)^{n+3} a_n \cdot a_3 & \cdots & a_n^2 + a_0 \end{vmatrix}_n = \tag{15}$$

$$= a_0^{n-1} \cdot \left(a_0 + \sum_{i=1}^{n} a_i^2 \right),$$

где $a_0, a_1, \ldots, a_n$ – заданные вещественные числа.

Доказательство леммы 1. Для доказательства данной леммы используем метод математической индукции. Сформируем базу индукции.

Для значения $n = 1$ определитель вычисляется тривиально:

$$\left| a_1^2 + a_0 \right| = a_1^2 + a_0.$$

Для значения $n = 2$:

$$\begin{vmatrix} a_1^2 + a_0 & -a_1 \cdot a_2 \\ -a_2 \cdot a_1 & a_2^2 + a_0 \end{vmatrix} = a_0 \cdot \left(a_1^2 + a_2^2 + a_0\right).$$

Для значения $n = 3$, применяя разложение по последней строке, получим

$$\begin{vmatrix} a_1^2 + a_0 & -a_1 \cdot a_2 & a_1 \cdot a_3 \\ -a_2 \cdot a_1 & a_2^2 + a_0 & -a_2 \cdot a_3 \\ a_3 \cdot a_1 & -a_3 \cdot a_2 & a_3^2 + a_0 \end{vmatrix} = a_0^2 \cdot \left(a_1^2 + a_2^2 + a_3^2 + a_0\right).$$

Выполненные вычисления позволяют сформулировать гипотезу: для произвольного натурального n будет справедлива формула (15).

Проверим гипотезу, вычислив определитель размерности $n + 1$, разлагая его по последней строке. После применения формулы (15) и упрощения, получим

$$D_{n+1} = \begin{vmatrix} a_1^2 + a_0 & -a_1 \cdot a_2 & a_1 \cdot a_3 & \cdots & (-1)^{n+2} a_1 \cdot a_{n+1} \\ -a_2 \cdot a_1 & a_2^2 + a_0 & -a_2 \cdot a_3 & \cdots & (-1)^{n+3} a_2 \cdot a_{n+1} \\ a_3 \cdot a_1 & -a_3 \cdot a_2 & a_3^2 + a_0 & \cdots & (-1)^{n+4} a_3 \cdot a_{n+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+2} a_{n+1} \cdot a_1 & (-1)^{n+3} a_{n+1} \cdot a_2 & (-1)^{n+4} a_{n+1} \cdot a_3 & \cdots & a_{n+1}^2 + a_0 \end{vmatrix}_{n+1} =$$

$$= \left(a_{n+1}^2 + a_0\right) \cdot a_0^{n-1} \cdot \left(a_0 + \sum_{i=1}^{n} a_i^2\right) +$$

$$+ a_{n+1}^2 \cdot a_1 \cdot \begin{vmatrix} -a_1 \cdot a_2 & a_1 \cdot a_3 & \cdots & (-1)^{n+1} a_1 \cdot a_n & (-1)^{n+2} a_1 \\ a_2^2 + a_0 & -a_2 \cdot a_3 & \cdots & (-1)^{n+2} a_2 \cdot a_n & (-1)^{n+3} a_2 \\ -a_3 \cdot a_2 & a_3^2 + a_0 & \cdots & (-1)^{n+3} a_3 \cdot a_n & (-1)^{n+4} a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (-1)^{n+2} a_n \cdot a_2 & (-1)^{n+3} a_n \cdot a_3 & \cdots & a_n^2 + a_0 & (-1)^{2n+1} a_n \end{vmatrix}_{n} + \cdots$$

$$\cdots + a_{n+1}^2 \cdot a_n \cdot \begin{vmatrix} a_1^2 + a_0 & -a_1 \cdot a_2 & \cdots & (-1)^{n} a_1 \cdot a_{n-1} & (-1)^{n+2} a_1 \\ -a_2 \cdot a_1 & a_2^2 + a_0 & \cdots & (-1)^{n+1} a_2 \cdot a_{n-1} & (-1)^{n+3} a_2 \\ a_3 \cdot a_1 & -a_3 \cdot a_2 & \cdots & (-1)^{n+2} a_3 \cdot a_{n-1} & (-1)^{n+4} a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (-1)^{n+1} a_n \cdot a_1 & (-1)^{n+2} a_n \cdot a_2 & \cdots & (-1)^{2n-1} a_n \cdot a_{n-1} & (-1)^{2n+1} a_n \end{vmatrix}_{n}.$$

Остается лишь вычислить полученные n определителей порядка n. Все они вычисляются одинаково. Продемонстрируем процедуру их вычисления на примере последнего определителя, обозначив его через D_n:

$$D_n = a_1 \cdot \begin{vmatrix} a_1 & -a_1 \cdot a_2 & \cdots & (-1)^{n} a_1 \cdot a_{n-1} & (-1)^{n+2} a_1 \\ -a_2 & a_2^2 + a_0 & \cdots & (-1)^{n+1} a_2 \cdot a_{n-1} & (-1)^{n+3} a_2 \\ a_3 & -a_3 \cdot a_2 & \cdots & (-1)^{n+2} a_3 \cdot a_{n-1} & (-1)^{n+4} a_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (-1)^{n+1} a_n & (-1)^{n+2} a_n \cdot a_2 & \cdots & (-1)^{2n-1} a_n \cdot a_{n-1} & (-1)^{2n+1} a_n \end{vmatrix}_{n} +$$

$$+ a_0 \cdot \begin{vmatrix} a_2^2 + a_0 & -a_2 \cdot a_3 & \cdots & (-1)^{n+1} a_2 \cdot a_{n-1} & (-1)^{n+3} a_2 \\ -a_3 \cdot a_2 & a_3^2 + a_0 & \cdots & (-1)^{n+2} a_3 \cdot a_{n-1} & (-1)^{n+4} a_3 \\ a_4 \cdot a_2 & -a_4 \cdot a_3 & \cdots & (-1)^{n+3} a_4 \cdot a_{n-1} & (-1)^{n+5} a_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (-1)^{n+2} a_n \cdot a_2 & (-1)^{n+3} a_n \cdot a_3 & \cdots & (-1)^{2n-1} a_n \cdot a_{n-1} & (-1)^{2n+1} a_n \end{vmatrix}_{n-1}.$$

Рассмотрим определитель, стоящий в первом слагаемом. Если из его последнего столбца вынести множитель $(-1)^n$, то у полученного определителя будут совпадать первый и последний столбцы, следовательно, он будет равняться нулю. Полученный в итоге определитель по структуре аналогичен исходному, поэтому проводя проделанные операции еще $n - 3$ раз, мы придем к виду

$$D_n = a_0^{n-2} \cdot \begin{vmatrix} a_0 & (-1)^{2n} a_{n-1} \\ 0 & (-1)^{2n+1} a_n \end{vmatrix} = -a_0^{n-1} a_n.$$

Таким же образом, можно вычислить и $n-1$ оставшихся определителей, которые все будут иметь отрицательный знак, как это видно при вычислении определителя для произвольного значения n. Таким образом, исходный определитель $n+1$ порядка может быть записан в виде

$$D_{n+1} = \left(a_{n+1}^2 + a_0\right) \cdot a_0^{n-1} \cdot \left(a_0 + \sum_{i=1}^{n} a_i^2\right) - a_0^{n-1} a_{n+1}^2 \sum_{i=1}^{n} a_i^2 = a_0^n \cdot \left(a_0 + \sum_{i=1}^{n+1} a_i^2\right).$$

По методу математической индукции формула (15) справедлива для всех натуральных значений n.

Доказательство леммы 1 завершено.

Используя лемму 1 можно доказать, что справедлива

Лемма 2. *Функция*

$$J(\alpha_1, \ldots, \alpha_{n+1}) = \frac{A}{\Delta^m} \tag{16}$$

достигает своего минимального значения при ограничении типа равенств (12) *в точке*

$$\alpha_k^* = \frac{(-1)^{k+1} \cdot \Delta_k}{\Delta}, \quad k = 1, \ldots, n+1, \tag{17}$$

если $A > 0$ и m принимает четные значения. Если к тому же $\Delta > 0$, то значения (17) *доставляют минимум функции* (16) *при $A > 0$ и любом натуральном m.*

В функции (16) *знаменатель задается формулой*

$$\Delta = \sum_{i=1}^{n+1} (-1)^{i+1} \alpha_i \Delta_i, \tag{18}$$

постоянные величины $A, \Delta_1, \ldots, \Delta_{n+1}$ не зависят от переменных $\alpha_1, \ldots, \alpha_{n+1}$. В выражениях (17) *значение знаменателя $\Delta = \pm\left(\Delta_1^2 + \cdots + \Delta_{n+1}^2\right)^{1/2}$.*

Доказательство леммы 2. Перепишем задачу минимизации функции (16) при ограничении типа равенств (12), прибегая к используемым в тензорной алгебре обозначениям:

$$\frac{A}{\Delta^m} \to \min_{\overline{\alpha} \in \mathbb{R}^{n+1}}, \quad \alpha_i \cdot \alpha_i = 1, \quad i = 1, \ldots, n+1, \tag{19}$$

где $\overline{\alpha} = (\alpha_1, \ldots, \alpha_{n+1})^\top$ – вектор переменных задачи. Здесь и далее будем предполагать суммирование по повторяющимся индексам в диапазоне их изменения, если не оговорено обратное.

Для решения задачи минимизации с ограничением типа равенств (19) воспользуемся методом множителей Лагранжа. Для этого составим функцию Лагранжа

$$L(\alpha_1, \ldots, \alpha_{n+1}, \gamma) = \frac{A}{\Delta^m} + \gamma \cdot (\alpha_i \cdot \alpha_i - 1), \quad i = 1, \ldots, n+1, \tag{20}$$

где γ – множитель Лагранжа.

Необходимое условие. Для проверки необходимых условий условного экстремума вычислим первые производные функции Лагранжа (20):

$$\frac{\partial L}{\partial \alpha_k} = -m \cdot \frac{A}{\Delta^{m+1}} \cdot (-1)^{k+1} \cdot \Delta_k + 2\gamma \cdot \alpha_k, \quad k = 1, \ldots, n+1. \tag{21}$$

Приравнивая производные (21) к нулю, получим систему из $n+1$ уравнения относительно $\alpha_1, \ldots, \alpha_{n+1}$ и γ:

$$-m \cdot \frac{A}{\Delta^{m+1}} \cdot (-1)^{k+1} \cdot \Delta_k + 2 \cdot \gamma \cdot \alpha_k = 0, \quad k = 1, \ldots, n+1. \tag{22}$$

Выразим α_k из системы (22):

$$\alpha_k = \frac{m \cdot A}{2 \cdot \gamma \cdot \Delta^{m+1}} \cdot (-1)^{k+1} \cdot \Delta_k, \quad k = 1, \ldots, n+1. \tag{23}$$

Пусть

$$\omega = \frac{m \cdot A}{2 \cdot \Delta^m}.$$

Тогда выражение (23) для α_k перепишется в виде

$$\alpha_k = \frac{\omega}{\gamma} \cdot \frac{(-1)^{k+1} \cdot \Delta_k}{\Delta}, \quad k = 1, \ldots, n+1. \tag{24}$$

Используем ограничение задачи (12):

$$\alpha_i \cdot \alpha_i = \frac{\omega^2}{\gamma^2} \cdot \frac{\Delta_i \cdot \Delta_i}{\Delta^2} = 1, \quad i = 1, \ldots, n+1. \tag{25}$$

Из равенства (25) получим

$$\gamma = \pm\omega \cdot \frac{(\Delta_i \cdot \Delta_i)^{1/2}}{\Delta}, \quad i = 1, \ldots, n+1. \tag{26}$$

Подставляя множитель Лагранжа (26) в соотношение (24), найдем

$$\alpha_k = \pm\frac{(-1)^{k+1} \cdot \Delta_k}{(\Delta_i \cdot \Delta_i)^{1/2}}, \quad i, k = 1, \ldots, n+1. \tag{27}$$

Упростим выражения (27). Подставим эти значения компонент α_k в формулу для определителя Δ (18):

$$\Delta = (-1)^{i+1} \cdot \alpha_i \cdot \Delta_i = \pm\frac{\Delta_i \cdot \Delta_i}{(\Delta_i \cdot \Delta_i)^{1/2}} = \pm(\Delta_i \cdot \Delta_i)^{1/2}, \quad i = 1, \ldots, n+1.$$

Подставляя полученное равенство в выражение для компонент α_k, найдем условно стационарную точку (17).

Достаточное условие. Как известно, характер экстремума задачи (19) связан со знакоопределенностью второго дифференциала функции Лагранжа (20)

$$d^2L = \sum_{i,j=1}^{n+1} \frac{\partial^2 L}{\partial\alpha_i \partial\alpha_j} d\alpha_i d\alpha_j. \tag{28}$$

В свою очередь знакоопределенность второго дифференциала функции Лагранжа зависит от знакоопределенности матрицы вторых производных функции Лагранжа. Вычислим эти производные, используя выражения для первых производных (21):

$$\frac{\partial^2 L}{\partial\alpha_k^2} = m \cdot (m+1) \cdot \frac{A}{\Delta^{m+2}} \cdot \Delta_k^2 + 2\gamma,$$
$$\frac{\partial^2 L}{\partial\alpha_k \partial\alpha_l} = m \cdot (m+1) \cdot \frac{A}{\Delta^{m+2}} \cdot (-1)^{k+l} \cdot \Delta_k \cdot \Delta_l, \quad i, k, l = 1, \ldots, n+1.$$

Учитывая, что, согласно достаточному условию условного экстремума второго рода, второй дифференциал функции Лагранжа (28) должен быть знакоопределенным в окрестности найденной условно стационарной точки, перепишем вторые производные в виде

$$\left.\frac{\partial^2 L}{\partial\alpha_k^2}\right|_{\alpha_k = \alpha_k^*} = \frac{A_1}{\Delta^{m+2}} \cdot \Delta_k^2 + 2\gamma^*,$$
$$\left.\frac{\partial^2 L}{\partial\alpha_k \partial\alpha_l}\right|_{\alpha_k = \alpha_k^*} = \frac{A_1}{\Delta^{m+2}} \cdot (-1)^{k+l} \cdot \Delta_k \cdot \Delta_l, \quad i, k, l = 1, \ldots, n+1,$$

где $\gamma^* = \omega = \dfrac{m \cdot A}{2 \cdot \Delta^m}$, $\Delta = \pm(\Delta_i \cdot \Delta_i)^{1/2}$, $A_1 = m \cdot (m+1) \cdot A$.

Таким образом, в окрестности полученной условно стационарной точки (17), матрица вторых производных функции Лагранжа (20) примет вид

$$\left[\frac{\partial^2 L}{\partial\alpha_i\partial\alpha_j}\right]_{i,j=1}^{n+1} =$$

$$= \begin{pmatrix} \frac{A_1}{\Delta^{m+2}}\cdot\Delta_1^2+2\gamma^* & -\frac{A_1}{\Delta^{m+2}}\cdot\Delta_1\cdot\Delta_2 & \cdots & \frac{(-1)^{n+2}\cdot A_1}{\Delta^{m+2}}\cdot\Delta_1\cdot\Delta_{n+1} \\ -\frac{A_1}{\Delta^{m+2}}\cdot\Delta_1\cdot\Delta_2 & \frac{A_1}{\Delta^{m+2}}\cdot\Delta_2^2+2\gamma^* & \cdots & \frac{(-1)^{n+3}\cdot A_1}{\Delta^{m+2}}\cdot\Delta_2\cdot\Delta_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(-1)^{n+2}\cdot A_1}{\Delta^{m+2}}\cdot\Delta_1\cdot\Delta_{n+1} & \frac{(-1)^{n+3}\cdot A_1}{\Delta^{m+2}}\cdot\Delta_2\cdot\Delta_{n+1} & \cdots & \frac{A_1}{\Delta^{m+2}}\cdot\Delta_{n+1}^2+2\gamma^* \end{pmatrix}. \tag{29}$$

Для анализа знакоопределенности матрицы (29) используем критерий Сильвестра. Рассмотрим M_k – угловой минор k-го порядка матрицы (29), равный

$$M_k = \left(\frac{A_1}{\Delta^{m+2}}\right)^k \begin{vmatrix} \Delta_1^2+\frac{\Delta^2}{m+1} & -\Delta_1\cdot\Delta_2 & \cdots & (-1)^{k+1}\cdot\Delta_1\cdot\Delta_k \\ -\Delta_1\cdot\Delta_2 & \Delta_2^2+\frac{\Delta^2}{m+1} & \cdots & (-1)^{k+2}\cdot\Delta_2\cdot\Delta_k \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{k+1}\cdot\Delta_1\cdot\Delta_k & (-1)^{k+2}\cdot\Delta_2\cdot\Delta_k & \cdots & \Delta_k^2+\frac{\Delta^2}{m+1} \end{vmatrix}_k.$$

Полученный определитель по структуре полностью удовлетворяет лемме 1 при $a_0 = \dfrac{\Delta^2}{m+1}$, $a_1 = \Delta_1, \ldots, a_k = \Delta_k$. Поскольку лемма 1 справедлива для любого натурального значения n, то все угловые миноры матрицы вторых производных (29) можно вычислить по формуле

$$M_k = \frac{m+1}{\Delta^2}\cdot\left(\frac{m\cdot A}{\Delta^m}\right)^k\cdot\left(\frac{\Delta^2}{m+1}+\sum_{i=1}^{k}\Delta_i^2\right), \quad k=1,\ldots,n+1. \tag{30}$$

Знак миноров (30) определяется множителем $\left(\dfrac{m\cdot A}{\Delta^m}\right)^k$. Этот множитель положителен, если $A>0$ и m принимает четные значения. Если же дополнительно $\Delta>0$, то $\left(\dfrac{m\cdot A}{\Delta^m}\right)^k>0$ при $A>0$ и любых натуральных m.

При указанных условиях все угловые миноры матрицы вторых производных функции Лагранжа (29) строго положительны и согласно критерию Сильвестра она будет положительно определена, а значит, и $d^2L>0$. Тогда, согласно достаточным условиям условного экстремума с ограничением типа равенств, найденная условно стационарная точка (17) является точкой условного минимума, доставляя функции (16) наименьшее значение при условии (12).

Доказательство леммы 2 завершено.

4. МИНИМАЛЬНОСТЬ КВАДРАТИЧНОЙ ПОГРЕШНОСТИ СИСТЕМЫ ПРОДОЛЖЕНИЯ РЕШЕНИЯ ПРИ МАЛЫХ ОДНОРОДНЫХ ВОЗМУЩЕНИЯХ ЕЕ МАТРИЦЫ

Структура системы продолжения решения (13) позволяет явно выписать формулы для компонент ее решения. Выпишем эти формулы, прежде чем перейти к возмущенному случаю.

Здесь и далее будем обозначать $x_1,\ldots,x_{n+1}$ невозмущенные неизвестные исходной системы (1), а производными $\dfrac{dx_1}{d\mu},\ldots,\dfrac{dx_{n+1}}{d\mu}$ будем обозначать компоненты решения невозмущенной системы продолжения решения (13).

Следуя монографии [8], используем метод Крамера для нахождения решения системы продолжения решения (13):

$$\frac{dx_i}{d\mu} = (-1)^{i+1}\frac{\Delta_i}{\Delta}, \quad i=1,\ldots,n+1. \tag{31}$$

Здесь Δ – определитель матрицы системы (13), который может быть вычислен по формуле (18), в которой $\Delta_1, \ldots, \Delta_{n+1}$ являются минорами, получаемыми из матрицы системы (13) вычеркиванием первой строки и столбца, номер которого соответствует индексу, т. е.

$$\Delta_i = \begin{vmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_{i-1}} & \dfrac{\partial F_1}{\partial x_{i+1}} & \cdots & \dfrac{\partial F_1}{\partial x_n} & \dfrac{\partial F_1}{\partial x_{n+1}} \\ \dfrac{\partial F_2}{\partial x_1} & \cdots & \dfrac{\partial F_2}{\partial x_{i-1}} & \dfrac{\partial F_2}{\partial x_{i+1}} & \cdots & \dfrac{\partial F_2}{\partial x_n} & \dfrac{\partial F_2}{\partial x_{n+1}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} & \cdots & \dfrac{\partial F_n}{\partial x_{i-1}} & \dfrac{\partial F_n}{\partial x_{i+1}} & \cdots & \dfrac{\partial F_n}{\partial x_n} & \dfrac{\partial F_n}{\partial x_{n+1}} \end{vmatrix}. \tag{32}$$

Будем полагать далее, что $\Delta_1^2 + \cdots + \Delta_{n+1}^2 \neq 0$, т. е. множество решений системы (1) не содержит существенно особых точек.

Обозначим далее $y_1, \ldots, y_{n+1}$ возмущенные неизвестные исходной системы (1), а производными $\dfrac{dy_1}{d\mu}, \ldots, \dfrac{dy_{n+1}}{d\mu}$ будем обозначать компоненты решения системы продолжения решения (13) после наложения на ее матрицу возмущений.

В данной статье под возмущениями некоторой строки (или столбца) матрицы системы продолжения решения (13) вида

$$(a_1, a_2, \ldots, a_{n+1})$$

будет пониматься строка (или столбец) вида

$$(\varepsilon_1 \cdot a_1, \varepsilon_2 \cdot a_2, \ldots, \varepsilon_{n+1} \cdot a_{n+1}),$$

которая прибавляется к исходной строке (или столбцу) при заданных значениях $\varepsilon_1, \ldots, \varepsilon_{n+1}$. Будем говорить, что возмущения однородные, если выполняется равенство

$$\varepsilon_1 = \cdots = \varepsilon_{n+1} = \varepsilon.$$

Рассмотрим малые возмущения, т. е. будем полагать, что слагаемыми, содержащими квадраты и более высокие степени $\varepsilon_1, \ldots, \varepsilon_{n+1}$, можно пренебречь по сравнению со слагаемыми, содержащими меньшие степени $\varepsilon_1, \ldots, \varepsilon_{n+1}$.

Естественным ограничением на значения $\varepsilon_1, \ldots, \varepsilon_{n+1}$ является их неравенство нулю. В противном случае, исходная и возмущенная системы продолжения решения будут совпадать.

В прикладных задачах наличие возмущений в системе продолжения решения (13) неизбежно. Поэтому важно иметь возможность прогнозирования их влияния на полученное решение. Возможны случаи, когда возмущения системы (13) полностью искажают искомое решение. Проанализируем влияние возмущений на решение системы (13). Введем погрешность компоненты решения возмущенной системы (13) в форме разности

$$\delta_i = \frac{dy_i}{d\mu} - \frac{dx_i}{d\mu}, \quad i = 1, \ldots, n+1. \tag{33}$$

В качестве меры погрешности будем использовать квадратичную погрешность

$$\delta = \sum_{i=1}^{n+1} \delta_i^2 = \sum_{i=1}^{n+1} \left(\frac{dy_i}{d\mu} - \frac{dx_i}{d\mu} \right)^2. \tag{34}$$

Разумеется квадратичная погрешность будет зависеть от выбора коэффициентов $\alpha_1, \ldots, \alpha_{n+1}$. В статье [7] сформулирована теорема о минимальности квадратичной погрешности решения возмущенной системы продолжения решения при выборе в окрестности каждой точки кривой множества решений в качестве значений $\alpha_1, \ldots, \alpha_{n+1}$ компонент вектора, касательного к кривой множества решений системы (1).

Полного доказательства этой теоремы в работе [7] не дано. Приведем его для случая однородных возмущений.

Прежде, чем переходить к полному доказательству, рассмотрим три частных случая: возмущение первой строки матрицы системы (13), возмущение любой строки матрицы системы (13), отличной от первой, и возмущение любого столбца матрицы системы (13). Это приводит к формулировке и доказательству трех теорем.

Придадим первой строчке матрицы системы (13) возмущения, прибавив к ней строку $\varepsilon\,(\alpha_1,\ldots,\alpha_{n+1})$, где ε – заданная малая величина, квадратом которой можно пренебречь по сравнению с меньшими степенями. Тогда возмущенная система продолжения решения запишется в виде

$$\begin{pmatrix} \alpha_1+\varepsilon\cdot\alpha_1 & \alpha_2+\varepsilon\cdot\alpha_2 & \cdots & \alpha_{n+1}+\varepsilon\cdot\alpha_{n+1} \\ \dfrac{\partial F_1}{\partial x_1} & \dfrac{\partial F_1}{\partial x_2} & \cdots & \dfrac{\partial F_1}{\partial x_{n+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_n} & \dfrac{\partial F_n}{\partial x_2} & \cdots & \dfrac{\partial F_n}{\partial x_{n+1}} \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} \\ \dfrac{dy_2}{d\mu} \\ \vdots \\ \dfrac{dy_{n+1}}{d\mu} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{35}$$

Для возмущенной системы продолжения решения (35) справедлива

Теорема 1. *Квадратичная погрешность решения возмущенной системы продолжения решения* (35)*, полученной из системы продолжения решения* (13) *наложением малых однородных возмущений на первую строку ее матрицы системы, достигает наименьшего значения в том случае, когда направляющий вектор* $\overline{\alpha}=(\alpha_1,\ldots,\alpha_{n+1})^{\top}$ *направлен по касательной к кривой множества решений системы* (1) *в рассматриваемой точке, т. е. компоненты вектора* $\overline{\alpha}$ *задаются в виде* (14).

Доказательство теоремы 1. По аналогии с формулами (31), (32), можно дать решение и для возмущенной системы (35):

$$\frac{dy_i}{d\mu}=(-1)^{i+1}\frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon},\quad i=1,\ldots,n+1. \tag{36}$$

Определитель Δ_ε может быть непосредственно вычислен вынесением общего множителя из первой строки:

$$\Delta_\varepsilon=(1+\varepsilon)\cdot\Delta, \tag{37}$$

а поскольку возмущению подвержена только первая строка матрицы системы (13), то справедливы равенства

$$\Delta_{i\varepsilon}=\Delta_i,\quad i=1,\ldots,n+1. \tag{38}$$

Подставляя равенства (37), (38) в формулы (36), получим

$$\frac{dy_i}{d\mu}=(-1)^{i+1}\frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon}=(-1)^{i+1}\frac{\Delta_i}{(1+\varepsilon)\cdot\Delta},\quad i=1,\ldots,n+1. \tag{39}$$

Воспользуемся теперь малостью возмущений. Домножим числитель и знаменатель правой части формулы (39) на $(1-\varepsilon)$ и отбросим слагаемое, включающее ε^2, перейдя к виду

$$\frac{dy_i}{d\mu}=(-1)^{i+1}\frac{(1-\varepsilon)\cdot\Delta_i}{\Delta},\quad i=1,\ldots,n+1. \tag{40}$$

Используя формулы (31) и (40), вычислим погрешности компонент решения возмущенной системы (33):

$$\delta_i=\frac{dy_i}{d\mu}-\frac{dx_i}{d\mu}=-(-1)^{i+1}\frac{\varepsilon\cdot\Delta_i}{\Delta},\quad i=1,\ldots,n+1. \tag{41}$$

Тогда, согласно формуле (34), квадратичная погрешность будет иметь вид

$$\delta=\sum_{i=1}^{n+1}\delta_i^2=\frac{\varepsilon^2}{\Delta^2}\sum_{i=1}^{n+1}\Delta_i^2. \tag{42}$$

Поставим задачу минимизации квадратичной ошибки (42), учитывая ограничение (12), накладываемое на выбор направления отсчета параметра продолжения:

$$\frac{\varepsilon^2\cdot\Delta_i\cdot\Delta_i}{\Delta^2}\to\min_{\overline{\alpha}\in\mathbb{R}^{n+1}},\quad \alpha_i\cdot\alpha_i=1,\quad i=1,\ldots,n+1. \tag{43}$$

Полученная задача (43) поиска минимума при ограничении типа равенств удовлетворяет условиям леммы 2 при $A=\varepsilon^2\cdot\Delta_i\cdot\Delta_i>0$, $i=1,\ldots,n+1$, и $m=2$. Это означает, что эта задача имеет точку минимума

с компонентами (17), совпадающими с компонентами касательного вектора к кривой множества решений в рассматриваемой точке:

$$\alpha_k = \alpha_k^* = \frac{dx_k}{d\mu}, \quad k = 1, \ldots, n+1. \tag{44}$$

В этой точке квадратичная погрешность (42) принимает наименьшее значение.

Доказательство теоремы 1 завершено.

Замечание 1. В данной работе внесение возмущений в матрицу системы продолжения решения (13) происходит путем добавления к некоторой строке (или столбцу) матрицы системы вида

$$(a_1, a_2, \ldots, a_{n+1})$$

возмущающей строки (или столбца) вида

$$(\varepsilon_1 \cdot a_1, \varepsilon_2 \cdot a_2, \ldots, \varepsilon_{n+1} \cdot a_{n+1}),$$

т. е. возмущение пропорционально самим компонентам возмущаемых строк (или столбцов) с коэффициентами пропорциональности $\varepsilon_1, \ldots, \varepsilon_{n+1}$. Такой способ наложения возмущений очень схож с процессом накопления вычислительной погрешности при использовании итерационных процессов решения нелинейных систем или пошаговых процессов решения систем обыкновенных дифференциальных уравнений.

В отличие от указанного способа, можно проводить возмущение матрицы системы путем добавления строки (или столбца) вида

$$(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n+1}).$$

В таком случае возмущения не зависят от возмущаемой строки (или столбца). Подобный вид возмущений можно сравнить с наложением неустранимой погрешности, возникающей на уровне построения модели, а не на уровне вычислений.

При всей схожести процессов наложения возмущений результаты их применения могут сильно отличаться, что является отдельной задачей для исследования.

Придадим $j+1$-й строке матрицы системы (13) возмущения, прибавив к ней строку $\varepsilon \cdot \left(\dfrac{\partial F_j}{\partial x_1}, \ldots, \dfrac{\partial F_j}{\partial x_{n+1}}\right)$, где ε – заданная малая величина, квадратом которой можно пренебречь по сравнению с меньшими степенями. Тогда возмущенная система продолжения решения запишется в виде

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_{n+1} \\ \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_{n+1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_j}{\partial x_1} + \varepsilon \cdot \dfrac{\partial F_j}{\partial x_1} & \cdots & \dfrac{\partial F_j}{\partial x_{n+1}} + \varepsilon \cdot \dfrac{\partial F_j}{\partial x_{n+1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} & \cdots & \dfrac{\partial F_n}{\partial x_{n+1}} \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} \\ \dfrac{dy_2}{d\mu} \\ \vdots \\ \dfrac{dy_{j+1}}{d\mu} \\ \vdots \\ \dfrac{dy_{n+1}}{d\mu} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{45}$$

Докажем вторую теорему.

Теорема 2. *Компоненты решения системы продолжения решения* (13) *при наложении малых однородных возмущений на любую строку ее матрицы системы, отличную от первой, не изменяются.*

Доказательство теоремы 2. По аналогии с доказательством теоремы 1, дадим решение для возмущенной системы (45):

$$\frac{dy_i}{d\mu} = (-1)^{i+1} \frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon}, \quad i = 1, \ldots, n+1. \tag{46}$$

Определитель Δ_ε может быть непосредственно вычислен вынесением общего множителя из $j+1$-й строки:

$$\Delta_\varepsilon = (1+\varepsilon) \cdot \Delta, \tag{47}$$

а для определителей $\Delta_{i\varepsilon}$ будут справедливы равенства

$$\Delta_{i\varepsilon} = (1+\varepsilon) \cdot \Delta_i, \quad i = 1, \ldots, n+1. \tag{48}$$

Подставляя равенства (47), (48) в формулы (46), получим

$$\frac{dy_i}{d\mu} = (-1)^{i+1}\frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon} = (-1)^{i+1}\frac{(1+\varepsilon)\cdot\Delta_i}{(1+\varepsilon)\cdot\Delta} = (-1)^{i+1}\frac{\Delta_i}{\Delta} = \frac{dx_i}{d\mu}, \quad i = 1,\ldots,n+1.$$

Таким образом, компоненты решения возмущенной системы (45) совпадают с компонентами решения исходной системы (13). При этом погрешности компонент решений $\delta_i \equiv 0,\ i = 1,\ldots,n+1$, и квадратичная погрешность $\delta \equiv 0$.

Доказательство теоремы 2 завершено.

Придадим j-му столбцу матрицы системы (13) возмущения, прибавив к нему столбец $\varepsilon\left(\alpha_j, \dfrac{\partial F_1}{\partial x_j}\ldots, \dfrac{\partial F_{n+1}}{\partial x_j}\right)^\top$, где ε – заданная малая величина, квадратом которой можно пренебречь по сравнению с меньшими степенями. Тогда возмущенная система продолжения решения запишется в виде

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_j + \varepsilon\cdot\alpha_j & \cdots & \alpha_{n+1} \\ \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_j} + \varepsilon\cdot\dfrac{\partial F_1}{\partial x_j} & \cdots & \dfrac{\partial F_1}{\partial x_{n+1}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} & \cdots & \dfrac{\partial F_n}{\partial x_j} + \varepsilon\cdot\dfrac{\partial F_n}{\partial x_j} & \cdots & \dfrac{\partial F_n}{\partial x_{n+1}} \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} \\ \dfrac{dy_2}{d\mu} \\ \vdots \\ \dfrac{dy_{n+1}}{d\mu} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{49}$$

Докажем третью теорему.

Теорема 3. *Квадратичная погрешность решения возмущенной системы продолжения решения* (49)*, полученной из системы продолжения решения* (13) *наложением малых однородных возмущений на любой j-й столбец ее матрицы системы при условии $\Delta_j \neq 0$, достигает наименьшего значения в том случае, когда вектор $\overline{\alpha} = (\alpha_1,\ldots,\alpha_{n+1})^\top$ направлен по касательной к кривой множества решений системы* (1) *в рассматриваемой точке, т. е. компоненты вектора $\overline{\alpha}$ задаются в виде* (14).

Доказательство теоремы 3. По аналогии с формулами (31), (32), найдем решение и для возмущенной системы (49) в виде (36). Определитель Δ_ε будет совпадать с выражением (37), а для миноров возмущенной матрицы системы (49) будут справедливы равенства

$$\Delta_{i\varepsilon} = (1+\varepsilon)\cdot\Delta_i, \quad i = 1,\ldots,n+1, \quad i \neq j, \quad \Delta_{j\varepsilon} = \Delta_j. \tag{50}$$

Подставляя равенства (37) и (50) в формулы (36), получим

$$\frac{dy_i}{d\mu} = (-1)^{i+1}\frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon} = (-1)^{i+1}\frac{\Delta_i}{\Delta} = \frac{dx_i}{d\mu}, \quad i = 1,\ldots,n+1, \quad i \neq j,$$
$$\frac{dy_j}{d\mu} = (-1)^{j+1}\frac{\Delta_{j\varepsilon}}{\Delta_\varepsilon} = (-1)^{i+1}\frac{\Delta_j}{(1+\varepsilon)\cdot\Delta}.$$

Как и в теореме 1, воспользовавшись малостью возмущений, получим

$$\frac{dy_j}{d\mu} = (-1)^{j+1}\frac{(1-\varepsilon)\cdot\Delta_j}{\Delta}. \tag{51}$$

Учитывая равенства $\dfrac{dy_i}{d\mu} = \dfrac{dx_i}{d\mu}$ при $i \neq j$ и (51), вычислим погрешности компонент решения (33):

$$\delta_i = 0, \quad i = 1,\ldots,n+1, \quad i \neq j, \quad \delta_j = -(-1)^{j+1}\frac{\varepsilon\cdot\Delta_j}{\Delta}.$$

Тогда, согласно формуле (34), квадратичная погрешность будет иметь вид

$$\delta = \frac{\varepsilon^2\cdot\Delta_j^2}{\Delta^2}. \tag{52}$$

Поставим задачу минимизации квадратичной ошибки (52) при ограничении типа равенств (12):

$$\frac{\varepsilon^2\Delta_j^2}{\Delta^2} \to \min_{\overline{\alpha}\in\mathbb{R}^{n+1}}, \quad \alpha_i\cdot\alpha_i = 1, \quad i = 1,\ldots,n+1. \tag{53}$$

Полученная задача поиска минимума при ограничении типа равенств (53) удовлетворяет условиям леммы 2 при $A = \varepsilon^2\Delta_j^2 > 0$ и $m = 2$. Это означает, что эта задача имеет точку минимума с компонентами (17), совпадающими с компонентами касательного вектора к кривой множества решений в рассматриваемой точке, т.е. удовлетворяющими равенствам (44). В этой точке квадратичная погрешность (52) принимает наименьшее значение.

Доказательство теоремы 3 завершено.

Замечание 2. Условие $\Delta_j \neq 0$ в теореме 3 является существенным, поскольку в противном случае квадратичная погрешность $\delta \equiv 0$ и постановка задачи ее минимизации теряет смысл.

Можно обобщить теоремы 2 и 3 на случай, когда возмущаются не одна строка или столбец, а произвольное их количество.

Теорема 4. *Компоненты решения системы продолжения решения* (13) *при малых однородных возмущениях любых k строк (где $k = 1, \ldots, n$) ее матрицы системы, отличных от первой, не изменяются.*

Теорема 5. *Квадратичная погрешность решения возмущенной системы продолжения решения, полученной из системы продолжения решения* (13) *наложением малых однородных возмущениях на любые k столбцов (где $k = 1, \ldots, n+1$) ее матрицы системы с номерами $i_1, \ldots, i_k$ при условии $\Delta_{i_1}^2 + \cdots + \Delta_{i_k}^2 \neq 0$, достигает наименьшего значения в том случае, когда вектор $\overline{\alpha} = (\alpha_1, \ldots, \alpha_{n+1})^\top$ направлен по касательной к кривой множества решений системы* (1) *в рассматриваемой точке, т. е. компоненты вектора $\overline{\alpha}$ задаются в виде* (14).

При этом доказательства этих утверждений аналогичны приведенным.

Для полного доказательства теоремы о минимальности квадратичной погрешности, возникающей при возмущении элементов матрицы системы продолжения решения (13), при малых однородных возмущениях остается показать, что это свойство также будет справедливо и при возмущении произвольных k строк и m столбцов матрицы системы, где $k, m = 1, \ldots, n+1$.

Поскольку система продолжения решения (13) является линейной, то можно провести ее декомпозицию на случаи, рассмотренные в теоремах 1-5, что и доказывает теорему о минимальной квадратичной погрешности. Однако целесообразно дать и конструктивное доказательство. Докажем теорему для случая, когда к каждой строке и столбцу матрицы системы (13) прибавлены они же, домноженные на $\varepsilon/2$, где ε – заданная малая величина, квадратом которой можно пренебречь. Тогда возмущенная система продолжения решения запишется в виде

$$\begin{pmatrix} \alpha_1 + \varepsilon\cdot\alpha_1 & \alpha_2 + \varepsilon\cdot\alpha_2 & \cdots & \alpha_{n+1} + \varepsilon\cdot\alpha_{n+1} \\ \dfrac{\partial F_1}{\partial x_1} + \varepsilon\cdot\dfrac{\partial F_1}{\partial x_1} & \dfrac{\partial F_1}{\partial x_2} + \varepsilon\cdot\dfrac{\partial F_1}{\partial x_2} & \cdots & \dfrac{\partial F_1}{\partial x_{n+1}} + \varepsilon\cdot\dfrac{\partial F_1}{\partial x_{n+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_n}{\partial x_1} + \varepsilon\cdot\dfrac{\partial F_n}{\partial x_1} & \dfrac{\partial F_n}{\partial x_2} + \varepsilon\cdot\dfrac{\partial F_n}{\partial x_2} & \cdots & \dfrac{\partial F_n}{\partial x_{n+1}} + \varepsilon\cdot\dfrac{\partial F_n}{\partial x_{n+1}} \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} \\ \dfrac{dy_2}{d\mu} \\ \vdots \\ \dfrac{dy_{n+1}}{d\mu} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{54}$$

В других случаях доказательство не будет изменяться, за исключением вида квадратичной ошибки для конкретного рассматриваемого случая.

Докажем общую теорему.

Теорема 6. *Квадратичная погрешность решения возмущенной системы продолжения решения, полученной из системы продолжения решения* (13) *наложением малых однородных возмущений на произвольные k строк и m столбцов (где $k, m = 1, \ldots, n+1$) матрицы системы, достигает наименьшего значения в том случае, когда вектор $\overline{\alpha} = (\alpha_1, \ldots, \alpha_{n+1})^\top$ направлен по касательной к кривой множества решений системы* (1) *в рассматриваемой точке, т. е. компоненты вектора $\overline{\alpha}$ задаются в виде* (14).

Доказательство теоремы 6. По аналогии с формулами (31), (32), найдем решение и для возмущенной системы (54) в виде (36). Определитель Δ_ε будет вычисляться по формуле

$$\Delta_\varepsilon = (1+\varepsilon)^{n+1}\cdot\Delta, \tag{55}$$

а для миноров возмущенной матрицы системы (54) будут справедливы равенства

$$\Delta_{i\varepsilon} = (1+\varepsilon)^n\cdot\Delta_i, \quad i = 1, \ldots, n+1. \tag{56}$$

Учитывая малость возмущений, можно упростить выражения (55), (56):

$$\Delta_\varepsilon = [1 + (n+1)\cdot\varepsilon]\cdot\Delta, \quad \Delta_{i\varepsilon} = (1 + n\cdot\varepsilon)\cdot\Delta_i, \quad i = 1, \ldots, n+1. \tag{57}$$

Подставляя равенства (57) в формулы (36), получим

$$\frac{dy_i}{d\mu} = (-1)^{i+1}\frac{\Delta_{i\varepsilon}}{\Delta_\varepsilon} = (-1)^{i+1}\frac{(1+n\cdot\varepsilon)\cdot\Delta_i}{[1+(n+1)\cdot\varepsilon]\cdot\Delta}, \quad i=1,\ldots,n+1.$$

Используя соотношения (33), вычислим

$$\delta_i = \frac{dy_i}{d\mu} - \frac{dx_i}{d\mu} = (-1)^{i+1}\frac{(1+n\cdot\varepsilon)\cdot\Delta_i}{[1+(n+1)\cdot\varepsilon]\cdot\Delta} - (-1)^{i+1}\frac{\Delta_i}{\Delta}, \quad i=1,\ldots,n+1.$$

Приведя дроби к общему знаменателю, домножая числитель и знаменатель полученной дроби на $1-(n+1)\cdot\varepsilon$ и отбрасывая все слагаемые, содержащие ε^2, получим $\delta_i, i=1,\ldots,n+1$, вида (41). Тогда, согласно формуле (34), квадратичная погрешность будет иметь вид (42), аналогичный полученному в теореме 1. Задача ее минимизации при ограничении типа равенств (12) будет совпадать с задачей (43). Она имеет точку минимума с компонентами (17), совпадающими с компонентами касательного вектора к кривой множества решений в рассматриваемой точке, т. е. удовлетворяющими равенствам (44). В этой точке квадратичная погрешность (42) принимает наименьшее значение.

Доказательство теоремы 6 завершено.

5. ЧИСЛЕННЫЙ ПРИМЕР

В качестве примера, иллюстрирующего полученные теоретические результаты, рассмотрим задачу построения лемнискаты Бернулли, задаваемой в декартовой системе координат уравнением

$$F(x_1,x_2) = \left(x_1^2+x_2^2\right)^2 - 2a^2\left(x_1^2-x_2^2\right) = 0, \tag{58}$$

где a – заданный вещественный параметр.

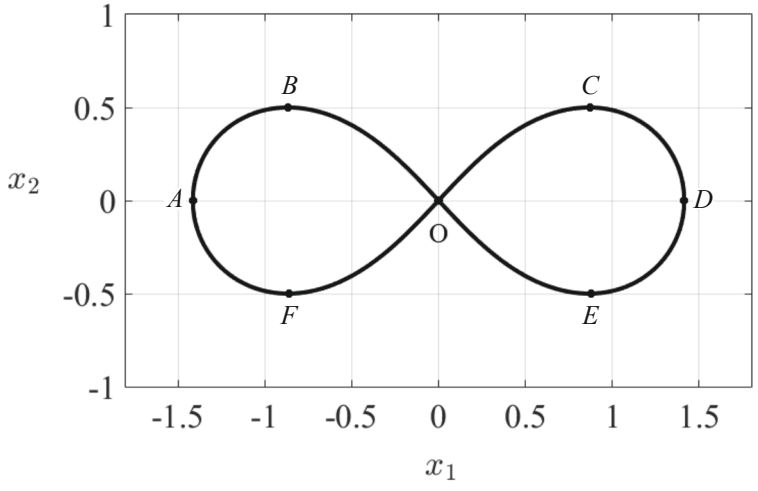
Для построения графика лемнискаты Бернулли удобнее воспользоваться ее уравнением в полярных координатах

$$r^2 = 2a^2\cos 2\varphi,$$

используя которое можно дать выражения для координат x_1 и x_2:

$$x_1 = r\cos\varphi = \pm\sqrt{2}a\sqrt{\cos 2\varphi}\cos\varphi,$$
$$x_2 = r\sin\varphi = \pm\sqrt{2}a\sqrt{\cos 2\varphi}\sin\varphi.$$

Здесь угол $\varphi\in[-\pi/4;\pi/4]$. Положительному знаку соответствует часть лемнискаты, расположенная в правой полуплоскости, а отрицательному – в левой. При $a=1$ график лемнискаты Бернулли изображен на фиг. 1.



Фиг. 1. Лемниската Бернулли.

Если рассматривать построение лемнискаты Бернулли как задачу численного решения уравнения (58), зависящего от двух неизвестных, то процесс решения при помощи алгоритмов, описанных в разд. 2, будет сопряжен со значительными трудностями. Это связано с наличием на лемнискате двух предельных особых точек

для параметра продолжения x_1 (точки A и D), четырех предельных особых точек для параметра продолжения x_2 (точки B, C, E и F) и одной существенно особой точки (точку ветвления), расположенной в начале координат O.

Построение лемнискаты Бернулли методом продолжения решения предполагает зависимость неизвестных уравнения (58) от параметра продолжения μ:

$$x_1 = x_1(\mu), \quad x_2 = x_2(\mu). \tag{59}$$

Параметр продолжения решения для данного уравнения задается локально, в окрестности каждой точки кривой множества решений, в форме

$$d\mu = \alpha_1 dx_1 + \alpha_2 dx_2.$$

Последнее соотношение можно переписать в виде

$$\alpha_1 \frac{dx_1}{d\mu} + \alpha_2 \frac{dx_2}{d\mu} = 1. \tag{60}$$

Если теперь продифференцировать уравнение (58) по параметру μ, учитывая зависимости (59), и дополнить его уравнением (60), то будет получена замкнутая система продолжения решения из двух дифференциальных уравнений относительно производных $\frac{dx_1}{d\mu}$ и $\frac{dx_2}{d\mu}$, которую в векторно-матричном виде можно записать в форме

$$\begin{pmatrix} \alpha_1 & \alpha_2 \\ F_1 & F_2 \end{pmatrix} \cdot \begin{pmatrix} \frac{dx_1}{d\mu} & \frac{dx_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{61}$$

где

$$F_1 = \frac{\partial F}{\partial x_1} = 2x_1\left(x_1^2 + x_2^2\right) - 4a^2 x_1, \quad F_2 = \frac{\partial F}{\partial x_2} = 2x_2\left(x_1^2 + x_2^2\right) + 4a^2 x_2.$$

Таким образом, построение кривой множества решений уравнения (58) из начальной точки $M_0(x_{10}, x_{20})$ по выбранному параметру продолжения решения (с заданными значениями α_1 и α_2), сводится к решению начальной задачи для системы уравнений (61) с начальными условиями

$$x_1(0) = x_{10},\ x_2(0) = x_{20}. \tag{62}$$

При построении лемнискаты Бернулли методом Давиденко используем два параметра продолжения: x_1, которому соответствуют значения $\alpha_1 = 1$ и $\alpha_2 = 0$; x_2, которому соответствуют значения $\alpha_1 = 0$ и $\alpha_2 = 1$. Для параметра продолжения x_1 начальная задача (61), (62) перепишется в виде

$$\frac{dx_2}{dx_1} = -\frac{F_1}{F_2}, \quad x_2(x_{10}) = x_{20}, \tag{63}$$

а для параметра продолжения x_2:

$$\frac{dx_1}{dx_2} = -\frac{F_2}{F_1}, \quad x_1(x_{20}) = x_{10}. \tag{64}$$

Обе задачи (63) и (64) решались для значения $a = 1$ численно явным методом Эйлера с постоянным шагом интегрирования $h = 10^{-4}$. В процессе решения приходилось проводить смену параметра продолжения решения четыре раза при специальном выборе начальной точки $x_{10} = 1,\ x_{20} = \sqrt{\sqrt{5} - 2}$. Из начальной точки задача (63) решалась до момента достижения параметром x_1 значения -1. Затем производилась смена параметра продолжения на x_2 и из последней полученной точки решалась задача (64) до достижения параметром x_2 значения $-\sqrt{\sqrt{5} - 2}$. После этого производился возврат к параметру x_1 и задача (63) решалась из последней полученной точки до достижения параметром x_1 значения 1. Последняя смена параметра производилась на x_2, после этого задача (64) решалась до возврата в начальную точку.

График лемнискаты, полученный методом Давиденко при $a = 1$, изображен на фиг. 2а.

Для оценки погрешности вычислений в точке будем использовать абсолютную погрешность

$$\epsilon(x_1, x_2) = \left| \left(x_1^2 + x_2^2\right)^2 - 2a^2\left(x_1^2 - x_2^2\right) \right|. \tag{65}$$

В табл. 1 даны значения погрешностей решения уравнения (58), полученного методом Давиденко, где ϵ_{av} – среднее значение погрешности в точке, ϵ_{med} – медиана значений погрешности, ϵ_{std} – среднее квадратическое отклонение значений погрешности от среднего значения.

Таблица 1. Погрешность ϵ вида (65), возникающая при численном построении лемнискаты Бернулли

	Метод Давиденко			Наилучшая парамеризация		
Невозмущенная система продолжения решения						
	ϵ_{av}	ϵ_{med}	ϵ_{std}	ϵ_{av}	ϵ_{med}	ϵ_{std}
	0.0011	0.0014	$8.0898 \cdot 10^{-4}$	$8.4093 \cdot 10^{-4}$	0.0010	$5.7732 \cdot 10^{-4}$
Возмущение первой строки матрицы системы продолжения решения						
$\varepsilon = 0.01$	0.0011	0.0014	$8.0125 \cdot 10^{-4}$	$8.3265 \cdot 10^{-4}$	$9.9890 \cdot 10^{-4}$	$5.7163 \cdot 10^{-4}$
$\varepsilon = 0.05$	0.0011	0.0014	$7.7066 \cdot 10^{-4}$	$8.0108 \cdot 10^{-4}$	$9.6099 \cdot 10^{-4}$	$5.4994 \cdot 10^{-4}$
$\varepsilon = 0.1$	0.0010	0.0013	$7.3585 \cdot 10^{-4}$	$7.6483 \cdot 10^{-4}$	$9.1747 \cdot 10^{-4}$	$5.2504 \cdot 10^{-4}$
Возмущение второго столбца матрицы системы продолжения решения						
$\varepsilon = 0.01$	0.0055	0.0049	0.0043	0.0051	0.0046	0.0040
$\varepsilon = 0.05$	0.0231	0.0200	0.0203	0.0221	0.0191	0.0192
$\varepsilon = 0.1$	0.0454	0.0382	0.0407	0.0434	0.0362	0.0388

Более эффективным способом решения уравнения (58) является применение наилучшего параметра, дифференциал которого для уравнения (58) удовлетворяет соотношению

$$(d\lambda)^2 = (dx_1)^2 + (dx_2)^2. \tag{66}$$

Наилучшему параметру соответствуют значения $\alpha_1 = \dfrac{dx_1}{d\mu}$ и $\alpha_2 = \dfrac{dx_2}{d\mu}$, входящие в систему продолжения решения (61) с заменой обозначения μ на λ.

Решение уравнения (58) с использованием наилучшего параметра при выборе положительного направления обхода кривой множества решений сводится к решению задачи Коши

$$\frac{dx_1}{d\lambda} = -\frac{F_2}{\sqrt{F_1^2 + F_2^2}},\ \frac{dx_2}{d\lambda} = \frac{F_1}{\sqrt{F_1^2 + F_2^2}},\quad x_1(0) = x_{10},\ x_2(0) = x_{20}. \tag{67}$$

Решение задачи (67) также проводилось для значения $a = 1$ явным методом Эйлера с постоянным шагом $l = 10^{-4}$. Можно видеть, что при решении начальной задачи (67) исчезает проблема прохождения предельных особых точек, что делает ненужным смену параметра продолжения решения. Однако остается проблема в прохождении существенно особой точки, лежащей в начале координат.

График лемнискаты, полученный методом наилучшей параметризации при $a = 1$ имеет вид, аналогичный изображенному на фиг. 2а.

В табл. 1 даны значения погрешностей решения уравнения (58), полученного методом наилучшей параметризации.

Проанализируем полученные решения. Основным неудобством при решении задач (63) и (64) является необходимость смены параметра продолжения. Но это не единственный недостаток. Из табл. 1 видно, что даже используя специально выбранные начальную точку и точки смены параметра, лежащие вне малых окрестностей предельных и существенно особых точек, погрешность решения задач (63) и (64) превосходит погрешность решения задачи (67). При специальном выборе начальной точки и точек смены параметра погрешности, конечно, различаются незначительно. Но при произвольном выборе начальной точки и момента смены параметра погрешность решения задач (63) и (64) может значительно вырасти, как в окрестности начальной точки, так и в окрестности точек смены параметра. Последнее происходит из-за трудности выявления предельных и существенно особых точек при численном счете. В прикладных задачах аналитическое решение возможно лишь в исключительных случаях, а методы оценки локальной погрешности могут давать искаженные результаты в окрестности предельных и существенно особых точек.

5.1. Возмущение первой строки матрицы продолжения решения

Рассмотрим наложение малых возмущений на первую строку матрицы системы продолжения решения. Тогда возмущенная система (61) запишется в виде

$$\begin{pmatrix} \alpha_1 + \varepsilon \cdot \alpha_1 & \alpha_2 + \varepsilon \cdot \alpha_2 \\ F_1 & F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{68}$$

где ε – малое заданное значение, квадратом которого можно пренебречь по сравнению с меньшими степенями. Для визуального разделения будем компоненты возмущенного решения обозначать y_1 и y_2.

Для трех рассмотренных ранее начальных задач (63), (64) и (67) можно получить возмущенные аналоги. Для возмущенного параметра $\mu = x_1$ система продолжения решения (68) перепишется в виде

$$\begin{pmatrix} 1 + \varepsilon & 0 \\ F_1 & F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

и сведется к начальной задаче

$$\frac{dy_1}{d\mu} = \frac{1}{1+\varepsilon}, \quad \frac{dy_2}{d\mu} = -\frac{1}{1+\varepsilon} \cdot \frac{F_1}{F_2}, \quad y_1(0) = x_{10},\ y_2(0) = x_{20}. \tag{69}$$

Для возмущенного параметра продолжения $\mu = x_2$ система продолжения решения (68) перепишется в виде

$$\begin{pmatrix} 0 & 1 + \varepsilon \\ F_1 & F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

и сведется к начальной задаче

$$\frac{dy_1}{d\mu} = -\frac{1}{1+\varepsilon} \cdot \frac{F_2}{F_1}, \quad \frac{dy_2}{d\mu} = \frac{1}{1+\varepsilon}, \quad y_1(0) = x_{10},\ y_2(0) = x_{20}. \tag{70}$$

При использовании возмущенного наилучшего параметра (66) система продолжения решения (68) перепишется в виде

$$\begin{pmatrix} (1+\varepsilon) \cdot \dfrac{dx_1}{d\lambda} & (1+\varepsilon) \cdot \dfrac{dx_2}{d\lambda} \\ F_1 & F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

и сведется к начальной задаче

$$\begin{aligned} &\frac{dy_1}{d\mu} = -\frac{1}{1+\varepsilon} \cdot \frac{F_2}{\sqrt{F_1^2 + F_2^2}}, \ \frac{dy_2}{d\mu} = \frac{1}{1+\varepsilon} \cdot \frac{F_1}{\sqrt{F_1^2 + F_2^2}}, \\ &y_1(0) = x_{10},\ y_2(0) = x_{20}. \end{aligned} \tag{71}$$

Отметим, что в функциях F_1 и F_2, входящих в начальные задачи (69), (70) и (71), переменными являются уже не x_1 и x_2, а y_1 и y_2.

Вычислим погрешности решения системы продолжения решения для случая выбора в качестве параметра продолжения возмущенной переменной $\mu = x_2$:

$$\delta_{21} = \frac{dx_1}{dx_2} - \frac{dy_1}{d\mu} = -\frac{\varepsilon}{1+\varepsilon} \cdot \frac{F_2}{F_1}, \quad \delta_{22} = \frac{dx_2}{dx_2} - \frac{dy_2}{d\mu} = \frac{\varepsilon}{1+\varepsilon}.$$

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_2 = \delta_{21}^2 + \delta_{22}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \left[1 + \left(\frac{F_2}{F_1}\right)^2\right].$$

Если выбрать в качестве параметра продолжения возмущенную переменную $\mu = x_1$, то

$$\delta_{11} = \frac{dx_1}{dx_1} - \frac{dy_1}{d\mu} = \frac{\varepsilon}{1+\varepsilon}, \quad \delta_{12} = \frac{dx_2}{dx_1} - \frac{dy_2}{d\mu} = -\frac{\varepsilon}{1+\varepsilon} \cdot \frac{F_1}{F_2}.$$

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_1 = \delta_{11}^2 + \delta_{12}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \left[1 + \left(\frac{F_1}{F_2}\right)^2\right].$$

Для возмущенного наилучшего параметра продолжения решения погрешности будут вычислены в виде

$$\delta_{\lambda 1} = \frac{dx_1}{d\lambda} - \frac{dy_1}{d\lambda} = -\frac{\varepsilon \cdot F_2}{(1+\varepsilon) \cdot \sqrt{F_1^2 + F_2^2}},$$
$$\delta_{\lambda 2} = \frac{dx_2}{d\lambda} - \frac{dy_2}{d\lambda} = -\frac{\varepsilon \cdot F_1}{(1+\varepsilon) \cdot \sqrt{F_1^2 + F_2^2}}.$$

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_\lambda = \delta_{\lambda 1}^2 + \delta_{\lambda 2}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2}. \tag{72}$$

Используя оценки сверху для квадратичной погрешности (72)

$$\frac{\varepsilon^2}{(1+\varepsilon)^2} \le \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \left[1 + \left(\frac{F_2}{F_1}\right)^2\right], \quad \frac{\varepsilon^2}{(1+\varepsilon)^2} \le \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \left[1 + \left(\frac{F_1}{F_2}\right)^2\right],$$

можно получить неравенства

$$\delta_\lambda \le \delta_1, \quad \delta_\lambda \le \delta_2,$$

т. е. квадратичная погрешность возмущенной системы продолжения решения (68) при выборе наилучшего параметра не превосходит значений, полученных с использованием параметров продолжения x_1 и x_2. Согласно полученным теоретическим результатам, эти неравенства будут справедливы и для других параметров продолжения решения, отличных от x_1 и x_2.

Эти результаты позволяют утверждать, что и для уравнения (58) воздействие возмущений (вычислительных погрешностей) будет наименьшим при использовании для его решения наилучшего параметра. Продемонстрируем это на примере численного решения возмущенных задач.

Возмущенные задачи (69), (70) и (71) решались для значения $a = 1$ явным методом Эйлера с постоянным шагом интегрирования, таким же как и для невозмущенных задач. Параметр возмущения $\varepsilon = 0.01; 0.05; 0.1$.

Графики лемнискаты для возмущенных задач при $a = 1$ и различных значениях ε, полученные методами Давиденко и наилучшей параметризации, имеют вид аналогичный изображенному на фиг. 1. Графики решений возмущенных задач близки к графику точного решения, за исключением существенной особой точки, лежащей в начале координат. В ее окрестности происходит малое отклонение от точного решения.

В табл. 1 даны значения погрешностей решений возмущенных задач для всех рассматриваемых случаев.

При однородных малых возмущениях первой строки матрицы системы продолжения решения, погрешность решения задач (69), (70) незначительно больше по сравнению с погрешностью решения задачи (71). Это связано, как и в невозмущенном случае, с удачным выбором начальной точки и точек смены параметра.

Отличительной особенностью этого случая возмущения является уменьшение погрешности по сравнению с невозмущенным случаем. Это можно объяснить тем, что возмущениям подвержена только первая строка матрицы системы продолжения решения, отвечающая за задание параметра μ. Поэтому возмущенный параметр продолжения решения μ можно рассматривать как удлиненный в $1 + \varepsilon$ раз параметр продолжения x_1 и x_2 для задач (69) и (70) соответственно, и удлиненный в $\sqrt{1+\varepsilon}$ раз наилучший параметр для задачи (71). Поскольку происходит удлинение параметра продолжения, то участки кривой множества решений большой кривизны удается преодолеть с меньшей погрешностью. Этим и объясняется уменьшение погрешности решения для данного класса возмущенных задач.

5.2. Возмущение столбца матрицы продолжения решения

Рассмотрим наложение малых возмущений на второй столбец матрицы системы продолжения решения. Тогда возмущенная система (61) запишется в виде

$$\begin{pmatrix} \alpha_1 & \alpha_2 + \varepsilon \cdot \alpha_2 \\ F_1 & F_2 + \varepsilon \cdot F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^\top = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{73}$$

Как и в п. 5.1, для трех рассмотренных ранее начальных задач (63), (64) и (67) получим возмущенные аналоги. Для параметра продолжения $\mu = x_1$ система продолжения решения (73) перепишется в виде

$$\begin{pmatrix} 1 & 0 \\ F_1 & F_2 + \varepsilon \cdot F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

и сведется к начальной задаче

$$\frac{dy_1}{d\mu} = 1, \quad \frac{dy_2}{d\mu} = -\frac{1}{1+\varepsilon} \cdot \frac{F_1}{F_2}, \quad y_1(0) = x_{10},\ y_2(0) = x_{20}, \tag{74}$$

эквивалентной начальной задаче

$$\frac{dy_2}{dx_1} = -\frac{1}{1+\varepsilon} \cdot \frac{F_1}{F_2}, \quad y_2(x_{10}) = x_{20}.$$

Для возмущенного параметра продолжения $\mu = x_2$ система продолжения решения (73) перепишется в виде

$$\begin{pmatrix} 0 & 1+\varepsilon \\ F_1 & F_2 + \varepsilon \cdot F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

и сведется к начальной задаче

$$\frac{dy_1}{d\mu} = -\frac{F_2}{F_1}, \quad \frac{dy_2}{d\mu} = \frac{1}{1+\varepsilon}, \quad y_1(0) = x_{10},\ y_2(0) = x_{20}. \tag{75}$$

При использовании возмущенного наилучшего параметра (66) система продолжения решения (73) перепишется в виде

$$\begin{pmatrix} \dfrac{dx_1}{d\lambda} & (1+\varepsilon) \cdot \dfrac{dx_2}{d\lambda} \\ F_1 & (1+\varepsilon) \cdot F_2 \end{pmatrix} \cdot \begin{pmatrix} \dfrac{dy_1}{d\mu} & \dfrac{dy_2}{d\mu} \end{pmatrix}^{\top} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

и сведется к начальной задаче

$$\frac{dy_1}{d\mu} = -\frac{F_2}{\sqrt{F_1^2 + F_2^2}}, \ \frac{dy_2}{d\mu} = \frac{F_1}{(1+\varepsilon) \cdot \sqrt{F_1^2 + F_2^2}}, \quad y_1(0) = x_{10},\ y_2(0) = x_{20}. \tag{76}$$

Вычислим погрешности решения системы продолжения решения для случая выбора в качестве параметра продолжения возмущенной переменной $\mu = x_2$:

$$\delta_{21} = \frac{dx_1}{dx_2} - \frac{dy_1}{d\mu} = 0, \quad \delta_{22} = \frac{dx_2}{dx_2} - \frac{dy_2}{d\mu} = \frac{\varepsilon}{1+\varepsilon}.$$

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_2 = \delta_{21}^2 + \delta_{22}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2}.$$

Если выбрать в качестве параметра продолжения переменную $\mu = x_1$, то

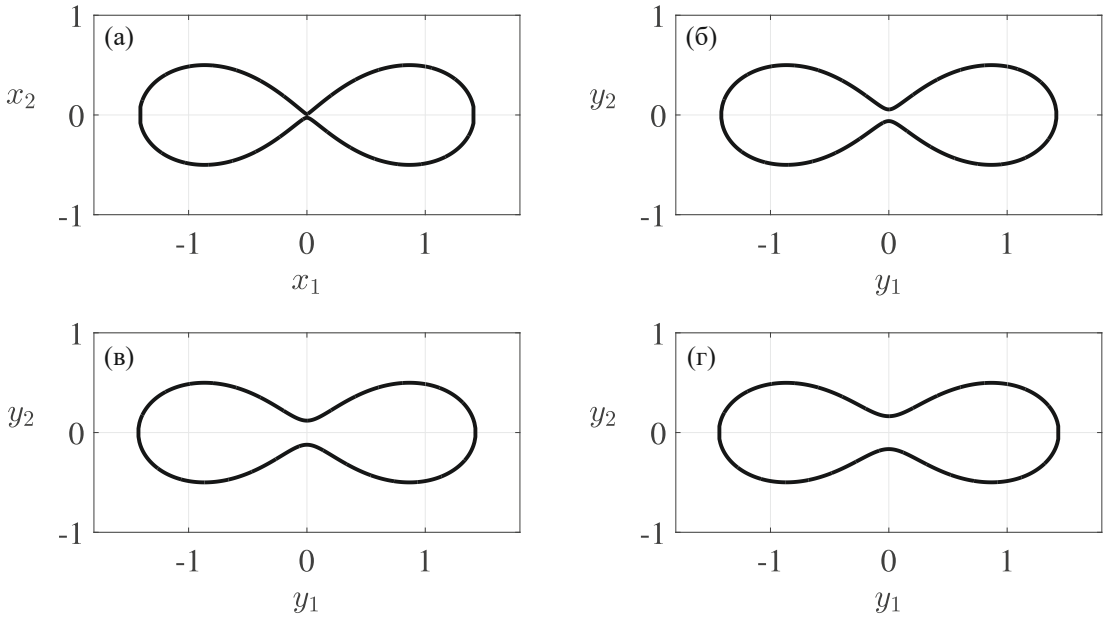
$$\delta_{11} = \frac{dx_1}{dx_1} - \frac{dy_1}{dy_1} = 0, \quad \delta_{12} = \frac{dx_2}{dx_1} - \frac{dy_2}{dy_1} = -\frac{\varepsilon}{1+\varepsilon} \cdot \frac{F_1}{F_2}.$$

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_1 = \delta_{11}^2 + \delta_{12}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \left(\frac{F_1}{F_2}\right)^2.$$

Для наилучшего параметра продолжения решения погрешности будут вычислены в виде

$$\delta_{\lambda 1} = \frac{dx_1}{d\lambda} - \frac{dy_1}{d\mu} = 0, \quad \delta_{\lambda 2} = \frac{dx_2}{d\lambda} - \frac{dy_2}{d\mu} = \frac{\varepsilon \cdot F_1}{(1+\varepsilon) \cdot \sqrt{F_1^2 + F_2^2}}.$$



Фиг. 2. Лемниската Бернулли, метод Давиденко, система продолжения решения с возмущенным вторым столбцом: (а) — для $\varepsilon = 0$, (б) — для $\varepsilon = 0.01$, (в) — для $\varepsilon = 0.05$, (г) — для $\varepsilon = 0.1$.

Тогда квадратичная погрешность для данного случая будет равна

$$\delta_\lambda = \delta_{\lambda 1}^2 + \delta_{\lambda 2}^2 = \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \frac{F_1^2}{F_1^2 + F_2^2}. \tag{77}$$

Используя оценки сверху для квадратичной погрешности (77)

$$\frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \frac{F_1^2}{F_1^2 + F_2^2} \le \frac{\varepsilon^2}{(1+\varepsilon)^2}, \quad \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \frac{F_1^2}{F_1^2 + F_2^2} \le \frac{\varepsilon^2}{(1+\varepsilon)^2} \cdot \frac{F_1^2}{F_2^2},$$

можно получить неравенства

$$\delta_\lambda \le \delta_1, \quad \delta_\lambda \le \delta_2.$$

Как и в предыдущем рассмотренном случае, квадратичная погрешность решения возмущенной системы продолжения решения (73) меньше при выборе наилучшего параметра, по сравнению с использованием параметров продолжения x_1 и x_2. Согласно полученным теоретическим результатам, эти неравенства будут справедливы и для других параметров продолжения решения, отличных от x_1 и x_2.

Возмущенные задачи (75), (74) и (76) решались для значения $a = 1$ явным методом Эйлера с постоянным шагом интегрирования, таким же как и для невозмущенных задач. Параметр возмущения $\varepsilon = 0.01; 0.05; 0.1$.

Графики лемнискаты, полученные для возмущенных задач при $a = 1$ для различных значений ε, изображены на фиг. 2б–2г. Так как погрешности решений задач (75), (74) и (76) мало отличаются, то на фиг. 2 приводится только графики, полученные методом Давиденко. Кривые множества решений, полученные с использованием наилучшего параметра имеют аналогичный вид.

В табл. 1 даны значения погрешностей решений возмущенных задач для всех рассматриваемых случаев. Все полученные расчетные данные полностью согласуются с теоретическими результатами.

6. ЗАКЛЮЧЕНИЕ

В данной работе исследуется одно из свойств наилучшего параметра, которое имеет важное значение в прикладных расчетах. Это свойство минимальности квадратичной погрешности, возникающей при возмущении элементов матрицы системы продолжения решения (13).

Целью данной работы было строгое доказательство минимальности квадратичной погрешности решения системы продолжения решения (13) с возмущенной матрицей системы при использовании наилучшего параметра. Это удалось доказать для случая малых однородных возмущений, значения которых одинаковы, а их квадратами можно пренебречь.

Все теоретические результаты были полностью подтверждены на примере численного построения лемнискаты Бернулли (численного решения нелинейного уравнения с двумя неизвестными). Построение лемнискаты Бернулли показывает, что при использовании в качестве параметров продолжения решения x_1 и x_2 даже в случае удачного выбора начальной точки и точек смены параметра (которые не должны попадать в окрестность предельных и существенно особых точек), полученная квадратичная погрешность превосходит значение, полученное при использовании наилучшего параметра. Это оказывается справедливым даже при наличии у лемнискаты существенной особой точки в начале координат, хотя ее наличия и не предполагалось при доказательстве утверждений статьи. При прохождении существенно особой точки эффективность использования наилучшего параметра падает. Более эффективные способы прохождения существенно особых точек с использованием наилучшего параметра рассмотрены в работах [9, 10].

При доказательстве теорем в статье предполагалось, что возмущения накладываются на элемент матрицы системы продолжения решения пропорционально его значению (с коэффициентом пропорциональности ε), т. е. возмущения являются зависимыми, что соответствует возникновению вычислительной погрешности при использовании приближенных методов решения. Таким образом, преобразование задачи к наилучшему параметру позволяет минимизировать воздействие на решение зависимых возмущений, в том числе и погрешность вычислений.

СПИСОК ЛИТЕРАТУРЫ

1. *Lahaye M. E.* Une metode de resolution d'une categorie d'equations transcendentes // Comptes Rendus hebdomadaires des seances de L'Academie des sciences. 1934. Vol. 198. No. 21. P. 1840–1842.
2. *Lahaye M. E.* Solution of system of transcendental equations // Acad. Roy. Belg. Bull. Cl. Sci. 1948. Vol. 5. P. 805–822.
3. *Давиденко Д. Ф.* Об одном новом методе численного решения систем нелинейных уравнений // Докл. АН СССР. 1953 Т. 88. № 4. С. 601–602.
4. *Давиденко Д. Ф.* О приближенном решении систем нелинейных уравнений // Украинский матем. ж. 1953 Т. 5. № 2. С. 196–206.
5. *Ворович И. И., Зипалова В. Ф.* К решению нелинейных краевых задач теории упругости методом перехода к задаче Коши // Прикл. матем. и механ. 1965. Т. 29. Вып. 5. С. 894–901.
6. *Рикс Э.* Применение метода Ньютона к задаче упругой устойчивости // Прикл. механ. 1972. № 5. С. 204–210.
7. *Кузнецов Е. Б., Шалашилин В. И.* Задача Коши как задача продолжения по наилучшему параметру // Дифференц. ур-ния. 1994. Т. 30. № 6. С. 964–971.
8. *Шалашилин В. И., Кузнецов Е. Б.* Метод продолжения решения по параметру и наилучшая параметризация в прикладной математике и механике. М.: Эдиториал УРСС, 1999.
9. *Красников С.Д., Кузнецов Е.Б.* Численное продолжение решения в особых точках коразмерности единица // Ж. вычисл. матем. и матем. физ. 2015. Т. 55. № 11. С. 1835–1856.
10. *Красников С.Д., Кузнецов Е.Б.* Численное продолжение решения в особых точках высокой коразмерности для систем нелинейных алгебраических или трансцендентных уравнений // Ж. вычисл. матем. и матем. физ. 2016. Т. 56. № 9. С. 1571–1585.

ON THE MINIMALITY OF SQUARED ERROR OF SOLUTIONS TO SYSTEMS OF EQUATIONS TRANSFORMED TO THE BEST PARAMETER UNDER SMALL HOMOGENEOUS PERTURBATIONS

E. B. Kuznetsov[a,*], **S. S. Leonov**[a,b,**]

[a] *125993 Moscow, Volokolamsk Highway 4, Moscow Aviation Institute (National Research University), Russia*
[b] *117198 6 Miklukho-Maklay str., Patrice Lumumba Peoples' Friendship University of Russia, Moscow, Russia*
**e-mail: kuznetsov@mai.ru*
***e-mail: powerandglory@yandex.ru*

Received: 05.05.2024
Revised: 05.08.2024
Accepted: 23.08.2024

Abstract. Solving of systems of nonlinear equations with a scalar parameter is studied. The set of solutions to such systems is a curve in the space of variables of the equation system and the parameter. Its construction is usually carried out using numerical methods and is associated with numerous difficulties arising due to the presence of limiting and essentially singular points on the curve of the set of solutions. To find such curves, the method of solution continuation with respect to a parameter and the best parameterization is used, which allows us to reduce the solution to the Cauchy problem for a system of differential equations of solution continuation. Stability of the solution to perturbations introduced into the continuation system is investigated. For the first time, the previously formulated proposition about the minimality of the squared error of the solution to the continuation system under homogeneous small perturbations of its matrix is completely proved. The theoretical results are illustrated by the example of the numerical construction of Bernoulli's lemniscate.

Keywords: systems of nonlinear equations, solution continuation with respect to a parameter, best parameterization, system of solution continuation, small perturbations, squared error.